UNIVERSITY OF CAMBRIDGE
Faculty of Economics

CAMBRIDGE WORKING PAPERS IN ECONOMICS

Forecasting with panel data: estimation uncertainty versus parameter heterogeneity

M. Hashem Pesaran
Andreas Pick
Allan Timmermann

Abstract

We provide a comprehensive examination of the predictive accuracy of panel forecasting methods based on individual, pooling, fixed effects, and Bayesian estimation, and propose optimal weights for forecast combination schemes. We consider linear panel data models, allowing for weakly exogenous regressors and correlated heterogeneity. We quantify the gains from exploiting panel data and demonstrate how forecasting performance depends on the degree of parameter heterogeneity, whether such heterogeneity is correlated with the regressors, the goodness of fit of the model, and the cross-sectional (N) and time (T) dimensions. Monte Carlo simulations and empirical applications to house prices and CPI inflation show that forecast combination and Bayesian forecasting methods perform best overall and rarely produce the least accurate forecasts for individual series.

Reference Details

CWPE	2219
Published	18 March 2022
Updated	18 April 2024
Key Words	Forecasting, Panel data, Heterogeneity, Pooled estimation; Forecast combination
JEL Codes	C33, C53
Website	www.econ.cam.ac.uk/cwpe

Forecasting with panel data: Estimation uncertainty versus parameter heterogeneity*

M. Hashem Pesaran† Andreas Pick‡ Allan Timmermann§

March 24, 2024

Abstract

We provide a comprehensive examination of the predictive accuracy of panel forecasting methods based on individual, pooling, fixed effects, and Bayesian estimation, and propose optimal weights for forecast combination schemes. We consider linear panel data models, allowing for weakly exogenous regressors and correlated heterogeneity. We quantify the gains from exploiting panel data and demonstrate how forecasting performance depends on the degree of parameter heterogeneity, whether such heterogeneity is correlated with the regressors, the goodness of fit of the model, and the cross-sectional (N) and time (T) dimensions. Monte Carlo simulations and empirical applications to house prices and CPI inflation show that forecast combination and Bayesian forecasting methods perform best overall and rarely produce the least accurate forecasts for individual series.

JEL codes: C33, C53

Keywords: Forecasting, Panel data, Heterogeneity, Pooled estimation; Forecast combination.

*We thank the editor and three anonymous reviewers for their helpful and constructive comments. We also thank Laura Liu, Mahrad Sharifvaghefi, Ron Smith, Cynthia Yang, Liying Yang and seminar participants at University of Pittsburgh, the Board of the Federal Reserve, ESEM, VTSS, and IAAE for helpful comments.

†University of Southern California and Trinity College, Cambridge. Email: pesaran@usc.edu

‡Erasmus University Rotterdam, Erasmus School of Economics, Burgemeester Oudlaan 50, 3000DR Rotterdam, and Tinbergen Institute. Email: andreas.pick@cantab.net

§UC San Diego, Rady School of Management, 9500 Gilman Drive, La Jolla CA 92093-0553. Email: atimmermann@ucsd.edu.

1 Introduction

Panel data sets on economic and financial variables are widely available at individual, firm, industry, regional, and country granularities and have been extensively used for estimation and inference. Yet, panel estimation methods have had a comparatively lower impact on common practices in economic forecasting, which remain dominated by unit-specific forecasting models or low-dimensional multivariate models such as vector autoregressions (Hsiao, 2022). The relative shortage of panel applications in the economic forecasting literature is, in part, a result of the absence of a deeper understanding of the determinants of forecasting performance for different panel estimation methods and the absence of guidelines on which methods work well in different settings.

In this paper, we examine existing approaches and develop novel forecast combination methods for panel data with possibly correlated heterogeneous parameters and conduct a systematic comparison of their predictive accuracy in settings with different cross-sectional (N) and time (T) dimensions and varying degrees of parameter heterogeneity, whether correlated or not. Our analysis provides a deeper understanding of the determinants of the performance of these methods across a variety of settings chosen for their relevance to economic forecasting problems. This includes the important choice of whether to use pooled versus individual estimates, or perhaps a combination of the two approaches, with a focus on forecasting rather than parameter estimation and inference.

We begin by exploring analytically the bias-variance trade-off between individual, fixed effects (FE), and pooled estimation for forecasting. Our analysis is conducted in a general setting that allows for weakly exogenous regressors and correlated heterogeneity, consistent with the type of dynamic panel models commonly used in empirical applications. We show how such effects contribute to the mean squared forecast error (MSFE) of forecasts based on individual, FE, and pooled estimates.

We next examine two forecast combination methods. Estimation errors are well-known to lead to imprecisely estimated combination weights for data with a small time-series dimension. Our combination schemes assume homogeneous weights across individual variables, which allows us to use cross-sectional information to reduce the effect of estimation error on the combination weights and stabilizes our combination weights compared to a scheme that lets the weights be individual-specific. Our first scheme combines forecasts from individual and pooled models. To handle cases where the pooling estimator imposes too much homogeneity, we propose a second combination scheme based

on forecasts from the individual-specific and fixed effect estimators.

Finally, we consider forecasts based on Bayesian estimators, namely the empirical Bayes approach of Hsiao et al. (1999) and the hierarchical Bayesian approach of Lindley and Smith (1972) using the Gibbs sampler of Gelfand et al. (1996). These are related to forecast combination and we show for the empirical Bayes estimator that it can be thought of as a weighted average of an estimator that allows for full heterogeneity and a pooled mean group estimator. The empirical Bayes scheme assigns greater weight to the pooled estimator, the lower the estimated degree of parameter heterogeneity and so adapts to the degree of parameter heterogeneity characterizing a given data set. The hierarchical Bayesian model has also been used by Lee and Griffith (1979) and Maddala et al. (1997).

We evaluate the predictive accuracy of these alternative panel forecasting methods through Monte Carlo simulations of a set of first-order autoregressive (ARX) panel models. These simulations explore the importance to forecasting performance of the degree of parameter heterogeneity, along with how correlated it is, whether it affects intercepts or slopes, and the dimensions of N and T. In the scenario with homogeneous parameters, forecasts based on pooled estimates are most accurate. Forecasts based on fixed or random effect estimates perform well, relative to other methods, when parameter heterogeneity is confined to the intercepts and does not affect slopes. Outside these cases, combination and Bayesian forecasts produce the most accurate forecasts since they are better able to handle parameter heterogeneity, whether correlated or not, while being more robust in cases with a small T than the individual-specific approach.

Next, we consider two empirical applications selected to represent varying degrees of heterogeneity and predictive power of the underlying forecasting models. We characterize the center of the cross-sectional loss distribution of the forecasts through the ratio of their average MSFE values relative to the average MSFE of the unit-specific benchmark. We also study the quantiles and the tail features of the loss distributions through the proportion of units for which the predictive accuracy of each approach is either best or worst among all methods considered.

Our first application considers predictability of house prices across 362 US metropolitan statistical areas (MSAs). The forecasting models for this application have a high pooled R^2 value above 0.8. In this application, individual-specific forecasts perform quite poorly, producing the highest MSFE values among all methods for up to 60% of the MSAs and the lowest MSFE values for less than 7% of MSAs. Forecasts based on pooled estimates perform notably better and reduce the average

MSFE value by 3% relative to the forecasts based on individual estimates. Forecast combinations and Bayesian forecasts work even better in this application, beating forecasts based on individual estimates for nearly 90% of MSAs while rarely generating the least accurate forecasts for individual series.

Our second application considers forecasts for a panel containing 187 subcategories of CPI inflation. Our forecasting models for this data have a substantially lower pooled R^2 in the range of 0.1 to 0.3. In this application, forecasts based on individual estimates generate the highest MSFE-values for 40% of the series and the lowest MSFE-values for less than 2% of the series. Forecasts based on pooled estimates also produce the highest MSFE-values for around 40% of the individual series but, conversely, generate the lowest MSFE-values for nearly 20% of the series. Combination forecasts are again more accurate than either of these methods as they improve on the average MSFE performance and reduce the MSFE of the individual-specific forecasts for 80% of the series. They also do not produce the largest MSFE for a single individual variable. Even better inflation forecasts are produced by the empirical Bayes method which is more accurate (in an MSFE sense) than the individual forecasts for 98% of the series, generates the lowest MSFE values for 35% of the individual variables, and never produces the worst MSFE performance among our methods.

Overall, forecasts that use only the information on a given unit tend to have loss distributions with wide dispersions across units. Their associated forecasts are therefore sometimes the best but far more often the worst, and their distribution of MSFE performance is often shifted to the right, implying larger losses on average than for other methods. Forecasts based on pooling, random effects (RE), or FE estimation tend to perform better, on average, than the individual-specific model whose forecast accuracy they beat for the majority of series. However, relative to the MSFE-values of the individual-specific forecasts, these approaches also tend to have a right-skewed distribution, suggesting a high risk of poor forecasting performance for individual series whose model parameters are very different from the average. Combinations and Bayesian forecasts have much narrower MSFE distributions across units, often shifted to the left as they are centered around a smaller average loss. These methods rarely produce the largest squared forecast error among all methods that we consider.

While the literature on forecasting with panel data has focused on panel data models developed for inference rather than forecasting, there are some notable exceptions. The review articles by Baltagi (2008, 2013) consider the forecasting performance of the best linear unbiased predictor

(BLUP) of Goldberger (1962) in models with either fixed effects or random effects. The BLUP estimator gives rise to a generalized least squares (GLS) predictor which Baltagi compares to models that allow for autoregressive moving average (ARMA) dynamics in innovations as well as models with spatial dependencies in the errors.

Trapani and Urga (2009) use Monte Carlo simulations to assess the forecasting performance of pooled, individual, and shrinkage estimators and find that the degree of heterogeneity is a key determinant of the accuracy of different forecasts. Brückner and Siliverstovs (2006) consider a similar group of methods to forecast migration data and find that fixed effects and shrinkage estimators perform best.

Wang et al. (2019) also propose forecast combination methods. However, their analysis does not allow for correlation of regressors and parameters or dynamics in the model. Additionally, their combination weights are determined from in-sample test statistics rather than the expected out-of-sample performance that we propose. In this sense, our approach is closer to the forecast based test for a structural break of Pesaran et al. (2013) and Boot and Pick (2020), where the target is also significant improvements in forecast accuracy rather than a significant change in parameters.

Liu, Moon and Schorfheide (2020) study forecasting for dynamic panel data models with a short time-series dimension. Though T exceeds the number of parameters that have to be estimated for each series, such estimates are typically very noisy and not consistent under large N, fixed T asymptotics. To handle estimation noise, like Lee and Griffith (1979), they adopt a Bayesian approach that shrinks the heterogeneous parameters to their mean, thus also exploiting cross-sectional information.[1] This is closely related to the idea of using forecast combinations to reduce the effect on the forecasts of noisy estimates of individual-specific parameters.

The outline of the rest of the paper is as follows. Section 2 introduces the model setup and our assumptions, while Section 3 derives analytical results on the predictive accuracy of individual, pooled, and FE forecasting schemes. Section 4 introduces our forecast combination schemes. Our theoretical results are summarized in four propositions. Section 5 describes the empirical and hierarchical Bayes estimators. Section 6 presents a set of Monte Carlo experiments designed to shed light on the determinants of the (relative) forecasting performance of the methods introduced in Sec-

[1] Our theoretical analysis focuses on the case with finite T and $N \to \infty$ and does not require that $\sqrt{N}/T \to 0$ as N and $T \to \infty$, jointly, which is often assumed in the literature.

tions 3 and 4. Section 7 reports results for our empirical applications. Finally, Section 8 concludes. Technical details are provided in appendices at the end of the paper and an online supplement.

2 Setup and assumptions

We begin by describing the panel regression setup and assumptions used in our analysis.

2.1 Panel regression model

Our analysis considers the following linear panel regression model:

$$y_{it} = \alpha_i + \boldsymbol{\beta}_i' \boldsymbol{x}_{it} + \varepsilon_{it} = \boldsymbol{\theta}_i' \boldsymbol{w}_{it} + \varepsilon_{it}, \quad \varepsilon_{it} \sim (0, \sigma_i^2), \tag{1}$$

where $i = 1, 2, \ldots, N$ refers to the individual units and $t = 1, 2, \ldots, T$ refers to the time period, y_{it} is the outcome of unit i at time t, $\boldsymbol{x}_{it}$ is a $k \times 1$ vector of regressors—or predictors—used to forecast y_{it}, $\boldsymbol{\beta}_i$ is the associated vector of regression coefficients, and ε_{it} is the disturbance of unit i in period t. The second equality in (1) introduces the notations $\boldsymbol{\theta}_i = (\alpha_i, \boldsymbol{\beta}_i')'$ and $\boldsymbol{w}_{it} = (1, \boldsymbol{x}_{it}')'$ which have dimensions $K \times 1$, with $K = k + 1$. For simplicity, we use the time subscript t for $\boldsymbol{x}_{it}$ and $\boldsymbol{w}_{it}$, but it is important to emphasize that this refers to the predicted time for the outcome variable, y_{it}. For a forecast horizon of h periods, all variables in $\boldsymbol{x}_{it}$ must therefore be known at time $t-h$. Our notation avoids explicitly referring to h everywhere, but it should be recalled throughout the analysis that $\boldsymbol{x}_{it}$ includes suitably lagged predictors. We will focus on the case of $h = 1$ but extensions to larger h are straightforward.

Notations: Stacking the time series of outcomes, regressors and disturbances, define $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$, $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}', \boldsymbol{x}_{i2}', \ldots, \boldsymbol{x}_{iT}')'$, $\boldsymbol{W}_i = (\boldsymbol{\tau}_T, \boldsymbol{X}_i)$, where $\boldsymbol{\tau}_T$ is a $T \times 1$ vector of ones, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})'$. Further, let $\boldsymbol{y} = (\boldsymbol{y}_1', \boldsymbol{y}_2', \ldots, \boldsymbol{y}_N')'$, $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2', \ldots, \boldsymbol{X}_N')'$, $\boldsymbol{W} = (\boldsymbol{W}_1', \boldsymbol{W}_2', \ldots, \boldsymbol{W}_N')'$, and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', \boldsymbol{\varepsilon}_2', \ldots, \boldsymbol{\varepsilon}_N')'$. Generic positive finite constants are denoted by C when large and c when small. They can take different values at different instances. $\lambda_{\max}(\boldsymbol{A})$ and $\lambda_{\min}(\boldsymbol{A})$ denote the maximum and minimum eigenvalues of matrix $\boldsymbol{A}$. $\boldsymbol{A} \succ \boldsymbol{0}$ and $\boldsymbol{A} \succeq \boldsymbol{0}$ denote that $\boldsymbol{A}$ is a positive definite and a non-negative definite matrix, respectively. $\|\boldsymbol{A}\| = \lambda_{\max}^{1/2}(\boldsymbol{A}'\boldsymbol{A})$ and $\|\boldsymbol{A}\|_1$ denote the spectral and column norms of matrix $\boldsymbol{A}$, respectively. $\|\boldsymbol{x}\|_p = [\mathrm{E}(\|\boldsymbol{x}\|^p)]^{1/p}$. If $\{f_n\}_{n=1}^{\infty}$ is any real sequence and $\{g_n\}_{n=1}^{\infty}$ is a sequence of positive real numbers, then $f_n = O(g_n)$,

if there exists a C such that $|f_n|/g_n \leq C$ for all n and $f_n = o(g_n)$ if $f_n/g_n \to 0$ as $n \to \infty$. Similarly, $f_n = O_p(g_n)$ if f_n/g_n is stochastically bounded and $f_n = o_p(g_n)$ if $f_n/g_n \xrightarrow{p} 0$. The operator $\xrightarrow{p}$ denotes convergence in probability, and $\xrightarrow{d}$ denotes convergence in distribution.

2.2 Assumptions

Our theoretical analysis builds on a set of standard assumptions about the underlying data generating process.

Assumption 1. *ε_{it} is serially independent with mean zero, a fixed variance σ_i^2 $(0 < c < \sigma_i^2 < C < \infty)$, and with $\sup_{i,t} \mathrm{E}\,|\varepsilon_{it}|^4 < C < \infty$.*

Assumption 2. *For all i and t, the following orthogonality conditions hold:*

$$\mathrm{E}\,(\varepsilon_{it}\,|\boldsymbol{w}_{is}\,) = 0, \text{ for } t \geq s, \text{ for } t = 1, 2, \ldots, T, T+1.$$

Assumption 3. *(a) $\{\boldsymbol{w}_{it}\}$ for $i = 1, 2, \ldots, N$ are covariance stationary with $\mathrm{E}(\boldsymbol{w}_{it}\boldsymbol{w}_{it}') = \boldsymbol{Q}_i$, $\sup_{i,t=\{1,2,\ldots,T\}} \mathrm{E}\,\|\boldsymbol{w}_{it}\|^4 < C$, $\sup_{i,T} \|\boldsymbol{w}_{i,T+1}\| < C$, and*

$$\sup_i \lambda_{\max}\,(\boldsymbol{Q}_i\,) < C < \infty, \text{ and } \sup_i \lambda_{\max}\left(\boldsymbol{Q}_i^{-1}\right) < C < \infty. \tag{2}$$

(b) The sample covariance matrices $\boldsymbol{Q}_{iT} = T^{-1}\boldsymbol{W}_i'\boldsymbol{W}_i = T^{-1}\sum_{t=1}^{T} \boldsymbol{w}_{it}\boldsymbol{w}_{it}'$, for $i = 1, 2, \ldots, N$ are positive definite.

Assumption 4. *There exists a fixed T_0 such that for all $T > T_0$*

$$\sup_i \mathrm{E}\left\|T^{-1/2}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\right\|^4 < C < \infty, \tag{3}$$

$$\sup_i \mathrm{E}\left[\lambda_{\max}^4\,(\boldsymbol{Q}_{iT}\,)\right] < C < \infty, \text{ and } \sup_i \mathrm{E}\left[\lambda_{\max}^4\left(\boldsymbol{Q}_{iT}^{-1}\right)\right] < C < \infty. \tag{4}$$

Under Assumption 1 the optimal forecast of $y_{i,T+1}$, in a mean squared error sense, is given by $\mathrm{E}\,(y_{i,T+1}\,|\boldsymbol{w}_{i,T+1}, \boldsymbol{W}_i\,) = \boldsymbol{\theta}_i'\boldsymbol{w}_{i.T+1}$. Note that $\boldsymbol{w}_{i.T+1}$ is known at time T, and is bounded under

Assumption 3. Assumption 2 allows the regressors to be weakly exogenous with respect to $\boldsymbol{\varepsilon}_i$ and therefore permits the inclusion of lagged dependent variables such as $y_{i,T}$ in $\boldsymbol{w}_{i,T+1}$. Assumption 3 is an identification assumption that allows consistent estimation of individual slope coefficients, $\boldsymbol{\theta}_i$. Assumption 4 is required when we compare average MSFEs based on individual and pooled estimators. It provides sufficient conditions under which (see Lemma A.1)

$$\mathrm{E}\left\|\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i-\boldsymbol{\theta}_i\right)\right\|^2=\mathrm{E}\left\|\boldsymbol{Q}_{iT}^{-1}\left(T^{-1/2}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\right)\right\|^2<C<\infty, \tag{5}$$

where $\hat{\boldsymbol{\theta}}_i=\left(\boldsymbol{W}_i'\boldsymbol{W}_i\right)^{-1}\boldsymbol{W}_i'\boldsymbol{y}_i$ is the least squares estimator of $\boldsymbol{\theta}_i$. The moment conditions in Assumption 4 can be relaxed when $\boldsymbol{w}_{it}$ is strictly exogenous. From covariance-stationarity in Assumption 3, we have $\|\boldsymbol{Q}_{iT}-\boldsymbol{Q}_i\|=O_p(T^{-1/2})$ and it is possible to show that there exists a finite T_0 such that for all $T>T_0$ conditions (2) and (4) will be met.

We next introduce assumptions that are required primarily for establishing the properties of pooled and fixed effects predictors.

Assumption 5. *(a)* $\boldsymbol{\theta}_i=\boldsymbol{\theta}+\boldsymbol{\eta}_i$ *with* $\|\boldsymbol{\theta}\|<C$, $\mathrm{E}\|\boldsymbol{\eta}_i\|<C$, $\mathrm{E}(\boldsymbol{\eta}_i)=0$, $\mathrm{E}(\boldsymbol{\eta}_i\boldsymbol{\eta}_i')=\boldsymbol{\Omega}_\eta$, *and* $\|\boldsymbol{\Omega}_\eta\|<C$. *(b) Let* $\boldsymbol{q}_{it}=\boldsymbol{w}_{it}\boldsymbol{w}_{it}'\boldsymbol{\eta}_i$, *then* $\mathrm{E}(\boldsymbol{q}_{it})=\boldsymbol{q}_i$ *(fixed),* $\sup_i\|\boldsymbol{q}_i\|<C$, $\sup_{i,t}\mathrm{E}\|\boldsymbol{q}_{it}\|^2<C$, *and* $\sup_i\mathrm{E}\left\|\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\right\|^2<C$.

Assumption 6. $\boldsymbol{\eta}_i$ *is distributed independently of* $\boldsymbol{\varepsilon}_i$, *for all* i.

Assumption 7. $\bar{\boldsymbol{\xi}}_{NT}=N^{-1}\sum_{i=1}^N\boldsymbol{\xi}_{iT}=O_p\left(N^{-1/2}\right)$, *where* $\boldsymbol{\xi}_{iT}=T^{-1}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i=T^{-1}\sum_{t=1}^T\boldsymbol{w}_{it}\varepsilon_{it}$.

Assumption 8. *There exists a fixed* T_0 *such that for all* $T>T_0$ *and* $N=1,2,\ldots,$ *the pooled covariance matrices* $\bar{\boldsymbol{Q}}_{NT}$ *and* $\bar{\boldsymbol{Q}}_N$, *defined in terms of* $\boldsymbol{Q}_{iT}=T^{-1}\boldsymbol{W}_i'\boldsymbol{W}_i$ *and* $\boldsymbol{Q}_i=\mathrm{E}(\boldsymbol{Q}_{iT})$,

$$\bar{\boldsymbol{Q}}_{NT}=N^{-1}\sum_{i=1}^N\boldsymbol{Q}_{iT},\ \textit{and}\ \bar{\boldsymbol{Q}}_N=\mathrm{E}\left(\bar{\boldsymbol{Q}}_{NT}\right)=N^{-1}\sum_{i=1}^N\boldsymbol{Q}_i, \tag{6}$$

are positive definite, $\left\|\bar{\boldsymbol{Q}}_N^{-1}\right\|<C$, *and*

$$\sup_{N,T}\mathrm{E}\left[\lambda_{\max}^2\left(\bar{\boldsymbol{Q}}_{NT}\right)\right]<C<\infty,\ \textit{and}\ \sup_{N,T}\mathrm{E}\left[\lambda_{\max}^2\left(\bar{\boldsymbol{Q}}_{NT}^{-1}\right)\right]<C<\infty.$$

Assumption 9. $(\boldsymbol{\varepsilon}_i,\boldsymbol{W}_i,\boldsymbol{\eta}_i)$ *are distributed independently over* i.

For pooled estimation of $\boldsymbol{\theta}$, the conditions on $\boldsymbol{Q}_{iT}$ can be relaxed and it is sufficient that $\bar{\boldsymbol{Q}}_{NT}$ is positive definite, and $\sup_{N,T} \mathrm{E}\left\|\boldsymbol{Q}_{NT}^{-1}\right\|^2 < C$. Assumptions 5 and 6 identify the population mean of $\boldsymbol{\theta}_i$ denoted by $\boldsymbol{\theta}$, but allow for correlated heterogeneity.[2] The degree of parameter heterogeneity is measured by the norm of $\boldsymbol{\Omega}_\eta$, and the extent to which heterogeneity is correlated is measured by the norm of $\boldsymbol{q}_i$.[3]

Assumptions 5–8 are not required for forecasts based on the individual estimates and the associated MSFE. The assumption of cross-sectional independence for ε_{it} (or w_{it}) is not needed to establish results on the MSFE of individual forecasts. However, we do require some degree of uncorrelatedness over i when the objective is to compute the MSFE averaged across all N units under consideration or over a sub-group of the units. In particular, to ensure that the cross-sectional average MSFE tends to a non-random limit, the units under consideration must satisfy the law of large numbers. To this end, we need the units to be cross-sectionally weakly correlated, possibly conditional on known (or estimated) common factors. The situation is different when we consider pooled or Bayesian forecasts. Optimality of these forecasts *does* depend on the assumption of cross-sectional independence, or at least some form of weak cross-sectional dependence. A comprehensive analysis of the implications of cross-sectional dependence for forecast combinations and comparisons of predictive accuracy are beyond the scope of the present paper, however.

2.3 Correlated heterogeneity

We measure the degree of correlated heterogeneity for unit i at time t by $\boldsymbol{q}_i = \mathrm{E}\left(\boldsymbol{w}_{it}\boldsymbol{w}_{it}'\boldsymbol{\eta}_i\right)$ and, on average, by

$$\bar{\boldsymbol{q}}_{NT} = N^{-1}T^{-1}\sum_{i=1}^{N}\boldsymbol{W}_i'\boldsymbol{W}_i\boldsymbol{\eta}_i = N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{w}_{it}\boldsymbol{w}_{it}'\boldsymbol{\eta}_i. \tag{7}$$

Taking expectations,

$$\mathrm{E}\left(\bar{\boldsymbol{q}}_{NT}\right) = \bar{\boldsymbol{q}}_N = N^{-1}\sum_{i=1}^{N}\boldsymbol{q}_i. \tag{8}$$

[2]Here we are simplifying the notations and use $\boldsymbol{\theta}$, rather than $\boldsymbol{\theta}_0$, to denote the population mean which is technically more appropriate.

[3]Under Assumption 2, $\mathrm{E}\left(\boldsymbol{\xi}_{iT}\right) = T^{-1}\sum_{t=1}^{T}\mathrm{E}\left(\boldsymbol{w}_{it}\varepsilon_{it}\right) = \boldsymbol{0}$ and $\mathrm{E}\left(\boldsymbol{\xi}_{NT}\right) = \boldsymbol{0}$. Note that ε_{it} and $\boldsymbol{w}_{it}$ are uncorrelated but not independently distributed. We also note that under Assumption 3, $\left\|\bar{\boldsymbol{Q}}_{NT}\right\| \leq \sup_i \left\|\boldsymbol{Q}_{iT}\right\| < C$, and $\left\|\bar{\boldsymbol{Q}}_N\right\| \leq \sup_i \left\|\boldsymbol{Q}_i\right\| < C$.

Assumptions 5 and 6 accommodate correlated heterogeneity and allow for non-zero values of $\mathrm{E}\left(\boldsymbol{W}_i'\boldsymbol{W}_i\boldsymbol{\eta}_i\right)$. In the context of fixed effects models, the intercepts α_i in (1) are allowed to have non-zero correlation with the regressors, but optimality of forecasts based on pooled estimates of $\boldsymbol{\beta}$ requires Assumption 6 and the condition $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\mathrm{E}\left(\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{X}_i\boldsymbol{\eta}_{i\beta}\right)=\mathbf{0}$, where $\boldsymbol{\eta}_{i\beta}=\boldsymbol{\beta}_i-\boldsymbol{\beta}$, $\boldsymbol{M}_T=\boldsymbol{I}_T-\boldsymbol{\tau}_T\left(\boldsymbol{\tau}_T'\boldsymbol{\tau}_T\right)^{-1}\boldsymbol{\tau}_T'$, $\boldsymbol{\tau}_T$ is a $T\times 1$ vector of ones, and $\boldsymbol{I}_T$ is a $T\times T$ identity matrix.[4]

3 Theoretical results on forecasting performance

We next use the setup and assumptions from Section 2 to establish theoretical results on the forecasting performance of different modeling approaches. Section 3.1 discusses forecasts based on individual and pooled estimation and, building on this, Section 3.2 covers fixed effects forecasts.

We note that our theoretical framework can also be applied to forecasts across groups instead of individuals, when there are *a priori* known groups such as industries or states within a given country. Pooled regressions can be applied to any given, a priori known group, so long as the number of units within the group is sufficiently large and the cross-sectional dependence of units within the group is sufficiently weak. Failure of the latter assumption implies that there are missing pervasive (strong) common factors that must also be taken into account. Such extensions are beyond the scope of the present paper and are topics for future research.

3.1 Forecasts based on individual and pooled estimation

Suppose we are interested in forecasting $y_{i,T+1}$ conditional on the information known at time T which we denote by $\boldsymbol{w}_{i,T+1}$ to clarify the correspondence to $y_{i,T+1}$. Without loss of generality, given the conditional nature of the forecasting exercise, we are assuming that $\sup_{i,T}\|\boldsymbol{w}_{i,T+1}\|<C$.[5] Forecasts based on individual estimators take the form

$$\hat{y}_{i,T+1}=\hat{\boldsymbol{\theta}}_i'\boldsymbol{w}_{i,T+1},\ i=1,2,\ldots,N, \tag{9}$$

[4] See Pesaran and Yang (2023). Note that $\mathrm{E}\left(\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{X}_i\boldsymbol{\eta}_{i\beta}\right)=\mathbf{0}$, is sufficient but not necessary for the validity of fixed effects estimation. This condition is not met if x_{it} includes lagged values of y_{it}, even if $T\to\infty$.

[5] See part (a) of Assumption 3.

where $\hat{\boldsymbol{\theta}}_i = (\boldsymbol{W}_i'\boldsymbol{W}_i)^{-1}\boldsymbol{W}_i'\boldsymbol{y}_i$, is the least squares estimator of $\boldsymbol{\theta}_i$. Similarly, forecasts based on the pooled estimator are given by

$$\tilde{y}_{i,T+1} = \tilde{\boldsymbol{\theta}}'\boldsymbol{w}_{i,T+1},\ i = 1, 2, \ldots, N, \tag{10}$$

where $\tilde{\boldsymbol{\theta}} = (\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{y}$. Using (6), (7), and the definition of $\bar{\boldsymbol{\xi}}_{NT}$ in Assumption 7,

$$\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_i = -\boldsymbol{\eta}_i + \bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT} + \bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{\xi}}_{NT}. \tag{11}$$

Forecast errors from these schemes take the form

$$\hat{e}_{i,T+1} = y_{iT+1} - \hat{y}_{i,T+1} = \varepsilon_{i,T+1} - (\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)'\boldsymbol{w}_{i,T+1}, \tag{12}$$

$$\tilde{e}_{i,T+1} = y_{iT+1} - \tilde{y}_{i,T+1} = \varepsilon_{i,T+1} - (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_i)'\boldsymbol{w}_{i,T+1}. \tag{13}$$

Forecasts based on individual estimation

Noting that $(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)'\boldsymbol{w}_{i,T+1} = \boldsymbol{\varepsilon}_i'\boldsymbol{W}_i(\boldsymbol{W}_i'\boldsymbol{W}_i)^{-1}\boldsymbol{w}_{i,T+1}$, it is easily seen that the forecasts based on the individual estimates result in the following average MSFE:

$$N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}^2 = N^{-1}\sum_{i=1}^{N}\varepsilon_{i,T+1}^2 + T^{-1}S_{NT} - 2R_{NT}, \tag{14}$$

where $S_{NT} = N^{-1}\sum_{i=1}^{N} s_{iT}$, $R_{NT} = N^{-1}\sum_{i=1}^{N} r_{iT}$, with elements

$$r_{iT} = \left(\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i(\boldsymbol{W}_i'\boldsymbol{W}_i)^{-1}\boldsymbol{w}_{i,T+1}\right)\varepsilon_{i,T+1}, \tag{15}$$

$$s_{iT} = \boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_{iT}^{-1}\left(T^{-1}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i\right)\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}. \tag{16}$$

It is now easily seen that, under Assumptions 1 and 3, $\mathrm{E}(r_{iT}) = 0$ and $\sup_{i,T}\mathrm{E}|r_{iT}| < C$. Similarly, $\sup_{i,T}\mathrm{E}|s_{iT}| < C$ and

$$\mathrm{E}(s_{iT}) = \mathrm{E}\left[\boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_{iT}^{-1}\left(\frac{\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i}{T}\right)\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}\right].$$

Therefore, under cross-sectional independence (Assumption 9), $R_{NT} = O_p(N^{-1/2})$, and $S_{NT} = \mathrm{E}(S_{NT}) + O_p(N^{-1/2})$, and we obtain the results summarized in the following proposition for the average MSFE of the forecasts based on the individual estimates (for a detailed proof see Section A.2.1 of the Appendix):

Proposition 1. *Suppose that Assumptions 1–4 and 9 hold. Then for a fixed T_0 such that $T > T_0$: (a) the average MSFE resulting from individual-specific estimation of the parameters has the following representation*

$$N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}^2 = N^{-1}\sum_{i=1}^{N}\varepsilon_{i,T+1}^2 + T^{-1}h_{NT} + O_p(N^{-1/2}) + O_p\left(N^{-1/2}T^{-1}\right), \tag{17}$$

where

$$h_{NT} = N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[\boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_{iT}^{-1}\left(\frac{\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i}{T}\right)\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}\right], \tag{18}$$

$\boldsymbol{Q}_{iT} = T^{-1}\boldsymbol{W}_i'\boldsymbol{W}_i$, $h_{NT} > 0$, and $h_{NT} = O(1)$.

(b) If $\boldsymbol{W}_i$ is strictly exogenous, h_{NT} simplifies to

$$h_{NT} = N^{-1}\sum_{i=1}^{N}\sigma_i^2\mathrm{E}\left(\boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}\right). \tag{19}$$

The h_{NT} term captures the cost associated with the error in estimation of $\hat{\boldsymbol{\theta}}_i$. For typical panel data sets, T is not large and parameter estimation uncertainty captured by the $O\left(T^{-1}\right)$ term $T^{-1}h_{NT}$ in (17) can therefore be important. Parameter heterogeneity, in contrast, does not affect the accuracy of the forecast in (17). The magnitude of h_{NT} plays an important role in comparisons of forecasts based on individual and pooled estimates and depends on how far the predictors are from their mean. As an example, consider the simple case where $\boldsymbol{w}_{it} = (1, x_{it})'$ and x_{it} is strictly exogenous. Then

$$h_{NT} = \bar{\sigma}_N^2 + N^{-1}\sum_{i=1}^{N}\sigma_i^2\mathrm{E}\left[\frac{(x_{i,T+1} - \bar{x}_{iT})^2}{s_{iT}^2}\right],$$

where $\bar{\sigma}_N^2 = N^{-1}\sum_{i=1}^{N}\sigma_i^2$, $s_{iT}^2 = T^{-1}\sum_{t=1}^{T}(x_{it} - \bar{x}_{iT})^2$, and $\bar{x}_{iT} = T^{-1}\sum_{t=1}^{T}x_{it}$. It is clear that

h_{NT} is minimized when $x_{i,T+1} = \bar{x}_{iT}$, for all i. But in general where $x_{i,T+1} \neq \bar{x}_{iT}$ for most i, then we must have T sufficiently large such that $sup_i \mathrm{E}\left[\left(x_{i,T+1} - \bar{x}_{iT}\right)^2 / s_{iT}^2\right] < C$.

Forecasts based on pooled estimation

While the forecast accuracy results for the individual regressions do not depend on the degree of parameter heterogeneity, whether correlated or not, the degree of correlated heterogeneity does matter for consistency of the pooled estimator. Using (11) in (13) we can express the squared forecast error when pooled estimates are used as follows:

$$\tilde{e}_{i,T+1}^2 = \varepsilon_{i,T+1}^2 + \boldsymbol{w}_{i,T+1}' \boldsymbol{d}_{i,NT} \boldsymbol{d}_{i,NT}' \boldsymbol{w}_{i,T+1} - 2\boldsymbol{d}_{i,NT}' \boldsymbol{w}_{i,T+1} \varepsilon_{i,T+1},$$

where $\boldsymbol{d}_{i,NT} = -\boldsymbol{\eta}_i + \bar{\boldsymbol{Q}}_{NT}^{-1} \bar{\boldsymbol{q}}_{NT} + \bar{\boldsymbol{Q}}_{NT}^{-1} \bar{\boldsymbol{\xi}}_{NT}$, $\bar{\boldsymbol{Q}}_{NT}$ and $\bar{\boldsymbol{q}}_{NT}$ are defined by (6) and(7), and $\bar{\boldsymbol{\xi}}_{NT}$ is defined under Assumption 7. After some algebra, and averaging over i, we have

$$\begin{aligned} N^{-1} \sum_{i=1}^{N} \tilde{e}_{i,T+1}^2 &= N^{-1} \sum_{i=1}^{N} \varepsilon_{i,T+1}^2 + N^{-1} \sum_{i=1}^{N} \boldsymbol{w}_{i,T+1}' \boldsymbol{\eta}_i \boldsymbol{\eta}_i' \boldsymbol{w}_{i,T+1} \\ &\quad + \tilde{S}_{N,T+1} + 2\tilde{R}_{N,T+1}, \end{aligned} \tag{20}$$

where $\tilde{S}_{N,T+1}$, and $\tilde{R}_{N,T+1}$ are defined by equations (A.10) and (A.11) in Section A.2.2 of the Appendix. It can be shown that $\tilde{R}_{N,T+1} = O_p(N^{-1/2})$, and $\tilde{S}_{N,T+1} = -\bar{\boldsymbol{q}}_N' \bar{\boldsymbol{Q}}_N^{-1} \bar{\boldsymbol{q}}_N + O_p\left(N^{-1/2}\right)$, where $\bar{\boldsymbol{Q}}_N$ and $\bar{\boldsymbol{q}}_N$ are defined by (6) and (8), respectively. The limiting properties of the average MSFE based on pooled estimates are summarized in the following proposition.

Proposition 2. *(a) Under Assumptions 1–9, the MSFE for the forecasts based on pooled estimation of the parameters is*

$$N^{-1} \sum_{i=1}^{N} \tilde{e}_{i,T+1}^2 = N^{-1} \sum_{i=1}^{N} \varepsilon_{i,T+1}^2 + \Delta_{NT} + O_p(N^{-1/2}), \tag{21}$$

where

$$\Delta_{NT} = N^{-1} \sum_{i=1}^{N} \mathrm{E}\left(\boldsymbol{w}_{i,T+1}' \boldsymbol{\eta}_i \boldsymbol{\eta}_i' \boldsymbol{w}_{i,T+1}\right) - \bar{\boldsymbol{q}}_N' \bar{\boldsymbol{Q}}_N^{-1} \bar{\boldsymbol{q}}_N. \tag{22}$$

(b) In the special case of uncorrelated heterogeneity, the MSFE simplifies to

$$N^{-1}\sum_{i=1}^{N}\tilde{e}_{i,T+1}^{2}=N^{-1}\sum_{i=1}^{N}\varepsilon_{i,T+1}^{2}+\text{tr}(\bar{\boldsymbol{Q}}_{N}\boldsymbol{\Omega}_{\eta})+O_{p}(N^{-1/2}). \tag{23}$$

(c) Parameter heterogeneity (whether correlated or uncorrelated) increases the MSFE of the forecasts based on the pooled estimator, namely $\Delta_{NT}>0$.

Parts (a) and (b) of Proposition 2 are established in Appendix A.2.2. To establish part (c) note that the first term of Δ_{NT}, $N^{-1}\sum_{i=1}^{N}\text{E}\left(\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_{i}\boldsymbol{\eta}_{i}'\boldsymbol{w}_{i,T+1}\right)=N^{-1}\sum_{i=1}^{N}\text{E}\left(\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_{i}\right)^{2}\geq 0$, and arises irrespective of whether heterogeneity is correlated or not. The second term, $\bar{\boldsymbol{q}}_{N}\bar{\boldsymbol{Q}}_{N}^{-1}\bar{\boldsymbol{q}}_{N}'$, enters only if heterogeneity is correlated. The balance of the two terms, namely Δ_{NT}, can be signed under stationarity where $\text{E}\left(\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_{i}\boldsymbol{\eta}_{i}'\boldsymbol{w}_{i,T+1}\right)=\text{E}\left(\boldsymbol{w}_{it}'\boldsymbol{\eta}_{i}\boldsymbol{\eta}_{i}'\boldsymbol{w}_{it}\right)$. In this case, we have

$$\Delta_{N}=N^{-1}\sum_{i=1}^{N}\text{E}\left(\boldsymbol{w}_{it}'\boldsymbol{\eta}_{i}\boldsymbol{\eta}_{i}'\boldsymbol{w}_{it}\right)-\bar{\boldsymbol{q}}_{N}'\bar{\boldsymbol{Q}}_{N}^{-1}\bar{\boldsymbol{q}}_{N}. \tag{24}$$

To establish that the net effect of the two terms in Δ_{NT} is non-negative, we first show that the sample estimate of Δ_{NT} can be obtained as the sum of squares of the residuals from the pooled panel regression of $\boldsymbol{\eta}_{i}'\boldsymbol{w}_{it}$ on $\boldsymbol{w}_{it}$. Consider the panel regression $\boldsymbol{\eta}_{i}'\boldsymbol{w}_{it}=\boldsymbol{\gamma}'\boldsymbol{w}_{it}+\nu_{it}$, and note that the pooled estimator of $\boldsymbol{\gamma}$ is given by

$$\hat{\boldsymbol{\gamma}}_{NT}=\left(N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{w}_{it}\boldsymbol{w}_{it}'\right)^{-1}N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{w}_{it}\boldsymbol{w}_{it}'\boldsymbol{\eta}_{i}=\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT},$$

which yields the residual sum of squares

$$N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\nu}_{it}^{2}=N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{\eta}_{i}'\boldsymbol{w}_{it}-\hat{\boldsymbol{\gamma}}_{NT}'\boldsymbol{w}_{it}\right)^{2}=\hat{\Delta}_{NT},$$

and, by construction, $\hat{\Delta}_{NT}$ is non-negative and is given by

$$\hat{\Delta}_{NT}=T^{-1}N^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\boldsymbol{w}_{it}'\boldsymbol{\eta}_{i}\boldsymbol{\eta}_{i}'\boldsymbol{w}_{it}-\bar{\boldsymbol{q}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT}\geq 0.$$

This result also holds for a fixed T and as $N \to \infty$ (applying Slutsky's theorem to the second term):

$$\lim_{N\to\infty} \Delta_{NT} = \operatorname*{plim}_{N\to\infty} N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\nu}_{it}^2 \geq 0.$$

The impact on the MSFE from neglected heterogeneity, $\hat{\Delta}_{NT}$, does not vanish even if both N and $T \to \infty$. This is in line with the early results obtained in Pesaran and Smith (1995) who established the large N and T inconsistency of fixed effects estimators of heterogeneous dynamic panels because heterogeneity is always correlated in dynamic panels.[6]

A comparison of forecasts based on individual and pooled estimates

Next, we consider the difference in the average MSFE performance of the forecasts based on the pooled versus individual parameter estimates. Proposition 1 shows that the MSFE from the forecasts based on the individual estimates will be affected by an estimation error term of the form

$$h_{NT} = N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[\boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_{iT}^{-1}\left(\frac{\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i}{T}\right)\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}\right] > 0. \tag{25}$$

While the forecasts from the pooled estimates are more robust to estimation errors, they are in turn affected by correlated and uncorrelated heterogeneity as captured by the term

$$\Delta_{NT} = N^{-1}\sum_{i=1}^{N}\mathrm{E}\left(\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\boldsymbol{w}_{i,T+1}\right) - \bar{\boldsymbol{q}}_N'\bar{\boldsymbol{Q}}_N^{-1}\bar{\boldsymbol{q}}_N. \tag{26}$$

We compare the difference in the average MSFE of the forecasts from the pooled versus individual estimates as a ratio measured relative to the MSFE of the forecasts from the individual estimates as a benchmark:

$$\begin{aligned} & \frac{N^{-1}\sum_{i=1}^{N}\tilde{e}_{i,T+1}^2 - N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}^2}{N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}^2} \\ = \; & \frac{\Delta_{NT} - T^{-1}h_{NT} + O_p(N^{-1/2}) + O_p(N^{-1/2}T^{-1}).}{N^{-1}\sum_{i=1}^{N}\varepsilon_{i,T+1}^2 + T^{-1}h_{NT} + O_p(N^{-1/2})}. \end{aligned}$$

[6]This latter property is illustrated by a simple example in Section A.3 of the Appendix.

Hence, there exists a T_0 such that, for a fixed $T > T_0$, and as $N \to \infty$

$$\frac{N^{-1}\sum_{i=1}^{N}\tilde{e}_{i,T+1}^2 - N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}^2}{N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}^2} \xrightarrow{p} \frac{\Delta - T^{-1}h_T}{\bar{\sigma}^2 + T^{-1}h_T},$$

where $h_T = \lim_{N\to\infty} h_{NT} \geq 0$, $\Delta = \lim_{N\to\infty} \Delta_N \geq 0$, and $\bar{\sigma}^2 = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N}\sigma_i^2 > 0$**.** It follows that when T is fixed and N is large, generally it will not be possible to rank the two forecasting schemes. The outcome will depend on the sign and the magnitude of $\Delta - T^{-1}h_T$.[7]

For large values of T, however, we can show that the individual forecasts generate the lowest MSFE values. Specifically, for a fixed N and as $T \to \infty$

$$\frac{N^{-1}\sum_{i=1}^{N}\tilde{e}_{i,T+1}^2 - N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}^2}{N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}^2} \xrightarrow{p} \frac{\Delta_N}{\bar{\sigma}^2} + O_p(N^{-1/2}).$$

Similarly, when both N and $T \to \infty$ (in any order)

$$\frac{N^{-1}\sum_{i=1}^{N}\tilde{e}_{i,T+1}^2 - N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}^2}{N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}^2} \xrightarrow{p} \Delta/\bar{\sigma}^2 \geq 0,$$

where $\Delta = \lim_{T\to\infty}(\Delta_T)$. Therefore, on average, individual-specific estimates lead to more accurate forecasts as compared to the pooled estimates when T is sufficiently large.

3.2 Forecasts based on fixed effects estimation

The comparison of forecasts based on individual or pooled estimates can be extended to intermediate cases where a sub-set of the parameters are allowed to vary across units. A prominent example is the FE forecast

$$\hat{y}_{i,T+1}^{\text{FE}} = \hat{\alpha}_{i,\text{FE}} + \hat{\boldsymbol{\beta}}_{\text{FE}}'\boldsymbol{x}_{i,T+1}, \tag{27}$$

[7] In comparing Δ_T with $T^{-1}h_T$, it is also important to bear in mind that h_T is well defined if moments of $\hat{\boldsymbol{\theta}}_i$ (at least up to second order) exist (see the moment condition (5)). This in turn requires that $T > T_0$ for some finite T_0. The value of T_0 depends on the nature of the $(\boldsymbol{w}_{it}, \varepsilon_{it})$ process and its distributional properties.

where $\hat{\alpha}_{i,\text{FE}} = \boldsymbol{\tau}_T'(\boldsymbol{y}_i - \hat{\boldsymbol{\beta}}_{\text{FE}}'\boldsymbol{X}_i)/T$ and $\hat{\boldsymbol{\beta}}_{\text{FE}} = \left(\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{X}_i\right)^{-1}\sum_{i=1}^N \boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{y}_i$. The associated FE forecast error is given by

$$\hat{e}_{i,T+1}^{\text{FE}} = \tilde{\varepsilon}_{i,T+1} - (\hat{\boldsymbol{\beta}}_{\text{FE}} - \boldsymbol{\beta}_i)'\tilde{\boldsymbol{x}}_{i,T+1}, \tag{28}$$

where $\tilde{\varepsilon}_{i,T+1} = \varepsilon_{i,T+1} - \bar{\varepsilon}_{iT}$, $\tilde{\boldsymbol{x}}_{i,T+1} = \boldsymbol{x}_{i,T+1} - \bar{\boldsymbol{x}}_{iT}$, $\bar{\varepsilon}_{iT} = T^{-1}\sum_{t=1}^T \varepsilon_{it}$, and $\bar{\boldsymbol{x}}_{iT} = T^{-1}\sum_{t=1}^T \boldsymbol{x}_{it}$. Following the derivations for the pooled estimates, it is easily seen that

$$\hat{\boldsymbol{\beta}}_{\text{FE}} - \boldsymbol{\beta}_i = -\boldsymbol{\eta}_{i,\beta} + \bar{\boldsymbol{Q}}_{NT,\beta}^{-1}\bar{\boldsymbol{q}}_{NT,\beta} + \bar{\boldsymbol{Q}}_{NT,\beta}^{-1}\bar{\boldsymbol{\xi}}_{NT,\beta},$$

where $\boldsymbol{\eta}_{i,\beta} = \boldsymbol{\beta}_i - \boldsymbol{\beta}$, $\bar{\boldsymbol{\xi}}_{NT,\beta} = N^{-1}\sum_{i=1}^N T^{-1}\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{\varepsilon}_i$,

$$\bar{\boldsymbol{Q}}_{NT,\beta} = N^{-1}\sum_{i=1}^N T^{-1}\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{X}_i, \text{ and } \bar{\boldsymbol{q}}_{NT,\beta} = N^{-1}\sum_{i=1}^N \left(T^{-1}\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{X}_i\right)\boldsymbol{\eta}_{i,\beta}.$$

With one exception, the derivation of the average MSFE for the FE estimation closely parallels the case of the pooled estimator with $\boldsymbol{\eta}_{i,\beta}$ in place of $\boldsymbol{\eta}_i$, $\bar{\boldsymbol{Q}}_{NT,\beta}$ replacing $\bar{\boldsymbol{Q}}_{NT}$, $\bar{\boldsymbol{q}}_{NT,\beta}$ replacing $\bar{\boldsymbol{q}}_{NT}$, $\bar{\boldsymbol{\xi}}_{NT,\beta}$ replacing $\bar{\boldsymbol{\xi}}_{NT}$, and $\tilde{\boldsymbol{x}}_{i,T+1} = \boldsymbol{x}_{i,T+1} - \bar{\boldsymbol{x}}_{iT}$ in place of $\boldsymbol{x}_{i,T+1}$. The exception arises due to the fact that in the case of weakly exogenous regressors, $\bar{\varepsilon}_{iT}$ (and hence $\tilde{\varepsilon}_{i,T+1}$) is not distributed independently of $(\hat{\boldsymbol{\beta}}_{\text{FE}} - \boldsymbol{\beta}_i)'\tilde{\boldsymbol{x}}_{i,T+1}$. To account for this dependence, we first note that, under Assumption 7, $\bar{\boldsymbol{\xi}}_{NT,\beta} = O_p\left(N^{-1/2}\right)$, and

$$\begin{aligned} N^{-1}\sum_{i=1}^N \left(\hat{\boldsymbol{\beta}}_{\text{FE}} - \boldsymbol{\beta}_i\right)'\tilde{\boldsymbol{x}}_{i,T+1}\bar{\varepsilon}_{iT} &= N^{-1}\sum_{i=1}^N \left(-\boldsymbol{\eta}_{i,\beta} + \bar{\boldsymbol{Q}}_{NT,\beta}^{-1}\bar{\boldsymbol{q}}_{NT,\beta} + \bar{\boldsymbol{Q}}_{NT,\beta}^{-1}\bar{\boldsymbol{\xi}}_{NT,\beta}\right)'\tilde{\boldsymbol{x}}_{i,T+1}\bar{\varepsilon}_{iT} \\ = \quad -N^{-1}\sum_{i=1}^N \boldsymbol{\eta}_{i,\beta}'\tilde{\boldsymbol{x}}_{i,T+1}\bar{\varepsilon}_{iT} &+ \bar{\boldsymbol{q}}_{NT,\beta}'\bar{\boldsymbol{Q}}_{NT,\beta}^{-1}\left(N^{-1}\sum_{i=1}^N \tilde{\boldsymbol{x}}_{i,T+1}\bar{\varepsilon}_{iT}\right) + O_p\left(N^{-1/2}\right). \end{aligned}$$

Also, under Assumptions 4 and 9 we have

$$\frac{1}{N}\sum_{i=1}^N \left(\hat{\boldsymbol{\beta}}_{\text{FE}} - \boldsymbol{\beta}_i\right)'\tilde{\boldsymbol{x}}_{i,T+1}\bar{\varepsilon}_{iT} = c_{NT}^{\text{FE}} + O_p(N^{-1/2}). \tag{29}$$

The expression for c_{NT}^{FE} simplifies somewhat by noting that under Assumption 2, $\mathrm{E}\left(\boldsymbol{x}_{iT+1}\bar{\varepsilon}_{iT}\right)=\mathbf{0}$, and using Lemma A.1 we have $\bar{\boldsymbol{q}}'_{NT,\beta}\bar{\boldsymbol{Q}}_{NT,\beta}^{-1}=\bar{\boldsymbol{q}}'_{N,\beta}\bar{\boldsymbol{Q}}_{N,\beta}^{-1}+O_p\left(N^{-1/2}\right)$. Hence,

$$c_{NT}^{\text{FE}}=-N^{-1}\sum_{i=1}^{N}\mathrm{E}\left(\boldsymbol{\eta}'_{i,\beta}\tilde{\boldsymbol{x}}_{i,T+1}\bar{\varepsilon}_{iT}\right)+\bar{\boldsymbol{q}}'_{N,\beta}\bar{\boldsymbol{Q}}_{N,\beta}^{-1}\left[N^{-1}\sum_{i=1}^{N}\mathrm{E}\left(\bar{\boldsymbol{x}}_{iT}\bar{\varepsilon}_{iT}\right)\right]. \tag{30}$$

c_{NT}^{FE} tends to zero for T sufficiently large or if $\boldsymbol{x}_{it}$ is strictly exogenous. Note that under Assumption 6, $\boldsymbol{\eta}_{i,\beta}$ and ε_{it} are independently distributed. Using these results, the MSFE under fixed effects estimation is

$$N^{-1}\sum_{i=1}^{N}\left(\hat{e}_{i,T+1}^{\text{FE}}\right)^2=N^{-1}\sum_{i=1}^{N}\tilde{\varepsilon}_{i,T+1}^2+\Delta_{NT}^{\text{FE}}-2c_{NT}^{\text{FE}}+O_p(N^{-1/2}), \tag{31}$$

where

$$\Delta_{NT}^{\text{FE}}=N^{-1}\sum_{i=1}^{N}\mathrm{E}(\tilde{\boldsymbol{x}}'_{i,T+1}\boldsymbol{\eta}_{i,\beta}\boldsymbol{\eta}'_{i,\beta}\tilde{\boldsymbol{x}}_{i,T+1})-\bar{\boldsymbol{q}}'_{N,\beta}\bar{\boldsymbol{Q}}_{N,\beta}^{-1}\bar{\boldsymbol{q}}_{N,\beta}. \tag{32}$$

A comparison of forecasts based on individual and fixed effects estimates

To compare the FE forecast to the individual forecasts, rewrite (12) as $\hat{e}_{i,T+1}=\varepsilon_{i,T+1}-(\hat{\alpha}_i-\alpha_i)-\boldsymbol{x}'_{i,T+1}(\hat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i)$, and note that $\hat{\alpha}_i-\alpha_i=\bar{\varepsilon}_{iT}-\bar{\boldsymbol{x}}'_{iT}\left(\hat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i\right)$. Therefore,

$$\hat{e}_{i,T+1}=\tilde{\varepsilon}_{i,T+1}-\tilde{\boldsymbol{x}}'_{i,T+1}(\hat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i). \tag{33}$$

The derivation of the average MSFE, $N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}^2$ can now proceed as before, except that under weak exogeneity the two components of $\hat{e}_{i,T+1}$, in (33), are no longer independently distributed and, as in the FE estimation, we need to consider the additional term

$$\begin{aligned}
&N^{-1}\sum_{i=1}^{N}\tilde{\boldsymbol{x}}'_{i,T+1}(\hat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i)\tilde{\varepsilon}_{i,T+1}=N^{-1}\sum_{i=1}^{N}\tilde{\boldsymbol{x}}'_{i,T+1}(\boldsymbol{X}'_i\boldsymbol{M}_T\boldsymbol{X}_i)^{-1}\boldsymbol{X}'_i\boldsymbol{M}_T\boldsymbol{\varepsilon}_i\tilde{\varepsilon}_{i,T+1}\\
=\;&-N^{-1}\sum_{i=1}^{N}\tilde{\boldsymbol{x}}'_{i,T+1}(\boldsymbol{X}'_i\boldsymbol{M}_T\boldsymbol{X}_i)^{-1}\boldsymbol{X}'_i\boldsymbol{M}_T\boldsymbol{\varepsilon}_i\bar{\varepsilon}_{iT}+O_p(N^{-1/2}).
\end{aligned}$$

Using this, we have

$$N^{-1}\sum_{i=1}^{N}\tilde{\boldsymbol{x}}'_{i,T+1}(\hat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i)\tilde{\varepsilon}_{i,T+1}=c_{NT,\beta}+O_p(N^{-1/2}), \tag{34}$$

where

$$c_{NT,\beta}=N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[\tilde{\boldsymbol{x}}'_{i,T+1}(\boldsymbol{X}'_i\boldsymbol{M}_T\boldsymbol{X}_i)^{-1}\boldsymbol{X}'_i\boldsymbol{M}_T\boldsymbol{\varepsilon}_i\bar{\varepsilon}_{iT}\right]. \tag{35}$$

Taking this term into account we obtain

$$N^{-1}\sum_{i=1}^{N}\hat{e}^2_{i,T+1}=N^{-1}\sum_{i=1}^{N}\tilde{\varepsilon}^2_{i,T+1}+h_{NT,\beta}-2c_{NT,\beta}+O_p(N^{-1/2}), \tag{36}$$

where

$$h_{NT,\beta}=N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[\tilde{\boldsymbol{x}}'_{i,T+1}\boldsymbol{Q}^{-1}_{iT,\beta}\left(\frac{\boldsymbol{X}'_i\boldsymbol{M}_T\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_i\boldsymbol{M}_T\boldsymbol{X}_i}{T}\right)\boldsymbol{Q}^{-1}_{iT,\beta}\tilde{\boldsymbol{x}}_{i,T+1}\right], \tag{37}$$

and $\boldsymbol{Q}_{iT,\beta}=T^{-1}\left(\boldsymbol{X}'_i\boldsymbol{M}_T\boldsymbol{X}_i\right)$. As with the term c^{FE}_{NT} in the average MSFE of the FE forecasts, $c_{NT,\beta}=0$ when $\boldsymbol{x}_{it}$ is strictly exogenous. To see why this is so, note that in this case, $\mathrm{E}\left(\boldsymbol{\varepsilon}_i\bar{\varepsilon}_{iT}\,|\boldsymbol{X}_i\right)=(\sigma^2_i/T)\boldsymbol{\tau}_T$ and

$$\mathrm{E}\left[\tilde{\boldsymbol{x}}'_{i,T+1}(\boldsymbol{X}'_i\boldsymbol{M}_T\boldsymbol{X}_i)^{-1}\boldsymbol{X}'_i\boldsymbol{M}_T\boldsymbol{\varepsilon}_i\bar{\varepsilon}_{iT}\,|\boldsymbol{X}_i\right]=\tilde{\boldsymbol{x}}'_{i,T+1}(\boldsymbol{X}'_i\boldsymbol{M}_T\boldsymbol{X}_i)^{-1}\boldsymbol{X}'_i\boldsymbol{M}_T\mathrm{E}\left[\boldsymbol{\varepsilon}_i\bar{\varepsilon}_{iT}\,|\boldsymbol{X}_i,\tilde{\boldsymbol{x}}_{i,T+1}\right]=0,$$

so unconditionally $\mathrm{E}\left[\tilde{\boldsymbol{x}}'_{i,T+1}(\boldsymbol{X}'_i\boldsymbol{M}_T\boldsymbol{X}_i)^{-1}\boldsymbol{X}'_i\boldsymbol{M}_T\boldsymbol{\varepsilon}_i\bar{\varepsilon}_{iT}\right]=0$, and $c_{NT,\beta}=0$.

Apart from the error term, $\varepsilon_{i,T+1}-\bar{\varepsilon}_{iT}$, which is common to the individual and FE forecasts, the squared forecast errors are analogous to those in the comparison of individual and pooled forecasts except that we work with demeaned data and allow for the additional terms c^{FE}_{NT} and $c_{NT,\beta}$ if the regressors are weakly exogenous. Further, similar to the case of the individual and pooled forecasts, for T finite and N large, the ranking of the two forecasts will depend on the relative magnitudes of $\Delta^{\mathrm{FE}}_{NT}+c^{\mathrm{FE}}_{NT}$ and $h_{NT,\beta}+c_{NT,\beta}$. Also, for $T\to\infty$ the individual forecasts will be more precise than the FE forecasts.

4 Forecast combinations

We next consider approaches that combine the forecasts from Section 3 to minimize the MSFE.

4.1 Combinations of individual and pooled forecasts

Given the MSFE trade-off associated with the forecasts in (9) and (10), combining the forecasts based on the individual and pooled estimates, $\hat{y}_{i,T+1}$ and $\tilde{y}_{i,T+1}$, may be desirable. As noted in the literature (e.g., Timmermann, 2006), forecast combinations tend to perform particularly well, relative to the underlying forecasts, if the forecast errors are weakly correlated and have MSFE values of a similar magnitude. Correlations between forecast errors based on the individual and pooled estimation schemes tend to be lower for (i) greater differences in the estimates of $\boldsymbol{\theta}_i$ resulting from larger estimation errors (small T); (ii) greater heterogeneity (large $\|\boldsymbol{\Omega}_\eta\|$), and (iii) greater bias of the pooled estimator due to correlated heterogeneity.

If the level of parameter heterogeneity is either very large or very small, one of the individual or pooled estimation approaches will be dominant, reducing potential gains from forecast combination. Similarly, if T is very small but N is large and there is little parameter heterogeneity, we would expect pooled estimation to dominate individual estimation by a sufficiently large margin that forecast combination offers small, if any, gains. Conversely, if T is very large, forecasts using individual estimates will dominate forecasts using pooled estimates by a sufficient margin that renders forecast combination less attractive. Building on these observations, consider combining the two forecasts $\hat{y}_{i,T+1}$ and $\tilde{y}_{i,T+1}$ using common weights, ω, to obtain[8]

$$y^*_{i,T+1}(\omega) = \omega\hat{y}_{i,T+1} + (1-\omega)\tilde{y}_{i,T+1}, \tag{38}$$

with associated forecast error $e^*_{i,T+1}(\omega) = \omega\hat{e}_{i,T+1}+(1-\omega)\tilde{e}_{i,T+1}$. The average MSFE of the combined

[8]We focus here on a simple constant-coefficient linear combination scheme. Lahiri, Peng, and Zhao (2017) discuss a broader range of combination methods and Elliott (2017) provides an analysis of the effect on the combination weights and forecasting performance from having a large common component in the forecast errors.

forecast is given by

$$N^{-1}\sum_{i=1}^{N} e_{i,T+1}^{*2}(\omega) = \omega^2\left(N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}^2\right)+(1-\omega)^2\left(N^{-1}\sum_{i=1}^{N}\tilde{e}_{i,T+1}^2\right) +2\omega(1-\omega)\left(N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}\tilde{e}_{i,T+1}\right).$$

The value of ω that minimizes the average MSFE is therefore given by

$$\omega_{NT}^{*}=\frac{N^{-1}\sum_{i=1}^{N}\tilde{e}_{i,T+1}^2-\left(N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}\tilde{e}_{i,T+1}\right)}{\left(N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}^2\right)+\left(N^{-1}\sum_{i=1}^{N}\tilde{e}_{i,T+1}^2\right)-2\left(N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}\tilde{e}_{i,T+1}\right)}. \tag{39}$$

Approximate expressions for $N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}^2$ and $N^{-1}\sum_{i=1}^{N}\tilde{e}_{i,T+1}^2$ are given by (17) and (21), respectively. We obtain a similar expression for $N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}\tilde{e}_{i,T+1}$, with $N^{-1}\sum_{i=1}^{N}\varepsilon_{i,T+1}^2$ cancelling out from ω_{NT}^{*}. The result is summarized in the following proposition with proofs provided in Section A.2.3 of the Appendix.

Proposition 3. *(a) Under Assumptions 1–9, the optimal combination weight that minimizes the MSFE of the forecast combination in (38) is given by*

$$\omega_{NT}^{*}=\frac{\Delta_{NT}-T^{-1/2}\psi_{NT}}{\Delta_{NT}+T^{-1}h_{NT}-2T^{-1/2}\psi_{NT}}+O_p(N^{-1/2})+O_p(N^{-1/2}T^{-1/2}), \tag{40}$$

where h_{NT} and Δ_{NT} are defined in equations (25) and (26), respectively, and ψ_{NT} is given by

$$\begin{aligned}\psi_{NT}=&\left[N^{-1}\sum_{i=1}^{N}\mathrm{E}\left(T^{-1/2}\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\right)\right]\bar{\boldsymbol{Q}}_N^{-1}\bar{\boldsymbol{q}}_N \\ &-N^{-1}\sum_{i=1}^{N}\mathrm{E}\left(T^{-1/2}\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\right).\end{aligned} \tag{41}$$

(b) Under uncorrelated heterogeneity, $\psi_{NT}=0$, and Δ_{NT} and h_{NT} will be affected accordingly.

For small to moderate values of T and large N, we expect $\omega_{NT}^{*}<1$, with a non-zero weight placed on the forecasts based on the pooled estimate.

Our forecast combination scheme does not attempt to estimate weights specific to the individual units, ω_i^{*} which require estimation of $\mathrm{E}\left(\tilde{e}_{i,T+1}^2\right)$, $\mathrm{E}\left(\hat{e}_{i,T+1}^2\right)$, and $\mathrm{E}(\hat{e}_{i,T+1}\tilde{e}_{i,T+1})$ for each i separately, and their estimates will depend on on the individual estimates such as $\hat{\theta}_i$ and $\hat{\sigma}_i^2$ and thus require

large T for consistency. Instead, we use the cross-section to estimate ω^*_{NT}. This requires consistent estimation of h_{NT}, Δ_{NT}, and ψ_{NT} which is achieved under Assumption 9 and requires large N as long as $T > T_0$, where T_0 is finite. See Sub-section 4.3 below.

4.2 Combining individual and fixed effect forecasts

Combination weights can also be determined for the case where the pooled forecast is replaced with the FE forecasts. In this case, the combined forecast is given by

$$y^*_{i,T+1}(\omega_{\text{FE}}) = \omega_{\text{FE}}\hat{y}_{i,T+1} + (1-\omega_{\text{FE}})\hat{y}_{i,T+1,\text{FE}}, \tag{42}$$

yielding the optimal weight

$$\omega^*_{FE,NT} = \frac{N^{-1}\sum_{i=1}^{N}\left(\hat{e}^{\text{FE}}_{i,T+1}\right)^2 - \left(N^{-1}\sum_{i=1}^{N}\hat{e}^{\text{FE}}_{i,T+1}\hat{e}_{i,T+1}\right)}{\left(N^{-1}\sum_{i=1}^{N}\hat{e}^2_{i,T+1}\right) + N^{-1}\sum_{i=1}^{N}\left(\hat{e}^{\text{FE}}_{i,T+1}\right)^2 - 2\left(N^{-1}\sum_{i=1}^{N}\hat{e}^{\text{FE}}_{i,T+1}\hat{e}_{i,T+1}\right)}. \tag{43}$$

The expressions for $N^{-1}\sum_{i=1}^{N}\left(\hat{e}^{\text{FE}}_{i,T+1}\right)^2$ and $N^{-1}\sum_{i=1}^{N}\hat{e}^2_{i,T+1}$ are given by (31) and (36), respectively, and the expression for $N^{-1}\sum_{i=1}^{N}\hat{e}^{\text{FE}}_{i,T+1}\hat{e}_{i,T+1}$ can be similarly obtained. In this case, the shared term $\sum_{i=1}^{N}(\varepsilon_{i,T+1}-\bar{\varepsilon}_{iT})^2/N$ cancels out and we have the result summarized in the following proposition with proofs provided in Section A.2.4 of the Appendix.

Proposition 4. *(a) Under Assumptions 1–9, the optimal combination weight that minimizes the MSFE of the forecast combination in (42) is given by*

$$\omega^*_{FE,NT} = \frac{\Delta^{FE}_{NT} - T^{-1/2}\psi^{FE}_{NT} - \left(c^{FE}_{NT} - c_{NT,\beta}\right)}{\Delta^{FE}_{NT} + T^{-1}h_{NT,\beta} - 2T^{-1/2}\psi^{FE}_{NT}} + O_p(N^{-1/2}) + O_p(N^{-1/2}T^{-1/2}), \tag{44}$$

where Δ^{FE}_{NT} is defined in (32) and $h_{NT,\beta}$ in (37), respectively. Moreover,

$$\begin{aligned}\psi^{FE}_{NT} &= N^{-1}\sum_{i=1}^{N}\text{E}\left[\left(T^{-1/2}\boldsymbol{\varepsilon}_i'\boldsymbol{M}_T\boldsymbol{X}_i\right)\boldsymbol{Q}^{-1}_{iT,\beta}\left(\tilde{\boldsymbol{x}}_{i,T+1}\tilde{\boldsymbol{x}}'_{i,T+1}\right)\right]\bar{\boldsymbol{Q}}^{-1}_{N,\beta}\bar{\boldsymbol{q}}_{N,\beta} \\ &\quad -N^{-1}\sum_{i=1}^{N}\text{E}\left[\left(T^{-1/2}\boldsymbol{\varepsilon}_i'\boldsymbol{M}_T\boldsymbol{X}_i\right)\boldsymbol{Q}^{-1}_{iT,\beta}\left(\tilde{\boldsymbol{x}}_{i,T+1}\tilde{\boldsymbol{x}}'_{i,T+1}\right)\boldsymbol{\eta}_{i,\beta}\right],\end{aligned} \tag{45}$$

and c^{FE}_{NT} and $c_{NT,\beta}$ are defined by (30) and (35), respectively.

(b)Under uncorrelated heterogeneity, $\psi_{NT}^{FE}=0$, and Δ_{NT}^{FE} and $h_{NT,\beta}$ will be affected accordingly.

4.3 Estimation of combination weights

To compute the weights for the forecast combination in Proposition 3, we need estimates of h_{NT}, Δ_{NT}, and ψ_{NT}. Under Assumption 9, these terms can be estimated by their sample means with unknown parameters replaced by their estimates. Specifically, we have

$$\hat{h}_{NT}=N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}^{\prime}\boldsymbol{Q}_{iT}^{-1}\hat{\boldsymbol{H}}_{iT}\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}, \tag{46}$$

where $\hat{\boldsymbol{H}}_{iT}=T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^{2}\left(\boldsymbol{w}_{it}\boldsymbol{w}_{it}^{\prime}\right)$, and $\hat{\varepsilon}_{it}=y_{it}-\hat{\boldsymbol{\theta}}_{i}^{\prime}\boldsymbol{w}_{it},\ t=1,2,\ldots,T$,

$$\hat{\Delta}_{NT}=N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}^{\prime}\hat{\boldsymbol{\eta}}_{i}\hat{\boldsymbol{\eta}}_{i}^{\prime}\boldsymbol{w}_{i,T+1} \tag{47}$$

where $\hat{\boldsymbol{\eta}}_{i}=\hat{\boldsymbol{\theta}}_{i}-\tilde{\boldsymbol{\theta}}$. Finally, we set $\psi_{NT}=0$. We do this because errors in estimating ψ_{NT} are or order $T^{-1/2}$ so we cannot expect to obtain accurate estimates of this term in cases with small T. Since $\psi_{NT}=0$ in the absence of correlated heterogeneity, effectively this means that we ignore such effects when estimating ψ_{NT} although, of course, our theory captures such effects. We also provide Monte Carlo results that show our estimated weights are quite close to the oracle weights that make use of true parameters values, under heterogeneity. Accordingly, we use the following estimate of ω_{NT}^{*}

$$\hat{\omega}_{NT}^{*}=\frac{\hat{\Delta}_{NT}}{\hat{\Delta}_{NT}+T^{-1}\hat{h}_{NT}}, \tag{48}$$

which is guaranteed to lie in the range (0,1].

Similarly, when estimating the weight in Proposition 4 we set $\psi_{NT}^{\mathrm{FE}}=0$ and $c_{NT}^{\mathrm{FE}}-c_{NT,\beta}=0$ and estimate $\omega_{FE,NT}^{*}$ as

$$\hat{\omega}_{FE,NT}^{*}=\frac{\hat{\Delta}_{NT}^{\mathrm{FE}}}{\hat{\Delta}_{NT}^{\mathrm{FE}}+T^{-1}\hat{h}_{NT,\beta}}, \tag{49}$$

where

$$\hat{\Delta}_{NT}^{\text{FE}} = \frac{1}{N}\sum_{i=1}^{N} \tilde{\boldsymbol{x}}_{i,T+1}' \hat{\boldsymbol{\eta}}_{i,\beta} \hat{\boldsymbol{\eta}}_{i,\beta}' \tilde{\boldsymbol{x}}_{i,T+1}, \tag{50}$$

$\hat{\boldsymbol{\eta}}_{i,\beta} = \hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}}_{\text{FE}}$, $\tilde{\boldsymbol{x}}_{i,T+1} = \boldsymbol{x}_{iT+1} - \bar{\boldsymbol{x}}_{iT}$,

$$\hat{h}_{NT,\beta} = N^{-1}\sum_{i=1}^{N} \tilde{\boldsymbol{x}}_{i,T+1}' \boldsymbol{Q}_{iT,\beta}^{-1} \left(\frac{\boldsymbol{X}_i' \boldsymbol{M}_T \hat{\boldsymbol{\varepsilon}}_i \hat{\boldsymbol{\varepsilon}}_i' \boldsymbol{M}_T \boldsymbol{X}_i}{T} \right) \boldsymbol{Q}_{iT,\beta}^{-1} \tilde{\boldsymbol{x}}_{i,T+1}, \tag{51}$$

and $\hat{\varepsilon}_{it} = y_{it} - \hat{\boldsymbol{\theta}}_i' \boldsymbol{w}_{it}$.

5 Bayesian Forecasts

Bayesian panel forecasts are becoming increasingly common in empirical applications and constitute an alternative approach to the frequentist forecasts discussed so far. We consider two such approaches here, namely empirical Bayes and hierarchical Bayesian forecasts. The empirical Bayes (EB) forecast uses the estimator of Hsiao et al. (1999) and takes the form

$$\hat{y}_{i,T+1}^{EB} = \hat{\boldsymbol{\theta}}_{i,EB}' \boldsymbol{w}_{i,T+1},$$

where

$$\hat{\boldsymbol{\theta}}_{i,EB} = (\hat{\sigma}_i^{-2} \boldsymbol{W}_i' \boldsymbol{W}_i + \hat{\boldsymbol{\Omega}}_\theta^{-1})^{-1} (\hat{\sigma}_i^{-2} \boldsymbol{W}_i' \boldsymbol{y}_i + \hat{\boldsymbol{\Omega}}_\theta^{-1} \, \bar{\hat{\boldsymbol{\theta}}}), \tag{52}$$

$$\bar{\hat{\boldsymbol{\theta}}} = N^{-1}\sum_{i=1}^{N} \hat{\boldsymbol{\theta}}_i, \quad \hat{\sigma}_i^2 = (T-K)^{-1} \hat{\boldsymbol{\varepsilon}}_i' \hat{\boldsymbol{\varepsilon}}_i,$$

and

$$\hat{\boldsymbol{\Omega}}_\theta = \frac{1}{N}\sum_{i=1}^{N} (\hat{\boldsymbol{\theta}}_i - \bar{\hat{\boldsymbol{\theta}}})(\hat{\boldsymbol{\theta}}_i - \bar{\hat{\boldsymbol{\theta}}})',$$

where $\hat{\boldsymbol{\varepsilon}}_i = \boldsymbol{y}_i - \boldsymbol{W}_i \hat{\boldsymbol{\theta}}_i$, and $\hat{\boldsymbol{\theta}}_i = (\boldsymbol{W}_i' \boldsymbol{W}_i)^{-1} \boldsymbol{W}_i' \boldsymbol{y}_i$.[9]

The EB estimator can also be written as a weighted average of $\hat{\boldsymbol{\theta}}_i$, which allows for full hetero-

[9] It is necessary that $N > T$ for $\hat{\boldsymbol{\Omega}}_\theta$ to be positive definite.

geneity, and the mean group estimator, $\bar{\hat{\boldsymbol{\theta}}}$:

$$\hat{\boldsymbol{\theta}}_{i,EB} = \boldsymbol{\mathcal{W}}_{iT}\hat{\boldsymbol{\theta}}_i + (\boldsymbol{I}_k - \boldsymbol{\mathcal{W}}_{iT})\,\bar{\hat{\boldsymbol{\theta}}}, \tag{53}$$

where, recalling that $\boldsymbol{Q}_{iT} = T^{-1}\boldsymbol{W}_i'\boldsymbol{W}_i$, the weight matrix $\boldsymbol{\mathcal{W}}_{iT}$ is given by

$$\boldsymbol{\mathcal{W}}_{iT} = \left(\boldsymbol{I}_k + T^{-1}\hat{\sigma}_i^2\boldsymbol{Q}_{iT}^{-1}\,\hat{\boldsymbol{\Omega}}_\theta^{-1}\right)^{-1}, \tag{54}$$

The weights on the heterogeneous estimates are larger, the greater the degree of heterogeneity, as measured by the norm of $\hat{\boldsymbol{\Omega}}_\theta$, with $\hat{\boldsymbol{\theta}}_{i,EB} \to \hat{\boldsymbol{\theta}}_i$ as $\left\|\hat{\boldsymbol{\Omega}}_\theta\right\| \to \infty$. Also, since $\hat{\sigma}_i^2\boldsymbol{Q}_{iT}^{-1}\,\hat{\boldsymbol{\Omega}}_\theta^{-1}$ is bounded in T, $\hat{\boldsymbol{\theta}}_{i,EB}$ converges *numerically* to $\hat{\boldsymbol{\theta}}_i$, as $T \to \infty$. Hence, one would expect the EB estimator to perform well when T is relatively small and the degree of heterogeneity is not too large. Note that the EB weights vary across i unlike our forecast combination schemes which assume homogeneous weighting across all series.[10]

We also consider forecasts from the hierarchical Bayesian model of Lindley and Smith (1972). These assume $\varepsilon_{it} \sim iid\mathrm{N}(0,\sigma^2)$ with the following priors:

$$\begin{aligned}
\boldsymbol{\theta}_i &\sim \mathrm{N}(\bar{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}),\\
\bar{\boldsymbol{\theta}} &\sim \mathrm{N}(\boldsymbol{d}, \boldsymbol{S}_{\bar{\theta}}),\\
\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} &\sim \mathrm{Wishart}(\nu_\Sigma, (\nu_\Sigma \boldsymbol{S}_\Sigma)^{-1}),\\
\sigma^2 &\sim \mathrm{invGamma}(\nu_\sigma/2, \nu_\sigma s^2/2).
\end{aligned}$$

Draws from the parameter distribution are generated using the Gibbs sampler. We use proper priors that are weakly informative: $\boldsymbol{d} = \boldsymbol{0}$, $\boldsymbol{S}_{\bar{\theta}} = \boldsymbol{I}_K 10^6$, $\boldsymbol{S}_\Sigma = \boldsymbol{I}_K$, $\nu_\Sigma = K$, $\nu_\sigma = 0.1$, and $s^2 = 0.1$. This avoids the use of uninformative priors that appear to be difficult to attain in hierarchical models (Gelman, 2006). Further details are provided in Section S.3 in the Online Supplement.

[10] While the EB estimator in (52) is fully parametric, other studies pursue a non-parametric approach to the distribution of $\hat{\boldsymbol{\theta}}_i$; see, e.g., Brown and Greenshtein (2009) and Gu and Koenker (2017) and, more recently, Liu (2023) and Liu, Moon, and Schorfheide (2023).

6 Monte Carlo experiments

Having developed theory for the determinants of the predictive performance of the individual, pooled, FE, and combined forecasts, we next use Monte Carlo simulations to examine their finite-sample performance. We do so in the context of a dynamic heterogeneous panel data model. We begin with a pure panel autoregressive model and then move on to a dynamic heterogeneous panel model with an additional regressor.[11] We allow for dynamics, parameter heterogeneity, and correlations between the regressors and coefficients. We also consider the nature of the trade-off between heterogeneity and estimation uncertainty under different degrees of fit of the underlying panel regressions.

For each data generating process (DGP) we use the following forecasting methods: (1) pooled estimation, (2) individual estimation, (3) random effects, (4) fixed effects, (5) combination of individual and pooled forecasts, (6) combination of individual and FE forecasts, (7) empirical Bayes forecasts, and (8) hierarchical Bayes forecasts. We add the random effects forecasts to the experiments given their widespread use in applied work.[12]

Given our focus on large N panels, we set $N = \{50, 100, 1000\}$ and consider different time dimensions, namely $T = \{20, 50, 100\}$, for all MC experiments. The values of the parameters used in the simulations are reported in Table S.3 in Appendix S.2.

6.1 Panel AR model

Our first DGP is a panel autoregressive model given by

$$y_{it} = \alpha_i + \beta_i y_{i,t-1} + \varepsilon_{it}, \quad i = 1, 2, \ldots, N, \; t = 1, 2, \ldots, T + 1, \tag{55}$$

where $\varepsilon_{it} = \sigma_i(z_{it}^2 - 1)/\sqrt{2}$ with $z_{it} \sim iid\mathrm{N}(0,1)$, $\sigma_i^2 \sim 0.5 + 0.5\chi_1^2$, $\alpha_i \sim \mathrm{N}(\alpha_{0i}, \sigma_\alpha^2)$, $\beta_i = \beta_0 + \eta_{i\beta}$, $\eta_{i\beta} \sim \mathrm{Uniform}(-a_\beta/2, a_\beta/2)$, and a_β, together with N and T, are parameters that are varied over the Monte Carlo experiments. We examine four settings:

- $\alpha_{0i} = 1$, and $\sigma_\alpha^2 = a_\beta = 0$ (homogeneous case);
- $\alpha_{0i} = 2/3$ if $i \leq N/2$, $\alpha_{0i} = 4/3$ if $i > N/2$, $\sigma_\alpha^2 = 0.5$, $a_\beta = 0$ (heterogeneous fixed effects);

[11] Further analytical results for the panel AR(1) model is provided in Section A.3 of the Appendix.
[12] Additional results for equal weighted combinations and oracle weights are in Section S.4 of the Online Supplement.

- $\alpha_{0i} = 2/3$ if $i \leq N/2$, $\alpha_{0i} = 4/3$ if $i > N/2$, $\sigma_\alpha^2 = 0.5$, and $a_\beta = 0.5$ (medium heterogeneity);
- $\alpha_{0i} = 2/3$ if $i \leq N/2$, $\alpha_{0i} = 4/3$ if $i > N/2$, $\sigma_\alpha^2 = 1$, and $a_\beta = 1$ (strong heterogeneity).

The values of β_0 (reported in Table S.3) are calibrated such that we achieve values of the pooled R^2 (PR^2) (further described in Section S.2 of the Online Supplement) of approximately 0.2 and 0.6. To ensure that $\mathrm{E}\left(\frac{1}{1-\beta_i^2}\right) < C$, it is required that $0 \leq a_\beta < 2(1 - |\beta_0|)$.[13]

We initialize the DGP at $T = 0$ and draw y_{i0} from a normal distribution with mean $\kappa\alpha_i$ and variance $\sigma_i^2/(1 - \beta_i^2)$, where κ is a constant. When κ is not unity, departures from a stationary distribution of y_{it} could be important especially when T is small. Results for $\kappa \neq 1$ are provided in the Online Supplement as they remain qualitatively identical to those for $\kappa = 1$.

6.2 Panel ARX model

Our second DGP adds a regressor, x_{it}, to the panel AR(1) model:

$$y_{it} = \alpha_i + \beta_i y_{i,t-1} + \gamma_i x_{it} + \varepsilon_{it}. \tag{56}$$

Again $\varepsilon_{it} = \sigma_i(z_{it}^2 - 1)/\sqrt{2}$ with $z_{it} \sim iid\mathrm{N}(0,1)$, and $\sigma_i^2 \sim \mathrm{iid}\left(1 + \chi_1^2\right)/2$. The DGP for x_{it} is

$$x_{it} = \mu_{xi} + \xi_{it}, \tag{57}$$

where

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

$\mu_{xi} = (z_i^2 - 1)/\sqrt{2}$, $z_i \sim \mathrm{iidN}\left(0,1\right)$, and $\sigma_{xi}^2 \sim \mathrm{iid}\left(1 + \chi_1^2\right)/2$, for individual units $i = 1, 2, \ldots, N$, and observation periods t. The autocorrelation coefficient of x_{it} is $\rho_{xi} \sim \mathrm{iid\,Uniform}(0, 0.95)$, thus allowing for a high degree of dynamic heterogeneity in the regressors.

As for the pure panel AR model, the coefficients of the lagged dependent variables, $y_{i,t-1}$, are generated as $\beta_i = \beta_0 + \eta_{i\beta}$, with $\eta_{i\beta} \sim \mathrm{iidUniform}(-a_\beta/2, a_\beta/2)$, and as before $0 \leq a_\beta < 2(1 - |\beta_0|)$.

[13] See equations(S.9) and (S.10) in Section A.3 of the Appendix. Also note that $\mathrm{E}\left(\beta_i\right) = \beta_0$ and $\mathrm{Var}\left(\beta_i\right) = a_\beta^2/24$.

To allow for correlated heterogeneity, we set

$$\alpha_i = \alpha_{0i} + \phi\mu_{xi} + \sigma_\eta\eta_i, \text{ and } \gamma_i = \gamma_{0i} + \pi\mu_{xi} + \sigma_\zeta\zeta_i, \tag{58}$$

where $\eta_i, \zeta_i \sim \text{iidN}(0,1)$ and $\alpha_0 = \text{E}(\alpha_i) = \alpha_{0i} + \phi\text{E}(\mu_{xi}) = \alpha_{0i}$. Again, we examine four settings:

- $\alpha_{0i} = 1$, $\gamma_{0i} = 0.1$ and $\sigma_\alpha^2 = \sigma_\gamma^2 = a_\beta = 0$ (homogeneity);
- $\alpha_{0i} = 2/3$ if $i \leq N/2$, $\alpha_{0i} = 4/3$ if $i > N/2$, $\sigma_\alpha^2 = 0.5$, $\gamma_{0i} = 0.1$, and $\sigma_\gamma^2 = a_\beta = 0$
- $\alpha_{0i} = 2/3$ if $i \leq N/2$, $\alpha_{0i} = 4/3$ if $i > N/2$, $\sigma_\alpha^2 = 0.5$, $\gamma_{0i} = 0.2/3$ if $i \leq N/2$, $\gamma_{0i} = 0.4/3$ if $i > N/2$, $\sigma_\gamma^2 = 0.1$, and $a_\beta = 0.5$
- $\alpha_{0i} = 2/3$ if $i \leq N/2$, $\alpha_{0i} = 4/3$ if $i > N/2$, $\sigma_\alpha^2 = 1$, $\gamma_{0i} = 0.2/3$ if $i \leq N/2$, $\gamma_{0i} = 0.4/3$ if $i > N/2$, $\sigma_\gamma^2 = 0.2$, and $a_\beta = 1$

Note that non-zero correlations need not bias the pooled estimates. What matters for these is the correlation between $y_{i,t-1}^2$ and x_{it}^2 and the individual coefficients.

Using (57) and (58) we have

$$\text{E}\left[x_{it}\left(\gamma_i - \gamma_0\right)\right] = \text{E}\left[\left(\mu_{xi} + \xi_{it}\right)\left(\pi\mu_{xi} + \sigma_\zeta\zeta_i\right)\right] = \pi\text{E}\left(\mu_{xi}^2\right) \neq 0,$$

$$\begin{aligned}\text{E}\left[x_{it}^2\left(\gamma_i - \gamma_0\right)\right] &= \text{E}\left[\left(\mu_{xi} + \xi_{it}\right)^2\left(\pi\mu_{xi} + \sigma_\zeta\zeta_i\right)\right] \\ &= \pi\text{E}\left(\mu_{xi}^3\right).\end{aligned}$$

Therefore, $\text{E}\left[x_{i,t-1}^2\left(\gamma_i - \gamma_0\right)\right] = 0$ if μ_{xi} are draws from a symmetric distribution around 0. To rule out this possibility, we draw μ_{xi} from a chi-square distribution.

To control the degree of correlated heterogeneity, we first note that (taking expectations with respect to both i and t)

$$\begin{aligned}&\text{E}\left(\gamma_i\right) = \gamma_0, \quad \text{Var}(\gamma_i) = \pi^2 + \sigma_\zeta^2, \\ &\text{E}\left(x_{it}\right) = \text{E}\left(\mu_{xi} + \xi_{it}\right) = 0, \quad \text{Var}\left(x_{it}\right) = \text{E}\left(x_{it} - \mu_{xi}\right)^2 = \sigma_{xi}^2,\end{aligned}$$

and $\mathrm{E}\left[\mathrm{Var}\left(x_{it}\right)\right] = \mathrm{E}\left(1+\chi_1^2\right)/2 = 1$. Also, since ν_{it} is distributed independently of η_j and ζ_j for all t, i and j, $\mathrm{Cov}\left(\gamma_i, x_{it}\right) = \pi$ and $\mathrm{Corr}\left(\gamma_i, x_{it}\right) = \pi\left(\sigma_\zeta^2 + \pi^2\right)^{-1/2}$. To achieve a given level of $\mathrm{Corr}(\gamma_i, x_{it}) = \rho_{\gamma x}$, we set

$$\pi = \frac{\rho_{\gamma x}\sigma_\zeta}{\left(1-\rho_{\gamma x}^2\right)^{1/2}}. \tag{59}$$

Similarly, to achieve $\mathrm{Corr}(\alpha_i, x_{i,t-1}) = \rho_{\alpha x}$, we set

$$\phi = \frac{\rho_{\alpha x}\sigma_\eta}{(1-\rho_{\alpha x}^2)^{1/2}}. \tag{60}$$

Defining $\sigma_\gamma^2 = \mathrm{Var}(\gamma_i) = \pi^2 + \sigma_\zeta^2$, we can use (59) to see that $\pi = \rho_{\gamma x}\sigma_\gamma$. An equivalent result emerges for ϕ where, for $\sigma_\alpha^2 = \mathrm{Var}(\alpha_i)$, we have $\phi = \rho_{\alpha x}\sigma_\alpha$. We thus use the parameters σ_α^2, σ_γ^2, and a_β to vary the degree of parameter heterogeneity in α_i, γ_i and β_i, respectively.

For comparability, we use the values of a_β from the pure panel AR model in Section 6.1. Note, however, that $PR^2_{ARX} > PR^2_{AR}$ whenever $\mathrm{E}(\gamma_i^2) \neq 0$ since

$$PR^2_{ARX} = PR^2_{AR} + \frac{\mathrm{E}(\gamma_i^2)(1-PR^2_{AR})}{1+\mathrm{E}(\gamma_i^2)}$$

for details see Section S.2 of the Online Supplement.

We initialize y_{i0} as $y_{i0} \sim \text{iidN}\left(\kappa\mu_{iy0}, \sigma^2_{iy0}\right)$, with

$$\mu_{iy0} = \frac{\alpha_i + \gamma_i\mu_{xi}}{1-\beta_i^2}, \quad \sigma^2_{iy0} = \frac{\gamma_i^2\sigma^2_{xi} + \sigma_i^2}{1-\beta_i^2},$$

and set $\xi_{i0} = 0$. We consider initialization schemes both with $\kappa = 1$ and $\kappa \neq 1$. When $\kappa \neq 1$, we depart from the stationary distribution of y_{it}, which could be important when T is small. Since the results are qualitatively unchanged for these initialization schemes, we relegate them to the online supplement.

6.3 Forecasts and measures of forecast performance

The following panel models are fitted in the simulations to compute forecasts of $y_{i,T+1}$:

$$\boldsymbol{y}_i = \boldsymbol{W}_i\boldsymbol{\theta}_i + \boldsymbol{\varepsilon}_i,$$

where $\boldsymbol{y}_i = (y_{i1}, y_{i2}, y_{i3}, \ldots, y_{iT})'$, $\boldsymbol{W}_i = (\boldsymbol{\iota}_T, \boldsymbol{y}_{i,-1})$ in the case of the panel AR model and $\boldsymbol{W}_i = (\boldsymbol{\iota}_T, \boldsymbol{y}_{i,-1}, \boldsymbol{x}_i)$ in the case of the panel ARX model, $\boldsymbol{\iota}_T$ is a $T\times 1$ vector of ones, $\boldsymbol{y}_{i,-1} = (y_{i0}, y_{i1}, \ldots, y_{i,T-1})'$, $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{i,T})'$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})'$, and $\boldsymbol{\theta}_i = (\alpha_i, \beta_i, \gamma_i)'$.

The resulting forecasts are evaluated using the ratio of the average MSFE of each method measured relative to that of the reference individual forecasts

$$\text{rMSFE}_j = \frac{\frac{1}{NR}\sum_{i=1}^{N}\sum_{r=1}^{R}(y_{i,T+1,r} - \hat{y}_{i,T+1,j,r})^2}{\frac{1}{NR}\sum_{i=1}^{N}\sum_{r=1}^{R}(y_{i,T+1,r} - \hat{y}_{i,T+1,ref,r})^2},$$

where j denotes the methods—pooling, fixed effects, random effects, combination—and *ref* denotes the reference forecast which is the individual forecast. Replications are denoted by $r = 1, 2, \ldots, R$, where $R = 10,000$. An exception is the results for the hierarchical Bayesian forecasts which are based on $R = 500$ replications due to the computational intensity of this approach.

Additionally, we report quantiles $\alpha = (0.01, 0.05, 0.1, 0.5, 0.9, 0.95, 0.99)$ of the relative forecast accuracy over the N units in the panel

$$\text{quantile}_j(\alpha) = \left[\frac{\frac{1}{R}\sum_{r=1}^{R}(y_{i,T+1,r} - \hat{y}_{i,T+1,j,r})^2}{\frac{1}{R}\sum_{r=1}^{R}(y_{i,T+1,r} - \hat{y}_{i,T+1,ref,r})^2}\right]_{\alpha}.$$

6.4 Results

The results for the ratio of average MSFEs from the panel AR and ARX models are reported in Tables 1 and 2. In each table, we vary the cross-sectional dimension (N) across three blocks of results and the time-series dimension (T) along the columns. Each row assumes a different combination of the two hyperparameters that determine heterogeneity, a_β and σ_α^2, with the homogeneous case ($a_\beta = \sigma_\alpha^2 = 0$) in the top row. The top part of the table sets PR^2 to approximately 0.2 while the bottom part sets PR^2 to approximately 0.6.

With no heterogeneity and a small time-series dimension, T, consistent with Propositions 1

Table 1: Monte Carlo results for panel AR model

a_β	σ_α^2	Pooled			RE			FE			Empirical Bayes			Hier. Bayes*			Comb. (pool)			Comb. (FE)		
	T	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100
$PR^2 = 0.2$																						
$N = 50$																						
0.0	0.0	0.856	0.957	0.980	0.886	0.966	0.984	0.904	0.976	0.990	0.883	0.967	0.985	0.886	0.979	0.994	0.906	0.970	0.986	0.941	0.984	0.993
0.0	0.5	1.031	1.156	1.182	0.897	0.975	0.990	0.904	0.976	0.990	0.906	0.978	0.992	0.906	0.985	0.996	0.950	0.994	0.999	0.941	0.984	0.993
0.5	0.5	1.104	1.236	1.263	0.921	0.999	1.013	0.927	1.001	1.013	0.916	0.985	0.996	0.911	0.988	0.997	0.958	0.996	0.999	0.949	0.991	0.998
1.0	1.0	1.323	1.477	1.508	0.991	1.076	1.088	0.998	1.077	1.089	0.934	0.992	0.998	0.930	0.992	0.999	0.973	0.999	1.000	0.965	0.997	0.999
$N = 100$																						
0.0	0.0	0.849	0.955	0.980	0.889	0.973	0.989	0.897	0.975	0.990	0.877	0.965	0.985	0.904	0.977	0.991	0.902	0.968	0.985	0.936	0.983	0.993
0.0	0.5	1.052	1.180	1.212	0.891	0.974	0.989	0.897	0.975	0.990	0.900	0.978	0.991	0.927	0.986	0.995	0.950	0.994	0.999	0.936	0.983	0.993
0.5	0.5	1.111	1.244	1.278	0.916	1.000	1.016	0.922	1.001	1.016	0.911	0.985	0.996	0.933	0.988	0.997	0.956	0.996	0.999	0.945	0.990	0.998
1.0	1.0	1.334	1.484	1.525	0.995	1.083	1.100	1.002	1.085	1.101	0.932	0.992	0.998	0.951	0.993	0.998	0.971	0.998	1.000	0.962	0.996	0.999
$N = 1000$																						
0.0	0.0	0.848	0.955	0.979	0.883	0.969	0.987	0.895	0.975	0.989	0.873	0.965	0.984	0.865	0.965	0.986	0.898	0.968	0.985	0.931	0.982	0.992
0.0	0.5	1.047	1.179	1.209	0.888	0.973	0.989	0.895	0.975	0.989	0.897	0.977	0.991	0.901	0.979	0.992	0.947	0.993	0.998	0.931	0.982	0.992
0.5	0.5	1.099	1.237	1.268	0.912	0.999	1.015	0.919	1.001	1.016	0.907	0.985	0.996	0.914	0.987	0.997	0.953	0.995	0.999	0.941	0.990	0.997
1.0	1.0	1.321	1.480	1.518	0.991	1.083	1.100	0.999	1.085	1.101	0.928	0.992	0.998	0.939	0.995	0.999	0.969	0.997	0.999	0.960	0.996	0.999
$PR^2 = 0.6$																						
$N = 50$																						
0.0	0.0	0.843	0.954	0.979	0.892	0.971	0.985	0.900	0.975	0.989	0.885	0.967	0.984	0.884	0.977	0.992	0.898	0.968	0.985	0.936	0.983	0.992
0.0	0.5	0.927	1.048	1.075	0.889	0.973	0.988	0.900	0.975	0.989	0.906	0.978	0.991	0.893	0.983	0.995	0.930	0.988	0.997	0.936	0.983	0.992
0.5	0.5	1.001	1.126	1.153	0.925	1.016	1.032	0.937	1.020	1.033	0.918	0.984	0.995	0.908	0.985	0.995	0.946	0.993	0.998	0.949	0.992	0.998
1.0	1.0	1.181	1.322	1.352	1.014	1.126	1.148	1.028	1.130	1.150	0.934	0.989	0.997	0.928	0.990	0.997	0.967	0.997	0.999	0.966	0.997	1.000
$N = 100$																						
0.0	0.0	0.838	0.952	0.978	0.888	0.972	0.988	0.893	0.973	0.989	0.880	0.965	0.984	0.900	0.978	0.991	0.894	0.966	0.984	0.931	0.981	0.992
0.0	0.5	0.928	1.052	1.082	0.882	0.971	0.988	0.893	0.973	0.989	0.901	0.977	0.991	0.912	0.984	0.995	0.927	0.988	0.997	0.931	0.981	0.992
0.5	0.5	0.996	1.124	1.157	0.919	1.018	1.039	0.932	1.022	1.040	0.911	0.984	0.995	0.922	0.986	0.996	0.942	0.992	0.998	0.945	0.992	0.998
1.0	1.0	1.187	1.326	1.365	1.019	1.146	1.186	1.036	1.153	1.189	0.935	0.991	0.997	0.942	0.991	0.998	0.963	0.997	0.999	0.963	0.997	1.000
$N = 1000$																						
0.0	0.0	0.838	0.951	0.978	0.887	0.970	0.987	0.893	0.973	0.989	0.878	0.965	0.984	0.865	0.966	0.987	0.890	0.965	0.984	0.928	0.980	0.992
0.0	0.5	0.926	1.051	1.081	0.881	0.970	0.988	0.893	0.973	0.989	0.899	0.976	0.991	0.901	0.977	0.992	0.923	0.987	0.997	0.928	0.980	0.992
0.5	0.5	0.993	1.123	1.155	0.917	1.017	1.037	0.931	1.021	1.039	0.909	0.984	0.996	0.918	0.985	0.996	0.939	0.991	0.998	0.941	0.991	0.998
1.0	1.0	1.181	1.329	1.366	1.014	1.143	1.179	1.032	1.150	1.182	0.930	0.991	0.998	0.937	0.992	0.998	0.961	0.996	0.999	0.961	0.997	0.999

Note: The table reports the ratio of average MSFE for a given forecasting method over the average MSFE of the forecasts based on individual estimates. The forecasts are based on the following estimates: 'Pooled' for the pooled estimation, 'RE' for the random effects estimation, 'FE' for the fixed effects estimation, 'Empirical Bayes' is the empirical Bayes estimation, 'Hier.Bayes' the hierarchical Bayesian estimation, 'Comb. (pool)' denotes the combination of forecasts based on individual and pooled estimation, and 'Comb. (FE)' the combination of forecasts based on individual and fixed effects estimation. The parameters a_β and σ_α^2 determine the heterogeneity of the slope coefficient and the intercept. Reults in the top and bottom panels are for PR^2 of approximately 0.2 and 0.6, respectively. The DGP is more fully specified in Section 6.1. Results for all but the hierarchical Bayesian method are based on $R = 10.000$ replications. *The hierarchical Bayesian results are based on $R = 500$ replications.

Table 2: Monte Carlo results for panel ARX

a_β	σ^2_α	Pooled			RE			FE			Empirical Bayes			Hier. Bayes*			Comb. (pool)			Comb. (FE)		
	T	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100
$PR^2 = 0.2$																						
$N = 50$, $\rho_{\gamma x} = 0$																						
0.0	0.0	0.799	0.934	0.969	0.835	0.950	0.978	0.843	0.953	0.979	0.849	0.953	0.978	0.854	0.971	0.991	0.870	0.954	0.978	0.900	0.967	0.985
0.0	0.5	0.995	1.160	1.208	0.837	0.952	0.979	0.843	0.953	0.979	0.866	0.962	0.984	0.868	0.976	0.993	0.924	0.988	0.997	0.900	0.967	0.985
0.5	0.5	1.082	1.257	1.308	0.921	1.044	1.075	0.929	1.045	1.075	0.893	0.980	0.995	0.900	0.984	0.996	0.938	0.991	0.998	0.928	0.986	0.996
1.0	1.0	1.316	1.528	1.592	1.049	1.189	1.227	1.060	1.191	1.228	0.918	0.989	0.998	0.927	0.990	0.998	0.957	0.995	0.999	0.949	0.993	0.998
$N = 50$, $\rho_{\gamma x} = 0.5$																						
0.0	0.5	0.984	1.148	1.195	0.837	0.952	0.979	0.843	0.953	0.979	0.866	0.962	0.984	0.869	0.976	0.994	0.923	0.987	0.997	0.900	0.967	0.985
0.5	0.5	1.089	1.268	1.321	0.917	1.040	1.071	0.924	1.041	1.071	0.889	0.979	0.995	0.896	0.983	0.996	0.938	0.992	0.998	0.926	0.986	0.996
1.0	1.0	1.329	1.543	1.611	1.045	1.182	1.220	1.055	1.184	1.221	0.916	0.988	0.998	0.924	0.990	0.998	0.958	0.995	0.999	0.948	0.992	0.998
$N = 100$, $\rho_{\gamma x} = 0$																						
0.0	0.0	0.792	0.931	0.969	0.823	0.944	0.975	0.836	0.951	0.979	0.843	0.952	0.978	0.857	0.966	0.987	0.865	0.952	0.978	0.895	0.966	0.985
0.0	0.5	0.955	1.120	1.169	0.830	0.949	0.979	0.836	0.951	0.979	0.856	0.960	0.983	0.878	0.974	0.991	0.914	0.985	0.997	0.895	0.966	0.985
0.5	0.5	1.073	1.249	1.302	0.936	1.068	1.102	0.946	1.071	1.103	0.884	0.978	0.994	0.902	0.983	0.996	0.934	0.990	0.998	0.929	0.987	0.997
1.0	1.0	1.344	1.553	1.617	1.106	1.258	1.299	1.120	1.261	1.300	0.914	0.988	0.997	0.930	0.990	0.998	0.957	0.995	0.999	0.954	0.994	0.999
$N = 100$, $\rho_{\gamma x} = 0.5$																						
0.0	0.5	0.951	1.115	1.164	0.830	0.950	0.979	0.836	0.951	0.979	0.857	0.960	0.984	0.878	0.974	0.991	0.913	0.985	0.996	0.895	0.966	0.985
0.5	0.5	1.104	1.285	1.339	0.943	1.074	1.107	0.951	1.076	1.108	0.883	0.978	0.994	0.901	0.982	0.996	0.938	0.991	0.998	0.930	0.988	0.997
1.0	1.0	1.395	1.613	1.679	1.116	1.267	1.307	1.127	1.270	1.308	0.913	0.988	0.997	0.928	0.989	0.998	0.960	0.996	0.999	0.954	0.994	0.999
$N = 1000$, $\rho_{\gamma x} = 0$																						
0.0	0.0	0.786	0.931	0.968	0.819	0.945	0.975	0.830	0.950	0.978	0.840	0.952	0.978	0.812	0.946	0.978	0.859	0.951	0.977	0.889	0.965	0.984
0.0	0.5	0.971	1.149	1.196	0.824	0.949	0.978	0.830	0.950	0.978	0.852	0.960	0.983	0.843	0.959	0.984	0.914	0.986	0.997	0.889	0.965	0.984
0.5	0.5	1.093	1.284	1.335	0.930	1.067	1.100	0.940	1.069	1.101	0.879	0.977	0.994	0.885	0.981	0.995	0.934	0.990	0.998	0.925	0.987	0.997
1.0	1.0	1.374	1.603	1.666	1.098	1.257	1.296	1.110	1.261	1.297	0.908	0.987	0.997	0.922	0.991	0.998	0.957	0.995	0.999	0.951	0.994	0.999
$N = 1000$, $\rho_{\gamma x} = 0.5$																						
0.0	0.5	0.959	1.135	1.181	0.824	0.949	0.978	0.830	0.950	0.978	0.853	0.960	0.983	0.844	0.960	0.984	0.911	0.985	0.996	0.889	0.965	0.984
0.5	0.5	1.119	1.315	1.367	0.929	1.065	1.097	0.937	1.066	1.098	0.878	0.977	0.994	0.884	0.980	0.995	0.937	0.991	0.998	0.924	0.987	0.997
1.0	1.0	1.406	1.642	1.707	1.093	1.252	1.291	1.104	1.254	1.291	0.906	0.987	0.997	0.921	0.990	0.998	0.959	0.995	0.999	0.950	0.994	0.999
$PR^2 = 0.6$																						
$N = 50$, $\rho_{\gamma x} = 0$																						
0.0	0.0	0.788	0.931	0.967	0.834	0.950	0.977	0.839	0.952	0.978	0.851	0.953	0.977	0.858	0.968	0.989	0.864	0.952	0.976	0.897	0.966	0.984
0.0	0.5	0.874	1.031	1.073	0.828	0.949	0.977	0.839	0.952	0.978	0.867	0.962	0.983	0.864	0.974	0.992	0.897	0.977	0.994	0.897	0.966	0.984
0.5	0.5	0.986	1.151	1.200	0.925	1.057	1.091	0.940	1.062	1.092	0.897	0.980	0.994	0.886	0.982	0.994	0.925	0.987	0.997	0.929	0.987	0.997
1.0	1.0	1.207	1.400	1.461	1.069	1.230	1.281	1.089	1.238	1.284	0.924	0.988	0.997	0.915	0.989	0.997	0.953	0.994	0.998	0.953	0.994	0.999
$N = 50$, $\rho_{\gamma x} = 0.5$																						
0.0	0.5	0.870	1.026	1.068	0.827	0.949	0.977	0.839	0.952	0.978	0.867	0.963	0.983	0.864	0.974	0.992	0.896	0.977	0.993	0.897	0.966	0.984
0.5	0.5	0.980	1.145	1.195	0.920	1.053	1.087	0.936	1.058	1.088	0.895	0.979	0.994	0.880	0.981	0.994	0.923	0.987	0.997	0.928	0.986	0.996
1.0	1.0	1.189	1.382	1.445	1.061	1.224	1.274	1.082	1.232	1.277	0.919	0.987	0.997	0.914	0.989	0.997	0.951	0.994	0.998	0.951	0.993	0.999
$N = 100$, $\rho_{\gamma x} = 0$																						
0.0	0.0	0.781	0.927	0.967	0.826	0.946	0.976	0.832	0.948	0.978	0.844	0.951	0.978	0.853	0.964	0.986	0.858	0.949	0.976	0.891	0.964	0.984
0.0	0.5	0.860	1.020	1.067	0.822	0.946	0.977	0.832	0.948	0.978	0.857	0.959	0.983	0.865	0.971	0.991	0.888	0.975	0.993	0.891	0.964	0.984
0.5	0.5	0.994	1.168	1.221	0.941	1.085	1.124	0.956	1.091	1.126	0.883	0.977	0.994	0.889	0.980	0.994	0.922	0.987	0.997	0.928	0.988	0.997
1.0	1.0	1.256	1.461	1.526	1.128	1.314	1.378	1.151	1.325	1.383	0.911	0.985	0.996	0.921	0.987	0.997	0.953	0.994	0.999	0.955	0.995	0.999
$N = 100$, $\rho_{\gamma x} = 0.5$																						
0.0	0.5	0.857	1.016	1.062	0.822	0.946	0.977	0.832	0.948	0.978	0.858	0.959	0.983	0.865	0.971	0.991	0.887	0.974	0.993	0.891	0.964	0.984
0.5	0.5	1.004	1.180	1.232	0.946	1.091	1.129	0.961	1.096	1.131	0.884	0.977	0.994	0.886	0.979	0.995	0.923	0.988	0.997	0.929	0.988	0.997
1.0	1.0	1.276	1.484	1.548	1.139	1.330	1.394	1.162	1.340	1.399	0.912	0.986	0.996	0.918	0.986	0.997	0.954	0.994	0.999	0.955	0.995	0.999
$N = 1000$, $\rho_{\gamma x} = 0$																						
0.0	0.0	0.779	0.927	0.967	0.823	0.945	0.976	0.829	0.948	0.977	0.846	0.952	0.978	0.814	0.947	0.979	0.853	0.948	0.976	0.886	0.963	0.983
0.0	0.5	0.863	1.027	1.072	0.818	0.945	0.976	0.829	0.948	0.977	0.857	0.959	0.982	0.842	0.957	0.984	0.886	0.975	0.993	0.886	0.963	0.983
0.5	0.5	0.995	1.176	1.226	0.935	1.084	1.122	0.951	1.089	1.124	0.881	0.976	0.993	0.884	0.979	0.995	0.920	0.987	0.997	0.925	0.987	0.997
1.0	1.0	1.247	1.461	1.521	1.121	1.317	1.378	1.146	1.328	1.383	0.911	0.985	0.996	0.917	0.988	0.997	0.950	0.993	0.998	0.953	0.994	0.999
$N = 1000$, $\rho_{\gamma x} = 0.5$																						
0.0	0.5	0.858	1.021	1.065	0.818	0.945	0.976	0.829	0.948	0.977	0.857	0.959	0.982	0.843	0.957	0.984	0.884	0.974	0.993	0.886	0.963	0.983
0.5	0.5	0.996	1.176	1.227	0.932	1.081	1.119	0.948	1.087	1.121	0.882	0.976	0.993	0.883	0.978	0.994	0.920	0.987	0.997	0.924	0.987	0.997
1.0	1.0	1.243	1.456	1.517	1.114	1.311	1.372	1.139	1.321	1.377	0.912	0.986	0.996	0.916	0.987	0.997	0.950	0.993	0.998	0.952	0.994	0.999

Note: The DGP is given in Section 6.2. See the footnote of Table 1 for further details.

and 2, pooling yields MSFE-values around 15% lower than those of the individual forecasts. This finding is quite robust to the values of N and PR^2, though the advantage of pooling over individual estimation improves by a further 1-2% as these parameters are increased from their lowest to their highest values in our setting. Conversely, the advantage of pooling over the benchmark is rapidly reduced once T increases from $T = 20$ (improvement of 15%) to $T = 50$ (improvement of 4-5%) and $T = 100$ (improvement of only 2%).

Even modest degrees of heterogeneity in the intercept or persistence parameter of the autoregressive process in (55) result in a substantial deterioration in the predictive accuracy of the forecasts based on the pooled estimator relative to individual estimation. Outside special cases such as $T = 20$, $PR^2 = 0.6$ and heterogeneity only in the intercept, the individual forecasts dominate the pooled estimator along the MSFE ratio metric in most scenarios with parameter heterogeneity, particularly when T is relatively large.

The RE approach produces the best overall predictive accuracy only for a narrow subset of cases with heterogeneity only in the intercept and homogeneous slopes ($\sigma_\alpha^2 = 0.5$, $a_\beta = 0$). The predictive accuracy of the FE approach is similar to that of the RE approach with MSFE ratios typically less than 0.5% higher. Both approaches perform quite poorly under strong parameter heterogeneity, however, with MSFE-values around 7-18% higher than the benchmark for $T = 50, 100$.

Combining the forecasts from the pooled and individual-specific estimators leads to good overall performance, reducing the MSFE of the benchmark by about 10% under homogeneous parameters when $T = 20$. Strong parameter heterogeneity reduces the gain of this combination to 3-4%, again assuming $T = 20$. The scope for improvements in MSFE-ratios is markedly smaller for $T = 50$ or $T = 100$ but, in contrast with the pooled, RE, and FE approaches, this forecast combination never generates MSFE-ratios that exceed one.

Under parameter homogeneity, the standard pooled-individual specific combination also performs better than the combination of forecasts from the FE and individual-specific models. However, in the presence of parameter heterogeneity, the FE-individual forecast combination performs even better than the baseline combination of pooled and individual-specific forecasts with additional MSFE reductions of 0-2%. Such gains arise because the FE forecasts are more accurate than the pooled forecasts in this case.[14]

[14]These findings are sensitive to the assumed value for PR^2 and in some cases reverse when PR^2 is raised to 0.6.

The EB estimator produces the best overall performance with many MSFE values substantially lower than the individual-specific approach. Relative to this baseline, the EB approach reduces the MSFE ratio by 6–13% when $T = 20$, with smaller improvements for higher degrees of parameter heterogeneity. Improvements in MSFE-values for the EB estimator over the baseline shrink to 0.8–3.5% for $T = 50$ and are further reduced to 0.2–1.6% when $T = 100$. Like the forecast combinations, the EB approach also does not produce an MSFE ratio above unity in any of our simulations.

The hierarchical Bayes approach generates MSFE ratios that are typically a little worse than for the EB approach, though differences are quite small (0-2%).[15] Given the higher computational cost of implementing the hierarchical approach, the empirical Bayes forecasts appear to be preferable based on these results.

MSFE ratios for the pooled combination approaches are generally 1-5% higher that those of the EB approach when $T = 20$, 0–3% higher when $T = 50$, and less than 1% higher when $T = 100$. In the majority of the simulations, the hierarchical Bayes approach also tends to be a little better than the forecast combinations, though differences are very small for the larger sample sizes $T = 50, 100$.

Consistent with Proposition 1, in the setting with the largest time-series dimension ($T = 100$), although some of the panel-based forecasting models generate mean MSFE ratios less than one, improvements in the predictive accuracy tend to be small, rarely exceeding 1.5%, and typically being much smaller outside the scenario with homogeneous parameters.

Table 2 shows results for the ARX case with first-order autoregressive dynamics and an additional regressor, x_{it}, included. Though heterogeneity is also correlated in the AR panel (Pesaran and Smith, 1995), this setup allows us to study further the role of correlated heterogeneity by varying the correlation between the coefficient γ_i and x_{it} as measured by $\rho_{\gamma x}$. Every second block of results in the table assume $\rho_{\gamma x} = 0$ interchanged with blocks raising this parameter to $\rho_{\gamma x} = 0.5$—the latter only contains the three rows with results for heterogeneous parameters as homogenous parameters are non-random and therefore have zero correlation with the regressors.

In the scenario where $\rho_{\gamma x} = 0$, the results from Table 1 continue to hold: The pooled estimator produces the most accurate forecasts in the absence of any parameter heterogeneity, while the RE approach is best with heterogeneity only in the intercept but not in the slope parameter. We also

[15] In a few of the simulations, forecasts from the hierarchical approach are marginally more accurate than forecasts based on the EB approach. However, this is likely attributable to simulation errors given the smaller simulation size used for the hierarchical Bayes model.

note that the RE and FE approaches now underperform by an even bigger margin relative to the benchmark under the highest degrees of parameter heterogeneity considered here ($a_\beta = 1$, $\sigma_\alpha^2 = 1$).

Raising the level of correlated parameter heterogeneity ($\rho_{\gamma x} = 0.5$) leads to a marginal deterioration in the performance of the pooled estimation approach in the scenario with the highest values of a_β and σ_α^2 when $PR^2 = 0.2$, though the opposite holds in some cases for $PR = 0.6$. The value of $\rho_{\gamma x}$ only has a modest impact on the predictive accuracy of the combination, RE, and FE approaches.

Table 2 also shows that the EB and hierarchical Bayesian approaches continue to produce the best overall predictive accuracy in most scenarios for the panel ARX simulations followed by the combination approaches.

Tables 3 to 5 report the quantiles of the ratios of MSFEs over the individual units. For brevity, we focus on the case of $N = 1000$ and $T = 50$. The quantiles for the panel AR model in Table 3 show that under parameter homogeneity all methods have distributions of forecasts that are below unity for all quantiles so that even the 99th percentile of the MSFE ratio distribution favors a panel-based forecast over the individual forecast.[16] Unsurprisingly, under homogenous parameters the distribution of MSFE ratios is furthest to the left for the pooled forecasts. For this scenario, the EB forecasts deliver the second largest gains after the pooled forecast although the pooled combination forecasts reduce MSFE ratios the most in the right tail regardless of the goodness of fit of the model.

Introducing parameter heterogeneity in the intercept only ($a_\beta = 0$, $\sigma_\alpha^2 = 0.5$), the pooled estimator now produces larger MSFE values than the individual approach across all quantiles. For this case, the RE and FE forecasts generate smaller MSFE values than the individual forecasts across all quantiles and provide the most precise forecasts at any quantile with the RE marginally more precise than the FE forecast. The EB forecasts are the second most accurate across all quantiles closely followed by the combination of individual and FE forecasts.

When also the slope parameter is heterogeneous, the RE and FE forecasts are less precise than the individual forecasts in the right quantiles but still more precise in the left quantiles. The combination and EB forecasts improve over the individual forecasts for all but the very largest quantiles. Even at the 99th percentile of the MSFE ratio distribution, these approaches only perform marginally worse than the individual forecasts and so offer insurance against large forecast errors.

[16] Results for the hierarchical Bayes forecasts are somewhat erratic particularly for the tail quantiles due to the much smaller number of replications used for this method and so are omitted from here.

Table 3: Monte Carlo results, quantiles of ratios of MSFEs for panel AR models, $N = 1000$ and $T = 50$

	$PR^2 = 0.2$							$PR^2 = 0.6$						
Quantiles	0.01	0.05	0.10	0.50	0.90	0.95	0.99	0.01	0.05	0.10	0.50	0.90	0.95	0.99
$a_\beta = 0.0$, $\sigma_\alpha^2 = 0.0$														
Pooled	0.935	0.946	0.948	0.955	0.962	0.963	0.967	0.936	0.942	0.944	0.951	0.958	0.959	0.962
RE	0.952	0.962	0.964	0.970	0.975	0.977	0.980	0.957	0.963	0.965	0.971	0.976	0.977	0.980
FE	0.958	0.967	0.969	0.975	0.980	0.982	0.985	0.959	0.965	0.968	0.973	0.978	0.980	0.983
Emp.Bayes	0.950	0.957	0.960	0.967	0.972	0.973	0.976	0.951	0.959	0.960	0.967	0.972	0.973	0.975
Comb. (pool)	0.957	0.962	0.963	0.968	0.971	0.972	0.974	0.956	0.960	0.961	0.965	0.968	0.970	0.971
Comb. (FE)	0.972	0.977	0.979	0.982	0.986	0.986	0.988	0.972	0.976	0.977	0.980	0.983	0.984	0.986
$a_\beta = 0.0$, $\sigma_\alpha^2 = 0.5$														
Pooled	1.076	1.087	1.092	1.137	1.412	1.542	1.852	1.018	1.026	1.029	1.045	1.098	1.125	1.185
RE	0.957	0.966	0.967	0.974	0.979	0.981	0.984	0.956	0.962	0.965	0.970	0.975	0.977	0.980
FE	0.958	0.967	0.969	0.975	0.980	0.982	0.985	0.959	0.965	0.968	0.973	0.978	0.980	0.983
Emp.Bayes	0.963	0.970	0.972	0.978	0.983	0.985	0.987	0.965	0.969	0.972	0.977	0.982	0.983	0.985
Comb. (pool)	0.988	0.990	0.991	0.993	0.998	1.001	1.007	0.981	0.983	0.984	0.987	0.991	0.992	0.996
Comb. (FE)	0.972	0.977	0.979	0.982	0.986	0.986	0.988	0.972	0.976	0.977	0.980	0.983	0.984	0.986
$a_\beta = 0.5$, $\sigma_\alpha^2 = 0.5$														
Pooled	1.007	1.034	1.063	1.234	1.480	1.618	1.902	0.971	0.988	1.002	1.112	1.262	1.286	1.325
RE	0.964	0.971	0.973	0.992	1.036	1.046	1.062	0.963	0.969	0.972	1.002	1.076	1.115	1.187
FE	0.964	0.971	0.974	0.992	1.042	1.057	1.077	0.964	0.969	0.972	1.000	1.099	1.162	1.252
Emp.Bayes	0.971	0.976	0.978	0.985	0.994	0.998	1.004	0.962	0.971	0.975	0.984	0.994	1.000	1.010
Comb. (pool)	0.989	0.991	0.992	0.995	0.998	1.000	1.005	0.982	0.985	0.986	0.991	0.997	0.998	1.001
Comb. (FE)	0.980	0.984	0.985	0.989	0.996	0.999	1.001	0.983	0.986	0.987	0.990	0.999	1.004	1.010
$a_\beta = 1.0$, $\sigma_\alpha^2 = 1.0$														
Pooled	1.008	1.059	1.108	1.420	1.990	2.106	2.315	0.958	0.985	1.007	1.260	1.743	1.837	1.955
RE	0.966	0.974	0.978	1.043	1.248	1.299	1.404	0.967	0.973	0.977	1.074	1.355	1.504	2.134
FE	0.967	0.974	0.978	1.044	1.248	1.309	1.451	0.966	0.973	0.977	1.065	1.347	1.613	2.349
Emp.Bayes	0.977	0.982	0.984	0.992	1.001	1.004	1.011	0.957	0.974	0.983	0.991	1.000	1.004	1.011
Comb. (pool)	0.993	0.994	0.995	0.997	1.000	1.001	1.003	0.989	0.991	0.992	0.995	1.000	1.002	1.004
Comb. (FE)	0.988	0.991	0.992	0.995	1.002	1.005	1.009	0.990	0.993	0.994	0.995	1.002	1.007	1.014

Note: This table reports the quantiles for the cross-sectional distribution of ratios of MSFEs. The quantiles are based on MSFE values calculated from 10,000 replications. Hierarchical Bayesian forecasts are omitted as they are only available for 500 replications. For further details, see the footnote of Table 1.

Table 4: Monte Carlo results, quantiles of ratio of MSFEs for panel ARX models, $PR^2 = 0.2$, $N = 1000$ and $T = 50$

	$\rho_{\gamma x} = 0$							$\rho_{\gamma x} = 0.5$						
Quantiles	0.01	0.05	0.10	0.50	0.90	0.95	0.99	0.01	0.05	0.10	0.50	0.90	0.95	0.99
$a_\beta = 0.0$, $\sigma^2_\alpha = 0.0$														
Pooled	0.911	0.919	0.922	0.932	0.940	0.942	0.946	0.911	0.919	0.922	0.932	0.940	0.942	0.946
RE	0.925	0.933	0.937	0.946	0.953	0.955	0.959	0.925	0.933	0.937	0.946	0.953	0.955	0.959
FE	0.931	0.939	0.942	0.951	0.959	0.960	0.964	0.931	0.939	0.942	0.951	0.959	0.960	0.964
Emp.Bayes	0.930	0.941	0.944	0.956	0.965	0.967	0.970	0.930	0.941	0.944	0.956	0.965	0.967	0.970
Comb. (pool)	0.938	0.943	0.945	0.951	0.956	0.957	0.960	0.938	0.943	0.945	0.951	0.956	0.957	0.960
Comb. (FE)	0.953	0.957	0.960	0.965	0.970	0.971	0.973	0.953	0.957	0.960	0.965	0.970	0.971	0.973
$a_\beta = 0.0$, $\sigma^2_\alpha = 0.5$														
Pooled	1.046	1.056	1.063	1.107	1.352	1.545	1.943	1.038	1.048	1.054	1.099	1.330	1.503	1.894
RE	0.930	0.937	0.941	0.950	0.958	0.960	0.963	0.929	0.937	0.941	0.950	0.958	0.960	0.963
FE	0.931	0.939	0.942	0.951	0.959	0.960	0.964	0.931	0.939	0.942	0.951	0.959	0.960	0.964
Emp.Bayes	0.940	0.948	0.953	0.963	0.971	0.973	0.976	0.941	0.949	0.953	0.963	0.971	0.973	0.977
Comb. (pool)	0.978	0.980	0.981	0.985	0.993	0.999	1.014	0.977	0.979	0.981	0.984	0.992	0.997	1.013
Comb. (FE)	0.953	0.957	0.960	0.965	0.970	0.971	0.973	0.953	0.957	0.960	0.965	0.970	0.971	0.973
$a_\beta = 0.5$, $\sigma^2_\alpha = 0.5$														
Pooled	0.993	1.040	1.080	1.271	1.639	1.818	2.670	1.030	1.066	1.100	1.286	1.592	1.745	2.726
RE	0.947	0.956	0.962	1.021	1.284	1.446	2.362	0.945	0.954	0.962	1.021	1.266	1.490	2.228
FE	0.947	0.957	0.962	1.021	1.289	1.459	2.418	0.945	0.954	0.962	1.021	1.271	1.502	2.315
Emp.Bayes	0.957	0.965	0.968	0.980	0.991	0.996	1.002	0.957	0.965	0.968	0.980	0.990	0.994	1.002
Comb. (pool)	0.983	0.985	0.986	0.990	0.998	1.001	1.013	0.984	0.986	0.987	0.991	0.996	0.999	1.013
Comb. (FE)	0.977	0.980	0.981	0.985	0.998	1.005	1.048	0.976	0.979	0.981	0.985	0.997	1.007	1.040
$a_\beta = 1.0$, $\sigma^2_\alpha = 1.0$														
Pooled	1.005	1.087	1.151	1.554	2.267	2.690	4.003	1.032	1.093	1.149	1.531	2.313	2.581	4.441
RE	0.952	0.968	0.985	1.165	1.728	2.015	4.652	0.952	0.967	0.984	1.164	1.695	2.142	3.816
FE	0.952	0.968	0.985	1.167	1.748	2.080	4.825	0.952	0.967	0.988	1.162	1.701	2.199	3.844
Emp.Bayes	0.964	0.975	0.978	0.989	1.000	1.003	1.008	0.966	0.975	0.978	0.989	0.999	1.002	1.007
Comb. (pool)	0.989	0.990	0.991	0.994	1.000	1.002	1.009	0.990	0.991	0.992	0.995	0.999	1.002	1.010
Comb. (FE)	0.987	0.989	0.990	0.993	1.000	1.005	1.030	0.987	0.989	0.990	0.993	1.000	1.006	1.026

Note: For details see the footnote of Table 3.

Table 5: Monte Carlo results, quantiles of ratio of MSFEs for panel ARX models, $PR^2 = 0.6$, $N = 1000$ and $T = 50$

	$\rho_{\gamma x} = 0$							$\rho_{\gamma x} = 0.5$						
Quantiles	0.01	0.05	0.10	0.50	0.90	0.95	0.99	0.01	0.05	0.10	0.50	0.90	0.95	0.99
$a_\beta = 0.0$, $\sigma_\alpha^2 = 0.0$														
Pooled	0.907	0.914	0.917	0.927	0.935	0.938	0.942	0.907	0.914	0.917	0.927	0.935	0.938	0.942
RE	0.927	0.934	0.937	0.946	0.954	0.956	0.959	0.927	0.934	0.937	0.946	0.954	0.956	0.959
FE	0.930	0.936	0.939	0.949	0.956	0.958	0.962	0.930	0.936	0.939	0.949	0.956	0.958	0.962
Emp.Bayes	0.930	0.939	0.944	0.956	0.965	0.967	0.971	0.930	0.939	0.944	0.956	0.965	0.967	0.971
Comb. (pool)	0.935	0.940	0.942	0.948	0.953	0.954	0.957	0.935	0.940	0.942	0.948	0.953	0.954	0.957
Comb. (FE)	0.950	0.955	0.958	0.963	0.968	0.969	0.971	0.950	0.955	0.958	0.963	0.968	0.969	0.971
$a_\beta = 0.0$, $\sigma_\alpha^2 = 0.5$														
Pooled	0.990	0.997	1.002	1.022	1.076	1.120	1.208	0.987	0.993	0.998	1.016	1.067	1.111	1.210
RE	0.927	0.933	0.936	0.946	0.954	0.956	0.959	0.927	0.933	0.936	0.946	0.954	0.956	0.959
FE	0.930	0.936	0.939	0.949	0.956	0.958	0.962	0.930	0.936	0.939	0.949	0.956	0.958	0.962
Emp.Bayes	0.940	0.947	0.951	0.962	0.970	0.972	0.976	0.940	0.947	0.951	0.962	0.970	0.972	0.976
Comb. (pool)	0.966	0.969	0.970	0.975	0.980	0.983	0.992	0.965	0.968	0.969	0.974	0.979	0.982	0.992
Comb. (FE)	0.950	0.955	0.958	0.963	0.968	0.969	0.971	0.950	0.955	0.958	0.963	0.968	0.969	0.971
$a_\beta = 0.5$, $\sigma_\alpha^2 = 0.5$														
Pooled	0.954	0.988	1.013	1.158	1.441	1.616	2.065	0.959	0.984	1.004	1.155	1.422	1.543	2.277
RE	0.945	0.953	0.962	1.043	1.306	1.478	2.470	0.945	0.952	0.963	1.039	1.275	1.508	2.374
FE	0.944	0.954	0.962	1.042	1.314	1.485	2.768	0.944	0.954	0.962	1.040	1.294	1.528	2.363
Emp.Bayes	0.950	0.961	0.966	0.978	0.992	0.997	1.007	0.952	0.960	0.966	0.979	0.991	0.996	1.007
Comb. (pool)	0.976	0.978	0.980	0.986	0.996	1.002	1.018	0.976	0.978	0.980	0.987	0.996	1.000	1.020
Comb. (FE)	0.977	0.980	0.981	0.986	0.997	1.004	1.053	0.976	0.980	0.981	0.985	0.997	1.006	1.038
$a_\beta = 1.0$, $\sigma_\alpha^2 = 1.0$														
Pooled	0.951	0.997	1.038	1.399	2.071	2.450	3.595	0.946	0.991	1.030	1.390	2.030	2.334	3.859
RE	0.952	0.969	0.990	1.225	1.815	2.173	4.506	0.950	0.970	0.993	1.213	1.778	2.088	4.058
FE	0.951	0.968	0.990	1.218	1.853	2.223	4.599	0.951	0.969	0.992	1.206	1.841	2.169	3.983
Emp.Bayes	0.948	0.968	0.976	0.988	0.999	1.002	1.008	0.955	0.972	0.977	0.988	0.998	1.002	1.006
Comb. (pool)	0.986	0.987	0.988	0.993	1.000	1.004	1.015	0.986	0.987	0.988	0.993	1.000	1.003	1.015
Comb. (FE)	0.988	0.990	0.990	0.993	1.001	1.004	1.021	0.988	0.990	0.990	0.993	1.000	1.004	1.024

Note: For details see the footnote of Table 3.

The quantiles of the MSFE ratios for the panel ARX model are reported in Tables 4 ($PR^2 = 0.2$) and 5 ($PR^2 = 0.6$). Columns on the left assume the parameters and regressors are uncorrelated while columns on the right are for the correlated case. With homogeneous parameters, the results are identical. In this case, the pooled forecasts are again most precise across all quantiles followed by the RE and FE forecasts. Combination and EB forecasts continue to be more precise than the individual forecasts for all quantiles.

When the intercept varies across individuals, the pooled forecasts again produce larger MSFE-values than the individual forecasts, whereas the RE and FE estimators generate the most precise forecasts across all quantiles. When also the slope coefficient is heterogeneous, the RE and FE forecasts have a larger MSFE than the individual forecast for quantiles to the right of the median. The RE and FE estimators can produce very poor forecasts with MSFE ratios between 2.2 and almost 5 at the 99th quantile. In contrast, the combination and EB forecasts continue to generate MSFE ratios below unity for most quantiles and only marginally underperform the benchmark forecasts in the right tail of the distribution of MSFE ratios.

In Section S.4 of the Supplementary Appendix (Tables S.5-S.8) we also report a complete suite of Monte Carlo simulation results based on an equal-weighted combination scheme for our two combination schemes. The predictive accuracy of the equal-weighted combination scheme is comparable to that of the combinations based on estimated weights in the presence of modest levels of parameter heterogeneity. Conversely, equal-weighted combinations underperform forecast combinations with estimated weights when the level of parameter heterogeneity is either very low or very high. In either case, one approach (individual estimation or pooling) dominates the other by a sufficiently large margin that equal-weighting becomes sub-optimal.

We also considered the performance of an (infeasible) oracle combination scheme that uses the true parameter values to compute the optimal combination weights. Compared against our feasible estimates of the combination weights, this oracle scheme shows the impact of parameter estimation error on forecasting performance. We find that the cost of estimation error is only sizeable if T is small ($T = 20$) and the parameters are homogeneous. For this case, the oracle scheme reduces the MSFE of the pooled-individual combination by 0.051 (0.906 versus 0.856) and by 0.037 for the FE-individual combination. Differences are much smaller (0.005 and 0.011) in the heterogeneous case even when $T = 20$ and are further reduced for $T = 100$ where, in many cases, only the third

decimal of the MSFE ratio is affected.

We also considered combination schemes that allow the weights to vary across individual series. These schemes perform marginally better in the case with completely homogeneous parameters but worse in many cases under parameter heterogeneity. Increasing the correlation between the coefficients and regressors also leads to a deterioration in the performance of these combinations relative to the cross-sectionally pooled combination schemes.

7 Empirical applications

We next apply our set of panel forecasting methods to two empirical applications on house price inflation in U.S. metropolitan areas and inflation in CPI sub-indices. These applications represent quite different levels of in-sample fit: For the CPI data the pooled R^2 (PR^2) of our models is around 0.2 while for house prices it exceeds 0.8.

7.1 Measures of forecasting performance

Our empirical applications compute the out-of-sample MSFE as

$$\mathrm{MSFE}_{ij} = \frac{1}{T-T_1}\sum_{t=T_1}^{T-1}(y_{i,t+1} - \hat{y}_{i,j,t+1})^2,$$

where $\hat{y}_{i,j,t+1}$ is the forecast of $y_{i,t+1}$ using method j and information known at time t. Each forecast in the test sample, $\hat{y}_{i,j,t+1}$, is generated using a rolling estimation window of observations $t-w+1, t-w, \ldots, t$, where w is the length of the rolling window. Details of the size of these samples are reported with each application.

We report the ratio of the average MSFE of method j relative to

the average MSFE for the reference forecasts from the individual-specific model

$$\mathrm{rMSFE}_j = \frac{\frac{1}{N}\sum_{i=1}^{N}\mathrm{MSFE}_{ij}}{\frac{1}{N}\sum_{i=1}^{N}\mathrm{MSFE}_{i,ref}}$$

We also report the proportion of units in the cross-section for which a given method produces a smaller MSFE than the reference forecast. Letting $\mathrm{I}(\cdot)$ be the indicator function that equals unity

if the expression inside the operator is true and is zero otherwise, we have

$$\frac{1}{N}\sum_{i=1}^{N} \mathrm{I}\left[\mathrm{MSFE}_{ij} < \mathrm{MSFE}_{i,ref}\right].$$

Next, we report the proportion of units in the cross-section for which a given method has the smallest or largest MSFE value computed as

$$\frac{1}{N}\sum_{i=1}^{N} \mathrm{I}\left[\mathrm{MSFE}_{ij} = \min_{l} \mathrm{MSFE}_{il}\right],$$
$$\frac{1}{N}\sum_{i=1}^{N} \mathrm{I}\left[\mathrm{MSFE}_{ij} = \max_{l} \mathrm{MSFE}_{il}\right].$$

These measures help us understand both the risk of underperformance and, on the upside, the possibility of superior predictive accuracy.[17] Finally, we report the quantiles $\alpha = (0.01, 0.05, 0.1, 0.5, 0.9, 0.95, 0.99)$ of the relative forecast accuracy over the N units in the panel

$$\mathrm{quantile}_j(\alpha) = \left[\frac{\mathrm{MSFE}_{ij}}{\mathrm{MSFE}_{i,ref}}\right]_{\alpha}.$$

To give more detailed insights into the distribution of the forecast performance across different units, we also provide density plots of the individual ratios of MSFEs.

We examine the significance of any differences in forecast accuracy using the Diebold and Mariano (1995) (DM) test of predictive accuracy both for the panel as a whole and for the individual series. First, we use the panel version of the DM test proposed by Pesaran et al. (2013). This tests the null that the MSFE generated by the individual forecasts, averaged both across time and units, is equal in expectation to the equivalent MSFE generated by the panel models. Second, we apply the DM test to the N forecasts for individual units in the sample and report the number of significant values in either direction and the number of insignificant test statistics. The tests are set up so that negative values indicate that the panel forecasts are more accurate than the individual forecasts, while positive values of the DM tests indicate that the individual forecasts are more accurate.

[17] These proportions can add up to more than one due to ties between forecasting methods.

7.2 U.S. house prices

Our first application uses quarterly data on real house price inflation in 377 U.S. Metropolitan Statistical Areas (MSAs) from the first quarter of 1975 to the first quarter of 2023, which we obtain from the Freddie Mac website.[18] Our forecasts focus on the one-quarter-ahead MSA-level rate of house price changes. Forecasts start in 2000Q1 and end in 2023Q1, a total of 93 quarterly periods. We use a rolling window of 60 quarters to estimate the model parameters.

Our prediction model for the house price inflation rate in quarter t for MSA i, y_{it}, takes the form

$$y_{it} = \alpha_i + \beta_i y_{i,t-1} + \beta_i^* y_{i,t-1}^* + \gamma_{Ri} \bar{y}_{i,t-1}^{(R)} + \gamma_{Ci} \bar{y}_{t-1}^{(C)} + \varepsilon_{it}, \tag{61}$$

where $i = 1, 2, \ldots, N$ denotes individual MSAs and $t = 1, 2, \ldots, T$ refers to the time period, $y_{it}^* = \sum_{k=1, k\neq i}^{N} \omega_{ik} y_{kt}$ is the spatial effect for a set of spatial weights ω_{ik}, $\bar{y}_{it}^{(R)}$ is the average house price inflation in the region of unit i, and $\bar{y}_t^{(C)}$ is the country-wide average house price inflation. The weights, ω_{ik}, measure the spatial effect of house prices in MSA k on house prices in MSA i and are based on geographic distance, that is $\omega_{ik} = v_{ik} / \sum_{k=1}^{N} v_{ik}$ and $v_{ik} = 1$ if MSAs (i, k) are at most 100 miles apart and is zero otherwise. We obtain the weights from the data set of Yang (2021) and exclude MSAs that do not have any neighbors within 100 miles, which leaves 362 MSAs in our sample.

We consider two forecasting models. The first, denoted SAR, is a spatial autoregressive model that excludes the regional and country-wide averages, such that in (61) $\gamma_{Ri} = \gamma_{Ci} = 0$. The second, denoted SARX, is the model in (61) with all coefficients left unrestricted.

Table 6 reports the results. In the first two columns, the first row shows the cross-sectional average MSFE value for the forecasts based on individual estimates. Subsequent rows report ratios of the mean of the individual MSFE for the respective methods relative to the benchmark forecasts. Values below unity show that the ratio of average MSFE performance (across MSAs) is better for the method listed in the row than for the benchmark while values above unity indicate the opposite. The next two columns headed 'freq. beating benchmark' report the proportion of MSAs for which the respective methods have a smaller MSFE than the benchmark, while the columns headed 'freq.

[18] For each MSA house prices are calculated by deflating the Freddie Mac house price index by the CPI.

Table 6: House price inflation forecasts

Forecast methods	Ratio of ave. MSFE		Freq. beating benchmark		Freq. smallest MSFE		Freq. largest MSFE	
	SAR	SARX	SAR	SARX	SAR	SARX	SAR	SARX
Individual	3.253	3.248	–	–	0.069	0.047	0.613	0.384
Pooled	0.969	0.990	0.660	0.417	0.229	0.127	0.202	0.365
RE	0.973	0.993	0.685	0.428	0.110	0.058	0.014	0.014
FE	0.983	1.002	0.682	0.450	0.221	0.080	0.160	0.229
Emp.Bayes	0.961	0.935	0.884	0.878	0.163	0.260	0.000	0.003
Hier.Bayes	0.984	0.967	0.859	0.840	0.047	0.169	0.008	0.006
Comb. (pool)	0.963	0.944	0.865	0.859	0.108	0.171	0.003	0.000
Comb. (FE)	0.970	0.948	0.867	0.865	0.052	0.088	0.000	0.000

Note: SAR denotes the spatial autoregressive model and SARX the same model with regional and nationwide house price averages added. Under 'Ratio of ave. MSFE', the entry under 'Individual' is the average MSFE of forecasts based on individual estimates and the remaining rows are the ratios of MSFEs for the methods listed in the first column (given in the footnote of Table 1). The columns 'Freq. beating benchmark' report, for the particular model, the proportions of MSAs for which the method in the row has a lower MSFE than the benchmark. Under 'Freq. smallest (largest) MSFE' are, for the particular model, the proportions of MSAs for which the method in the row has a lowest (highest) MSFE of all methods. The results in this table are for geographic spatial weighting (within a 100km neighborhood) and are based on estimation windows of 60 observations.

smallest MSFE' and 'freq. largest MSFE' show the proportion of MSAs for which the respective methods have the smallest or largest MSFE among all forecasting methods.

For the SAR model fitted to house prices, the panel-based forecasting approaches reduce MSFE values by 2–4% over the individual-specific forecasts. The combination and Bayesian approaches deliver the best results with the RE and FE approaches being slightly worse, though differences in average MSFE ratios are quite small.

We find greater variation in the rates at which the different forecasting approaches produce more accurate forecasts than the baseline individual-specific approach. While the pooled, RE, and FE approaches are more accurate than the baseline forecasts for roughly two-thirds of the housing price series, the Bayesian and combination approaches improve on the baseline for 86–88% of the series.

For the SAR model, the pooled and FE approaches lead the way when it comes to the highest frequency of cases with the smallest MSFE values (22%) followed by the EB approach (16%) and the RE and pooled-individual combination approaches (10%).

The ability of the pooled and FE forecasts to be most accurate for the highest proportion of individual series comes at the expense of also producing a higher risk of being the least accurate forecast for 20% and 16% of the series, respectively. In contrast, the Bayesian and combination

approaches rarely (or never for combinations) produce the worst forecasts for the individual series.

These results point to a much lower cross-sectional dispersion in (relative) MSFE values for the Bayesian and combination approaches compared with the pooled and FE approaches. In contrast, the individual-specific approach produces by far the highest frequency of forecasts with the highest MSFE values (61%), while only producing the lowest MSFE values for 7% of the series. This unattractive risk profile highlights significant scope for improvements in forecast accuracy by using panel information.

For the SARX model, the average of the MSFE ratios exceeds unity for the FE approach while conversely the Bayesian and combination approaches reduce average MSFE values by 5–7% relative to the baseline. Further, the Bayesian and combination approaches reduce the MSFE value of the baseline forecasts for a much greater proportion of series (84-88%) than the pooled, RE, and FE approaches (41-45%). The Bayesian approaches produce particularly high proportions of cases with the most accurate forecasts (17% and 26%, respectively) while rarely producing the worst forecasts for individual series.

The pooled-individual combination approach produces the best forecasts for 17% of the series while never producing the worst forecasts. Pooling alone generates the worst forecasts for 36% of the series while the individual-specific forecasts are the worst for 38% of the series, vividly illustrating the benefits from forecast combination.

Table 7 reports quantiles of the ratio of MSFE values for the housing price data. First consider the results for the SAR model (upper panel). Consistent with the results in Table 6, the cross-sectional distribution is more dispersed for the pooled, RE, and FE approaches with a spread between the 1st and 99th quantiles around 0.40 versus 0.15 for the Bayesian and combination approaches. While the distribution of MSFE ratios for the pooled, RE and FE approaches is right-skewed – with higher values above unity (worse MSFE performance) than below unity (better performance) – the opposite holds for the Bayesian and combination approaches. For the latter, we see improvements in MSFE performance of 10–13% at the first percentile versus deterioration in (relative) performance of 3–6% at the 99th percentile. Our results for the 90th quantile show that while close to 90 percent of the Bayesian and combination forecasts are as accurate as the individual specific forecasts (MSFE ratios at one), at this percentile the pooled, RE, and FE forecasts produce MSFE ratios that are 8–10% worse than the baseline.

Table 7: Quantiles of ratio of MSFEs for house price inflation over MSAs

Quantiles	0.01	0.05	0.10	0.50	0.90	0.95	0.99
House prices: SAR							
Pooled	0.821	0.865	0.879	0.970	1.079	1.135	1.226
RE	0.830	0.866	0.884	0.966	1.085	1.130	1.220
FE	0.851	0.880	0.893	0.967	1.097	1.143	1.224
Emp.Bayes	0.876	0.925	0.933	0.971	1.002	1.017	1.031
Hier.Bayes	0.939	0.952	0.964	0.987	1.004	1.012	1.032
Comb. (pool)	0.872	0.918	0.935	0.974	1.004	1.014	1.038
Comb. (FE)	0.881	0.931	0.946	0.976	1.004	1.011	1.032
House prices: SARX							
Pooled	0.764	0.821	0.854	1.024	1.198	1.241	1.420
RE	0.769	0.825	0.856	1.016	1.195	1.237	1.406
FE	0.777	0.833	0.855	1.015	1.200	1.252	1.441
Emp.Bayes	0.830	0.877	0.902	0.963	1.004	1.015	1.066
Hier.Bayes	0.900	0.915	0.934	0.972	1.007	1.019	1.051
Comb. (pool)	0.873	0.896	0.912	0.965	1.006	1.015	1.033
Comb. (FE)	0.885	0.901	0.916	0.967	1.003	1.011	1.027

Note: The table reports the quantiles of the distribution MSFE ratios computed across individual MSAs. SAR denotes the spatial autoregressive model and SARX the same model with regional and nationwide house price averages added. The forecasting methods are listed in the footnote of Table 1.

These results show that the Bayesian and combination approaches offer the attractive feature of not only improving on the MSFE values of the baseline "on average" but, equally importantly, rarely producing markedly worse forecasts than the baseline and very often generating substantially better results.

Qualitatively similar but even stronger results are obtained for the SARX model (lower panel). For this model we find an even larger spread around 0.65 between the 1st and 99th quantiles of the MSFE ratio distribution among the pooled, RE, and FE forecasts versus values ranging from 0.15 to 0.23 for the Bayesian and combination approaches. Again, we see the same asymmetries in the MSFE ratio distribution which is heavily right-tailed for the pooled, RE, and FE forecasts but left-tailed for the Bayesian and combination approaches. For the latter, MSFE ratios at the 99th quantile exceed unity only by 2–7% which is far lower than for the pooled, RE, and FE approaches (40–44%). Interestingly, the combination approaches perform notably better than the Bayesian forecasts at the 99th percentiles for the SARX model.[19]

[19]Equal-weighted combination also performs very well in both of the empirical applications, although they do not reduce the risk of generating high MSFE-values for individual units in the cross-section to the same extent as our optimal forecast combination scheme.

Our findings are summarized visually in Figure 1 which shows probability density and cumulative density plots fitted to the cross-sectional distribution of MSFE ratios for our forecasting methods. The figure shows a widely dispersed, right-skewed distribution of MSFE ratios for the pooled, RE and FE forecasting approaches compared to the Bayesian and combination approaches whose distributions are far more peaked and centered just below unity. The right skew for the pooled, FE, and RE forecasts is particularly clear for the SARX model displayed in the right panels. This feature is highly undesirable as it raises the likelihood of very poor forecasts for an individual housing price series compared with that of the Bayesian and combination approaches.[20]

Table 8 reports DM test statistics for our two empirical applications. In each panel, the first row shows the outcome of the panel DM test while the subsequent three rows show the number of cross-sectional units with a DM test below -1.96 (panel forecasts are significantly more accurate), the number of insignificant cases with a DM test statistic within $(-1.96; 1.96)$, and the number of variables with a DM test above 1.96 (individual-specific forecasts are significantly more accurate).

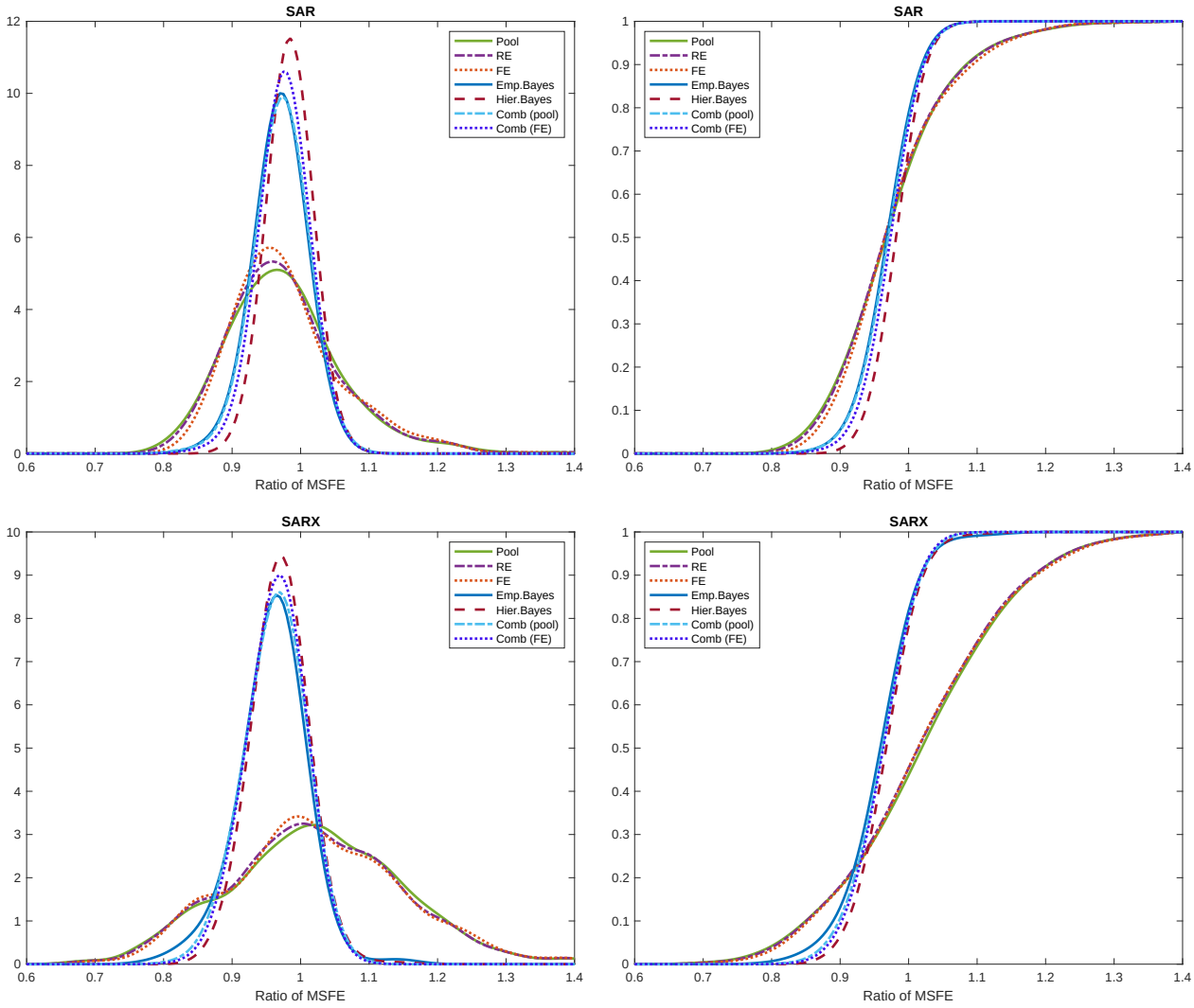
For the panel DM test applied to the SAR model, the null of equal predictive accuracy is strongly rejected for all panel forecasts, suggesting that the panel forecasts are significantly more accurate, on average, than the individual forecasts. Test statistics are particularly large for the combination and Bayesian forecasts. Conversely, for the SARX model, we fail to reject the null of equal (average) predictive accuracy of the individual forecasts versus the pooled, RE, and FE forecasts. In contrast, the combination and Bayesian forecasts lead to strong rejections of the null of equal forecast accuracy.

Turning to the SAR forecasts of the individual house price series, although the DM test fails to reject the null of equal predictive accuracy for at least 70% of the variables, we find many more cases in which the combination and Bayesian forecasts are significantly more accurate than the baseline (89–95 cases) than cases where the pooled, RE, or FE forecasts beat the baseline forecasts by a significant margin (24–27 cases). The individual forecasts are significantly more accurate than the combination or Bayesian forecasts for only two or fewer of the MSAs versus up to nine MSAs for the FE forecasts.

For the SARX model, there are only slightly more cases for which the pooled, RE, and FE forecasts significantly improve on the benchmark (17) than MSAs where the opposite holds (11-14).

[20]The impressive performance of the EB approach for the tail groups is consistent with Efron (2011) and is a point that may carry over to the forecast combinations.

Figure 1: Distributions of ratios of MSFE for house price inflation across MSAs



Note: The figure shows density (left panels) and cumulative density (right panels) plots of the ratios of MSFE values for the SAR top panel) and SARX model (bottom). The density estimates use a normal kernel with bandwidth 0.03. The cumulative densities are normalized to 1 at the right tail. The forecasting methods are listed in the footnote of Tables 1.

Table 8: Diebold-Mariano test statistics for equal predictive accuracy

	Pooled	RE	FE	Emp.Bay.	Hier.Bay.	Comb(pool)	Comb(FE)
House Prices: SAR							
Panel DM	−4.69	−4.04	−2.60	−16.73	−18.66	−19.08	−17.08
DM< −1.96	25	27	24	95	93	94	89
Insign.	330	327	329	267	267	267	272
DM> 1.96	7	8	9	0	2	1	1
House Prices: SARX							
Panel DM	−1.05	−0.71	0.23	−15.31	−16.83	−22.08	−21.57
DM< −1.96	17	17	17	116	87	107	109
Insign.	331	332	334	242	274	255	253
DM> 1.96	14	13	11	4	1	0	0
CPI: AR							
Panel DM	−3.82	−3.75	−3.44	−6.75	−15.41	−7.75	−6.12
DM< −1.96	8	5	1	50	44	24	9
Insign.	77	92	96	51	55	73	92
DM> 1.96	16	4	4	0	2	4	0
CPI: AR-PC							
Panel DM	−3.95	−3.91	−3.61	−6.90	−14.11	−8.00	−6.38
DM< −1.96	5	1	1	55	42	31	19
Insign.	82	97	96	46	59	70	82
DM> 1.96	14	3	4	0	0	0	0
CPI: ARX							
Panel DM	−7.95	−7.78	−7.56	−11.25	−17.85	−13.41	−12.22
DM< −1.96	24	24	24	71	68	55	55
Insign.	57	61	63	30	33	41	46
DM> 1.96	20	16	14	0	0	5	0

Note: The row "Panel DM" reports the results of the panel version of the Diebold-Mariano test of Pesaran, Pick and Pranovich (2013). Remaining rows report unit by unit Diebold-Mariano test results. The row "DM< −1.96" reports the number of units with a DM test statistic smaller than −1.96; the row "Insign." reports the number of units with a DM test statistic between −1.96 and 1.96, and the row "DM> 1.96" shows the number of units whose test statistic exceeds 1.96. Each test is for the null hypothesis that the forecasting method in the columns has equal forecast accuracy as the forecasts based on individual estimates. The forecasting methods are listed in the footnote of Table 1.

In contrast, there are between 87 and 116 MSAs for which the Bayesian or combination forecasts are significantly more accurate than the baseline forecasts and only 0-4 cases where the reverse holds.

7.3 CPI inflation of sub-indices

Our second application covers inflation rates for up to 187 sub-indices of the US consumer price index (CPI) obtained from the FRED database. The data is measured at the monthly frequency and spans the period from January 1967 to December 2022. Again, we use rolling estimation windows with 60 observations and require each estimation sample to be balanced, excluding individual series without a complete set of observations in a given window. After accounting for the necessary pre-samples, we generate 599 forecasts for each series, with the first forecast computed for February 1973.

We consider three forecasting models: (i) a purely autoregressive specification with lags 1, 2, and 12, denoted AR; (ii) the same AR specification augmented with the lagged value of the first principal component of the data, denoted AR-PC; and (iii) the AR-PC model augmented with the lagged default yield and lagged term spread, denoted ARX.

The first three columns of Table 9 show that the ratio of average MSFEs of the individual forecasts are beaten by all methods. The pooled forecasts are 6–12% more accurate and beat the baseline forecasts for 44–46% of the series. Pooled forecasts are most accurate for 19–20% of the series but least accurate for 40–44% of the series, which is clearly an unattractive performance.

Individual forecasts of CPI inflation also perform very poorly and generate the most accurate forecasts for less than 2% of the inflation series while conversely generating the least accurate forecasts for more than 40% of the series.

RE and FE forecasts perform similarly to the pooled forecasts on average. They produce more accurate forecasts than the benchmark model for 50–60% of the CPI series and produce the best overall performance for 5% or fewer of the series. While the RE forecasts are the worst overall for 0.5% or fewer of the CPI series, the FE forecasts produce the highest MSFE ratios for 13–17% of the series.

The EB approach records the best performance with improvements in MSFE ratios between 7 and 11%, a 98% frequency of beating the baseline forecasts for the individual series, and the highest frequency (37-48%) with the smallest MSFE values while never producing the least accurate forecasts.

Table 9: CPI inflation forecasting results

Forecast method	Ratio of ave. MSFE			Freq. beating benchmark			Freq. smallest MSFE			Freq. largest MSFE		
	AR	AR-PC	ARX	AR	AR-PC	ARX	AR	AR-PC	ARX	AR	AR-PC	ARX
Individual	14.376	14.401	15.501	–	–	–	0.016	0.016	0.005	0.417	0.433	0.439
Pooled	0.938	0.936	0.878	0.449	0.460	0.444	0.187	0.187	0.203	0.439	0.439	0.396
RE	0.940	0.936	0.880	0.535	0.567	0.508	0.053	0.043	0.032	0.005	0.000	0.000
FE	0.944	0.941	0.883	0.583	0.604	0.508	0.048	0.053	0.000	0.139	0.128	0.166
Emp.Bayes	0.934	0.930	0.892	0.984	0.984	0.984	0.374	0.401	0.481	0.000	0.000	0.000
Hier.Bayes	0.992	0.992	0.968	0.904	0.925	0.936	0.139	0.118	0.096	0.000	0.000	0.000
Comb. (pool)	0.954	0.952	0.917	0.797	0.818	0.775	0.118	0.091	0.139	0.000	0.000	0.000
Comb. (FE)	0.965	0.964	0.927	0.850	0.866	0.850	0.064	0.091	0.043	0.000	0.000	0.000

Note: The column denoted 'AR' reports the results for a panel AR specification, 'AR-PC' adds the first principle component of the panel of sub-indices, and 'ARX' further adds the default yield and the term spread to the model. 'Individual' denotes the average MSFE of forecasts based on individual estimates. The remaining rows report the ratios of MSFEs for the methods listed in the footnote of Table 1. The results are based on estimation windows of 60 observations.

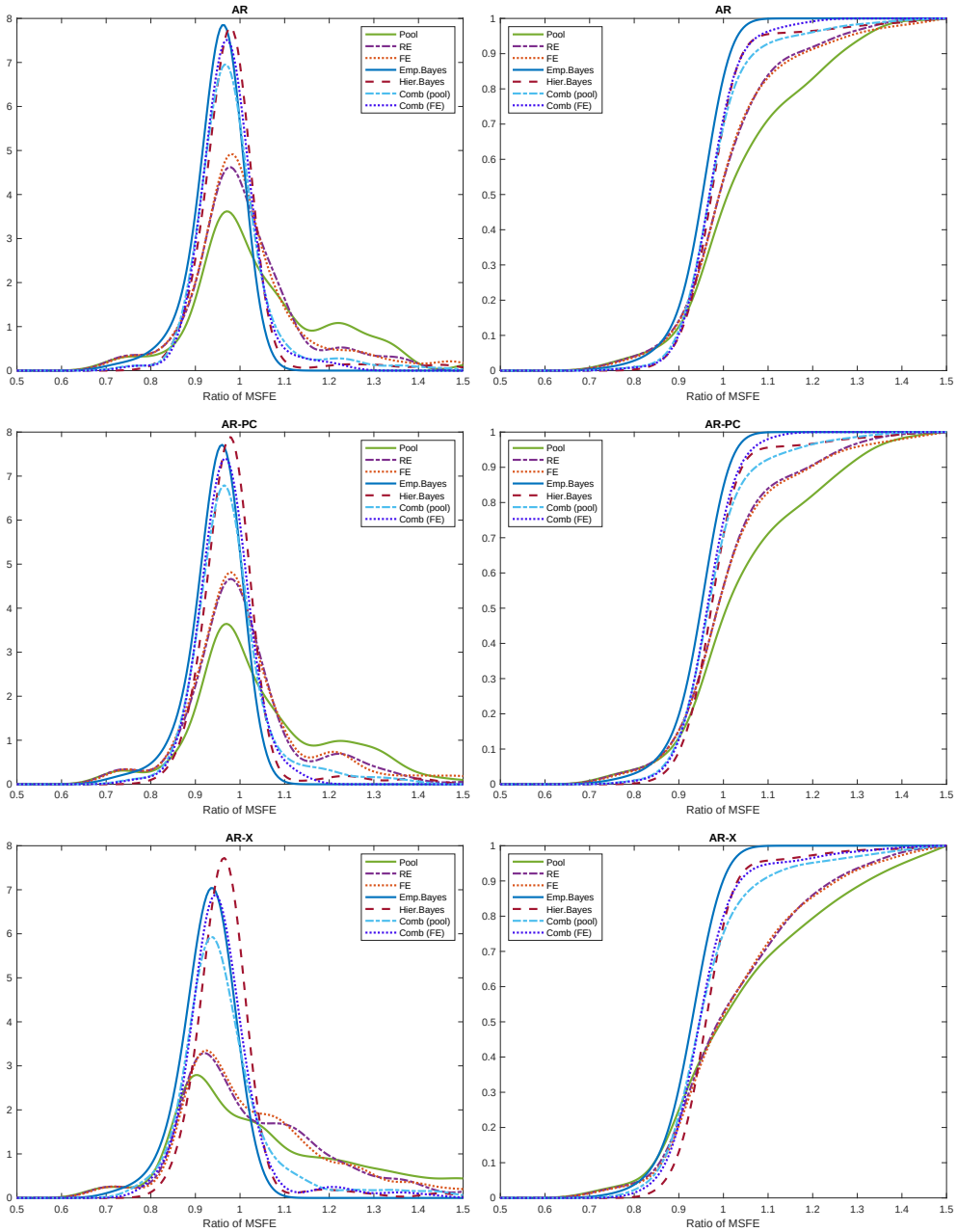
Forecasts from the hierarchical Bayesian and combination schemes beat the benchmark for 78%–94% of the individual series. Forecasts from the pooled combination approach are most accurate for 9–14% of the series versus 4-9% for the FE combination scheme and 10–14% for the hierarchical Bayes approach. These forecasts do not produce the least accurate forecasts for any of the series.

Table 10 shows the quantiles of the distribution of MSFE ratios for the CPI inflation series. The range between the right-most and left-most quantiles is again much wider for the pooled, RE, and FE forecasts than for the EB and combination approaches. The key difference from the housing price application is that the hierarchical Bayes approach now performs very poorly for some of the individual CPI inflation series as witnessed by its right-skewed MSFE ratio distribution. The EB approach is the clear standout model for the CPI inflation data as it improves on the forecasting performance of the baseline approach for more than 95% of the series and underperforms the benchmark by a very small margin even at the 99th percentile of the MSFE ratio distribution.

Our evidence is summarized in Figure 2 which, for the AR model, shows probability density and cumulative density plots of the MSFE ratios across the different CPI sub-indices. The figure clearly highlights the pronounced dispersion and thick right tails of the MSFE-ratio distribution for the pooled, RE, FE, and hierarchical Bayes forecasts. The distributions of MSFE ratios of the EB and combination approaches are far more concentrated and less asymmetrical. The EB approach stands out as having, by far, the thinnest right tail and, hence, the lowest probability of generating forecasts less accurate than those from the individual-specific benchmark.

The DM test results for the CPI data in Table 8 show qualitatively similar findings as for the house price data. All panel models generate significantly negative DM panel test statistics and so their associated CPI inflation forecasts are significantly more accurate, on average, than the baseline forecasts. For the Bayesian and combination approaches, we continue to see many instances with significant improvements in forecast accuracy for individual CPI series relative to the baseline and only five or fewer series for which the reverse holds. In contrast, for the AR and AR-PC model specifications, the baseline forecasts are significantly more accurate than the pooled, RE, and FE forecasts for even more of the individual CPI series than instances where the reverse holds.

Figure 2: Distributions of ratios of MSFE for CPI subindices



Note: These graphs show density (left panels) and cumulative density plots (right panels) of the ratios of MSFEs for the panel AR model in the top row, the panel AR with principal components added in the second row, and the model with all regressors in the third row. The density estimates use a normal kernel with a bandwidth 0.04. The cumulative densities are normalized to 1 at the right tail. The forecasting methods are listed in the footnote of Tables 1.

Table 10: Quantiles for ratios of MSFEs for CPI inflation over subindices

Quantiles	0.01	0.05	0.10	0.50	0.90	0.95	0.99
CPI: AR							
Pooled	0.713	0.833	0.912	1.025	1.517	2.081	2.575
RE	0.722	0.832	0.891	0.984	1.230	1.360	2.036
FE	0.727	0.842	0.893	0.985	1.196	1.347	1.647
Emp.Bayes	0.726	0.852	0.891	0.962	0.982	0.995	1.011
Hier.Bayes	0.863	0.896	0.931	0.984	0.999	1.061	1.521
Comb. (pool)	0.763	0.902	0.916	0.972	1.049	1.156	1.359
Comb. (FE)	0.773	0.902	0.927	0.976	1.018	1.042	1.193
CPI: AR-PC							
Pooled	0.711	0.845	0.906	1.020	1.447	2.101	2.514
RE	0.720	0.851	0.892	0.985	1.238	1.355	1.900
FE	0.725	0.861	0.890	0.984	1.211	1.322	1.490
Emp.Bayes	0.725	0.853	0.878	0.956	0.983	0.993	1.001
Hier.Bayes	0.875	0.911	0.928	0.984	0.997	1.029	1.446
Comb. (pool)	0.773	0.894	0.914	0.970	1.054	1.164	1.290
Comb. (FE)	0.782	0.900	0.918	0.970	1.009	1.037	1.113
CPI: ARX							
Pooled	0.681	0.826	0.874	1.032	1.543	2.412	3.098
RE	0.691	0.832	0.886	0.996	1.370	1.796	2.850
FE	0.694	0.843	0.891	0.996	1.387	1.627	2.921
Emp.Bayes	0.729	0.837	0.866	0.935	0.981	0.990	1.004
Hier.Bayes	0.863	0.896	0.907	0.963	0.987	1.005	1.340
Comb. (pool)	0.786	0.850	0.893	0.940	1.091	1.255	1.450
Comb. (FE)	0.810	0.856	0.902	0.945	1.020	1.079	1.353

Note: The table reports the quantiles of the distribution of the ratios of MSFEs for subindices. The models are listed in the footnote of Table 9 and the forecasting methods are listed in the footnote of Table 1.

8 Conclusion

We provide a comprehensive examination of the out-of-sample predictive accuracy of a large set of novel and existing panel forecasting methods, including individual estimation, pooled estimation, random effects, fixed effects, empirical and hierarchical Bayes, and forecast combination. Our analysis characterizes analytically the determinants of squared error performance as it relates to bias and estimation error variance components. We provide insights into how parameter heterogeneity, predictive power, and sample sizes regulate the bias-variance trade-off that determines predictive accuracy. To quantify the importance of these theoretical points in practice, we study Monte Carlo simulations and consider two empirical applications to house prices and CPI inflation.

Our main findings can be summarized in three points. First, we find that many panel forecasting approaches perform systematically better than forecasts based on individual estimates. For panels with a small or medium-sized time-series dimension T–a setting relevant to many empirical applications in economics–our Monte Carlo simulations and empirical applications demonstrate sizeable gains from exploiting panel information to obtain forecasts that are more accurate than forecasts from individual-specific models both on average and also for the majority of individual units.

Second, our analytical results and Monte Carlo simulations show that one should not expect a single forecasting approach to be uniformly dominant across applications that differ in terms of the cross-sectional and time-series dimensions, strength of predictive power, and degree of heterogeneity in intercept and slope coefficients along with how correlated this heterogeneity is.

Forecasts based on pooled estimates are most accurate only in situations with little or no parameter heterogeneity and a small T dimension, while forecasts based on FE and RE estimates perform relatively well mainly when heterogeneity is confined to model intercepts and T is small. Neither of these approaches perform well in settings with high levels of heterogeneity where individual-specific forecasts tend to perform better, particularly if T is relatively large. By over-weighting forecasts that perform well and underweighting forecasts that perform poorly, forecast combination and empirical Bayes methods manage to produce the most accurate forecasts across a broad range of settings.

Third, the panel forecasting methods clearly differ in terms of their risk profiles, particularly their ability to reduce the probability of generating very poor forecasts for individual units in a cross-section. While the individual, pooled, random and fixed effect estimation methods perform poorly in some of the simulations and empirical applications, the forecast combination and empirical Bayes methods rarely generate the least accurate forecasts for individual units and retain some probability of being the best forecasting method. These panel forecasting approaches therefore come out on top of our analysis.

In a nutshell, our simulations and empirical applications suggest that forecast combinations and Bayesian panel methods offer insurance against poor performance. Compared to the alternative forecasting methods we consider, this better "risk-return" trade-off makes the combination and Bayes methods attractive in forecast applications with panel data.

References

Baltagi, B.H. (2008) "Forecasting with panel data" *Journal of Forecasting* 27, 153–173.

Baltagi, B.H. (2013) "Panel data forecasting" Ch. 18 in Elliott, G. and A. Timmermann, *Handbook of Economic Forecasting*, volume 2B. North Holland: Elsevier.

Boot, T., and A. Pick (2020) "Does modeling a structural break improve forecast accuracy?" *Journal of Econometrics* 215, 35–59.

Brown, L, and Greenshtein, E., 2009, Non-parametric empirical Bayes and compound decision approaches to estimation of a high-dimensional vector of normal means. *Annals of Statistics* 37, 1685–1704.

Brückner, H., and B. Siliverstovs (2006) "On the estimation and forecasting of international migration: how relevant is heterogeneity across countries" *Empirical Economics* 31, 735–754.

Diebold, F.X., and R.S. Mariano, (1995) 'Comparing predictive accuracy.' *Journal of Business & Economic Statistics* 13, 253–263.

Efron, B. (2011). Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106, 1602–1614.

Elliott, G. (2017). Forecast combination when outcomes are difficult to predict. *Empirical Economics*, 53, 7–20.

Gelfand, A.E., S.E. Hills, A. Racine-Poon, A.F.M. Smith (1996). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, 85, 972–985.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1, 515–533.

Goldberger, A.S. (1962) "Best linear unbiased prediction in the generalized linear regression model" *Journal of the American Statistical Association* 57, 369–375.

Gu, J., and R. Koenker, 2017, Unobserved heterogeneity in income dynamics: An empirical Bayes perspective. *Journal of Business & Economic Statistics* 35, 1–16.

Hsiao, C. (2022) *Analysis of Panel Data*, Fourth Edition. Econometric Society Monograph, Cambridge University Press, Cambridge.

Lahiri, K., Peng, H., & Zhao, Y. (2017). Online learning and forecast combination in unbalanced panels. *Econometric Reviews*, 36, 257–288.

Lee, L.-F., and W.E. Griffiths (1979) "The prior likelihood and best linear unbiased prediction in stochastic coefficient linear models" Center for Economic Research, University of Minnesota, Discussion paper 79–107.

Lindley D.V. and A.F.M. Smith (1972) Bayesian estimates for the linear model. *Journal of the Royal Statistical Society, Series B* 34, 1–41.

Liu, L. (2023) Density Forecasts in panel data models: A semiparametric Bayesian perspective. *Journal of Business & Economic Statistics* 41, 349–363.

Liu, L., H.R. Moon, and F. Schorfheide (2020) "Forecasting with dynamic panel data models" *Econometrica* 88, 171–201.

Liu, L., H.R. Moon, and F. Schorfheide (2023) Forecasting with a panel Tobit model. *Quantitative Economics* 14, 117–159.

Maddala, G.S., R.P. Trost, H. Li, F. Joutz (1997) "Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage priors" *Journal of Business & Economic Statistics* 15, 90–100.

Hsiao C., M.H. Pesaran and A.K. Tahmiscioglu (1999) Bayes estimation of short-run coefficients in dynamic panel data models. Ch.11 in C. Hsiao, K. Lahiri, L.-F. Lee and M.H.Pesaran *A*nalysis of panels and limited dependent variable models, Cambridge: Cambridge University Press.

Pesaran, M.H., A. Pick, and M. Pranovich (2013) "Optimal forecasts in the presence of structural breaks" *Journal of Econometrics* 177, 79–113.

Pesaran, M.H., R. Smith (1995) "Estimating long-run relationships from dynamic heterogeneous panels" *Journal of Econometrics* 68, 134–152.

Pesaran, M.H., and L. Yang, 2023, "Trimmed mean group estimation of average treatment effects in ultra short T panels under correlated heterogeneity" https://arxiv.org/abs/2310.11680

Timmermann, A. (2006) "Forecast combinations"Ch. 4 in G. Elliott, C. W. J. Granger and A. Timmermann (Eds.) *Handbook of Economic Forecasting*, volume. 1, North Holland: Elsevier.

Trapani, L. and G. Urga (2009) "Optimal forecasting with heterogeneous panels: A Monte Carlo Study"*International Journal of Forecasting* 25, 567–586.

Wang, W., X. Zhang, and R. Paap (2019) "To pool of not to pool: What is a good strategy for parameter estimation and forecasting in panels" *Journal of Applied Econometrics* 34, 724–745.

Yang, Cynthia F. (2021) "Common factors and spatial dependence: An application to US house prices" *Econometric Reviews* 40, 14–50.

Mathematical Appendix

A.1 Lemmas

Lemma 1 Suppose that Assumptions 8 and 9 hold, then for a fixed $T > T_0$ we have

$$\bar{\boldsymbol{Q}}_{NT} - \mathrm{E}\left(\bar{\boldsymbol{Q}}_{NT}\right) = O_p(N^{-1/2}), \text{ and } \bar{\boldsymbol{q}}_{NT} - \mathrm{E}\left(\bar{\boldsymbol{q}}_{NT}\right) = O_p\left(N^{-1/2}\right), \tag{A.1}$$

and

$$\bar{\boldsymbol{Q}}_{NT}^{-1} - \mathrm{E}\left(\bar{\boldsymbol{Q}}_{NT}\right)^{-1} = O_p(N^{-1/2}), \tag{A.2}$$

where $\bar{\boldsymbol{Q}}_{NT} = N^{-1}\sum_{i=1}^{N}\boldsymbol{Q}_{iT}$, $\bar{\boldsymbol{q}}_{NT} = N^{-1}\sum_{i=1}^{N}\boldsymbol{q}_{iT}$, $\boldsymbol{Q}_{iT} = T^{-1}\sum_{t=1}^{T}\boldsymbol{w}_{it}\boldsymbol{w}_{it}'$, and $\boldsymbol{q}_{iT} = T^{-1}\sum_{t=1}^{T}\boldsymbol{w}_{it}\boldsymbol{w}_{it}'\boldsymbol{\eta}_i$. Further, under Assumptions 3 and 5

$$\mathrm{E}\left(\bar{\boldsymbol{Q}}_{NT}\right) = \bar{\boldsymbol{Q}}_N, \text{ and } \mathrm{E}\left(\bar{\boldsymbol{q}}_{NT}\right) = \bar{\boldsymbol{q}}_N, \tag{A.3}$$

where $\bar{\boldsymbol{Q}}_N = N^{-1}\sum_{i=1}^{N}\boldsymbol{Q}_i$, $\bar{\boldsymbol{q}}_N = N^{-1}\sum_{i=1}^{N}\boldsymbol{q}_i$, $\boldsymbol{Q}_i = \mathrm{E}(\boldsymbol{w}_{it}\boldsymbol{w}_{it}')$, and $\boldsymbol{q}_i = \mathrm{E}(\boldsymbol{w}_{it}\boldsymbol{w}_{it}'\boldsymbol{\eta}_i)$.

Proof Note that

$$\bar{\boldsymbol{Q}}_{NT} - \mathrm{E}\left(\bar{\boldsymbol{Q}}_{NT}\right) = N^{-1}\sum_{i=1}^{N}\left[\boldsymbol{Q}_{iT} - E\left(\boldsymbol{Q}_{iT}\right)\right], \text{ and } \bar{\boldsymbol{q}}_{NT} - \mathrm{E}\left(\bar{\boldsymbol{q}}_{NT}\right) = N^{-1}\sum_{i=1}^{N}\left[\boldsymbol{q}_{iT} - E\left(\boldsymbol{q}_{iT}\right)\right].$$

Under Assumptions 3 and 9, the elements of $\boldsymbol{Q}_{iT} - \mathrm{E}\left(\boldsymbol{Q}_{iT}\right)$ and $\boldsymbol{q}_{iT} - \mathrm{E}\left(\boldsymbol{q}_{iT}\right)$ are independently distributed with mean zero and finite variances. Therefore, (A.1) follows. Also

$$\begin{aligned}\left\|\bar{\boldsymbol{Q}}_{NT}^{-1} - \mathrm{E}\left(\bar{\boldsymbol{Q}}_{NT}\right)^{-1}\right\| &= \left\|\bar{\boldsymbol{Q}}_{NT}^{-1}\left[\bar{\boldsymbol{Q}}_{NT} - \mathrm{E}\left(\bar{\boldsymbol{Q}}_{NT}\right)\right]\mathrm{E}\left(\bar{\boldsymbol{Q}}_{NT}\right)^{-1}\right\| \\ &\leq \left\|\bar{\boldsymbol{Q}}_{NT}^{-1}\right\|\left\|\bar{\boldsymbol{Q}}_{NT} - \mathrm{E}\left(\bar{\boldsymbol{Q}}_{NT}\right)\right\|\left\|\mathrm{E}\left(\bar{\boldsymbol{Q}}_{NT}\right)^{-1}\right\|,\end{aligned}$$

and, by Assumption 8, $\left\|\bar{\boldsymbol{Q}}_{NT}^{-1}\right\| = \lambda_{\max}\left(\bar{\boldsymbol{Q}}_{NT}^{-1}\right) < C$, and $\left\|\mathrm{E}\left(\bar{\boldsymbol{Q}}_{NT}\right)^{-1}\right\| = \left\|\bar{\boldsymbol{Q}}_{N}^{-1}\right\| = O(1)$. Hence, $\left\|\bar{\boldsymbol{Q}}_{NT}^{-1} - \mathrm{E}\left(\bar{\boldsymbol{Q}}_{NT}\right)^{-1}\right\|$ has the same order as $\left\|\bar{\boldsymbol{Q}}_{NT}^{-1} - \mathrm{E}\left(\bar{\boldsymbol{Q}}_{NT}\right)^{-1}\right\| = O_p(N^{-1/2})$, as required. Result (A.3) follows from the stationarity properties, $\boldsymbol{Q}_i = \mathrm{E}(\boldsymbol{w}_{it}\boldsymbol{w}_{it}')$ and

$\boldsymbol{q}_i = \mathrm{E}(\boldsymbol{w}_{it}\boldsymbol{w}_{it}'\boldsymbol{\eta}_i)$.

Lemma 2 Under Assumptions 1-9

$$\sup_{i,T} \mathrm{E}\left\|\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)\right\|^s < C, \ s = 1, 2, \tag{A.4}$$

where $\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i = (\boldsymbol{W}_i'\boldsymbol{W}_i)^{-1}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i$, and

$$\sup_{i,T} \mathrm{E}\left\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_i\right\| < C, \tag{A.5}$$

$\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_i = -\boldsymbol{\eta}_i + \bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT} + \bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{\xi}}_{NT}$.

Proof Since $\left\|\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)\right\| \leq \left\|\boldsymbol{Q}_{iT}^{-1}\right\| \left\|T^{-1/2}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\right\|$, then

$$\left\|\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)\right\|^2 \leq \left\|\boldsymbol{Q}_{iT}^{-1}\right\|^2 \left\|T^{-1/2}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\right\|^2,$$

and by the Cauchy–Schwarz inequality

$$\begin{aligned}
\sup_{i,T} \mathrm{E}\left\|\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)\right\|^2 &\leq \left(\sup_{i,T} \mathrm{E}\left\|\boldsymbol{Q}_{iT}^{-1}\right\|^4\right)^{1/2} \left(\sup_{i,T} \mathrm{E}\left\|T^{-1/2}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\right\|^4\right)^{1/2} \\
&= \left\{\sup_{i,T} \mathrm{E}\left[\lambda_{\max}^4\left(\boldsymbol{Q}_{iT}^{-1}\right)\right]\right\}^{1/2} \left(\sup_{i,T} \mathrm{E}\left\|T^{-1/2}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\right\|^4\right)^{1/2}.
\end{aligned}$$

Both of the terms on the right hand side of the above are bounded under Assumption 4, and we have $\sup_{i,T}\mathrm{E}\left\|\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)\right\|^2 < C$. This result in turn implies $\sup_{i,T}\mathrm{E}\left\|\sqrt{T}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)\right\| < C$, and result (A.4) follows. Regarding $\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_i$, we first note that

$$\left\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_i\right\| \leq \|\boldsymbol{\eta}_i\| + \left\|\bar{\boldsymbol{Q}}_{NT}^{-1}\right\| \|\bar{\boldsymbol{q}}_{NT}\| + \left\|\bar{\boldsymbol{Q}}_{NT}^{-1}\right\| \left\|\bar{\boldsymbol{\xi}}_{NT}\right\|,$$

and

$$\mathrm{E}\left\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_i\right\| \leq \mathrm{E}\|\boldsymbol{\eta}_i\| + \left(\mathrm{E}\left\|\bar{\boldsymbol{Q}}_{NT}^{-1}\right\|^2\right)^{1/2} \left(\mathrm{E}\|\bar{\boldsymbol{q}}_{NT}\|^2\right)^{1/2} + \left(\mathrm{E}\left\|\bar{\boldsymbol{Q}}_{NT}^{-1}\right\|^2\right)^{1/2} \left(\mathrm{E}\left\|\bar{\boldsymbol{\xi}}_{NT}\right\|^2\right)^{1/2}. \tag{A.6}$$

Under Assumption 5, $\mathrm{E}\|\boldsymbol{\eta}_i\| < C$ and $\sup_{i,t}\mathrm{E}\|\boldsymbol{w}_{it}\boldsymbol{w}_{it}'\boldsymbol{\eta}_i\|^2 < C$. Also by the Cauchy-Schwarz inequality

$$\mathrm{E}\left\|\boldsymbol{w}_{it}\varepsilon_{it}\right\|^2 \leq \left(\mathrm{E}\left\|\boldsymbol{w}_{it}\right\|^4\right)^{1/2}\left(\left|\varepsilon_{it}\right|^4\right)^{1/2},$$

and, under Assumptions 1 and 3 we have $sup_{i,t}\mathrm{E}\|\boldsymbol{w}_{it}\varepsilon_{it}\|^2 < C$. Then, applying Minkowski's inequality to $\bar{\boldsymbol{\xi}}_{NT} = N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{w}_{it}\varepsilon_{it}$,

$$\mathrm{E}\left\|\bar{\boldsymbol{\xi}}_{NT}\right\|_2 = \left(\mathrm{E}\left\|\bar{\boldsymbol{\xi}}_{NT}\right\|^2\right)^{1/2} \leq N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}\left\|\boldsymbol{w}_{it}\varepsilon_{it}\right\|_2 \leq \sup_{i,t}\left(\mathrm{E}\left\|\boldsymbol{w}_{it}\varepsilon_{it}\right\|^2\right)^{1/2},$$

and it follows that $\mathrm{E}\left\|\bar{\boldsymbol{\xi}}_{NT}\right\|^2 < C$. Similarly, since $\bar{\boldsymbol{q}}_{NT} = N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{w}_{it}\boldsymbol{w}_{it}'\boldsymbol{\eta}_i$ and $\sup_{i,t}\mathrm{E}\left\|\boldsymbol{w}_{it}\boldsymbol{w}_{it}'\boldsymbol{\eta}_i\right\|^2 < C$, then $\mathrm{E}\left\|\bar{\boldsymbol{q}}_{NT}\right\|^2 < C$. Also, by Assumption 8, $\left\|\bar{\boldsymbol{Q}}_{NT}^{-1}\right\|^2 = \lambda_{\max}\left(\bar{\boldsymbol{Q}}_{NT}^{-2}\right) < C$. Using these results in (A.6) now yields (A.5).

A.2 Proofs of the propositions

A.2.1 Proof of Proposition 1

Let $\boldsymbol{P}_i = \boldsymbol{W}_i(\boldsymbol{W}_i'\boldsymbol{W}_i)^{-1}$. Using (15), note that

$$\mathrm{E}\left(r_{iT}\left|\boldsymbol{\varepsilon}_i, \boldsymbol{W}_i, \boldsymbol{w}_{i,T+1}\right.\right) = \left(\boldsymbol{\varepsilon}_i'\boldsymbol{P}_i\boldsymbol{w}_{i,T+1}\right)\mathrm{E}\left(\varepsilon_{i,T+1}\left|\boldsymbol{\varepsilon}_i, \boldsymbol{W}_i, \boldsymbol{w}_{i,T+1}\right.\right),$$

and, under Assumptions 1 and 2, $\mathrm{E}\left(\varepsilon_{i,T+1}\left|\boldsymbol{\varepsilon}_i, \boldsymbol{W}_i, \boldsymbol{w}_{i,T+1}\right.\right) = 0$, for all i. Hence, unconditionally $\mathrm{E}\left(r_{iT}\right.) = 0$. Furthermore, $|r_{iT}| \leq \|\boldsymbol{\varepsilon}_i'\boldsymbol{P}_i\|\,\|\boldsymbol{w}_{i,T+1}\|\,|\varepsilon_{i,T+1}|$ and $|\varepsilon_{i,T+1}|$ is distributed independently of $\boldsymbol{w}_{i,T+1}$ and $T^{-1}\boldsymbol{\varepsilon}_i'\boldsymbol{P}_i$. Hence by the Cauchy–Schwarz inequality

$$\mathrm{E}\left|r_{iT}\right| \leq \left[\mathrm{E}\left\|\boldsymbol{\varepsilon}_i'\boldsymbol{P}_i\right\|^2\right]^{1/2}\left(\mathrm{E}\left\|\boldsymbol{w}_{i,T+1}\right\|^2\right)^{1/2}\mathrm{E}\left|\varepsilon_{i,T+1}\right|.$$

Again, under Assumption 1, $\sup_{i,T} \mathrm{E}\,|\varepsilon_{i,T+1}| < C$ and $\sup_{i,T} \mathrm{E}\,\|\boldsymbol{w}_{i,T+1}\|^2 < C$. Also, since $\boldsymbol{Q}_{iT}^{-1}$ is symmetric, $\left\|\boldsymbol{Q}_{iT}^{-1}\right\|^2 = \lambda^2_{\max}\left(\boldsymbol{Q}_{iT}^{-1}\right)$ and we have

$$\begin{aligned} \left\|\boldsymbol{\varepsilon}_i'\boldsymbol{P}_i\right\|^2 &= \left\|T^{-1}\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i(T^{-1}\boldsymbol{W}_i'\boldsymbol{W}_i)^{-1}\right\|^2 \leq T^{-1}\left\|\boldsymbol{Q}_{iT}^{-1}\right\|^2 \left\|T^{-1/2}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\right\|^2 \\ &\leq \lambda^2_{\max}\left(\boldsymbol{Q}_{iT}^{-1}\right)\left\|T^{-1}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\right\|^2. \end{aligned} \tag{A.7}$$

By the Cauchy-Schwarz inequality and under Assumption 4

$$\sup_{i,T} \mathrm{E}\left\|\boldsymbol{\varepsilon}_i'\boldsymbol{P}_i\right\|^2 \leq \left\{\sup_{i,T} \mathrm{E}\left[\lambda^2_{\max}\left(\boldsymbol{Q}_{iT}^{-1}\right)\right]\right\}^{1/2}\left[\sup_{i,T}\left\|T^{-1}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\right\|^4\right]^{1/2} < C,$$

and $\sup_{i,T} \mathrm{E}\,|r_{iT}| < C$. Finally, under Assumption 9, r_{iT} are independently distributed over i. Then, by the law of large numbers for independently distributed processes with zero means we have

$$R_{NT} = O_p(N^{-1/2}). \tag{A.8}$$

Consider now S_{NT} and note that

$$S_{NT} = N^{-1}\sum_{i=1}^{N} E(s_{iT}) + N^{-1}\sum_{i=1}^{N}\left[s_{iT} - E(s_{iT})\right],$$

where s_{iT} is given by (16). Under Assumption 9, s_{iT} is distributed independently across i, and the second term of S_{NT} will be $O_p(N^{-1/2})$ if $\sup_{i,T} \mathrm{E}\,|s_{iT}| < C$. Also

$$|s_{iT}| \leq \|\boldsymbol{w}_{i,T+1}\|^2\left\|\boldsymbol{Q}_{iT}^{-1}\right\|^2\left\|T^{-1/2}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\right\|^2,$$

and $\sup_{i,T}\|\boldsymbol{w}_{i,T+1}\| < C$. Hence, $\sup_{i,T} \mathrm{E}\,|s_{iT}| < C$ follows if

$$\sup_{i,T} \mathrm{E}\left[\left\|\boldsymbol{Q}_{iT}^{-1}\right\|^2\left\|T^{-1/2}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\right\|^2\right] < C.$$

This condition is satisfied by Assumptions 1 and 4, noting that by the Cauchy–Schwarz inequality

$$\mathrm{E}\left[\left\|\boldsymbol{Q}_{iT}^{-1}\right\|^2\left\|T^{-1/2}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\right\|^2\right] \leq \left[\mathrm{E}\left\|\boldsymbol{Q}_{iT}^{-1}\right\|^4\right]^{1/2}\left[\mathrm{E}\left\|T^{-1/2}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\right\|^4\right]^{1/2},$$

and $\left\|\boldsymbol{Q}_{iT}^{-1}\right\|^4 = \lambda_{\max}^4\left(\boldsymbol{Q}_{iT}^{-1}\right)$. Therefore, $S_{NT} = \mathrm{E}\left(S_{NT}\right) + O_p(N^{-1/2})$, where $\mathrm{E}\left(S_{NT}\right) = N^{-1}\sum_{i=1}^{N}\mathrm{E}(s_{iT}) = h_{NT}$, and the result in equation (17) follows, with h_{NT} given by (18). When the regressors $\boldsymbol{w}_{it}$ are strictly exogenous, using (16), we have (under Assumption 1)

$$\begin{aligned}\mathrm{E}(s_{iT}) &= \mathrm{E}([\mathrm{E}\left(s_{iT}\left|\boldsymbol{w}_{i,T+1},\boldsymbol{W}_i\right.\right)] = \mathrm{E}\left[\boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_{iT}^{-1}\left(\frac{\boldsymbol{W}_i' E\left(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\right)\boldsymbol{W}_i}{T}\right)\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}\right] \\ &= E\left[\sigma_i^2\boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_{iT}^{-1}\left(\frac{\boldsymbol{W}_i'\boldsymbol{W}_i}{T}\right)\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}\right] = \sigma_i^2 E\left(\boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}\right),\end{aligned}$$

and the result in equation (19) follows.

A.2.2 Proof of Proposition 2

The average MSFE of forecasts based on pooled estimates is given by (20) which we reproduce here for convenience.

$$N^{-1}\sum_{i=1}^{N}\tilde{e}_{i,T+1}^2 = N^{-1}\sum_{i=1}^{N}\varepsilon_{i,T+1}^2 + N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\boldsymbol{w}_{i,T+1} + \tilde{S}_{N,T+1} + 2\tilde{R}_{N,T+1}, \tag{A.9}$$

where

$$\begin{aligned}\tilde{S}_{N,T+1} &= \bar{\boldsymbol{q}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{Q}}_{N,T+1}\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT} + \bar{\boldsymbol{\xi}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{Q}}_{N,T+1}\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{\xi}}_{NT} \\ &\quad -2\bar{\boldsymbol{q}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\,\bar{\boldsymbol{q}}_{N,T+1} - 2\bar{\boldsymbol{\xi}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{N,T+1} + 2\bar{\boldsymbol{\xi}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{Q}}_{N,T+1}\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT},\end{aligned} \tag{A.10}$$

$$\tilde{R}_{N,T+1} = N^{-1}\sum_{i=1}^{N}\boldsymbol{\eta}_i'\boldsymbol{w}_{i,T+1}\varepsilon_{i,T+1} - \left(\bar{\boldsymbol{q}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1} + \bar{\boldsymbol{\xi}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\right)\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}\varepsilon_{i,T+1}\right), \tag{A.11}$$

and

$$\bar{\boldsymbol{Q}}_{N,T+1} = N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}', \text{ and } \bar{\boldsymbol{q}}_{N,T+1} = N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i. \tag{A.12}$$

Under Assumption 7, $\bar{\boldsymbol{\xi}}_{NT} = O_p\left(N^{-1/2}\right)$. Using Lemma A.1, we have $\bar{\boldsymbol{Q}}_{NT}^{-1} = \bar{\boldsymbol{Q}}_{N}^{-1} + O_p\left(N^{-1/2}\right) = O_p(1)$, and similarly $\bar{\boldsymbol{q}}_{NT} = O_p(1)$ and $\bar{\boldsymbol{q}}_{N,T+1} = O_p(1)$. Using these results in (A.10) we now have

$$\tilde{S}_{N,T+1} = \bar{\boldsymbol{q}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{Q}}_{N,T+1}\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT} - 2\bar{\boldsymbol{q}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{N,T+1} + O_p\left(N^{-1/2}\right). \tag{A.13}$$

Note that under stationarity (see Assumptions 3 and 5), $\mathrm{E}\left(\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\right) = \boldsymbol{q}_i$, $\mathrm{E}\left(\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\right) = \boldsymbol{Q}_i$, and consider

$$\begin{aligned}\bar{\boldsymbol{q}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{Q}}_{N,T+1}\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT} = (\boldsymbol{\Delta}_{q,NT} + \bar{\boldsymbol{q}}_N)'\left(\boldsymbol{\Delta}_{Q,NT} + \bar{\boldsymbol{Q}}_N^{-1}\right)\left(\bar{\boldsymbol{Q}}_{N,T+1} - \bar{\boldsymbol{Q}}_N + \bar{\boldsymbol{Q}}_N\right) \\ \times \left(\boldsymbol{\Delta}_{Q,NT} + \bar{\boldsymbol{Q}}_N^{-1}\right)\left(\bar{\boldsymbol{q}}_{N,T+1} - \bar{\boldsymbol{q}}_N + \bar{\boldsymbol{q}}_N\right),\end{aligned}$$

where (by Lemma A.1)

$$\boldsymbol{\Delta}_{Q,NT} = \bar{\boldsymbol{Q}}_{NT}^{-1} - \bar{\boldsymbol{Q}}_N^{-1} = O_p(N^{-1/2}),\ \boldsymbol{\Delta}_{q,NT} = \bar{\boldsymbol{q}}_{NT} - \bar{\boldsymbol{q}}_N = O_p(N^{-1/2}),$$

and $\bar{\boldsymbol{Q}}_N$ and $\bar{\boldsymbol{q}}_N$ are defined by (6) and (8), respectively. Also note that

$$\begin{aligned}\bar{\boldsymbol{Q}}_{N,T+1} &= N^{-1}\sum_{i=1}^{N}\mathrm{E}\left(\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\right) + O_p(N^{-1/2}) = \bar{\boldsymbol{Q}}_N + O_p(N^{-1/2}), \\ \bar{\boldsymbol{q}}_{N,T+1} &= N^{-1}\sum_{i=1}^{N}\mathrm{E}\left(\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\right) + O_p(N^{-1/2}) = \bar{\boldsymbol{q}}_N + O_p(N^{-1/2}).\end{aligned}$$

Hence, it readily follows that

$$\bar{\boldsymbol{q}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{Q}}_{N,T+1}\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT} = \bar{\boldsymbol{q}}_N'\bar{\boldsymbol{Q}}_N^{-1}\bar{\boldsymbol{q}}_N + O_p\left(N^{-1/2}\right). \tag{A.14}$$

Similarly, $\bar{\boldsymbol{q}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{N,T+1} = \bar{\boldsymbol{q}}_N'\bar{\boldsymbol{Q}}_N^{-1}\bar{\boldsymbol{q}}_N + O_p\left(N^{-1/2}\right)$, and as a result

$$\tilde{S}_{N,T+1} = -\bar{\boldsymbol{q}}_N'\bar{\boldsymbol{Q}}_N^{-1}\bar{\boldsymbol{q}}_N + O_p\left(N^{-1/2}\right).$$

Finally, since $\varepsilon_{i,T+1}$ (which has zero mean) is distributed independently of $\boldsymbol{w}_{i,T+1}$ and $\boldsymbol{\eta}_i$, under Assumption 9,

$$N^{-1}\sum_{i=1}^{N}\boldsymbol{\eta}_i'\boldsymbol{w}_{i,T+1}\varepsilon_{i,T+1}=O_p(N^{-1/2}),\text{ and }N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}\varepsilon_{i,T+1}=O_p(N^{-1/2}),$$

and $\tilde{R}_{N,T+1}=O_p(N^{-1/2})$, noting that $\left(\bar{\boldsymbol{q}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}+\bar{\boldsymbol{\xi}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\right)=O_p(1)$. Using this result and (A.14) in (20) now yields

$$N^{-1}\sum_{i=1}^{N}\tilde{e}_{i,T+1}^2=N^{-1}\sum_{i=1}^{N}\varepsilon_{i,T+1}^2+N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\boldsymbol{w}_{i,T+1}-\bar{\boldsymbol{q}}_N'\bar{\boldsymbol{Q}}_N^{-1}\bar{\boldsymbol{q}}_N+O_p\left(N^{-1/2}\right).\quad\text{(A.15)}$$

Also, under Assumption 9, $\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\boldsymbol{w}_{i,T+1}$ is independently distributed over i and we have, noting that under Assumption 5, $\sup_{i,T}\mathrm{E}\left|\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\boldsymbol{w}_{i,T+1}\right|=\sup_{i,T}\mathrm{E}\left\|\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\right\|^2<C$,

$$N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\boldsymbol{w}_{i,T+1}=N^{-1}\sum_{i=1}^{N}\mathrm{E}\left(\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\boldsymbol{w}_{i,T+1}\right)+O_p\left(N^{-1/2}\right).$$

Using this result in (A.15) now yields equation (21). To establish (23), note that when the heterogeneity is uncorrelated, $\bar{\boldsymbol{q}}_N=0$, and $\mathrm{E}\left(\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\boldsymbol{w}_{i,T+1}\right)=\mathrm{E}\left[\mathrm{tr}\left(\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\right)\right]=\mathrm{tr}\left[\mathrm{E}\left(\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\right)\mathrm{E}\left(\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\right)\right]$. Also, under Assumptions 3 and 5, $\mathrm{E}\left(\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\right)\mathrm{E}\left(\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\right)=\boldsymbol{Q}_i\boldsymbol{\Omega}_\eta$, and $N^{-1}\sum_{i=1}^{N}\mathrm{E}\left(\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\boldsymbol{w}_{i,T+1}\right)=\mathrm{tr}\left(\bar{\boldsymbol{Q}}_N\boldsymbol{\Omega}_\eta\right)$, as required.

A.2.3 Proof of Proposition 3

Using (12) and (13),

$$\begin{aligned}N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}\tilde{e}_{i,T+1} &= N^{-1}\sum_{i=1}^{N}\varepsilon_{i,T+1}^2+N^{-1}\sum_{i=1}^{N}(\hat{\boldsymbol{\theta}}_i-\boldsymbol{\theta}_i)'\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_i)\qquad\text{(A.16)}\\ &\quad -N^{-1}\sum_{i=1}^{N}(\hat{\boldsymbol{\theta}}_i-\boldsymbol{\theta}_i)'\boldsymbol{w}_{i,T+1}\varepsilon_{i,T+1}-N^{-1}\sum_{i=1}^{N}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_i)'\boldsymbol{w}_{i,T+1}\varepsilon_{i,T+1},\end{aligned}$$

where $\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_i=-\boldsymbol{\eta}_i+\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT}+\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{\xi}}_{NT}$, and $\hat{\boldsymbol{\theta}}_i-\boldsymbol{\theta}_i=(\boldsymbol{W}_i'\boldsymbol{W}_i)^{-1}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i$. Noting that the third term in the above, apart from the minus sign, is the same as R_{NT} defined below (14), by (A.8) it

follows that

$$N^{-1}\sum_{i=1}^{N}(\hat{\boldsymbol{\theta}}_i-\boldsymbol{\theta}_i)'\boldsymbol{w}_{i,T+1}\varepsilon_{i,T+1}=N^{-1}\sum_{i=1}^{N}r_{iT}=R_{NT}=O_p(N^{-1/2}). \tag{A.17}$$

Further,

$$\begin{aligned}N^{-1}\sum_{i=1}^{N}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_i)'\boldsymbol{w}_{i,T+1}\varepsilon_{i,T+1} &= -N^{-1}\sum_{i=1}^{N}\boldsymbol{\eta}_i'\boldsymbol{w}_{i,T+1}\varepsilon_{i,T+1}+\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}\varepsilon_{i,T+1}\right)\\ &\quad+\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{\xi}}_{NT}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}\varepsilon_{i,T+1}\right).\end{aligned}$$

By Lemma A.1, $\bar{\boldsymbol{Q}}_{NT}^{-1}=O_p(1)$ and $\bar{\boldsymbol{q}}_{NT}=O_p(1)$, and by Assumption 7 ,$\bar{\boldsymbol{\xi}}_{NT}=O_p(N^{-1/2})$. Also, under Assumptions 1 and 6, $\boldsymbol{\eta}_i'\boldsymbol{w}_{i,T+1}\varepsilon_{i,T+1}$ and $\boldsymbol{w}_{i,T+1}\varepsilon_{i,T+1}$ have mean zero and first order moments. Hence, given Assumption 9 we have

$$N^{-1}\sum_{i=1}^{N}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_i)'\boldsymbol{w}_{i,T+1}\varepsilon_{i,T+1}=O_p\left(N^{-1/2}\right). \tag{A.18}$$

Consider now the second term of (A.16):

$$\begin{aligned}&N^{-1}\sum_{i=1}^{N}(\hat{\boldsymbol{\theta}}_i-\boldsymbol{\theta}_i)'\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_i)\\ =\;&N^{-1}\sum_{i=1}^{N}\left(-\boldsymbol{\eta}_i+\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT}+\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{\xi}}_{NT}\right)'\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'(\boldsymbol{W}_i'\boldsymbol{W}_i)^{-1}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\\ =\;&T^{-1/2}\left[N^{-1}\sum_{i=1}^{N}\left(-\boldsymbol{\eta}_i'+\bar{\boldsymbol{q}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\right)\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'(T^{-1}\boldsymbol{W}_i'\boldsymbol{W}_i)^{-1}T^{-1/2}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\right]\\ &+T^{-1/2}\bar{\boldsymbol{\xi}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\left[N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'(T^{-1}\boldsymbol{W}_i'\boldsymbol{W}_i)^{-1}T^{-1/2}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\right],\end{aligned}$$

where, as noted above, $\bar{\boldsymbol{\xi}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}=O_p\left(N^{-1/2}\right)$. Also, under stationarity (Assumption 3) and using (A.1) and (A.2) (See Lemma A.1), $\bar{\boldsymbol{q}}_{NT}=\bar{\boldsymbol{q}}_N+O_p\left(N^{-1/2}\right)$ and $\bar{\boldsymbol{Q}}_{NT}^{-1}=\bar{\boldsymbol{Q}}_N^{-1}+O_p\left(N^{-1/2}\right)$, and we have

$$N^{-1}\sum_{i=1}^{N}(\hat{\boldsymbol{\theta}}_i-\boldsymbol{\theta}_i)'\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_i)=T^{-1/2}\left(g_{1,nT}-g_{2,nT}\right)+O_p(T^{-1/2}N^{-1/2}),$$

where

$$
\begin{aligned}
g_{1,NT} &= \left[N^{-1}\sum_{i=1}^{N}\left(T^{-1/2}\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i\right)\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\right]\bar{\boldsymbol{Q}}_N^{-1}\bar{\boldsymbol{q}}_N,\\
g_{2,NT} &= N^{-1}\sum_{i=1}^{N}\left(T^{-1/2}\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i\right)\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i.
\end{aligned}
$$

We also note that under Assumptions 3, 4, 8 and 9

$$
g_{1,NT}=\mathrm{E}\left(g_{1,nT}\right)+O_p\left(N^{-1/2}\right),\text{ and }g_{2,NT}=\mathrm{E}\left(g_{2,NT}\right)+O_p\left(N^{-1/2}\right).
$$

Hence,

$$
N^{-1}\sum_{i=1}^{N}(\hat{\boldsymbol{\theta}}_i-\boldsymbol{\theta}_i)'\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_i)=T^{-1/2}\left[\mathrm{E}\left(g_{1,nT}\right)-\mathrm{E}\left(g_{2,NT}\right)\right]+O_p\left(N^{-1/2}\right)+O_p\left(T^{-1/2}N^{-1/2}\right).
$$

Substituting this result together with (A.17) and (A.18) in (A.16), we obtain

$$
N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}\tilde{e}_{i,T+1}=N^{-1}\sum_{i=1}^{N}\varepsilon_{i,T+1}^2+T^{-1/2}\psi_{NT}+O_p(N^{-1/2})+O_p\left(T^{-1/2}N^{-1/2}\right),\quad\text{(A.19)}
$$

where $\psi_{NT}=\mathrm{E}(g_{1,nT})-\mathrm{E}(g_{2,NT})$, or more specifically,

$$
\begin{aligned}
\psi_{NT} &= N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[T^{-1/2}\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\right]\bar{\boldsymbol{Q}}_N^{-1}\bar{\boldsymbol{q}}_N \qquad\text{(A.20)}\\
&\quad -N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[T^{-1/2}\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\right].
\end{aligned}
$$

Using (A.19), together with (17) and (21), in (39) now yields (40).

A.2.4 Proof of Proposition 4

Recall from (33) and (28) that $\hat{e}_{i,T+1}=\tilde{\varepsilon}_{i,T+1}-\tilde{\boldsymbol{x}}_{i,T+1}'(\hat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i)$ and $\hat{e}_{i,T+1}^{\mathrm{FE}}=\tilde{\varepsilon}_{i,T+1}-(\hat{\boldsymbol{\beta}}_{\mathrm{FE}}-\boldsymbol{\beta}_i)'\tilde{\boldsymbol{x}}_{i,T+1}$, where $\tilde{\varepsilon}_{i,T+1}=\varepsilon_{i,T+1}-\bar{\varepsilon}_{iT}$ and $\tilde{\boldsymbol{x}}_{i,T+1}=\boldsymbol{x}_{iT+1}-\bar{\boldsymbol{x}}_{iT}$,

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i &= (\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{\varepsilon}_i=\boldsymbol{Q}_{iT,\beta}^{-1}\boldsymbol{\xi}_{iT,\beta},\\
\hat{\boldsymbol{\beta}}_{\mathrm{FE}}-\boldsymbol{\beta}_i &= -\boldsymbol{\eta}_{i,\beta}+\bar{\boldsymbol{Q}}_{NT,\beta}^{-1}\bar{\boldsymbol{q}}_{NT,\beta}+\bar{\boldsymbol{Q}}_{NT,\beta}^{-1}\bar{\boldsymbol{\xi}}_{NT,\beta},
\end{aligned}
$$

$\boldsymbol{Q}_{iT,\beta} = (T^{-1}\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{X}_i)^{-1}$, $\boldsymbol{\xi}_{iT,\beta} = T^{-1/2}\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{\varepsilon}_i$, and $\bar{\boldsymbol{\xi}}_{NT,\beta} = N^{-1}\sum_{i=1}^{N}\boldsymbol{\xi}_{iT,\beta} = O_p(N^{-1/2})$.

Hence,

$$\begin{aligned} N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}^{\mathrm{FE}}\hat{e}_{i,T+1} &= N^{-1}\sum_{i=1}^{N}\tilde{\varepsilon}_{i,T+1}^2 \\ &\quad + N^{-1}\sum_{i=1}^{N}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)'\tilde{\boldsymbol{x}}_{i,T+1}\tilde{\boldsymbol{x}}_{i,T+1}'(\hat{\boldsymbol{\beta}}_{\mathrm{FE}} - \boldsymbol{\beta}_i) \\ &\quad - N^{-1}\sum_{i=1}^{N}(\varepsilon_{i,T+1} - \bar{\varepsilon}_{iT})\tilde{\boldsymbol{x}}_{i,T+1}'\left[\left(\hat{\boldsymbol{\beta}}_{\mathrm{FE}} - \boldsymbol{\beta}_i\right) + (\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)\right]. \end{aligned} \tag{A.21}$$

Using (29) and (34), we have

$$N^{-1}\sum_{i=1}^{N}\tilde{\varepsilon}_{i,T+1}\tilde{\boldsymbol{x}}_{i,T+1}'\left[\left(\hat{\boldsymbol{\beta}}_{\mathrm{FE}} - \boldsymbol{\beta}_i\right) + (\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)\right] = c_{NT}^{\mathrm{FE}} + c_{NT,\beta} + O_p(N^{-1/2}). \tag{A.22}$$

Also

$$\begin{aligned} & N^{-1}\sum_{i=1}^{N}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)'\left(\tilde{\boldsymbol{x}}_{i,T+1}\tilde{\boldsymbol{x}}_{i,T+1}'\right)(\hat{\boldsymbol{\beta}}_{\mathrm{FE}} - \boldsymbol{\beta}_i) \\ = \; & T^{-1/2}N^{-1}\sum_{i=1}^{N}\left(T^{-1/2}\boldsymbol{\varepsilon}_i'\boldsymbol{M}_T\boldsymbol{X}_i\right)\boldsymbol{Q}_{iT,\beta}^{-1}\left(\tilde{\boldsymbol{x}}_{i,T+1}\tilde{\boldsymbol{x}}_{i,T+1}'\right)\left(-\boldsymbol{\eta}_{i,\beta} + \bar{\boldsymbol{Q}}_{NT,\beta}^{-1}\bar{\boldsymbol{q}}_{NT,\beta} + \bar{\boldsymbol{Q}}_{NT,\beta}^{-1}\bar{\boldsymbol{\xi}}_{NT,\beta}\right). \end{aligned}$$

First, $\bar{\boldsymbol{\xi}}_{NT,\beta} = O_p(N^{-1/2})$ and $\bar{\boldsymbol{Q}}_{NT,\beta}^{-1} = \bar{\boldsymbol{Q}}_{N,\beta}^{-1} + O_p(N^{-1/2})$, where $\bar{\boldsymbol{Q}}_{N,\beta} = E\left(\bar{\boldsymbol{Q}}_{NT,\beta}\right)$, (see Lemma A.1). Hence, for a fixed $T > T_0$

$$\left[N^{-1}\sum_{i=1}^{N}\left(T^{-1/2}\boldsymbol{\varepsilon}_i'\boldsymbol{M}_T\boldsymbol{X}_i\right)\boldsymbol{Q}_{iT,\beta}^{-1}\left(\tilde{\boldsymbol{x}}_{i,T+1}\tilde{\boldsymbol{x}}_{i,T+1}'\right)\right]\bar{\boldsymbol{Q}}_{NT,\beta}^{-1}\bar{\boldsymbol{\xi}}_{NT,\beta} = O_p(N^{-1/2}).$$

Also, under Assumption 9,

$$\begin{aligned} & N^{-1}\sum_{i=1}^{N}\left(T^{-1/2}\boldsymbol{\varepsilon}_i'\boldsymbol{M}_T\boldsymbol{X}_i\right)\boldsymbol{Q}_{iT,\beta}^{-1}\left(\tilde{\boldsymbol{x}}_{i,T+1}\tilde{\boldsymbol{x}}_{i,T+1}'\right)\boldsymbol{\eta}_{i,\beta} \\ = \; & N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[\left(T^{-1/2}\boldsymbol{\varepsilon}_i'\boldsymbol{M}_T\boldsymbol{X}_i\right)\boldsymbol{Q}_{iT,\beta}^{-1}\left(\tilde{\boldsymbol{x}}_{i,T+1}\tilde{\boldsymbol{x}}_{i,T+1}'\right)\boldsymbol{\eta}_{i,\beta}\right] + O_p(N^{-1/2}), \end{aligned}$$

and

$$
\begin{aligned}
& N^{-1}\sum_{i=1}^{N}\left(T^{-1/2}\boldsymbol{\varepsilon}_i'\boldsymbol{M}_T\boldsymbol{X}_i\right)\boldsymbol{Q}_{iT,\beta}^{-1}\left(\tilde{\boldsymbol{x}}_{i,T+1}\tilde{\boldsymbol{x}}_{i,T+1}'\right) \\
= \;& N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[\left(T^{-1/2}\boldsymbol{\varepsilon}_i'\boldsymbol{M}_T\boldsymbol{X}_i\right)\boldsymbol{Q}_{iT,\beta}^{-1}\left(\tilde{\boldsymbol{x}}_{i,T+1}\tilde{\boldsymbol{x}}_{i,T+1}'\right)\right]+O_p(N^{-1/2}).
\end{aligned}
$$

Then

$$
N^{-1}\sum_{i=1}^{N}(\hat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i)'\tilde{\boldsymbol{x}}_{i,T+1}\tilde{\boldsymbol{x}}_{i,T+1}'(\hat{\boldsymbol{\beta}}_{\mathrm{FE}}-\boldsymbol{\beta}_i)=T^{-1/2}\psi_{NT}^{\mathrm{FE}}+O_p(T^{-1/2}N^{-1/2}), \tag{A.23}
$$

where

$$
\begin{aligned}
\psi_{NT}^{\mathrm{FE}} \;=\; & N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[\left(T^{-1/2}\boldsymbol{\varepsilon}_i'\boldsymbol{M}_T\boldsymbol{X}_i\right)\boldsymbol{Q}_{iT,\beta}^{-1}\left(\tilde{\boldsymbol{x}}_{i,T+1}\tilde{\boldsymbol{x}}_{i,T+1}'\right)\right]\bar{\boldsymbol{Q}}_{N,\beta}^{-1}\bar{\boldsymbol{q}}_{N,\beta} \\
& -N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[\left(T^{-1/2}\boldsymbol{\varepsilon}_i'\boldsymbol{M}_T\boldsymbol{X}_i\right)\boldsymbol{Q}_{iT,\beta}^{-1}\left(\tilde{\boldsymbol{x}}_{i,T+1}\tilde{\boldsymbol{x}}_{i,T+1}'\right)\boldsymbol{\eta}_{i,\beta}\right].
\end{aligned} \tag{A.24}
$$

Using (A.22) and (A.23) in (A.21) yields

$$
\begin{aligned}
N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}^{\mathrm{FE}}\hat{e}_{i,T+1} \;=\; & N^{-1}\sum_{i=1}^{N}\tilde{\varepsilon}_{i,T+1}^2 \\
& +T^{-1/2}\psi_{NT}^{\mathrm{FE}}-\left(c_{NT}^{\mathrm{FE}}+c_{NT,\beta}\right)+O_p(N^{-1/2})+O_p(T^{-1/2}N^{-1/2}).
\end{aligned}
$$

Substituting this result together with (36) and (31) in (43) now yields equation (44).

A.3 Panel AR(1): An example of correlated heterogeneity

Correlated heterogeneity can arise in many contexts. One important example is dynamic panel data models where, barring special cases, heterogeneity is correlated by design. As a simple example, consider the stationary panel AR(1) case where $y_{it}=\beta_i y_{i,t-1}+\varepsilon_{it}$, for $t=\ldots-2-1,0,1,\ldots,T,T+1,\ldots$, and $\sup_i|\beta_i|\leq c$ for some positive $c<1$, and β_i follows a random coefficient model $\beta_i=\beta_0+\eta_i$, where $\beta_0=E(\beta_i)$, and η_i is suitably truncated such that the stationary condition $\sup_i|\beta_i|\leq c$ is met.

Suppose our objective is to forecast y_{iT+1} based on the observations $\{y_{it},t=0,1,2,\ldots,T\}$.[21] In

[21] The assumption that the process for y_{it} has started a long time prior to date 0, is equivalent to assuming that y_{i0}

the context of the general linear model analyzed in the paper, $\boldsymbol{w}_{it} = y_{i,t-1}$ and $\boldsymbol{\theta}_i = \beta_i$. It is easily verified that our Assumptions 1-9 cover the dynamic case where one or more elements of $\boldsymbol{w}_{it}$ are lagged values of y_{it}. Forecasts based on pooled estimates, which incorrectly assume $\beta_i = \beta_0$ generate a heterogeneity bias, Δ_N, given by (24). In the present example $q_i = \mathrm{E}\left(y_{i,t-1}^2 \eta_i\right)$, $Q_i = \mathrm{E}\left(y_{i,t-1}^2\right)$, and

$$\Delta_N = N^{-1} \sum_{i=1}^{N} \mathrm{E}\left(y_{it}^2 \eta_i^2\right) - \frac{\left[N^{-1} \sum_{i=1}^{N} E\left(y_{i,t-1}^2 \eta_i\right)\right]^2}{N^{-1} \sum_{i=1}^{N} E\left(y_{i,t-1}^2\right)},$$

where q_i measures the degree of correlated heterogeneity. To derive q_i for the AR model, note that

$$y_{it} = \sum_{s=0}^{\infty} \beta_i^s \varepsilon_{i,t-s} = \sum_{s=0}^{\infty} \left(\beta_0 + \eta_i\right)^s \varepsilon_{i,t-s}, \tag{A.25}$$

so y_{it} is a non-linear function of η_i, and, in general, $q_i = \mathrm{E}\left(y_{i,t-1}^2 \eta_i\right) \neq 0$. This shows that heterogeneity in panel AR models generates correlated heterogeneity as is also implicit in the analysis of Pesaran and Smith (1995). Using (A.25) we have

$$\mathrm{E}(y_{it}) = 0, \; Q_i = \mathrm{E}(y_{it}^2) = \mathrm{E}\left(y_{i,t-1}^2\right) = \mathrm{E}\left(\frac{\sigma_i^2}{1-\beta_i^2}\right), \text{ for all } t,$$

$$q_i = \mathrm{E}\left(y_{i,t-1}^2 \eta_i\right) = \sum_{s=0}^{\infty} \mathrm{E}\left(\beta_i^{2s} \eta_i \sigma_i^2\right) = \mathrm{E}\left(\frac{\eta_i \sigma_i^2}{1-\beta_i^2}\right), \text{ and } \mathrm{E}\left(y_{it}^2 \eta_i^2\right) = \mathrm{E}\left(\frac{\eta_i^2 \sigma_i^2}{1-\beta_i^2}\right).$$

In this simple example, heterogeneity is uncorrelated only if $\beta_0 = 0$ and η_i is symmetrically distributed around 0. This follows since when $\beta_0 = 0$ we have $q_i = \mathrm{E}\left(\frac{\eta_i \sigma_i^2}{1-\eta_i^2}\right)$ and under symmetry $\eta_i \sigma_i^2 / \left(1 - \eta_i^2\right)$ is an odd function of η_i, which yields $q_i = 0$. But when $\beta_0 \neq 0$, then $q_i \neq 0$ even if η_i has a symmetric distribution. The expression for Δ_N is strictly positive irrespective of whether $q_i = 0$ or not. Under stationarity, Δ_N simplifies to

$$\begin{aligned} \Delta_{\mathrm{AR}} &= \mathrm{E}\left(y_{it}^2 \eta_i^2\right) - \frac{\left[\mathrm{E}\left(y_{i,t-1}^2 \eta_i\right)\right]^2}{\mathrm{E}\left(y_{i,t-1}^2\right)} \\ &= \frac{\mathrm{E}\left(\frac{\eta_i^2 \sigma_i^2}{1-\beta_i^2}\right) \mathrm{E}\left(\frac{\sigma_i^2}{1-\beta_i^2}\right) - \left[\mathrm{E}\left(\frac{\eta_i \sigma_i^2}{1-\beta_i^2}\right)\right]^2}{\mathrm{E}\left(\frac{\sigma_i^2}{1-\beta_i^2}\right)}. \end{aligned} \tag{A.26}$$

is drawn from a distribution with zero mean and variance $\sigma_i^2/(1-\beta_i^2)$.

Let $f_i = \sigma_i \eta_i / \sqrt{1-\beta_i^2}$ and $g_i = \sigma_i / \sqrt{1-\beta_i^2}$, and note that the numerator of Δ_{AR} can be written as $\text{E}(f_i^2)\text{E}(g_i^2) - [E(f_i g_i)]^2 \geq 0$, which establishes that $\Delta_{\text{AR}} \geq 0$, in line with part (c) of Proposition 2.

The magnitude of Δ_{AR} depends on the joint distribution of β_i and σ_i^2. As an example, consider the case where σ_i^2 and β_i are independently distributed, $\text{E}(\sigma_i^2) = \sigma^2$ and $\eta_i \sim \text{Uniform}(-a/2, a/2)$, for $a > 0$.[22] Then,

$$q_i = \sigma^2 \text{E}\left(\frac{\eta_i}{1-\beta_i^2}\right) = \frac{\sigma^2}{2}\left[E\left(\frac{\eta_i}{1-\beta_0-\eta_i}\right) + E\left(\frac{\eta_i}{1+\beta_0+\eta_i}\right)\right]$$

To derive the expectation in this above expression note that for a given B, such that $B^2 - a^2/4 > 0$, we have

$$\text{E}\left(\frac{\eta_i}{B+\eta_i}\right) = \frac{1}{a}\int_{-a/2}^{a/2}\left(\frac{\eta}{B+\eta}\right)d\eta = 1 - \left(\frac{B}{a}\right)\ln\left(\frac{B+a/2}{B-a/2}\right). \tag{A.27}$$

Using this result, and setting $B = 1+\beta_0$, we have, for $(1+\beta_0)^2 > a^2/4$,

$$\text{E}\left(\frac{\eta_i}{1+\beta_0+\eta_i}\right) = 1 - \left(\frac{1+\beta_0}{a}\right)\ln\left(\frac{1+\beta_0+a/2}{1+\beta_0-a/2}\right).$$

Similarly, again for $(\beta_0 - 1)^2 > a^2/4$,

$$\text{E}\left(\frac{\eta_i}{1-\beta_0-\eta_i}\right) = -\text{E}\left(\frac{\eta_i}{\beta-1+\eta_i}\right) = -\left[1 - \left(\frac{\beta_0-1}{a}\right)\ln\left(\frac{\beta_0-1+a/2}{\beta_0-1-a/2}\right)\right].$$

Overall, assuming that $a/2 < 1 - |\beta_0|$, we have

$$\begin{aligned}\text{E}\left(y_{i,t-1}^2\eta_i\right) &= \frac{\sigma^2}{2}\left[\left(\frac{\beta_0-1}{a}\right)\ln\left(\frac{\beta_0-1+a/2}{\beta_0-1-a/2}\right) - \left(\frac{1+\beta_0}{a}\right)\ln\left(\frac{1+\beta_0+a/2}{1+\beta_0-a/2}\right)\right] \\ &= \frac{\sigma^2}{2a}\left[-(1-\beta_0)\ln\left(\frac{1-\beta_0-a/2}{1-\beta_0+a/2}\right) - (1+\beta_0)\ln\left(\frac{1+\beta_0+a/2}{1+\beta_0-a/2}\right)\right]. \end{aligned} \tag{A.28}$$

To ensure that $|\beta_i| = |\beta_0 + \eta_i| < 1$, we require that a is sufficiently small relative to β_0 and $|\beta_0| < 1$. A sufficient condition for this to hold is that

$$|\beta_i| = |\beta_0 + \eta_i| \leq |\beta_0| + |\eta_i| = |\beta_0| + a/2 < 1.$$

[22] Note in this case η_i is symmetrically distributed around 0.

Table A.2: Numerical values for $\mathrm{E}\left(y_{i,t-1}^2\eta_i\right)$ and Δ_{AR} for the panel AR(1) model

β_0	$\mathrm{E}(y_{i,t-1}^2\eta_i)$	Δ_{AR}
0.3	0.100	0.116
0.45	0.316	0.163
0.49	0.657	0.211
0.4999	1.830	0.322

Note: The numerical values are based on $a = \sigma^2 = 1$.

We can now calculate $\mathrm{E}\left(y_{i,t-1}^2\eta_i\right)$ for a range of values for $\beta_0 < 1/2$. Using $a = 1$ and $\sigma^2 = 1$, we obtain the values given in Table A.2.

In general for $a > 0$, $\mathrm{E}\left(y_{i,t-1}^2\eta_i\right) \neq 0$, and for any $|\beta_0| < 1$, it follows that $\mathrm{E}\left(y_{i,t-1}^2\eta_i\right) \to 0$ as $a \to 0$. Since $\eta_i \sim iid\mathrm{Uniform}(-a/2, a/2)$ is symmetrically distributed, then $\mathrm{E}\left(y_{i,t-1}^2\eta_i\right) = 0$ for $\beta_0 = 0$. But $\mathrm{Cov}(y_{i,t-1}^2, \eta_i^2) \neq 0$, even under symmetry and $y_{i,t-1}^2$ and η_i are not independently distributed. For example, when $\beta_0 = 0$, we have

$$\mathrm{E}(y_{it}^2\eta_i^2) = \sigma^2\mathrm{E}\left(\frac{\eta_i^2}{1-\eta_i^2}\right) \neq \mathrm{E}(y_{it}^2)\mathrm{E}(\eta_i^2) = \sigma^2\mathrm{E}\left(\frac{1}{1-\eta_i^2}\right)\mathrm{E}(\eta_i^2).$$

When β_i and σ_i^2 are independently distributed, using (A.26), we have

$$\sigma^{-2}\Delta_{\mathrm{AR}} = \frac{\mathrm{E}\left(\frac{\eta_i^2}{1-\beta_i^2}\right)\mathrm{E}\left(\frac{1}{1-\beta_i^2}\right) - \left[\mathrm{E}\left(\frac{\eta_i}{1-\beta_i^2}\right)\right]^2}{\mathrm{E}\left(\frac{1}{1-\beta_i^2}\right)}.$$

We can derive an analytical expression for $\mathrm{E}\left(\frac{1}{1-\beta_i^2}\right)$, noting that

$$\mathrm{E}\left(\frac{1}{B+\eta_i}\right) = \frac{1}{a}\int_{-a/2}^{a/2}\left(\frac{1}{B+\eta}\right)d\eta = \frac{1}{a}\ln\left(\frac{B+a/2}{B-a/2}\right).$$

Hence,

$$\begin{aligned}\mathrm{E}\left(\frac{1}{1-\beta_i^2}\right) &= \frac{1}{2}\left[-\mathrm{E}\left(\frac{1}{-1+\beta_0+\eta_i}\right) + \mathrm{E}\left(\frac{1}{1+\beta_0+\eta_i}\right)\right] \\ &= -\frac{1}{2a}\ln\left(\frac{\beta_0-1+a/2}{\beta_0-1-a/2}\right) + \frac{1}{2a}\ln\left(\frac{\beta_0+1+a/2}{\beta_0+1-a/2}\right),\end{aligned}$$

or

$$\mathrm{E}\left(\frac{1}{1-\beta_i^2}\right) = \frac{1}{2a}\left[\ln\left(\frac{1+\beta_0+a/2}{1+\beta_0-a/2}\right) - \ln\left(\frac{1-\beta_0-a/2}{1-\beta_0+a/2}\right)\right]. \tag{A.29}$$

Using (A.29) and simulated values of $\mathrm{E}\left(\frac{\eta_i}{1-\beta_i^2}\right)$ and $\mathrm{E}\left(\frac{\eta_i^2}{1-\beta_i^2}\right)$, we obtain the values of Δ_{AR} for $\alpha = 1$ and $\sigma^2 = 1$ that are reported in Table A.2.

Online Supplement

Forecasting with panel data: Estimation uncertainty versus parameter heterogeneity

by

M. Hashem Pesaran[23]

Andreas Pick[24]

and

Allan Timmermann[25]

---

[23]University of Southern California and Trinity College, Cambridge. Email: pesaran@usc.edu

[24]Erasmus University Rotterdam, Erasmus School of Economics, Burgemeester Oudlaan 50, 3000DR Rotterdam, and Tinbergen Institute. Email: andreas.pick@cantab.net

[25]UC San Diego, Rady School of Management, 9500 Gilman Drive, La Jolla CA 92093-0553. Email: atimmermann@ucsd.edu.

S.1 Introduction

This supplementary appendix provides additional material underpinning the analysis in the main paper along with a set of extensions to the Monte Carlo simulations and empirical results. We begin by deriving in Section S.2 the pooled R-squared, PR_N^2, used in the Monte Carlo simulations to target the predictive power of our panel forecasting models. We characterize PR_N^2 as a function of the underlying parameters of the DGPs and use this to calibrate the parameters used in the simulations. Next, Section S.3 provides details of how we implement the estimators used in our analysis. Section S.4 provides additional simulation and empirical results from equal-weighted forecast combination schemes, while Section S.5 examines the robustness of our Monte Carlo simulation results with regards to how we set the initial condition of the autoregressive processes.

S.2 Derivation of the pooled R-squared PR_N^2

Consider the panel data model

$$y_{it} = \alpha_i + \beta_i y_{i,t-1} + \gamma_i x_{it} + \varepsilon_{it}, \tag{S.1}$$

$$x_{it} = \mu_{xi} + \xi_{it}, \quad \xi_{it} = \rho_{xi}\xi_{i,t-1} + \sigma_{xi}\sqrt{1-\rho_{xi}^2}\nu_{it}.$$

Further, $\text{Var}(\varepsilon_{it}) = 1$, and $\text{Var}(\nu_{it}) = 1$ as set out in further detail in Section 6. To simplify the derivations, we treat x_{it} as strictly exogenous (no feedback from $y_{i,t-1}$) and assume that y_{it} is stationary and started a long time in the past. To deal with the heterogeneity across the different equations in the panel, we use the following average measure of fit, for a given N,

$$PR_N^2 = 1 - \frac{N^{-1}\sum_{i=1}^N \text{Var}\left(\varepsilon_{it}\,|\boldsymbol{\theta}_i,\, x_{it}\,\right)}{N^{-1}\sum_{i=1}^N \text{Var}(y_{it}\,|\boldsymbol{\theta}_i,\, x_{it},\,)}, \tag{S.2}$$

where as before $\boldsymbol{\theta}_i = (\alpha_i, \beta_i, \gamma_i)'$. For the numerator we have

$$\text{Var}\left(\varepsilon_{it} \,\middle|\, \boldsymbol{\theta}_i,\, \sigma_i^2,\, x_{it}\;\right) = \sigma_i^2. \tag{S.3}$$

To derive $\text{Var}(y_{it}\,|\boldsymbol{\theta}_i,\, x_{it}\,)$, we note that

$$\begin{aligned}
\text{Var}(y_{it}\,\big|\boldsymbol{\theta}_i, \sigma_i^2, x_{it}\,) &= \text{E}\left[\text{Var}(y_{it}\,\big|\boldsymbol{\theta}_i, \sigma_i^2, y_{i,t-1}, x_{it}\,)\right] + \text{Var}\left[\text{E}(y_{it}\,\big|\boldsymbol{\theta}_i, \sigma_i^2, y_{i,t-1}, x_{it}\,)\right], \\
\text{E}(y_{it}\,\big|\boldsymbol{\theta}_i, \sigma_i^2, y_{i,t-1}, x_{it}\,) &= \alpha_i + \beta_i y_{i,t-1} + \gamma_i x_{it},\ \text{Var}(y_{it}\,\big|\boldsymbol{\theta}_i, \sigma_i^2, y_{i,t-1}, x_{it}\,) = \sigma_i^2, \\
\text{Var}\left[\text{E}(y_{it}\,\big|\boldsymbol{\theta}_i, \sigma_i^2, y_{i,t-1}, x_{it}\,)\right] &= \beta_i^2 \text{Var}(y_{it}\,\big|\boldsymbol{\theta}_i, \sigma_i^2, x_{it}\,) + \gamma_i^2 \text{Var}\,(x_{it})\,.
\end{aligned}$$

Hence,

$$\text{Var}(y_{it}\,\big|\boldsymbol{\theta}_i, \sigma_i^2, x_{it}\,) = \frac{\gamma_i^2 \text{Var}(\xi_{it}) + \sigma_i^2}{1-\beta_i^2}. \tag{S.4}$$

Now using (S.3) and (S.4) in (S.2), we obtain

$$PR_N^2 = 1 - \left(\frac{N^{-1}\sum_{i=1}^N \sigma_i^2}{N^{-1}\sum_{i=1}^N \frac{\gamma_i^2\sigma_{xi}^2+\sigma_i^2}{1-\beta_i^2}} \right),$$

where $\sigma_{xi}^2 = \text{Var}(\xi_{it})$. After some simplifications we have

$$PR_N^2 = \frac{b_N + (c_N - a_N)}{b_N + c_N}, \tag{S.5}$$

where $a_N = N^{-1}\sum_{i=1}^N \sigma_i^2$, $\quad b_N = N^{-1}\sum_{i=1}^N \frac{\gamma_i^2\sigma_{xi}^2}{1-\beta_i^2}$, and $c_N = N^{-1}\sum_{i=1}^N \frac{\sigma_i^2}{1-\beta_i^2}$.

When these parameters are distributed independently, as $N \to \infty$, we obtain

$$\begin{aligned}
a_N &\xrightarrow{p} \text{E}(\sigma_i^2),\ b_N \xrightarrow{p} \text{E}(\gamma_i^2)\text{E}(\sigma_{xi}^2)\text{E}\left(\frac{1}{1-\beta_i^2}\right), \\
c_N &\xrightarrow{p} \text{E}(\sigma_i^2)\text{E}\left(\frac{1}{1-\beta_i^2}\right).
\end{aligned}$$

Hence, using (S.5), we note that (as $N \to \infty$)

$$PR_N^2 \to PR^2 = \frac{\text{E}(\gamma_i^2)\text{E}(\sigma_{xi}^2)\text{E}\left(\frac{1}{1-\beta_i^2}\right) + \left[\text{E}(\sigma_i^2)\text{E}\left(\frac{1}{1-\beta_i^2}\right) - \text{E}(\sigma_i^2)\right]}{\text{E}(\gamma_i^2)\text{E}(\sigma_{xi}^2)\text{E}\left(\frac{1}{1-\beta_i^2}\right) + \text{E}(\sigma_i^2)\text{E}\left(\frac{1}{1-\beta_i^2}\right)}.$$

Under our design $\mathrm{E}(\sigma_i^2) = 1$, $\mathrm{E}(\sigma_{xi}^2) = 1$, and the above expression simplifies to

$$PR^2 = \frac{\mathrm{E}(\gamma_i^2)\mathrm{E}\left(\frac{1}{1-\beta_i^2}\right) + \left[\mathrm{E}\left(\frac{1}{1-\beta_i^2}\right) - 1\right]}{\mathrm{E}(\gamma_i^2)\mathrm{E}\left(\frac{1}{1-\beta_i^2}\right) + \mathrm{E}\left(\frac{1}{1-\beta_i^2}\right)}. \tag{S.6}$$

For the pure panel AR model (55), where $\gamma_i = 0,\ \forall i$, this reduces to

$$PR^2_{\mathrm{AR}} = 1 - \frac{1}{\mathrm{E}\left(\frac{1}{1-\beta_i^2}\right)}. \tag{S.7}$$

When β_i is homogeneous such that $\beta_i = \beta_0$, $\mathrm{E}\left(\frac{1}{1-\beta_i^2}\right) = 1/(1-\beta_0^2)$, this simplifies to the familiar condition $PR^2_{\mathrm{AR}} = \beta_0^2$.

Additionally, we can now write

$$PR^2 = PR^2_{\mathrm{AR}} + \frac{\mathrm{E}(\gamma_i^2)(1 - PR^2_{\mathrm{AR}})}{1 + \mathrm{E}(\gamma_i^2)}. \tag{S.8}$$

In the general case where σ_i^2 is not distributed independently of β_i and N is finite we have

$$PR^2_N > 1 - a_N/c_N = 1 - \frac{N^{-1}\sum_{i=1}^N \sigma_i^2}{N^{-1}\sum_{i=1}^N \frac{\sigma_i^2}{1-\beta_i^2}}.$$

In the case where $\beta_i = \beta_0 + \eta_{i\beta}$, and $\eta_{i\beta} \sim$iid$Uniform(-\alpha_\beta/2, \alpha_\beta/2)$, $\alpha_\beta > 0$, we have (see also (A.29) in the Appendix to the paper):

$$\begin{aligned}
\mathrm{E}\left(\tfrac{1}{1-\beta_i^2}\right) &= \tfrac{1}{a_\beta}\int_{-a_\beta/2}^{a_\beta/2} \tfrac{1}{1-\left(\beta_0+\eta_\beta\right)^2} d\eta_\beta \\
&= \tfrac{1}{2a_\beta}\int_{-a_\beta/2}^{a_\beta/2} \left[\tfrac{1}{1+\beta_0+\eta_\beta} + \tfrac{1}{1-\beta_0-\eta_\beta}\right] d\eta_\beta \\
&= \tfrac{1}{2a_\beta}\left[ln(1+\beta_0+\eta_\beta) - ln(1-\beta_0-\eta_\beta)\right]_{-a_\beta/2}^{a_\beta/2} \\
&= \tfrac{1}{2a_\beta}\left[ln\left(\tfrac{1+\beta_0+a_\beta/2}{1+\beta_0-a_\beta/2}\right) - ln\left(\tfrac{1-\beta_0-a_\beta/2}{1-\beta_0+a_\beta/2}\right)\right],
\end{aligned} \tag{S.9}$$

assuming that

$$(1+\beta_0+a_\beta/2)(1+\beta_0-a_\beta/2) > 0 \text{ and } (1-\beta_0-a_\beta/2)(1-\beta_0+a_\beta/2) > 0,$$

or if

$$0 \leq a_\beta < 2\left(1 - |\beta_0|\right). \tag{S.10}$$

It is easily established that $\mathrm{E}\left(\frac{1}{1-\beta_i^2}\right) \to \frac{1}{1-\beta_0^2}$, as $a_\beta \to 0$.

Now using (S.9) in (S.7) we have

$$PR^2_{\mathrm{AR}}(a_\beta, \beta_0) = 1 - \frac{2a_\beta}{\left[\ln\left(\frac{1+\beta_0+a_\beta/2}{1+\beta_0-a_\beta/2}\right) - \ln\left(\frac{1-\beta_0-a_\beta/2}{1-\beta_0+a_\beta/2}\right)\right]}. \tag{S.11}$$

Numerical values for PR^2_{AR} in (S.11) for different values of a_β and β_0 are given in Table S.2. Note that under heterogeneity ($a_\beta > 0$), PR^2_{AR} exceeds its homogeneous counterpart.

Table S.2: Values of $PR^2_{\mathrm{AR}}(a_\beta, \beta_0)$

β_0	$PR^2_{\mathrm{AR}}(a_\beta = 0, \beta_0)$	$PR^2_{\mathrm{AR}}(a_\beta = 0.5, \beta_0)$	$PR^2_{\mathrm{AR}}(a_\beta = 1, \beta_0)$
0	0	0.021	0.090
0.1	0.01	0.021	0.105
0.2	0.04	0.065	0.150
0.3	0.09	0.202	0.232
0.4	0.16	0.199	0.364
0.49	0.25	0.292	0.624

Note: The table reports numerical values for PR^2_{AR} in (S.11) for different values of a_β and β_0.

Our Monte Carlo simulations target two values of PR^2, namely 0.2 and 0.6, for the panel AR model. We do so by calibrating the values of the a_β and β_0 parameters. The values of the parameters used to this end are reported in Table S.3. For the same parameters, the PR^2 for the panel ARX model are similar, albeit somewhat higher.

S.3 Details of the estimators

This section provides details on the implementation of the estimators and forecasts used in the Monte Carlo experiments and empirical applications. Recall that the DGP, (56), in the Monte Carlo experiments is

$$y_{it} = \alpha_i + \beta_i y_{i,t-1} + \gamma_i x_{it} + \varepsilon_{it} = \alpha_i + \boldsymbol{\beta}_i' \mathbf{x}_{it} + \varepsilon_{it} = \boldsymbol{\theta}_i' \boldsymbol{w}_{it} + \varepsilon_{it}, \quad \varepsilon_{it} \sim (0, \sigma_i^2), \tag{S.12}$$

Table S.3: PR^2 for parameters of Monte Carlo models

a_β	β_0	PR^2_{AR}	$PR^2_{ARX}(\rho_{\gamma x}=0)$	$PR^2_{ARX}(\rho_{\gamma x}=0.5)$
0	0.447	0.200	0.209	0.209
0.5	0.401	0.200	0.280	0.301
1	0.267	0.200	0.340	0.374
0	0.775	0.601	0.605	0.605
0.5	0.688	0.600	0.640	0.651
1	0.486	0.599	0.669	0.686

Note: The table reports the parameters for a_β and β_0 in the first two columns and the implied values for PR^2 in the remaining columns. The values of PR^2 are obtained by simulation using the DGP in Section 6 with 10.000 replications.

for $t = 1, 2, \ldots, T$ and $i = 1, 2, \ldots, N$, where $\boldsymbol{\beta}_i = (\beta_i, \gamma_i)'$, $\boldsymbol{\theta}_i = (\alpha_i, \boldsymbol{\beta}_i')'$, $\mathbf{x}_{it} = (y_{i,t-1}, x_{it})'$, and $\boldsymbol{w}_{it} = (1, \mathbf{x}_{it}')'$. Here we consider a more general case where the dimension of $\boldsymbol{x}_{it}$ is $k \times 1$ and that of $\boldsymbol{w}_{it}$ is $K \times 1$, where $K = k + 1$. In principle, $\mathbf{x}_{it}$ could include highered lags of y_{it} and x_{it}, and other covariates. As in the main analysis, for simplicity we do not explicitly refer to the forecast horizon, h, but it is assumed that $\boldsymbol{x}_{it}$ contains information known at time $t - h$. Below we assume a forecast horizon of $h = 1$.

Individual forecasts The individual-specific forecasts based on the data of a given cross-sectional unit are

$$\hat{y}_{i,T+1} = \hat{\alpha}_{i,T} + \hat{\boldsymbol{\beta}}_{i,T}' \boldsymbol{x}_{i,T+1} = \hat{\boldsymbol{\theta}}_{i,T}' \boldsymbol{w}_{i,T+1} \tag{S.13}$$

The parameters are estimated using the estimation sample containing T observations: $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$ and $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{iT})'$. In matrix notation, the model is

$$\boldsymbol{y}_i = \alpha_i \boldsymbol{\tau}_T + \boldsymbol{X}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i = \boldsymbol{W}_i \boldsymbol{\theta}_i + \boldsymbol{\varepsilon}_i,$$

where $\boldsymbol{\tau}$ is a $T \times 1$ unit vector, $\boldsymbol{W}_i = (\boldsymbol{w}_{i1}, \boldsymbol{w}_{i2}, \ldots, \boldsymbol{w}_{iT})'$, $\boldsymbol{w}_{it} = (1, \mathbf{x}_{it}')'$, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})'$. The parameters are estimated as

$$\hat{\boldsymbol{\beta}}_{i,T} = \left(\boldsymbol{X}_i' \boldsymbol{M}_T \boldsymbol{X}_i\right)^{-1} \boldsymbol{X}_i \boldsymbol{M}_T \boldsymbol{y}_i,$$

$$\hat{\alpha}_{i,T} = \left(\boldsymbol{\tau}_T' \boldsymbol{M}_{ix} \boldsymbol{\tau}_T\right)^{-1} \boldsymbol{\tau}_T' \boldsymbol{M}_{ix} \boldsymbol{y}_i,$$

$$\boldsymbol{M}_T = \boldsymbol{I}_T - \boldsymbol{\tau}_T \left(\boldsymbol{\tau}_T' \boldsymbol{\tau}_T\right)^{-1} \boldsymbol{\tau}_T', \; \boldsymbol{M}_{ix} = \boldsymbol{I}_T - \boldsymbol{X}_i \left(\boldsymbol{X}_i' \boldsymbol{X}_i\right)^{-1} \boldsymbol{X}_i'.$$

Written in more compact form, we have

$$\hat{\boldsymbol{\theta}}_{i,T} = \left(\boldsymbol{W}_i' \boldsymbol{W}_i\right)^{-1} \boldsymbol{W}_i' \boldsymbol{y}_i. \tag{S.14}$$

The "individual" forecasts in (S.13), for $i = 1, 2, \ldots, N$, will be used as the reference forecast and the MSFE of all other methods are reported as ratios relative to the MSFE of this forecast, defined by

$$\text{MSFE}_{ref} = N^{-1} \sum_{i=1}^{N} \left(y_{i,T+1} - \hat{\boldsymbol{\theta}}_{i,T}' \boldsymbol{w}_{i,T+1}\right)^2. \tag{S.15}$$

Pooled forecasts The forecasts that use the pooled information of all units in the panel are

$$\tilde{y}_{i,T+1} = \tilde{\boldsymbol{\theta}}_{\text{pool}}' \boldsymbol{w}_{i,T+1}, \tag{S.16}$$

where

$$\tilde{\boldsymbol{\theta}}_{\text{pool}} = \left(\boldsymbol{W}' \boldsymbol{W}\right)^{-1} \boldsymbol{W} \boldsymbol{y} = \left(\sum_{i=1}^{N} \boldsymbol{W}_i' \boldsymbol{W}_i\right)^{-1} \sum_{i=1}^{N} \boldsymbol{W}_i' \boldsymbol{y}_i, \tag{S.17}$$

and $\boldsymbol{W} = (\boldsymbol{W}_1', \boldsymbol{W}_2', \ldots, \boldsymbol{W}_N')'$ and $\boldsymbol{y} = (\boldsymbol{y}_1', \boldsymbol{y}_2', \ldots, \boldsymbol{y}_N')'$.

Fixed effects forecast The FE forecasts are given by

$$\hat{y}_{i,T+1}^{FE} = \hat{\alpha}_{i,\text{FE}} + \hat{\boldsymbol{\beta}}_{\text{FE}}' \boldsymbol{x}_{i,T+1}, \tag{S.18}$$

where

$$\hat{\boldsymbol{\beta}}_{\text{FE}} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_T \boldsymbol{X}_i\right)^{-1} \sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_T \boldsymbol{y}_i,$$

and

$$\hat{\alpha}_{i,\mathrm{FE}} = \boldsymbol{\tau}_T'(\boldsymbol{y}_i - \hat{\boldsymbol{\beta}}_{\mathrm{FE}}'\boldsymbol{X}_i)/T$$

Goldberger's random effects BLUP This forecast uses the best linear unbiased predictor (BLUP) of Goldberger (1962). For this forecast, the model is assumed to be as follows:

$$y_{i,t+1} = \alpha + \boldsymbol{\beta}'\boldsymbol{x}_{i,t+1} + \varepsilon_{i,t+1},$$

where $\varepsilon_{i,t+1} = \eta_i + u_{i,t+1}$. The BLUP forecasts are given as

$$\hat{y}_{i,T+1}^{RE} = \hat{\alpha}_{\mathrm{RE}} + \hat{\boldsymbol{\beta}}_{\mathrm{RE}}'\boldsymbol{x}_{i,T+1} + \frac{T\hat{\sigma}_\eta^2}{T\hat{\sigma}_\eta^2 + \hat{\sigma}_u^2}\bar{\hat{\varepsilon}}_i, \tag{S.19}$$

where $\bar{\hat{\varepsilon}}_i = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{it}$ and $\hat{\varepsilon}_{it} = y_{it} - \hat{\alpha}_{\mathrm{RE}} - \boldsymbol{x}_{it}'\hat{\boldsymbol{\beta}}_{\mathrm{RE}}$. $\hat{\alpha}_{\mathrm{RE}}$, and $\hat{\boldsymbol{\beta}}_{\mathrm{RE}}$ are estimated by GLS using

$$\hat{\boldsymbol{\Sigma}}^{-1} = \hat{\sigma}_u^{-2}\left(\boldsymbol{M}_T + \hat{\rho}\boldsymbol{P}_T\right)$$

where $\boldsymbol{P}_T = \boldsymbol{I}_T - \boldsymbol{M}_T$, $\quad \hat{\rho} = \hat{\sigma}_u^2/(T\hat{\sigma}_\eta^2 + \hat{\sigma}_u^2)$,

$$\hat{\sigma}_u^2 = \frac{1}{N(T-1)-K}\sum_{i=1}^{N}(\boldsymbol{y}_i - \hat{\alpha}_{i,\mathrm{FE}} - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_{\mathrm{FE}})'\boldsymbol{M}_T(\boldsymbol{y}_i - \hat{\alpha}_{i,\mathrm{FE}} - \boldsymbol{X}_i\hat{\boldsymbol{\beta}}_{\mathrm{FE}})$$

$$\hat{\sigma}_\eta^2 = \frac{1}{N-K}\sum_{i=1}^{N}(\bar{y}_i - \hat{\boldsymbol{\beta}}_{\mathrm{FE}}'\bar{\boldsymbol{x}}_i)^2 - \hat{\sigma}_u^2/T,$$

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\mathrm{RE}} \;=\; & \left[\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{X}_i + \frac{\hat{\rho}}{N}\sum_{i=1}^{N}(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})'\right]^{-1} \times \\ & \left[\frac{1}{NT}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{y}_i + \frac{\hat{\rho}}{N}\sum_{i=1}^{N}(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})(\bar{y}_i - \bar{y})'\right],\end{aligned}$$

and

$$\hat{\alpha}_{\text{RE}} = \bar{y} - \hat{\boldsymbol{\beta}}'_{\text{RE}}\bar{\boldsymbol{x}},$$

where

$$\bar{\boldsymbol{x}}_i = T^{-1}\sum_{t=1}^{T}\boldsymbol{x}_{i,t}, \quad \bar{\boldsymbol{x}} = N^{-1}\sum_{i=1}^{N}\bar{\boldsymbol{x}}_i, \quad \bar{y}_i = T^{-1}\sum_{t=1}^{T}y_{it}, \quad \bar{y} = N^{-1}\sum_{i=1}^{N}\bar{y}_i.$$

See Baltagi (2013, pp. 999–1001) and Pesaran (2015, pp. 646–649) for further details.

Combination of individual and pooled forecasts

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

where $\hat{y}_{i,T+1}$ and $\tilde{y}_{i,T+1}$ are the individual and pooled forecasts in (S.13) and (S.16) with weights

$$\hat{\omega}^{*}_{NT} = \frac{\hat{\Delta}_{NT}}{\hat{\Delta}_{NT} + T^{-1}\hat{h}_{NT}},$$

where $\hat{\Delta}_{NT}$ and $\hat{h}_{NT}$ are given by (47) and (46).

Combination of individual and FE forecasts

$$y^{*}_{i,T+1}(\hat{\omega}^{*}_{FE,NT}) = \hat{\omega}^{*}_{FE,NT}\hat{y}_{i,T+1} + (1 - \hat{\omega}^{*}_{FE,NT})\hat{y}_{i,T+1,\text{FE}},$$

where $\hat{y}_{i,T+1}$ and $\hat{y}_{i,T+1,\text{FE}}$ are the individual and FE forecasts in (S.13) and (S.18) with the weight

$$\hat{\omega}^{*}_{FE,NT} = \frac{\hat{\Delta}^{\text{FE}}_{NT}}{\hat{\Delta}^{\text{FE}}_{NT} + T^{-1}\hat{h}_{NT,\beta}}.$$

$\hat{\Delta}^{\text{FE}}_{NT}$ are $\hat{h}_{NT,\beta}$ are given by (50) and (51).

Empirical Bayes forecast The empirical Bayes forecast using the estimator of Hsiao et al. (1999)

is

$$\hat{y}_{i,T+1}^{EB} = \hat{\boldsymbol{\theta}}'_{i,EB}\boldsymbol{w}_{i,T+1},$$

where

$$\hat{\boldsymbol{\theta}}'_{i,EB} = (\hat{\sigma}_i^{-2}\boldsymbol{W}_i'\boldsymbol{W}_i + \hat{\boldsymbol{\Omega}}_\theta^{-1})^{-1}(\hat{\sigma}_i^{-2}\boldsymbol{W}_i'\boldsymbol{y}_i + \hat{\boldsymbol{\Omega}}_\theta^{-1}\ \bar{\hat{\boldsymbol{\theta}}}),$$

$$\bar{\hat{\boldsymbol{\theta}}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\theta}}_{i,T}, \quad \hat{\sigma}_i^2 = \widehat{\boldsymbol{\varepsilon}}_i'\widehat{\boldsymbol{\varepsilon}}_i/(T-K),$$

$$\hat{\boldsymbol{\Omega}}_\theta = \frac{1}{N}\sum_{i=1}^{N}(\hat{\boldsymbol{\theta}}_{i,T} - \bar{\hat{\boldsymbol{\theta}}})(\hat{\boldsymbol{\theta}}_{i,T} - \bar{\hat{\boldsymbol{\theta}}})',$$

and $\widehat{\boldsymbol{\varepsilon}} = \boldsymbol{y}_i - \boldsymbol{W}_i\hat{\boldsymbol{\theta}}_{i,T}$ with $\hat{\boldsymbol{\theta}}_{i,T}$ given in (S.14).

Hierarchical Bayesian forecast We apply the hierarchical Bayesian model of Lindley and Smith (1972) which assumes $\varepsilon_{it} \sim iid\mathrm{N}(0,\sigma^2)$, using the following priors:

$$\begin{aligned}
\boldsymbol{\theta}_i &\sim \mathrm{N}(\bar{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}),\\
\bar{\boldsymbol{\theta}} &\sim \mathrm{N}(\boldsymbol{d}, \boldsymbol{S}_{\bar{\theta}}),\\
\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} &\sim \mathrm{Wishart}(\nu_\Sigma, (\nu_\Sigma \boldsymbol{S}_\Sigma)^{-1}),\\
\sigma^2 &\sim \mathrm{invGamma}(\nu_\sigma/2, \nu_\sigma s^2/2).
\end{aligned}$$

The Gibbs sampler uses the conditional posteriors (Gelfand et al., 1990) as set out below, where $|\cdot$ denotes conditional on the other parameters in the Gibbs sampler, for $r_b = 1, 2, \ldots, R_b$, where R_b denotes the number of random draws used in the Gibbs sampler:

- $\boldsymbol{\theta}_{i,r_b}|\cdot \sim \mathrm{N}(\boldsymbol{b}_\mathrm{i}, \boldsymbol{S}_\mathrm{i})$, where $\boldsymbol{b}_i = \boldsymbol{S}_i\left(\sigma_{r_b-1}^{-2}\boldsymbol{W}_i'\boldsymbol{y}_i + \boldsymbol{\Sigma}_{\boldsymbol{\theta},\,r_b-1}^{-1}\bar{\boldsymbol{\theta}}_{r_b-1}\right)$,
  and $\boldsymbol{S}_i = \left(\sigma_{r_b-1}^{-2}\boldsymbol{W}_i'\boldsymbol{W}_i + \boldsymbol{\Sigma}_{\boldsymbol{\theta},\,r_b-1}^{-1}\right)^{-1}$
- $\sigma_{r_b}^2|\cdot \sim \mathrm{invGamma}\left([NT+\nu_\sigma]/2, \frac{1}{2}\left[\sum_{i=1}^{N}(\boldsymbol{y}_i - \boldsymbol{W}_i\boldsymbol{\theta}_{i,r_b})'(\boldsymbol{y}_i - \boldsymbol{W}_i\boldsymbol{\theta}_{i,r_b}) + \nu_\sigma s^2\right]\right)$
- $\bar{\boldsymbol{\theta}}_{r_b}|\cdot \sim \mathrm{N}(\boldsymbol{h}, \boldsymbol{S}_h)$, where $\boldsymbol{h} = \boldsymbol{S}_h\left(\boldsymbol{\Sigma}_{\boldsymbol{\theta},r_b-1}^{-1}\sum_{i=1}^{N}\boldsymbol{\theta}_{i,r_b} + \boldsymbol{S}_{\bar{\theta}}^{-1}\boldsymbol{d}\right)$ and $\boldsymbol{S}_h = \left(N\boldsymbol{\Sigma}_{\boldsymbol{\theta},r_b-1}^{-1} + \boldsymbol{S}_{\bar{\theta}}^{-1}\right)^{-1}$
- $\boldsymbol{\Sigma}_{\boldsymbol{\theta},r_b}^{-1}|\cdot \sim \mathrm{Wishart}\left(N+\nu_\Sigma,\ \left[\sum_{i=1}^{N}\left(\boldsymbol{\theta}_{i,r_b} - \bar{\boldsymbol{\theta}}_{r_b}\right)\left(\boldsymbol{\theta}_{i,r_b} - \bar{\boldsymbol{\theta}}_{r_b}\right)' + \nu_\Sigma\boldsymbol{S}_\Sigma\right]^{-1}\right)$

Table S.4: Results for alternative priors in the applications

	House price		CPI		
Priors: a, b, s^2	SAR	SARX	AR	AR-PC	ARX
2,2,0.1	0.995	0.987	0.998	0.999	0.987
4,4,0.1	1.000	1.000	1.000	1.000	1.000
6,0,1	0.970	0.946	0.981	0.980	0.951

Note: The table reports the ratio of MSFEs for the hierarchical Bayesian model for different priors. In the first columns are the priors, where the first number is the exponent a for $\boldsymbol{S}_{\bar{\theta}} = \boldsymbol{I}_K 10^a$, the second number the exponent b for $\boldsymbol{S}_{\boldsymbol{\Sigma}} = \boldsymbol{I}_K 10^b$, and the final number the prior for s^2. For further details see the footnote of Table 6 and 9.

The Gibbs sampler draws iteratively from the conditional posterior distributions, starting with the following initial values ($r_b = 0$)

$$\sigma_0^2 = \widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}/(NT-K), \quad \widehat{\boldsymbol{\varepsilon}} = (\widehat{\boldsymbol{\varepsilon}}_1, \widehat{\boldsymbol{\varepsilon}}_2, \ldots, \widehat{\boldsymbol{\varepsilon}}_N)', \quad \widehat{\boldsymbol{\varepsilon}}_i = \boldsymbol{y}_i - \boldsymbol{W}_i \hat{\boldsymbol{\theta}}_{i,T}$$

$$\bar{\boldsymbol{\theta}}_0 = \frac{1}{N}\sum_{i=1}^{N} \hat{\boldsymbol{\theta}}_{i,T}, \quad \text{and} \quad \boldsymbol{\Sigma}_{\bar{\boldsymbol{\theta}},0}^{-1} = \frac{1}{N}\sum_{i=1}^{N} (\hat{\boldsymbol{\theta}}_{i,T} - \bar{\boldsymbol{\theta}}_0)(\hat{\boldsymbol{\theta}}_{i,T} - \bar{\boldsymbol{\theta}}_0)'.$$

Estimates from the Gibbs sampler are obtained from 1500 iterations with the first 500 discarded as a burn-in sample. In each iteration, we calculate

$$\hat{y}_{i,T+1,r_b}^{HB} = \hat{\boldsymbol{\theta}}_{i,r_b}' \boldsymbol{w}_{i,T+1}, \tag{S.20}$$

for $i = 1, 2, \ldots, N$ and the forecast is then $\hat{y}_{i,T+1}^{HB} = \frac{1}{R_b}\sum_{r_b=1}^{R_b} \hat{y}_{i,T+1,r_b}^{HB}$.

We use the following hyperpriors: $\boldsymbol{d} = \boldsymbol{0}$, $\boldsymbol{S}_{\bar{\theta}} = \boldsymbol{I}_K 10^6$, $\boldsymbol{S}_{\Sigma} = \boldsymbol{I}_K$, $\nu_{\Sigma} = K$, $\nu_{\sigma} = 0.1$, and $s^2 = 0.1$, which are proper priors that are weakly informative and avoid the use of uninformative priors that appear to be difficult to attain in hierarchical models (Gelman, 2006). Results for alternative priors are reported in Table S.4. The results suggest that the choice of prior for the error variance has relatively little influence, whereas the prior choices for the parameter covariances can substantially alter the forecast accuracy.

S.4 Additional Monte Carlo applications and empirical results

As a practical alternative to the combination forecasts in Section 4, which are based on estimates of the optimal combination weights, forecast combinations using equal weights have a long history in the literature (Timmermann, 2006). We therefore considered how this forecast combination scheme performs both in the Monte Carlo simulations and for the empirical applications. As in the paper, we separately consider combination schemes for the individual-pooled forecasts and for the individual-FE forecasts.

Columns 3-8 of Tables S.5 and S.6 report results for the two DGPs in our Monte Carlo experiments. Relative to the benchmark individual forecasts, the equal-weighted combination of the individual-pooled forecasts performs well at lower levels of heterogeneity when the time-series dimension T is small. However, forecasts from this scheme quickly become inferior to the individual forecasts as the level of parameter heterogeneity rises. This result is unsurprising: As the level of parameter heterogeneity increases, the pooled forecasts start to be dominated by the rising bias term induced by parameter heterogeneity. The more sophisticated optimal combination scheme introduced in the paper can adjust to this by reducing the weight on the pooled forecasts. However, the equal-weighted forecast cannot adapt to the increasingly poor performance of the pooled forecasts which is why we see the rise in the MSFE ratios for this scheme at higher levels of parameter heterogeneity. Moreover, comparing the results in Tables S.5 and S.6 to those from the main paper, we see that even for the case with homogeneous parameters, the optimal combination scheme performs better than the equal-weighted individual-pooled combination. The reason is that when the parameters are homogeneous, it is actually optimal to put a much higher weight on the pooled forecasts than the equal-weighted scheme does and this is achieved by our optimal combination scheme.

The equal-weighted combination of the individual and FE forecasts performs a little worse than the individual-pooled combination only under the scenario with homogenous parameters which of course favors the pooled forecasts. However, as the level of parameter heterogeneity increases, the equal-weighted individual-FE combination adapts much better than the individual-pooled combination and avoids MSFE-ratios that are notably higher than unity. This happens because the FE forecasts can adapt to parameter heterogeneity in a way that the forecasts based on the pooled estimates fail to do. Still, compared to the optimal combination weights introduced in the paper,

using equal-weights leads to a notable deterioration in the performance of the individual-FE forecast combination as the level of parameter heterogeneity goes up.

Tables S.7 and S.8 report the quantiles of the MSFE ratios. In the scenario with homogeneous parameters, the quantiles are comparable, if slightly worse in the sense that the quantiles are shifted to the right, than those of the forecasts based on estimated combination weights. However, as the level of parameter heterogeneity increases, the distributions of ratios become relatively wider for the equal-weighted combination scheme and the performance of these forecasts, particularly in the right tail, is notably worse than that of the optimal combination scheme.

Overall, we conclude from these Monte Carlo simulations that the optimal forecast combination scheme introduced in our paper produces more accurate forecasts that are notably more robust to parameter heterogeneity than the equal-weighted combination schemes considered here.

Table S.9 shows the performance of the equal-weighted forecasts for the application to house price inflation. For comparison, we also show the forecasting results for our optimal combination scheme. In this application pooling beats individual forecasts, which suggests a low degree of parameter heterogeneity. The equal-weighted forecast combinations perform correspondingly well. In fact, the combination of individual and pooled forecasts has the lowest average MSFE, offers the most precise forecasts for 10.2% (SAR model) and 14.9% (SARX) of MSAs and never produces the worst forecast. This performance is marginally better than that of the optimal combination schemes with estimated weights.

The quantiles of the ratios of MSFEs in Table S.10 reveal that the equal-weighted combined forecasts have a wider distribution than the forecasts based on the optimal combinations with estimated weights. Hence, compared to our optimal forecast combinations the equal-weighted combination has a higher chance of producing either very good forecasts (low MSFE ratios) or very poor forecasts(high MSFE ratios) for individual units in the cross-section.

The results for the CPI application in Table S.11 show that in a similar fashion the equal-weighted combination provides precise forecasts, which are more accurate, on average, than the optimal forecast combination, though beaten by a small margin by the empirical Bayes forecasts. Further, Table S.12 shows that, again, the distribution of MSFE ratios is wider that the corresponding distribution using estimated weights.

Table S.13 shows the results from the panel and individual DM test statistics. For both applica-

Table S.5: Monte Carlo results: panel AR, equal and oracle weighted combinations

a_β	σ^2_α	Eq. w. (pool)			Eq. w. (FE)			Oracle w. (pool)			Oracle w. (FE)		
	T	20	50	100	20	50	100	20	50	100	20	50	100
$PR^2 = 0.2$													
$N = 50$													
0.0	0.0	0.892	0.968	0.985	0.928	0.982	0.992	0.856	0.957	0.980	0.904	0.976	0.990
0.0	0.5	0.938	1.019	1.035	0.928	0.982	0.992	0.939	0.993	0.998	0.904	0.976	0.990
0.5	0.5	0.957	1.039	1.055	0.934	0.988	0.998	0.950	0.995	0.999	0.923	0.988	0.997
1.0	1.0	1.011	1.099	1.116	0.954	1.008	1.017	0.968	0.997	0.999	0.954	0.996	0.999
$N = 100$													
0.0	0.0	0.887	0.967	0.985	0.922	0.981	0.992	0.849	0.955	0.980	0.897	0.975	0.990
0.0	0.5	0.941	1.024	1.043	0.922	0.981	0.992	0.940	0.993	0.998	0.897	0.975	0.990
0.5	0.5	0.956	1.040	1.059	0.929	0.988	0.999	0.949	0.995	0.999	0.917	0.988	0.997
1.0	1.0	1.013	1.100	1.121	0.951	1.009	1.020	0.968	0.997	0.999	0.952	0.996	0.999
$N = 1000$													
0.0	0.0	0.886	0.966	0.985	0.921	0.981	0.992	0.838	0.952	0.978	0.893	0.973	0.989
0.0	0.5	0.939	1.023	1.042	0.921	0.981	0.992	0.903	0.987	0.997	0.893	0.973	0.989
0.5	0.5	0.952	1.038	1.057	0.927	0.988	0.999	0.927	0.991	0.998	0.921	0.990	0.998
1.0	1.0	1.009	1.098	1.119	0.949	1.009	1.020	0.958	0.996	0.999	0.955	0.997	0.999
$PR^2 = 0.6$													
$N = 50$													
0.0	0.0	0.883	0.965	0.984	0.923	0.981	0.992	0.843	0.954	0.979	0.900	0.975	0.989
0.0	0.5	0.910	0.991	1.008	0.923	0.981	0.992	0.904	0.986	0.996	0.900	0.975	0.989
0.5	0.5	0.932	1.010	1.027	0.934	0.992	1.002	0.932	0.991	0.998	0.928	0.991	0.998
1.0	1.0	0.977	1.060	1.077	0.959	1.021	1.032	0.959	0.997	0.999	0.958	0.997	0.999
$N = 100$													
0.0	0.0	0.879	0.964	0.984	0.918	0.980	0.991	0.848	0.955	0.979	0.895	0.975	0.989
0.0	0.5	0.909	0.991	1.011	0.918	0.980	0.991	0.938	0.993	0.998	0.895	0.975	0.989
0.5	0.5	0.927	1.010	1.029	0.927	0.992	1.004	0.947	0.994	0.999	0.915	0.988	0.997
1.0	1.0	0.976	1.060	1.081	0.956	1.025	1.042	0.967	0.997	0.999	0.949	0.996	0.999
$N = 1000$													
0.0	0.0	0.878	0.963	0.984	0.917	0.979	0.991	0.838	0.951	0.978	0.893	0.973	0.989
0.0	0.5	0.908	0.991	1.010	0.917	0.979	0.991	0.902	0.986	0.997	0.893	0.973	0.989
0.5	0.5	0.926	1.009	1.029	0.927	0.992	1.004	0.927	0.991	0.998	0.920	0.990	0.998
1.0	1.0	0.974	1.061	1.081	0.954	1.024	1.040	0.957	0.996	0.999	0.953	0.997	0.999

Note: 'Eq. w.' refer to combinations with equal weights. 'Oracle w.' refer to combinations where the weights use the true parmeters $\boldsymbol{\theta}_i$ in the DGP. For further details see the footnote of Table 1.

tions, the panel DM test show significant improvements over the individual forecasts. For the house price applications, somewhat fewer forecasts for MSAs are significantly better than the individual forecast compared to what we find for the optimal combination scheme. For the CPI application, in contrast, the pooled forecast with equal weights is significantly more precise than the benchmark for slightly more series than under the optimal combination scheme.

Table S.6: Monte Carlo results: panel ARX, equal weights combinations

a_β	σ^2_α	Eq. w. (pool)			Eq. w. (FE)			Oracle w. (pool)			Oracle w. (FE)		
	T	20	50	100	20	50	100	20	50	100	20	50	100
$PR^2 = 0.2$													
$N = 50$, $\rho_{\gamma x} = 0$													
0.0	0.0	0.849	0.950	0.977	0.882	0.964	0.984	0.799	0.934	0.969	0.843	0.953	0.979
0.0	0.5	0.903	1.008	1.036	0.882	0.964	0.984	0.903	0.986	0.996	0.843	0.953	0.979
0.5	0.5	0.927	1.033	1.061	0.906	0.988	1.008	0.925	0.990	0.997	0.903	0.985	0.996
1.0	1.0	0.988	1.101	1.132	0.941	1.025	1.046	0.952	0.994	0.999	0.942	0.993	0.998
$N = 50$, $\rho_{\gamma x} = 0.5$													
0.0	0.5	0.900	1.005	1.033	0.882	0.964	0.984	0.900	0.985	0.996	0.843	0.953	0.979
0.5	0.5	0.928	1.036	1.064	0.904	0.987	1.007	0.926	0.990	0.997	0.900	0.984	0.996
1.0	1.0	0.991	1.105	1.137	0.940	1.023	1.044	0.953	0.994	0.999	0.940	0.993	0.998
$N = 100$, $\rho_{\gamma x} = 0$													
0.0	0.0	0.844	0.949	0.977	0.876	0.963	0.984	0.792	0.931	0.969	0.836	0.951	0.979
0.0	0.5	0.888	0.997	1.027	0.876	0.963	0.984	0.888	0.983	0.996	0.836	0.951	0.979
0.5	0.5	0.921	1.030	1.060	0.907	0.993	1.015	0.920	0.989	0.998	0.904	0.986	0.997
1.0	1.0	0.991	1.106	1.139	0.953	1.041	1.064	0.951	0.994	0.999	0.945	0.994	0.999
$N = 100$, $\rho_{\gamma x} = 0.5$													
0.0	0.5	0.887	0.996	1.026	0.876	0.963	0.984	0.887	0.983	0.996	0.836	0.951	0.979
0.5	0.5	0.929	1.039	1.070	0.908	0.995	1.016	0.925	0.990	0.998	0.906	0.987	0.997
1.0	1.0	1.004	1.121	1.155	0.955	1.044	1.066	0.954	0.995	0.999	0.946	0.994	0.999
$N = 1000$, $\rho_{\gamma x} = 0$													
0.0	0.0	0.840	0.948	0.976	0.872	0.963	0.984	0.786	0.931	0.968	0.830	0.950	0.978
0.0	0.5	0.891	1.005	1.034	0.872	0.963	0.984	0.892	0.985	0.996	0.830	0.950	0.978
0.5	0.5	0.926	1.039	1.068	0.903	0.993	1.014	0.922	0.990	0.998	0.900	0.986	0.997
1.0	1.0	0.999	1.119	1.151	0.948	1.041	1.063	0.953	0.994	0.999	0.942	0.993	0.999
$N = 1000$, $\rho_{\gamma x} = 0.5$													
0.0	0.5	0.888	1.001	1.030	0.872	0.963	0.984	0.888	0.984	0.996	0.830	0.950	0.978
0.5	0.5	0.933	1.047	1.077	0.902	0.992	1.013	0.927	0.991	0.998	0.898	0.985	0.997
1.0	1.0	1.008	1.129	1.162	0.947	1.039	1.062	0.955	0.995	0.999	0.941	0.993	0.999
$PR^2 = 0.6$													
$N = 50$, $\rho_{\gamma x} = 0$													
0.0	0.0	0.841	0.948	0.975	0.877	0.963	0.983	0.788	0.931	0.967	0.839	0.952	0.978
0.0	0.5	0.870	0.975	1.002	0.877	0.963	0.983	0.855	0.973	0.993	0.839	0.952	0.978
0.5	0.5	0.904	1.006	1.034	0.908	0.991	1.011	0.904	0.985	0.996	0.904	0.985	0.996
1.0	1.0	0.963	1.069	1.099	0.949	1.036	1.059	0.944	0.993	0.998	0.945	0.993	0.999
$N = 50$, $\rho_{\gamma x} = 0.5$													
0.0	0.5	0.869	0.973	1.001	0.877	0.963	0.983	0.853	0.972	0.992	0.839	0.952	0.978
0.5	0.5	0.902	1.004	1.033	0.906	0.990	1.010	0.902	0.985	0.996	0.902	0.985	0.996
1.0	1.0	0.957	1.064	1.095	0.945	1.034	1.058	0.941	0.994	0.998	0.941	0.993	0.999
$N = 100$, $\rho_{\gamma x} = 0$													
0.0	0.0	0.835	0.946	0.975	0.871	0.961	0.983	0.781	0.927	0.967	0.832	0.948	0.978
0.0	0.5	0.862	0.971	1.001	0.871	0.961	0.983	0.844	0.971	0.993	0.832	0.948	0.978
0.5	0.5	0.901	1.009	1.040	0.906	0.997	1.020	0.901	0.986	0.997	0.904	0.986	0.997
1.0	1.0	0.971	1.083	1.116	0.958	1.056	1.085	0.945	0.993	0.998	0.946	0.994	0.999
$N = 100$, $\rho_{\gamma x} = 0.5$													
0.0	0.5	0.861	0.970	1.000	0.871	0.961	0.983	0.842	0.970	0.992	0.832	0.948	0.978
0.5	0.5	0.904	1.012	1.043	0.907	0.999	1.022	0.904	0.986	0.997	0.906	0.987	0.997
1.0	1.0	0.976	1.089	1.122	0.961	1.060	1.089	0.947	0.994	0.998	0.948	0.994	0.999
$N = 1000$, $\rho_{\gamma x} = 0$													
0.0	0.0	0.834	0.945	0.975	0.869	0.960	0.983	0.779	0.927	0.967	0.829	0.948	0.977
0.0	0.5	0.863	0.973	1.002	0.869	0.960	0.983	0.846	0.972	0.993	0.829	0.948	0.977
0.5	0.5	0.902	1.011	1.041	0.903	0.996	1.019	0.902	0.986	0.997	0.901	0.986	0.997
1.0	1.0	0.969	1.083	1.115	0.955	1.056	1.084	0.944	0.993	0.998	0.944	0.994	0.999
$N = 1000$, $\rho_{\gamma x} = 0.5$													
0.0	0.5	0.861	0.971	1.001	0.869	0.960	0.983	0.843	0.971	0.992	0.829	0.948	0.977
0.5	0.5	0.902	1.011	1.041	0.902	0.996	1.018	0.902	0.986	0.997	0.900	0.986	0.997
1.0	1.0	0.967	1.082	1.114	0.953	1.055	1.082	0.944	0.994	0.998	0.943	0.994	0.999

Note: See footnotes of Tables 1 and S.5 for details.

Table S.7: Monte Carlo results: quantiles of ratios of MSFEs for panel AR models, $N = 1000$ and $T = 50$

	$PR^2 = 0.2$							$PR^2 = 0.6$						
Quantiles	0.01	0.05	0.10	0.50	0.90	0.95	0.99	0.01	0.05	0.10	0.50	0.90	0.95	0.99
$a_\beta = 0.0$, $\sigma_\alpha^2 = 0.0$														
Comb. (pool, eq.w.)	0.953	0.960	0.962	0.966	0.970	0.971	0.973	0.953	0.958	0.959	0.964	0.967	0.969	0.970
Comb. (FE, eq.w.)	0.969	0.976	0.977	0.981	0.985	0.986	0.987	0.970	0.974	0.976	0.979	0.983	0.983	0.985
Comb. (pool, Oracle)	0.935	0.946	0.948	0.955	0.962	0.963	0.967	0.936	0.942	0.944	0.951	0.958	0.959	0.962
Comb. (FE, Oracle)	0.958	0.967	0.969	0.975	0.980	0.982	0.985	0.959	0.965	0.968	0.973	0.978	0.980	0.983
$a_\beta = 0.0$, $\sigma_\alpha^2 = 0.5$														
Comb. (pool, eq.w.)	0.994	0.998	1.000	1.014	1.082	1.115	1.192	0.978	0.981	0.983	0.990	1.004	1.011	1.027
Comb. (FE, eq.w.)	0.969	0.976	0.977	0.981	0.985	0.986	0.987	0.970	0.974	0.976	0.979	0.983	0.983	0.985
Comb. (pool, Oracle)	0.987	0.989	0.990	0.993	1.000	1.003	1.010	0.979	0.981	0.982	0.986	0.992	0.994	1.000
Comb. (FE, Oracle)	0.958	0.967	0.969	0.975	0.980	0.982	0.985	0.959	0.965	0.968	0.973	0.978	0.980	0.983
$a_\beta = 0.5$, $\sigma_\alpha^2 = 0.5$														
Comb. (pool, eq.w.)	0.979	0.987	0.993	1.036	1.100	1.130	1.201	0.965	0.972	0.977	1.007	1.045	1.052	1.063
Comb. (FE, eq.w.)	0.973	0.978	0.979	0.986	1.000	1.005	1.012	0.972	0.977	0.978	0.986	1.016	1.032	1.054
Comb. (pool, Oracle)	0.989	0.990	0.991	0.994	0.999	1.001	1.007	0.981	0.983	0.984	0.991	0.998	1.000	1.003
Comb. (FE, Oracle)	0.973	0.978	0.980	0.986	0.999	1.005	1.011	0.976	0.980	0.981	0.986	1.007	1.017	1.031
$a_\beta = 1.0$, $\sigma_\alpha^2 = 1.0$														
Comb. (pool, eq.w.)	0.979	0.993	1.005	1.083	1.227	1.256	1.314	0.961	0.970	0.978	1.045	1.165	1.189	1.217
Comb. (FE, eq.w.)	0.975	0.980	0.982	0.998	1.053	1.071	1.106	0.975	0.978	0.981	1.002	1.076	1.152	1.326
Comb. (pool, Oracle)	0.993	0.994	0.994	0.996	1.000	1.001	1.004	0.985	0.987	0.988	0.995	1.006	1.008	1.012
Comb. (FE, Oracle)	0.988	0.990	0.991	0.994	1.003	1.008	1.014	0.989	0.991	0.992	0.994	1.004	1.013	1.026

Note: See footnotes of Tables 1 and S.5 for details.

Table S.8: Monte Carlo results: quantiles of ratio of MSFEs for panel ARX models, $N = 1000$ and $T = 50$

	$\rho_{x\beta} = 0$							$\rho_{\gamma x} = 0.5$						
Quantiles	0.01	0.05	0.10	0.50	0.90	0.95	0.99	0.01	0.05	0.10	0.50	0.90	0.95	0.99
$PR^2 = 0.2$														
$a_\beta = 0.0$, $\sigma_\alpha^2 = 0.0$														
Comb. (pool, eq.w.)	0.935	0.940	0.942	0.949	0.954	0.955	0.958	0.935	0.940	0.942	0.949	0.954	0.955	0.958
Comb. (FE, eq.w.)	0.950	0.954	0.957	0.964	0.968	0.970	0.972	0.950	0.954	0.957	0.964	0.968	0.970	0.972
Comb. (pool, Oracle)	0.911	0.919	0.922	0.932	0.940	0.942	0.946	0.911	0.919	0.922	0.932	0.940	0.942	0.946
Comb. (FE, Oracle)	0.931	0.939	0.942	0.951	0.959	0.960	0.964	0.931	0.939	0.942	0.951	0.959	0.960	0.964
$a_\beta = 0.0$, $\sigma_\alpha^2 = 0.5$														
Comb. (pool, eq.w.)	0.972	0.979	0.981	0.995	1.057	1.102	1.206	0.971	0.976	0.979	0.993	1.051	1.092	1.197
Comb. (FE, eq.w.)	0.950	0.954	0.957	0.964	0.968	0.970	0.972	0.950	0.954	0.957	0.964	0.968	0.970	0.972
Comb. (pool, Oracle)	0.975	0.977	0.979	0.983	0.996	1.005	1.027	0.974	0.976	0.978	0.983	0.995	1.003	1.027
Comb. (FE, Oracle)	0.931	0.939	0.942	0.951	0.959	0.960	0.964	0.931	0.939	0.942	0.951	0.959	0.960	0.964
$a_\beta = 0.5$, $\sigma_\alpha^2 = 0.5$														
Comb. (pool, eq.w.)	0.962	0.976	0.986	1.035	1.129	1.177	1.382	0.973	0.983	0.992	1.039	1.118	1.157	1.396
Comb. (FE, eq.w.)	0.956	0.962	0.965	0.981	1.050	1.090	1.337	0.955	0.962	0.965	0.980	1.044	1.105	1.304
Comb. (pool, Oracle)	0.981	0.983	0.985	0.990	0.999	1.003	1.019	0.983	0.984	0.986	0.990	0.997	1.001	1.019
Comb. (FE, Oracle)	0.971	0.975	0.977	0.983	1.003	1.016	1.088	0.970	0.974	0.976	0.982	1.002	1.020	1.080
$a_\beta = 1.0$, $\sigma_\alpha^2 = 1.0$														
Comb. (pool, eq.w.)	0.966	0.987	1.005	1.108	1.289	1.388	1.713	0.973	0.989	1.004	1.102	1.297	1.366	1.816
Comb. (FE, eq.w.)	0.960	0.967	0.971	1.017	1.161	1.255	1.937	0.959	0.966	0.972	1.017	1.152	1.272	1.695
Comb. (pool, Oracle)	0.988	0.990	0.990	0.994	1.000	1.003	1.012	0.989	0.990	0.991	0.994	1.000	1.003	1.012
Comb. (FE, Oracle)	0.986	0.988	0.989	0.992	1.002	1.007	1.042	0.985	0.987	0.988	0.992	1.002	1.010	1.038
$PR^2 = 0.6$														
$a_\beta = 0.0$, $\sigma_\alpha^2 = 0.0$														
Comb. (pool, eq.w.)	0.932	0.936	0.939	0.946	0.951	0.952	0.955	0.932	0.936	0.939	0.946	0.951	0.952	0.955
Comb. (FE, eq.w.)	0.948	0.952	0.955	0.961	0.966	0.967	0.969	0.948	0.952	0.955	0.961	0.966	0.967	0.969
Comb. (pool, Oracle)	0.907	0.914	0.917	0.927	0.935	0.938	0.942	0.907	0.914	0.917	0.927	0.935	0.938	0.942
Comb. (FE, Oracle)	0.930	0.936	0.939	0.949	0.956	0.958	0.962	0.930	0.936	0.939	0.949	0.956	0.958	0.962
$a_\beta = 0.0$, $\sigma_\alpha^2 = 0.5$														
Comb. (pool, eq.w.)	0.956	0.961	0.964	0.973	0.986	0.997	1.021	0.956	0.960	0.962	0.971	0.984	0.994	1.022
Comb. (FE, eq.w.)	0.948	0.952	0.955	0.961	0.966	0.967	0.969	0.948	0.952	0.955	0.961	0.966	0.967	0.969
Comb. (pool, Oracle)	0.959	0.963	0.965	0.972	0.981	0.988	1.004	0.957	0.962	0.963	0.971	0.980	0.987	1.007
Comb. (FE, Oracle)	0.930	0.936	0.939	0.949	0.956	0.958	0.962	0.930	0.936	0.939	0.949	0.956	0.958	0.962
$a_\beta = 0.5$, $\sigma_\alpha^2 = 0.5$														
Comb. (pool, eq.w.)	0.950	0.960	0.967	1.007	1.078	1.126	1.231	0.950	0.959	0.966	1.007	1.073	1.107	1.287
Comb. (FE, eq.w.)	0.953	0.960	0.963	0.984	1.056	1.101	1.429	0.953	0.960	0.963	0.983	1.053	1.109	1.320
Comb. (pool, Oracle)	0.973	0.976	0.978	0.986	0.998	1.005	1.025	0.973	0.976	0.978	0.986	0.998	1.003	1.029
Comb. (FE, Oracle)	0.970	0.974	0.975	0.982	1.005	1.019	1.120	0.970	0.973	0.975	0.982	1.004	1.021	1.085
$a_\beta = 1.0$, $\sigma_\alpha^2 = 1.0$														
Comb. (pool, eq.w.)	0.949	0.963	0.975	1.068	1.239	1.333	1.609	0.947	0.961	0.973	1.068	1.228	1.308	1.683
Comb. (FE, eq.w.)	0.959	0.966	0.971	1.028	1.189	1.285	1.880	0.960	0.966	0.971	1.025	1.185	1.278	1.732
Comb. (pool, Oracle)	0.981	0.983	0.985	0.993	1.005	1.011	1.030	0.990	0.991	0.991	0.994	0.998	0.999	1.004
Comb. (FE, Oracle)	0.985	0.987	0.988	0.992	1.003	1.009	1.041	0.986	0.988	0.989	0.992	1.002	1.007	1.034

Note: The table reports the quantiles of the ratios of MSFEs over the different individuals. See footnotes of Tables 1 and S.5 for further details.

Table S.9: House price inflation forecasts, including equal-weighted combinations

Forecast methods	Ratio of ave. MSFE		Freq. beating benchmark		Freq. smallest MSFE		Freq. largest MSFE	
	SAR	SARX	SAR	SARX	SAR	SARX	SAR	SARX
Individual	3.253	3.248	–	–	0.066	0.047	0.613	0.384
Pooled	0.969	0.990	0.660	0.417	0.213	0.086	0.202	0.365
RE	0.973	0.993	0.685	0.428	0.105	0.041	0.014	0.014
FE	0.983	1.002	0.682	0.450	0.191	0.061	0.160	0.229
Emp.Bayes	0.961	0.935	0.884	0.878	0.130	0.215	0.000	0.003
Hier.Bayes	0.984	0.967	0.859	0.840	0.025	0.135	0.008	0.006
Comb. (pool)	0.963	0.944	0.865	0.859	0.039	0.069	0.003	0.000
Comb. (FE)	0.970	0.948	0.867	0.865	0.033	0.030	0.000	0.000
Comb. (eq.weight,pool)	0.949	0.923	0.845	0.768	0.105	0.157	0.000	0.000
Comb. (eq.weight,FE)	0.957	0.930	0.862	0.757	0.094	0.157	0.000	0.000

Note: See the footnotes of Tables 1, 6 and S.5 for further details.

Table S.10: Quantiles of ratio of MSFEs for house price inflation over MSAs, equal-weighted combinations

Quantiles	0.01	0.05	0.10	0.50	0.90	0.95	0.99
House prices: SAR							
Comb. (eq.weight,pool)	0.853	0.884	0.906	0.962	1.010	1.025	1.065
Comb. (eq.weight,FE)	0.861	0.899	0.914	0.963	1.010	1.033	1.052
House prices: SARX							
Comb. (eq.weight,pool)	0.796	0.835	0.862	0.957	1.031	1.053	1.101
Comb. (eq.weight,FE)	0.811	0.840	0.866	0.959	1.025	1.047	1.097

Note: The table reports the quantiles of the distribution of MSFE ratios across MSAs. See Tables 7 and S.7 for further details.

S.5 Choice of initial condition

Tables S.14 to S.17 repeat the Monte Carlo experiments reported in Tables 1 to 2 for alternative values of κ to check whether initializing the DGP out of equilibrium matters for the results. We consider two values, namely $\kappa = 1/2$ and $\kappa = 2$. For efficiency, we restrict ourselves to $N = 50$ and 100 and omit the hierarchical Bayesian forecasts. The results indicate that the choice of κ has a mild influence on the MSFE mainly when $T = 20$ but that the choice of initial condition does not have a large impact on the overall conclusions from the Monte Carlo experiments.

Table S.11: CPI inflation forecasting results, including equal-weighted combinations

Forecast method	Ratio of ave. MSFE			Freq. beating benchmark			Freq. smallest MSFE			Freq. largest MSFE		
	AR	AR-PC	ARX	AR	AR-PC	ARX	AR	AR-PC	ARX	AR	AR-PC	ARX
Individual	14.376	14.401	15.501	–	–	–	0.016	0.016	0.005	0.417	0.433	0.439
Pooled	0.938	0.936	0.878	0.449	0.460	0.444	0.187	0.187	0.193	0.439	0.439	0.396
RE	0.940	0.936	0.880	0.535	0.567	0.508	0.048	0.037	0.016	0.005	0.000	0.000
FE	0.944	0.941	0.883	0.583	0.604	0.508	0.037	0.053	0.000	0.139	0.128	0.166
Emp.Bayes	0.934	0.930	0.892	0.984	0.984	0.984	0.257	0.289	0.380	0.000	0.000	0.000
Hier.Bayes	0.992	0.992	0.968	0.904	0.925	0.936	0.123	0.112	0.075	0.000	0.000	0.000
Comb. (pool)	0.954	0.952	0.917	0.797	0.818	0.775	0.075	0.032	0.027	0.000	0.000	0.000
Comb. (FE)	0.965	0.964	0.927	0.850	0.866	0.850	0.027	0.011	0.011	0.000	0.000	0.000
Comb. (eq.weight,pool)	0.935	0.932	0.895	0.807	0.797	0.802	0.118	0.134	0.214	0.000	0.000	0.000
Comb. (eq.weight,FE)	0.944	0.940	0.900	0.872	0.882	0.856	0.112	0.128	0.080	0.000	0.000	0.000

Note: The column denoted 'AR' reports the results for an autoregressive specification, 'AR-PC' adds the first principal component of the panel of sub-indices, and 'ARX' further adds the default yield and the term spread to the model. For further details see the footnotes of Tables 1, 6 and S.5.

Table S.12: Quantiles of ratio of MSFEs for CPI inflation over subindices, equal-weighted combinations

Quantiles	0.01	0.05	0.10	0.50	0.90	0.95	0.99
CPI: AR							
Comb. (eq.weight,pool)	0.745	0.838	0.900	0.970	1.055	1.195	1.367
Comb. (eq.weight,FE)	0.759	0.843	0.904	0.968	1.011	1.054	1.169
CPI: AR-PC							
Comb. (eq.weight,pool)	0.731	0.834	0.891	0.968	1.079	1.201	1.312
Comb. (eq.weight,FE)	0.743	0.839	0.892	0.962	1.007	1.030	1.095
CPI: ARX							
Comb. (eq.weight,pool)	0.719	0.793	0.852	0.935	1.062	1.263	1.433
Comb. (eq.weight,FE)	0.733	0.809	0.867	0.935	1.022	1.108	1.439

Note: The table reports the quantiles of the distribution of ratios of MSFEs across subindices. See Tables 10 and S.5 for further details.

Table S.13: Diebold-Mariano test statistics: equal-weighted combinations

	Comb(eq.weights,Pool)	Comb(eq.weights,FE)
House Prices: SAR		
Panel DM	−14.876	−12.825
DM< −1.96	77	77
Insign.	283	283
DM> 1.96	2	2
House Prices: SARX		
Panel DM	−15.674	−14.525
DM< −1.96	72	69
Insign.	289	292
DM> 1.96	1	1
CPI: AR		
Panel DM	−6.606	−5.665
DM< −1.96	28	23
Insign.	71	78
DM> 1.96	2	0
CPI: AR-PC		
Panel DM	−6.793	−5.887
DM< −1.96	30	32
Insign.	70	69
DM> 1.96	1	0
CPI: ARX		
Panel DM	−11.529	−10.762
DM< −1.96	59	59
Insign.	41	42
DM> 1.96	1	0

Note: See footnotes of Table 8 and S.5 for details.

Table S.14: Monte Carlo results: panel AR, $\kappa = 0.5$

a_β	σ^2_α		Pooled			RE			FE			Empirical Bayes			Comb. (pool)			Comb. (FE)		
		T	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100
$PR^2 = 0.2$																				
$N = 50$																				
0.0	0.0		0.855	0.957	0.980	0.884	0.966	0.984	0.903	0.976	0.990	0.883	0.967	0.985	0.907	0.970	0.986	0.941	0.984	0.993
0.0	0.5		1.036	1.156	1.182	0.900	0.975	0.990	0.907	0.976	0.990	0.909	0.979	0.992	0.953	0.995	0.999	0.943	0.984	0.993
0.5	0.5		1.109	1.237	1.263	0.924	1.000	1.013	0.929	1.001	1.013	0.919	0.986	0.996	0.962	0.997	0.999	0.951	0.991	0.998
1.0	1.0		1.331	1.478	1.508	1.001	1.077	1.089	1.002	1.077	1.089	0.938	0.992	0.998	0.976	0.999	1.000	0.966	0.996	0.999
$N = 100$																				
0.0	0.0		0.850	0.956	0.980	0.889	0.973	0.989	0.898	0.975	0.990	0.877	0.966	0.985	0.903	0.969	0.985	0.937	0.983	0.993
0.0	0.5		1.056	1.181	1.212	0.894	0.974	0.989	0.900	0.976	0.990	0.903	0.978	0.991	0.953	0.995	0.999	0.938	0.984	0.993
0.5	0.5		1.117	1.245	1.278	0.920	1.001	1.016	0.925	1.002	1.016	0.914	0.986	0.996	0.959	0.996	0.999	0.947	0.991	0.998
1.0	1.0		1.343	1.486	1.525	1.006	1.085	1.101	1.009	1.085	1.101	0.937	0.993	0.998	0.974	0.998	1.000	0.963	0.996	0.999
$PR^2 = 0.6$																				
$N = 50$																				
0.0	0.0		0.839	0.955	0.979	0.882	0.971	0.985	0.891	0.975	0.989	0.879	0.967	0.984	0.897	0.968	0.985	0.933	0.983	0.992
0.0	0.5		0.939	1.051	1.076	0.891	0.974	0.989	0.898	0.976	0.989	0.907	0.979	0.991	0.937	0.990	0.997	0.938	0.984	0.993
0.5	0.5		1.018	1.130	1.155	0.947	1.025	1.035	0.953	1.025	1.034	0.919	0.986	0.996	0.954	0.995	0.999	0.952	0.992	0.998
1.0	1.0		1.202	1.328	1.354	1.091	1.179	1.180	1.100	1.178	1.179	0.936	0.991	0.998	0.973	0.998	0.999	0.970	0.997	0.999
$N = 100$																				
0.0	0.0		0.838	0.953	0.979	0.882	0.972	0.988	0.888	0.974	0.989	0.877	0.966	0.985	0.895	0.967	0.985	0.931	0.982	0.992
0.0	0.5		0.943	1.056	1.083	0.886	0.972	0.988	0.893	0.975	0.989	0.904	0.978	0.991	0.936	0.990	0.997	0.934	0.983	0.992
0.5	0.5		1.014	1.129	1.158	0.945	1.030	1.043	0.952	1.031	1.043	0.913	0.986	0.996	0.950	0.994	0.999	0.948	0.992	0.998
1.0	1.0		1.207	1.335	1.369	1.140	1.275	1.305	1.164	1.286	1.308	0.932	0.992	0.998	0.969	0.998	0.999	0.969	0.998	1.000

Note: The DGP uses $\kappa = 0.5$, see Section 6.1. For further details see the footnote of Table 1.

Table S.15: Monte Carlo results: panel ARX, $\kappa = 0.5$

a_β	σ^2_α	Pooled			RE			FE			Empirical Bayes			Comb. (pool)			Comb. (FE)		
	T	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100
$PR^2 = 0.2$																			
$N = 50$, $\rho_{\gamma x} = 0$																			
0.0	0.0	0.799	0.934	0.969	0.835	0.951	0.978	0.843	0.953	0.979	0.848	0.953	0.978	0.870	0.954	0.978	0.901	0.968	0.985
0.0	0.5	0.998	1.161	1.208	0.839	0.952	0.979	0.845	0.953	0.979	0.867	0.963	0.984	0.927	0.988	0.997	0.902	0.968	0.985
0.5	0.5	1.087	1.258	1.308	0.925	1.044	1.075	0.932	1.046	1.075	0.896	0.981	0.995	0.941	0.991	0.998	0.930	0.986	0.996
1.0	1.0	1.327	1.529	1.593	1.060	1.190	1.227	1.067	1.192	1.228	0.926	0.990	0.998	0.962	0.996	0.999	0.953	0.993	0.998
$N = 50$, $\rho_{\gamma x} = 0.5$																			
0.0	0.5	0.988	1.149	1.195	0.840	0.952	0.979	0.845	0.953	0.979	0.868	0.963	0.984	0.925	0.988	0.997	0.902	0.968	0.985
0.5	0.5	1.097	1.269	1.321	0.924	1.041	1.071	0.929	1.042	1.071	0.895	0.979	0.995	0.943	0.992	0.998	0.929	0.986	0.996
1.0	1.0	1.341	1.544	1.612	1.057	1.184	1.221	1.061	1.185	1.221	0.923	0.988	0.998	0.963	0.996	0.999	0.951	0.993	0.998
$N = 100$, $\rho_{\gamma x} = 0$																			
0.0	0.0	0.792	0.932	0.969	0.823	0.945	0.976	0.837	0.951	0.979	0.843	0.952	0.978	0.865	0.952	0.978	0.896	0.966	0.980
0.0	0.5	0.959	1.121	1.169	0.832	0.950	0.979	0.839	0.951	0.979	0.858	0.960	0.983	0.916	0.985	0.997	0.897	0.966	0.980
0.5	0.5	1.077	1.250	1.302	0.939	1.069	1.103	0.948	1.071	1.103	0.886	0.978	0.994	0.936	0.991	0.998	0.930	0.988	0.997
1.0	1.0	1.349	1.554	1.618	1.111	1.259	1.299	1.123	1.262	1.300	0.917	0.988	0.997	0.959	0.995	0.999	0.954	0.994	0.999
$N = 100$, $\rho_{\gamma x} = 0.5$																			
0.0	0.5	0.955	1.116	1.165	0.833	0.950	0.979	0.839	0.951	0.979	0.859	0.961	0.984	0.916	0.985	0.997	0.897	0.966	0.985
0.5	0.5	1.109	1.286	1.339	0.946	1.075	1.107	0.953	1.077	1.108	0.886	0.978	0.994	0.940	0.992	0.998	0.931	0.988	0.997
1.0	1.0	1.402	1.614	1.679	1.122	1.269	1.307	1.130	1.271	1.308	0.916	0.988	0.997	0.962	0.996	0.999	0.955	0.994	0.999
$PR^2 = 0.6$																			
$N = 50$, $\rho_{\gamma x} = 0$																			
0.0	0.0	0.787	0.933	0.967	0.829	0.951	0.977	0.834	0.953	0.978	0.847	0.954	0.978	0.864	0.952	0.977	0.896	0.967	0.984
0.0	0.5	0.887	1.034	1.074	0.832	0.951	0.977	0.839	0.953	0.978	0.869	0.964	0.983	0.904	0.979	0.994	0.899	0.967	0.984
0.5	0.5	0.997	1.155	1.201	0.939	1.064	1.093	0.949	1.066	1.093	0.899	0.981	0.995	0.930	0.988	0.997	0.931	0.987	0.997
1.0	1.0	1.223	1.407	1.464	1.133	1.292	1.323	1.152	1.297	1.321	0.926	0.989	0.998	0.958	0.995	0.999	0.958	0.994	0.999
$N = 50$, $\rho_{\gamma x} = 0.5$																			
0.0	0.5	0.883	1.029	1.069	0.832	0.951	0.977	0.840	0.953	0.978	0.870	0.964	0.983	0.903	0.979	0.994	0.899	0.967	0.984
0.5	0.5	0.998	1.150	1.196	0.941	1.062	1.090	0.949	1.062	1.089	0.897	0.980	0.995	0.931	0.988	0.997	0.930	0.986	0.996
1.0	1.0	1.214	1.390	1.447	1.127	1.271	1.304	1.139	1.272	1.303	0.923	0.988	0.997	0.956	0.994	0.999	0.956	0.994	0.998
$N = 100$, $\rho_{\gamma x} = 0$																			
0.0	0.0	0.781	0.928	0.968	0.821	0.946	0.976	0.828	0.949	0.978	0.841	0.952	0.978	0.858	0.950	0.977	0.890	0.965	0.984
0.0	0.5	0.872	1.023	1.068	0.825	0.947	0.978	0.832	0.950	0.978	0.859	0.960	0.983	0.895	0.977	0.994	0.893	0.965	0.984
0.5	0.5	1.006	1.171	1.221	0.954	1.090	1.126	0.965	1.093	1.127	0.886	0.978	0.994	0.927	0.989	0.997	0.930	0.988	0.997
1.0	1.0	1.269	1.464	1.527	1.179	1.357	1.409	1.200	1.364	1.410	0.913	0.986	0.997	0.957	0.995	0.999	0.957	0.995	0.999
$N = 100$, $\rho_{\gamma x} = 0.5$																			
0.0	0.5	0.870	1.019	1.063	0.826	0.948	0.978	0.833	0.950	0.978	0.861	0.960	0.983	0.895	0.976	0.994	0.893	0.965	0.984
0.5	0.5	1.020	1.184	1.233	0.963	1.097	1.131	0.973	1.100	1.132	0.886	0.979	0.994	0.930	0.989	0.998	0.931	0.988	0.997
1.0	1.0	1.294	1.489	1.550	1.201	1.367	1.415	1.220	1.372	1.416	0.912	0.987	0.997	0.958	0.995	0.999	0.958	0.995	0.999

Note: See the footnote of Table S.14 for details.

Table S.16: Monte Carlo results: panel AR, $\kappa = 2$

a_β	σ^2_α	Pooled			RE			FE			Empirical Bayes			Comb. (pool)			Comb. (FE)		
	T	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100
$PR^2 = 0.2$																			
$N = 50$																			
0.0	0.0	0.882	0.959	0.980	0.911	0.968	0.984	0.930	0.978	0.990	0.906	0.968	0.985	0.921	0.971	0.986	0.955	0.985	0.993
0.0	0.5	1.068	1.160	1.183	0.925	0.978	0.990	0.931	0.979	0.991	0.930	0.981	0.992	0.963	0.995	0.999	0.956	0.986	0.993
0.5	0.5	1.142	1.241	1.264	0.947	1.002	1.013	0.953	1.003	1.014	0.938	0.987	0.996	0.971	0.997	0.999	0.964	0.993	0.998
1.0	1.0	1.367	1.484	1.509	1.031	1.082	1.090	1.033	1.081	1.090	0.953	0.993	0.998	0.982	0.999	1.000	0.976	0.997	0.999
$N = 100$																			
0.0	0.0	0.880	0.958	0.980	0.919	0.975	0.989	0.928	0.978	0.990	0.905	0.968	0.985	0.919	0.970	0.986	0.953	0.985	0.993
0.0	0.5	1.092	1.185	1.213	0.920	0.977	0.990	0.925	0.978	0.990	0.926	0.981	0.992	0.964	0.995	0.999	0.952	0.985	0.993
0.5	0.5	1.153	1.249	1.279	0.945	1.003	1.016	0.950	1.004	1.017	0.936	0.988	0.996	0.969	0.996	0.999	0.961	0.992	0.998
1.0	1.0	1.382	1.492	1.527	1.039	1.090	1.102	1.043	1.090	1.102	0.953	0.994	0.998	0.980	0.998	1.000	0.975	0.997	0.999
$PR^2 = 0.6$																			
$N = 50$																			
0.0	0.0	0.888	0.961	0.980	0.928	0.976	0.987	0.938	0.981	0.991	0.920	0.972	0.985	0.924	0.972	0.986	0.959	0.987	0.993
0.0	0.5	1.002	1.067	1.080	0.925	0.979	0.991	0.931	0.981	0.991	0.934	0.983	0.992	0.957	0.992	0.998	0.955	0.987	0.994
0.5	0.5	1.063	1.148	1.161	0.991	1.042	1.044	1.008	1.042	1.043	0.937	0.988	0.996	0.965	0.996	0.999	0.971	0.995	0.998
1.0	1.0	1.234	1.342	1.360	1.155	1.235	1.235	1.183	1.242	1.234	0.948	0.992	0.998	0.977	0.998	0.999	0.982	0.999	1.000
$N = 100$																			
0.0	0.0	0.891	0.961	0.981	0.933	0.980	0.990	0.939	0.981	0.991	0.924	0.972	0.986	0.924	0.972	0.986	0.959	0.987	0.993
0.0	0.5	1.017	1.075	1.089	0.924	0.979	0.990	0.929	0.981	0.991	0.935	0.983	0.992	0.959	0.992	0.998	0.954	0.987	0.994
0.5	0.5	1.065	1.146	1.165	1.000	1.051	1.055	1.021	1.054	1.054	0.937	0.989	0.997	0.964	0.995	0.999	0.971	0.995	0.999
1.0	1.0	1.242	1.338	1.371	1.235	1.319	1.339	1.289	1.347	1.352	0.949	0.992	0.998	0.977	0.998	1.000	0.984	0.999	1.000

Note: The DGP uses $\kappa = 2$, see Section 6.2. For further details see the footnote of Table 1.

Table S.17: Monte Carlo results: panel ARX, $\kappa = 2$

a_β	σ^2_α	Pooled			RE			FE			Empirical Bayes			Comb. (pool)			Comb. (FE)		
	T	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100
$PR^2 = 0.2$																			
$N = 50$, $\rho_{\gamma x} = 0$																			
0.0	0.0	0.827	0.937	0.970	0.863	0.953	0.979	0.871	0.955	0.980	0.874	0.955	0.978	0.885	0.955	0.978	0.916	0.969	0.985
0.0	0.5	1.031	1.165	1.209	0.863	0.954	0.979	0.868	0.955	0.980	0.888	0.964	0.984	0.938	0.989	0.997	0.914	0.969	0.985
0.5	0.5	1.117	1.262	1.309	0.947	1.047	1.075	0.954	1.048	1.076	0.913	0.982	0.995	0.949	0.992	0.998	0.941	0.987	0.997
1.0	1.0	1.355	1.533	1.593	1.078	1.192	1.228	1.087	1.194	1.228	0.936	0.990	0.998	0.966	0.996	0.999	0.959	0.993	0.999
$N = 50$, $\rho_{\gamma x} = 0.5$																			
0.0	0.5	1.017	1.152	1.196	0.862	0.954	0.979	0.866	0.955	0.980	0.886	0.964	0.984	0.936	0.988	0.997	0.913	0.969	0.985
0.5	0.5	1.127	1.274	1.322	0.942	1.043	1.071	0.947	1.044	1.072	0.908	0.981	0.995	0.950	0.992	0.998	0.938	0.987	0.996
1.0	1.0	1.374	1.551	1.613	1.078	1.187	1.222	1.081	1.187	1.221	0.931	0.989	0.998	0.967	0.996	0.999	0.957	0.993	0.998
$N = 100$, $\rho_{\gamma x} = 0$																			
0.0	0.0	0.816	0.934	0.970	0.847	0.947	0.976	0.860	0.953	0.980	0.864	0.953	0.979	0.878	0.953	0.978	0.909	0.968	0.985
0.0	0.5	0.987	1.125	1.170	0.854	0.952	0.979	0.860	0.954	0.980	0.877	0.962	0.984	0.927	0.986	0.997	0.909	0.968	0.985
0.5	0.5	1.102	1.253	1.302	0.958	1.071	1.103	0.967	1.073	1.104	0.900	0.980	0.995	0.943	0.991	0.998	0.940	0.988	0.997
1.0	1.0	1.372	1.558	1.618	1.126	1.260	1.299	1.140	1.264	1.301	0.926	0.989	0.997	0.963	0.995	0.999	0.961	0.995	0.999
$N = 100$, $\rho_{\gamma x} = 0.5$																			
0.0	0.5	0.983	1.120	1.165	0.854	0.953	0.979	0.860	0.954	0.980	0.878	0.963	0.984	0.926	0.986	0.997	0.908	0.968	0.985
0.5	0.5	1.139	1.291	1.341	0.966	1.077	1.108	0.972	1.079	1.108	0.900	0.980	0.995	0.947	0.992	0.998	0.941	0.989	0.997
1.0	1.0	1.432	1.620	1.681	1.141	1.271	1.308	1.150	1.273	1.308	0.925	0.988	0.997	0.966	0.996	0.999	0.961	0.995	0.999
$PR^2 = 0.6$																			
$N = 50$, $\rho_{\gamma x} = 0$																			
0.0	0.0	0.834	0.940	0.970	0.873	0.957	0.979	0.878	0.959	0.980	0.887	0.959	0.979	0.888	0.957	0.978	0.920	0.971	0.986
0.0	0.5	0.951	1.051	1.080	0.865	0.956	0.979	0.870	0.958	0.980	0.894	0.967	0.985	0.926	0.982	0.995	0.916	0.970	0.986
0.5	0.5	1.036	1.167	1.205	0.977	1.075	1.100	0.997	1.077	1.098	0.913	0.982	0.995	0.941	0.989	0.997	0.948	0.989	0.997
1.0	1.0	1.240	1.408	1.466	1.201	1.340	1.374	1.248	1.360	1.374	0.934	0.990	0.998	0.962	0.995	0.999	0.969	0.995	0.999
$N = 50$, $\rho_{\gamma x} = 0.5$																			
0.0	0.5	0.940	1.044	1.074	0.863	0.956	0.979	0.867	0.957	0.980	0.892	0.966	0.985	0.923	0.981	0.994	0.914	0.970	0.985
0.5	0.5	1.042	1.170	1.204	0.980	1.075	1.098	0.997	1.075	1.096	0.909	0.981	0.995	0.941	0.989	0.997	0.946	0.988	0.997
1.0	1.0	1.245	1.405	1.454	1.179	1.315	1.349	1.208	1.323	1.349	0.929	0.988	0.997	0.960	0.995	0.998	0.965	0.995	0.999
$N = 100$, $\rho_{\gamma x} = 0$																			
0.0	0.0	0.824	0.935	0.969	0.862	0.952	0.978	0.869	0.955	0.980	0.879	0.956	0.979	0.882	0.954	0.978	0.914	0.969	0.985
0.0	0.5	0.930	1.038	1.072	0.858	0.954	0.979	0.864	0.955	0.980	0.886	0.964	0.984	0.918	0.980	0.994	0.911	0.969	0.985
0.5	0.5	1.046	1.181	1.225	0.992	1.098	1.129	1.013	1.104	1.129	0.903	0.980	0.995	0.939	0.989	0.998	0.948	0.990	0.998
1.0	1.0	1.296	1.468	1.528	1.251	1.397	1.443	1.299	1.421	1.453	0.922	0.986	0.997	0.962	0.995	0.999	0.969	0.996	0.999
$N = 100$, $\rho_{\gamma x} = 0.5$																			
0.0	0.5	0.926	1.034	1.068	0.857	0.953	0.979	0.862	0.955	0.980	0.885	0.964	0.984	0.916	0.979	0.994	0.910	0.969	0.985
0.5	0.5	1.069	1.199	1.238	1.009	1.108	1.135	1.030	1.113	1.136	0.903	0.980	0.995	0.942	0.990	0.998	0.949	0.990	0.998
1.0	1.0	1.328	1.498	1.553	1.277	1.408	1.448	1.327	1.425	1.454	0.920	0.986	0.997	0.963	0.995	0.999	0.970	0.996	0.999

Note: See the footnote of Table S.16 for details.

References

Baltagi, B.H. (2013) "Panel data forecasting" Ch. 18 in Elliott, G. and A. Timmermann, *Handbook of Economic Forecasting*, volume 2B. North Holland: Elsevier.

Gelfand, A.E., S.E. Hills, A. Racine-Poon, A.F.M. Smith (1996). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, 85, 972–985.

Goldberger, A.S. (1962) "Best linear unbiased prediction in the generalized linear regression model" *Journal of the American Statistical Association* 57, 369–375.

Hsiao C., M.H. Pesaran and A.K. Tahmiscioglu (1999) Bayes estimation of short-run coefficients in dynamic panel data models. Ch.11 in C. Hsiao, K. Lahiri, L.-F. Lee and M.H.Pesaran *A*nalysis of panels and limited dependent variable models, Cambridge: Cambridge University Press.

Lindley D.V. and A.F.M. Smith (1972) Bayesian estimates for the linear model. *Journal of the Royal Statistical Society, Series B* 34, 1-41.

Pesaran, M.H. (2015) *Time Series and Panel Data Econometrics*, Oxford University Press.

Pesaran, M.H., R. Smith (1995) "Estimating long-run relationships from dynamic heterogeneous panels" *Journal of Econometrics* 68, 134–152.

Timmermann, A. (2006) "Forecast combinations" Ch. 4 in G. Elliott, C. W. J. Granger and A. Timmermann (Eds.) *Handbook of Economic Forecasting*, volume. 1, North Holland: Elsevier.