Forecasting with Dynamic Panel Data Models

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

[1]University of Pennsylvania

[2]University of Southern California

[3]University of Pennsylvania, CEPR, and NBER

Progress Report: March 2014

Introduction

- Goal: develop methods to generate forecasts from a panel data model of the form

$$Y_{it} = \beta' X_{it} + \lambda_i' W_t + U_{it}, \quad t = 1, \ldots, T; i = 1, \ldots, N. \tag{1}$$

- Here X_{it} may contain lags of Y_{it}.
- We consider a large N and small T environment.
- Empirical context:
  - Monitoring of banking sector; stress tests: forecasts of capital-asset ratios, charge-offs, etc.
  - Why large N? Track individual banks or bank holding companies.
  - Why small T? Mergers; changes in regulatory environments; lack of variation in Y's and X's in normal times.

Dynamic Panel Model

- Consider a simple dynamic panel data model:

$$Y_{it} = \rho Y_{it-1} + \lambda_i + U_{it}, \tag{2}$$

  where $U_{it} \sim iid\,(0, 1)$ and λ_i represents the unobserved individual heterogeneity.

- For a given ρ, the optimal forecast of Y_{iT+1} at time T is

$$\mathbb{E}(Y_{iT+1}|Y, \rho) = \rho Y_{iT} + \mathbb{E}(\lambda_i|Y, \rho).$$

- In the dynamic panel literature, the focus has been to find a consistent estimate of ρ in the presence of the incidental parameters λ_i to avoid the incidental parameter problems.
- Our interest is to have a good forecast that requires to use "good" estimates of both ρ and λ_i's with small T panel.

Introduction

- Selection bias poses a challenge for short time span panel data:
- the usual panel data estimate of the fixed effects (QMLE) tends to over-predict (under-predict) the future capital-asset ratios for the banks with high (low) current capital-asset ratio.

Monte Carlo Illustration

$$\text{Model: } Y_{it} = \rho Y_{i,t-1} + \lambda_i + U_{it}, \quad t = 1, \dots, T; i = 1, \dots, N.$$

$$\text{Design: } T = 3,\ N = 1,000,\ \rho = 0.8,\ \lambda_i \sim U[0,1], \\ Y_{i0} \sim N\big(\lambda_i/(1-\rho), 1/(1-\rho^2)\big).$$

		Bottom 20		Middle 20		Top 20		All	
$\hat{\rho}$	$\hat{\lambda}_i$	MSE	Med	MSE	Med	MSE	Med	MSE	Med
				No Shrinkage					
GMM (AB)	QMLE		0.97		0.02		-0.99		0.00
GMM (BB)	QMLE		0.97		0.05		-0.99		0.00

- Forecast errors:

$$\left(Y_{i,T+1}^{(s)} - \hat{Y}_{i,T+1}^{c,(s)}(\hat{\theta}_{0:T,<-i>}) \right)^2. \tag{3}$$

- Relatively large selection bias for top and bottom groups.

Efron (2011)'s Motivation

- Our paper was inspired by work by Efron (2011). Consider the following question:
  - Want to predict the US Masters golf tournament final scores (the average score after four rounds) after the first round.
  - The first round score, Y_i, consists of true skill, λ_i, and (unpredicable) luck, U_i.
  - If the scores Y_i are independent across i, the natural estimator of λ_i appears to be Y_i, the first round score.
  - Question: "Can we estimate λ_i more precisely, by using the other players' scores of the first round, $(Y_1, ..., Y_N)$?"
- This question arises more generally in dynamic panel data models.

Introduction

- We employ an empirical Bayes approach to combine cross sectional and time series information together and thus obtain "better" forecasts for banks with extreme capital asset ratios.

- We estimate $\mathbb{E}(\lambda_i|Y,\rho)$ as the posterior mean of λ_i.

Related Literature

- Tweedie's formula and its use: e.g., Robbins (1951), Brown (2008), Brown and Greenshtein (2009), Efron (2011), Gu and Koenker (2013).

- Consistent estimation of ρ in dynamic panel data models with fixed effects when T is small:

  - IV/GMM: e.g. Anderson and Hsiao (1982), Arellano and Bond (1991), Arellano and Bond (1995), Blundell and Bond (1998), and Alvarez and Arellano (2003).

  - Bayesian: e.g. Lancaster (2002) - (informational) orthogonal parameterization.

- Bayesian inference in panel data models

- Correlated random effect models

PS: Maurice Tweedie = British medical physicist and statistician, born in 1919 and died in 1996.

Road Map

1. Introduction
2. Decision-theoretic Considerations
3. Two Empirical Bayes Predictors:
    - Parametric Family of Distributions for λ_i
    - Nonparametric $p(\lambda_i)$ and Tweedie's Formula
4. Simulations
5. Empirical Application
6. Conclusion.

Decision-Theoretic Considerations

- Simple model $Y_{it} = \lambda_i + \rho Y_{i,t-1} + U_{it}$.
- $\hat{Y}^{T+1} = [\hat{Y}_{1,T+1}, \dots, \hat{Y}_{N,T+1}]'$ is vector of forecasts.
- Compound L_2 loss function:

$$L_N(\hat{Y}^{T+1}, Y^{T+1}) = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{Y}^{T+1} - Y^{T+1} \right)^2. \tag{4}$$

- Expected compound loss:

$$\begin{aligned} & \mathbb{E}_{(\rho,\lambda)} L_N \left(\hat{Y}^{T+1}, Y^{T+1} \right) \\ = \quad & \mathbb{E}_{(\rho,\lambda)} \left[\frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_{(\rho,\lambda)} \left[\left(\hat{Y}_{i,T+1} - Y_{i,T+1} \right)^2 \middle| Y^{0:T} \right] \right] \\ = \quad & \mathbb{E}_{(\rho,\lambda)} \left[\frac{1}{N} \sum_{i=1}^{N} \left(\hat{Y}_{i,T+1} - \lambda_i - \rho Y_{iT} \right)^2 \right] + 1 \end{aligned} \tag{5}$$

Decision-Theoretic Considerations

- Consider the class of additively separable forecasts $\hat{Y}_{i,T+1} = \hat{\lambda}_i + \hat{\rho} Y_{iT}$ where $\hat{\lambda}_i$ and $\hat{\rho}$ are estimators of λ_i and ρ.
- Decision space:

$$\mathcal{D} = \left\{ \left(\hat{\lambda}_1 + \hat{\rho} Y_{1T}, ..., \hat{\lambda}_N + \hat{\rho} Y_{NT} \right) \;\middle|\; (\hat{\lambda}, \hat{\rho}) \in \mathcal{F}_{0:T} \right\}. \tag{6}$$

- Find asymptotically optimal forecast in the class $\mathcal{D}$ that minimizes the expected compound loss (as $N \longrightarrow \infty$):

$$\begin{aligned} & \mathbb{E}_{(\rho,\lambda)} \left[L_N(\hat{Y}^{T+1}_{opt}, Y^{T+1}) \right] \\ \leq & \inf_{\hat{Y}_{i,T+1} \in \mathcal{D}} \mathbb{E}_{(\rho,\lambda)} \left[L_N(\hat{Y}^{T+1}, Y^{T+1}) \right] + o(1). \end{aligned} \tag{7}$$

Decision-Theoretic Considerations

- Suppose ρ is known...
- Then forecast simplifies to $\hat{Y}_{i,T+1} = \hat{\lambda}_i + \rho Y_{iT}$.
- Finding an optimal forecast is equivalent to constructing an optimal estimator of λ:

$$\inf_{\hat{\lambda}} \mathbb{E}_{\lambda} \left[\frac{1}{N} \sum_{i=1}^{N} (\hat{\lambda}_i - \lambda_i)^2 \right]. \tag{8}$$

- This estimator is constructed from $Z_{it} = Z_{it}(\rho) = Y_{it} - \rho Y_{i,t-1}$.
- For $T = 1$ see Robbins (1951, 1956).

Decision-Theoretic Considerations

- Suppose $T = 1$ and $\hat{\lambda}_i = g(Z_{i1})$.
- Expected compound loss becomes integrated risk with empirical distribution of λ as prior:

$$\begin{aligned}
\mathbb{E}_\lambda \left[\frac{1}{N} \sum_{i=1}^{N} (\hat{\lambda}_i - \lambda_i)^2 \right] &= \frac{1}{N} \sum_{i=1}^{N} \int \big(g(z) - \lambda_i \big)^2 \phi(z - \lambda_i) dz \\
&= \int \left[\int \big(g(z) - \lambda \big)^2 \phi(z - \lambda) dz \right] dG_N(\lambda_i) \\
&= \mathbb{E}_{G_N} \left[(g(Z) - \lambda_i)^2 \right].
\end{aligned}$$

- Optimal estimator:

$$g^*_{G_N}(z) = \frac{\int \lambda_i \phi(z - \lambda_i) dG_N(\lambda_i)}{\int \phi(z - \lambda_i) dG_N(\lambda_i)}. \qquad (9)$$

- To implement this estimator we need to generate an estimate of $G_N(\lambda_i)$ based on cross-sectional information.

Decision-Theoretic Considerations

- Idea: Approximate $g^*_{G_N}(z)$ by $\hat{g}^*(z)$ such that

$$\mathbb{E}_{G_N}\left[(\hat{g}^*(z)-\lambda_i)^2\right] \leq \mathbb{E}_{G_N}\left[(g^*_{G_N}(z)-\lambda_i)^2\right] + o(1). \tag{10}$$

- There are some results in the statistics literature, e.g., Zhang (2003), Brown and Greenshtein (2009), and Jiang and Zhang (2009).
- TO DO: extend THEORETICAL results to panel data application.
- FOR NOW: we consider two different implementations of the basic idea:
  - Treat G_N parametrically (indexed by finite-dimensional hyperparameter):

$$\lambda_i \sim N\big(0, \omega^2\big) \quad \text{or} \quad \lambda_i \sim N\big(\phi_0 + \phi_1 Y_{i0}, \omega^2\big).$$

  - Treat G_N nonparametrically: use some general density $p(\lambda_i | Y_{i0})$.

Use cross-sectional information to estimate relevant features of G_N.

Next Steps

- To fix ideas, we will consider the simple model:

$$Y_{it} = \rho Y_{i,t-1} + \lambda_i + U_{it}, \quad U_{it}|(Y_{i,t-1}, \lambda_i) \sim N(0,1), \tag{11}$$

- For now we will assume that λ_i is independent of Y_{i0}.
- Step 1: parametric Bayesian analysis with family of priors $\lambda_i | Y_{i0} \sim N(0, \omega^2)$.
- Step 2: treat $p(\lambda)$ nonparametrically – realizing that the Bayes estimator of λ_i depends on $p(\lambda_i)$ only through the marginal distribution of $Z_i(\rho) = \frac{1}{T}\sum_{t=1}^{T}(Y_{it} - \rho Y_{i,t-1})$. Tweedie's Formula!.

Step 1: Parametric Analysis

- Y^t is $N \times 1$; Y is $N \times T$.
- X is $N \times T$, λ is $N \times 1$.
- Likelihood function:

$$
\begin{aligned}
& p(\rho, \lambda | Y^{0:T}) \\
\propto \quad & \prod_{i=1}^{N} p(Y_i^{1:T} | \rho, \lambda_i, Y_{i0}) p(\lambda_i) \\
\propto \quad & \exp\left\{ -\frac{1}{2} \left(tr\left[(Y - X\rho - \lambda\iota_T')(Y - X\rho - \lambda\iota_T')' \right] + \omega^{-2}\lambda'\lambda \right) \right\}
\end{aligned} \tag{12}
$$

Parametric Analysis: Posterior Distribution

- Posterior of $\lambda|\rho$:

$$\lambda_i|(\rho, Y^{0:T}) \sim N\big(\mu_{\lambda_i}(\rho), \sigma_\lambda^2\big), \; i = 1, \ldots, N. \quad (13)$$

where

$$\begin{aligned} \mu_\lambda(\rho) &= \sigma_\lambda^2(Y - X\rho)\iota_T \\ \sigma_\lambda^2 &= (T + \omega^{-2})^{-1}. \end{aligned}$$

Parametric Analysis: Forecasting

- The one-step-ahead predictive density for $Y_{i,T+1}$ is given by

$$p(Y_{i,T+1}|Y^{0:T},\rho) = \int p(Y_{i,T+1}|Y_{iT},\rho,\lambda_i)p(\lambda_i|\rho,Y^{0:T})d\lambda_i. \tag{14}$$

- The mean of this predictive density can be written as

$$\mathbb{E}[Y_{i,T+1}|Y^{0:T},\rho] = \rho Y_{iT} + \mathbb{E}[\lambda_i|Y^{0:T},\rho]. \tag{15}$$

- Define the MLE of λ_i conditional on ρ as

$$Z_i(\rho) = \frac{1}{T}\sum_{i=1}^{T}(Y_{iT} - \rho Y_{i,T-1}). \tag{16}$$

- Then the posterior mean of λ_i can be decomposed as follows:

$$\mathbb{E}[\lambda_i|Y^{0:T},\rho] = \underbrace{Z_i(\rho)}_{\text{MLE}} \underbrace{- \frac{1}{(1+\omega^2 T)}Z_i(\rho)}_{\text{Bayes Correction}}. \tag{17}$$

Parametric Analysis: Estimate Hyperparameters

- We will estimate the common parameters, e.g. ρ, jointly with the hyperparameter ω that serves as an index for $p(\lambda)$ using the cross-sectional information:

$$(\hat{\omega}, \hat{\rho}) = \text{argmax } \ln p(Y^{1:T}|Y^0, \rho, \omega), \tag{18}$$

where

$$\ln p(Y^{1:T}|Y^0, \rho, \omega) = \ln \int p(Y^{1:T}|Y^0, \rho, \lambda, \omega) p(\rho, \lambda|\omega, Y^0) d(\rho, \lambda).$$

- The posterior mean predictor with data-driven hyperparameter choice becomes (now making the dependence on ω explicit):

$$\mathbb{E}[\lambda_i|Y^{0:T}, \hat{\rho}, \hat{\omega}] = Z_i(\hat{\rho}) - \frac{1}{(1+\hat{\omega}^2 T)} Z_i(\hat{\rho}). \tag{19}$$

- Note: we could replace $\hat{\rho}$ by the posterior mean $\mathbb{E}[\rho|Y^{0:T}, \hat{\omega}]$.
- Generalization: condition on Y_{i0}: $\lambda_i \sim N(\phi_0 + \phi_1 Y_{i0}, \omega^2)$.

Step 2: Tweedie's Formula

- Replace $p(\lambda|\omega)$ by more general family of distributions $p(\lambda)$.
- Recall $Z_i(\rho) = \frac{1}{T}(Y_i - X_i\rho)\iota_T$.
- Our simple model implies that $Z_i(\rho)|\rho \sim N(\lambda_i, 1/T)$.
- Under our distributional assumptions we obtain

$$q(Z_i(\rho)|\lambda_i) = (2\pi/T)^{-1/2} \exp\left\{-\frac{T}{2}(Z_i(\rho) - \lambda_i)^2\right\}. \quad (20)$$

- Write the Gaussian density using the following exponential-family representation:

$$q(Z_i(\rho)|\lambda_i) = \exp\left\{\lambda_i T Z_i(\rho) - \psi(\lambda_i)\right\} q_0(Z_i(\rho)). \quad (21)$$

where

$$\psi(\lambda_i) = \frac{T}{2}\lambda_i^2 \quad \text{and} \quad q_0(Z_i(\rho)) = (2\pi/T)^{-1/2} \exp\left\{-\frac{T}{2}Z_i^2(\rho)\right\}$$

Tweedie's Formula

- Posterior of λ conditional on ρ:

$$p(\lambda|Y^{0:T}, \rho) = \prod_{i=1}^{N} \frac{\exp\{\lambda_i T Z_i(\rho) - \psi(\lambda_i)\} p(\lambda_i)}{\int \exp\{\lambda_i T Z_i(\rho) - \psi(\lambda_i)\} p(\lambda_i) d\lambda_i} \qquad (22)$$

- Now focus on the posterior of λ_i and write

$$p(\lambda_i|Y^{0:T}, \rho) = \exp\{\lambda_i T Z_i(\rho) - \chi(Z_i(\rho))\} p(\lambda_i) \exp\{-\psi(\lambda_i)\}.$$

where

$$\chi(Z_i) = \ln \int \exp\{\lambda_i T Z_i(\rho) - \psi(\lambda_i)\} p(\lambda_i) d\lambda_i.$$

- Since the posterior density integrates to one, we obtain

$$\begin{aligned} 0 &= \frac{\partial}{\partial Z_i} \int \exp\{\lambda_i T Z_i - \chi(Z_i)\} p(\lambda_i) \exp\{-\psi(\lambda_i)\} d\lambda_i \\ &= T \int \lambda_i p(\lambda_i|Y^{0:T}, \rho) d\lambda_i - \chi'(Z_i). \end{aligned}$$

Tweedie's Formula

- Tweedie's formula:

$$\mathbb{E}[\lambda_i | Y^{0:T}, \rho] = \frac{1}{T}\chi'(Z_i(\rho)). \tag{23}$$

- Using the definition of $q_0(Z_i)$ we can write

$$\chi(Z_i) = \ln \int q(Z_i|\lambda_i) p(\lambda_i) d\lambda_i + \frac{1}{2}\ln(2\pi/T) + \frac{T}{2} Z_i^2.$$

- This leads to

$$\mathbb{E}[\lambda_i | Y^{0:T}, \rho] = \underbrace{Z_i(\rho)}_{\text{MLE}} + \underbrace{\frac{1}{T}\frac{\partial \ln q(Z_i)}{\partial Z_i}}_{\text{Bayes Correction}} \Big|_{Z_i = Z_i(\rho)}. \tag{24}$$

- NOTE: we only need to estimate the marginal density of Z_i. We do not need to estimate $p(\lambda_i)$!
- Generalization: condition on Y_{i0}.

Tweedie's Formula: Implementation

- First we find a consistent estimate of ρ, say $\hat{\rho}$.
- Second, compute QMLE for λ:

$$Z_i(\hat{\rho}) = \frac{1}{T}\sum_{t=1}^{T}(Y_{it} - \hat{\rho}Y_{it-1}). \qquad (25)$$

- Third, nonparametric correction based on Tweedie's formula:

$$\hat{\lambda}_i = Z_i(\hat{\rho}) + \frac{1}{T}\frac{1}{\hat{q}(Z_i(\hat{\rho}))}\frac{\partial \hat{q}(z)}{\partial z}|_{z=Z_i(\hat{\rho})}, \qquad (26)$$

where $\hat{p}(z)$ is a nonparametric density estimate of Z_i.

Tweedie's Formula: Implementation – $\hat{\rho}$

Arellano & Bover (95) ("GMM (AB)"):

- Moment conditions based on Orthogonal Forward Demeaning: $\mathbb{E}\left(W_{it}' U_{it}^*\right) = 0$, where $W_{it} = \left(Y_{i0}, \cdots, Y_{i,t-1}\right)$, $U_{it}^* = \sqrt{\frac{T-t}{T-t+1}} \left[U_{it} - \frac{U_{i,t+1} + \cdots + U_{iT}}{T-t}\right]$, $t = 1, \cdots, T-1$.
- Under homoskedasticity, one-step estimator as it's already an asymptotically efficient GMM estimator.
- Better finite sample properties than Arellano and Bond (91) estimator based on the first difference when ρ is close to 1.

Blundell & Bond (98) ("GMM (BB)"):

- Moment conditions:
  $\mathbb{E}\left(W_{it}' \Delta U_{it}\right) = 0$
  $\mathbb{E}\left(\Delta Y_{i,t-1}\left(\lambda_i + U_{it}\right)\right) = 0$
  where $W_{it} = \left(Y_{i0}, \cdots, Y_{i,t-2}\right)$, $t = 2, \cdots, T$.
- Two-step estimator.
- Need $\mathbb{E}\left(\Delta Y_{i,t-1}\lambda_i\right) = 0$ or $\mathbb{E}\left(\lambda_i\left(Y_{i0} - \frac{\lambda_i}{1-\rho}\right)\right) = 0$. Stationary initial condition.
- Better dealing with weak IV problem when ρ is close to 1 when the initial condition is stationary.

Tweedie's Formula: Implementation – $\hat{q}(Z)$

- Lindsey's method:

$$\hat{q}_{Lindsey}(z) = \exp\left\{ \sum_{j=0}^{J} \gamma_j z^j \right\}$$

Estimate γ_i's by Poisson regression.

- Kernel smoothing:

$$\hat{q}_{kernel}(z) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{Z_i - z}{h} \right)$$

- Note: in the application we use densities that are conditional on Y_{i0}.

A Small Simulation Experiment

- Model: $Y_{it} = \lambda_i + \rho Y_{i,t-1} + U_{it}$ where $U_{it} \sim iidN(0,1)$.
- $\lambda_i | Y_{i0} \sim iidU[0,1]$.
- Y_{i0} distribution:

$$\text{Design 1} \quad : \quad Y_{i0}|(\lambda_i, \rho) \sim N\left(\frac{\lambda_i}{1-\rho}, \ \frac{1}{1-\rho^2}\right). \tag{27}$$

$$\text{Design 2} \quad : \quad Y_{i0}|(\lambda_i, \rho) \sim N\left(0, \ 0.1^2\right). \tag{28}$$

- Autoregressive coefficient: $\rho = 0.8$.
- $N = 1,000, \ T = 3$.

Simulation: Performance Statistics

- Forecast errors:

$$\left(Y_{i,T+1}^{(s)} - \hat{Y}_{i,T+1}^{c,(s)}(\hat{\theta}_{0:T,<-i>}) \right)^2. \tag{29}$$

- We consider four different groups of observations:
  - *Bottom*: 20 smallest Y_{iT}'s (out of 1,000)
  - *Middle*: 20 Y_{iT}'s around the median
  - *Top*: 20 largest Y_{iT}'s
  - *All*: all Y_{iT}'s

- We compute mean-squared forecast errors and median forecast errors.

Monte Carlo: Design 1

$$\text{Model: } Y_{it} = \rho Y_{i,t-1} + \lambda_i + U_{it}, \quad t = 1, \dots, T; i = 1, \dots, N.$$

$$\text{Design: } T = 3,\ N = 1,000,\ \rho = 0.8,\ \lambda_i \sim U[0,1],$$
$$Y_{i0} \sim N\big(\lambda_i/(1-\rho), 1/(1-\rho^2)\big).$$

		Bottom 20		Middle 20		Top 20		All	
$\hat{\rho}$	$\hat{\lambda}_i$	MSE	Med	MSE	Med	MSE	Med	MSE	Med
				No Shrinkage					
GMM (AB)	QMLE	2.12	0.97	1.21	0.02	2.22	-0.99	1.34	0.00
GMM (BB)	QMLE	2.14	0.97	1.19	0.05	2.20	-0.99	1.33	0.00

- AB and BB estimators perform very similarly.
- Relatively large selection bias for top and bottom groups.

Monte Carlo: Design 1

Model: $Y_{it} = \rho Y_{i,t-1} + \lambda_i + U_{it}, \quad t = 1, \dots, T; i = 1, \dots, N.$

Design: $T = 3$, $N = 1,000$, $\rho = 0.8$, $\lambda_i \sim U[0,1]$,
$Y_{i0} \sim N\big(\lambda_i/(1-\rho), 1/(1-\rho^2)\big)$.

		Bottom 20		Middle 20		Top 20		All	
$\hat{\rho}$	$\hat{\lambda}_i$	MSE	Med	MSE	Med	MSE	Med	MSE	Med
No Shrinkage									
GMM (AB)	QMLE	2.12	0.97	1.21	0.02	2.22	-0.99	1.34	0.00
Tweedie's Formula									
GMM (AB)	Lindsey	1.28	-0.06	1.05	0.05	1.35	-0.08	1.10	0.00
GMM (AB)	Kernel	1.33	0.00	1.04	0.02	1.45	-0.05	1.10	0.00

- Tweedie's formula is able to correct selection bias.

Monte Carlo: Design 1

$$\text{Model: } Y_{it} = \rho Y_{i,t-1} + \lambda_i + U_{it}, \quad t = 1, \dots, T; i = 1, \dots, N.$$

$$\text{Design: } T = 3,\ N = 1,000,\ \rho = 0.8,\ \lambda_i \sim U[0,1], \\ Y_{i0} \sim N\big(\lambda_i/(1-\rho), 1/(1-\rho^2)\big).$$

		Bottom 20		Middle 20		Top 20		All	
$\hat{\rho}$	$\hat{\lambda}_i$	MSE	Med	MSE	Med	MSE	Med	MSE	Med
No Shrinkage									
GMM (AB)	QMLE	2.12	0.97	1.21	0.02	2.22	-0.99	1.34	0.00
Tweedie's Formula									
GMM (AB)	Lindsey	1.28	-0.06	1.05	0.05	1.35	-0.08	1.10	0.00
GMM (AB)	Kernel	1.33	0.00	1.04	0.02	1.45	-0.05	1.10	0.00
Empirical Bayes Forecast with Parametric Model									
Max of Marg. LH		1.05	0.07	1.01	0.04	1.12	-0.11	1.05	0.00

- The parametric Bayes model works even better.

Monte Carlo: Design 2

Model: $Y_{it} = \rho Y_{i,t-1} + \lambda_i + U_{it}, \quad t = 1, \dots, T; i = 1, \dots, N.$

Design: $T = 3$, $N = 1,000$, $\rho = 0.8$, $\lambda_i \sim U[0,1]$, $Y_{i0} \sim N\left(0, 0.1^2\right)$.

		Bottom 20		Middle 20		Top 20		All	
$\hat{\rho}$	$\hat{\lambda}_i$	MSE	Med	MSE	Med	MSE	Med	MSE	Med
No Shrinkage									
GMM (AB)	QMLE	4.75	0.70	1.97	0.24	13.57	-0.10	3.13	0.25
GMM (BB)	QMLE	3.97	1.68	1.11	-0.14	5.28	-2.04	1.68	-0.18

- Under this design the GMM(BB) estimator is preferable.
- Relatively large selection bias for top and bottom groups.

Monte Carlo: Design 2

Model: $Y_{it} = \rho Y_{i,t-1} + \lambda_i + U_{it}, \quad t = 1, \ldots, T; i = 1, \ldots, N.$

Design: $T = 3,\ N = 1,000,\ \rho = 0.8,\ \lambda_i \sim U[0,1],\ Y_{i0} \sim N\big(0, 0.1^2\big).$

		Bottom 20		Middle 20		Top 20		All	
$\hat{\rho}$	$\hat{\lambda}_i$	MSE	Med	MSE	Med	MSE	Med	MSE	Med
		No Shrinkage							
GMM (BB)	QMLE	3.97	1.68	1.11	-0.14	5.28	-2.04	1.68	-0.18
		Tweedie's Formula							
GMM (BB)	Lindsey	1.55	0.25	1.12	-0.14	1.63	-0.53	1.14	-0.18
GMM (BB)	Kernel	1.62	0.11	1.14	-0.13	1.91	-0.41	1.20	-0.18

- Tweedie's formula is able to correct selection bias.

Monte Carlo: Design 2

Model: $Y_{it} = \rho Y_{i,t-1} + \lambda_i + U_{it}, \quad t = 1, \dots, T; i = 1, \dots, N.$

Design: $T = 3$, $N = 1{,}000$, $\rho = 0.8$, $\lambda_i \sim U[0,1]$, $Y_{i0} \sim N(0, 0.1^2)$.

		Bottom 20		Middle 20		Top 20		All	
$\hat{\rho}$	$\hat{\lambda}_i$	MSE	Med	MSE	Med	MSE	Med	MSE	Med
No Shrinkage									
GMM (BB)	QMLE	3.97	1.68	1.11	-0.14	5.28	-2.04	1.68	-0.18
Tweedie's Formula									
GMM (BB)	Lindsey	1.55	0.25	1.12	-0.14	1.63	-0.53	1.14	-0.18
GMM (BB)	Kernel	1.62	0.11	1.14	-0.13	1.91	-0.41	1.20	-0.18
Empirical Bayes Forecast with Parametric Model									
Max of Marg. LH		1.09	0.02	1.08	0.04	1.13	-0.03	1.08	0.00

- The parametric Bayes model works even better.

Application

- In the aftermath of the 2007-09 global financial crisis bank stress tests have become an important tool used by central banks and other regulators to conduct macroprudential regulation and supervision.
- Stress tests come in many flavors, one of them is to predict the evolution of bank balance sheets conditional on economic conditions.
- Bank-level forecasts can then be aggregated into industry-wide losses and revenues.
- Initially, we tried to focus on forecasts of charge-offs and revenues which can be mapped into forecasts of capital-asset ratios.
- However, charge-offs have very non-Gaussian features and for now we switched to direct forecasts of capital-asset ratios.
- Stress tests condition on extreme counterfactual economic conditions, whereas in our forecast exercise we condition on actual economic conditions.

Application

- We follow Covas, Rump, and Zakrajsek (CRZ, 2013) in terms of capital-asset ratio definitions.
- Regulators pay attention to the so-called tier-1-common ratio:

$$\mathrm{T1CR}_{it} = \frac{E_{it} - \mathrm{Deductions}_{it}}{\mathrm{RWA}_{it}}.$$

- Tier-1 common equity is the highest quality component of bank capital. The denominator RWA is the Basel I risk-weighted assets.
- CRZ decompose the evolution of equity as

$$\begin{aligned} E_{it} &= E_{i,t-1} + (1-\tau)\Big[\sum_j \mathrm{PPNR}^j_{it} \times \mathrm{Assets}^j_{it} \\ &\quad - \sum_l \mathrm{NCO}^l_{it} \times \mathrm{Loans}^l_{it} \Big] - \mathrm{Equity\ Payouts}_{it} \end{aligned}$$

where PPNR are net revenues and NCO are net charge-offs.

Data Sources

- Bank balance sheet data are available through the Call Reports at quarterly frequency from the Federal Reserve Bank of Chicago.
- We multiply T1CR by 100 and take logs.
- We will relate T1CR to local economic conditions, e.g., house prices and unemployment. Thus, we will focus on small banks (assets less than 1 billion \$).
- We use the Summary of Deposits data from the Federal Deposit Insurance Corporation to determine the local market for each bank.
- Currently: local market = state.
- We collect
  - state-level housing price index (all transactions, not seasonally adjusted) from the Federal Housing Finance Agency;
  - state-level unemployment rate (monthly data averaged to quarterly freq, seasonally adjusted) from the Bureau of Labor Statistics.

Model Specification

- Basic panel data model

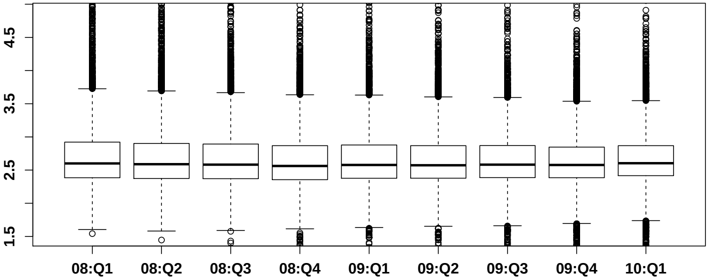
$$\ln(100 \cdot \text{T1CR}_{it}) = \lambda_i + \beta_1 \ln(100 \cdot \text{T1CR}_{i,t-1}) + \beta_2 \text{UR}_{it} + \beta_3 \ln \text{HPI}_{it} + U_{it} \quad (30)$$

- $U_{it} \sim iidN(0, \sigma^2)$.
- Parametric prior for λ_i:

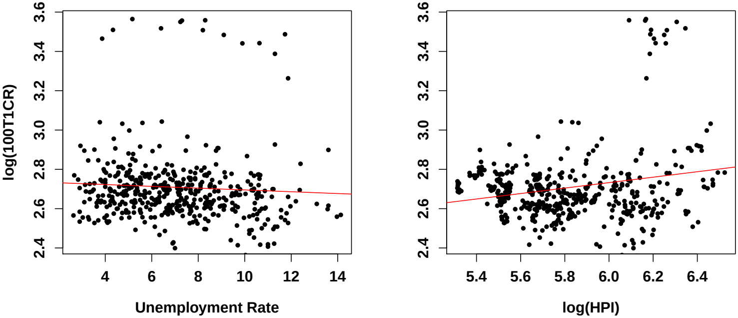
$$\lambda_i | (\text{T1CR}_{i0}, \phi, \omega^2) \sim iidN\big(\phi_0 + \phi_1 \ln(100 \cdot \text{T1CR}_{i0}), \omega^2\big) \quad (31)$$

- Sample period: $t = 0$ corresponds to 2008:Q1, $t = T$ is 2009:Q4.
- Forecast period: $t = T + 1$ is 2010:Q1.
- Sample size is $N = 6,066$.

$\log(100 \cdot \text{T1CR})$ Data



T1CR Data versus Unemployment and House Prices

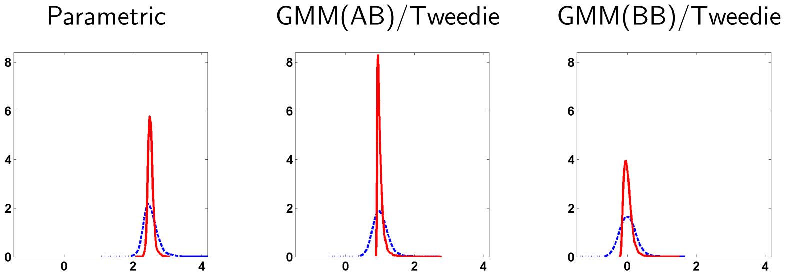


Note: capital asset ratios are averaged across time for each bank and across banks within the same state.

Parameter Estimates

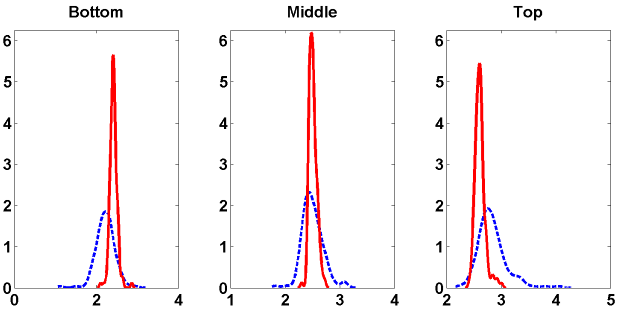
Parameter	Max of Marg. LH	GMM(AB)	GMM(BB)
$\ln(100 \cdot \text{T1CR}_{i,t-1})$	0.0497	0.0456	0.0385
$\ln \text{HPI}_{it}$	0.0172	0.2784	0.4548
$\text{UR}_{i,t}$	-0.0095	-0.0066	-0.0061
$\hat{\sigma}^2$	0.2223	0.2219	0.2221
$\hat{\phi}_0$	2.2644		
$\hat{\phi}_1$	0.0910		
$\hat{\omega}^2$	0.0650		

Shrinkage Effects: Estimates of $Z_i(\hat{\rho})$ (blue, dashed) and $\hat{\lambda}_i$ (red, solid)



- The empirical Bayes procedures induce a substantial amount of shrinkage: $\hat{\lambda}_i$ densities are much more concentrated than $Z_i(\hat{\rho})$ densities.

Implicit Bias Correction: Parametric Bayesian Model



- The empirical Bayes procedures induces a bias correction for the bottom and top groups.

Forecast Results

$\hat{\beta}$	$\hat{\lambda}_i$	Bottom 2% MSE	Bottom 2% Med	Middle 2% MSE	Middle 2% Med	Top 2% MSE	Top 2% Med	All MSE	All Med
				No Shrinkage					
GMM (AB)	QMLE	0.48	0.28	0.22	0.02	0.60	-0.34	0.25	-0.01
GMM (BB)	QMLE	0.47	0.25	0.23	0.02	0.58	-0.32	0.26	0.00

- GMM(BB) and GMM(AB) estimators perform similarly.
- Relatively large selection bias for top and bottom groups.

Forecast Results

$\hat{\beta}$	$\hat{\lambda}_i$	Bottom 2% MSE	Bottom 2% Med	Middle 2% MSE	Middle 2% Med	Top 2% MSE	Top 2% Med	All MSE	All Med
No Shrinkage									
GMM (AB)	QMLE	0.48	0.28	0.22	0.02	0.60	-0.34	0.25	-0.01
Tweedie's Formula									
GMM (AB)	Lindsey	0.32	-0.04	0.19	-0.02	0.53	-0.22	0.22	-0.03
GMM (AB)	Kernel	0.39	-0.02	0.20	0.01	0.58	-0.24	0.24	-0.03

- Tweedie's formula is able to correct the selection bias.

Forecast Results

$\hat{\beta}$	$\hat{\lambda}_i$	Bottom 2% MSE	Bottom 2% Med	Middle 2% MSE	Middle 2% Med	Top 2% MSE	Top 2% Med	All MSE	All Med
				No Shrinkage					
GMM (AB)	QMLE	0.48	0.28	0.22	0.02	0.60	-0.34	0.25	-0.01
				Tweedie's Formula					
GMM (AB)	Lindsey	0.32	-0.04	0.19	-0.02	0.53	-0.22	0.22	-0.03
GMM (AB)	Kernel	0.39	-0.02	0.20	0.01	0.58	-0.24	0.24	-0.03
				Empirical Bayes Forecast with Parametric Model					
Max of Marg. LH		0.34	0.03	0.19	-0.04	0.49	-0.19	0.22	-0.05

- Similar performance of parametric approach and Tweedie's formula.

Conclusions

- To forecast dynamic panel data model, it's important to have a "good" estimates of the individual effects λ_i.
- "Selection" bias: repeated positive shocks (U_{it}) lead to overestimation of their corresponding λ_i's, especially when T is small.
- Shrinkage estimators can offset the selection bias and improve the forecasts:
  - Empirical Bayes estimator of parametric model; essentially a random effects model.
  - Plug-in predictor based on Tweedie's formula
- Both methods lead to improvements in forecast accuracy in simulations and in an application to capital-asset ratio forecasts.
- Work in progress... Many extensions.