Forecasting with panel data: Estimation uncertainty versus parameter heterogeneity*

M. Hashem Pesaran† Andreas Pick‡ Allan Timmermann§

August 15, 2025

Abstract

We provide a comprehensive examination of the predictive accuracy of panel forecasting methods based on individual, pooling, fixed effects, and empirical Bayes 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 dimensions of the data. Monte Carlo simulations and empirical applications to house prices and CPI inflation show that empirical Bayes and forecast combination 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, Empirical Bayes; 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, IAAE, and at 2024 NBER-NSF Time Series Conference held at University of Pennsylvania for helpful comments.

†University of Cambridge, UK, and University of Southern California, USA Email: mhp1@cam.ac.uk

‡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. We 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 forecast combination methods. Estimation errors are well-known to lead to imprecisely estimated combination weights for data with a small time-series dimension. Our main combination scheme assumes homogeneous weights across individual variables. This allows us to use cross-sectional information to reduce the effect of estimation error on the combination weights compared to the conventional combination scheme that lets the weights be individual-specific, which we also consider. To handle cases where the pooling estimator imposes too much homogeneity, we also consider combinations based on forecasts from the individual-specific and fixed effect estimators.

Our theoretical analysis of the individual and pooled estimation schemes 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. The estimation of the combination weights, however, requires $T \to \infty$, but at a much slower rate compared to N.

Finally, we consider forecasts based on the empirical Bayesian (EB) approach of Hsiao et al. (1999). 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.[1]

We evaluate the predictive accuracy of these alternative panel forecasting methods through Monte Carlo simulations. The simulations explore the importance to forecasting performance of the degree of parameter heterogeneity, how correlated it is, whether it affects intercepts or slopes, the value of the regressors in the forecast period, and 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, empirical Bayes and forecast combinations produce the most accurate forecasts and 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. Our first application considers predictability of house prices across 362 US metropolitan statistical areas (MSAs). In this application, individual-specific forecasts perform quite poorly, producing the highest MSFE values among all methods for more than 50% of the MSAs. Forecasts based on pooled estimates perform notably better and, across all forecasts, reduce the average MSFE value by 8% relative to the forecasts based on individual estimates, though this gets reversed if the regressor set is close to the sample average. Empirical Bayes and forecast combinations work even better in this application, beating forecasts

[1]In the Online Supplement we also report results based on the hierarchical Bayesian approach of Lindley and Smith (1972), Lee and Griffith (1979), and Maddala et al. (1997).

based on individual estimates for over 90% of MSAs while almost never generating the least accurate forecasts for individual series.

Our second application considers forecasts for a panel containing 187 subcategories of CPI inflation. In this application, forecasts based on individual estimates generate the highest MSFE-values for 44% of the series. Forecasts based on pooled estimates produce the highest MSFE-values for 40% of the individual series but, conversely, generate the lowest MSFE-values for 20% of the series. Combination forecasts produce lower MSFE values for 73-97% of the individual CPI series than the individual forecasts while almost never generating the largest MSFE value. Even better inflation forecasts are produced by the empirical Bayes method which is more accurate (in the MSFE sense) than the individual forecasts for 98% of the series and generates the lowest MSFE values for 39% of the individual variables while never generating the largest MSFE value.

Overall, forecasts that use only the information on a given unit tend to have loss distributions with wide dispersion 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, or fixed effects estimation tend to perform better, on average, than the individual forecasts whose accuracy they beat for the majority of series. However, relative to the individual-specific forecasts, these approaches also tend to have a right-skewed MSFE distribution, suggesting a high risk of poor forecasting performance for individual series whose model parameters are very different from the average. Combinations and Empirical Bayes forecasts have narrower MSFE distributions across units, often shifted to the left as they are centered around a smaller average loss, and rarely produce the largest squared forecast error among all methods we consider.

Related literature: 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 parameter 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. See Pick and Timmermann (2024) for a review of the literature.

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, they adopt a nonparametric Bayesian approach that shrinks the heterogeneous parameters towards local patterns in the distribution. 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.

Outline: The rest of the paper is organized 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. Section 5 describes the empirical Bayes estimator. Section 6 presents Monte Carlo experiments, Section 7 reports our empirical applications, and Section 8 concludes. Technical details are provided in appendices at the end of the paper and in 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} (including, possibly, factors), $\boldsymbol{\beta}_i$ 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. *$\{\varepsilon_{it}\}$ for $i = 1, 2, \ldots, N$ are martingale difference processes with respect to the* *filtration,* $\mathcal{I}_{it} = (\boldsymbol{w}_{it}, \boldsymbol{w}_{i,t-1}, \ldots)$, *so that:*

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

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

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

(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* $i = 1, 2, \ldots, N$, *satisfy the conditions* $\sup_i \lambda_{\max}\left(\boldsymbol{Q}_{iT}\right) < C < \infty$, *and* $\sup_i \lambda_{\max}\left(\boldsymbol{Q}_{iT}^{-1}\right) < C < \infty$.

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

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

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

Under Assumption 1 the optimal forecast of $y_{i,T+1}$, in a mean squared error sense, is given by $\mathrm{E}\left(y_{i,T+1} \left| \boldsymbol{w}_{i,T+1}, \boldsymbol{W}_i \right.\right) = \boldsymbol{\theta}_i'\boldsymbol{w}_{i.T+1}$. 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 $\boldsymbol{w}_{i,T+1}$. Part (a) of Assumption 3 is standard in the forecasting literature and requires the regressors to be stationary. Part (b) is an identification assumption that allows estimation of individual slope coefficients, $\boldsymbol{\theta}_i$, by least squares. 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.

Under weakly exogenous regressors the least squares estimator has a small T bias, and $\mathrm{E}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right) = O\left(T^{-1}\right)$. Under strictly exogenous regressors, in contrast, $\mathrm{E}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right) = \boldsymbol{0}$. We also note that, under Assumption 3 and 4, $\|\boldsymbol{Q}_{iT} - \boldsymbol{Q}_i\| = O_p(T^{-1/2})$, and $\left\|\boldsymbol{Q}_{iT}^{-1} - \boldsymbol{Q}_i^{-1}\right\| = O_p(T^{-1/2})$. These results, which hold for each i, are used in the implementation of our combination forecasts below. For proof of consistency of the weights in the combined forecasts discussed in Section 4 below, we need the stronger conditions

$$\sup_i \|\boldsymbol{Q}_{iT} - \boldsymbol{Q}_i\| = O_p\left(\frac{\ln(N)}{\sqrt{T}}\right), \text{ and } \sup_i \left\|\boldsymbol{Q}_{iT}^{-1} - \boldsymbol{Q}_i^{-1}\right\| = O_p\left(\frac{\ln(N)}{\sqrt{T}}\right), \tag{6}$$

still allowing N to rise much faster than T.[2]

Finally, let $\boldsymbol{g}_{it} = \boldsymbol{w}_{it}\varepsilon_{it}$, and note that $T^{-1/2}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i = T^{-1/2}\sum_{t=1}^T \boldsymbol{g}_{it}$. Also under Assumption 2 $\boldsymbol{g}_{it}$ is a martingale difference process with respect to $\mathcal{I}_{it} = (\boldsymbol{w}_{it}, \boldsymbol{w}_{i,t-1}, \ldots)$, and we have $\mathrm{E}\left(\boldsymbol{g}_{it}\right) = \boldsymbol{0}$,

$$\mathrm{Var}\left(T^{-1/2}\sum_{t=1}^T \boldsymbol{g}_{it}\right) = T^{-1}\sum_{t=1}^T \mathrm{E}\left(\boldsymbol{g}_{it}\boldsymbol{g}_{it}'\right) = T^{-1}\mathrm{E}\left(\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i\right) = T^{-1}\sum_{t=1}^T \sigma_i^2\mathrm{E}\left(\boldsymbol{w}_{it}\boldsymbol{w}_{it}'\right).$$

Further, under Assumption 3, $\mathrm{E}\left(\boldsymbol{w}_{it}\boldsymbol{w}_{it}'\right) = \boldsymbol{Q}_i$, and it follows that

$$\mathrm{E}\left(T^{-1}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i\right) = \sigma_i^2\boldsymbol{Q}_i. \tag{7}$$

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* $\|\boldsymbol{\theta}\| < C$, $\mathrm{E}\|\boldsymbol{\eta}_i\| < C$, $\mathrm{E}\left(\boldsymbol{\eta}_i\right) = 0$, $\mathrm{E}\left(\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\right) = \boldsymbol{\Omega}_\eta$, *and* $\|\boldsymbol{\Omega}_\eta\| < C$. *(b) Let* $\boldsymbol{q}_{it} = \boldsymbol{w}_{it}\boldsymbol{w}_{it}'\boldsymbol{\eta}_i$, *then* $\mathrm{E}\left(\boldsymbol{q}_{it}\right) = \boldsymbol{q}_i$ *(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. $\bar{\boldsymbol{\xi}}_{NT} = N^{-1}\sum_{i=1}^N \boldsymbol{\xi}_{iT} = O_p\left(N^{-1/2}T^{-1/2}\right)$, *where* $\boldsymbol{\xi}_{iT} = T^{-1}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i = T^{-1}\sum_{t=1}^T \boldsymbol{w}_{it}\varepsilon_{it} = O_p(T^{-1/2})$.

Assumption 8. *There exists a fixed* T_0 *such that for all* $T > T_0$ *and* $N = 1, 2, \ldots,$ *the pooled*

[2]As noted by Fan et al. (2015, Section 3.1), this stronger condition is typically satisfied for strictly stationary data that satisfy strong mixing conditions.

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* $\boldsymbol{Q}_i = \mathrm{E}\left(\boldsymbol{Q}_{iT}\right)$*,*

$$\bar{\boldsymbol{Q}}_{NT} = N^{-1}\sum_{i=1}^{N}\boldsymbol{Q}_{iT}, \text{ and } \bar{\boldsymbol{Q}}_N = \mathrm{E}\left(\bar{\boldsymbol{Q}}_{NT}\right) = N^{-1}\sum_{i=1}^{N}\boldsymbol{Q}_i, \tag{8}$$

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

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

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.[3] 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 $\boldsymbol{q}_i$.[4]

Assumptions 5–8 are not required for forecasts based on the individual estimates and the associated MSFE. Assumption 9 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.[5]

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,

[3] We simplify the notations and use $\boldsymbol{\theta}$, rather than $\boldsymbol{\theta}_0$, to denote the population mean which is technically more appropriate.

[4] 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. Under Assumption 3, $\left\|\bar{\boldsymbol{Q}}_{NT}\right\| \leq \sup_i\left\|\boldsymbol{Q}_{iT}\right\| < C$, and $\left\|\bar{\boldsymbol{Q}}_N\right\| \leq \sup_i\left\|\boldsymbol{Q}_i\right\| < C$.

[5] Cross-sectional dependence in forecast errors can be exploited by using interactive time effects (latent factors) or spatial (network) effects, see, e.g., Chudik et al. 2016.

on average, by

$$\bar{\boldsymbol{q}}_{NT} = N^{-1}T^{-1}\sum_{i=1}^{N}\boldsymbol{W}_i'\boldsymbol{W}_i\boldsymbol{\eta}_i = N^{-1}T^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{w}_{it}\boldsymbol{w}_{it}'\boldsymbol{\eta}_i. \tag{9}$$

Taking expectations,

$$\mathrm{E}\left(\bar{\boldsymbol{q}}_{NT}\right) = \bar{\boldsymbol{q}}_N = N^{-1}\sum_{i=1}^{N}\boldsymbol{q}_i. \tag{10}$$

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 $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\mathrm{E}\left(\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{X}_i\boldsymbol{\eta}_{i\beta}\right) = \boldsymbol{0}$, 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.[6]

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.

Note that our theoretical framework can be equally 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 but lies beyond the scope of the present paper.

3.1 Forecasts based on individual and pooled estimation

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

[6]See Pesaran and Yang (2024b). Note that $\mathrm{E}\left(\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{X}_i\boldsymbol{\eta}_{i\beta}\right) = \boldsymbol{0}$ is sufficient but not necessary for the validity of fixed effects estimation. This condition is not met if $\boldsymbol{x}_{it}$ includes lagged values of y_{it}, even if $T\to\infty$.

conditional nature of the forecasting exercise, we assume that $\sup_{i,T} \|\boldsymbol{w}_{i,T+1}\| < C$.[7] Forecasts based on individual estimators take the form

$$\hat{y}_{i,T+1} = \hat{\boldsymbol{\theta}}_i' \boldsymbol{w}_{i,T+1},\ i = 1, 2, \ldots, N, \tag{11}$$

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

$$\tilde{y}_{i,T+1} = \tilde{\boldsymbol{\theta}}' \boldsymbol{w}_{i,T+1},\ i = 1, 2, \ldots, N, \tag{12}$$

where $\tilde{\boldsymbol{\theta}} = (\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{y}$. Using (8), (9) and the definition of $\bar{\boldsymbol{\xi}}_{NT}$ in Assumption 7,

$$\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_i = -\boldsymbol{\eta}_i + \bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT} + \bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{\xi}}_{NT}. \tag{13}$$

Forecast errors from these schemes take the form

$$\hat{e}_{i,T+1} = y_{iT+1} - \hat{y}_{i,T+1} = \varepsilon_{i,T+1} - (\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)'\boldsymbol{w}_{i,T+1}, \tag{14}$$

$$\tilde{e}_{i,T+1} = y_{iT+1} - \tilde{y}_{i,T+1} = \varepsilon_{i,T+1} - (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_i)'\boldsymbol{w}_{i,T+1}. \tag{15}$$

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 generate the following average MSFE:

$$N^{-1}\sum_{i=1}^{N} \hat{e}_{i,T+1}^2 = N^{-1}\sum_{i=1}^{N} \varepsilon_{i,T+1}^2 + T^{-1}S_{NT} - 2R_{NT}, \tag{16}$$

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

$$r_{iT} = \left(\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i(\boldsymbol{W}_i'\boldsymbol{W}_i)^{-1}\boldsymbol{w}_{i,T+1}\right)\varepsilon_{i,T+1}, \tag{17}$$

$$s_{iT} = \boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_{iT}^{-1}\left(T^{-1}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i\right)\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}. \tag{18}$$

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

Under Assumptions 1 and 3, $\mathrm{E}(r_{iT}) = 0$ and $\sup_{i,T} \mathrm{E}|r_{iT}| < C$, and under cross-sectional independence (Assumption 9) we have $R_{NT} = O_p(N^{-1/2})$. Similarly, $\sup_{i,T} \mathrm{E}|s_{iT}| < C$,

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

$S_{NT} = \mathrm{E}(S_{NT}) + O_p(N^{-1/2})$, 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 $T > T_0$, the average MSFE resulting from individual-specific estimation of the parameters, given by (16), has the following representation*

$$N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}^2 = N^{-1}\sum_{i=1}^{N}\varepsilon_{i,T+1}^2 + T^{-1}h_{NT} + O_p(N^{-1/2}), \tag{19}$$

where

$$h_{NT} = N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[\boldsymbol{w}'_{i,T+1}\boldsymbol{Q}_{iT}^{-1}\left(\frac{\boldsymbol{W}'_i\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_i\boldsymbol{W}_i}{T}\right)\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}\right], \tag{20}$$

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

(b) If $\boldsymbol{W}_i$ is strictly exogenous, h_{NT} simplifies to $h_{NT} = N^{-1}\sum_{i=1}^{N}\sigma_i^2\mathrm{E}\left(\boldsymbol{w}'_{i,T+1}\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}\right)$.

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 (19) can therefore be important. Parameter heterogeneity, in contrast, does not affect the accuracy of the forecast in (19). The magnitude of h_{NT} plays an important role in the comparisons of forecasts based on individual and pooled estimates and depends on how far the predictors are from their mean. For example, when $\boldsymbol{w}_{it} = (1, x_{it})'$ and x_{it} is strictly exogenous,

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

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 $\bar{x}_{iT} = T^{-1}\sum_{t=1}^{T}x_{it}$. Hence, h_{NT} is minimized when $x_{i,T+1} = \bar{x}_{iT}$, for all i. When $x_{i,T+1} \neq \bar{x}_{iT}$ for most i, T must be sufficiently large such that $sup_i \mathrm{E}\Big[(x_{i,T+1}-\bar{x}_{iT})^2/s_{iT}^2\Big] < C$.

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 (13) in (15) 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 (8) and (9), and $\bar{\boldsymbol{\xi}}_{NT}$ is defined under Assumption 7. After some algebra, and averaging over i, we have

$$\begin{aligned} N^{-1}\sum_{i=1}^{N} \tilde{e}_{i,T+1}^2 &= N^{-1}\sum_{i=1}^{N} \varepsilon_{i,T+1}^2 + N^{-1}\sum_{i=1}^{N} \boldsymbol{w}_{i,T+1}' \boldsymbol{\eta}_i \boldsymbol{\eta}_i' \boldsymbol{w}_{i,T+1} \\ &\quad + \tilde{S}_{N,T+1} + 2\tilde{R}_{N,T+1}, \end{aligned} \tag{21}$$

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 $\tilde{S}_{N,T+1} = -\bar{\boldsymbol{q}}_N' \bar{\boldsymbol{Q}}_N^{-1} \bar{\boldsymbol{q}}_N + O_p\left(N^{-1/2}\right)$. 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, given by (21), is*

$$N^{-1}\sum_{i=1}^{N} \tilde{e}_{i,T+1}^2 = N^{-1}\sum_{i=1}^{N} \varepsilon_{i,T+1}^2 + \Delta_{NT} + O_p(N^{-1/2}), \tag{22}$$

where

$$\Delta_{NT} = N^{-1}\sum_{i=1}^{N} \mathrm{E}\left(\boldsymbol{w}_{i,T+1}' \boldsymbol{\eta}_i \boldsymbol{\eta}_i' \boldsymbol{w}_{i,T+1}\right) - \bar{\boldsymbol{q}}_N' \bar{\boldsymbol{Q}}_N^{-1} \bar{\boldsymbol{q}}_N. \tag{23}$$

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

Note that the impact on the MSFE from neglected heterogeneity, Δ_{NT}, does not vanish even if both N and $T \to \infty$, which is similar to the finding by Pesaran and Smith (1995) for heterogeneous dynamic panels since heterogeneity is always correlated in dynamic panels.[8]

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

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

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

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

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:

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

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

$$\frac{N^{-1}\sum_{i=1}^{N} \tilde{e}_{i,T+1}^2 - N^{-1}\sum_{i=1}^{N} \hat{e}_{i,T+1}^2}{N^{-1}\sum_{i=1}^{N} \hat{e}_{i,T+1}^2} \xrightarrow{p} \frac{\Delta - T^{-1} h_T}{\bar{\sigma}^2 + T^{-1} h_T}, \tag{24}$$

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, the ranking of the two forecasting schemes will depend on the sign and magnitude of $\Delta - T^{-1} h_T$.[9]

[8] This latter property is illustrated by a simple panel AR(1) model with heterogeneous AR coefficients in Section A.4 of the Appendix. See also Pesaran and Yang (2024a) where estimation of such models with short T panels is considered.

[9] 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.

For large values of T, however, 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, vanishing estimation uncertainty implied by large T means that, on average, individual forecasts are at least as precise as pooled forecasts irrespective of N.

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

$$\hat{y}_{i,T+1}^{\mathrm{FE}} = \hat{\alpha}_{i,\mathrm{FE}} + \hat{\boldsymbol{\beta}}_{\mathrm{FE}}'\boldsymbol{x}_{i,T+1}, \tag{25}$$

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

$$\hat{e}_{i,T+1}^{\mathrm{FE}} = \bar{\bar{\varepsilon}}_{i,T+1} - (\hat{\boldsymbol{\beta}}_{\mathrm{FE}} - \boldsymbol{\beta}_i)'\bar{\bar{\boldsymbol{x}}}_{i,T+1}, \tag{26}$$

where $\bar{\bar{\varepsilon}}_{i,T+1} = \varepsilon_{i,T+1} - \bar{\varepsilon}_{iT}$, $\bar{\bar{\boldsymbol{x}}}_{i,T+1} = \boldsymbol{x}_{i,T+1} - \bar{\boldsymbol{x}}_{iT}$, $\bar{\varepsilon}_{iT} = T^{-1}\sum_{t=1}^{T}\varepsilon_{it}$, and $\bar{\boldsymbol{x}}_{iT} = T^{-1}\sum_{t=1}^{T}\boldsymbol{x}_{it}$. Section S.5 in the Online Supplement provides details of the derivation of the MSFE under fixed effects estimation:

$$N^{-1}\sum_{i=1}^{N}\left(\hat{e}_{i,T+1}^{\mathrm{FE}}\right)^2 = N^{-1}\sum_{i=1}^{N}\bar{\bar{\varepsilon}}_{i,T+1}^{2} + \Delta_{NT}^{\mathrm{FE}} - 2c_{NT}^{\mathrm{FE}} + O_p(N^{-1/2}), \tag{27}$$

where

$$\Delta_{NT}^{\mathrm{FE}} = N^{-1}\sum_{i=1}^{N}\mathrm{E}(\bar{\bar{\boldsymbol{x}}}'_{i,T+1}\boldsymbol{\eta}_{i,\beta}\boldsymbol{\eta}'_{i,\beta}\bar{\bar{\boldsymbol{x}}}_{i,T+1}) - \bar{\boldsymbol{q}}'_{N,\beta}\bar{\boldsymbol{Q}}_{N,\beta}^{-1}\bar{\boldsymbol{q}}_{N,\beta}, \tag{28}$$

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

$$c_{NT}^{\mathrm{FE}} = -N^{-1}\sum_{i=1}^{N}\mathrm{E}\left(\boldsymbol{\eta}'_{i,\beta}\bar{\bar{\boldsymbol{x}}}_{i,T+1}\bar{\varepsilon}_{iT}\right) + \bar{\boldsymbol{q}}'_{N,\beta}\bar{\boldsymbol{Q}}_{N,\beta}^{-1}\left[N^{-1}\sum_{i=1}^{N}\mathrm{E}\left(\bar{\boldsymbol{x}}_{iT}\bar{\varepsilon}_{iT}\right)\right]. \tag{29}$$

c_{NT}^{FE} tends to zero for T sufficiently large or if $\boldsymbol{x}_{it}$ is strictly exogenous.

Similar to the case of the individual and pooled forecasts, for T finite and N large, the ranking of the individual and FE forecasts will depend on the relative magnitudes estimation error and parameter heterogeneity. Precise expressions can be found in the Online Supplement. 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 (11) and (12), 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, we combine the two forecasts $\hat{y}_{i,T+1}$ and $\tilde{y}_{i,T+1}$ using common weights, ω, to obtain[10]

$$y^*_{i,T+1}(\omega) = \omega \hat{y}_{i,T+1} + (1-\omega)\tilde{y}_{i,T+1}, \tag{30}$$

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 forecast is given by

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

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

$$\omega^*_{NT} = \frac{N^{-1}\sum_{i=1}^{N}\tilde{e}^2_{i,T+1} - \left(N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}\tilde{e}_{i,T+1}\right)}{\left(N^{-1}\sum_{i=1}^{N}\hat{e}^2_{i,T+1}\right) + \left(N^{-1}\sum_{i=1}^{N}\tilde{e}^2_{i,T+1}\right) - 2\left(N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}\tilde{e}_{i,T+1}\right)}. \tag{31}$$

Expressions for $N^{-1}\sum_{i=1}^{N}\hat{e}^2_{i,T+1}$ and $N^{-1}\sum_{i=1}^{N}\tilde{e}^2_{i,T+1}$ are given by (19) and (22), respectively. We obtain a similar expression for $N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}\tilde{e}_{i,T+1}$, with $N^{-1}\sum_{i=1}^{N}\varepsilon^2_{i,T+1}$ cancelling out from ω^*_{NT}. The result is summarized in the following proposition (proven in Appendix Section A.2.3):

Proposition 3. *(a) Under Assumptions 1–9, and for a given value of* $\mathbf{w}_{i,T+1}$*, the optimal combination weight that minimizes the MSFE of the forecast combination in (30) is given by*

$$\omega^*_{NT} = \frac{\Delta_{NT} - T^{-1}\psi_{NT}}{\Delta_{NT} + T^{-1}h_{NT} - 2T^{-1}\psi_{NT}} + O_p(N^{-1/2}), \tag{32}$$

[10] We focus here on a simple constant-coefficient linear combination scheme. Lahiri et al. (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.

where

$$h_{NT} = N^{-1} \sum_{i=1}^{N} \mathrm{E}\left[\boldsymbol{w}_{i,T+1}' \boldsymbol{Q}_{iT}^{-1} \left(\frac{\boldsymbol{W}_i' \boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i' \boldsymbol{W}_i}{T} \right) \boldsymbol{Q}_{iT}^{-1} \boldsymbol{w}_{i,T+1} \right] > 0, \tag{33}$$

$$\Delta_{NT} = N^{-1} \sum_{i=1}^{N} \mathrm{E}\left(\boldsymbol{w}_{i,T+1}' \boldsymbol{\eta}_i \boldsymbol{\eta}_i' \boldsymbol{w}_{i,T+1} \right) - \bar{\boldsymbol{q}}_N' \bar{\boldsymbol{Q}}_N^{-1} \bar{\boldsymbol{q}}_N > 0, \tag{34}$$

and

$$\begin{aligned} \psi_{NT} = {} & TN^{-1} \sum_{i=1}^{N} \mathrm{E}\left[\boldsymbol{\varepsilon}_i' \boldsymbol{W}_i (\boldsymbol{W}_i' \boldsymbol{W}_i)^{-1} \boldsymbol{w}_{i,T+1} \boldsymbol{w}_{i,T+1}' \right] \bar{\boldsymbol{Q}}_N^{-1} \bar{\boldsymbol{q}}_N \\ & - TN^{-1} \sum_{i=1}^{N} \mathrm{E}\left[\boldsymbol{\varepsilon}_i' \boldsymbol{W}_i (\boldsymbol{W}_i' \boldsymbol{W}_i)^{-1} \boldsymbol{w}_{i,T+1} \boldsymbol{w}_{i,T+1}' \boldsymbol{\eta}_i \right]. \end{aligned} \tag{35}$$

(b) Under strict exogeneity, irrespective of whether heterogeneity is correlated, we have $\psi_{NT} = 0$, $h_{NT} = N^{-1} \sum_{i=1}^{N} \sigma_i^2 \mathrm{E}\left(\boldsymbol{w}_{i,T+1}' \boldsymbol{Q}_{iT}^{-1} \boldsymbol{w}_{i,T+1} \right)$*, and* $\Delta_{NT} = N^{-1} \sum_{i=1}^{N} \mathrm{E}\left(\boldsymbol{w}_{i,T+1}' \boldsymbol{\Omega}_\eta \boldsymbol{w}_{i,T+1} \right)$.

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.

Forecast combinations with individual weights

Pesaran et al. (2022) show that, under strict exogeneity of the regressors and uncorrelated heterogeneity, optimal weights can be obtained that are specific to the individual unit. The combination of individual and pooled forecast is then

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

where the optimal value of ω_i is given by

$$\omega_i^* = \frac{\boldsymbol{w}_{i,T+1}' \boldsymbol{\Omega}_\eta \boldsymbol{w}_{i,T+1}}{\boldsymbol{w}_{i,T+1}' (T^{-1} \sigma_i^2 \boldsymbol{Q}_{iT}^{-1} + \boldsymbol{\Omega}_\eta) \boldsymbol{w}_{i,T+1}} \tag{36}$$

The weights again depend on the variances and covariances of the underlying forecast errors. Related to this, Giacomini et al. (2023) develop a random effects approach for linear panels that similarly combines univariate and pooled forecasts in a way that minimizes minimax-regret and MSFE.

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

$$y^*_{i,T+1}(\omega_{\text{FE}}) = \omega_{\text{FE}}\hat{y}_{i,T+1} + (1-\omega_{\text{FE}})\hat{y}_{i,T+1,\text{FE}}, \tag{37}$$

yielding the optimal pooled weight

$$\omega^*_{\text{FE},NT} = \frac{N^{-1}\sum_{i=1}^N\left(\hat{e}^{\text{FE}}_{i,T+1}\right)^2 - \left(N^{-1}\sum_{i=1}^N \hat{e}^{\text{FE}}_{i,T+1}\hat{e}_{i,T+1}\right)}{\left(N^{-1}\sum_{i=1}^N \hat{e}^2_{i,T+1}\right) + N^{-1}\sum_{i=1}^N\left(\hat{e}^{\text{FE}}_{i,T+1}\right)^2 - 2\left(N^{-1}\sum_{i=1}^N \hat{e}^{\text{FE}}_{i,T+1}\hat{e}_{i,T+1}\right)}. \tag{38}$$

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 (27) and (19), 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 (37) is given by*

$$\omega^*_{FE,NT} = \frac{\Delta^{FE}_{NT} - T^{-1}\psi^{FE}_{NT} - \left(c^{FE}_{NT} - c_{NT,\beta}\right)}{\Delta^{FE}_{NT} + T^{-1}h_{NT,\beta} - 2T^{-1}\psi^{FE}_{NT}} + O_p(N^{-1/2}), \tag{39}$$

where Δ^{FE}_{NT} and c^{FE}_{NT} are defined in (28) and (29), respectively,

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

$$\begin{aligned}\psi^{FE}_{NT} \;=\; & TN^{-1}\sum_{i=1}^N \mathrm{E}\left[\left(\hat{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta}_i\right)'\bar{\bar{\boldsymbol{x}}}_{i,T+1}\bar{\bar{\boldsymbol{x}}}'_{i,T+1}\right]\bar{\boldsymbol{Q}}^{-1}_{N,\beta}\bar{\boldsymbol{q}}_{N,\beta}\\ & -TN^{-1}\sum_{i=1}^N \mathrm{E}\left[\left(\hat{\boldsymbol{\beta}}_{FE}-\boldsymbol{\beta}_i\right)'\bar{\bar{\boldsymbol{x}}}_{i,T+1}\bar{\bar{\boldsymbol{x}}}'_{i,T+1}\boldsymbol{\eta}_{i,\beta}\right],\end{aligned}$$

and

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

(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

Estimates of the weights for the forecast combination in Proposition 3 require estimates of Δ_{NT}, h_{NT}, and ψ_{NT}. Under Assumption 9, these terms can be estimated by their sample means with unknown parameters replaced by their estimates. We summarize the estimators here with details in Appendix A.3. Using (34) and (33), the estimators of Δ_{NT} and h_{NT} are given by

$$\hat{\Delta}_{NT} = N^{-1}\sum_{i=1}^{N}\boldsymbol{w}'_{i,T+1}\tilde{\boldsymbol{\eta}}_i\tilde{\boldsymbol{\eta}}'_i\boldsymbol{w}_{i,T+1}, \tag{40}$$

where $\tilde{\boldsymbol{\eta}}_i = \tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_i$, and

$$\hat{h}_{NT} = N^{-1}\sum_{i=1}^{N}\boldsymbol{w}'_{i,T+1}\boldsymbol{Q}_{iT}^{-1}\hat{\boldsymbol{H}}_{iT}\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}, \tag{41}$$

where $\hat{\boldsymbol{H}}_{iT} = \hat{\sigma}_i^2 T^{-1}\sum_{t=1}^{T}\boldsymbol{w}_{it}\boldsymbol{w}'_{it}$, $\hat{\sigma}_i^2 = \sum_{t=1}^{T}\hat{\varepsilon}_{it}^2/(T-K)$, and $\hat{\varepsilon}_{it} = y_{it} - \hat{\boldsymbol{\theta}}'_i\boldsymbol{w}_{it}$. We show in Appendix A.3 that

$$\begin{aligned}\hat{\Delta}_{NT} - \Delta_{NT} &= O_p\left(N^{-1/2}\right) + O_p(T^{-1}),\\ \hat{h}_{NT} - h_{NT} &= O_p(N^{-1/2}) + O_p\left(\frac{\ln(N)}{\sqrt{T}}\right).\end{aligned}$$

In the case of strictly exogenous regressors, $\hat{\Delta}_{NT}$ is a consistent estimator of Δ_{NT} for fixed T as $N \to \infty$.

Consider now ψ_{NT}, given by (35), and recall that $\psi_{NT} = 0$ under uncorrelated heterogeneity. To estimate ψ_{NT} under correlated heterogeneity, we first note that the approach of replacing expectations by sample moments and then estimating $\boldsymbol{\varepsilon}_i$ from the OLS residuals, $\hat{\boldsymbol{\varepsilon}}_i = \left(\mathbf{y}_i - \mathbf{W}_i\hat{\boldsymbol{\theta}}_i\right)$ will not work in the case of ψ_{NT}, since $\mathbf{W}'_i\hat{\varepsilon}_i = \mathbf{W}'_i\left(\mathbf{y}_i - \mathbf{W}_i\hat{\boldsymbol{\theta}}_i\right) = \mathbf{0}$ for all i. If used in (35), this results in $\hat{\psi}_{NT} = 0$, which is not a consistent estimator of ψ_{NT} under correlated heterogeneity. To overcome this problem, we replace $\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i(\boldsymbol{W}_i'\boldsymbol{W}_i)^{-1}$ by $\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)'$ and note that ψ_{NT} can be written equivalently as (noting that $\boldsymbol{w}_{i,T+1}$ for $i = 1, 2, \ldots, N$ are given)

$$\psi_{NT} = N^{-1}\sum_{i=1}^{N} \mathrm{E}\left[T\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)'\right] \mathrm{E}\left(\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\right) \bar{\boldsymbol{Q}}_N^{-1}\bar{\boldsymbol{q}}_N \qquad (42)$$
$$- N^{-1}\sum_{i=1}^{N} \mathrm{E}\left[T\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)'\right] \mathrm{E}\left(\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\right).$$

We now employ a half-jackknife estimator of $\boldsymbol{\theta}_i$ (Dhaene and Jochmans, 2015, Chudik et al., 2018) to estimate $T\mathrm{E}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)$, which is the small sample bias of $\hat{\boldsymbol{\theta}}_i$ in the case of weakly exogenous regressors. The half-jackknife estimator of $\boldsymbol{\theta}_i$ is defined by $\hat{\boldsymbol{\theta}}_{i,JK} = 2\hat{\boldsymbol{\theta}}_i - \frac{1}{2}\left(\hat{\boldsymbol{\theta}}_{ia} + \hat{\boldsymbol{\theta}}_{ib}\right)$, where $\hat{\boldsymbol{\theta}}_{ia}$ and $\hat{\boldsymbol{\theta}}_{ib}$ are least squares estimators of $\boldsymbol{\theta}_i$ based on two equal halves of the sample of size $T_h = T/2$ (omitting an observation in the case of uneven T), namely $\hat{\boldsymbol{\theta}}_{ia} = \left(\sum_{t=1}^{T_h}\boldsymbol{w}_{it}\boldsymbol{w}_{it}'\right)^{-1}\sum_{t=1}^{T_h}\boldsymbol{w}_{it}y_{it}$ and $\hat{\boldsymbol{\theta}}_{ib} = \left(\sum_{t=T_h+1}^{T}\boldsymbol{w}_{it}\boldsymbol{w}_{it}'\right)^{-1}\sum_{t=T_h+1}^{T}\boldsymbol{w}_{it}y_{it}$. Then $\mathrm{E}\left[T\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)\right]$ can be estimated by $T\left[\frac{1}{2}\left(\hat{\boldsymbol{\theta}}_{ia} + \hat{\boldsymbol{\theta}}_{ib}\right) - \hat{\boldsymbol{\theta}}_i\right]$ and ψ_{NT} by

$$\hat{\psi}_{NT} = \left[TN^{-1}\sum_{i=1}^{N}\left[\frac{1}{2}\left(\hat{\boldsymbol{\theta}}_{ia} + \hat{\boldsymbol{\theta}}_{ib}\right) - \hat{\boldsymbol{\theta}}_i\right]'\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\right]\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT}\left(\hat{\boldsymbol{\eta}}\right) \qquad (43)$$
$$-TN^{-1}\sum_{i=1}^{N}\left[\frac{1}{2}\left(\hat{\boldsymbol{\theta}}_{ia} + \hat{\boldsymbol{\theta}}_{ib}\right) - \hat{\boldsymbol{\theta}}_i\right]'\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\hat{\boldsymbol{\eta}}_i,$$

where $\hat{\boldsymbol{\eta}}_i = \hat{\boldsymbol{\theta}}_i - N^{-1}\sum_{i=1}^{N}\hat{\boldsymbol{\theta}}_i$, and $\bar{\boldsymbol{q}}_{NT}\left(\hat{\boldsymbol{\eta}}\right) = (NT)^{-1}\sum_{i=1}^{N}\boldsymbol{W}_i'\boldsymbol{W}_i\hat{\boldsymbol{\eta}}_i$. Consistency of $\hat{\psi}_{NT}$ as an estimator of ψ_{NT} is established as $N, T \to \infty$, since $T\mathrm{E}\left(\hat{\boldsymbol{\theta}}_{i,JK} - \boldsymbol{\theta}_i\right) = O\left(T^{-1}\right)$.[11] Thus, to use the half-jackknife method for models with weakly exogenous regressors we need T large, although it is not required that $\sqrt{T}/N$ tends to zero, as it is in the case of large N and T asymptotics.

The components of the weights in Proposition 4 that combine individual and fixed effects forecasts can be estimated in a similar fashion, with details provided in Appendix A.3.

4.4 Empirical Bayes Forecasts

Bayesian panel forecasts are becoming increasingly common in empirical applications and constitute an alternative approach to the frequentist forecasts discussed so far. Due to their resemblance to our forecast combination schemes and their recent popularity (e.g., Armstrong et al., 2022, and Efron,

[11]Note that $\hat{\psi}_{NT} - \psi_{NT}$ depends on $\mathrm{E}\left[\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right) - \left(\frac{1}{2}\left(\hat{\boldsymbol{\theta}}_{ia} + \hat{\boldsymbol{\theta}}_{ib}\right) - \hat{\boldsymbol{\theta}}_i\right)\right] = \mathrm{E}\left(\hat{\boldsymbol{\theta}}_{i,JK} - \boldsymbol{\theta}_i\right)$.

2016), we focus on empirical Bayes (EB) methods. The EB forecast uses the estimator of Hsiao et al. (1999) and takes the form $\hat{y}^{EB}_{i,T+1} = \hat{\boldsymbol{\theta}}'_{i,EB}\boldsymbol{w}_{i,T+1}$, where

$$\hat{\boldsymbol{\theta}}_{i,EB} = (\hat{\sigma}_i^{-2}\boldsymbol{W}_i'\boldsymbol{W}_i + \hat{\boldsymbol{\Omega}}_\eta^{-1})^{-1}(\hat{\sigma}_i^{-2}\boldsymbol{W}_i'\boldsymbol{y}_i + \hat{\boldsymbol{\Omega}}_\eta^{-1}\,\bar{\hat{\boldsymbol{\theta}}}), \tag{44}$$

$\bar{\hat{\boldsymbol{\theta}}} = N^{-1}\sum_{i=1}^N \hat{\boldsymbol{\theta}}_i$, $\quad \hat{\sigma}_i^2 = (T-K)^{-1}\hat{\varepsilon}_i'\hat{\varepsilon}_i$, and $\hat{\boldsymbol{\Omega}}_\eta = \frac{1}{N}\sum_{i=1}^N(\hat{\boldsymbol{\theta}}_i - \bar{\hat{\boldsymbol{\theta}}})(\hat{\boldsymbol{\theta}}_i - \bar{\hat{\boldsymbol{\theta}}})'$, where $\hat{\boldsymbol{\varepsilon}}_i = \boldsymbol{y}_i - \boldsymbol{W}_i\hat{\boldsymbol{\theta}}_i$, and $\hat{\boldsymbol{\theta}}_i = (\boldsymbol{W}_i'\boldsymbol{W}_i)^{-1}\boldsymbol{W}_i'\boldsymbol{y}_i$.[12] $\hat{\boldsymbol{\theta}}_{i,EB}$ can also be written as a weighted average of $\hat{\boldsymbol{\theta}}_i$, that allows for full heterogeneity, and the mean group estimator, $\bar{\hat{\boldsymbol{\theta}}}$, namely $\hat{\boldsymbol{\theta}}_{i,EB} = \boldsymbol{\mathcal{W}}_{iT}\hat{\boldsymbol{\theta}}_i + (\boldsymbol{I}_k - \boldsymbol{\mathcal{W}}_{iT})\,\bar{\hat{\boldsymbol{\theta}}}$, with the weight matrix $\boldsymbol{\mathcal{W}}_{iT}$ given by

$$\boldsymbol{\mathcal{W}}_{iT} = \left(\boldsymbol{I}_k + T^{-1}\hat{\sigma}_i^2\boldsymbol{Q}_{iT}^{-1}\,\hat{\boldsymbol{\Omega}}_\eta^{-1}\right)^{-1}, \tag{45}$$

recalling that $\boldsymbol{Q}_{iT} = T^{-1}\boldsymbol{W}_i'\boldsymbol{W}_i$ is invertible under Assumption 3. The weights on the heterogeneous estimates are larger, the greater the degree of heterogeneity, as measured by the norm of $\hat{\boldsymbol{\Omega}}_\eta$, with $\hat{\boldsymbol{\theta}}_{i,EB} \to \hat{\boldsymbol{\theta}}_i$ as $\left\|\hat{\boldsymbol{\Omega}}_\eta\right\| \to \infty$. Also, since $\hat{\sigma}_i^2\boldsymbol{Q}_{iT}^{-1}\,\hat{\boldsymbol{\Omega}}_\eta^{-1}$ 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 even when T is relatively small and the degree of heterogeneity is not too large. For large T, EB and individual forecasts coincide and both methods will work well.

The EB weights do *not* depend on $\boldsymbol{w}_{i,T+1}$ and are derived assuming uncorrelated heterogeneity and strictly exogenous regressors. They have the desirable feature of placing more weights on individual estimates if they are precisely estimated relative to the degree of parameter heterogeneity measured by $\hat{\boldsymbol{\Omega}}_\eta$. The individual optimum weights in (36) fall somewhere between the common optimal weights and the EB weights.[13] Like the EB weights, consistent estimation of individual weights require strict exogeneity and uncorrelated heterogeneity.

The EB weights are comparable to the unit-specific weights given by (36) and the two sets of weights coincide only when $\mathbf{w}_{i,T+1}$ is an scalar. To see this note that the estimates of the unit

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

[13] While the EB estimator in (44) 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 et al. (2023).

specific weights can be written as

$$\hat{\omega}_{iT}^{*} = \left[1 + \hat{\sigma}_i^2 T^{-1} \frac{\mathbf{w}_{i,T+1}' \mathbf{Q}_{iT}^{-1} \mathbf{w}_{i,T+1}}{\mathbf{w}_{i,T+1}' \hat{\boldsymbol{\Omega}}_{\eta} \mathbf{w}_{i,T+1}}\right]^{-1}, \tag{46}$$

and reduces to the EB weights only when $K = 1$ and $\hat{\omega}_{iT}^{*}$ no longer depend on $\mathbf{w}_{i,T+1}$. But in general the estimates of the unit-specific weights differ from the EB weights.

An alternative to the EB forecast is a hierarchical Bayesian approach as proposed by Lindley and Smith (1972). The full Bayesian treatment would require choices of the priors of each component, including the parameter covariance matrix. In the Online Supplement, we provide Monte Carlo results that shows that the resulting forecast performance is highly sensitive to the choice of prior.

5 Monte Carlo experiments

We examine the finite-sample performance of the panel forecasting schemes in the context of a dynamic heterogeneous panel data model using Monte Carlo experiments.[14] We allow for dynamics, parameter heterogeneity, and correlations between the regressors and coefficients. The forecasting methods are: (1) individual estimation which serves as the benchmark against which other methods are compared, (2) pooled estimation, (3) random effects, (4) fixed effects, (5) combination of individual and pooled forecasts using the weights in (32), (6) combination of individual and FE forecasts using the weights in (39), (7) individual forecast combination weights, and (8) EB forecasts.[15]

Results do not vary greatly along the N dimension, so we focus on the case with $N = 100$ and provide results for $N = 1000$ in the Online Supplement. The T dimension of the panel is more important, so we consider three different values, $T = \{20, 50, 100\}$. The values of the parameters used in the simulations are reported in Table S.1 in Appendix S.2.

5.1 Data Generating Process

Our DGP augments a panel AR(1) model with an additional regressor,

$$y_{it} = \alpha_i + \beta_i y_{i,t-1} + \gamma_i x_{it} + \varepsilon_{it}, \tag{47}$$

[14]Further analytical results for a simple panel AR(1) model are provided in Section A.4 of the Appendix.

[15]Additional results for equal weighted combinations and oracle weights are in Section S.4 of the Online Supplement.

where $\varepsilon_{it} = \sigma_i(z_{it}^2 - 1)/\sqrt{2}$ with $z_{it} \sim iid\text{N}(0,1)$, $\sigma_i^2 \sim \text{iid}\left(1+\chi_1^2\right)/2$, and x_{it} is generated as

$$x_{it} = \mu_{xi} + \xi_{it}, \tag{48}$$

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

The coefficients of the lagged dependent variables, $y_{i,t-1}$, are generated as $\beta_i = \beta_0 + \eta_{i\beta}$, with $\eta_{i\beta} \sim \text{iidUniform}(-a_\beta/2, a_\beta/2)$ and $0 \leq a_\beta < 2(1 - |\beta_0|)$.

To allow for correlated heterogeneity, we set

$$\alpha_i = \alpha_{0i} + \phi\mu_{xi} + \sigma_\eta\eta_i, \text{ and } \gamma_i = \gamma_{0i} + \pi\mu_{xi} + \sigma_\zeta\zeta_i, \tag{49}$$

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

- $\alpha_{0i} = 2/3$ 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 pooled estimates is the correlation between $y_{i,t-1}^2$ and x_{it}^2 and the individual coefficients.

Using (48) and (49), we have

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

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

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)

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

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 $\mathrm{Corr}\left(\gamma_i, x_{it}\right) = \pi\left(\sigma_\zeta^2 + \pi^2\right)^{-1/2}$. While heterogeneity is generally correlated in AR panel models (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 $\rho_{\gamma x}$. To achieve a given level of $\mathrm{Corr}(\gamma_i, x_{it}) = \rho_{\gamma x}$, we set

$$\pi = \frac{\rho_{\gamma x}\sigma_\zeta}{\left(1 - \rho_{\gamma x}^2\right)^{1/2}}. \tag{50}$$

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

$$\phi = \frac{\rho_{\alpha x}\sigma_\eta}{(1 - \rho_{\alpha x}^2)^{1/2}}. \tag{51}$$

Defining $\sigma_\gamma^2 = \mathrm{Var}(\gamma_i) = \pi^2 + \sigma_\zeta^2$, we can use (50) 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.

We set $\xi_{i0} = 0$ and initialize y_{i0} as $y_{i0} \sim \mathrm{iidN}\left(\mu_{iy0}, \sigma_{iy0}^2\right)$ with $\mu_{iy0} = \frac{\alpha_i + \gamma_i\mu_{xi}}{1-\beta_i^2}$, $\sigma_{iy0}^2 = \frac{\gamma_i^2\sigma_{xi}^2 + \sigma_i^2}{1-\beta_i^2}$, We also experimented with initialization schemes that started the DGP on values away from the long run equilibrium, which did not change the results qualitatively.

Since the forecast combinations use $\boldsymbol{w}_{i,T+1} = (1, y_{iT}, x_{i,T})'$ as an input, in the simulations we set $\boldsymbol{w}_{i,T+1}$ as $\boldsymbol{w}_{i,T+1} = \left(1, E\left(y_{it}\right) + \kappa_i\sqrt{\mathrm{Var}(y_{it})}, \mu_{xi} + \kappa_i\sigma_{xi}\right)'$, where $E\left(y_{it}\right)$ and $\mathrm{Var}(y_{it})$ are derived by assuming y_{it} is stationary and conditional on the model's parameters.[16]

The panel forecasts are evaluated using the ratio of the average MSFE of method j (pooling, fixed effects, random effects, empirical Bayes, and the forecast combinations) measured relative to

[16]It is easily established that $\mathrm{E}\left(y_{it}\right) = \frac{\alpha_i + \beta_i\mu_{xi}}{1-\beta_i}$ and $\mathrm{Var}(y_{it}) = \frac{\sigma_i^2}{1-\beta_i^2} + \left(\frac{\gamma_i^2\sigma_{xi}^2}{1-\beta_i^2}\right)\left(1 + \frac{2\beta_i\rho_{xi}}{1-\beta_i\rho_{xi}}\right)$.

that of the reference individual forecasts

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

where b denotes the benchmark forecast which is the individual forecast. Replications are denoted by $r = 1, 2, \ldots, R$, where $R = 10,000$.

5.2 Simulation Results

Monte Carlo simulation results are reported in Table 1. Our theoretical analysis shows that the term h_{NT} that adversely affects forecasts from the individual estimates depends on the value of $\|\boldsymbol{w}_{i,T+1} - \text{E}[\boldsymbol{w}_{i,T+1}]\|$, with small values of these deviations leading to better forecasting performance for the individual estimates. To examine this effect, we present two sets of conditional forecasting performance results, namely for $\kappa_i = 0$, that is when $\boldsymbol{w}_{i,T+1}$ is set to its mean $\text{E}(\boldsymbol{w}_{it}) = (1, \text{E}(y_{it}), \mu_{xi} + \kappa_i \sigma_{xi})$ in the top panel and when $\boldsymbol{w}_{i,T+1}$ deviates from its mean by generating forecasts conditional on $\boldsymbol{w}_{i,T+1} = \left(1, \text{E}(y_{it}) + \kappa_i \sqrt{\text{Var}(y_{it})}, \mu_{xi} + \kappa_i \sigma_{xi}\right)'$ in the bottom panel. We set $\kappa_i = 1$ for $i \leq N/2$, and $\kappa_i = -1$, for $i > N/2$.

We vary the parameter that controls correlation in heterogeneity ($\rho_{\gamma x}$) across three blocks of results and examine different combinations of the two hyperparameters that determine heterogeneity, a_β and σ_α^2. Finally, we vary the time-series dimension (T) along the columns.

With little heterogeneity and a small time-series dimension, $T = 20$, consistent with Propositions 1 and 2, pooling yields an MSFE up to 25% lower than the individual forecast with the gain being largest when the predictor is far from its mean ($\kappa_i = \pm 1$). However, the advantage of the pooled forecasts over the individual forecasts vanishes quickly for the two larger values of T and turns to distinctly worse performance under larger parameter heterogeneity—particularly when the predictors are away from their mean.

The RE estimator produces the most accurate forecasts when parameter heterogeneity is limited to the intercept ($a_\beta = 0, \sigma_\alpha^2 = 0.5$) and the predictor is far from its mean. When slope coefficients are heterogeneous, this method yields quite poor forecasting performance that deteriorates with T. Similar findings hold for the forecasts based on the FE method. Forecast accuracy for both methods tends to worsen (relative to the benchmark forecasts) under correlated heterogeneity.

Table 1: Monte Carlo results

a_β	σ^2_α	Pooled			RE			FE			Empirical Bayes			Comb. (pool)			Comb. (FE)			Comb. ω_i^*		
	T	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100
Conditional on $\kappa_i = 0$																						
$\rho_{\gamma x} = 0$																						
0.0	0.5	0.864	0.985	1.010	0.911	0.985	0.996	0.923	0.987	0.997	0.935	0.989	0.997	0.913	0.982	0.995	0.955	0.992	0.998	0.947	0.994	0.999
0.5	0.5	0.860	0.981	1.006	0.953	1.004	1.005	0.978	1.009	1.007	0.944	0.992	0.998	0.911	0.981	0.995	0.979	0.999	1.000	0.939	0.994	0.999
1.0	1.0	0.819	0.964	0.994	1.089	1.096	1.061	1.167	1.119	1.068	0.937	0.990	0.998	0.895	0.975	0.992	1.022	1.006	1.000	0.900	0.986	0.998
$\rho_{\gamma x} = 0.5$																						
0.0	0.5	0.862	0.982	1.007	0.910	0.985	0.996	0.923	0.987	0.997	0.937	0.989	0.997	0.912	0.981	0.995	0.955	0.992	0.998	0.950	0.994	0.999
0.5	0.5	0.856	0.977	1.001	0.950	1.003	1.005	0.977	1.009	1.007	0.946	0.993	0.998	0.910	0.980	0.994	0.979	0.999	1.000	0.942	0.994	0.999
1.0	1.0	0.820	0.964	0.994	1.098	1.103	1.065	1.174	1.125	1.073	0.941	0.991	0.998	0.895	0.975	0.992	1.024	1.006	1.000	0.900	0.986	0.998
Conditional on $\kappa_i = \pm 1$																						
$\rho_{\gamma x} = 0$																						
0.0	0.5	0.744	0.997	1.065	0.733	0.924	0.971	0.752	0.928	0.972	0.804	0.945	0.979	0.806	0.948	0.985	0.842	0.951	0.980	0.800	0.957	0.989
0.5	0.5	0.867	1.163	1.243	0.910	1.113	1.162	0.949	1.122	1.164	0.849	0.969	0.992	0.832	0.964	0.991	0.913	0.985	0.996	0.808	0.965	0.992
1.0	1.0	1.049	1.455	1.572	1.274	1.504	1.529	1.388	1.540	1.540	0.881	0.977	0.995	0.855	0.975	0.995	0.983	0.999	1.000	0.786	0.963	0.993
$\rho_{\gamma x} = 0.5$																						
0.0	0.5	0.764	1.023	1.093	0.731	0.923	0.971	0.752	0.928	0.972	0.803	0.945	0.979	0.809	0.950	0.986	0.842	0.951	0.980	0.801	0.958	0.989
0.5	0.5	0.894	1.196	1.278	0.912	1.120	1.168	0.955	1.130	1.171	0.846	0.968	0.991	0.836	0.966	0.992	0.914	0.986	0.997	0.809	0.965	0.992
1.0	1.0	1.068	1.479	1.597	1.287	1.515	1.535	1.401	1.552	1.547	0.878	0.974	0.993	0.856	0.975	0.995	0.986	0.999	1.000	0.785	0.960	0.992

Notes: 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: 'Pooled' based on pooled estimation, 'RE' based on the random effects estimation, 'FE' based on the fixed effects estimation, 'Empirical Bayes' based on the empirical Bayes estimation, 'Comb. (pool)' refers to the combination of forecasts based on individual and pooled estimation, 'Comb. (FE)' the combination of forecasts based on individual and fixed effects estimation, and 'Comb. ω_i^*' the combination forecasts using the individual weights of Pesaran et al. (2022). The parameters a_β and σ^2_α determine the heterogeneity of the slope coefficient and the intercept. The results in the upper panel are for $\kappa_i = 0$ where $\boldsymbol{w}_{i,T+1}$ equals the expected values of the regressors. The results in the lower panel are for $\kappa_i = \pm 1$ where $\boldsymbol{w}_{i,T+1}$ equals the expected values of the regressors plus or minus one standard deviation. Results are for PR^2 of approximately 0.6 and $N = 100$. The DGP is set out in Section 5.1.

Regardless of the level of heterogeneity in parameters (whether correlated or not), the empirical Bayes forecasts perform very well particularly for the smallest sample size ($T = 20$). Unlike forecasts based on the pooled, RE or FE estimators, the empirical Bayes forecasts have the attractive feature that they never perform worse, on average, than the benchmark. These forecasts perform particularly well when the predictor is away from its mean value.

Among the three forecast combinations, the cross-sectional averaging scheme that combines the pooled and individual forecasts generally performs better than the fixed effect combination scheme and also, in some cases, improves on the EB forecasts. When $\boldsymbol{w}_{i,T+1}$ is far away from its mean, T is small, and parameter heterogeneity is high, the combination scheme with individual weights performs particularly well, including relative to the EB forecast.

6 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.

6.1 Measures of forecasting performance

Our empirical applications compute the out-of-sample MSFE as $\mathrm{MSFE}_{ij} = (T-T_1)^{-1}\sum_{t=T_1}^{T-1}(y_{i,t+1} - \hat{y}_{i,j,t+1})^2$, 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, which we set to $w = 60$ in both applications. As in the simulations, we report the ratio of the average MSFE of method j relative to the average MSFE for the benchmark forecasts (b) from the individual-specific model $\mathrm{rMSFE}_j = \left(N^{-1}\sum_{i=1}^{N}\mathrm{MSFE}_{ij}\right) / \left(N^{-1}\sum_{i=1}^{N}\mathrm{MSFE}_{ib}\right)$. We also report the proportion of units in the cross-section for which each method produces a smaller MSFE than the benchmark along with the proportion of units in the cross-section for which each method has the smallest or largest MSFE value.

Similar to the simulation study, we distinguish between forecasts where the regressors are close to their mean or one standard deviation away from their mean. Unlike in the Monte Carlo experiments, the parameters are unknown in the two applications and we, therefore, select forecasts based on $d_{i,T+1} = \hat{\boldsymbol{\theta}}_i' \boldsymbol{w}_{i,T+1}$. Regressors are said to be in the neighborhood of the mean of d_{it} when $|d_{i,T+1} - \bar{d}_i - \kappa_i s_d| < c\sigma_d$, where $\bar{d}_i$ is the mean and s_d the standard deviation of d_{it} in the estimation sample, $t = 1, 2, \ldots, T$ and $c = 0.1$. $\kappa_i = 0$ then gives the results where the predictors are close to their mean and $\kappa_i = \pm 1$ shows the results when the predictors one standard deviation away from the mean. Additionally, we report results for all forecasts.

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.[17] 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. For simplicity, we report results for all forecasts.

6.2 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 target the one-quarter-ahead MSA-level rate of house price log changes. After accounting for the necessary pre-sample and the estimation window, the first forecast is for 1991Q2 and the last for 2023Q1, a total of 128 forecasts per MSA.

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

$$y_{it} = \alpha_i + \beta_i y_{i,t-1} + \beta_i^* y_{i,t-1}^* + \gamma_{Ri} \bar{y}_{i,t-1}^{(R)} + \gamma_{Ci} \bar{y}_{t-1}^{(C)} + \varepsilon_{it}, \tag{52}$$

[17]The panel DM test first computes the difference between the cross-sectional average squared forecast error at a given point in time for the benchmark versus competing model. It then uses the time series of these average squared forecast errors to compute Newey-West HAC standard errors that account for serial dependencies.

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

where $i = 1, 2, \ldots, N$ denotes individual MSAs and $t = 1, 2, \ldots, T$ refers to the time period, $y^*_{it} = \sum_{k=1,k\neq i}^{N} \omega^s_{ik} y_{kt}$ is the spatial effect for a set of spatial weights ω^s_{ik}, $\bar{y}^{(R)}_{it}$ is the average house price inflation in the region of unit i, and $\bar{y}^{(C)}_{t}$ 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 $\omega^s_{ik} = v_{ik}/\sum_{k=1}^{N} v_{ik}$ 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 without neighbors within 100 miles, which leaves 362 MSAs in our sample.

The top panel in Table 2 reports the results. The column labeled "all" shows results averaged across the full test sample, while columns labeled $\kappa_i = 0$ and $\kappa_i = \pm 1$ show results for sub-samples in which the predictor vector is close to the mean and one standard deviation away from the mean, respectively. In the first three 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 three 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. 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.

Across the full sample, the average MSFE ratio below one for the pooled, RE, and FE forecasts. However, these methods do notably worse than the forecasts based on individual estimates when the predictors are close to their mean ($\kappa_i = 0$). Empirical Bayes forecast produce the best overall MSFE performance, reducing the MSFE of the benchmark by 10%, followed by reductions of 6-8% among the three forecast combination schemes. The EB forecasts perform particularly well when the predictors are far away from their mean.

While the proportional reductions in MSFE ratios may not seem very large, they translate into very high frequencies of beating the benchmark. The EB forecasts produce lower MSFE values than the benchmark for 94% of the housing price series followed by 92-94% for the forecast combinations but only 59-62% for the pooled, RE, and FE forecasts.

Turning to evidence of individual forecasts being "best" or "worst", for the full test sample the

Table 2: Results for the applications

	Ratio of ave. MSFE			Freq. beating benchmark			Freq. smallest MSFE			Freq. largest MSFE		
Observations	all	$\kappa_i = 0$	$\kappa_i = \pm 1$	all	$\kappa_i = 0$	$\kappa_i = \pm 1$	all	$\kappa_i = 0$	$\kappa_i = \pm 1$	all	$\kappa_i = 0$	$\kappa_i = \pm 1$
House price inflation forecasts												
Individual	2.822	2.520	3.542	–	–	–	0.008	0.273	0.146	0.569	0.282	0.420
Pooled	0.920	1.162	0.947	0.613	0.381	0.536	0.116	0.171	0.249	0.157	0.188	0.160
RE	0.924	1.166	0.960	0.619	0.376	0.528	0.108	0.022	0.047	0.003	0.019	0.008
FE	0.936	1.186	0.980	0.591	0.381	0.517	0.055	0.108	0.105	0.251	0.376	0.296
Emp.Bayes	0.901	0.955	0.881	0.942	0.519	0.652	0.185	0.157	0.127	0.003	0.064	0.052
Comb. (pool)	0.920	0.961	0.932	0.939	0.522	0.688	0.157	0.077	0.099	0.000	0.011	0.017
Comb. (FE)	0.937	0.977	0.940	0.917	0.494	0.677	0.019	0.072	0.069	0.011	0.039	0.044
Comb. (ω_i^*)	0.921	0.957	0.909	0.936	0.541	0.713	0.044	0.028	0.052	0.006	0.003	0.006
CPI inflation forecasts												
Individual	15.501	10.451	11.295	–	–	–	0.005	0.134	0.070	0.439	0.214	0.316
Pooled	0.878	1.013	0.971	0.444	0.374	0.417	0.203	0.118	0.112	0.396	0.406	0.380
RE	0.880	1.001	0.957	0.508	0.390	0.401	0.016	0.070	0.064	0.000	0.037	0.032
FE	0.883	0.992	0.959	0.508	0.401	0.401	0.000	0.086	0.102	0.166	0.230	0.225
Emp.Bayes	0.892	0.991	0.926	0.984	0.652	0.818	0.390	0.278	0.278	0.000	0.070	0.011
Comb. (pool)	0.930	0.987	0.953	0.733	0.481	0.572	0.128	0.091	0.123	0.000	0.027	0.016
Comb. (FE)	0.935	0.980	0.967	0.791	0.524	0.583	0.053	0.064	0.070	0.000	0.016	0.021
Comb. (ω_i^*)	0.897	0.972	0.931	0.973	0.695	0.813	0.203	0.160	0.182	0.000	0.000	0.000

results for the house price application and the bottom panel reports the results for the CPI subindices application. The first three columns report the ratio of average MSFE of the respective method in the row relative to that of the individual forecast. The exception is the individual forecast, which reports the average MSFE (times 10^5 in the case of CPI). The second three columns report the proportion of cross-section units for which the respective method in the rows have a lower MSFE than the individual forecast. The third three columns report the proportion of cross-section units for which the respective method in the row has the lowest MSFE. The last three columns report the proportion of units for which the respective method in the row has the highest MSFE. In each block, the first column averages over all forecasts, the second over the forecasts for which $d_{i,T+1} = \hat{\boldsymbol{\theta}}_i' \boldsymbol{w}_{i,T+1}$ is close to its mean in the estimation sample. The third column averages over the forecast for which $d_{i,T+1}$ is close to plus or minus one standard deviation from its mean in the estimation sample. The methods in the rows are listed in the footnote of Table 1.

benchmark forecasts only produce the smallest MSFE for 1% of the variables versus 19% for the EB and 16% for pooled forecast combination schemes. Using this metric, again the benchmark forecasts perform much better when the predictors are close to their sample mean for which they are most accurate for 27% of the MSAs versus 16% and 8% for the EB and pooled combinations, respectively. Conversely, forecasts based on individual estimates are worst overall for 57% of the variables versus 1% or less for the EB and forecast combination schemes.

These results show that the EB 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 often generating substantially better results. Interestingly, the risk of producing the highest MSFE value is notably lower for the pooled combination and individual weighted combination than for the EB forecasts when the predictors are close to their mean.[19]

Figure 1 summarizes our findings visually through density plots fitted to the cross-sectional distribution of MSFE ratios for our forecasting methods.[20] MSFE ratios have a widely dispersed, right-skewed distribution for the pooled forecasts compared to the Bayesian and combination approaches whose distributions are far more peaked and centered just below unity. 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.[21]

The first and second rows of Table 3 reports panel DM test statistics and the number of cross-sectional units with a DM test below -1.96 (panel forecasts are significantly more accurate) or above 1.96 (individual-specific forecasts are significantly more accurate), respectively, for each application.

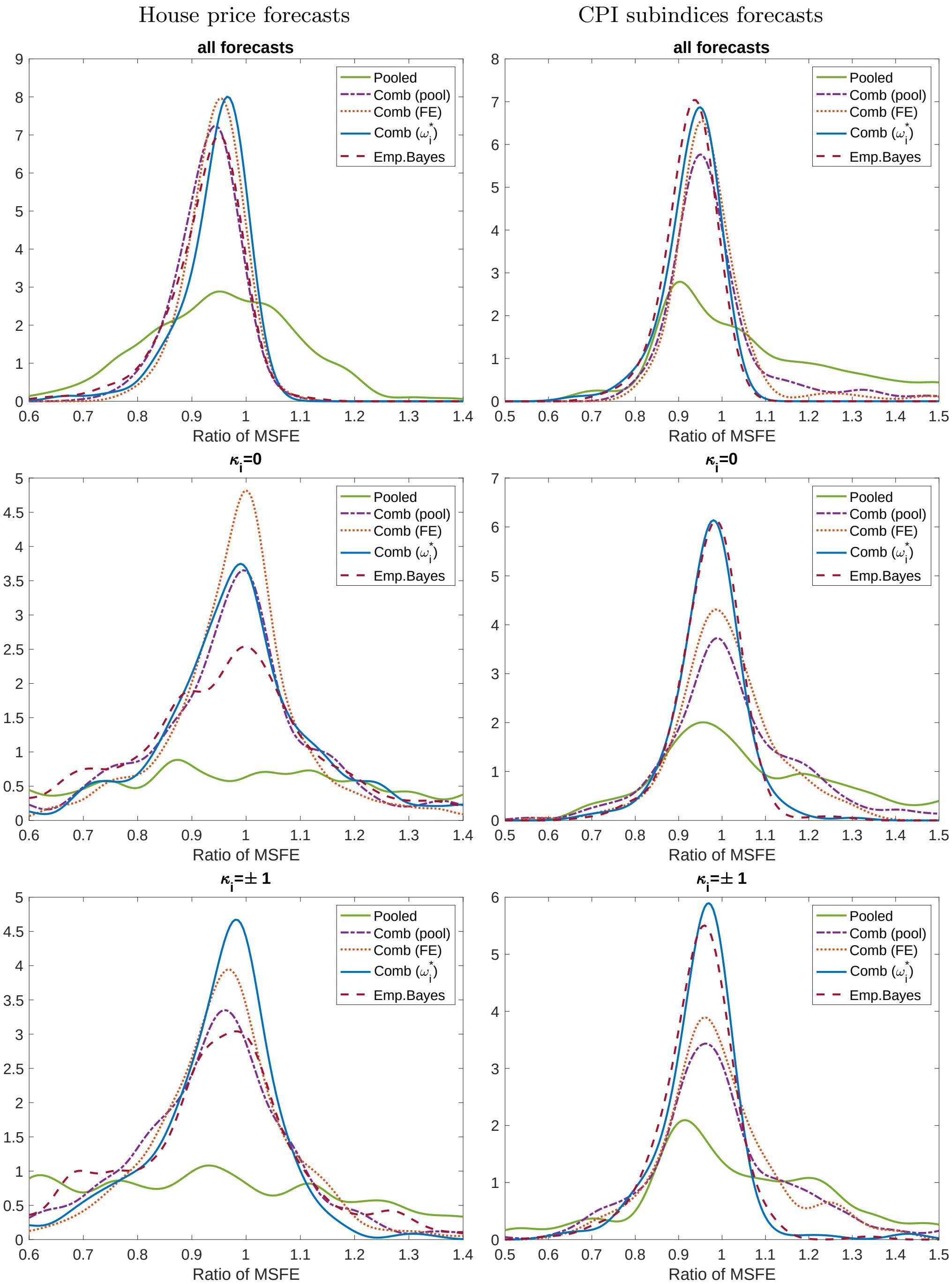
The panel DM tests show that the EB and combination forecasts are significantly more accurate than the individual forecasts "on average" as well as for a large portion of the individual series (between 169 and 240 MSAs), while the opposite only happens for two individual MSAs in the case of the EB forecasts. Pooled, RE and FE panel forecasts are also significantly more accurate than the individual forecasts on average as well as for between 57 and 62 of the individual MSAs and significantly less accurate for very few MSAs.

[19]Equal-weighted combinations also performs quite well in both of the empirical applications, which is a known feature in the forecast combination literature.

[20]To reduce the number of lines, we do not plot the densities for the FE and RE approaches which are very similar to those from pooling.

[21]The impressive performance of the EB approach for the tail groups is consistent with Efron (2011).

Figure 1: Distributions of ratios of MSFEs



Notes: The graphs show density plots of the ratios of MSFEs for the house price application in the left column and those for the CPI subindices application in the right column. In the first row are the density plots for the MSFEs from all forecasts, in the second row for the forecasts for which $d_{i,T+1} = \hat{\boldsymbol{\theta}}_i' \boldsymbol{w}_{i,T+1}$ is close to its mean in the estimation sample, and in the third row for the forecast for which $d_{i,T+1}$ is close to plus or minus one standard deviation from its mean in the estimation sample. The density estimates use a normal kernel with a bandwidth 0.04. The forecasting methods are listed in the footnote of Tables 1.

Table 3: Diebold-Mariano test statistics for equal predictive accuracy

	Pooled	RE	FE	Emp.Bay.	Comb(pool)	Comb(FE)	Comb(ω_i^*)
House Prices: all forecasts							
Panel DM	−9.45	−9.11	−7.57	−24.63	−22.93	−21.19	−27.59
DM < −1.96/DM > 1.96	60/6	62/6	57/8	209/2	189/0	169/0	240/0
CPI: all forecasts							
Panel DM	−7.95	−7.78	−7.56	−11.25	−11.67	−10.59	−11.48
DM < −1.96/DM > 1.96	35/60	33/45	32/42	134/0	56/23	50/10	137/0

Notes: The row "Panel DM" reports the results of the panel version of the Diebold-Mariano test of Pesaran et al. (2013). The second row report unit by unit Diebold-Mariano test results: "DM < -1.96" reports the number of units with a DM test statistic smaller than -1.96 and "DM > 1.96" shows the number of units whose test statistic exceeds 1.96. The remaining units have insignificant DM test statistics. In total the house prices panel consist of 362 units and the CPI panel of 187 units. 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.

6.3 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 up to 599 forecasts for each series, with the first forecast computed for February 1973.

We consider an autoregressive forecasting specification with lags 1, 2, and 12 augmented with lagged values of the first principal component of the data, the default yield and term spread.

The bottom panel of Table 2 shows that, for the full test sample, all forecasting methods produce lower MSFE values than the benchmark. The pooled, RE, FE and EB forecasts reduce the average MSFE of the benchmark by around 12%, while the forecast combination methods reduce it by 7-10%. Interestingly, when the predictors are close to their sample mean, the lowest MSFE ratios are produced by the three forecast combination methods, while conversely the EB scheme performs best when the predictors are further removed from their mean.

The EB forecasting scheme performs particularly well overall, beating the benchmark model's accuracy for 98% of the variables followed by 97% for the individual weights, 73-79% for the pooled and FE combinations and around 50% for the RE and FE schemes. As in the first application, these percentages are notably lower for predictors close to their mean and higher further away.

The EB forecasts also produce by far the highest frequency with the smallest MSFE values overall

(39%) followed by 20% for the pooled and individual forecast combination scheme. This is matched by very low probabilities of producing the worst forecast which never occurs in our sample for the EB method or any of the three forecast combination schemes but is far more likely to occur for the benchmark (43.9%) and pooled forecasts (39.6%).

Our evidence is summarized the probability density plots for the MSFE ratios in the right panels of Figure 1. The figure clearly highlights the pronounced dispersion and thick right tails of the MSFE-ratio distribution for the pooled forecasts. The distributions of MSFE ratios of the EB and combination approaches are far more concentrated and less asymmetrical. For values of the predictors farther away from the mean, the tails of the densities are somewhat thicker, with the EB approach standing out as having the thinnest right tail and, hence, the lowest probability of generating forecasts less accurate than those from the individual-specific benchmark.

Turning to the DM test results for the CPI inflation data in Table 3, all panel models generate significantly negative DM panel test statistics and so their associated forecasts are significantly more accurate, on average, than the individual forecasts. The pooled, RE, and FE models perform somewhat worse in this application, as the number of individual CPI series for which their forecasts are significantly more accurate than the individual-specific forecasts is smaller than those for which the opposite holds. Conversely, the EB and combination forecasts continue to be significantly more accurate than the benchmark forecasts for between 50 and 137 of the individual CPI series and are only significantly less accurate for between zero and 23 series. The EB and individual combination approaches perform particularly well in this application.

7 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 Bayes, and forecast combinations.

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 both on average and for the majority of individual units from exploiting panel information.

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 differ in terms of 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

Armstrong, T.B., M. Kolesár, and M. Plagborg-Møller (2022). Robust empirical Bayes confidence intervals. *Econometrica*, 90, 2567–2602.

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 (eds.) *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 E. Greenshtein (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.

Chudik, A., V. Grossman, and M.H. Pesaran (2016). A multi-country approach to forecasting output growth using PMIs. *Journal of Econometrics*, 192, 349–365.

Chudik, A., M.H. Pesaran and J.-C. Yang (2018). Half-panel jackknife fixed effects estimation of linear panels with weakly exogenous regressors. *Journal of Applied Econometrics*, 33, 816–836

Dhaene, G., and K. Jochmans (2015). Split-panel jackknife estimation of fixed effect models. *Review of Economic Studies*, 82, 991–1030

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.

Efron, B. (2016). Empirical Bayes deconvolution estimates. *Biometrika* 103, 1–20.

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

Fan, J. and Y. Liao and J. Yao (2015). Power enhancement in high-dimensional cross-sectional tests. *Econometrica*, 83, 1497–1541.

Gelfand, A.E., S.E. Hills, A. Racine-Poon and 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.

Giacomini, R., S. Lee, and S. Sarpietro (2023). A robust method for microforecasting and estimation of random effects. Unpublished Working Paper.

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.

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 (eds.) *Analysis of Panels and Limited Dependent Variable Models*, Cambridge: Cambridge University Press.

Lahiri, K., H. Peng and Y. Zhao (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 and 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.

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., A. Pick, and A. Timmermann (2022). Forecasting with panel data: estimation uncertainty versus parameter heterogeneity. *CEPR Discussion Paper* 17123.

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 (2024a). Heterogeneous autoregressions in short panel data models. *Journal of Applied Econometrics*, 39, 1359–1378.

Pesaran, M.H., and L. Yang (2024b). Trimmed mean group estimation of average treatment effects in ultra short T panels under correlated heterogeneity. `https://arxiv.org/abs/2310.11680`.

Pick, A., and A. Timmermann (2024). Panel data forecasting. Ch. 4 in M. Clements and A. Galvao (eds.) *Handbook of Research Methods and Applications in Macroeconomic Forecasting*, Cheltenham: Edgar Elgar.

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

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

and by the Cauchy–Schwarz inequality

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

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

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

By the Cauchy-Schwarz inequality and under Assumption 4

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

and $\sup_{i,T} \mathrm{E}\left|r_{iT}\right| < C$. 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

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

where s_{iT} is given by (18). 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 $\sup_{i,T} \mathrm{E}\left|s_{iT}\right| < C$. Also

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

and $\sup_{i,T}\left\|\boldsymbol{w}_{i,T+1}\right\| < C$. Hence, $\sup_{i,T} \mathrm{E}\left|s_{iT}\right| < C$ 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 (19) follows, with h_{NT} given by (20).

A.2.2 Proof of Proposition 2

The average MSFE of forecasts based on pooled estimates is given by (21) 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

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

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 (8) and (10), 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 (21) now yields

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

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 (22).

To establish part (b) of Proposition 2 note that the first term of Δ_{NT}, $N^{-1}\sum_{i=1}^{N}\mathrm{E}\left(\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\boldsymbol{w}_{i,T+1}\right) = N^{-1}\sum_{i=1}^{N}\mathrm{E}\left(\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\right)^2 \geq 0$, 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 (Δ_{NT}), can be signed under stationarity where $\mathrm{E}\left(\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\boldsymbol{w}_{i,T+1}\right) = \mathrm{E}\left(\boldsymbol{w}_{it}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\boldsymbol{w}_{it}\right)$. In this case, we have

$$\Delta_N = N^{-1}\sum_{i=1}^{N}\mathrm{E}\left(\boldsymbol{w}_{it}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\boldsymbol{w}_{it}\right) - \bar{\boldsymbol{q}}_N'\bar{\boldsymbol{Q}}_N^{-1}\bar{\boldsymbol{q}}_N. \tag{A.16}$$

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

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

which yields the residual sum of squares

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

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

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

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

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

A.2.3 Proof of Proposition 3

Using (14) and (15),

$$\begin{aligned} N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}\tilde{e}_{i,T+1} &= N^{-1}\sum_{i=1}^{N}\varepsilon_{i,T+1}^{2}+N^{-1}\sum_{i=1}^{N}(\hat{\boldsymbol{\theta}}_i-\boldsymbol{\theta}_i)'\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_i) \\ &\quad -N^{-1}\sum_{i=1}^{N}(\hat{\boldsymbol{\theta}}_i-\boldsymbol{\theta}_i)'\boldsymbol{w}_{i,T+1}\varepsilon_{i,T+1}-N^{-1}\sum_{i=1}^{N}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_i)'\boldsymbol{w}_{i,T+1}\varepsilon_{i,T+1}, \end{aligned} \tag{A.17}$$

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 (16), by (A.8) it follows that

$$N^{-1}\sum_{i=1}^{N}(\hat{\boldsymbol{\theta}}_i-\boldsymbol{\theta}_i)'\boldsymbol{w}_{i,T+1}\varepsilon_{i,T+1}=N^{-1}\sum_{i=1}^{N}r_{iT}=R_{NT}=O_p(N^{-1/2}). \tag{A.18}$$

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

$$N^{-1}\sum_{i=1}^{N}(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_i)'\boldsymbol{w}_{i,T+1}\varepsilon_{i,T+1}=O_p\left(N^{-1/2}\right). \tag{A.19}$$

Consider now the second term of (A.17):

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

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

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

where

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

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,

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

Substituting this result together with (A.18) and (A.19) in (A.17), we obtain

$$N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}\tilde{e}_{i,T+1}=N^{-1}\sum_{i=1}^{N}\varepsilon_{i,T+1}^2+T^{-1}\psi_{NT}+O_p(N^{-1/2})+O_p\left(T^{-1/2}N^{-1/2}\right), \quad \text{(A.20)}$$

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

$$\begin{aligned}\psi_{NT} &= TN^{-1}\sum_{i=1}^{N}\mathrm{E}\left[\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i(\boldsymbol{W}_i'\boldsymbol{W}_i)^{-1}\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\right]\bar{\boldsymbol{Q}}_N^{-1}\bar{\boldsymbol{q}}_N \\ &\quad -TN^{-1}\sum_{i=1}^{N}\mathrm{E}\left[\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i(\boldsymbol{W}_i'\boldsymbol{W}_i)^{-1}\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\right].\end{aligned} \tag{A.21}$$

Finally, using (A.20), together with (19) and (22), in (31) now yields (32).

A.2.4 Proof of Proposition 4

To compare the FE forecast to the individual forecasts, rewrite (14) 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, $\hat{e}_{i,T+1} = \bar{\bar{\varepsilon}}_{i,T+1} - \bar{\bar{\boldsymbol{x}}}_{i,T+1}'(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)$. Furthermore, $\hat{e}_{i,T+1}^{\mathrm{FE}} = \bar{\bar{\varepsilon}}_{i,T+1} - (\hat{\boldsymbol{\beta}}_{\mathrm{FE}} - \boldsymbol{\beta}_i)'\bar{\bar{\boldsymbol{x}}}_{i,T+1}$, where $\bar{\bar{\varepsilon}}_{i,T+1} = \varepsilon_{i,T+1} - \bar{\varepsilon}_{iT}$ and $\bar{\bar{\boldsymbol{x}}}_{i,T+1} = \boldsymbol{x}_{iT+1} - \bar{\boldsymbol{x}}_{iT}$,

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

Hence,

$$\begin{aligned}N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}^{\mathrm{FE}}\hat{e}_{i,T+1} &= N^{-1}\sum_{i=1}^{N}\bar{\bar{\varepsilon}}_{i,T+1}^2 \\ &\quad +N^{-1}\sum_{i=1}^{N}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)'\bar{\bar{\boldsymbol{x}}}_{i,T+1}\bar{\bar{\boldsymbol{x}}}_{i,T+1}'(\hat{\boldsymbol{\beta}}_{\mathrm{FE}} - \boldsymbol{\beta}_i) \\ &\quad -N^{-1}\sum_{i=1}^{N}(\varepsilon_{i,T+1} - \bar{\varepsilon}_{iT})\bar{\bar{\boldsymbol{x}}}_{i,T+1}'\left[\left(\hat{\boldsymbol{\beta}}_{\mathrm{FE}} - \boldsymbol{\beta}_i\right) + (\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)\right].\end{aligned} \tag{A.22}$$

Under Assumptions 4 and 9 we have

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

Additionally,

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

where $c_{NT,\beta}=N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[\bar{\bar{\boldsymbol{x}}}_{i,T+1}'(\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{\varepsilon}_i\bar{\varepsilon}_{iT}\right]$. Details are in the Online Supplement, where we also show that

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

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

Using this, we have

$$N^{-1}\sum_{i=1}^{N}\bar{\bar{\varepsilon}}_{i,T+1}\bar{\bar{\boldsymbol{x}}}_{i,T+1}'\left[\left(\hat{\boldsymbol{\beta}}_{\mathrm{FE}}-\boldsymbol{\beta}_i\right)+(\hat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i)\right]=c_{NT}^{\mathrm{FE}}+c_{NT,\beta}+O_p(N^{-1/2}). \tag{A.23}$$

Also

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

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 $T>T_0$

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

Also, under Assumption 9,

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

and

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

Then

$$
N^{-1} \sum_{i=1}^{N} (\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)' \bar{\bar{\boldsymbol{x}}}_{i,T+1} \bar{\bar{\boldsymbol{x}}}_{i,T+1}' (\hat{\boldsymbol{\beta}}_{\mathrm{FE}} - \boldsymbol{\beta}_i) = T^{-1/2} \psi_{NT}^{\mathrm{FE}} + O_p(T^{-1/2} N^{-1/2}), \tag{A.24}
$$

where

$$
\begin{aligned}
\psi_{NT}^{\mathrm{FE}} \quad = \quad & N^{-1} \sum_{i=1}^{N} \mathrm{E} \left[\left(T^{-1/2} \boldsymbol{\varepsilon}_i' \boldsymbol{M}_T \boldsymbol{X}_i \right) \boldsymbol{Q}_{iT,\beta}^{-1} \left(\bar{\bar{\boldsymbol{x}}}_{i,T+1} \bar{\bar{\boldsymbol{x}}}_{i,T+1}' \right) \right] \bar{\boldsymbol{Q}}_{N,\beta}^{-1} \bar{\boldsymbol{q}}_{N,\beta} \\
& - N^{-1} \sum_{i=1}^{N} \mathrm{E} \left[\left(T^{-1/2} \boldsymbol{\varepsilon}_i' \boldsymbol{M}_T \boldsymbol{X}_i \right) \boldsymbol{Q}_{iT,\beta}^{-1} \left(\bar{\bar{\boldsymbol{x}}}_{i,T+1} \bar{\bar{\boldsymbol{x}}}_{i,T+1}' \right) \boldsymbol{\eta}_{i,\beta} \right].
\end{aligned} \tag{A.25}
$$

Using (A.23) and (A.24) in (A.22) yields

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

Substituting this result together with (S.22) and (27) in (38) now yields equation (39).

A.3 Estimation of Combination Weights

There are three components in the forecast combination weights, given by (32), namely Δ_{NT}, h_{NT} and ψ_{NT}. To establish that $\hat{\Delta}_{NT}(\tilde{\boldsymbol{\eta}}) = N^{-1} \sum_{i=1}^{N} \boldsymbol{w}_{i,T+1}' \tilde{\boldsymbol{\eta}}_i \tilde{\boldsymbol{\eta}}_i' \boldsymbol{w}_{i,T+1}$ is a consistent estimator of Δ_{NT}, recall that

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

where

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

Then

$$\begin{aligned}
\hat{\Delta}_{NT}(\tilde{\boldsymbol{\eta}}) &= N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\tilde{\boldsymbol{\eta}}_i\tilde{\boldsymbol{\eta}}_i'\boldsymbol{w}_{i,T+1} \\
&= N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\left(\boldsymbol{\eta}_i + \boldsymbol{Q}_{iT}^{-1}\boldsymbol{\xi}_{iT} - \bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT} - \bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{\xi}}_{NT}\right) \\
&\quad \times\left(\boldsymbol{\eta}_i + \boldsymbol{Q}_{iT}^{-1}\boldsymbol{\xi}_{iT} - \bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT} - \bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{\xi}}_{NT}\right)'\boldsymbol{w}_{i,T+1} \\
&= 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}\boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_{iT}^{-1}\boldsymbol{\xi}_{iT}\boldsymbol{\xi}_{iT}'\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1} \\
&\quad + N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT}\bar{\boldsymbol{q}}_{NT}\bar{\boldsymbol{Q}}_{NT}^{-1}\boldsymbol{w}_{i,T+1} + N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{\xi}}_{NT}\bar{\boldsymbol{\xi}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\boldsymbol{w}_{i,T+1} \\
&\quad + 2N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\boldsymbol{\xi}_{iT}'\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1} - 2N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\bar{\boldsymbol{q}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\boldsymbol{w}_{i,T+1} \\
&\quad - 2N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\bar{\boldsymbol{\xi}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\boldsymbol{w}_{i,T+1} - 2N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_{iT}^{-1}\boldsymbol{\xi}_{iT}\bar{\boldsymbol{q}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\boldsymbol{w}_{i,T+1} \\
&\quad - N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_{iT}^{-1}\boldsymbol{\xi}_{iT}\bar{\boldsymbol{\xi}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\boldsymbol{w}_{i,T+1} + 2N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT}\bar{\boldsymbol{\xi}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\boldsymbol{w}_{i,T+1},
\end{aligned}$$

and we have that

$$\begin{aligned}
N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_{iT}^{-1}\boldsymbol{\xi}_{iT}\boldsymbol{\xi}_{iT}'\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1} &= N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\mathrm{E}\left(\boldsymbol{Q}_{iT}^{-1}\boldsymbol{\xi}_{iT}\boldsymbol{\xi}_{iT}'\boldsymbol{Q}_{iT}^{-1}\right)\boldsymbol{w}_{i,T+1} + O_p\left(N^{-1/2}\right) \\
\mathrm{E}\left(\boldsymbol{Q}_{iT}^{-1}\boldsymbol{\xi}_{iT}\boldsymbol{\xi}_{iT}'\boldsymbol{Q}_{iT}^{-1}\right) &= T^{-2}\mathrm{E}\left(\boldsymbol{Q}_{iT}^{-1}\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i'\boldsymbol{Q}_{iT}^{-1}\right) = T^{-1}\sigma_i^2\boldsymbol{Q}_i^{-1} \\
N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\boldsymbol{\xi}_{iT}\boldsymbol{\xi}_{iT}'\boldsymbol{w}_{i,T+1} &= N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\mathrm{E}\left(\boldsymbol{Q}_{iT}^{-1}\boldsymbol{\xi}_{iT}\boldsymbol{\xi}_{iT}'\boldsymbol{Q}_{iT}^{-1}\right)\boldsymbol{w}_{i,T+1} + O_p\left(N^{-1/2}\right) \\
&= O_p\left(N^{-1/2}\right) + O_p(T^{-1}),
\end{aligned}$$

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

$$N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{\xi}}_{NT}\bar{\boldsymbol{\xi}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\boldsymbol{w}_{i,T+1}=\bar{\boldsymbol{\xi}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\right)\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{\xi}}_{NT}=O_p(N^{-1}T^{-1}),$$

and

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

Also since $\mathrm{E}\big(\boldsymbol{Q}_{iT}^{-1}\boldsymbol{\xi}_{iT}\,|\boldsymbol{w}_{i,T+1},\boldsymbol{\eta}_i\big)=0$,

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

$$N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\bar{\boldsymbol{\xi}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\boldsymbol{w}_{i,T+1}=N^{-3}\sum_{i=1}^{N}\boldsymbol{\xi}_{iT}'\boldsymbol{Q}_{iT}^{-1}\left(\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\right)+O_p\left(N^{-3}\right)=O_p\left(N^{-5/2}\right),$$

$$N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_{iT}^{-1}\boldsymbol{\xi}_{iT}\bar{\boldsymbol{q}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\boldsymbol{w}_{i,T+1}=\bar{\boldsymbol{q}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\left(N^{-1}\sum_{i=1}^{N}\left(\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\right)\boldsymbol{Q}_{iT}^{-1}\boldsymbol{\xi}_{iT}\right)=O_p\left(N^{-1/2}\right),$$

$$N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_{iT}^{-1}\boldsymbol{\xi}_{iT}\bar{\boldsymbol{\xi}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\boldsymbol{w}_{i,T+1}=\bar{\boldsymbol{\xi}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_{iT}^{-1}\boldsymbol{\xi}_{iT}\right)=O_p(N^{-1}T^{-1/2}),$$

and

$$N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT}\bar{\boldsymbol{\xi}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\boldsymbol{w}_{i,T+1}=\bar{\boldsymbol{\xi}}_{NT}'\bar{\boldsymbol{Q}}_{NT}^{-1}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\right)\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT}=O_p(T^{-1/2}N^{-1/2}).$$

Overall

$$\hat{\Delta}_{NT}\left(\tilde{\boldsymbol{\eta}}\right)=N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\boldsymbol{\eta}_i\boldsymbol{\eta}_i'\boldsymbol{w}_{i,T+1}-\bar{\boldsymbol{q}}_N\bar{\boldsymbol{Q}}_N^{-1}\bar{\boldsymbol{q}}_N+O_p\left(N^{-1/2}\right)+O_p(T^{-1}),$$

or equivalently

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

In combination, $\hat{\Delta}_{NT}\left(\tilde{\boldsymbol{\eta}}\right)$ is a consistent estimator of Δ_{NT} for large N and T. In the case of strictly exogenous regressors, $\hat{\Delta}_{NT}\left(\tilde{\boldsymbol{\eta}}\right)$ is a consistent estimator of Δ_{NT} for a fixed T, so long as $E\left(\hat{\boldsymbol{\theta}}_i\right)$ exits, since in that case $\mathrm{E}\left(\hat{\boldsymbol{\theta}}_j-\boldsymbol{\theta}_j\right)=\mathbf{0}$ for all j.

Now consider the second component of the weights, namely h_{NT}, and We will show that a consistent estimator is

$$\hat{h}_{NT} = N^{-1}\sum_{i=1}^{N} \boldsymbol{w}'_{i,T+1}\boldsymbol{Q}_{iT}^{-1}\hat{\boldsymbol{H}}_{iT}\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1} = h_{NT} + O_p\left(N^{-1/2}\right) + O_p\left(\frac{\ln(N)}{\sqrt{T}}\right),$$

where $\hat{\boldsymbol{H}}_{iT} = \hat{\sigma}_i^2 T^{-1}\sum_{t=1}^{T}\boldsymbol{w}_{it}\boldsymbol{w}'_{it}$.where $\hat{\sigma}_i^2 = \sum_{t=1}^{T}\hat{\varepsilon}_{it}^2/(T-K)$, instead of $T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2\left(\boldsymbol{w}_{it}\boldsymbol{w}'_{it}\right)$. Note that

$$\hat{h}_{NT} - h_{NT} = N^{-1}\sum_{i=1}^{N} \boldsymbol{w}'_{i,T+1}\boldsymbol{Q}_{iT}^{-1}\hat{\boldsymbol{H}}_{iT}\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1}, -N^{-1}\sum_{i=1}^{N}\mathrm{E}\left(s_{iT}\ \right),$$

where s_{iT} is defined by (18). Since $N^{-1}\sum_{i=1}^{N}\mathrm{E}\left(s_{iT}\ \right) = N^{-1}\sum_{i=1}^{N} s_{iT} + O_p\left(N^{-1/2}\right)$,

$$\begin{aligned}
\hat{h}_{NT} - h_{NT} &= N^{-1}\sum_{i=1}^{N}\boldsymbol{w}'_{i,T+1}\boldsymbol{Q}_{iT}^{-1}\left[\hat{\sigma}_i^2 T^{-1}\sum_{t=1}^{T}\boldsymbol{w}_{it}\boldsymbol{w}'_{it} - \frac{\boldsymbol{W}'_i\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_i\boldsymbol{W}_i}{T}\right]\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1} + O_p\left(N^{-1/2}\right) \\
&= N^{-1}\sum_{i=1}^{N}\hat{\sigma}_i^2\boldsymbol{w}'_{i,T+1}\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1} - N^{-1}\sum_{i=1}^{N}\boldsymbol{w}'_{i,T+1}\boldsymbol{Q}_{iT}^{-1}\left(\frac{\boldsymbol{W}'_i\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}'_i\boldsymbol{W}_i}{T}\right)\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1} + O_p\left(N^{-1/2}\right) \\
&= D_{1,NT} - D_{2,NT} + O_p\left(N^{-1/2}\right).
\end{aligned}$$

Now consider the decomposition

$$\begin{aligned}
D_{1,NT} &= N^{-1}\sum_{i=1}^{N}\hat{\sigma}_i^2\boldsymbol{w}'_{i,T+1}\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1} = N^{-1}\sum_{i=1}^{N}\hat{\sigma}_i^2\boldsymbol{w}'_{i,T+1}\left(\boldsymbol{Q}_{iT}^{-1} - \boldsymbol{Q}_i^{-1} + \boldsymbol{Q}_i^{-1}\right)\boldsymbol{w}_{i,T+1} \\
&= N^{-1}\sum_{i=1}^{N}\hat{\sigma}_i^2\boldsymbol{w}'_{i,T+1}\boldsymbol{Q}_i^{-1}\boldsymbol{w}_{i,T+1} + N^{-1}\sum_{i=1}^{N}\hat{\sigma}_i^2\boldsymbol{w}'_{i,T+1}\left(\boldsymbol{Q}_{iT}^{-1} - \boldsymbol{Q}_i^{-1}\right)\boldsymbol{w}_{i,T+1}
\end{aligned}$$

Similarly

$$\left\| N^{-1}\sum_{i=1}^{N}\hat{\sigma}_i^2\boldsymbol{w}'_{i,T+1}\left(\boldsymbol{Q}_{iT}^{-1} - \boldsymbol{Q}_i^{-1}\right)\boldsymbol{w}_{i,T+1}\right\| \leq \sup_i \|\boldsymbol{w}_{i,T+1}\|^2 \sup_i \hat{\sigma}_i^2 \sup_i \left\|\left(\boldsymbol{Q}_{iT}^{-1} - \boldsymbol{Q}_i^{-1}\right)\right\| = O_p\left(\frac{\ln(N)}{\sqrt{T}}\right),$$

and $\sup_i \left(\hat{\sigma}_i^2 - \sigma_i^2\right) = O_p\left(\frac{\ln(N)}{\sqrt{T}}\right)$. Hence

$$\begin{aligned}
D_{1,NT} &= N^{-1}\sum_{i=1}^{N}\hat{\sigma}_i^2 \boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_i^{-1}\boldsymbol{w}_{i,T+1} + O_p\left(\frac{\ln(N)}{\sqrt{T}}\right) \\
&= N^{-1}\sum_{i=1}^{N}\sigma_i^2 \boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_i^{-1}\boldsymbol{w}_{i,T+1} + N^{-1}\sum_{i=1}^{N}\left(\hat{\sigma}_i^2 - \sigma_i^2\right)\boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_i^{-1}\boldsymbol{w}_{i,T+1} + O_p\left(\frac{\ln(N)}{\sqrt{T}}\right) \\
&= N^{-1}\sum_{i=1}^{N}\sigma_i^2 \boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_i^{-1}\boldsymbol{w}_{i,T+1} + O_p\left(\frac{\ln(N)}{\sqrt{T}}\right)
\end{aligned}$$

Similarly, and noting that $\mathrm{E}\left(\frac{\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i}{T}\right) = \sigma_i^2\boldsymbol{Q}_i$, we have

$$\begin{aligned}
D_{2,NT} &= N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_{iT}^{-1}\left(\frac{\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i}{T}\right)\boldsymbol{Q}_{iT}^{-1}\boldsymbol{w}_{i,T+1} \\
&= N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_i^{-1}\left(\frac{\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i}{T}\right)\boldsymbol{Q}_i^{-1}\boldsymbol{w}_{i,T+1} + O_p\left(\frac{\ln(N)}{\sqrt{T}}\right) \\
&= N^{-1}\sum_{i=1}^{N}\boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_i^{-1}\mathrm{E}\left(\frac{\boldsymbol{W}_i'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{W}_i}{T}\right)\boldsymbol{Q}_i^{-1}\boldsymbol{w}_{i,T+1} + O_p(N^{-1/2}) + O_p\left(\frac{\ln(N)}{\sqrt{T}}\right) \\
&= N^{-1}\sum_{i=1}^{N}\sigma_i^2\boldsymbol{w}_{i,T+1}'\boldsymbol{Q}_i^{-1}\boldsymbol{w}_{i,T+1} + O_p(N^{-1/2}) + O_p\left(\frac{\ln(N)}{\sqrt{T}}\right).
\end{aligned}$$

Hence,

$$\hat{h}_{NT} - h_{NT} = O_p(N^{-1/2}) + O_p\left(\frac{\ln(N)}{\sqrt{T}}\right),$$

as desired.

Finally, turn to ψ_{NT}. The asymptotic bias, $\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i$, for each i can then be estimated using bootstrap or half-jackknifing. The sieve bootstrap could be used for a pure panel AR model but, generally, not with weakly exogenous regressors. However, the half-jackknife estimator can work more generally. For a give T, split the sample in two equal parts—one observation is dropped if T is an odd number. Denote the estimators based on the two sub-samples by $\hat{\boldsymbol{\theta}}_{ia}$ and $\hat{\boldsymbol{\theta}}_{ib}$. Then $\mathrm{E}\left(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right)$ can be estimated by

$$\hat{\boldsymbol{\theta}}_i - \left[2\hat{\boldsymbol{\theta}}_i - \frac{1}{2}\left(\hat{\boldsymbol{\theta}}_{ia} + \hat{\boldsymbol{\theta}}_{ib}\right)\right] = \left[\frac{1}{2}\left(\hat{\boldsymbol{\theta}}_{ia} + \hat{\boldsymbol{\theta}}_{ib}\right) - \hat{\boldsymbol{\theta}}_i\right]$$

A consistent estimator of ψ_{NT} is then given by

$$\begin{aligned}\hat{\psi}_{NT} &= \left[TN^{-1}\sum_{i=1}^{N}\left[\frac{1}{2}\left(\hat{\boldsymbol{\theta}}_{ia}+\hat{\boldsymbol{\theta}}_{ib}\right)-\hat{\boldsymbol{\theta}}_{i}\right]'\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\right]\bar{\boldsymbol{Q}}_{NT}^{-1}\bar{\boldsymbol{q}}_{NT}\left(\mathring{\boldsymbol{\eta}}\right)\\ &\quad -TN^{-1}\sum_{i=1}^{N}\left[\frac{1}{2}\left(\hat{\boldsymbol{\theta}}_{ia}+\hat{\boldsymbol{\theta}}_{ib}\right)-\hat{\boldsymbol{\theta}}_{i}\right]'\boldsymbol{w}_{i,T+1}\boldsymbol{w}_{i,T+1}'\mathring{\boldsymbol{\eta}}\end{aligned}$$

Estimating the weights in combination of individual and fixed effects forecasts

Similar to the derivations above, it can be shown that the components of the weights in Proposition 4 can be estimated as follows. First,

$$\hat{\Delta}_{NT}^{\mathrm{FE}}=\frac{1}{N}\sum_{i=1}^{N}\bar{\bar{\boldsymbol{x}}}_{i,T+1}'\mathring{\boldsymbol{\eta}}_{i,\beta}\mathring{\boldsymbol{\eta}}_{i,\beta}'\bar{\bar{\boldsymbol{x}}}_{i,T+1}-\hat{\bar{\boldsymbol{q}}}_{NT,\beta}'\left(\mathring{\boldsymbol{\eta}}_{i,\beta}\right)\hat{\bar{\boldsymbol{Q}}}_{NT,\beta}^{-1}\hat{\bar{\boldsymbol{q}}}_{NT,\beta}\left(\mathring{\boldsymbol{\eta}}_{i,\beta}\right), \tag{A.26}$$

$\mathring{\boldsymbol{\eta}}_{i,\beta}=\hat{\boldsymbol{\beta}}_i-\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_i$, $\bar{\bar{\boldsymbol{x}}}_{i,T+1}=\boldsymbol{x}_{iT+1}-\bar{\boldsymbol{x}}_{iT}$, $\hat{\bar{\boldsymbol{Q}}}_{NT,\beta}=N^{-1}T^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{X}_i$ and $\hat{\bar{\boldsymbol{q}}}_{NT,\beta}\left(\mathring{\boldsymbol{\eta}}_{i,\beta}\right)=N^{-1}T^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{X}_i\mathring{\boldsymbol{\eta}}_{i,\beta}$

Next,

$$\hat{h}_{NT,\beta}=N^{-1}\sum_{i=1}^{N}\bar{\bar{\boldsymbol{x}}}_{i,T+1}'\boldsymbol{Q}_{iT,\beta}^{-1}\mathring{\boldsymbol{H}}_{iT,\beta}\boldsymbol{Q}_{iT,\beta}^{-1}\bar{\bar{\boldsymbol{x}}}_{i,T+1}, \tag{A.27}$$

$\mathring{\boldsymbol{H}}_{iT,\beta}=\hat{\sigma}_i^2\frac{1}{T}\sum_{t=1}^{T}\bar{\bar{\boldsymbol{x}}}_{it}\bar{\bar{\boldsymbol{x}}}_{it}'$, $\hat{\sigma}_i^2=\hat{\boldsymbol{\varepsilon}}_i'\hat{\boldsymbol{\varepsilon}}_i/(T-K)$ and $\hat{\varepsilon}_{it}=y_{it}-\hat{\boldsymbol{\theta}}_i'\boldsymbol{w}_{it}$. Furthermore,

$$\begin{aligned}\hat{\psi}_{NT}^{\mathrm{FE}} &= TN^{-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\beta}}_{\mathrm{FE}}-\hat{\boldsymbol{\beta}}_{\mathrm{FEJK}}\right)'\bar{\bar{\boldsymbol{x}}}_{i,T+1}\bar{\bar{\boldsymbol{x}}}_{i,T+1}'\bar{\boldsymbol{Q}}_{N,\beta}^{-1}\bar{\boldsymbol{q}}_{N,\beta}(\mathring{\boldsymbol{\eta}}_{i,\beta})\\ &\quad -TN^{-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\beta}}_{\mathrm{FE}}-\hat{\boldsymbol{\beta}}_{\mathrm{FEJK}}\right)'\bar{\bar{\boldsymbol{x}}}_{i,T+1}\bar{\bar{\boldsymbol{x}}}_{i,T+1}'\mathring{\boldsymbol{\eta}}_{i,\beta}\end{aligned} \tag{A.28}$$

where $\hat{\boldsymbol{\beta}}_{\mathrm{FEJK}}=2\hat{\boldsymbol{\beta}}_{\mathrm{FE}}-\frac{1}{2}\left(\hat{\boldsymbol{\beta}}_{\mathrm{FE},a}+\hat{\boldsymbol{\beta}}_{\mathrm{FE},b}\right)$ is the half-jackknife estimator of Chudik et al. (2018).

Finally, the weights include the difference between

$$c_{NT}^{\mathrm{FE}}=\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\beta}}_{\mathrm{FE}}-\boldsymbol{\beta}_i\right)'\bar{\bar{\boldsymbol{x}}}_{i,T+1}\bar{\hat{\varepsilon}}_{iT} \quad \text{and} \quad \hat{c}_{NT,\beta}=\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i\right)'\bar{\bar{\boldsymbol{x}}}_{i,T+1}\bar{\hat{\varepsilon}}_{iT}.$$

However,

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

A.4 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 $\{y_{it}, t = 0, 1, 2, \ldots, T\}$.[22] In 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 (A.16). In the present example $q_i = \text{E}\left(y_{i,t-1}^2\eta_i\right)$, $Q_i = \text{E}\left(y_{i,t-1}^2\right)$, and

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

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

$$y_{it} = \sum_{s=0}^{\infty}\beta_i^s\varepsilon_{i,t-s} = \sum_{s=0}^{\infty}\left(\beta_0 + \eta_i\right)^s\varepsilon_{i,t-s}, \tag{A.29}$$

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

[22] The assumption that the process for y_{it} has started a long time prior to date 0, is equivalent to assuming that y_{i0} is drawn from a distribution with zero mean and variance $\sigma_i^2/(1-\beta_i^2)$.

of Pesaran and Smith (1995). Using (A.29) 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

$$\begin{aligned}\Delta_{\mathrm{AR}} &= \mathrm{E}\left(y_{it}^2\eta_i^2\right) - \frac{\left[\mathrm{E}\left(y_{i,t-1}^2\eta_i\right)\right]^2}{\mathrm{E}\left(y_{i,t-1}^2\right)} \\ &= \frac{\mathrm{E}\left(\frac{\eta_i^2\sigma_i^2}{1-\beta_i^2}\right)\mathrm{E}\left(\frac{\sigma_i^2}{1-\beta_i^2}\right) - \left[\mathrm{E}\left(\frac{\eta_i\sigma_i^2}{1-\beta_i^2}\right)\right]^2}{\mathrm{E}\left(\frac{\sigma_i^2}{1-\beta_i^2}\right)}. \end{aligned} \tag{A.30}$$

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 $\mathrm{E}(f_i^2)\mathrm{E}(g_i^2) - [E(f_ig_i)]^2 \geq 0$, which establishes that $\Delta_{\mathrm{AR}} \geq 0$, 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, $\mathrm{E}(\sigma_i^2) = \sigma^2$ and $\eta_i \sim \mathrm{Uniform}(-a/2, a/2)$, for $a > 0$.[23] Then,

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

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

$$\mathrm{E}\left(\frac{\eta_i}{B+\eta_i}\right) = \frac{1}{a}\int_{-a/2}^{a/2}\left(\frac{\eta}{B+\eta}\right)d\eta = 1 - \left(\frac{B}{a}\right)\ln\left(\frac{B+a/2}{B-a/2}\right). \tag{A.31}$$

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

Table A.1: 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	0.117
0.45	0.316	0.163
0.49	0.657	0.211
0.4999	1.783	0.328

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

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

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

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

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

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

$$\begin{aligned} \mathrm{E}\left(y_{i,t-1}^2\eta_i\right) &= \frac{\sigma^2}{2}\left[\left(\frac{\beta_0-1}{a}\right)\ln\left(\frac{\beta_0-1+a/2}{\beta_0-1-a/2}\right) - \left(\frac{1+\beta_0}{a}\right)\ln\left(\frac{1+\beta_0+a/2}{1+\beta_0-a/2}\right)\right] \\ &= \frac{\sigma^2}{2a}\left[-(1-\beta_0)\ln\left(\frac{1-\beta_0-a/2}{1-\beta_0+a/2}\right) - (1+\beta_0)\ln\left(\frac{1+\beta_0+a/2}{1+\beta_0-a/2}\right)\right]. \end{aligned} \quad \text{(A.32)}$$

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

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

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.1.

In general, for $a > 0$ and $|\beta_0| < 1$, $\mathrm{E}\left(y_{i,t-1}^2\eta_i\right) \neq 0$, $\mathrm{E}\left(y_{i,t-1}^2\eta_i\right) \to 0$, only if $a \to 0$. 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.30), 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

$$\mathrm{E}\left(\frac{1}{1-\beta_i^2}\right) = \frac{1}{2a}\left[\ln\left(\frac{1+\beta_0+a/2}{1+\beta_0-a/2}\right) - \ln\left(\frac{1-\beta_0-a/2}{1-\beta_0+a/2}\right)\right]. \quad \text{(A.33)}$$

Using (A.33) 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.1 for $10,000$ replications.

Online Supplement

Forecasting with panel data: Estimation uncertainty versus parameter heterogeneity

by

M. Hashem Pesaran[24]

Andreas Pick[25]

and

Allan Timmermann[26]

---

[24]University of Cambridge, UK, and University of Southern California, USA Email: mhp1@cam.ac.uk

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

[26]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.

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 5. 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 $\mathrm{Var}(y_{it}\,|\boldsymbol{\theta}_i,\, x_{it}\,)$, we note that

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

Hence,

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

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 $\sigma_{xi}^2 = \mathrm{Var}(\xi_{it})$. 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

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

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

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

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

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

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

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

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.33) in the Appendix to the paper):

$$\begin{aligned}
\mathrm{E}\left(\frac{1}{1-\beta_i^2}\right) &= \frac{1}{a_\beta}\int_{-a_\beta/2}^{a_\beta/2} \frac{1}{1-\left(\beta_0+\eta_\beta\right)^2} d\eta_\beta \\
&= \frac{1}{2a_\beta}\int_{-a_\beta/2}^{a_\beta/2} \left[\frac{1}{1+\beta_0+\eta_\beta} + \frac{1}{1-\beta_0-\eta_\beta}\right] d\eta_\beta \\
&= \frac{1}{2a_\beta}\left[ln(1+\beta_0+\eta_\beta) - ln(1-\beta_0-\eta_\beta)\right]_{-a_\beta/2}^{a_\beta/2} \\
&= \frac{1}{2a_\beta}\left[ln\left(\frac{1+\beta_0+a_\beta/2}{1+\beta_0-a_\beta/2}\right) - ln\left(\frac{1-\beta_0-a_\beta/2}{1-\beta_0+a_\beta/2}\right)\right],
\end{aligned} \tag{S.7}$$

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

$$0 \le a_\beta < 2\left(1-|\beta_0|\right). \tag{S.8}$$

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

Our Monte Carlo simulations target an PR^2 of 0.6. 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.1.

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, (47), in the Monte Carlo

Table S.1: PR^2 for parameters of Monte Carlo models

a_β	β_0	$PR^2_{ARX}(\rho_{\gamma x}=0)$	$PR^2_{ARX}(\rho_{\gamma x}=0.5)$
0	0.775	0.605	0.605
0.5	0.688	0.640	0.651
1	0.486	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.

experiments is

$$y_{it} = \alpha_i + \beta_i y_{i,t-1} + \gamma_i x_{it} + \varepsilon_{it} = \alpha_i + \boldsymbol{\beta}_i' \mathbf{x}_{it} + \varepsilon_{it} = \boldsymbol{\theta}_i' \boldsymbol{w}_{it} + \varepsilon_{it}, \quad \varepsilon_{it} \sim (0, \sigma_i^2), \tag{S.9}$$

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 higher order 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

$$\hat{y}_{i,T+1} = \hat{\alpha}_{i,T} + \hat{\boldsymbol{\beta}}_{i,T}' \boldsymbol{x}_{i,T+1} = \hat{\boldsymbol{\theta}}_{i,T}' \boldsymbol{w}_{i,T+1} \tag{S.10}$$

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

$$\hat{\boldsymbol{\theta}}_{i,T} = \left(\boldsymbol{W}_i' \boldsymbol{W}_i\right)^{-1} \boldsymbol{W}_i' \boldsymbol{y}_i. \tag{S.11}$$

The "individual" forecasts in (S.10), 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

$$\text{MSFE}_{ref} = N^{-1} \sum_{i=1}^{N} \left(y_{i,T+1} - \hat{\boldsymbol{\theta}}_{i,T}' \boldsymbol{w}_{i,T+1}\right)^2. \tag{S.12}$$

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

$$\tilde{y}_{i,T+1} = \tilde{\boldsymbol{\theta}}_{\text{pool}}' \boldsymbol{w}_{i,T+1}, \tag{S.13}$$

where

$$\tilde{\boldsymbol{\theta}}_{\text{pool}} = \left(\boldsymbol{W}' \boldsymbol{W}\right)^{-1} \boldsymbol{W} \boldsymbol{y} = \left(\sum_{i=1}^{N} \boldsymbol{W}_i' \boldsymbol{W}_i\right)^{-1} \sum_{i=1}^{N} \boldsymbol{W}_i' \boldsymbol{y}_i, \tag{S.14}$$

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

$$\hat{y}_{i,T+1}^{FE} = \hat{\alpha}_{i,\text{FE}} + \hat{\boldsymbol{\beta}}_{\text{FE}}' \boldsymbol{x}_{i,T+1}, \tag{S.15}$$

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

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

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

$$\hat{y}^{RE}_{i,T+1} = \hat{\alpha}_{\text{RE}} + \hat{\boldsymbol{\beta}}'_{\text{RE}}\boldsymbol{x}_{i,T+1} + \frac{T\hat{\sigma}^2_\eta}{T\hat{\sigma}^2_\eta + \hat{\sigma}^2_u}\bar{\widehat{\varepsilon}}_i, \tag{S.16}$$

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

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

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

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

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

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

and $\hat{\alpha}_{\mathrm{RE}} = \bar{y} - \hat{\boldsymbol{\beta}}'_{\mathrm{RE}}\bar{\boldsymbol{x}}$, 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.10) and (S.13) with weights

$$\hat{\omega}^*_{NT} = \frac{\hat{\Delta}_{NT} - T^{-1}\hat{\psi}_{NT}}{\hat{\Delta}_{NT} + T^{-1}\hat{h}_{NT} - 2T^{-1}\hat{\psi}_{NT}},$$

where $\hat{\Delta}_{NT}$, $\hat{\psi}_{NT}$, and $\hat{h}_{NT}$ are given by (40), (43) and (41).

Combination of individual and FE forecasts

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

where $\hat{y}_{i,T+1}$ and $\hat{y}_{i,T+1,\mathrm{FE}}$ are the individual and FE forecasts in (S.10) and (S.15) with the weight

$$\hat{\omega}^*_{\mathrm{FE},NT} = \frac{\hat{\Delta}^{\mathrm{FE}}_{NT} - T^{-1}\hat{\psi}^{\mathrm{FE}}_{NT}}{\hat{\Delta}^{\mathrm{FE}}_{NT} + T^{-1}\hat{h}_{NT,\beta} - 2T^{-1}\hat{\psi}^{\mathrm{FE}}_{NT}}.$$

$\hat{\Delta}^{\mathrm{FE}}_{NT}$, $\hat{\psi}^{\mathrm{FE}}_{NT}$, and $\hat{h}_{NT,\beta}$ are given by (A.26), (A.28), and (A.27).

Combination with individual weights

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

where $\hat{y}_{i,T+1}$ and $\tilde{y}_{i,T+1}$ are the individual and pooled forecasts in (S.10) and (S.13) with weights

$$\omega_i^* = \frac{\boldsymbol{w}'_{i,T+1}\hat{\boldsymbol{\Omega}}_\eta \boldsymbol{w}_{i,T+1}}{\boldsymbol{w}'_{i,T+1}(T^{-1}\hat{\sigma}_i^2 \boldsymbol{Q}_{iT}^{-1} + \hat{\boldsymbol{\Omega}}_\eta)\boldsymbol{w}_{i,T+1}}$$

where $\hat{\boldsymbol{\Omega}}_\eta = N^{-1}\sum_{i=1}^N (\hat{\boldsymbol{\theta}}_i - \bar{\hat{\boldsymbol{\theta}}})(\hat{\boldsymbol{\theta}}_i - \bar{\hat{\boldsymbol{\theta}}})'$, $\bar{\hat{\boldsymbol{\theta}}} = N^{-1}\sum_{i=1}^N \hat{\boldsymbol{\theta}}_i$, and the estimator of $\hat{\sigma}_i^2$ is given in Pesaran et al. (2022).

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

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

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

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

and $\widehat{\boldsymbol{\varepsilon}} = \boldsymbol{y}_i - \boldsymbol{W}_i\hat{\boldsymbol{\theta}}_{i,T}$ with $\hat{\boldsymbol{\theta}}_{i,T}$ given in (S.11).

Hierarchical Bayesian forecast In this supplement, we additionally 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 $r_b = 1,2,\dots,R_b$, 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 $\boldsymbol{S}_i = \left(\sigma_{r_b-1}^{-2}\boldsymbol{W}'_i\boldsymbol{W}_i + \boldsymbol{\Sigma}_{\boldsymbol{\theta},\,r_b-1}^{-1}\right)^{-1}$

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

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

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

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

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

$$\hat{y}^{HB}_{i,T+1,r_b} = \hat{\boldsymbol{\theta}}'_{i,r_b}\boldsymbol{w}_{i,T+1}, \tag{S.17}$$

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

We use the following hyperpriors: $\boldsymbol{d} = \boldsymbol{0}$, $\nu_\Sigma = K$, $\nu_\sigma = 0.1$, and $s^2 = 0.1$. For the prior covariance matrices $\boldsymbol{S}_{\bar{\theta}}$ and $\boldsymbol{S}_\Sigma$ we provide the results for three settings: (1) $\boldsymbol{S}_{\bar{\theta}} = \boldsymbol{I}_K 10^6$, $\boldsymbol{S}_\Sigma = \boldsymbol{I}_K 10$, (2) $\boldsymbol{S}_{\bar{\theta}} = \boldsymbol{I}_K 10^2$, $\boldsymbol{S}_\Sigma = \boldsymbol{I}_K 10^2$, and (3) $\boldsymbol{S}_{\bar{\theta}} = \boldsymbol{I}_K$, $\boldsymbol{S}_\Sigma = \boldsymbol{I}_K$. These are proper, weakly informative priors that avoid the use of uninformative priors that appear to be difficult to attain in hierarchical models (Gelman, 2006).

Monte Carlo results for the hierarchical Bayesian model are given in Table S.2. Since the MCMC approach to the hierarchical Bayesian model is computationally quite expensive, we restrict the Monte Carlo experiments to 1000 iterations and report some of the remaining methods as a reference. It can be seen from the table that the accuracy of the forecasts largely depends on the serendipitous choice of the prior.

Results for the applications are reported in Table S.5. 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.

Table S.2: Monte Carlo results including hierarchical Bayesian forecasts

a_β	σ^2_α	Pooled			RE			FE			Empirical Bayes			hier. Bayes: prior 1			hier. Bayes: prior 2			hier. Bayes: prior 3			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	20	50	100	20	50	100
Conditional on $\kappa_i = 0$																												
$N = 100$, $\rho_{\gamma x} = 0$																												
0.0	0.5	0.864	0.984	1.008	0.912	0.984	0.996	0.925	0.987	0.997	0.936	0.989	0.997	0.937	0.992	0.998	0.970	0.997	0.999	0.906	0.986	0.997	0.913	0.982	0.995	0.956	0.992	0.998
0.5	0.5	0.860	0.980	1.004	0.956	1.003	1.006	0.982	1.009	1.007	0.946	0.992	0.998	0.932	0.992	0.998	0.963	0.997	0.999	0.908	0.988	0.998	0.911	0.981	0.994	0.981	0.999	1.000
1.0	1.0	0.815	0.963	0.992	1.085	1.096	1.064	1.162	1.119	1.072	0.932	0.991	0.998	0.891	0.987	0.998	0.918	0.992	0.999	0.878	0.986	0.997	0.893	0.974	0.991	1.018	1.005	1.001
$N = 100$, $\rho_{\gamma x} = 0.5$																												
0.0	0.5	0.862	0.981	1.005	0.911	0.984	0.996	0.925	0.987	0.997	0.937	0.989	0.997	0.934	0.991	0.998	0.968	0.997	0.999	0.903	0.986	0.997	0.912	0.981	0.994	0.956	0.992	0.998
0.5	0.5	0.855	0.974	1.000	0.953	1.003	1.006	0.981	1.009	1.008	0.947	0.992	0.998	0.927	0.991	0.998	0.959	0.996	0.999	0.902	0.987	0.997	0.909	0.979	0.994	0.980	0.999	1.000
1.0	1.0	0.818	0.962	0.992	1.098	1.104	1.068	1.174	1.127	1.076	0.940	0.991	0.998	0.897	0.987	0.997	0.924	0.992	0.999	0.884	0.985	0.997	0.894	0.974	0.991	1.022	1.006	1.001
$N = 1000$, $\rho_{\gamma x} = 0$																												
0.0	0.5	0.857	0.981	1.006	0.904	0.984	0.996	0.918	0.987	0.996	0.934	0.989	0.997	0.901	0.986	0.997	0.932	0.992	0.998	0.879	0.983	0.996	0.908	0.981	0.995	0.951	0.992	0.998
0.5	0.5	0.847	0.975	0.998	0.935	1.001	1.004	0.966	1.007	1.006	0.939	0.992	0.998	0.895	0.987	0.997	0.919	0.990	0.998	0.881	0.985	0.997	0.904	0.979	0.993	0.972	0.998	1.000
1.0	1.0	0.837	0.965	0.990	1.044	1.072	1.050	1.117	1.094	1.058	0.941	0.990	0.997	0.893	0.985	0.997	0.904	0.986	0.997	0.890	0.985	0.997	0.900	0.975	0.991	0.998	1.000	1.000
$N = 1000$, $\rho_{\gamma x} = 0.5$																												
0.0	0.5	0.857	0.982	1.006	0.904	0.984	0.996	0.918	0.987	0.996	0.935	0.989	0.997	0.900	0.986	0.997	0.931	0.991	0.998	0.877	0.982	0.996	0.908	0.981	0.995	0.951	0.992	0.998
0.5	0.5	0.842	0.969	0.993	0.930	0.999	1.004	0.965	1.007	1.006	0.942	0.992	0.998	0.892	0.986	0.997	0.917	0.990	0.998	0.876	0.985	0.997	0.902	0.977	0.992	0.972	0.998	1.000
1.0	1.0	0.835	0.963	0.987	1.040	1.070	1.049	1.116	1.093	1.057	0.948	0.992	0.998	0.890	0.985	0.997	0.902	0.987	0.997	0.887	0.985	0.997	0.899	0.975	0.990	0.999	1.000	1.000
Conditional on $\kappa_i = \pm 1$																												
$N = 100$, $\rho_{\gamma x} = 0$																												
0.0	0.5	0.743	0.997	1.062	0.733	0.924	0.972	0.751	0.928	0.973	0.803	0.945	0.979	0.834	0.971	0.993	0.920	0.989	0.998	0.756	0.949	0.987	0.806	0.948	0.984	0.842	0.951	0.981
0.5	0.5	0.872	1.164	1.239	0.917	1.117	1.163	0.956	1.126	1.166	0.850	0.970	0.992	0.839	0.974	0.994	0.915	0.989	0.998	0.788	0.964	0.992	0.832	0.965	0.991	0.914	0.985	0.996
1.0	1.0	1.051	1.459	1.567	1.273	1.514	1.537	1.386	1.551	1.549	0.874	0.978	0.995	0.806	0.971	0.994	0.868	0.983	0.997	0.788	0.970	0.994	0.854	0.975	0.995	0.980	0.999	1.000
$N = 100$, $\rho_{\gamma x} = 0.5$																												
0.0	0.5	0.763	1.023	1.090	0.730	0.923	0.971	0.751	0.928	0.973	0.801	0.944	0.979	0.829	0.971	0.994	0.916	0.988	0.998	0.752	0.949	0.987	0.808	0.950	0.985	0.842	0.951	0.981
0.5	0.5	0.898	1.196	1.273	0.918	1.123	1.169	0.962	1.133	1.171	0.847	0.968	0.992	0.831	0.972	0.994	0.907	0.988	0.998	0.779	0.961	0.992	0.836	0.966	0.991	0.916	0.986	0.997
1.0	1.0	1.073	1.481	1.591	1.290	1.525	1.542	1.404	1.563	1.555	0.875	0.974	0.993	0.809	0.968	0.993	0.867	0.980	0.996	0.788	0.965	0.993	0.856	0.976	0.995	0.984	0.999	1.000
$N = 1000$, $\rho_{\gamma x} = 0$																												
0.0	0.5	0.746	1.001	1.065	0.732	0.925	0.971	0.753	0.929	0.972	0.810	0.947	0.979	0.759	0.951	0.987	0.833	0.972	0.993	0.720	0.933	0.978	0.804	0.949	0.984	0.840	0.952	0.980
0.5	0.5	0.880	1.188	1.265	0.910	1.140	1.191	0.956	1.151	1.194	0.835	0.965	0.990	0.772	0.959	0.990	0.819	0.969	0.993	0.758	0.957	0.989	0.829	0.965	0.991	0.908	0.986	0.997
1.0	1.0	1.120	1.516	1.619	1.323	1.583	1.616	1.456	1.628	1.632	0.856	0.969	0.991	0.788	0.962	0.991	0.804	0.964	0.991	0.788	0.962	0.991	0.863	0.978	0.995	0.976	0.998	0.999
$N = 1000$, $\rho_{\gamma x} = 0.5$																												
0.0	0.5	0.760	1.020	1.086	0.730	0.925	0.971	0.753	0.929	0.972	0.811	0.947	0.979	0.756	0.950	0.987	0.830	0.971	0.993	0.721	0.933	0.978	0.806	0.951	0.985	0.840	0.952	0.980
0.5	0.5	0.877	1.184	1.262	0.888	1.119	1.169	0.941	1.131	1.172	0.837	0.965	0.989	0.765	0.957	0.989	0.813	0.968	0.992	0.754	0.954	0.989	0.827	0.964	0.991	0.905	0.985	0.996
1.0	1.0	1.090	1.478	1.579	1.275	1.531	1.560	1.413	1.578	1.576	0.864	0.971	0.991	0.780	0.961	0.990	0.798	0.964	0.991	0.779	0.961	0.990	0.859	0.977	0.995	0.976	0.998	0.999

Notes: The hierarchical Bayesian forecasts differ in the following priors: (1) $\boldsymbol{S}_{\bar{\theta}} = \boldsymbol{I}_K 10^6$, $\boldsymbol{S}_\Sigma = \boldsymbol{I}_K 10$, (2) $\boldsymbol{S}_{\bar{\theta}} = \boldsymbol{I}_K 10^2$, $\boldsymbol{S}_\Sigma = \boldsymbol{I}_K 10^2$, and (3) $\boldsymbol{S}_{\bar{\theta}} = \boldsymbol{I}_K$, $\boldsymbol{S}_\Sigma = \boldsymbol{I}_K$. The results are based on 1000 replications. For further details see the footnote of Table 1.

S.4 Additional Monte Carlo applications and empirical results

In Section 5, we restricted our analysis to the case of $N = 100$. The results for $N = 1000$ can be found in Table S.3. The results from $N = 100$ clearly carry over and the influence of the number of cross-section units is small.

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.

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 sizable 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.

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

Table S.3: Monte Carlo results, $N = 1000$

a_β	σ^2_α	Pooled			RE			FE			Empirical Bayes			Comb. (pool)			Comb. (FE)			Comb. ω_i^*		
	T	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100	20	50	100
Conditional on $\kappa_i = 0$																						
$\rho_{\gamma x} = 0$																						
0.0	0.5	0.859	0.981	1.006	0.905	0.984	0.996	0.919	0.986	0.997	0.935	0.989	0.997	0.909	0.981	0.995	0.951	0.992	0.998	0.951	0.995	0.999
0.5	0.5	0.849	0.974	0.999	0.936	1.000	1.004	0.966	1.007	1.006	0.939	0.991	0.998	0.905	0.979	0.994	0.972	0.998	1.000	0.942	0.994	0.999
1.0	1.0	0.839	0.964	0.990	1.042	1.071	1.050	1.114	1.094	1.058	0.942	0.990	0.998	0.901	0.975	0.992	0.998	1.000	1.000	0.930	0.991	0.999
$\rho_{\gamma x} = 0.5$																						
0.0	0.5	0.859	0.981	1.006	0.905	0.984	0.996	0.919	0.986	0.997	0.936	0.989	0.997	0.909	0.981	0.995	0.951	0.992	0.998	0.953	0.995	0.999
0.5	0.5	0.845	0.968	0.993	0.931	0.999	1.004	0.965	1.007	1.006	0.943	0.992	0.998	0.903	0.977	0.993	0.972	0.998	1.000	0.947	0.994	0.999
1.0	1.0	0.837	0.962	0.988	1.039	1.070	1.049	1.114	1.094	1.057	0.949	0.991	0.998	0.900	0.974	0.991	0.998	1.000	1.000	0.935	0.992	0.999
Conditional on $\kappa_i = \pm 1$																						
$\rho_{\gamma x} = 0$																						
0.0	0.5	0.750	1.000	1.065	0.735	0.924	0.971	0.755	0.929	0.972	0.812	0.947	0.980	0.806	0.949	0.984	0.841	0.951	0.980	0.819	0.963	0.990
0.5	0.5	0.883	1.186	1.264	0.912	1.138	1.190	0.957	1.149	1.193	0.836	0.964	0.990	0.830	0.964	0.991	0.909	0.985	0.997	0.820	0.966	0.992
1.0	1.0	1.121	1.514	1.617	1.323	1.579	1.613	1.456	1.625	1.628	0.856	0.968	0.991	0.863	0.977	0.995	0.977	0.997	0.999	0.827	0.968	0.994
$\rho_{\gamma x} = 0.5$																						
0.0	0.5	0.764	1.019	1.085	0.733	0.924	0.971	0.755	0.929	0.972	0.813	0.947	0.980	0.808	0.950	0.985	0.841	0.951	0.980	0.821	0.964	0.990
0.5	0.5	0.880	1.182	1.261	0.890	1.116	1.168	0.942	1.129	1.171	0.838	0.964	0.990	0.828	0.963	0.991	0.906	0.985	0.996	0.828	0.966	0.992
1.0	1.0	1.091	1.476	1.576	1.275	1.527	1.557	1.412	1.573	1.573	0.864	0.970	0.992	0.859	0.976	0.995	0.977	0.998	0.999	0.836	0.968	0.993

Notes: The results are for $N = 1000$. For further details see Table 1.

Table S.4: Monte Carlo results for equally weighted and Oracle forecasts

a_β	σ^2_α	eq.weight(pool)			eq.weight(pool)			Oracle(pool)			Oracle(FE)		
	T	20	50	100	20	50	100	20	50	100	20	50	100
Conditional on $\kappa_i = 0$													
$N = 100$, $\rho_{\gamma x} = 0$													
0.0	0.5	0.888	0.977	0.995	0.950	0.992	0.998	0.872	0.983	0.997	0.944	0.995	0.999
0.5	0.5	0.886	0.976	0.994	0.970	0.999	1.001	0.869	0.980	0.996	0.968	0.999	1.003
1.0	1.0	0.860	0.968	0.990	1.038	1.036	1.020	0.857	0.974	0.994	1.017	1.021	1.016
$N = 100$, $\rho_{\gamma x} = 0.5$													
0.0	0.5	0.887	0.976	0.995	0.950	0.992	0.998	0.870	0.982	0.997	0.944	0.995	0.999
0.5	0.5	0.885	0.975	0.993	0.970	0.999	1.001	0.870	0.980	0.996	0.968	0.999	1.003
1.0	1.0	0.860	0.968	0.990	1.041	1.038	1.022	0.853	0.973	0.994	1.020	1.022	1.017
$N = 1000$, $\rho_{\gamma x} = 0$													
0.0	0.5	0.886	0.976	0.994	0.947	0.991	0.998	0.870	0.983	0.997	0.938	0.993	0.999
0.5	0.5	0.881	0.974	0.992	0.963	0.997	1.000	0.861	0.979	0.996	0.961	0.998	1.003
1.0	1.0	0.877	0.970	0.990	1.016	1.024	1.015	0.876	0.975	0.993	1.006	1.015	1.015
$N = 1000$, $\rho_{\gamma x} = 0.5$													
0.0	0.5	0.886	0.976	0.994	0.947	0.991	0.998	0.868	0.982	0.997	0.938	0.993	0.999
0.5	0.5	0.880	0.972	0.991	0.963	0.998	1.000	0.864	0.978	0.995	0.961	0.998	1.003
1.0	1.0	0.877	0.970	0.989	1.017	1.024	1.015	0.895	0.978	0.993	1.006	1.015	1.015
Conditional on $\kappa_i = \pm 1$													
$N = 100$, $\rho_{\gamma x} = 0$													
0.0	0.5	0.744	0.935	0.989	0.826	0.948	0.980	0.712	0.934	0.982	0.758	0.929	0.973
0.5	0.5	0.773	0.974	1.032	0.885	0.997	1.026	0.771	0.966	0.994	0.881	0.987	1.008
1.0	1.0	0.806	1.042	1.113	1.037	1.120	1.129	0.861	0.991	1.002	0.997	1.045	1.049
$N = 100$, $\rho_{\gamma x} = 0.5$													
0.0	0.5	0.745	0.939	0.994	0.826	0.948	0.980	0.719	0.939	0.984	0.758	0.929	0.973
0.5	0.5	0.778	0.981	1.040	0.888	1.000	1.028	0.782	0.972	0.996	0.884	0.989	1.009
1.0	1.0	0.810	1.048	1.119	1.043	1.125	1.131	0.847	0.989	1.002	1.003	1.048	1.051
$N = 1000$, $\rho_{\gamma x} = 0$													
0.0	0.5	0.748	0.936	0.988	0.829	0.949	0.980	0.717	0.936	0.982	0.761	0.930	0.973
0.5	0.5	0.774	0.978	1.036	0.884	1.003	1.033	0.777	0.969	0.995	0.883	0.990	1.011
1.0	1.0	0.831	1.059	1.123	1.042	1.133	1.146	0.914	1.126	1.163	0.997	1.045	1.052
$N = 1000$, $\rho_{\gamma x} = 0.5$													
0.0	0.5	0.748	0.939	0.993	0.829	0.949	0.980	0.723	0.939	0.984	0.761	0.930	0.973
0.5	0.5	0.771	0.976	1.035	0.881	0.998	1.028	0.785	0.995	1.024	0.879	0.988	1.009
1.0	1.0	0.823	1.049	1.113	1.032	1.121	1.132	0.918	1.105	1.136	0.993	1.041	1.047

Notes: The results are for equal weighted combinations of individual and pooled forecasts and for individual and FE forecasts, and for combinations using oracle weights, which use the disturbances and parameteres for the construction of the weights. For further details see the footnote of Table 1.

parameter heterogeneity than the equal-weighted combination schemes considered here.

Table S.5 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 results for the CPI application in Table S.5 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.

Table S.6 shows the results from the panel and individual DM test statistics. For both applications, 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.

S.5 Derivation of results for fixed effect estimation

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

Table S.5: Results for the applications, including hierarchical Bayesian and equal weights forecasts

	Ratio of ave. MSFE			Freq. beating benchmark			Freq. smallest MSFE			Freq. largest MSFE		
Observations	all	$\kappa_i = 0$	$\kappa_i = \pm 1$	all	$\kappa_i = 0$	$\kappa_i = \pm 1$	all	$\kappa_i = 0$	$\kappa_i = \pm 1$	all	$\kappa_i = 0$	$\kappa_i = \pm 1$
House price inflation forecasts												
Individual	2.822	2.520	3.542	–	–	–	0.003	0.218	0.127	0.569	0.235	0.376
Pooled	0.920	1.162	0.947	0.613	0.381	0.536	0.110	0.155	0.243	0.157	0.185	0.152
RE	0.924	1.166	0.960	0.619	0.376	0.528	0.099	0.022	0.044	0.003	0.017	0.008
FE	0.936	1.186	0.980	0.591	0.381	0.517	0.041	0.108	0.102	0.251	0.370	0.296
Emp.Bayes	0.901	0.955	0.881	0.942	0.519	0.652	0.077	0.066	0.061	0.003	0.039	0.019
Hier.Bayes (1)	0.940	0.973	0.939	0.928	0.536	0.677	0.014	0.050	0.017	0.000	0.030	0.014
Hier.Bayes (2)	0.978	0.987	0.977	0.953	0.528	0.657	0.014	0.039	0.019	0.000	0.039	0.036
Hier.Bayes (3)	0.908	0.969	0.915	0.914	0.511	0.627	0.265	0.122	0.097	0.000	0.017	0.041
Comb. (pool)	0.920	0.961	0.932	0.939	0.522	0.688	0.119	0.064	0.088	0.000	0.011	0.014
Comb. (FE)	0.937	0.977	0.940	0.917	0.494	0.677	0.014	0.058	0.066	0.011	0.036	0.041
indiv.weights	0.921	0.957	0.909	0.936	0.541	0.713	0.025	0.022	0.041	0.006	0.003	0.006
eq.weights (pool)	0.878	0.982	0.879	0.914	0.453	0.638	0.160	0.033	0.064	0.000	0.000	0.000
eq.weights (FE)	0.887	0.994	0.898	0.906	0.431	0.624	0.058	0.030	0.041	0.000	0.000	0.000
CPI inflation forecasts												
Individual	15.501	10.451	11.295	–	–	–	0.000	0.064	0.037	0.433	0.134	0.278
Pooled	0.878	1.013	0.971	0.444	0.374	0.417	0.193	0.112	0.091	0.396	0.406	0.374
RE	0.880	1.001	0.957	0.508	0.390	0.401	0.005	0.070	0.064	0.000	0.037	0.032
FE	0.883	0.992	0.959	0.508	0.401	0.401	0.000	0.059	0.096	0.166	0.225	0.219
Emp.Bayes	0.892	0.991	0.926	0.984	0.652	0.818	0.225	0.187	0.193	0.000	0.064	0.005
Hier.Bayes (1)	0.970	0.982	0.979	0.904	0.610	0.738	0.027	0.064	0.037	0.000	0.016	0.016
Hier.Bayes (2)	0.987	0.993	0.992	0.909	0.551	0.706	0.005	0.075	0.032	0.005	0.070	0.032
Hier.Bayes (3)	0.951	0.973	0.962	0.872	0.599	0.695	0.155	0.096	0.118	0.000	0.021	0.005
Comb. (pool)	0.930	0.987	0.953	0.733	0.481	0.572	0.037	0.064	0.064	0.000	0.016	0.016
Comb. (FE)	0.935	0.980	0.967	0.791	0.524	0.583	0.021	0.043	0.032	0.000	0.011	0.021
indiv.weights	0.897	0.972	0.931	0.973	0.695	0.813	0.128	0.102	0.118	0.000	0.000	0.000
eq.weights (pool)	0.895	0.984	0.945	0.802	0.519	0.647	0.166	0.053	0.096	0.000	0.000	0.000
eq.weights (FE)	0.900	0.982	0.942	0.856	0.535	0.647	0.037	0.011	0.021	0.000	0.000	0.000

Notes: The table reports the results for the methods in the paper and for five additional forecasts: hierarchical Bayesian forecasts for three priors: (1) $\boldsymbol{S}_{\bar{\theta}} = \boldsymbol{I}_K 10^6$, $\boldsymbol{S}_\Sigma = \boldsymbol{I}_K 10$, (2) $\boldsymbol{S}_{\bar{\theta}} = \boldsymbol{I}_K 10^2$, $\boldsymbol{S}_\Sigma = \boldsymbol{I}_K 10^2$, and (3) $\boldsymbol{S}_{\bar{\theta}} = \boldsymbol{I}_K$, $\boldsymbol{S}_\Sigma = \boldsymbol{I}_K$, the forecast that is an equal weighted average of the individual and the pooled forecasts and, finally, the forecast that is an equal weighted average of the individual and the FE forecasts. For further details see the footnote of Table 2.

Table S.6: Diebold-Mariano test statistics for equal predictive accuracy: equal weights forecasts

	hier.Bayes (1)	hier.Bayes (2)	hier.Bayes (3)	eq.weights(pool)	eq.weights(FE)
House Prices: all forecasts					
Panel DM	−29.58	−29.83	−28.00	−26.76	−25.20
DM < −1.96/DM > 1.96	202/1	207/1	186/1	158/0	148/1
CPI: all forecasts					
Panel DM	−13.15	−10.12	−17.78	−11.53	−10.76
DM < −1.96/DM > 1.96	112/8	90/3	94/8	79/14	79/9

Notes: The table reports the DM statistics for the three hierarchical Bayesian forecasts and the two equally weighted forecasts, where the first combines individual and pooled forecasts and the second individual and FE forecasts. For further details see the footnote of Table 3.

$\bar{\boldsymbol{q}}_{NT}$, $\bar{\boldsymbol{\xi}}_{NT,\beta}$ replacing $\bar{\boldsymbol{\xi}}_{NT}$, and $\bar{\bar{\boldsymbol{x}}}_{i,T+1} = \boldsymbol{x}_{i,T+1} - \bar{\boldsymbol{x}}_{iT}$ 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 $\bar{\bar{\varepsilon}}_{i,T+1}$) is not distributed independently of $(\hat{\boldsymbol{\beta}}_{\text{FE}} - \boldsymbol{\beta}_i)'\bar{\bar{\boldsymbol{x}}}_{i,T+1}$. 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

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

Also, under Assumptions 4 and 9 we have

$$\frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\beta}}_{\text{FE}} - \boldsymbol{\beta}_i\right)'\bar{\bar{\boldsymbol{x}}}_{i,T+1}\bar{\varepsilon}_{iT} = c_{NT}^{\text{FE}} + O_p(N^{-1/2}). \tag{S.18}$$

The expression for c_{NT}^{FE} simplifies somewhat by noting that under Assumption 2, $\text{E}\left(\boldsymbol{x}_{iT+1}\bar{\varepsilon}_{iT}\right) = \boldsymbol{0}$, and using Lemma A.1 we have $\bar{\boldsymbol{q}}_{NT,\beta}'\bar{\boldsymbol{Q}}_{NT,\beta}^{-1} = \bar{\boldsymbol{q}}_{N,\beta}'\bar{\boldsymbol{Q}}_{N,\beta}^{-1} + O_p\left(N^{-1/2}\right)$. Note that under Assumption 6, $\boldsymbol{\eta}_{i,\beta}$ and ε_{it} are independently distributed. Using these results, the MSFE under fixed effects estimation in (27) follows.

A comparison of forecasts based on individual and fixed effects estimates

Since,

$$\hat{e}_{i,T+1} = \bar{\bar{\varepsilon}}_{i,T+1} - \bar{\bar{\boldsymbol{x}}}_{i,T+1}'(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i). \tag{S.19}$$

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 (S.19), are no longer independently distributed and, as in the FE estimation, we need to consider the additional term

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

Using this, we have

$$
N^{-1}\sum_{i=1}^{N}\bar{\bar{\boldsymbol{x}}}_{i,T+1}'(\hat{\boldsymbol{\beta}}_i-\boldsymbol{\beta}_i)\bar{\bar{\varepsilon}}_{i,T+1}=c_{NT,\beta}+O_p(N^{-1/2}), \tag{S.20}
$$

where

$$
c_{NT,\beta}=N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[\bar{\bar{\boldsymbol{x}}}_{i,T+1}'(\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{\varepsilon}_i\bar{\varepsilon}_{iT}\right]. \tag{S.21}
$$

Taking this term into account we obtain

$$
N^{-1}\sum_{i=1}^{N}\hat{e}_{i,T+1}^{2}=N^{-1}\sum_{i=1}^{N}\bar{\bar{\varepsilon}}_{i,T+1}^{2}+h_{NT,\beta}-2c_{NT,\beta}+O_p(N^{-1/2}), \tag{S.22}
$$

where

$$
h_{NT,\beta}=N^{-1}\sum_{i=1}^{N}\mathrm{E}\left[\bar{\bar{\boldsymbol{x}}}_{i,T+1}'\boldsymbol{Q}_{iT,\beta}^{-1}\left(\frac{\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\boldsymbol{M}_T\boldsymbol{X}_i}{T}\right)\boldsymbol{Q}_{iT,\beta}^{-1}\bar{\bar{\boldsymbol{x}}}_{i,T+1}\right], \tag{S.23}
$$

and $\boldsymbol{Q}_{iT,\beta}=T^{-1}(\boldsymbol{X}_i'\boldsymbol{M}_T\boldsymbol{X}_i)$. As with the term c_{NT}^{FE} 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, $\mathrm{E}(\boldsymbol{\varepsilon}_i\bar{\varepsilon}_{iT}\,|\boldsymbol{X}_i)=(\sigma_i^2/T)\boldsymbol{\tau}_T$ and

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

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

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 and 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 (eds.) *Analysis 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., and 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.