Forecasting with panel data: Estimation uncertainty versus parameter heterogeneity

Hashem Pesaran - *USC and Trinity College, Cambridge*
Andreas Pick - *Erasmus University Rotterdam, TI, DNB*
Allan Timmermann - *UCSD, CEPR and CREATES*

ESEM, August 2022

Focus and agenda

- Panel literature focuses on parameter estimation and inference
    - Surprisingly few studies on using panel estimation to forecast individual units, see Baltagi (2013)
- We examine existing approaches and develop new forecasting methods for panel data with heterogeneous parameters
    - new forecast poolability test which we contrast with parameter homogeneity tests
    - new forecast combination methods
- We compare the predictive accuracy of individual forecasts for different cross-sectional (N) and time (T) dimensions, and varying degrees of parameter heterogeneity

Panel data models with parameter heterogeneity

- Baseline panel model

$$y_{it} = \boldsymbol{\beta}_i' \boldsymbol{x}_{it} + \varepsilon_{it}, \quad i = 1, 2, \ldots, N, \ t = 1, 2, \ldots, T$$

- We study forecast of individual units, y_{it}, given the ex-ante known predictors, x_{it}
- Forecasts can be computed using individual estimates of $\boldsymbol{\beta}_i$ or using the pooled estimate $\boldsymbol{\beta}_i = \boldsymbol{\beta}$

Comparing MSFE values: Individual vs. pooled

- Forecasts based on individual estimates:

$$\begin{aligned} \hat{y}_{i,T+1} &= \hat{\beta}_i' \mathbf{x}_{i,T+1}, \qquad \hat{\beta}_i = (\mathbf{X}_i'\mathbf{X}_i)^{-1}\mathbf{X}_i'\mathbf{y}_i, \\ \hat{e}_{i,T+1} &= y_{i,T+1} - \hat{y}_{i,T+1}. \end{aligned}$$

- Forecasts based on pooled estimator:

$$\begin{aligned} \tilde{y}_{i,T+1} &= \tilde{\beta}' \mathbf{x}_{i,T+1}, \qquad \tilde{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \\ \tilde{e}_{i,T+1} &= y_{i,T+1} - \tilde{y}_{i,T+1}. \end{aligned}$$

Proposition 1

(i) The MSFE from individual-specific parameter estimation is

$$\begin{aligned}\mathrm{Var}(\hat{e}_{i,T+1}|\boldsymbol{X}_i,\boldsymbol{x}_{i,T+1}) &= \sigma_i^2 + T^{-1}\sigma_i^2 \boldsymbol{x}'_{i,T+1}\boldsymbol{Q}_{iT}^{-1}\boldsymbol{x}_{i,T+1} \\ &= \sigma_i^2 + O_p\left(T^{-1}\right),\end{aligned}$$

where $\boldsymbol{Q}_{iT} = T^{-1}\boldsymbol{X}'_i\boldsymbol{X}_i$.
(ii) The MSFE from pooled parameter estimation is

$$\mathrm{Var}(\tilde{e}_{i,T+1}|\boldsymbol{X}_i,\boldsymbol{x}_{i,T+1}) = \sigma_i^2 + \boldsymbol{x}'_{i,T+1}\boldsymbol{\Omega}_\eta\boldsymbol{x}_{i,T+1} + O_p\left(N^{-1}\right).$$

Interpretation of Proposition

Remark 1 For small T, parameter estimation uncertainty can be important for the *individual* forecasts. Parameter heterogeneity, in contrast, does not affect the accuracy of the individual forecasts. For large T, forecasts based on individual estimation will have a lower MSFE than forecasts based on pooled estimation.

Remark 2 The accuracy of forecasts that use pooled estimates depends both on the degree of parameter heterogeneity and the dispersion of the predictors through $\boldsymbol{x}'_{i,T+1}\boldsymbol{\Omega}_{\eta}\boldsymbol{x}_{i,T+1}$.

Remark 3 Forecasts based on *individual* estimates have large T optimality properties even if predictors are weakly exogenous. In contrast, forecasts based on *pooled* regressions require strict exogeneity.

Remark 4 Parameter heterogeneity could be particularly problematic if $\boldsymbol{\eta}_i = \boldsymbol{\beta}_i - \mathrm{E}(\boldsymbol{\beta}_i)$ and $\boldsymbol{X}'_i\boldsymbol{X}_i$ are correlated. *Individual* estimates of $\boldsymbol{\beta}_i$ are not affected by parameter heterogeneity even if heterogeneity is correlated with $\boldsymbol{x}_{it}$. The same is *not* true for the *pooled* estimates: $\underset{N\to\infty}{\text{plim}}\left(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = \boldsymbol{\Psi}^{-1}\boldsymbol{b}$, where

$$\boldsymbol{\Psi} = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\mathrm{E}\left(\boldsymbol{X}'_i\boldsymbol{X}_i\right), \quad \boldsymbol{b} = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\mathrm{E}\left(\boldsymbol{X}'_i\boldsymbol{X}_i\boldsymbol{\eta}_i\right).$$

For the pooled estimator to be unbiased, need that $\mathrm{E}\left(\boldsymbol{X}'_i\boldsymbol{X}_i\boldsymbol{\eta}_i\right) = \boldsymbol{0}$.

Combination forecasts

Consider the combined forecast

$$y^*_{i,T+1} = \omega_i \hat{y}_{i,T+1} + (1 - \omega_i)\tilde{y}_{i,T+1},$$

with forecast error

$$e^*_{i,T+1} = \omega_i \hat{e}_{i,T+1} + (1 - \omega_i)\tilde{e}_{i,T+1}.$$

The optimal ω_i^*, which minimizes the MSFE of the combined forecast is

$$\omega_i^* = \frac{\text{Var}(\tilde{e}_{i,T+1}) - \text{Cov}(\hat{e}_{i,T+1}, \tilde{e}_{i,T+1})}{\text{Var}(\hat{e}_{i,T+1}) + \text{Var}(\tilde{e}_{i,T+1}) - 2\text{Cov}(\hat{e}_{i,T+1}, \tilde{e}_{i,T+1})}.$$

Optimal combination weights

Proposition 2

For fixed $T > T_0$, the optimal combination weights that minimize the MSFE conditional on $\boldsymbol{X}_i$ and $\boldsymbol{x}_{i,T+1}$, $\boldsymbol{\Omega}_\eta$ and σ_i^2 are given by (for $i = 1, 2, \ldots, N$)

$$\omega_i^* = \frac{\boldsymbol{x}'_{i,T+1}\boldsymbol{\Omega}_\eta \boldsymbol{x}_{i,T+1}}{\boldsymbol{x}'_{i,T+1}\left[T^{-1}\sigma_i^2\mathbf{Q}_{iT}^{-1} + \boldsymbol{\Omega}_\eta\right]\boldsymbol{x}_{i,T+1}} + O_p\left(N^{-1}\right)$$

Bias-adjusted combination weights

We use the following bias-adjusted plug-in estimator for ω_i^*

$$\tilde{\omega}_i^* = \frac{\boldsymbol{x}'_{i,T+1}\left[\hat{\boldsymbol{\Omega}}_\eta - \frac{1}{NT}\sum_{i=1}^N \hat{\sigma}_i^2 \mathbf{Q}_{iT}^{-1}\right]\boldsymbol{x}_{i,T+1}}{\boldsymbol{x}'_{i,T+1}\left[\frac{1}{T}\hat{\sigma}_i^2\mathbf{Q}_{iT}^{-1} + \hat{\boldsymbol{\Omega}}_\eta - \frac{1}{NT}\sum_{i=1}^N \hat{\sigma}_i^2 \mathbf{Q}_{iT}^{-1}\right]\boldsymbol{x}_{i,T+1}}$$

Weights are restricted to lie between zero and one

Forecast-based tests for pooling

From the expressions for the MSFE of the individual and pooled forecasts

$$\text{MSFE}(\hat{y}_{i,T+1}) - \text{MSFE}(\tilde{y}_{i,T+1}) \\ = T^{-1}\sigma_i^2 \boldsymbol{x}'_{i,T+1}\boldsymbol{Q}_{iT}^{-1}\boldsymbol{x}_{i,T+1} - \boldsymbol{x}'_{i,T+1}\,\boldsymbol{\Omega}_{\eta}\boldsymbol{x}_{i,T+1} + O_p\left(N^{-1}\right).$$

New pre-test for pooling

Proposition 3

Suppose that $\boldsymbol{\eta}_i$ and $\boldsymbol{\varepsilon}_i$ are normally distributed, and $\boldsymbol{\eta}_i$ are cross-sectionally independent. Then, under the null of equal forecast accuracy defined by

$$H_{0,PF} : T^{-1}\sigma_i^2\mathbf{x}'_{i,T+1}\mathbf{Q}_{iT}^{-1}\mathbf{x}_{i,T+1} = \mathbf{x}'_{i,T+1}\boldsymbol{\Omega}_\eta\mathbf{x}_{i,T+1}, \quad \forall i,$$

there exists a finite T_0 such that for all $T > T_0$ and as $N \to \infty$

$$PF_{NT} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\omega_{i,NT}^2 - 1}{\sqrt{2}}\right) \xrightarrow{d} \mathrm{N}(0,1),$$

where

$$\omega_{i,NT}^2 = \frac{T\left[\mathbf{x}'_{i,T+1}(\hat{\boldsymbol{\beta}}_i - \bar{\hat{\boldsymbol{\beta}}})\right]^2}{2\sigma_{i,NT}^2\mathbf{x}'_{i,T+1}\mathbf{Q}_{iT}^{-1}\mathbf{x}_{i,T+1}}.$$

Monte Carlo: Design

$$y_{i,t+1} = \alpha_i + \rho_i y_{it} + \gamma_i x_{it} + \kappa \sigma_i \varepsilon_{i,t+1},$$

$$\varepsilon_{i,t+1} \sim \text{iidN}(0,1)\,, \quad \sigma_i^2 \sim \text{iid}\left(1+\chi_1^2\right)/2$$

$$x_{it} = \mu_{xi} + \xi_{it}, \quad \mu_{xi} = (z_i^2 - 1)/\sqrt{2}, \quad z_i \sim \text{iidN}\,(0,1)\,,$$

$$\xi_{it} = \rho_{xi}\xi_{i,t-1} + \sigma_{xi}\left(1-\rho_{xi}^2\right)^{1/2}\nu_{it}, \quad \nu_{it} \sim \text{iidN}\,(0,1)\,,$$

$$\sigma_{xi}^2 \sim \text{iid}\left(1+\chi_1^2\right)/2.$$

$N = 500$ and $T = \{20, 50, 100\}$.
Number of replications: $\mathcal{R} = 10,000$.

Monte Carlo (cont.)

- Autocorrelations of x_{it}: $\rho_{xi} \sim$ iid Uniform$(0, 0.95)$
- Coefficient of $y_{i,t-1}$: $\rho_i \sim$ iid Uniform$(0, \bar{\rho})$, where we vary $\bar{\rho}$ to capture different degrees of dynamic heterogeneity. The value of $\bar{\rho}$ depends on the value of pooled R^2.
- We consider cases where the regressors and the coefficients are correlated:

$$\alpha_i = \phi\mu_{xi} + \sigma_\eta \eta_i, \text{ and } \gamma_i = 1 + \theta\mu_{xi} + \sigma_\zeta \zeta_i,$$

where $\eta_i, \zeta_i \sim$ iidN$(0, 1)$.

List of Panel Forecasting Methods

- Individual estimation
- Pooled estimation
- Random effects estimator of Goldberger (1962)
- Median group estimator (new)
- Forecast combination with bias-unadjusted and bias-adjusted weights
- Forecasts based on the pre-test (poolability test)
- Forecasts based on the shrinkage estimators (Maddala et al. (1997)):
  - prior likelihood
  - Bayesian
  - empirical Bayes

Measures of Predictive Accuracy: Median of individual MSFEs

Some Monte Carlo Results

- Pooled estimates perform best:
  - under parameter homogeneity
  - when average R^2 is low and T is relatively small, $T = 20$, even in the presence of parameter heterogeneity.
  - in the absence of correlated heterogeneity
- Forecast combination using $\hat{\boldsymbol{\Omega}}_\eta$ is the most precise method more often than any other methods, and is otherwise close to the best forecast. It is also robust to correlation between parameters and regressors.
- Pre-test forecasts, using the poolability test, consistently chooses best of individual and pooled forecasts.
- Shrinkage forecasts perform well under parameter homogeneity, but have a mixed performance under parameter heterogeneity
- Conclusion: No uniform ordering of the forecasting methods due to the tradeoffs between estimation uncertainty and parameter heterogeneity

Empirical Applications

Three empirical applications:

1. House price inflation across U.S. metropolitan areas
2. inflation of CPI sub-indices
3. stock returns on U.S. firms.

Contrasting in-sample fit:

- individual stock returns: in-sample $R^2 < 0.01$
- CPI: in-sample $R^2 \approx 0.2$
- house price inflation in-sample $R^2 \approx 0.8$

House price changes

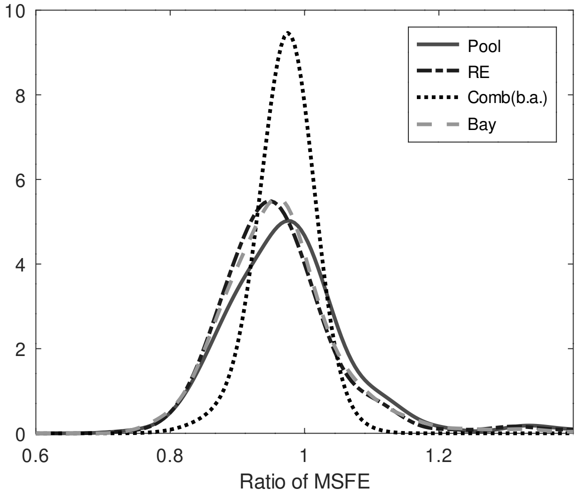
- Quarterly data on annual real house price inflation on 377 U.S. Metropolitan Statistical Areas (MSAs), 1975Q1–2014Q4
- Forecasts 1995Q1–2014Q4, rolling window of 60 observations.

$$y_{it} = \alpha_i + \rho_i y_{i,t-1} + \rho_i^* y_{i,t-1}^* + \gamma_{Ri} \bar{y}_{i,t-1}^{(R)} + \gamma_{Ci} \bar{y}_{t-1} + \varepsilon_{it},$$

$$y_{it}^* = \sum_{k=1,k\neq i}^{N} \omega_{ik} y_{jt}$$

- ω_{ik} are given spatial weights
- $\bar{y}_{it}^{(R)}$: average house price inflation in region R of MSA i
- $\bar{y}_t$: country-wide average house price inflation.
- SAR: $\gamma_{Ri} = \gamma_{Ci} = 0$; SARX: $\gamma_{Ri} \neq \gamma_{Ci} \neq 0$

Cross-sectional distribution of MSFE ratios for house price changes



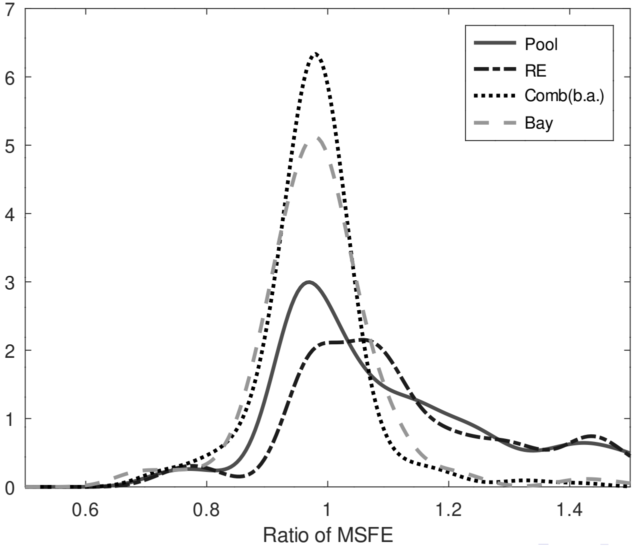
Forecasting results for U.S. House price changes

Forecast methods	Median MSFE		freq. beating benchmark		freq. smallest MSFE		freq. largest MSFE	
	SAR	SARX	SAR	SARX	SAR	SARX	SAR	SARX
Individual	2.536	2.569	–	–	0.077	0.036	0.450	0.401
Pooled	0.971	0.989	0.660	0.539	0.030	0.011	0.119	0.202
RE	0.952	0.952	0.754	0.682	0.221	0.091	0.041	0.044
Median group	0.952	0.941	0.727	0.688	0.312	0.318	0.050	0.069
Optimal combination								
Naive	0.980	0.975	0.876	0.934	0.019	0.047	0.000	0.000
Bias adj.	0.974	0.966	0.859	0.914	0.069	0.119	0.006	0.006
Pre-test								
PF	0.984	0.974	0.608	0.691	0.102	0.185	0.213	0.091
Shrinkage								
Prior lik.	0.970	0.963	0.715	0.622	0.047	0.088	0.105	0.149
Bayes.	0.960	0.948	0.749	0.699	0.058	0.047	0.006	0.003
Emp. Bayes.	0.956	0.954	0.754	0.652	0.064	0.058	0.011	0.036

CPI inflation subindices

- Inflation rates for 202 sub-indices of the US consumer price index (CPI)
- Monthly frequency, Nov 1958–Dec 2018
- Rolling estimation windows of 60 observations
- We generate 590 forecasts for each series, with the first forecast computed for November 1969
- Two forecasting models:
  - AR: a purely autoregressive specification with lags 1, 2, and 12;
  - AR-PC: AR specification augmented with the lagged value of the first PC;

Cross-sectional distribution of MSFE ratios for CPI inflation



Results for CPI inflation subindices

Forecast method	Median MSFE AR	Median MSFE AR-PC	freq. beating benchmark AR	freq. beating benchmark AR-PC	freq. smallest MSFE AR	freq. smallest MSFE AR-PC	freq. largest MSFE AR	freq. largest MSFE AR-PC
Individual	1.568	1.573	–	–	0.059	0.054	0.064	0.054
Pooled	1.076	1.077	0.351	0.347	0.144	0.153	0.119	0.124
RE	1.153	1.155	0.213	0.218	0.015	0.015	0.579	0.564
Median group	1.038	1.038	0.337	0.342	0.030	0.030	0.124	0.119
Optimal combination								
Naive	0.975	0.974	0.936	0.936	0.317	0.297	0.000	0.000
Bias adj.	0.973	0.971	0.678	0.673	0.074	0.074	0.000	0.000
Pre-test								
PF	1.000	1.000	0.356	0.485	0.030	0.030	0.0.03	0.025
Shrinkage								
Prior lik.	0.991	0.989	0.574	0.554	0.069	0.059	0.054	0.020
Bayes.	0.982	0.980	0.644	0.683	0.114	0.104	0.000	0.000
Emp. Bayes.	0.994	0.996	0.584	0.550	0.173	0.188	0.035	0.094

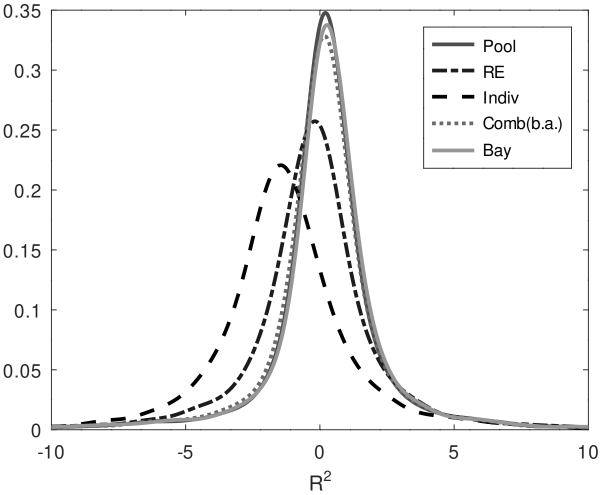
Stock returns

- Panel of 23,121 individual US firm-level monthly stock returns
- Sample from January 1977 through December 2017
- Rolling estimation window of 120 observations; out-of-sample forecasts from January 1987 through December 2017
  given period, between 1,116 and 2,726 stocks are included

$$y_{i,t+1} = \alpha_i + \beta_i x_{it} + \varepsilon_{it+1},$$

- x_{it} is the 6-month momentum of stock i, measured using cumulative returns up to the previous month

Cross-sectional distribution of OOS R^2 for stock return forecasts



Results for stock return forecasts

Forecast method	Median R-squared	freq. beating prevail.mean	freq. beating individual	freq. smallest MSFE	freq. largest MSFE
Individual	-1.399	0.352	–	0.088	0.583
Prevailing mean	–	–	0.648	0.164	0.089
Pooled	0.232	0.609	0.648	0.106	0.018
RE	-0.220	0.439	0.647	0.200	0.206
Median Group	0.302	0.623	0.648	0.151	0.011
Optimal combination					
Naive	−0.244	0.414	0.647	0.049	0.002
Bias adj.	0.184	0.577	0.648	0.117	0.026
Pre-test					
PF	0.232	0.609	0.648	0.106	0.018
Shrinkage					
Prior lik.	0.274	0.602	0.648	0.136	0.074
Bayes.	0.295	0.622	0.648	0.033	0.002
Emp. Bayes.	0.295	0.623	0.648	0.003	0.000

Conclusion

- Comprehensive examination of the out-of-sample predictive accuracy of a large set of existing and novel panel data methods
- Main findings:
  1. Several approaches perform systematically better than individual forecasts: sizeable gains from exploiting panel information
  2. No single forecasting approach is uniformly dominant across applications. Yet, combination and (Bayesian) shrinkage forecasts are more precise than the pooled and individual forecasts
  3. Methods differ in risk profiles: individual, pooled, random effect, and median group methods do poorly in at least one application
- Forecast combinations and shrinkage methods offer insurance against poor performance. Forecast combinations, in particular, perform well across the board, the performance of shrinkage methods tends to vary a bit more across applications