Forecasting Charge-Off Rates with a Panel Tobit Model: The Role of Uncertainty

Xin Sheng*, Rangan Gupta** and Qiang Ji***

Abstract

Based on a large panel dataset of small commercial banks in the United States, this paper employs a dynamic panel Tobit model to analyze the role of uncertainty in forecasting charge-off rates on loans for credit card (CC) and residential real estate (RRE). When compared to other standard predictors, such as house prices and unemployment rates, we find thatthe effect of uncertainty changes on charge-off rates is more pronounced. Furthermore, it is evident that including heteroskedasticity in the model specification leads to more accurate forecasts.

JEL Codes: C11, C23, C53, G21

Keywords: loan charge-offs, panel data, Tobit model, forecasting

1. Introduction

The boom and bust of the housing market of the United States (US) in 2007 led to the worst (global) financial crisis and severe recession, called the "Great Recession" since the Great Depression. The US federal government responded to the downturn in the fall of 2008 by providing extraordinary assistance, including bailouts to hundreds of financial institutions, with the estimated support for the financial sector of approximately $12.6 trillion (Atkinson et al., 2013). In an attempt to prevent similar episodes from occurring in the future, the government enacted the Dodd-Frank Wall Street Reform Act (Dodd-Frank Act) in July 2010, which is the most comprehensive financial reform since the 1930s, and aims to promote a safer and sounder financial system via the implementation of stricter regulations and supervisory practices to prevent another system-wide banking crisis. Indeed, given that banks are engaged in risky activities, these institutions will always incur some losses. However, the goal of banks and their regulators is to ensure that these losses are not so large and/or widespread that the entire banking sector is in trouble and government bailouts become necessary.

* Lord Ashcroft International Business School, Anglia Ruskin University, Chelmsford, CM1 1SQ, United Kingdom. Email: xin.sheng@anglia.ac.uk.

** Department of Economics, University of Pretoria, Pretoria, 0002, South Africa. Email: rangan.gupta@up.ac.za.

*** Corresponding author. Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190, China. Email: jqwxnjq@163.com or jqwxnjq@casipm.ac.cn.

Against this backdrop, in a recent paper Liu et al. (2019) forecast losses or charge-off rates on loans for credit card (CC) and residential real estate (RRE) of a panel of small[1] commercial banks of the US, based on different versions of the dynamic panel Tobit model. In particular, Liu et al., (2019) develops a framework to analyze a panel of censored data with a large cross-sectional and a short time-series dimension, with the latter resulting due to mergers and acquisitions, changing business models, and changes in regulatory environments. Since these banks operate in local markets, the authors include local changes in house prices and the unemployment rate as predictors in the empirical model. We aim to extend the work of Liu et al. (2019) by considering, for the first time, the role of local (state-level) uncertainty in forecasting the loan charge-off rates, in comparison to the changes in house prices and the unemployment rate. Our decision to look at uncertainty as an additional predictor emanates from the recent works of Mumtaz (2018), and Mumtaz et al., (2018), whereby these studies highlight the leading role of state-level uncertainty in driving a wide array of other state-level variables associated with the real sector and the housing market. Given this, our hypothesis is that, state-level uncertainty, since it encompasses the information content of multiple variables, is likely to outperform housing prices and unemployment rate in forecasting the loan charge-off rates of the small commercial banks. This presumption is also driven by the findings of Barth et al. (2018), who indicate the importance of several hundred predictor variables in accurately forecasting the loan charge-off rates of two very big banks (Citigroup and Wells Fargo & Company) and two very small banks (First Busey Corporation and Capital City Bank Group), based on factor and machine learning (ridge, LASSO, and elastic net) models (designed to handle large data sets).

The remainder of the paper is organized as follows: Section 2 briefly outlines the data and methodology, while Section 3 presents the results, with Section 4 concluding the paper.

2. Data and Methodology

The raw panel data of charge-offs on loans for credit card (CC) and residential real estate (RRE) of small banks are available at a quarterly frequency from the website of the Federal Reserve Bank of Chicago, with the cross-section of banks, i.e., N, =2576 and 561 respectively.[2] The

[1] The small-size banks are defined to be those banks that have total assets of less than one billion dollars.

[2] The raw data can be downloaded from https://www.chicagofed.org/banking/financial-institution-reports/commercial-bank-data. The charge-off rates are calculated as dividing charge-offs by the stock of loans. See the online appendix of Liu et al. (2019) for the details of data construction.

state-level unemployment and house price index (HPI) data are from Liu et al. (2019).[3] We use overall measures of state-level economic uncertainty (at forecasting horizons of 1-quarter-ahead) derived by Mumtaz (2018)[4] by improving the data-rich environment of Jurado et al. (2015) used to derive measures of uncertainty level for the overall US. It must be pointed out that, while the Jurado et al., (2015)-based measures of uncertainty is the average time-varying variance in the unpredictable component of the real and financial time-series, Mumtaz (2018) improves the estimates by filtering out the effects of idiosyncratic uncertainty and measurement error. Note that, the state-level uncertainty measures uses 21 macroeconomic and financial data series.[5] The rolling panel datasets range from 2001:Q2 to 2018:Q4 and are constructed with a window period of 12 quarters.[6]

Following the Bayesian approach of Liu et al. (2019), we specify a dynamic panel Tobit model with the flexible correlated random effects and heteroskedasticity as follows.

$$Y_{it} = Y_{it}^{*}\mathbb{I}\{Y_{it}^{*} \geq 0\}, i = 1,2, \dots N, T = 1,2, \dots T \tag{1}$$

$$Y_{it}^{*} = \lambda_i + \rho Y_{it-1}^{*} + \beta_1 LnHPI_{it-1} + \beta_2 UR_{it-1} + \beta_3 UC_{it-1} + u_{it}, \tag{2}$$

where $\mathbb{I}\{Y_{it}^{*} \geq 0\}$ is the indicator function which equals one if $Y_{it}^{*} \geq 0$, and zero otherwise. Y_{it} are charge-off rates, HPI_{it} represent the housing price index, UR_{it} are unemployment rates and UC_{it} are uncertainty. The model considers a heteroskedastic law of motion for Y_{it}^{*} with $u_{it} \sim iid\ N(0, \sigma_i^2), Y_{i0}^{*} \sim (u_{i*}, \sigma_{i*}^2)$. The heterogeneous parameters λ_i and σ_i^2 are modelled as the correlated random effects (CRE) and the distribution of (λ_i, σ_i^2) is represented by a mixture of normal distributions.

Subsequently, a variety of special cases of this model specification is considered. The Pooled Tobit specification treats λ_i as a homogeneous parameter and ignores the heterogeneity in λ_i, i.e., sets $\lambda_i = \lambda$ for all *i*. The Pooled Linear specification ignores the heterogeneity in λ_i and the censoring of the observations, i.e., sets $Y_{it} = Y_{it}^{*}$ and $\lambda_i = \lambda$ for all *i*. The model is linear

[3] The data along with the computer codes are available for download from: https://web.sas.upenn.edu/schorf/working-papers/.
[4] The reader is referred to the computer codes available at to obtain the measures of uncertainty: https://sites.google.com/site/hmumtaz77/research-papers?authuser=0.
[5] The variables include: real personal income and its components (social insurance, dividends, benefits and other income), overall employment, unemployment rate, real house prices, non-performing loans and net assets of banks, leading indicator, coincident indicator, all employees in health and education, financial services, government, information, leisure and hospitality, manufacturing, non-farm, professional and business services, and other services.
[6] There are 12 observations in each sample: one observation for the initialization of estimation, 10 observations for the estimation sample, and one observation for the evaluation of forecast.

since the censoring is absent. Other model specifications are also considered with the differences in their assumptions about the distributions of the heterogeneous parameters (normal versus flexible), the variances of the error terms (homoskedastic versus heteroskedastic), and the correlations between heterogeneous parameters and initial conditions (the random effects (RE) versus correlated random effects (CRE)).

3. Results

3.1. Estimates of Common Parameters

In Table 1, our results show that the effect of uncertainty changes is more pronounced than changes in unemployment rates and house prices for both RRE loans and CC loans. It is noteworthy that the size of the estimated parameters for uncertainty is much larger than that for unemployment and HPI. For example, the size of the estimated common parameter for uncertainty under the CRE Flexible, heteroskedastic specification is 0.50, which is 10 times greater than that for HPI and unemployment. This finding indicates that uncertainty plays an important role in forecasting charge-off rates. In addition, we find strong evidence that high uncertainty is associated with large charge-off rates if the heteroskedasticity is considered in the model specification. It is also evident that a fall in house prices and a rise in unemployment rates lead to an increase in the charge-off rates, which is consistent with the narrative of the great recession during the Global Financial Crisis (GFC) period.

3.2. Density Forecast Performance

Table 2 reports the average log predictive scores (LPS) and continuous-ranked probability scores (CRPS) to evaluate the performance of density forecasts for various specifications of the Tobit models.[7] Our results show that LPS (CRPS) for all heteroskedastic specifications are higher (lower) than that of homoscedastic specifications in both panels, indicating that including heteroskedasticity is important for density forecasting and it leads to more accurate forecasts for both CC and RRE samples. In contrast, accuracy differentials between the CRE versus RE and the flexible versus normal are relatively small. In viewing of these results, our subsequent full sample analysis focuses on the heteroskedastic versus homoskedastic specifications with flexible CRE.

[7] For the model specification M, we report the average log predictive scores $LPS_h(M) = \frac{1}{N}\sum_{i=1}^{N} ln(\mathbb{I}\{Y_{iT+h} = 0\}\mathbb{P}_{Y_{1:N,0:T}}^{Y_{iT+h}}\{Y_{iT+h} = 0|M\} + \mathbb{I}\{Y_{iT+h} > 0\}\, p\left(Y_{iT+h}|Y_{1:N,0:T}\right)$ and the continuous ranked probability scores $\mathrm{CRPS_h(M)} = \frac{1}{N}\sum_{i=1}^{N}\int_0^\infty (F_{Y_{1:N,0:T}}^{Y_{iT+h}}(Y_i|M) - \mathbb{I}\{Y_{iT+h} \leq Y_i\})^2\, dy$ as in Liu et al. (2019).

Table 1. Estimates of Common Parameters

Panel A: Credit Card (CC) samples

Model Specifications	AR(1)	HPI	Unemployment	Uncertainty
CRE, Flexible,	0.42	-0.05	0.05	0.50
Heterosk	[0.36, 0.48]	[-0.09, -0.01]	[0.00, 0.12]	[0.39, 0.59]
CRE, Normal,	0.34	-0.06	0.04	0.56
Heterosk	[0.31, 0.37]	[-0.09, -0.02]	[0.01, 0.07]	[0.47, 0.63]
RE, Flexible,	0.46	-0.03	0.07	0.40
Heterosk	[0.41, 0.49]	[-0.09, 0.01]	[0.01, 0.14]	[0.19, 0.58]
RE, Normal,	0.38	-0.03	0.03	0.46
Heterosk	[0.34, 0.44]	[-0.05, -0.00]	[0.00, 0.09]	[0.39, 0.59]
CRE, Flexible,	0.76	-0.41	0.27	0.11
Homosk	[0.71, 0.80]	[-0.63, -0.18]	[0.11, 0.44]	[-0.44, 0.61]
CRE, Normal,	0.44	-0.36	0.31	-0.26
Homosk	[0.42, 0.46]	[-0.54, -0.19]	[0.15, 0.49]	[-0.84, 0.19]
RE, Flexible,	0.83	-0.22	0.18	0.38
Homosk	[0.81, 0.86]	[-0.39, -0.06]	[0.03, 0.31]	[-0.10, 0.68]
RE, Normal,	0.52	-0.25	0.28	0.15
Homosk	[0.49, 0.54]	[-0.42, -0.06]	[0.10, 0.43]	[-0.36, 0.96]
Pooled Tobit,	0.86	-0.24	0.21	0.48
Homosk	[0.84, 0.88]	[-0.38, -0.08]	[0.06, 0.42]	[0.25, 0.70]
Pooled OLS,	0.80	-0.14	0.05	0.51
Homosk	[0.78, 0.81]	[-0.25, -0.03]	[-0.10, 0.19]	[0.32, 0.71]

Panel B: Residential Real Estate (RRE) samples

Model Specificaitons	AR(1)	HPI	Unemployment	Uncertainty
CRE, Flexible,	0.21	-0.01	0.01	0.03
Heterosk	[0.17, 0.22]	[-0.03, -0.01]	[0.01, 0.01]	[0.02, 0.04]
CRE, Normal,	0.08	-0.02	0.01	0.02
Heterosk	[0.07, 0.09]	[-0.03, -0.01]	[0.01, 0.01]	[0.01, 0.03]
RE, Flexible,	0.01	-0.02	0.01	0.02
Heterosk	[0.01, 0.01]	[-0.03, -0.01]	[0.01, 0.02]	[0.00, 0.03]
RE, Normal,	0.11	-0.01	0.01	0.02
Heterosk	[0.09, 0.12]	[-0.03, -0.01]	[0.01, 0.01]	[0.01, 0.04]
CRE, Flexible,	0.57	-0.16	0.20	-0.53
Homosk	[0.42, 0.69]	[-0.18, -0.14]	[0.17, 0.24]	[-0.83, -0.23]
CRE, Normal,	0.39	-0.15	0.22	-1.08
Homosk	[0.37, 0.42]	[-0.18, -0.12]	[0.18, 0.25]	[-1.22, -0.79]
RE, Flexible,	0.63	-0.17	0.18	-0.32
Homosk	[0.47, 0.74]	[-0.20, -0.14]	[0.16, 0.21]	[-0.48, -0.20]
RE, Normal,	0.43	-0.18	0.21	-0.42
Homosk	[0.40, 0.47]	[-0.20, -0.15]	[0.17, 0.23]	[-0.49, -0.34]
Pooled Tobit,	0.84	-0.14	0.18	-0.04
Homosk	[0.75, 0.88]	[-0.16, -0.12]	[0.15, 0.20]	[-0.07, -0.01]
Pooled OLS,	0.53	-0.07	0.06	-0.01
Homosk	[0.52, 0.54]	[-0.08, -0.06]	[0.05, 0.07]	[-0.02, 0.00]

Notes: The table contains posterior means and 90% credible intervals in brackets. The baseline estimation sample covers the Global Financial Crisis period (GFC) and ranges from 2007Q2 (t = 0) to 2009Q4 (t = T = 10). Heterosk refers to the heteroskedastic specification, which allows for heteroskedastic innovations in the model, while Homosk refers to the homoskedastic specification.

Table 2. Density Forecast Performance

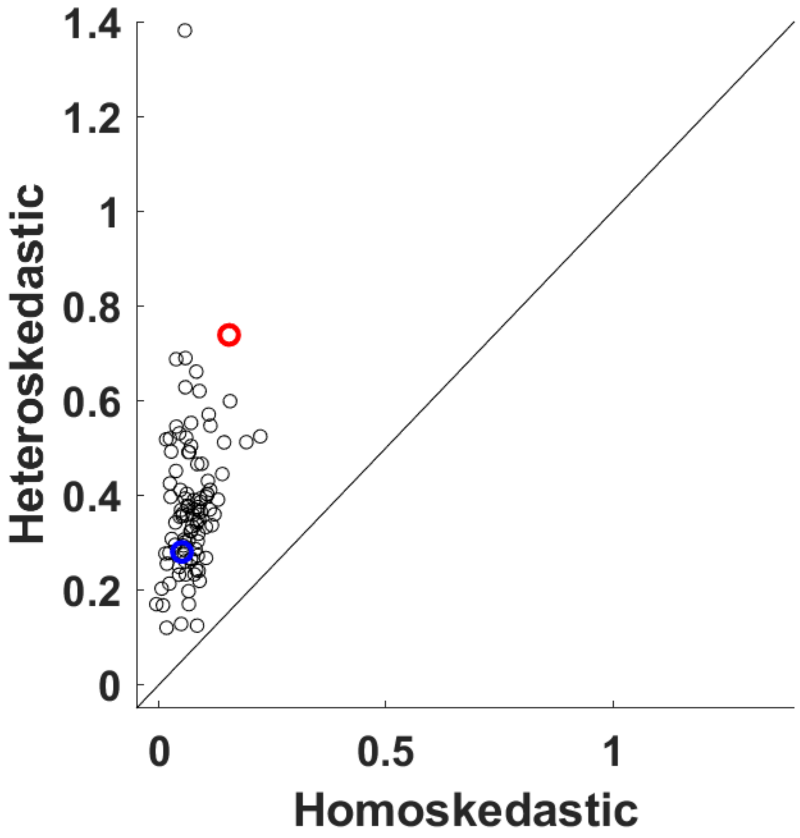
Model Specificaitons	LPS CC samples	CRPS CC samples	LPS RRE samples	CRPS RRE samples
CRE, Flexible, Heterosk	-0.55	0.25	-1.92	1.91
CRE, Normal, Heterosk	-0.57	0.26	-1.91	1.90
RE, Flexible, Heterosk	-0.59	0.26	-1.95	2.01
RE, Normal, Heterosk	-0.56	0.25	-1.91	1.92
CRE, Flexible, Homosk	-0.78	0.28	-2.50	2.57
CRE, Normal, Homosk	-0.77	0.28	-2.48	2.37
RE, Flexible, Homosk	-0.79	0.29	-2.65	2.68
RE, Normal, Homosk	-0.77	0.27	-2.56	2.42
Pooled Tobit, Homosk	-0.83	0.30	-2.65	2.66
Pooled OLS, Homosk	-1.59	0.38	-3.04	2.84

Note: Table 2 reports LPS (larger values indicate better results) and CRPS (smaller values indicate better results). The baseline estimation sample ranges from 2007Q2 (t=0) to 2009Q4 (t=T=10). We forecast 2010Q1 observations.

Figure 1 shows the LPS comparisons for all 111 samples ranging from 2007 to 2016. It compares predictive scores from the heteroskedastic versus homoskedastic specifications with flexible CRE. The figure shows that the results for the baseline samples (reported in Table 2) are qualitatively representative and including heteroskedasticity is very important for improving the density forecasts across all samples.[8]

[8] We further examined the forecast performances of our models following the methods of Liu et al. (2019) using the uncertainty data of Mumtaz (2018) at longer horizons, i.e., 2-, 3-, and 4-quarter-ahead. The results provide further evidence that the heteroskedastic specifications deliver more accurate forecasts, with these results are not reported but are available upon request.

Figure 1. Log Predictive Density Scores: All Samples



4.

5. **Note:** The figure provides the comparisons of log predictive scores (LPS) from the heteroskedastic versus homoskedastic specifications with flexible CRE, which are depicted as LPS differentials relative to the pooled Tobit specification. The blue and red circles correspond to baseline RRE and CC samples.

4. Conclusions

The prediction of charge-off rates is interesting from a regulator's perspective because charge-offs generate losses on loan portfolios. If these charge-offs are large, the bank may be entering a period of distress and require additional capital. Given this, in this paper, we use a dynamic panel Tobit model to analyse the importance of state-level uncertainty, relative to standard predictors, i.e., changes in house prices and unemployment rates, used recently by Liu et al. (2019). We find that uncertainty plays an important role in forecasting charge-off rates and the effect of uncertainty changes is more pronounced than changes in unemployment rates and house prices for both credit card (CC) and residential real estate (RRE) loans. Furthermore, it is evident that considering heteroskedasticity in the model specification leads to more accurate forecasts. Understandably, our results have important policy implications, and suggests that policy authorities need to reduce uncertainty via the adoption of expansionary monetary and fiscal policies to prevent the possibility of a system-wide banking crisis.

References

Atkinson, T., Luttrell, D., and Rosenblum, H. (2013). Assessing the costs and consequences of the 2007–09 financial crisis and its aftermath. Economic Letter, Federal Reserve Bank of Dallas, 8(7), 1-4.

Barth, J.R., Han, S., Joo, S., Lee, K.B., Maglic, S., and Shen, X. (2018). Forecasting net charge-off rates of banks: What model works best? Quantative Finance and Economics, 2(3): 554–589.

Jurado, K., Ludvigson, S.C., and Ng, S. (2015). Measuring Uncertainty. The American Economic Review, 105(3), 1177-1215.

Liu, L., Moon, H.R. and Schorfheide, F. (2019). Forecasting with a Panel Tobit Model. NBER Working Paper No. 26569.

Mumtaz, H. (2018). Does uncertainty affect real activity? Evidence from state-level data. Economics Letters, 167, 127-130.

Mumtaz, H., Sunder-Plassmann, L., and Theophilopoulou, A. (2018). The State-Level Impact of Uncertainty Shocks. Journal of Money, Credit and Banking, 50(8), 1879-1899.