Econometric Theory, **29**, 2013, 699–734.
doi:10.1017/S0266466612000679

MULTISTEP PREDICTION OF PANEL VECTOR AUTOREGRESSIVE PROCESSES

Ryan Greenaway-McGrevy
U.S. Bureau of Economic Analysis

This paper considers the conventional recursive (otherwise known as plug-in) and direct multistep forecasts in a panel vector autoregressive framework. We derive asymptotic expressions for the mean square prediction error (MSPE) of both forecasts as N (cross sections) and T (time periods) grow large. Both the bias and variance of the least squares fitting are manifest in the MSPE. Using these expressions, we consider the effect of model specification on predictor accuracy. When the fitted lag order (q) is equal to or exceeds the true lag order (p), the direct MSPE is larger than the recursive MSPE. On the other hand, when the fitted lag order is underspecified, the direct MSPE is smaller than the recursive MSPE. The recursive MSPE is increasing in q for all $q \geq p$. In contrast, the direct MSPE is not monotonic in q within the permissible parameter space. Extensions to bias-corrected least squares estimators are considered.

1. INTRODUCTION

The effect of model specification on the out-of-sample fit of estimators has been the subject of extensive research in econometrics and statistics. For example Shibata (1980), Kunitomo and Yamamoto (1985), Bhansali (1996,1997), Ing (2003), Ing and Wei (2005), and Schorfheide (2005), among others, derive asymptotic expressions for the mean square prediction error (MSPE) of the least squares (LS) fitting for stationary time series processes. Such expressions are useful because they can be used to design selection criteria for choosing the optimal lag order for prediction. Akaike information criteria (AIC) and final prediction error (FPE), for example, are optimal in the sense that they asymptotically minimize the MSPE of a class of autoregressions (Shibata). In the multistep horizon framework these MSPE expressions can be used not only to select between different model specifications (Bhansali, 1996), but also between different types of LS predictors (Bhansali, 1997; Ing; Schorfheide). Two commonly studied LS predictors are

The views expressed herein are those of the author and not necessarily those of the Bureau of Economic Analysis or the Department of Commerce. The author thanks Donggyu Sul, the co-editor, and three anonymous referees for their helpful suggestions and comments. Address feedback and comments to Ryan Greenaway-McGrevy, Bureau of Economic Analysis, 1441 L Street NW, Washington, D.C. 20230, USA; e-mail: ryan.greenaway-mcgrevy@bea.gov.

the “recursive” predictor, obtained from the quasi-maximum likelihood estimator (QMLE) of the forecasting regression, and the “direct” predictor, fitted by LS minimization of the multistep quadratic loss of the forecasting regression.

However, this research largely focuses on prediction of time series processes, and little, if any, research has focused on deriving expressions for the MSPE of fitted panel data regressions. Much of the extant theoretical panel research focuses on the effect of cross-sectional heterogeneity on panel MSPE while assuming strict exogeneity of the regressors (see, among others, Taub, 1979; Baltagi and Li, 1992). Although these assumptions are appropriate for a certain set of models, many important models that are used in practice fall outside this set. For example, Mark and Sul (2001) and Rapach and Wohar (2004) predict a panel of exchange rate returns using a fixed effects panel regression in which the deviation of the exchange rate from an economic fundamental is a regressor. Within each time series, the regressor is correlated with all past realizations of the regression error. Similarly, the forecasting regressions found in Baltagi and Griffin (1997), Baltagi, Griffin, and Xiong (2000), and Baltagi, Bresson, and Pirotte (2002) include lagged dependent variables in the fitted regression model, so that the regressors are correlated with past regression errors by construction. In general these panel forecasting regressions are nested within a panel vector autoregression (VAR) of the form

$$y_{i,t} = \beta_i + \sum_{s=1}^{q} \alpha_s' y_{i,t-s} + u_{i,t}, \tag{1}$$

where $y_{i,t}$ denotes an $m \times 1$ vector containing the predictand of interest and the associated regressors used in the forecasting equation; α_s is an $m \times m$ matrix of coefficients; and β_i is an $m \times 1$ vector of (cross section) fixed effect coefficients.

For a given forecasting horizon $h \geq 1$, (an element of) $\{y_{i,T+h}\}_{i=1}^{N}$ is forecast by estimating a model nested by (1) using the observed data $\{y_{i,t}\}_{i,t=1}^{N,T}$. Baltagi and Griffin (1997), Baltagi, Griffin, and Xiong (2000), and Baltagi, Bresson, and Pirotte (2002) employ a version of the recursive predictor, whereby the “one-step” model (1) is first estimated. The h-step forecast is obtained by recursively forecasting forward one step at a time, taking the predicted cross section at each step and plugging it back into the fitted model at each iteration. Mark and Sul (2001) and Rapach and Wohar (2004) use the direct forecast, whereby the fitted coefficients minimize the in-sample h-step quadratic loss of (1). This is equivalent to estimating the h-step model

$$y_{i,t} = \beta_{i,h} + \sum_{s=1}^{q} \alpha_{s,h}' y_{i,t-s-h+1} + e_{i,t+h}$$

by LS. In the h-step model the regressand is advanced h time periods to the regressors. The forecast of $\{y_{i,T+h}\}_{i=1}^{N}$ is then obtained by plugging the observed panel $\{y_{i,T-s+1}\}_{i,s=1}^{N,q}$ back into the fitted model. (The recursive and direct fittings thus only differ for multistep horizons $h \geq 2$.) Because coefficients are obtained by in-sample minimization of the multistep quadratic loss of the forecasting regression, the direct fitting is sometimes referred to as the loss-function estimator (LFE) (see, e.g., Schorfheide, 2005).

If the goal is to obtain the most accurate forecast of $\{y_{i,T+h}\}_{i=1}^{N}$ by estimating a model nested by (1), there are two key questions to be addressed. First, for a given lag order q, under what conditions should we favor the recursive forecast over the direct forecast? Following Schorfheide (2005), we refer to this as the *predictor selection* problem. In the time series context it is known that if the model is correctly specified, the recursive forecast exhibits a smaller MSPE (Bhansali, 1997; Ing, 2003). This is because the recursive predictor asymptotically exhibits less variance than the direct predictor. In contrast, when the model is misspecified (in the sense that the true lag order exceeds the fitted lag order) the direct forecast exhibits a smaller MSPE (Findley, 1983). However, results that enable a comparison of the recursive and direct fittings in the panel data context are lacking. Second, how does the fitted lag order q affect the MSPE? We refer to this issue as *model selection*. In the time series context it is known that if the true data generating process (DGP) has finite dimension, then asymptotically the MSPE-minimizing lag order is the true lag order for both the recursive and the direct predictors (Ing). Thus, attaining the most accurate out-of-sample LS forecast is the same problem as determining the true lag order of the DGP (provided the true lag order is finite). Again, it is unknown whether these results hold for panel regressions such as (1).[1] In particular, it may not be the case that the optimal q for the recursive predictor is the same as that for the direct predictor.

In this paper we address these questions by deriving asymptotic expressions for the MSPE of the recursive and direct panel predictors as both N (number of cross sections) and T (number of time periods) grow large. In empirical applications it is common for forecasts based on regressions with homogenous coefficients to outperform regressions with cross-sectionally heterogenous coefficients, provided a limited degree of heterogeneity is permitted through either fixed or random effects (see the detailed literature review in the "Forecasting applications" section of Baltagi, 2008). This result holds even when the time series dimension of the panel is large, and for this reason it is of interest to study models with homogenous slope coefficients such as (1) in a large T framework. However, we do not impose any restriction on the relative rate of growth in N relative to T, so we may permit restrictions such as $T/N \to 0$, for example. The unobserved DGP comes from the set of stationary, finite-order vector autoregressions. The asymptotic MSPEs are then used to analyze the effect of lag specification q on out-of-sample model fit, as well as to compare the relative accuracy of the recursive and direct forecasts for a given q. To ensure that differences between asymptotic MSPEs are $O(1)$, where necessary we scale the MSPEs by a function of the sample size. In the time series case the appropriate scaling factor is linear in T when the fitted lag order exceeds the true lag order of the DGP (see, among others, Bhansali, 1997; Ing, 2003). We show that in the panel context the appropriate scaling factor is proportional to the minimum of T^2 and NT, and thus depends on the size of N relative to T. We follow Ing and use "second-order MSPE" when referring specifically to these scaled MSPEs.

We demonstrate that both the finite-sample bias and the variance of the fitted predictor are potentially manifest in the scaled MSPEs.[2] This finding contrasts with the time series case, in which only the variance is manifest in the asymptotic MSPEs.[3] Whether the bias and the variance terms are manifest in the scaled MSPE depends on the size of N relative to T. If T is sufficiently large relative to N (e.g., if $N/T \to 0$), then only the variance dominates the MSPE; whereas if N and T are of comparable size, then both the bias and variance are manifest in the MSPE. If N is sufficiently large relative to T (e.g., if $T/N \to 0$), then only the bias terms and the variance terms associated with the fitted fixed effects are present in the MSPE.

Having derived asymptotic expressions for the MSPE, we address the issues of model selection and predictor selection. When we consider how model specification affects forecast accuracy, we find that the MSPE of the direct predictor is decreasing in the fitted lag order q for some regions of the permissible parameter space, even when q exceeds the true lag order of the unobserved DGP (see Theorem 3.3 and Proposition 3.3 in Section 3 below). Thus, in the multi-step panel context, estimation of the true lag order and selection of the MSPE-minimizing lag order can be two separate issues. In contrast, for the recursive forecast, the MSPE-minimizing lag order is the true lag order in all regions of the parameter space (Theorem 3.4 and Proposition 3.4). When we consider predictor selection, we show that for q exceeding the true lag order, the recursive predictor exhibits a smaller MSPE than the direct predictor (Corollary 3.2 and Proposition 3.5). We show that both the variance and the bias terms of the direct predictor are larger than the corresponding variance and bias terms of the recursive predictor, so that the result holds regardless of the size of N relative to T. When the model is misspecified (so that the true lag order exceeds the fitted lag order), the direct forecast has a smaller MSPE than the recursive forecast (Corollary 3.3 below).

The remainder of the paper is organized as follows. Section 2 contains our model and assumptions. In Section 3 we define the direct and recursive panel data predictors, derive the asymptotic expressions for the MSPE of the predictors, and compare the direct and recursive MSPE expressions. We also consider bias-corrected least squares estimators (see, e.g., Hahn and Kuersteiner, 2002) in the predictive setting in this section. Section 4 gives a small Monte Carlo study to verify our results, and Section 5 concludes. Throughout, ":=" is used as the definitional equality, and $\|\cdot\|$ denotes the spectral norm, i.e., $\|A\| = \lambda_{\max}(A'A)^{1/2}$ for matrices A, where $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue of the argument, and $\|b\|_B$ denotes the norm $\lambda_{\max}(b'Bb)^{1/2}$ for square matrices B. We use $[A : B]$ to denote the horizontal concatenation of conformable matrices A and B. The $km \times jm$ matrix $\mathbf{J}_{k,j} := [\mathbf{I}_{jm} : \mathbf{0}'_{(k-j)m\times jm}]'$ for $k \geq j$, where $\mathbf{0}_{j\times k}$ is a $j \times k$ matrix of zeros and $\mathbf{I}_k$ is the $k \times k$ identity matrix; and 1_k denotes a $k \times 1$ vector of ones. We use $N, T \to \infty$ to indicate $N \to \infty$ and $T \to \infty$ jointly. Lastly $\mathrm{tr}(B)$ and

$\det(B)$ denote the trace and the determinant, respectively, of a square matrix B. Proofs are contained in the Appendix.

2. MODEL AND ASSUMPTIONS

Consider an $m \times 1$ panel vector $y_{i,t}$ generated by a $VAR(p)$ with (cross section) fixed effects,

$$y_{i,t} = \beta_i + \sum_{s=1}^{p} \alpha_s' y_{i,t-s} + u_{i,t}. \tag{2}$$

Under the assumptions stated below, the following equivalent representation of $y_{i,t}$ will prove useful throughout the paper:

$$y_{i,t} = \mu_i + x_{i,t}, \quad \mu_i := \left(I_m - \sum_{s=1}^{p} \alpha_s'\right)^{-1} \beta_i,$$

where $x_{i,t}$ is given by

$$x_{i,t} = \sum_{s=1}^{p} \alpha_s' x_{i,t-s} + u_{i,t}. \tag{3}$$

We impose the following assumptions.

Assumption A. (i) The vector $u_{i,t}$ is $\text{iid}(0, \Sigma_u)$ with $\Sigma_u > 0$ and $\mathrm{E}\|u_{i,t}\|^{4(h+1)(1+\epsilon)} < \infty$ for some $\epsilon > 0$ and finite integer $h \geq 1$; (ii) $\det(I_m - \sum_{s=1}^{p} z^s \alpha_s) \neq 0$ for $|z| \leq 1$; (iii) p is a fixed constant independent of N and T satisfying $p \geq 1$ and at least one element of α_p is nonzero; (iv) the probabilistic process for $x_{i,t}$ begins in the infinite past; (v) $\beta_i \in \mathbb{R}^m$ for all $i = 1, \ldots, n$ and $a_s \in \mathbb{R}^{m \times m}$ for $s = 1, \ldots, p$.

Recall that h denotes the forecast horizon. The moment bounds on the error given in Assumption A(i) permit us to use a panel data version of the first moment bound theorem of Findley and Wei (1993) to derive asymptotic expressions for the squared prediction error (see, e.g., Ing, 2003; Schorfheide, 2005). Assumption A ensures that $x_{i,t}$ is a well-defined covariance stationary process. Under Assumption A(ii) we can express $x_{i,t}$ as an infinite order vector moving average process (see, e.g., Lütkepohl, 2007),

$$x_{i,t} = \sum_{s=0}^{\infty} \theta_s' u_{i,t-s}, \ \theta_s := \mathbf{J}_{p,1}' \mathbf{A}_p^s \mathbf{J}_{p,1}, \tag{4}$$

where $\mathbf{A}_p$ is defined as

$$\underset{pm \times pm}{\mathbf{A}_p} := \left[\boldsymbol{\alpha} : \mathbf{J}_{p,p-1}\right], \qquad \underset{pm \times m}{\boldsymbol{\alpha}} := \left[\alpha_1' : \alpha_2' : \cdots : \alpha_p'\right]'. \tag{5}$$

Under Assumption A(ii) the sequence of matrices $\{\theta_s\}_{s=0}^{\infty}$ satisfies the conventional absolute summability condition. That is, each element of θ_s is absolutely summable; see eq. (10.2.7) in Hamilton (1994).

We also place the following assumption on the errors.

Assumption B. The distribution of $u_{i,t}$ is uniformly Lipschitz; $\sup_{\nu'\nu=1} P[w_1 < \nu' u_{i,t} < w_2] \leq M(w_2 - w_1)^{\psi_1}$ for some scalars $\psi_1 > 0$, $\psi_2 > 0$, w_1 and w_2 satisfying $0 < w_2 - w_1 \leq \psi_2$, and finite M.

The Lipschitz condition is discussed in more detail in Findley and Wei (2002). For example, all bounded density functions satisfy the uniformly Lipschitz condition. The assumption is used to bound the moments of the LS estimators to be defined below, and it is commonly adopted in the time series literature for this purpose (see, e.g., Ing, 2003; Schorfheide, 2005; Ing and Wei, 2005).

3. MULTISTEP PREDICTION

In this section we define the direct and recursive predictors and derive asymptotic expressions for the MSPEs of both forecasts. Although our main contribution is to derive MSPEs for the panel with fixed effects $\{y_{i,t}\}_{i,t=1}^{N,T}$, as an intermediate step we begin in Section 3.1 by briefly deriving results for the panel that lacks heterogeneity $\{x_{i,t}\}_{i,t=1}^{N,T}$. In Section 3.2 we then consider the panel with fixed effects.

3.1. Panel Regressions without Heterogeneity

In this section we consider forecasting the panel $\{x_{i,t}\}_{i,t=1}^{N,T}$ defined in (3) above. The direct predictor is based on the minimization of the in-sample h-step quadratic loss of the $VAR(q)$ model, where q denotes the fitted lag order. (In general, q will be different than p, the lag order of the DGP, and q will be fixed when deriving asymptotic expressions.) The direct forecast of $x_{i,T+h}$ is defined as

$$\hat{x}^{\mathrm{D}}_{i,T+h}(q) := \textstyle\sum_{s=1}^{q} \hat{\alpha}_s'(q,h)\, x_{i,T-s+1}, \tag{6}$$

for $i = 1, \ldots, N$, where $\left\{\hat{\alpha}_s(q,h)\right\}_{s=1}^{q}$ is the LS estimator of the h-step $VAR(q)$,[4]

$$\left\{\hat{\alpha}_s(q,h)\right\}_{s=1}^{q} := \arg \min_{a_s \in \mathbb{R}^{m\times m}} \sum_{i=1}^{N} \sum_{t=q}^{T-h} \left\| x_{i,t+h} - \sum_{s=1}^{q} a_s' x_{i,t-s+1} \right\|^2. \tag{7}$$

Meanwhile, the recursive predictor is based on fitting the one-step model by LS and recursively forecasting the system up to the hth horizon. That is,

$$\begin{aligned}
\hat{x}^{\mathrm{R}}_{i,T+1}(q) &:= \textstyle\sum_{s=1}^{q} \hat{\alpha}_s'(q,1)\, x_{i,T-s+1}, \\
\hat{x}^{\mathrm{R}}_{i,T+2}(q), &:= \textstyle\sum_{s=2}^{q} \hat{\alpha}_s'(q,1)\, x_{i,T-s+2} + \hat{\alpha}_1'(q,1)\, \hat{x}^{\mathrm{R}}_{i,T+1}(q), \\
&\vdots \\
\hat{x}^{\mathrm{R}}_{i,T+h}(q) &:= \textstyle\sum_{s=1}^{q} \hat{\alpha}_s'(q,1)\, \hat{x}^{\mathrm{R}}_{i,T+h-s}(q),
\end{aligned} \tag{8}$$

where $\hat{x}^{\mathrm{R}}_{i,T+h-s}(q) := x_{i,T+h-s}$ for $s \geq h$ in (8), and $\left\{\hat{\alpha}_s(q,1)\right\}_{s=1}^{q}$ are defined by setting $h = 1$ in (7). Note that the superscript "D" in (6) is used to distinguish the direct forecast from the recursive forecast in (8), where we use the superscript "R".

We evaluate the accuracy of the predictors using MSPE. However, we follow Schorfheide (2005) and normalize the predictors by the population h-step predictor conditional on the observed $\left\{x_{i,t}\right\}_{i,t=1}^{N,T}$. The population predictor is $\boldsymbol{\alpha}'_{p,h}\mathbf{x}_{i,T}(p)$, where $\mathbf{x}_{i,t}(p) := (x'_{i,t},\ldots,x'_{i,t-p+1})'$;

$$\boldsymbol{\alpha}_{q,h} := \left(\mathbf{J}'_{p,q}\mathbf{R}_p\mathbf{J}_{p,q}\right)^{-1}\mathbf{J}'_{p,q}\mathbf{R}_p\mathbf{A}_p^h\mathbf{J}_{p,1} \qquad \text{for } q \leq p;$$

$$\boldsymbol{\alpha}_{q,h} := \left[\alpha'_{p,h} : \mathbf{0}_{m\times m} : \cdots : \mathbf{0}_{m\times m}\right]' \qquad \text{for } q > p; \tag{9}$$

and $\mathbf{R}_p := \sum_{j=0}^{\infty}\mathbf{A}_p^{\prime j}\mathbf{J}_{p,1}\Sigma_u\mathbf{J}'_{p,1}\mathbf{A}_p^j$. Note that so-defined $\mathbf{R}_p$ is the covariance matrix of $\mathbf{x}_{i,t}(p)$. We then define

$$\hat{\mathrm{L}}_{\mathrm{g}}(q,h) := \mathbb{E}\left(N^{-1}\sum_{i=1}^{N}\left\|\hat{x}^{\mathrm{g}}_{i,T+h}(q) - \boldsymbol{\alpha}'_{p,h}\mathbf{x}_{i,T}(p)\right\|^2_{W_m}\right), \qquad \mathrm{g} = \mathrm{D},\mathrm{R}, \tag{10}$$

as our measure of out-of-sample model fit. Here $\mathrm{g} = \mathrm{R}$, D denotes the recursive (R) and direct (D) predictors, respectively, and W_m denotes a weighting matrix that reflects the interest of the practitioner. For example, if the forecast of the first element of $y_{i,T+h}$ is of sole interest, W_m would have one in the upper left element and zeros elsewhere. In general, W_m is chosen such that there exists $W_m^{1/2}$, not necessarily square, such that $W_m = W_m^{1/2}W_m^{1/2\prime}$ and hence $\mathrm{tr}(W_m B) = \mathrm{tr}(W_m^{1/2\prime}BW_m^{1/2}) \geq 0$ if B is positive semidefinite.

3.1.1. Direct Predictor. Theorem 3.1 below derives asymptotic expressions for $\hat{\mathrm{L}}_D(q,h)$. We first let $\mathbf{A}_q := [\boldsymbol{\alpha}_{q,1} : \mathbf{J}_{q,q-1}]$, where $\boldsymbol{\alpha}_{q,1}$ is defined by setting $h = 1$ in (9), and

$$\mathbf{F}_{q,h} := \sum_{j=0}^{h-1}\left(W_m\theta'_j\Sigma_u\theta_j \otimes \mathbf{I}_{mq}\right) + \sum_{j=1}^{h-1}\sum_{k=1}^{j}\left(W_m\theta'_j\Sigma_u\theta_{k-1}\otimes\mathbf{A}_q^k + W_m\theta'_{k-1}\Sigma_u\theta_j\otimes\mathbf{A}_q^{\prime k}\right). \tag{11}$$

Recall that θ_j is the jth coefficient matrix in the moving average representation for $x_{i,t}$ in (4), Σ_u denotes the covariance matrix of the innovations $u_{i,t}$, and further note that $\mathbf{A}_q$ is the matrix of coefficients from the companion form representation of a $VAR(q)$ in $x_{i,t}$. We then have the following result.

THEOREM 3.1. *For $\hat{\mathrm{L}}_D(q,h)$ defined in (10) and finite $h \geq 1$, under Assumptions A and B we have*

(i) for finite integers $q \geq p$, and $f_D(q,h) := \mathrm{tr}\left(\mathbf{F}_{q,h}\right)$,

$$\lim_{N,T\to\infty} NT\left(\hat{\mathrm{L}}_D(q,h)\right) = f_D(q,h); \tag{12}$$

(ii) for integers q satisfying $1 \le q < p$, and $\xi(q,h) := (\mathbf{J}_{p,q}\boldsymbol{\alpha}_{q,h} - \boldsymbol{\alpha}_{p,h})'\mathbf{R}_p(\mathbf{J}_{p,q}\boldsymbol{\alpha}_{q,h} - \boldsymbol{\alpha}_{p,h})$,

$$\lim_{N,T\to\infty} \hat{\mathrm{L}}_D(q,h) = \mathrm{tr}(W_m \xi(q,h)).$$

3.1.2. Recursive Predictor. For the recursive forecast defined in (8) we have the following result.

THEOREM 3.2. *For $\hat{\mathrm{L}}_R(q,h)$ defined in (10) and finite $h \ge 1$, under Assumptions A and B we have*

(i) for finite integers $q \ge p$ and $f_R(q,h) := \sum_{j,k=0}^{h-1} \mathrm{tr}(W_m \theta'_{h-j-1}\Sigma_u \theta_{h-k-1} \otimes \mathbf{R}_q \mathbf{A}_q^j \mathbf{R}_q^{-1}\mathbf{A}_q^{k\prime})$,

$$\lim_{N,T\to\infty} NT\left(\hat{\mathrm{L}}_R(q,h)\right) = f_R(q,h); \tag{13}$$

(ii) for integers q satisfying $1 \le q < p$,

$$\lim_{N,T\to\infty} \hat{\mathrm{L}}_R(q,h)$$
$$= \mathrm{tr}(W_m\xi(q,h)) + \mathrm{tr}\left(W_m\left(\boldsymbol{\alpha}_{q,h} - \mathbf{A}_q^h \mathbf{J}_{q,1}\right)'\mathbf{R}_q\left(\boldsymbol{\alpha}_{q,h} - \mathbf{A}_q^h\mathbf{J}_{q,1}\right)\right).$$

Theorems 3.1 and 3.2 are similar to results obtained by Ing (2003) and Bhansali (1997) for time series autoregressive processes. Indeed, if we set $m = 1$ and $W_m = 1$, the terms $f_D(q,h)$ and $f_R(q,h)$ simplify to terms already derived by these researchers. As discussed in Bhansali (1997), the terms $f_D(q,h)$ and $f_R(q,h)$ are a manifestation of the sampling variance of the fitted predictors. To see this, for $k = 1, h$, let

$$\mathbf{V}_{q,k} := \sum_{j=0}^{k-1}\left(\theta'_j\Sigma_u\theta_j \otimes \mathbf{R}_q\right)$$
$$+ \sum_{r=1}^{k-1}\sum_{j=1}^{k-r}\left(\theta'_{j-1}\Sigma_u\theta_{j+r-1}\otimes\mathbf{R}_q\mathbf{A}_q^r + \theta'_{j+r-1}\Sigma_u\theta_{j-1}\otimes\mathbf{A}_q^{\prime r}\mathbf{R}_q\right), \tag{14}$$

so that we have $f_D(q,h) = \mathrm{tr}(\mathbf{F}_{q,h}) = \mathrm{tr}((W_m\otimes\mathbf{R}_q)(I_m\otimes\mathbf{R}_q^{-1})\mathbf{V}_{q,h}(I_m\otimes\mathbf{R}_q^{-1}))$. Now

$$\left(I_m\otimes\mathbf{R}_q^{-1}\right)\mathbf{V}_{q,k}\left(I_m\otimes\mathbf{R}_q^{-1}\right)$$
$$= \lim_{N,T\to\infty} NT\cdot\mathbb{E}\left(\mathrm{vec}\left(\hat{\boldsymbol{\alpha}}_{q,k} - \boldsymbol{\alpha}_{q,k}\right)\mathrm{vec}\left(\hat{\boldsymbol{\alpha}}_{q,k} - \boldsymbol{\alpha}_{q,k}\right)'\right),$$

so that we can directly observe the role of LS estimator variance in the second-order MSPE of the direct forecast. Similarly, for the recursive MSPE we have

$$f_R(q,h) = \sum_{j=1}^{h}\sum_{r=1}^{h}\mathrm{tr}\left(\left(W_m\otimes\mathbf{R}_q\right)\left(\theta'_{h-j}\otimes\mathbf{A}_q^{j-1}\right)\right.$$
$$\left.\times\left[\left(I_m\otimes\mathbf{R}_q^{-1}\right)\mathbf{V}_{q,1}\left(I_m\otimes\mathbf{R}_q^{-1}\right)\right]\left(\theta_{h-r}\otimes\mathbf{A}_q^{r-1\prime}\right)\right).$$

The variance of the LS estimator is manifest in the MSPE through the term in square brackets.

For panel data we have two dimensions (N and T) of the sample that can be used to estimate the parameters. In contrast to the time series case, in (12) and (13) we scale the (normalized) MSPE by NT in order to ensure that the asymptotic MSPE is $O(1)$.

Ing (2003) considers how the terms $f_D(q,h)$ and $f_R(q,h)$ vary with the lag order q for the autoregressive case (i.e., for $m = 1$). Extending these results to the VAR, we have the following.

PROPOSITION 3.1. *For all $q \geq p$, (i) $f_D(q+1,h) - f_D(q,h) = m \sum_{j=0}^{h-1} \mathrm{tr}(W_m \theta_j' \Sigma_u \theta_j) > 0$, and (ii) $f_R(q+1,h) \geq f_R(q,h)$. The inequality in (ii) is strict iff either (a) $\theta_{h-1} \neq \mathbf{0}_{m\times m}$ or (b) $\sum_{k=1}^{h-1}(\theta_{h-k-1} \otimes \mathbf{l}_{k-1}) \neq \mathbf{0}_{m^2q\times m^2}$, where the $mq \times m$ vector $\mathbf{l}_{k-1} = [\theta_{k-1}' : \cdots : \theta_0' : \mathbf{0}_{m\times m} : \cdots : \mathbf{0}_{m\times m}]'$ for $k > q$ and $\mathbf{l}_{k-1} := [\theta_{k-1}' : \cdots : \theta_{k-q}']'$ for $k \leq q$.*

For the scalar ($m = 1$) case, the conditions on the parameter space for the strict inequality in (ii) are equivalent to those given in Theorem 3(i) of Ing (2003). Condition (b) is satisfied when $q \geq h-1$ because the $(h-1)$-th $m^2 \times m^2$ block of $(\theta_0 \otimes \mathbf{l}_{h-2})$ is $(\theta_0 \otimes \theta_0) = I_{m^2}$. As stated in Ing, condition (a) and/or (b) is satisfied for $h \geq 5$ when $q \geq p$.

The second-order MSPEs for the direct and recursive predictors are increasing in the lag order q for all $q \geq p$. Ing (2003) interprets this result as support for the "strict parsimony principle" of Findley and Wei (2002) in the multistep framework. That is, specifying additional lags over and above p increases the MSPE of the forecasts.

We can use the results given in Section 3.3 below to compare the MSPEs of the direct and recursive forecasts for a given q. However, by inspection of Theorems 3.1(ii) and 3.2(ii), it is clear that the asymptotic MSPE of the direct forecast is no greater than that of the recursive when $q < p$.

3.2. Panel Regressions with Fixed Effects

We now consider forecasting panels $\{y_{i,t}\}_{i,t=1}^{N,T}$ generated according to the fixed effects panel $VAR(p)$ given in (2). The direct forecast of $y_{i,T+h}$ is defined as

$$\tilde{y}_{i,T+h}^{\mathrm{D}}(q) := \tilde{\beta}_i(q,h) + \sum_{s=1}^{q} \tilde{\alpha}_s'(q,h)\, y_{i,T-s+1}, \tag{15}$$

where the coefficients are fitted by minimization of the in-sample h-step quadratic loss. That is,

$$\left(\{\tilde{\alpha}_s(q,h)\}_{s=1}^{q}, \{\tilde{\beta}_i(q,h)\}_{i=1}^{N}\right)$$
$$:= \arg \min_{a_s \in \mathbb{R}^{m\times m}, b_i \in \mathbb{R}^m} \sum_{i=1}^{N} \sum_{t=q}^{T-h} \left\| y_{i,t+h} - \sum_{s=1}^{q} a_s' y_{i,t-s+1} - b_i \right\|^2. \tag{16}$$

The fitted VAR coefficients $\{\tilde{\alpha}_s(q,h)\}_{s=1}^{q}$ and $\{\tilde{\beta}_i(q,h)\}_{i=1}^{N}$ correspond to equation-by-equation LS estimation of m separate h-step regressions. Thus the $\{\tilde{\beta}_i(q,h)\}_{i=1}^{N}$ are $m \times 1$ vectors of fitted fixed effects for each equation. As in (8), the recursive predictor is based on fitting a one-step model by LS and recursively forecasting the system up to the hth horizon. The recursive forecast $\tilde{y}_{i,T+h}^{\mathrm{R}}(q)$ is defined as

$$\tilde{y}_{i,T+j}^{\mathrm{R}}(q) := \tilde{\beta}_i(q,1) + \sum\nolimits_{s=1}^{q} \tilde{\alpha}_s'(q,1)\,\tilde{y}_{i,T+j-s}^{\mathrm{R}}(q), \qquad j = 1,\ldots,h, \tag{17}$$

where $\tilde{y}_{i,T+j-s}^{\mathrm{R}}(q) := y_{i,T+j-s}$ for $s \geq j$, and $\{\tilde{\alpha}_s(q,1)\}_{s=1}^{q}$ and $\{\tilde{\beta}_i(q,1)\}_{i=1}^{N}$ are defined by setting $h = 1$ in (16).

Before proceeding, note that $\{y_{i,T+h}\}_{i=1}^{N}$ and the data used in estimation of the regression model, namely $\{y_{i,t}\}_{i,t=1}^{N,T}$, are produced by the same process. This is conventionally referred to as "same-sample realization" (e.g., Ing and Wei, 2005), and it contrasts against the "independent realization" assumption, whereby the variable to be forecast is independent of the data used to estimate the regression model (see, among others, Shibata, 1980). While independent realization is rather unrealistic, it is mathematically convenient, and in the time series context it is largely innocuous in the sense that the second-order approximations of the asymptotic MSPE are identical under either framework (Ing, 2003). However, it is unknown if such an assumption would likewise be innocuous in the panel data context. We therefore work within the same-sample realization framework.

As in the previous subsection, accuracy is assessed by the MSPE of the forecast relative to the population h-step predictor based on p lags. Provided that $T \geq p$, the population predictor is $\mu_i + \boldsymbol{\alpha}_{p,h}'\mathbf{x}_{i,T}(p)$, and the normalized MSPE is

$$\tilde{\mathrm{L}}_{\mathrm{g}}(q,h) := \mathbb{E}\left(\frac{1}{N}\sum\nolimits_{i=1}^{N}\left\|\tilde{y}_{i,T+h}^{\mathrm{g}}(q) - \mu_i - \boldsymbol{\alpha}_{p,h}'\mathbf{x}_{i,T}(p)\right\|_{W_m}^{2}\right), \qquad \mathrm{g} = \mathrm{R},\ \mathrm{D}, \tag{18}$$

where $\mathrm{g} = \mathrm{R}, \mathrm{D}$ denotes the recursive (R) and direct (D) predictors, respectively. We also require an appropriate scaling factor that, when multiplied by the (normalized) MSPE, ensures that the scaled MSPE is $O(1)$ when $q \geq p$. Therefore we define δ_{NT} as a sequence of positive integers satisfying

$$\delta_{NT}^2\left((NT)^{-1} + T^{-2}\right) \to \delta, \qquad \text{where } 0 < \delta < \infty, \tag{19}$$

as both $N \to \infty$ and $T \to \infty$. Sequences δ_{NT} satisfying (19) ensure that $\delta_{NT}^2(l_1(NT)^{-1} + l_2T^{-2})$ is finite and bounded above zero for any finite scalars $l_1 > 0$ and $l_2 > 0$. We also define

$$\kappa_N := \lim_{N,T\to\infty}(NT)^{-1}\delta_{NT}^2, \qquad \kappa_T := \lim_{N,T\to\infty} T^{-2}\delta_{NT}^2$$

so that δ_{NT} satisfying (19) implies that $0 < (l_1\kappa_N + l_2\kappa_T) < \infty$.

3.2.1. Direct Predictor. Theorem 3.3 below derives the asymptotic expressions for the MSPE of the direct multistep predictor with fixed effects given in (15). First, it is instructive to define the $mq \times m$ matrix

$$\mathbf{B}_q := \mathbf{R}_q^{-1} \left(1_q \otimes \left(I_m - \sum_{s=1}^{p} \alpha_s' \right)^{-1} \right) \Sigma_u, \tag{20}$$

and the scalars

$$a(h) := \operatorname{tr}\left(W_m \Lambda_h' \Sigma_u \Lambda_h \right), \tag{21}$$

$$b(q,h) := \operatorname{tr}\left(W_m \Lambda_h' \mathbf{B}_q' \mathbf{R}_q \mathbf{B}_q \Lambda_h \right), \tag{22}$$

$$c(h) := \operatorname{tr}\left(W_m \sum_{j=1}^{h-1} \left(\Lambda_j' \Sigma_u \Lambda_j + \left(\Lambda_h - \Lambda_j \right)' \Sigma_u \left(\Lambda_h - \Lambda_j \right) \right) \right), \tag{23}$$

$$d(q,h) := \operatorname{tr}\left(W_m \sum_{j=1}^{h-1} \left(\Lambda_h' \mathbf{B}_q' \mathbf{A}_q^{j-1\prime} \mathbf{J}_{q,1} \Sigma_u \left(\Lambda_h - \Lambda_j \right) \right) \right), \tag{24}$$

where $\Lambda_j := \sum_{k=0}^{j-1} \theta_k$ for $j \geq 1$. (In the remarks following Theorem 3.3 below we give an interpretation of these terms.) We then have the following result.

THEOREM 3.3. *For $\tilde{\mathrm{L}}_D(q,h)$ defined in (18) and finite $h \geq 1$, under Assumptions A and B we have*

(i) for finite integers $q \geq p$, and $g_D(q,h) := b(q,h) + 2d(q,h) + c(h) + a(h)q$,

$$\lim_{N,T\to\infty} \delta_{NT}^2 \left(\tilde{\mathrm{L}}_D(q,h) - T^{-1} a(h) \right) = \kappa_N f_D(q,h) + \kappa_T g_D(q,h); \tag{25}$$

(ii) for integers q satisfying $1 \leq q < p$, $\lim_{N,T\to\infty} \tilde{\mathrm{L}}_D(q,h) = \operatorname{tr}(W_m \xi(q,h))$.

Under misspecification (i.e., $q < p$) we obtain the same result as that given in Theorem 3.1 above. However, when $q \geq p$ the results differ.

First note that for $q \geq p$ we subtract $T^{-1}a(h)$ from $\mathrm{L}_D(q,h)$ before scaling by δ_{NT}^2. The term $T^{-1}a(h)$ is independent of the lag order q, hence we must consider the smaller order terms when analyzing the effect of lag specification on MSPE. (Note that the terms on the right-hand side of (25) are functions of q and thus are informative when considering how changes in q affect MSPE.) In addition, these smaller order terms are required to compare the direct and recursive forecasts, because the direct and recursive MSPE share identical $O(T^{-1})$ terms. See Theorem 3.4 below.

Consider the terms on the right-hand side of (25). The term $f_D(q,h)$ also comprises the MSPE in the panel forecast without heterogeneity (see Theorem 3.1 above). As discussed in Section 3.1 above, this term is a manifestation of the variance of the LS fitting. In the following remarks we discuss the additional terms $g_D(q,h)$ and $a(h)$. In order to do so, it is instructive to re-express the forecast as

$$\tilde{y}_{i,T+h}^{\mathrm{D}}(q) = \tilde{\boldsymbol{\alpha}}_{q,h}' \mathbf{x}_{i,T}(q) + \tilde{\mu}_i(q,h), \tag{26}$$

where $\tilde{\boldsymbol{\alpha}}_{q,h} := [\tilde{\alpha}_1'(q,h) : \cdots : \tilde{\alpha}_q'(q,h)]'$ and $\{\tilde{\mu}_i(q,h)\}_{i=1}^N$ are obtained from[5]

$$\left(\{\tilde{\alpha}_s(q,h)\}_{s=1}^q, \{\tilde{\mu}_i(q,h)\}_{i=1}^N\right)$$
$$:= \arg \min_{a_s \in \mathbb{R}^{m\times m}, d_i \in \mathbb{R}^m} \sum_{i=1}^N \sum_{t=q}^{T-h} \left\| y_{i,t+h} - \sum_{s=1}^q a_s' x_{i,t-s+1} - d_i \right\|^2. \quad \textbf{(27)}$$

Here (26) holds for any finite $q \geq 1$. This re-expression of the forecast is also used in Sims, Stock, and Watson (1990) and Ing (2003) in the time series context, and it is easily generalized to the panel context.

Remark 1. The terms $a(h)$ and $c(h)$ that partially comprise $g_D(q,h)$ in (25) result from the variance in the estimated fixed effects $\{\tilde{\mu}_i(q,h)\}_{i=1}^N$. To see this, when $q \geq p$ the prediction error can be expressed as

$$\tilde{y}^{\mathrm{D}}_{i,T+h}(q) - \mu_i - \boldsymbol{\alpha}_{p,h}'\mathbf{x}_{i,T}(p) = (\tilde{\mu}_i(q,h) - \mu_i) + \left(\tilde{\boldsymbol{\alpha}}_{q,h} - \boldsymbol{\alpha}_{q,h}\right)' \mathbf{x}_{i,T}(q). \quad \textbf{(28)}$$

Variance in $\{\tilde{\mu}_i(q,h)\}_{i=1}^N$ affects the prediction error through the first term in (28). For $e_{i,t+h} := \sum_{r=1}^h \theta_{h-r} u_{i,t+r}$, this term can be decomposed further as

$$\tilde{\mu}_i(q,h) - \mu_i = \frac{1}{T-h-q+1} \sum\nolimits_{t=q}^{T-h} e_{i,t+h}$$
$$- \frac{1}{T-h-q+1} \sum\nolimits_{t=q}^{T-h} \sum\nolimits_{s=1}^q (\tilde{\alpha}_s(q,h) - \alpha_s(q,h))' x_{i,t-s+1}. \quad \textbf{(29)}$$

We define $\{\alpha_s(q,h)\}_{s=1}^q$ such that $\boldsymbol{\alpha}_{q,h}[\alpha_1'(q,h):\ldots:\alpha_q'(q,h)]'$. Consider the first term. Because $e_{i,t+h}$ is i.i.d. mean-zero, $\frac{1}{T-h-q+1}\mathbb{E}(e_{i,t+h}e_{i,t+h}')$ partially comprises the variance of $\{\tilde{\mu}_i(q,h)\}_{i=1}^N$, and consequently $\mathbb{E}\|\frac{1}{T-h-q+1}\sum_{t=q}^{T-h} e_{i,t+h}\|_{W_m}^2$ will be manifest in the MSPE. Then under the assumptions of the theorem it can be shown that for all i,

$$\mathbb{E}\left\| \frac{1}{T-h-q+1} \sum\nolimits_{t=q}^{T-h} e_{i,t+h} \right\|_{W_m}^2 = \frac{a(h)}{T} + \frac{c(h) + qa(h)}{T^2} + O\left(\frac{1}{T^3}\right).$$

The first-order effect is the $O(T^{-1})$ term in the MSPE, namely $T^{-1}a(h)$. Intuitively, $\{\mu_i(q,h)\}_{i=1}^N$ are estimated using a single times series for each cross section, so the estimators exhibit $O(T^{-1})$ variance. The term $T^{-1}a(h)$ is subtracted from the MSPE when constructing the second-order MSPE in (25), and the smaller order term $T^{-2}(c(h) + qa(h))$ is manifest in the second-order MSPE as $\kappa_T(c(h) + qa(h))$. Note that $c(h)$ satisfies $c(1) = 0$, so the term $c(h)$ is irrelevant for one-step forecasts.

Remark 2. It is well-established that LS estimators of dynamic panel data models exhibit $O_p(T^{-1})$ bias (Nickell, 1981; Hahn and Kuersteiner, 2002, 2011;

Phillips and Sul, 2007). Intuitively, the bias arises because the regressors are correlated with the past innovations and because the fixed effects are partialed out by removing the average from each time series in the panel. To illustrate the bias, we can use Lemma A.3(i)(a) in the Appendix to show that for $q \geq p$,

$$\lim_{N,T\to\infty} T\cdot\left(\mathbb{E}\left(\tilde{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}\right)\right) = -\mathbf{R}_q^{-1}\left(1_q\otimes\left(I_m-\sum_{s=1}^{p}\alpha_s'\right)^{-1}\right)\Sigma_u\Lambda_h$$
$$= -\mathbf{B}_q\Lambda_h, \tag{30}$$

where $\mathbf{B}_q$ is defined in (20). Thus $-\frac{1}{T}\mathbf{B}_q\Lambda_h$ is the $O(T^{-1})$ approximation of the bias for the h-step LS estimator $\tilde{\boldsymbol{\alpha}}_{q,h}$. The term simplifies to $-\frac{1}{T}\mathbf{B}_q$ when $h=1$, and if in addition $p=q=m=1$, the bias approximation is $-\frac{1}{T}(1+\alpha_1)$, which is the canonical Nickell-bias for the fixed effects $AR(1)$ model.

Remark 3. The terms $b(q,h)$ and $d(q,h)$ that partially comprise $g_D(q,h)$ in (25) are manifestations of the Nickell-bias of the fitted model. Note that the Nickell-bias in $\tilde{\boldsymbol{\alpha}}_{q,h}$ will affect both terms on the right-hand side of (28) since $\{\tilde{\mu}_i(q,h)\}_{i=1}^N$ are a function of the elements of $\tilde{\boldsymbol{\alpha}}_{q,h}$. Thus, to analyze the effect of the bias on the MSPE, it is instructive to substitute (29) into (28) so that the forecast error can be expressed as

$$\tilde{y}^{\mathrm{D}}_{i,T+h}(q)-\mu_i-\boldsymbol{\alpha}'_{p,h}\mathbf{x}_{i,T}(p) = \bar{e}_{i,h}(q)+\left(\tilde{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}\right)'\left(\mathbf{x}_{i,T}(q)-\bar{\mathbf{x}}_{i,h}(q)\right), \tag{31}$$

for $\bar{\mathbf{x}}_{i,h}(q) := \frac{1}{T-q-h+1}\sum_{t=q}^{T-h}\mathbf{x}_{i,t}(q)$ and $\bar{e}_{i,h}(q) := \frac{1}{T-q-h+1}\sum_{t=q}^{T-h}e_{i,t+h}$. The Nickell-bias becomes manifest in the MSPE through the second term in the above expression. Using the decomposition $\tilde{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h} = \left(\tilde{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}+\frac{1}{T}\mathbf{B}_q\Lambda_h\right)-\frac{1}{T}\mathbf{B}_q\Lambda_h$, we have

$$\mathbb{E}\left\|\left(\tilde{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}\right)'\left(\mathbf{x}_{i,T}(q)-\bar{\mathbf{x}}_{i,h}(q)\right)\right\|^2_{W_m}$$
$$= \mathbb{E}\left\|\frac{1}{T}\Lambda_h'\mathbf{B}_q'\left(\mathbf{x}_{i,T}(q)-\bar{\mathbf{x}}_{i,h}(q)\right)\right\|^2_{W_m}$$
$$+\mathbb{E}\left\|\left(\tilde{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}+\frac{1}{T}\mathbf{B}_q\Lambda_h\right)'\left(\mathbf{x}_{i,T}(q)-\bar{\mathbf{x}}_{i,h}(q)\right)\right\|^2_{W_m}+o\left(\delta_{NT}^{-2}\right).$$

The first term on the right-hand side of the equation thereby captures the effect of the Nickell-bias on the second-order MSPE. Recall that $\mathbf{R}_q$ denotes the covariance matrix of the regressors $\mathbf{x}_{i,T}(q)$ in the $VAR(q)$, so that as $N,T\to\infty$,

$$\delta^2_{NT}\mathbb{E}\left\|\frac{1}{T}\Lambda_h'\mathbf{B}_q'\left(\mathbf{x}_{i,T}(q)-\bar{\mathbf{x}}_{i,h}(q)\right)\right\|^2_{W_m}\to\kappa_T\,\mathrm{tr}(W_m\Lambda_h'\mathbf{B}_q'\mathbf{R}_q\mathbf{B}_q\Lambda_h) = \kappa_T b(q,h),$$

showing how the bias affects the MSPE through the $b(q,h)$ term. Incidentally, we also have

$$\delta_{NT}^2 \mathbb{E}\left\| \left(\tilde{\boldsymbol{\alpha}}_{q,h} - \boldsymbol{\alpha}_{q,h} + \frac{1}{T}\mathbf{B}_q \Lambda_h \right)' \left(\mathbf{x}_{i,T}(q) - \bar{\mathbf{x}}_{i,h}(q) \right) \right\|_{W_m}^2 \to \kappa_N f_D(q,h).$$

See Lemma A.4 in the Appendix for a more detailed derivation of these results. In addition, Lemma A.6 shows that the Nickell-bias also affects the MSPE through the cross product of the two terms in (31), giving rise to the $d(q,h)$ term in $g_D(q,h)$. That is, it is likewise demonstrable that as $N,T \to \infty$,

$$\delta_{NT}^2 \mathbb{E}\left(\operatorname{tr}\left(W_m \left(\tilde{\boldsymbol{\alpha}}_{q,h} - \boldsymbol{\alpha}_{q,h} \right)' \left(\mathbf{x}_{i,T}(q) - \bar{\mathbf{x}}_{i,h}(q) \right) \bar{e}_{i,h}(q) \right) \right) \to \kappa_T d(q,h).$$

Note that $d(q,h)$ satisfies $d(q,1) = 0$, so the term is irrelevant for one-step forecasts.

The $O(T^{-2})$ terms in the MSPE associated with the bias and variance resultant from including fixed effects in the regression will be present in the second-order MSPE expressions when the scaling factor δ_{NT}^2 satisfies $\delta_{NT}^2 = O(T^2)$ (i.e. when N is at least of the same order of magnitude as T). In Section 3.4 below we will consider other estimators that have asymptotic MSPE that lack the bias terms.

The term $g_D(q,h)$ in the second-order MSPE satisfies

PROPOSITION 3.2. $g_D(q,h) > 0$ *for all* $h \geq 1$ *and* $q \geq 1$.

While it is not the main focus of the present paper, we can consider the case where the true DGP has no fixed effects ($\beta_i = \mathbf{0}_{m\times 1}$ for all i in (2)), but the practitioner includes fixed effects in the fitted regression. Unless $\kappa_T = 0$, which occurs for example if $N/T \to 0$, Proposition 3.2 implies that including fixed effects in the fitted regression model when they are lacking in the DGP unambiguously increases the second-order direct MSPE when $q \geq p$.

As in the previous subsection, we consider how the terms on the right-hand side of (25) change with respect to the fitted lag order q. Proposition 3.1 above has already shown that $f_D(q,h)$ is increasing in $q \geq p$. The following remarks consider how the remaining components change with the lag order q.

PROPOSITION 3.3. *For all* $q \geq p$, *it holds that (i)*

$$b(q+1,h) - b(q,h) = \operatorname{tr}\left(W_m \Lambda_h' \Sigma_u \Lambda_h \right), \tag{32}$$

such that $b(q+1,h) > b(q,h)$; *and (ii)*

$$d(q+1,h) - d(q,h) = \sum_{j=1}^{h-1} \operatorname{tr}\left(W_m \Lambda_h' \Sigma_u \theta_j - W_m \Lambda_h' \mathbf{B}_q' \mathbf{A}_q'^{\,j-1} \mathbf{J}_{q,1} \Sigma_u \theta_j \right),$$

such that $d(q+1,h) \lesseqgtr d(q,h)$.

Proposition 3.3(ii) means that $d(q,h)$ is not monotonic in q across the permissible parameter space. In fact, in some regions of the permissible parameter space, the decrease in $2d(q,h)$ resultant from an increase in q can be greater than the increase in the other terms that comprise $g_D(q,h)$, so that overall $g_D(q,h)$ is decreasing in q. The following example demonstrates this.

Example 3.1
Consider a panel AR(1) with fixed effects; $y_{i,t} = \rho y_{i,t-1} + u_{i,t} + \beta_i$, $u_{i,t} \sim iid(0,\sigma_u^2)$. Let the forecast horizon $h = 2$. Consider $d(q,h)$ as the fitted lag order increases from 1 to 2. We have $d(2,2) - d(1,2) = -\rho^2(1+\rho)\sigma_u^2$, which is negative for all permissible ρ. Meanwhile, by (32) we have $b(2,2) - b(2,1) = \sigma_u^2(1+\rho)^2$, and $a(2) = (1+\rho)^2\sigma_u^2$, so that

$$g_D(2,2) - g_D(2,1) = 2\sigma_u^2\left((1+\rho)^2 - \rho^2(1+\rho)\right).$$

The above expression is negative for all $\rho \in (-1, \frac{1}{2} - \frac{1}{2}\sqrt{5})$, and positive for all $\rho \in (\frac{1}{2} - \frac{1}{2}\sqrt{5}, 1)$.

As the above example shows, the strict parsimony principle of Findley and Wei (2002) does not hold across the entire permissible parameter space for direct multistep forecasts. Due to the bias in the fitted model, increasing the lag order may in fact increase the out-of-sample accuracy. This contrasts against the findings of Bhansali (1997) and Ing (2003), who show that in the time series case the optimal lag order for the direct predictor is p, the true lag order. However this result only holds for $h \geq 2$. Note that $d(q,1) = c(1) = 0$, and thus we have the following.

COROLLARY 3.1. *Setting $h = 1$ in Theorem 3.3, we have*

$$\lim_{N,T\to\infty} \delta_{NT}^2\left(\tilde{\mathrm{L}}_D(q,1) - T^{-1}a(1)\right) = (\kappa_N + \kappa_T)a(1)q + \kappa_T b(q,1).$$

Note that under Proposition 3.3(i), $b(q+1,1) - b(q,1) > 0$ for all $q \geq p$, and thus for the one-step forecast, we see that adding in lags $q \geq p$ unambiguously increases the second-order MSPE.

3.2.2. Recursive Predictor. Theorem 3.4 below derives the asymptotic expressions for the MSPE of the recursive predictor (17) with fixed effects. However, it is instructive to first define the scalar

$$k(q,h) := \mathrm{tr}\left(W_m \sum_{j=1}^h \sum_{k=1}^h \theta_{k-1}' \mathbf{B}_q' \mathbf{A}_q'^{h-k} \mathbf{R}_q \mathbf{A}_q^{h-j} \mathbf{B}_q \theta_{j-1}\right). \tag{33}$$

We then have the following result.

THEOREM 3.4. *For $\tilde{\mathrm{L}}_R(q,h)$ defined in (18) and finite $h \geq 1$, under Assumptions A and B we have*

(i) for finite integers $q \geq p$, and $g_R(q,h) := qa(h) + k(q,h)$,

$$\lim_{N,T\to\infty} \delta_{NT}^2\left(\tilde{\mathrm{L}}_R(q,h) - T^{-1}a(h)\right) = \kappa_N f_R(q,h) + \kappa_T g_R(q,h); \tag{34}$$

(ii) for integers q satisfying $1 \le q < p$,

$$\lim_{N,T\to\infty} \tilde{\mathrm{L}}_R(q,h) = \operatorname{tr}(W_m \xi(q,h)) + \operatorname{tr}\left(W_m \left(\boldsymbol{\alpha}_{q,h} - \mathbf{A}_q^h \mathbf{J}_{q,1}\right)' \mathbf{R}_q \left(\boldsymbol{\alpha}_{q,h} - \mathbf{A}_q^h \mathbf{J}_{q,1}\right)\right).$$

For $q \ge p$ we again subtract the term $T^{-1}a(h)$ from $\tilde{\mathrm{L}}_R(q,h)$ before scaling by δ_{NT}^2. The term $a(h)$ is independent of the lag order q, hence as above we consider the smaller $O((NT)^{-1})$ and $O(T^{-2})$ terms of the MSPE. The following remarks discuss these terms in more detail. However, it is first instructive to express the recursive forecast as

$$\tilde{y}_{i,T+h}^{\mathrm{R}}(q) = \mathbf{J}_{q,1}' \tilde{\mathbf{A}}_q^{h\prime} \mathbf{x}_{i,T}(q) + \sum_{j=1}^{h-1} \mathbf{J}_{q,1}' \tilde{\mathbf{A}}_q^j \mathbf{J}_{q,1}(\tilde{\mu}_i(q,1) - \mu_i) + \tilde{\mu}_i(q,1) \quad \textbf{(35)}$$

for $q \ge 1$ and $h \ge 2$, where $\tilde{\mathbf{A}}_q := \left[\tilde{\boldsymbol{\alpha}}_{q,1} : \mathbf{J}_{q,q-1}\right]$. Note that $\tilde{\boldsymbol{\alpha}}_{q,1} := [\tilde{\alpha}_1'(q,1) : \cdots : \tilde{\alpha}_q'(q,1)]'$ and $\tilde{\mu}_i(q,1)$ are defined by setting $h = 1$ in (27) above.

Remark 4. The term $a(h)$ in (34) is a manifestation of the variance of the estimated fixed effects $\{\tilde{\mu}_i(q,1)\}_{i=1}^N$. Note that by using the expression for $\tilde{\mu}_i(q,1) - \mu_i$ in (29), it is clear that $\frac{1}{T-q}\mathbb{E}(u_{i,t+1}u_{i,t+1}') = \Sigma_u$ partially comprises the variance of $\{\tilde{\mu}_i(q,1)\}_{i=1}^N$. To see how $\{\tilde{\mu}_i(q,1)\}_{i=1}^N$ affects the prediction error of the h-step forecast, we subtract $\mu_i + \boldsymbol{\alpha}_{p,h}' \mathbf{x}_{i,T}(p)$ from (35) to obtain

$$\tilde{y}_{i,T+h}^{\mathrm{R}}(q) - \mu_i - \boldsymbol{\alpha}_{p,h}' \mathbf{x}_{i,T}(p) = \mathbf{J}_{q,1}' \left(\tilde{\mathbf{A}}_q^h - \mathbf{A}_q^h\right)' \mathbf{x}_{i,T}(q) + \sum_{j=0}^{h-1} \mathbf{J}_{q,1}' \tilde{\mathbf{A}}_q^{j\prime} \mathbf{J}_{q,1}(\tilde{\mu}_i(q,1) - \mu_i) \quad \textbf{(36)}$$

for $q \ge p$. The variance of $\{\tilde{\mu}_i(q,1)\}_{i=1}^N$ affects the prediction error through the second term. Now under the assumptions of the theorem $\mathbb{E}\|\mathbf{J}_{q,1}' \tilde{\mathbf{A}}_q^j \mathbf{J}_{q,1} - \theta_j\|$ is sufficiently small for each $j = 1, \ldots, h-1$, so that we can replace $\mathbf{J}_{q,1}' \tilde{\mathbf{A}}_q^j \mathbf{J}_{q,1}$ with θ_j and obtain

$$\mathbb{E}\left\| \Lambda_h' \frac{1}{T-q} \sum_{t=q}^{T-1} u_{i,t+1} \right\|_{W_m}^2 = \frac{a(h)}{T} + \frac{q a(h)}{T^2} + O\left(\frac{1}{T^3}\right).$$

The first-order effect on the MSPE is $T^{-1}a(h)$. Because the term $T^{-1}a(h)$ is subtracted from the MSPE when constructing the second-order MSPE in (34), the term $T^{-2} q a(h)$ is manifest in the second-order MSPE as $\kappa_T q a(h)$.

Remark 5. Nickell-bias in the fitted $VAR(q)$ is manifest in the second-order asymptotic MSPE through the term $k(q,h)$. As stated above under the remarks to Theorem 3.3, $-\frac{1}{T}\mathbf{B}_q$ is an approximation of the bias for $h = 1$ up to and including $O(T^{-1})$ terms when $q \ge p$. The bias will affect the point estimates $\tilde{\boldsymbol{\alpha}}_{q,1}$ as well as $\{\mu_i(q,1)\}_{i=1}^{N}$. Thus, to demonstrate how the Nickell-bias affects the MSPE, after some manipulation we can re-express the prediction error (36) as

$$\begin{aligned}\tilde{y}^{\mathrm{R}}_{i,T+h}(q) - \mu_i - \boldsymbol{\alpha}'_{p,h}\mathbf{x}_{i,t}(p) \\ = \Lambda'_h \bar{e}_{i,1}(q) + \Big[\big(I_m \otimes \mathbf{x}_{i,T}(q)\big)'\mathbf{L}_{q,h} - \big(\Lambda_h \otimes \bar{\mathbf{x}}_{i,1}(q)\big)' \\ - \big(I_m \otimes \mathbf{J}_{q,1}\bar{e}_{i,1}(q)\big)'\mathbf{M}_{q,h}\Big] \operatorname{vec}\big(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1}\big) + \upsilon_{NT},\end{aligned} \qquad \textbf{(37)}$$

where $\mathbf{L}_{q,h} := \sum_{j=0}^{h-1}(\theta'_j \otimes \mathbf{A}_q^{h-j-1})$, $\mathbf{M}_{q,h} := \sum_{j=1}^{h-1}\mathbf{L}_{q,j}$, $\bar{e}_{i,1}(q) := \frac{1}{T-q}\sum_{t=q}^{T-1} u_{i,t+1}$, and υ_{NT} is a smaller order term that can be ignored when deriving the MSPE. We refer the reader to Section A.3 in the Appendix for a more detailed derivation of (37). The Nickell-bias thereby affects the prediction error through the second term on the right-hand side of (37). Lemma A.7 in the Appendix shows that this term satisfies

$$\begin{aligned}\mathbb{E}\Big\|\Big[\big(I_m \otimes \mathbf{x}_{i,T}(q)\big)'\mathbf{L}_{q,h} - \big(\Lambda_h \otimes \bar{\mathbf{x}}_{i,1}(q)\big)' - \big(I_m \otimes \mathbf{J}_{q,1}\bar{e}_{i,1}(q)\big)'\mathbf{M}_{q,h}\Big] \\ \times \operatorname{vec}\big(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1}\big)\Big\|^2_{W_m} = \mathbb{E}\big\|\big(I_m \otimes \mathbf{x}'_{i,T}(q)\big)\mathbf{L}_{q,h}\operatorname{vec}\big(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1}\big)\big\|^2_{W_m} \\ + o\left(\delta_{NT}^{-2}\right).\end{aligned}$$

Using $\big(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1}\big) = \left(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1} + \frac{1}{T}\mathbf{B}_q\right) - \frac{1}{T}\mathbf{B}_q$, we have

$$\begin{aligned}\mathbb{E}\big\|\big(I_m \otimes \mathbf{x}'_{i,T}(q)\big)\mathbf{L}_{q,h}\operatorname{vec}\big(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1}\big)\big\|^2_{W_m} \\ = \mathbb{E}\left\|\frac{1}{T}\big(I_m \otimes \mathbf{x}'_{i,T}(q)\big)\mathbf{L}_{q,h}\operatorname{vec}\big(\mathbf{B}_q\big)\right\|^2_{W_m} \\ + \mathbb{E}\left\|\big(I_m \otimes \mathbf{x}'_{i,T}(q)\big)\mathbf{L}_{q,h}\operatorname{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1} + \frac{1}{T}\mathbf{B}_q\right)\right\|^2_{W_m} + o\left(\delta_{NT}^{-2}\right).\end{aligned}$$

The first term captures the effect of Nickell-bias on the second-order MSPE. As $N, T \to \infty$,

$$\delta_{NT}^2\,\mathbb{E}\left\|\frac{1}{T}\big(I_m \otimes \mathbf{x}'_{i,T}(q)\big)\mathbf{L}_{q,h}\operatorname{vec}\big(\mathbf{B}_q\big)\right\|^2_{W_m} \to \kappa_T k(q,h),$$

showing how the bias affects the MSPE. Meanwhile, the second term satisfies

$$\mathbb{E}\left\|\big(I_m \otimes \mathbf{x}'_{i,T}(q)\big)\mathbf{L}_{q,h}\operatorname{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1} + \frac{1}{T}\mathbf{B}_q\right)\right\|^2_{W_m} \to \kappa_N f_R(q,h).$$

See Lemma A.7 in the Appendix for a more detailed derivation of this result. In addition, Lemma A.9 shows that the cross products of the terms in (37) scaled by δ_{NT}^2 are zero in the limit.

Remark 6. When $h = 1$ the direct and recursive forecasts are equivalent. Accordingly, the definitions of $b(q,h)$ and $k(q,h)$ satisfy $b(q,1) = k(q,1)$ and thus $g_R(q,1) = g_D(q,1)$.

We now consider how $g_R(q,h)$ behaves as q is altered.

PROPOSITION 3.4. *For all $q \geq p$ it holds that $k(q+1,h) - k(q,h) + a(h) > 0$.*

Because $g_R(q,h) = qa(h) + k(q,h)$, Proposition 3.4 means that $g_R(q+1,h) - g_R(q,h) > 0$ for all $q \geq p$. Thus, adding in additional lags in addition to p increases the second-order MSPE of the recursive forecast. In contrast to the direct forecast, the recursive forecast therefore obeys the strict parsimony principle of Findley and Wei (2002).

3.3. Comparing Predictors

Much extant research has compared the MSPE of the direct and recursive multistep forecasts. For example, Ing (2003) and Bhansali (1997) show that for a given $q \geq p$, the recursive forecast has an asymptotic MSPE less than that of the direct predictor once the difference in MSPE is scaled by T, where T denotes the time series dimension of the sample used in estimation. Meanwhile, it has long been known that when $q < p$ the direct predictor has a smaller MSPE than the recursive in the limit (see, e.g., Findley, 1983).

For the multistep panel data forecasts considered herein, we have the following. From (25) and (34) we can decompose the difference in the scaled MSPEs into two terms.

COROLLARY 3.2. *Under Assumptions A and B, for $\tilde{\mathrm{L}}_D(q,h)$ and $\tilde{\mathrm{L}}_R(q,h)$ defined in (18) and finite $q \geq p$ and $h \geq 1$,*

$$\lim_{N,T\to\infty} \delta_{NT}^2 \left(\tilde{\mathrm{L}}_D(q,h) - \tilde{\mathrm{L}}_R(q,h)\right) = \kappa_N \left(f_D(q,h) - f_R(q,h)\right) + \kappa_T \left(g_D(q,h) - g_R(q,h)\right). \tag{38}$$

Bhansali (1997) compares $f_D(q,h)$ and $f_R(q,h)$ for autoregressive processes (i.e., $m = 1$ and $W_m = 1$). We extend this result to the VAR case in part (i) of the following proposition, while part (ii) compares $g_D(q,h)$ and $g_R(q,h)$.

PROPOSITION 3.5. *Let $h \geq 2$. Then (i) the terms $f_D(q,h)$ and $f_R(q,h)$ satisfy*

$$f_D(q,h) - f_R(q,h) = \sum_{l=1}^{h-1} \sum_{k,j=0}^{h-l-1} \mathrm{tr}\left(W_m \theta'_{h-k-l-1} \Sigma_u \theta_{h-j-l-1} \otimes \mathbf{A}_q'^{k} \mathbf{J}_{q,1} \Sigma_u \mathbf{J}'_{q,1} \mathbf{A}_q^{j} \mathbf{R}_q^{-1}\right) > 0;$$

and (ii) the terms $g_D(q,h)$ and $g_R(q,h)$ satisfy $g_D(q,h) - g_R(q,h) > 0$.

Under Proposition 3.5 the terms on the right-hand side of (38) are strictly positive across the entire permissible parameter space, since at least one of κ_N or κ_T is bounded above zero for δ_{NT} defined in (19). Thus, Corollary 3.2 together with Proposition 3.5 implies that when $q \geq p$, the direct forecast has a larger second-order asymptotic MSPE than the recursive predictor when there are fixed effects in the forecasting regression, regardless of the relative size of N and T. Although the same ordering of the MSPEs holds in the time series context, the ordering holds in the panel context for a different reason. Namely, both the bias and the variance terms in the MSPE of the direct predictor exceed the bias and the variance terms of the recursive predictor.

Following Ing (2003), we also consider whether the terms on the right-hand side of (38) are monotonic in h. Ing shows that $(f_D(q,h) - f_R(q,h))$ is increasing in h for the $m=1$ case, meaning that the advantage of the recursive predictor over the direct predictor is increasing with the forecast horizon h for all $q \geq p$. The result is easily generalized to the $VAR(p)$ case using the expression for $(f_D(q,h) - f_R(q,h))$ given in Proposition 3.5(i). That is,

$$\begin{aligned}&(f_D(q,h+1) - f_D(q,h)) - (f_R(q,h+1) - f_R(q,h))\\&\quad = \sum_{j=0}^{h-1}\sum_{k=0}^{h-1} \operatorname{tr}\left(W_m\theta'_{h-j-1}\Sigma_u\theta_{h-k-1} \otimes \mathbf{A}_q'^{j}\mathbf{J}_{q,1}\Sigma_u\mathbf{J}'_{q,1}\mathbf{A}_q^k\mathbf{R}_q^{-1}\right) \geq 0.\end{aligned}$$

However, there is no corresponding ordering of $g_D(q,h) - g_R(q,h)$ in h across the permissible parameter space. Thus, although the recursive MSPE is smaller than the direct MSPE when $q \geq p$, the difference in the MSPEs is not monotonic in h unless we impose further restrictions (e.g., if $\kappa_T = 0$, then as in Ing the difference in the MSPEs is monotonic in h).

When $q < p$, both MSPEs are $O(1)$ in the limit, so we need not scale the MSPEs. The following corollary compares the MSPEs for $q < p$.

COROLLARY 3.3. *Under Assumptions A and B, for* $\tilde{\mathrm{L}}_D(q,h)$ *and* $\tilde{\mathrm{L}}_R(q,h)$ *defined in (18) and for integers q satisfying* $1 \leq q < p$ *and finite* $h \geq 1$*,*

$$\begin{aligned}&\lim_{N,T\to\infty}(\tilde{\mathrm{L}}_R(q,h) - \tilde{\mathrm{L}}_D(q,h))\\&\quad = \operatorname{tr}\left(W_m^{1/2\prime}\left(\boldsymbol{\alpha}_{q,h} - \mathbf{A}_q^h\mathbf{J}_{q,1}\right)'\mathbf{R}_q\left(\boldsymbol{\alpha}_{q,h} - \mathbf{A}_q^h\mathbf{J}_{q,1}\right)W_m^{1/2}\right) \geq 0.\end{aligned}$$

The direct predictor has an MSPE no greater than that of the recursive predictor in the limit. The parameter $\boldsymbol{\alpha}_{q,h}$, defined in (9) above, attains the minimum out-of-sample squared prediction error among linear predictors with q lags. It is straightforward to show that $\|\tilde{\boldsymbol{\alpha}}_{q,h} - \boldsymbol{\alpha}_{q,h}\| = o_p(1)$, whereas the recursive predictor vector satisfies $\|\tilde{\mathbf{A}}_q^h\mathbf{J}_{q,1} - \boldsymbol{\alpha}_{q,h}\| = O_p(1)$.

3.4. Extension to Bias-Corrected Predictors

The use of analytic expressions for the LS bias such as (30) to eliminate or attenuate the bias has been the focus of much of the recent panel literature (see, e.g., Kiviet, 1995; Hahn and Kuersteiner, 2002, 2011; and Phillips and Sul, 2007). From the prediction perspective, these "bias-corrected" fittings offer potential reductions in the asymptotic MSPE of both forecasts, since, as discussed above, the bias of the fitted model is manifest in the asymptotic MSPEs.

In this subsection we consider whether asymptotic MSPE expressions can be derived for bias-corrected LS predictors within the framework of the model outlined in Section 2. Derivation of MSPE expressions requires that certain higher-order moments of the fitted model parameters are bounded. (For example, the one-step LS estimator, used for the recursive forecast, must have bounded $2(h+1)$-th moments, whereas the h-step LS estimator, used for the direct forecast, must have bounded 4th moments.) We can show that bias-corrected LS fittings likewise satisfy necessary moment bounds for a subset of models nested by the VAR in (2)—namely, autoregressive (AR) models—without strengthening the assumptions of the data generating process. This is because the $O(T^{-1})$ approximation of the LS bias in the AR model can be simplified to a linear function of the parameter vector $\boldsymbol{\alpha}$. For example, in the $AR(1)$ case the $O(T^{-1})$ bias approximation is $-\frac{1}{T-1}(1+\alpha_1)$, which is linear in α_1 (Nickell, 1981). An associated bias-corrected LS estimator in this case is $\tilde{\alpha}_1(1,1)+\frac{1+\tilde{\alpha}_1(1,1)}{T-1}$ (Hahn and Kuersteiner, 2002). In general we consider a bias-corrected LS estimator of the $AR(q)$ model of the form $\ddot{\boldsymbol{\alpha}}_{q,1} := \tilde{\boldsymbol{\alpha}}_{q,1}+\frac{1}{T-q}(1_q+\tilde{\Psi}_q)$, where $\tilde{\Psi}_q := (\tilde{\psi}_1,\ldots,\tilde{\psi}_q)'$ with $\tilde{\psi}_1 := \tilde{\alpha}_q(q,1)$ and $\tilde{\psi}_j := \tilde{\psi}_{j-1}+\tilde{\alpha}_{q-j+1}(q,1)-\tilde{\alpha}_{j-1}(q,1)$ for $j\geq 2$. Because $\tilde{\Psi}_q$ is a linear transformation of $\tilde{\boldsymbol{\alpha}}_{q,1}$, we can bound moments of the bias correction $\frac{1}{T-q}(1_q+\tilde{\Psi}_q)$ using bounds already established for the fitted LS estimator vector $\tilde{\boldsymbol{\alpha}}_{q,1}$.

For the direct forecast we require an analogous bias correction for the h-step LS estimator. To the best knowledge of the author, a bias-corrected multistep LS estimator has not yet been proposed in the extant literature. By inspection of (30), the h-step LS bias is postmultiplied by $\Lambda_h=\sum_{j=0}^{h-1}\theta_j$, so a bias-corrected h-step LS estimator requires an estimate of Λ_h. We consider a bias-corrected estimator of the form $\ddot{\boldsymbol{\alpha}}_{q,h} := \tilde{\boldsymbol{\alpha}}_{q,h}+\frac{1}{T-q-h+1}(1_q+\tilde{\Psi}_q)\sum_{j=0}^{h-1}\tilde{\theta}_j$, where $\tilde{\theta}_j=\sum_{s=0}^{j-1}\tilde{\theta}_s\tilde{\alpha}_{j-s,1}(q,1)$ and $\tilde{\theta}_0=1$. Thus the bias correction for the proposed h-step LS estimator is based solely on the estimated parameters from the one-step model, $\tilde{\boldsymbol{\alpha}}_{q,1}=[\tilde{\alpha}_1'(q,1):\cdots:\tilde{\alpha}_q'(q,1)]'$.[6]

For both the direct and recursive predictors, the bias-corrected fixed effects are $\ddot{\beta}_i(q,k) := \tilde{\beta}_i(q,k)+\frac{1}{T-k-q+1}\sum_{j=0}^{h-1}\tilde{\theta}_j(1_q+\tilde{\Psi}_q)'\sum_{t=q}^{T-k}y_{i,t}$ for $k=1,h$. The bias-corrected direct predictor, denoted $\ddot{y}_{i,T+h}^{\mathrm{D}}(q)$, is defined as above in (15) with $\ddot{\beta}_i(q,h)$ and $\ddot{\boldsymbol{\alpha}}_{q,h}$ replacing $\tilde{\beta}_i(q,h)$ and $\tilde{\boldsymbol{\alpha}}_{q,h}$, respectively, while the bias-corrected recursive predictor, denoted $\ddot{y}_{i,T+h}^{\mathrm{R}}(q)$, is defined as above in (17) with $\ddot{\beta}_i(q,1)$ and $\ddot{\boldsymbol{\alpha}}_{q,1}$ replacing $\tilde{\beta}_i(q,1)$ and $\tilde{\boldsymbol{\alpha}}_{q,1}$, respectively. Then, noting that $m=1$ in this case, for the MSPE defined as

$$\ddot{\mathrm{L}}_g(q,h) := \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\ddot{y}^{g}_{i,T+h}(q) - \mu_i - \boldsymbol{\alpha}'_{p,h}\mathbf{x}_{i,T}(p)\right)^2\right), \; g = \mathrm{R, D}, \qquad \textbf{(39)}$$

we have the following.

THEOREM 3.5. *For* $\ddot{\mathrm{L}}_D(q,h)$ *defined in (39) and finite* $h \geq 1$, *under Assumptions A and B we have*

(i) for finite integers $q \geq p$, $\lim_{N,T\to\infty}\delta^2_{NT}\ddot{\mathrm{L}}_D(q,h) = \kappa_N f_D(q,h) + \kappa_T(qa(h)+c(h))$;

(ii) for integers q satisfying $1 \leq q < p$, $\lim_{N,T\to\infty}\ddot{\mathrm{L}}_D(q,h) = \xi(q,h)$.

THEOREM 3.6. *For* $\ddot{\mathrm{L}}_R(q,h)$ *defined in (39) and finite* $h \geq 1$, *under Assumptions A and B we have*

(i) for finite integers $q \geq p$, $\lim_{N,T\to\infty}\delta^2_{NT}\ddot{\mathrm{L}}_R(q,h) = \kappa_N f_R(q,h) + \kappa_T qa(h)$;

(ii) for integers q satisfying $1 \leq q < p$,

$$\lim_{N,T\to\infty}\ddot{\mathrm{L}}_R(q,h) = \xi(q,h) + \left(\boldsymbol{\alpha}_{q,h} - \mathbf{A}^h_q\mathbf{J}_{q,1}\right)'\mathbf{R}_q\left(\boldsymbol{\alpha}_{q,h} - \mathbf{A}^h_q\mathbf{J}_{q,1}\right).$$

Remark 7. Because $k(q,h) > 0$, we see that for $q \geq p$ the second-order asymptotic MSPE of the bias-corrected recursive LS predictor is less than that of the conventional LS predictor for asymptotic sequences under which κ_T is nonzero. However, comparing Theorems 3.3(i) and 3.5(i), there is no corresponding ordinal ranking of the direct least squares and bias-corrected least squares MSPE for a given $h \geq 2$ and $q \geq p$, meaning that the bias correction does not necessarily reduce the MSPE of the direct least squares predictor. This is because $b(q,h) + 2d(q,h)$ can be negative in some regions of the parameter space when $h \geq 2$. For example, in the $AR(1)$ Example 3.1 above, $b(q,h) + 2d(q,h)$ is negative for all $\rho \in (-1, 1-\sqrt{2})$. (This is not that surprising, since the bias in the fitted AR(1) model approaches zero as $\rho \to 1$.) However, since the bias correction clearly improves forecast accuracy for $\rho > 0$ in this example, the bias correction is nonetheless advisable for the many empirical settings in which serial correlation is positive.

Remark 8. Since $f_R(q,h) > f_D(q,h)$ and $c(h) > 0$, the bias-corrected recursive MSPE is smaller than the bias-corrected direct MSPE in all regions of the parameter space when $q \geq p$. For $q < p$, the bias-corrected direct predictor has an MSPE no greater than that of the recursive predictor in the limit.

Remark 9. In contrast to the conventional direct LS predictor, the bias-corrected LS direct predictor obeys a parsimony principle in all regions of the parameter space, since $qa(h)$ is weakly increasing in q (recall $a(h) \geq 0$), and since $f_D(q,h)$ is increasing in q (see Proposition 3.1 above.) The bias-corrected recursive predictor likewise obeys a parsimony principle.

4. MONTE CARLO STUDY

We conduct a small Monte Carlo study in order to verify the expressions derived for the asymptotic MSPEs under Theorems 3.3 through 3.6 as well as the ordinal rankings of the MSPEs implied by Propositions 3.3, 3.4, and 3.5. As demonstrated in Theorems 3.3 and 3.4, when $q \geq p$ the second-order LS MSPEs are comprised of two terms, namely $\kappa_N f_{\mathrm{g}}(q,h)$ and $\kappa_T g_{\mathrm{g}}(q,h)$ for $\mathrm{g} = \mathrm{D}, \mathrm{R}$. We consider two different sequences of N and T in order to isolate each term in the simulated MSPEs. Under the first sequence $T/N \to 0$, so that the $g_{\mathrm{g}}(q,h)$ term comprises the MSPE in the limit, while under the second sequence $N/T \to 0$, so that the $f_{\mathrm{g}}(q,h)$ term comprises the MSPE. Our DGP is nested by an $AR(2)$ with standard normal errors; $y_{i,t} = \beta_i + \rho_1 y_{i,t-1} + \rho_2 y_{i,t-2} + u_{i,t}$, $u_{i,t} \sim iidN(0,1)$ and $\beta_i \sim iidN(0,1)$. We consider fitted lag orders $q = 1,2,3,4$.

Simulations with T/N $\to$ 0. Here we set $N = T^2$, so that $\kappa_N = 0$, and consider $T = 25, 50, 100, 200$. All simulations are replicated 4,000 times. In order to verify Proposition 3.3, we set $\rho_1 = -0.8$, $\rho_2 = 0$, and $h = 2$, since, as illustrated in Example 3.1 above, with these parameters the MSPE of the direct forecast can be increasing in the fitted lag order q. (Note that in this case the DGP is an $AR(1)$.) Table 1 exhibits the simulated scaled MSPEs of the predictors, as well as the asymptotic second-order MSPEs in bold. (We set the scaling factor $\delta_{NT}^2 = T^2$ so that the asymptotic MSPEs for the recursive and direct LS predictors are $g_R(q,h)$ and $g_D(q,h)$, respectively). Under this asymptotic sequence, the bias of the LS estimator is manifest in the MSPEs of the LS predictors. For comparison we therefore also report the MSPE of the bias-corrected least squares

TABLE 1. h-step MSPEs in a large N panel ($N = T^2$ and $h = 2$)

	Recursive forecast (least squares)					Direct forecast (least squares)				
lags (q)	T = 25	T = 50	T = 100	T = 200	∞	T = 25	T = 50	T = 100	T = 200	∞
1	0.45	0.39	0.36	0.34	**0.32**	2.05	1.82	1.72	1.67	**1.62**
2	1.05	0.87	0.80	0.75	**0.72**	2.11	1.74	1.59	1.52	**1.44**
3	1.27	1.02	0.91	0.85	**0.80**	2.43	1.91	1.71	1.62	**1.52**
4	1.53	1.17	1.02	0.95	**0.88**	2.76	2.09	1.84	1.72	**1.60**
	Recursive forecast (bias-corrected)					Direct forecast (bias-corrected)				
1	0.15	0.09	0.07	0.06	**0.04**	2.11	1.88	1.78	1.73	**1.68**
2	0.25	0.16	0.12	0.10	**0.08**	2.41	2.03	1.87	1.80	**1.72**
3	0.38	0.24	0.18	0.15	**0.12**	2.73	2.18	1.97	1.86	**1.76**
4	0.52	0.32	0.24	0.20	**0.16**	3.09	2.35	2.06	1.93	**1.80**

Notes: The DGP is an AR(2) with $(\rho_1, \rho_2) = (-0.8, 0)$. Table entries in standard font are the average of the simulated second-order MSPEs of the forecast $\hat{y}_{i,T+h}(q)$, namely $T^2 \times \left[\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_{i,T+h}(q) - y_{i,T+h} - \sum_{j=1}^{h}\theta_{h-j} u_{i,T+j})^2 - \frac{1}{T}a(h)\right]$, and table entries in bold are the asymptotic second-order MSPEs of the forecast.

TABLE 2. *h*-step MSPEs in a large N panel ($N = T^2$ and $h = 3$)

a. Fitted $AR(q)$ with $p \geq q$

	Recursive forecast (least squares)					Direct forecast (least squares)				
q (lags)	$T=25$	$T=50$	$T=100$	$T=200$	∞	$T=25$	$T=50$	$T=100$	$T=200$	∞
2	4.99	5.03	5.07	5.01	**4.94**	13.69	12.97	12.62	12.35	**12.10**
3	9.56	9.27	9.16	9.00	**8.87**	20.24	18.57	17.72	17.25	**16.75**
4	13.74	13.11	12.78	12.54	**12.29**	26.90	23.92	22.47	21.71	**20.91**
q (lags)	Recursive forecast (bias-corrected)					Direct forecast (bias-corrected)				
2	4.52	4.43	4.40	4.31	**4.21**	9.11	8.40	8.10	7.87	**7.66**
3	7.14	6.77	6.62	6.47	**6.31**	12.20	10.96	10.42	10.07	**9.76**
4	9.94	9.23	8.89	8.65	**8.41**	15.61	13.62	12.79	12.31	**11.87**

b. Fitted $AR(q)$ with $p < q$

	Recursive forecast (least squares)					Direct forecast (least squares)				
q (lags)	$T=25$	$T=50$	$T=100$	$T=200$	∞	$T=25$	$T=50$	$T=100$	$T=200$	∞
1	0.16	0.11	0.09	0.08	**0.07**	0.10	0.05	0.02	0.01	**0.00**
	Recursive forecast (bias-corrected)					Direct forecast (bias-corrected)				
1	0.18	0.12	0.09	0.08	**0.07**	0.09	0.04	0.02	0.01	**0.00**

Notes: The DGP is an AR(2) with$(\rho_1, \rho_2) = (0.5, -0.3)$. For part (a), table entries in standard font are the average simulated second-order MSPEs of the forecast $\hat{y}_{i,T+h}(q)$, namely $T^2\ \left[\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_{i,T+h}(q) - y_{i,T+h} - \sum_{j=1}^{h}\theta_{h-j}u_{i,T+j})^2 - \frac{1}{T}a(h)\right]$, and table entries in bold are the asymptotic second-order MSPEs of the forecast. For part (b), table entries in standard font are the average simulated MSPEs of the forecast $\hat{y}_{i,T+h}(q)$, namely $\frac{1}{N}$ $\sum_{i=1}^{N}(\hat{y}_{i,T+h}(q) - y_{i,T+h} - \sum_{j=1}^{h}\theta_{h-j}u_{i,T+j})^2$, and table entries in bold are the asymptotic MSPEs of the forecast.

predictors introduced in Section 3.4. (The asymptotic MSPEs for the recursive and direct bias-corrected least-squares predictors are $qa(h)$ and $qa(h) + c(h)$, respectively.)

The MSPE of the recursive forecast increases with q, in accordance with Proposition 3.4. For the direct predictor, as q increases from 1 to 2 the MSPE decreases, provided $T \geq 50$; as stated above, the direct predictor does not obey a parsimony principle in the sense of Findley and Wei (2002). In all simulations the direct MSPE exceeds the recursive MSPE, in accordance with Proposition 3.5(ii). Note that for a given q and T, the bias-corrected recursive predictors exhibit a smaller MSPE than the conventional LS recursive predictors. In contrast, both the simulated and the asymptotic MSPEs of the bias-corrected direct predictor are greater than the conventional LS direct predictor. As discussed in Section 3.4 above, the bias correction does not necessarily cause a reduction in the asymptotic MSPE of the direct predictor across the entire parameter space. Last, the absolute difference between the average simulated MSPEs and the asymptotic MSPEs is monotonically decreasing in T, in accordance with the theorems presented above.

Next we consider whether the expressions for the asymptotic MSPE hold for longer forecast horizons. We consider an AR(2) DGP that exhibits positive serial dependence. Because the DGP follows an AR(2), we can compare the performance of the direct and iterative forecasts under misspecification by fitting an AR(1) model. Table 2(a) presents results from simulations with $\rho_1 = 0.5$, $\rho_2 = -0.3$, $h = 3$, for $q = 2, 3, 4$. (The fitted lag orders are therefore equal to or greater than the true lag order. The case of $q = 1$ is addressed in Table 2(b).) Again the direct MSPE exceeds the recursive MSPE, in accordance with Proposition 3.5(ii). However, note that under this DGP both the direct and the recursive least squares MSPEs are increasing in q. In contrast to the $AR(1)$ DGP considered in Table 1, here we see that the bias correction reduces the MSPE of both the recursive and the direct LS predictors, both asymptotically and in the finite sample. As in Table 1, the absolute difference between the average simulated MSPEs and the asymptotic MSPEs is monotonically decreasing as T increases. Comparing the asymptotic MSPEs for this DGP to those given in Table 1, we can see that the longer forecast horizon and the positive serial dependence in the DGP have the effect of increasing the MSPEs.

Table 2(b) presents the MSPEs for the same DGP as given in panel (a), except the fitted lag order $q = 1$ so that the fitted model is misspecified. Note that the direct LS MSPE is less than that of the recursive for all T, in accordance with Corollary 3.3. (Similarly, the bias-corrected direct MSPE is less than that of the bias-corrected recursive MSPE.)

Simulations with N/T → 0. Here we set $T = N^2$ and consider $N = 5, 10, 20, 40$. We set $\rho_1 = 0.5$, $\rho_2 = -0.3$ (so that the fitted model is misspecified when $q = 1$), and $h = 2$. All simulations are replicated 40,000 times. (Compared to the $T/N \to 0$ simulations above, additional replications were necessary in order for the empirical averages to converge.) Because T is larger than N, the effect of estimator bias is negligible in these simulations, and hence the bias-corrected MSPEs are similar to conventional LS MSPEs. We therefore omit reporting the bias-corrected MSPEs in the interest of brevity.

Table 3(a) exhibits the scaled MSPEs for fitted lag orders $q = 2, 3, 4$, with the asymptotic MSPEs given in bold. We set the scaling factor $\delta_{NT}^2 = NT$ so that the asymptotic MSPEs for the recursive and direct LS predictors are exactly $f_R(q,h)$ and $f_D(q,h)$, respectively. Evidently the second-order MSPE of the direct predictor is greater than that of the recursive predictor, in accordance with Proposition 3.5(i). Both the direct and recursive MSPEs are increasing in q, in accordance with Proposition 3.1.

Table 3(b) displays the MSPEs of each predictor when $q = 1$, so all fitted models are misspecified. Evidently the direct MSPE is less than the recursive MSPE, in accordance with Corollary 3.3.

TABLE 3. h-step MSPEs in a large T panel ($T = N^2$ and $h = 2$)

	Recursive forecast (least squares)					Direct forecast (least squares)				
q (lags)	$N=5$	$N=10$	$N=20$	$N=40$	∞	$N=5$	$N=10$	$N=20$	$N=40$	∞
a. Fitted $AR(q)$ with $p \geq q$										
2	4.04	2.82	2.42	2.22	**2.09**	6.15	4.15	3.52	3.23	**3.00**
3	6.46	4.42	3.77	3.48	**3.25**	9.50	6.09	5.01	4.61	**4.25**
4	9.77	6.34	5.25	4.86	**4.50**	13.28	7.98	6.54	6.00	**5.50**
b. Fitted AR(1) with $p < q$										
1	0.21	0.13	0.12	0.11	**0.11**	0.15	0.05	0.03	0.03	**0.02**

Notes: The DGP is an AR(2) with$(\rho_1, \rho_2) = (0.5, -0.3)$. For part (a), table entries in standard font are the average simulated second-order MSPEs of the forecast $\hat{y}_{i,T+h}(q)$, namely $NT\ \left[\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_{i,T+h}(q) - y_{i,T+h} - \sum_{j=1}^{h}\theta_{h-j}u_{i,T+j})^2 - \frac{1}{T}a(h)\right]$, and table entries in bold are the asymptotic second-order MSPEs of the forecast. For part (b), table entries in standard font are the average simulated MSPEs of the forecast $\hat{y}_{i,T+h}(q)$, namely $\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_{i,T+h}(q) - y_{i,T+h} - \sum_{j=1}^{h}\theta_{h-j}u_{i,T+j})^2$, and table entries in bold are the asymptotic MSPEs of the forecast.

5. CONCLUDING REMARKS

Panel forecasting is increasingly common in applied econometrics. However, expressions for the out-of-sample fit of many important panel models have not yet been derived. In this paper we begin to address this gap by deriving asymptotic expressions for the multistep MSPE of a fixed effects panel VAR. We consider both the multistep LS fitting, yielding the so-called direct forecast, and the one-step LS fitting, yielding the recursive forecast. We use these expressions to study how the out-of-sample fit of the models varies with the fitted lag order q, as well as to compare the relative performance of the direct and recursive forecasts.

Several findings prove interesting. First, for fitted lag orders $q \geq p$, the bias of the LS estimator transmutes into the MSPE unless T is sufficiently large relative to N (e.g. if $N/T \to 0$). This is in contrast to the time series context, in which only the variance of the LS estimator is manifest in the MSPE. Our findings imply that selecting the best lag order for either predictor can be difficult when N is moderate or large relative to T. In particular, selection criteria should penalize bias (see Lee, 2012, for more discussion). Second, for $h \geq 2$ the direct predictor does not adhere to a parsimony principle (e.g., Findley and Wei, 2002); specifying additional lags beyond the true lag order can reduce the MSPE of the direct predictor in some regions of the parameter space. This contrasts with the time series case, in which the MSPE-minimizing lag order is the true lag order for the direct predictor (Ing, 2003). For the recursive predictor, the MSPE-minimizing lag order is the true lag order. Last, for a given q there is an ordering of the direct and recursive MSPE across the permissible parameter space. When $q \geq p$, the recursive forecast has a smaller MSPE than the direct forecast, whereas when $q < p$, the direct forecast has an MSPE no larger than that of the recursive forecast.

Several extensions to this research appear promising. First, the DGP had a fixed lag order in the asymptotic analysis, meaning that for $h = 1$ there is no trade-off between the effects of specification error (decreasing in q) and the bias and variance of the fitted model (both increasing in q) on the MSPE. (For example, the optimal lag order is p when $h = 1$.) In order to study such a trade-off, we could consider an infinite order DGP and have the maximum permitted lag order of the fitted model increase with the sample size at a certain rate (Shibata, 1980). Based on this analysis, model selection criteria can be designed to minimize the expected MSPE of the model. Second, we assume that the regression errors are cross section independent. Although cross section independence is a common assumption elsewhere in panel research on forecasting (e.g., Taub, 1979; Baltagi and Li, 1992), it is often too strong for many panel data sets used in practice. In the absence of an a priori ordering of the cross sections, such as that offered by a spatial autoregressive model, extending our results to consider crosssectionally correlated panels could be achieved by including a parametric structure of dependence in the regression equation. For example, Pesaran (2006), Bai (2009), and Greenaway-McGrevy, Han, and Sul (2012) incorporate a latent factor model into a linear panel regression model. We leave these interesting extensions for future research.

NOTES

1. Ing (2003) also gives detailed results for the time series autoregression when the true lag order is infinite but the fitted model has finite lag order q. In this case the direct MSPE is asymptotically no greater than the recursive MSPE.

2. Our use of "bias" here is different than how the term is often used elsewhere in the forecasting literature, where "bias" refers to the inconsistency resultant from misspecification of the fitted model (e.g., Schorfheide, 2005). It has long been established that LS estimators of panel data models with fixed effects and weakly exogenous regressors exhibit $O(T^{-1})$ bias (Nickell, 1981; Hahn and Kuersteiner, 2002, 2011; Phillips and Sul, 2007).

3. This is not to say that LS estimators of time series autoregressive models are not biased. Rather, the effect of the bias on the MSPE is of smaller order than the effect of the variance.

4. Recall that $\|\cdot\|$ denotes the spectral norm, and that for vectors b the spectral norm $\|b\|$ is equivalent to $\left(\mathrm{tr}\left(bb'\right)\right)^{1/2}$. Thus (7) denotes equation-by-equation LS estimation of the $VAR(q)$. We express the LS estimator as (7) in the interests of brevity.

5. Note that under Assumption A, $\{\tilde{\alpha}_s(q,h)\}_{s=1}^{q}$ defined in (16) is identical to that defined in (27).

6. Note that the h-step bias correction is not linear in $\tilde{\boldsymbol{\alpha}}_{q,1}$ for $h \geq 2$. However, because the h-step bias correction is a polynomial in $\tilde{\mathbf{A}}_q$ of finite order h, we can use moment bounds on powers of $\tilde{\mathbf{A}}_q$ already established for the proof of Theorem 3.4 to bound moments of the h-step bias correction.

REFERENCES

Bai, J. (2009) Panel data models with interactive fixed effects. *Econometrica* 77, 1229–1279.

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

Baltagi, B.H., G. Bresson, & A. Pirotte (2002) Comparison of forecast performance for homogeneous, heterogeneous and shrinkage estimators: Some empirical evidence from US electricity and natural-gas consumption. *Economics Letters* 76, 375–382.

Baltagi, B.H. & J.M. Griffin, (1997) Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline. *Journal of Econometrics* 77, 303–327.

Baltagi, B.H., J.M. Griffin, & W. Xiong (2000) To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand. *Review of Economics and Statistics* 82, 117–126.

Baltagi, B.H. & Q. Li (1992) Prediction in the one-way error component model with serial correlation. *Journal of Forecasting* 11, 561–567.

Bhansali, R.J. (1996) Asymptotically efficient autoregressive model selection for multistep prediction. *Annals of the Institute of Statistical Mathematics* 48, 577–602.

Bhansali, R.J. (1997) Direct autoregressive predictors for multistep prediction: Order selection and performance relative to plug-in predictors. *Statistica Sinica* 7, 425–449.

Findley, D.F. (1983) On the use of multiple models for multi-period forecasting. *American Statistical Association: Proceedings of Business and Economic Statistics*, 528–531.

Findley, D.F. & C.Z. Wei (1993) Moment bounds for deriving time series CLTs and model selection procedures. *Statistica Sinica* 3, 453–480.

Findley, D.F. & C.Z. Wei (2002) AIC, overfitting principles, and boundedness of moments of inverse matrices for vector autoregressive models fit to weakly stationary time series. *Journal of Multivariate Analysis* 83, 415–450.

Greenaway-McGrevy, R., C. Han, & D. Sul (2012). Asymptotic distribution of factor augmented estimators for panel regression. *Journal of Econometrics* 169, 48–53.

Hahn, J. & G. Kuersteiner (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects. *Econometrica* 70, 1639–1657.

Hahn, J. & G. Kuersteiner (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory* 27, 1152–1191

Hamilton, J.D. (1994) *Time Series Analysis*. Princeton University Press.

Ing, C.K. (2003) Multistep prediction in autoregressive processes. *Econometric Theory* 19, 254–279.

Ing, C.K. & C.Z. Wei (2005) Order selection for same-realization predictions in autoregressive processes. *Annals of Statistics* 33, 2423–2474.

Kiviet, J.F. (1995) On bias, inconsistency and efficiency of some estimators in dynamic panel data models. *Journal of Econometrics* 68, 53–78.

Kunitomo, N. & T. Yamamoto (1985) Properties of predictors in misspecified autoregressive time series. *Journal of the American Statistical Association* 80, 941–950.

Lee, Y. (2012) Model Selection in the Presence of Incidental Parameters. Mimeo, University of Michigan.

Lütkepohl, H. (2007) *New Introduction to Time Series Analysis.* Springer Verlag.

Nickell, S. (1981) Biases in dynamic models with fixed effects. *Econometrica* 49, 1417–1425.

Mark, N. & D. Sul (2001) Nominal exchange rates and monetary fundamentals: Evidence from a seventeen country panel. *Journal of International Economics* 53, 29–52.

Pesaran, H. (2006) Estimation and inference in large heterogenous panels with a multi factor error structure. *Econometrica* 74, 967–1012.

Phillips, P.C.B. & D. Sul (2007) Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics* 137, 162–188.

Rapach, D.E. & M. E. Wohar (2004) Testing the monetary model of exchange rate determination: A closer look at panels. *Journal of International Money and Finance* 23, 841–865.

Schorfheide, F. (2005) VAR forecasting under misspecification. *Journal of Econometrics* 128, 99–136.

Shibata, R. (1980) Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics* 8, 147–164.

Sims, C.A., J.H. Stock, & M. Watson (1990) Inference in linear time series models with some unit roots. *Econometrica* 58, 113–144.

Taub, A.J. (1979) Prediction in the context of the variance components model. *Journal of Econometrics* 10, 103–107.

APPENDIX

The proofs of Propositions 3.1 through 3.5 are provided online in the supplementary material to this article, available at Cambridge Journals Online (journals.cambridge.org/ect). The proofs for Theorems 3.1 and 3.2 are simpler than those for Theorems 3.3 and 3.4, and are omitted in the interest of brevity. The proofs for Theorems 3.1 and 3.2 can be obtained easily using the lemmas and arguments presented herein.

Throughout the Appendix, M denotes a generic positive constant; $\langle A \rangle_k := (\mathrm{E}\|A\|^k)^{1/k}$ for an arbitrary real-valued matrix A; and $\lambda_{\min}(\cdot)$ denotes the the minimum eigenvalue. Recall that $e_{i,t+k} := \sum_{s=1}^{k} \theta_{k-s}' u_{i,t+s}$ for integers $k = 1, \ldots, h$, so that $e_{i,t+k}$ is the regression error in a k-step $VAR(p)$; i.e., $e_{i,t+k} = y_{i,t} - \mu_i - \boldsymbol{\alpha}_{p,k}' \mathbf{x}_{i,t}(p)$. For brevity we also define $\mathbf{s}_{q,k} := \frac{1}{N(T-q-k+1)} \sum_{i=1}^{N} \sum_{t=q}^{T-k} (I_m \otimes \mathbf{x}_{i,t}(q)) e_{i,t+k}$, and

$$\mathbf{X}_T := \left[\mathbf{x}_{1,T}(q) : \cdots : \mathbf{x}_{i,T}(q) : \cdots : \mathbf{x}_{N,T}(q)\right]',$$
$$\bar{\mathbf{X}}_k := \left[\bar{\mathbf{x}}_{1,k}(q) : \cdots : \bar{\mathbf{x}}_{i,k}(q) : \cdots : \bar{\mathbf{x}}_{N,k}(q)\right]',$$
$$\bar{\mathbf{e}}_k := \left[\bar{e}_{1,k}(q) : \cdots : \bar{e}_{i,k}(q) : \cdots : \bar{e}_{N,k}(q)\right]',$$

where recall that $\bar{\mathbf{x}}_{i,k}(q) := \frac{1}{T-q-k+1} \sum_{t=q}^{T-k} \mathbf{x}_{i,t}(q)$ and $\bar{e}_{i,k}(q) := \frac{1}{T-q-k+1} \sum_{t=q}^{T-k} e_{i,t+k}$.

A.1. Auxiliary Lemmas. The following lemmas are used throughout the proofs. The proof of Lemma A.1 is straightforward and is omitted. The proofs of Lemmas A.2 and A.3 can be found in the supplementary material. The proofs rely on a generalization of the first moment bound theorem of Findley and Wei (1993) to the panel context, which can be found in the supplementary material. Throughout, k and j denote finite integers satisfying $1 \le k \le h$, $1 \le j \le h+1$.

LEMMA A.1. *Let $q \ge p$ be a finite integer. Then as $T \to \infty$ under Assumption A,*

(i) $$\mathbb{E}\left(\frac{T-k-q+1}{N} \sum_{i=1}^{N} \sum_{t=q}^{T-k} \bar{e}_{i,k}(q) \bar{e}_{i,k}'(q)\right) = \Lambda_k' \Sigma_u \Lambda_k - \sum_{l=1}^{k-1} \frac{h-l}{T-q-k+1} \sum_{r=1}^{h-l} (\theta_{r-1}' \Sigma_u \theta_{r+l-1} + \theta_{r+l-1}' \Sigma_u \theta_{r-1}),$$

(ii) $$\mathbb{E}\left(\frac{T}{N} \sum_{i=1}^{N} \sum_{t=q}^{T-k} \bar{\mathbf{x}}_{i,k}(q) \bar{\mathbf{x}}_{i,k}'(q)\right) = (\mathbf{I}_{mq} - \mathbf{A}_q')^{-1} \mathbf{R}_q + \mathbf{R}_q \mathbf{A}_q (\mathbf{I}_{mq} - \mathbf{A}_q)^{-1} + o(1),$$

(iii) $$\mathbb{E}\left(\frac{T}{N} \sum_{i=1}^{N} \sum_{t=q}^{T-k} \bar{\mathbf{x}}_{i,k}(q) \bar{e}_{i,k}'(q)\right) = \mathbf{R}_q \mathbf{B}_q \Lambda_k + o(1),$$

(iv) $$\mathbb{E}\left(\frac{T}{N} \sum_{i=1}^{N} \mathbf{x}_{i,T}(q) \sum_{t=q}^{T-k} \bar{\mathbf{x}}_{i,k}'(q)\right) = (\mathbf{I}_{mq} - \mathbf{A}_q')^{-1} \mathbf{A}_q^{k\prime} \mathbf{R}_q,$$

(v) $$\mathbb{E}\left(\frac{T}{N} \sum_{i} \mathbf{x}_{i,T}(q) \sum_{t=q}^{T-k} \bar{e}_{i,k}'(q)\right) = \mathbf{R}_q \mathbf{B}_q \Lambda_k - \sum_{r=1}^{k-1} \mathbf{A}_q^{r-1\prime} \mathbf{J}_{q,1} \Sigma_u (\Lambda_k - \Lambda_r) + o(1),$$

(vi) $$\mathbb{E}\left(\frac{1}{N} \sum_{i,r=1}^{N} \sum_{t=q}^{T-k} \sum_{s=q}^{T-k} \left(I_m \otimes \mathbf{x}_{i,T}(q)\right) e_{i,s+k} e_{r,t+k}' \left(I_m \otimes \mathbf{x}_{r,t}'(q)\right)\right) = O(1),$$

(vii) $$\mathbb{E}\left(\frac{1}{NT} \sum_{i,r=1}^{N} \sum_{t,s=q}^{T-k} \left(I_m \otimes \mathbf{x}_{i,t}(q)\right) e_{i,t+k} e_{r,s+k}' (I_m \otimes \mathbf{x}_{r,s}'(q))\right) = \mathbf{V}_{q,k} + o(1).$$

LEMMA A.2. *Let $q \ge p$ be a finite integer. Then as $N \to \infty$ and $T \to \infty$ under Assumption A,*

(i) $\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N}\bar{e}_{i,k}(q)\bar{e}'_{i,k}(q)-\frac{1}{T}\Lambda'_k\Sigma_u\Lambda_k\right\|^{2j(1+\epsilon)}=O\left(\left(NT^2\right)^{-j(1+\epsilon)}\right)$,

(ii) $\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N}\bar{\mathbf{x}}_{i,k}(q)\bar{\mathbf{x}}'_{i,k}(q)-\frac{1}{T}\left[(\mathbf{I}_{mq}-\mathbf{A}'_q)^{-1}\mathbf{R}_q+\mathbf{R}_q\mathbf{A}_q(\mathbf{I}_{mq}-\mathbf{A}_q)^{-1}\right]\right\|^{2j(1+\epsilon)}$
$=O\left(\left(NT^2\right)^{-j(1+\epsilon)}\right)$,

(iii) $\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N}\left(I_m\otimes\bar{\mathbf{x}}_{i,k}(q)\right)\bar{e}_{i,k}(q)-\frac{1}{T}\operatorname{vec}\left(\mathbf{R}_q\mathbf{B}_q\Lambda_k\right)\right\|^{2j(1+\epsilon)}$
$=O\left((NT^2)^{-(1+\epsilon)j}\right)$,

(iv) $\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_{i,T}(q)\bar{\mathbf{x}}'_{i,k}(q)-\frac{1}{T}\left(\mathbf{I}_{mq}-\mathbf{A}'_q\right)^{-1}\mathbf{A}_q^{k\prime}\mathbf{R}_q\right\|^{2j(1+\epsilon)}$
$=O\left((NT)^{-(1+\epsilon)j}\right)$,

(v) $\mathbb{E}\Big\|\frac{1}{N}\sum_{i=1}^{N}\left(I_m\otimes\mathbf{x}_{i,T}(q)\right)\bar{e}_{i,k}(q)-\frac{1}{T}\operatorname{vec}\Big(\mathbf{R}_q\mathbf{B}_q\Lambda_k-\sum_{r=1}^{k-1}\mathbf{A}_q^{r-1\prime}\mathbf{J}_{q,1}\Sigma_u\big(\Lambda_k$
$-\Lambda_r\big)\Big)\Big\|^{2j(1+\epsilon)}=O\left(\left(NT^2\right)^{-j(1+\epsilon)}\right)$,

(vi) $\mathbb{E}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=q}^{T-k}\left(I_m\otimes\mathbf{x}_{i,t}\right)e_{i,t+k}\right\|^{2j(1+\epsilon)}=O\left((NT)^{-j(1+\epsilon)}\right)$,

(vii) $\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_{i,T}(q)\mathbf{x}'_{i,T}(q)-\mathbf{R}_q\right\|^{2j(1+\epsilon)}=O\left(N^{-(1+\epsilon)j}\right)$ *for all finite* $q\geq 1$.

LEMMA A.3. *As $N\to\infty$ and $T\to\infty$ under Assumptions A and B,*

(i) for finite integers $q\geq p$,

(a) $\left\langle\operatorname{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,k}-\boldsymbol{\alpha}_{q,k}\right)-\left(I_m\otimes\mathbf{R}_q^{-1}\right)\mathbf{s}_{q,k}+\frac{1}{T}\operatorname{vec}(\mathbf{B}_q\Lambda_k)\right\rangle_j=O\left(\delta_{NT}^{-2}\right)$,

(b) $\left\langle\ddot{\boldsymbol{\alpha}}_{q,k}-\boldsymbol{\alpha}_{q,k}-\mathbf{R}_q^{-1}\mathbf{s}_{q,k}\right\rangle_2=O\left(\delta_{NT}^{-2}\right)$;

(ii) for finite integers $q\geq 1$,

(a) $\mathbb{E}\left\|\operatorname{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,k}-\boldsymbol{\alpha}_{q,k}\right)\right\|^{2j}=O\left(\delta_{NT}^{-2j}\right)$,

(b) $\mathbb{E}\left\|\ddot{\boldsymbol{\alpha}}_{q,k}-\boldsymbol{\alpha}_{q,k}\right\|^{2(h+1)/k}=O\left((NT)^{-(h+1)/k}\right)$.

A.2. Proof of Theorem 3.3 (Direct least squares predictor). Let $\omega_{i,a}(q,h):=\boldsymbol{\alpha}'_{q,h}\mathbf{x}_{i,T}(q)-\boldsymbol{\alpha}'_{p,h}\mathbf{x}_{i,T}(p)$. For any $q\geq 1$ the prediction error is

$$\tilde{y}^{\mathrm{D}}_{i,T+h}(q)-\mu_i-\boldsymbol{\alpha}'_{p,h}\mathbf{x}_{i,t}(p)=\left(\tilde{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}\right)'\mathbf{x}_{i,T}(q)+\left(\tilde{\mu}_i(q,h)-\mu_i\right)+\omega_{i,a}(q,h). \tag{A.1}$$

Theorem 3.3(i). For $q\geq p$, $\omega_{i,a}(q,h)=0$ and $\tilde{\mu}_i(q,h)-\mu_i=\bar{e}_{i,h}(q)-(\tilde{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h})'\bar{\mathbf{x}}_{i,h}(q)$, so that the prediction error (A.1) becomes

$$\begin{aligned}\tilde{y}^{\mathrm{D}}_{i,T+h}(q)-\mu_i-\boldsymbol{\alpha}'_{p,h}\mathbf{x}_{i,t}(p)&=\left(\tilde{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}\right)'\left(\mathbf{x}_{i,T}(q)-\bar{\mathbf{x}}_{i,h}(q)\right)+\bar{e}_{i,h}\\&=:\varepsilon_{i,a}(q,h)+\varepsilon_{i,b}(q,h).\end{aligned}$$

Define $Q_{rs}(q,h):=W_m\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{i,r}(q,h)\varepsilon'_{i,s}(q,h)$ for $r,s\in\{a,b\}$, so $\tilde{\mathrm{L}}_D(q,h)=\sum_{r,s\in\{a,b\}}\mathbb{E}(\operatorname{tr}(Q_{rs}(q,h)))$. The theorem straightforwardly follows from Lemmas A.4

to A.6, which solve for $\lim_{N,T\to\infty}\delta_{NT}^2\mathbb{E}(\mathrm{tr}(Q_{rs}(q,h)))$, $r,s\in\{a,b\}$. The basic strategy under each lemma is to decompose each term into a sum, and either bound the expectation of each addend using Lemmas A.2 and A.3, or solve for the expectation of the addend directly using Lemma A.1.

LEMMA A.4. *$\lim_{N,T\to\infty}\delta_{NT}^2\mathbb{E}(\mathrm{tr}(Q_{aa}(q,h))) = \kappa_N f_D(q,h) + \kappa_T b(q,h)$, where $b(q,h)$ is defined in (22) and $f_D(q,h)$ is defined in Theorem 3.1.*

Proof. Notation defined herein applies only within this proof. By the triangle inequality, Lemmas A.1(iv) and A.2(iv), and Lemmas A.1(ii) and A.2(ii), respectively, we have

$$\left\langle \frac{1}{N}\mathbf{X}_T'\bar{\mathbf{X}}_h \right\rangle_{2j} = O\left(T^{-1}\right),\ \left\langle \frac{1}{N}\bar{\mathbf{X}}_h'\bar{\mathbf{X}}_h \right\rangle_{2j} = O\left(T^{-1}\right), \quad \text{for all } j=1,\ldots,h+1, \tag{A.2}$$

where recall $\langle \frac{1}{N}\bar{\mathbf{X}}_h'\bar{\mathbf{X}}_h\rangle_{2j} = \left(\mathbb{E}\|\frac{1}{N}\bar{\mathbf{X}}_h'\bar{\mathbf{X}}_h\|^{2j}\right)^{1/2j}$, etc. Let $\delta_{NT}^2\mathrm{tr}(Q_{aa}(q,h)) = I + II$, where

$$I = \delta_{NT}^2\mathrm{tr}\left[(W_m\otimes\mathbf{R}_q)\mathrm{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}\right)\mathrm{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}\right)'\right],$$

$$II = \delta_{NT}^2\mathrm{tr}\left[\left(W_m\otimes\left(\frac{1}{N}(\mathbf{X}_T-\bar{\mathbf{X}}_h)'(\mathbf{X}_T-\bar{\mathbf{X}}_h)-\mathbf{R}_q\right)\right)\mathrm{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}\right)\right.$$
$$\left.\times\ \mathrm{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}\right)'\right].$$

By Hölder's inequality,

$$\mathbb{E}(II) \le M\cdot\left(\left\langle\frac{\mathbf{X}_T'\mathbf{X}_T}{N}-\mathbf{R}_q\right\rangle_2 + \left\langle\frac{2\mathbf{X}_T'\bar{\mathbf{X}}_h}{N}\right\rangle_2 + \left\langle\frac{\bar{\mathbf{X}}_h'\bar{\mathbf{X}}_h}{N}\right\rangle_2\right)$$
$$\times\left(\mathbb{E}\left\|\delta_{NT}\mathrm{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,h}-\alpha_{q,h}\right)\right\|^4\right)^{1/2},$$

so that $\mathbb{E}(II) = O(T^{-1}) + O(N^{-1/2})$ by (A.2), Lemmas A.3(ii)(a) and A.2(vii). For I decompose $\mathrm{vec}(\tilde{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}) = (I_m\otimes\mathbf{R}_q^{-1})\mathbf{s}_{q,h} - \frac{1}{T}\mathrm{vec}(\mathbf{B}_q\Lambda_h)+\rho_{q,h}$, where $\langle\rho_{q,h}\rangle_2 = O\left(\delta_{NT}^{-2}\right)$ by Lemma A.3(i)(a), so that by Hölder's inequality, Lemma A.2(vi), and noting that $\mathbb{E}(\mathbf{s}_{q,h}) = \mathbf{0}_{m^2q\times 1}$, we have

$$\mathbb{E}(I) = \delta_{NT}^2\mathbb{E}\left(\mathrm{tr}\left(\left(W_m\otimes\mathbf{R}_q^{-1}\right)\mathbf{s}_{q,h}\mathbf{s}_{q,h}'\left(I_m\otimes\mathbf{R}_q^{-1}\right)\right)\right) + \frac{\delta_{NT}^2}{T^2}\mathrm{tr}\left(W_m\Lambda_h'\mathbf{B}_q'\mathbf{R}_q\mathbf{B}_q\Lambda_h\right)$$
$$+o(1).$$

The stated result follows by using Lemma A.1(vii) to solve for $\delta_{NT}^2\mathbb{E}(\mathbf{s}_{q,h}\mathbf{s}_{q,h}')$ as $N,T\to\infty$. ■

LEMMA A.5. *$\lim_{N,T\to\infty}\delta_{NT}^2\mathbb{E}(\mathrm{tr}(Q_{bb}(q,h)) - \frac{1}{T}a(h)) = \kappa_T(c(h)+qa(h))$, where $a(h)$ and $c(h)$ are defined in (21) and (23), respectively.*

Proof. $Q_{bb}(q,h) = \mathrm{tr}(W_m\frac{1}{N}\sum_{i=1}^N \bar{e}_{i,h}(q)\bar{e}_{i,h}'(q))$. The result follows from Lemma A.1(i). ■

LEMMA A.6. *$lim_{N,T\to\infty}\delta_{NT}^{2}\mathbb{E}(\mathrm{tr}(Q_{ba}(q,h))) = \kappa_T d(q,h)$, where $d(q,h)$ is defined in (24).*

Proof. Notation defined herein applies only within this proof. First,

$$\delta_{NT}^{2}\mathrm{tr}(Q_{ba}(q,h)) = \delta_{NT}^{2}\left[\frac{\left(I_m\otimes\mathbf{X}_T'\right)\mathrm{vec}(\bar{\mathbf{e}}_h)}{N} - \frac{\left(I_m\otimes\bar{\mathbf{X}}_h'\right)\mathrm{vec}(\bar{\mathbf{e}}_h)}{N}\right]'\left(\Lambda_h W_m\otimes\mathbf{I}_{mq}\right)$$
$$\times\,\mathrm{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}\right).$$

By the triangle inequality and Lemmas A.1(v) and A.2(v), and A.1(iii) and A.2(iii), respectively,

$$\left\langle\frac{1}{N}\left(I_m\otimes\mathbf{X}_T'\right)\mathrm{vec}(\bar{\mathbf{e}}_h)\right\rangle_{2j} = O\left(\frac{1}{T}\right),\qquad \left\langle\frac{1}{N}\left(I_m\otimes\bar{\mathbf{X}}_h'\right)\mathrm{vec}(\bar{\mathbf{e}}_h)\right\rangle_{2j} = O\left(\frac{1}{T}\right),$$
$$j=1,\dots,h+1. \tag{A.3}$$

Then using $\mathrm{vec}(\tilde{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}) = (I_m\otimes\mathbf{R}_q^{-1})\mathbf{s}_{q,h} - \frac{1}{T}\mathrm{vec}(\mathbf{B}_q\Lambda_h)+\rho_{q,h}$, $\delta_{NT}^{2}\mathrm{tr}(Q_{ba}(q,h)) = I - II + III$, where

$$I = \delta_{NT}^{2}\frac{1}{N}\left(\left(I_m\otimes\mathbf{X}_T'\right)\mathrm{vec}(\bar{\mathbf{e}}_h) - \left(I_m\otimes\bar{\mathbf{X}}_h'\right)\mathrm{vec}(\bar{\mathbf{e}}_h)\right)'\left(\Lambda_h W_m\otimes\mathbf{R}_q^{-1}\right)\mathbf{s}_{q,h},$$
$$II = \delta_{NT}^{2}\frac{1}{N}\left(\left(I_m\otimes\mathbf{X}_T'\right)\mathrm{vec}(\bar{\mathbf{e}}_h) - \left(I_m\otimes\bar{\mathbf{X}}_h'\right)\mathrm{vec}(\bar{\mathbf{e}}_h)\right)'\frac{1}{T}\mathrm{vec}\left(\mathbf{R}_q\mathbf{B}_q\Lambda_h\right),$$

and where $\mathbb{E}(III)\le M\left(\left\langle\frac{1}{N}\mathbf{X}_T'\bar{\mathbf{e}}_h\right\rangle_2 + \left\langle\frac{1}{N}\bar{\mathbf{X}}_h'\bar{\mathbf{e}}_h\right\rangle_2\right)\cdot\delta_{NT}^{2}\langle\rho_{q,h}\rangle_2 = O\left(T^{-1}\right)$ by Lemmas A.3(i)(a), (A.3), and Hölder's inequality. We can straightforwardly solve for $\mathbb{E}(II) = -\kappa_T d(q,h)$ by Lemmas A.1(iii) and (v). Last, to show that $\mathbb{E}(I) = o(1)$,

$$I = \frac{\delta_{NT}^{2}}{N}\mathrm{tr}\left(\left(\Lambda_h W_m\otimes\mathbf{R}_q^{-1}\right)\mathbf{s}_{q,h}\mathrm{vec}(\bar{\mathbf{e}}_h)'(I_m\otimes\mathbf{X}_T)\right)$$
$$-\frac{\delta_{NT}^{2}}{T}\mathrm{vec}\left(\mathbf{R}_q\mathbf{B}_q\Lambda_h\right)'\left(\Lambda_h W_m\otimes\mathbf{R}_q^{-1}\right)\mathbf{s}_{q,h}$$
$$+\delta_{NT}^{2}\left(\frac{1}{N}\left(I_m\otimes\bar{\mathbf{X}}_h'\right)\mathrm{vec}(\bar{\mathbf{e}}_h) - \frac{1}{T}\mathrm{vec}\left(\mathbf{R}_q\mathbf{B}_q\Lambda_h\right)\right)'\left(\Lambda_h W_m\otimes\mathbf{R}_q^{-1}\right)\mathbf{s}_{q,h}$$
$$= IV + V + VI.$$

First note that $\mathbb{E}(V)=0$ because $\mathbb{E}\left(\mathbf{s}_{q,h}\right) = \mathbf{0}_{m^2q\times 1}$. Next, since $\langle\mathbf{s}_{q,h}\rangle_{2j} = O((NT)^{-1/2})$ for all $j=1,\dots,h+1$ by Lemma A.2(vi), it follows that $\mathbb{E}(VI) = O(\delta_{NT}^{2}T^{-2}N^{-1}) = O(\delta_{NT}^{-1})$ by Lemma A.2(iii) and Hölder's inequality. Last, $\mathbb{E}(IV) = O(\delta_{NT}^{2}N^{-1}T^{-2}) = O(\delta_{NT}^{-1})$ by Lemma A.1(vi). ■

Theorem 3.3(ii). By inspection of (A.1), since $\mathbb{E}\left\|\mathrm{vec}(\tilde{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h})\right\|^4 = o(\delta_{NT}^{-4})$ by Lemma A.3$(ii)(a)$ and since we can straightforwardly show that $\mathbb{E}\|\tilde{\mu}_i(q,h)-\mu_i\|^2 = o(1)$, by Hölder's inequality and Lemma A.2(vii) the MSPE is dominated by $\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\omega_{i,a}(q,h)\omega_{i,a}'(q,h)\right) = \mathrm{tr}\left(W_m\xi(q,h)\right)$.

A.3. Proof of Theorem 3.4 (Recursive least squares predictor). For the recursive predictor given in (35), for finite $q \geq 1$ the prediction error is

$$\tilde{y}^{R}_{i,T+h}(q) - \mu_i - \boldsymbol{\alpha}'_{p,h}\mathbf{x}_{i,t}(p) = \mathbf{J}'_{q,1}\left(\tilde{\mathbf{A}}^{h}_{q} - \mathbf{A}^{h}_{q}\right)'\mathbf{x}_{i,T}(q) + \mathbf{J}'_{q,1}\left(\sum_{j=0}^{h-1}\tilde{\mathbf{A}}^{j}_{q}\right)' \times \mathbf{J}_{q,1}(\tilde{\mu}_i(q,1) - \mu_i) + \omega_{i,b}(q,h), \quad \textbf{(A.4)}$$

where $\omega_{i,b}(q,h) := (\mathbf{J}'_{q,1}\mathbf{A}^{h\prime}_{q}\mathbf{x}_{i,T}(q) - \mathbf{J}'_{p,1}\mathbf{A}^{h\prime}_{p}\mathbf{x}_{i,T}(p))$. We repeatedly use

$$\tilde{\mathbf{A}}^{j}_{q} - \mathbf{A}^{j}_{q} = \sum_{k=0}^{j-1}\mathbf{A}^{j-k-1}_{q}\left(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1}\right)\mathbf{J}'_{q,1}\mathbf{A}^{k-1}_{q} + \mathbf{g}_{a,j} \quad \text{for all } j \geq 2, \quad \textbf{(A.5)}$$

where the $mq \times mq$ matrix $\mathbf{g}_{a,j}$ satisfies $\left\|\text{vec}(\mathbf{g}_{a,j})\right\| \leq M\sum_{k=2}^{j}\left\|\text{vec}(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1})\right\|^{k}$.

Theorem 3.4(i). For $q \leq p$, $\omega_{i,b}(q,h) = 0$ and $\tilde{\mu}_i(q,1) - \mu_i = \bar{e}_{i,1}(q) - (\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1})'\mathbf{x}_{i,1}(q)$ in (A.4). Then using (A.5) the prediction error (A.4) becomes

$$\tilde{y}^{R}_{i,T+h}(q) - \mu_i - \boldsymbol{\alpha}'_{p,h}\mathbf{x}_{i,t}(p) = \varepsilon_{i,c}(q,h) + \varepsilon_{i,d}(q,h) + \left(I_m \otimes \mathbf{x}'_{i,T}(q)\right)\text{vec}(\mathbf{g}_{a,h}) + g_{b,i,h}, \quad \textbf{(A.6)}$$

where $\varepsilon_{i,d}(q,h) := \Lambda'_h\bar{e}_{i,1}$,

$$\varepsilon_{i,c}(q,h) := \left[\left(I_m \otimes \mathbf{x}_{i,T}(q)\right)'\mathbf{L}_{q,h} - \left(\Lambda_h \otimes \bar{\mathbf{x}}_{i,1}(q)\right)' - \left(I_m \otimes \mathbf{J}_{q,1}\bar{e}_{i,1}(q)\right)'\mathbf{M}_{q,h}\right] \times \text{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1}\right),$$

and the $m \times 1$ vector $g_{b,i,h}$ satisfies

$$\mathbb{E}\left\|g_{b,i,h}\right\|^2 \leq M\left(\mathbb{E}\left\|\bar{e}_{i,1}\right\|^2\sum_{j=2}^{h-1}\mathbb{E}\left\|\text{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1}\right)\right\|^{2j} + \sum_{j=1}^{h}\mathbb{E}\left\|\text{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1}\right)\right\|^{2j}\mathbb{E}\left\|\bar{\mathbf{x}}_{i,1}(q)\right\|^2\right).$$

Now $\mathbb{E}\left\|\bar{e}_{i,1}\right\|^2 = O(T^{-1})$, while by Lemma A.3(*ii*)(*a*), $\mathbb{E}\left\|\text{vec}(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1})\right\|^{2j} = O(\delta_{NT}^{-2j})$ for $j = 1, \ldots, h+1$, and hence $\mathbb{E}\left\|g_{b,i,h}\right\|^2 = O(T^{-1}\delta_{NT}^{-2})$. In addition, by Hölder's inequality

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left\|\left(I_m \otimes \mathbf{x}'_{i,T}(q)\right)\text{vec}(\mathbf{g}_{a,h})\right\|^2 \leq M\sum_{j=2}^{h}\left(\mathbb{E}\left\|\frac{\mathbf{X}'_T\mathbf{X}_T}{N}\right\|^{(j+1)}\right)^{1/(j+1)} \times \left(\mathbb{E}\left\|\text{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1}\right)\right\|^{2(j+1)}\right)^{j/(j+1)}.$$

The first expectation is $O(1)$, while the second is $O(\delta_{NT}^{-4})$, so we can express $\tilde{\mathrm{L}}_R(q,h) = \sum_{r,s\in\{c,d\}}\mathbb{E}(\text{tr}(Q_{rs}(q,h))) + \upsilon_{NT}$, where υ_{NT} denotes $o\left(\delta_{NT}^{-2}\right)$ terms associated with the products of $\mathbf{g}'_{a,h}\mathbf{x}_{i,T}(q)$ and $g_{b,i,h}$ in (A.6). Lemmas A.7 through A.9 below solve for the $O(1)$ terms comprising $\lim_{N,T\to\infty}\delta_{NT}^{2}\mathbb{E}(\text{tr}(Q_{rs}(q,h)))$, $r,s \in \{c,d\}$, so that the $\upsilon_{NT} = o(\delta_{NT}^{-2})$ term can be ignored.

LEMMA A.7. *$lim_{N,T\to\infty}\delta_{NT}^2\mathbb{E}(\mathrm{tr}(Q_{cc}(q,h))) = \kappa_N f_R(q,h) + \kappa_T k(q,h)$, where $f_R(q,h)$ is defined in Theorem 3.2 and $k(q,h)$ is defined in (33).*

Proof. Notation defined herein applies only within this proof. We expand δ_{NT}^2 $\mathrm{tr}(Q_{cc}(q,h)) = I - II - (II)' + III$, where

$$I = \delta_{NT}^2 \mathrm{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1}\right)' \mathbf{L}_{q,h}' \left(W_m \otimes \mathbf{R}_q\right) \mathbf{L}_{q,h} \mathrm{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1}\right),$$

$$II = \delta_{NT}^2 \mathrm{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1}\right)' \mathbf{L}_{q,h}' \left[\left(W_m \Lambda_h' \otimes \frac{\mathbf{X}_T' \bar{\mathbf{X}}_1}{N}\right) + \left(W_m \otimes \frac{\mathbf{X}_T' \bar{\mathbf{e}}_1 \mathbf{J}_{q,1}'}{N}\right) \mathbf{M}_{q,h}\right]$$
$$\times \mathrm{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1}\right),$$

and

$$III = \delta_{NT}^2 \mathrm{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1}\right)'$$
$$\times \left[\mathbf{L}_{q,h}' \left(W_m \otimes \left(\frac{\mathbf{X}_T' \mathbf{X}_T}{N} - \mathbf{R}_q\right)\right) \mathbf{L}_{q,h} + \mathbf{M}_{q,h}' \left(I_m \otimes \frac{\mathbf{J}_{q,1} \bar{\mathbf{e}}_1' \bar{\mathbf{e}}_1 \mathbf{J}_{q,1}'}{N}\right) \mathbf{M}_{q,h}\right.$$
$$+ \left(\Lambda_h W_m \Lambda_h' \otimes \frac{\bar{\mathbf{X}}_1' \bar{\mathbf{X}}_1}{N}\right) + \mathbf{M}_{q,h}' \left(W_m \Lambda_h' \otimes \frac{\mathbf{J}_{q,1} \bar{\mathbf{e}}_1' \bar{\mathbf{X}}_1}{N}\right)$$
$$\left.+ \left(\Lambda_h W_m \otimes \frac{\bar{\mathbf{X}}_1' \bar{\mathbf{e}}_1 \mathbf{J}_{q,1}}{N}\right) \mathbf{M}_{q,h}\right] \mathrm{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1}\right).$$

We can solve for $\mathbb{E}(I) = \kappa_N f_R(q,h) + \kappa_T k(q,h) + o(1)$ using the same methods as used in the proof of Lemma A.4 above. Next, since $\left\|\mathbf{L}_{q,h}\right\| = O(1)$, $\|\Lambda_h\| = O(1)$, by Hölder's inequality

$$\mathbb{E}(II) \le M \cdot \left(\left\langle \frac{\mathbf{X}_T' \bar{\mathbf{e}}_1}{N} \right\rangle_2 + \left\langle \frac{\mathbf{X}_T' \bar{\mathbf{X}}_1}{N} \right\rangle_2\right) \cdot \left(\mathbb{E}\left\|\delta_{NT} \mathrm{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1}\right)\right\|^4\right)^{1/2} = O\left(T^{-1}\right)$$

by (A.2), (A.3), and Lemma A.3(ii)(a). For III, by Lemma A.1(i) and Lemma A.2(i)

$$\left\langle \frac{1}{N} \bar{\mathbf{e}}_1' \bar{\mathbf{e}}_1 \right\rangle_{2j} = O\left(T^{-1}\right) \quad \text{for } j = 1, \ldots, h+1. \tag{A.7}$$

Then using Hölder's inequality

$$\mathbb{E}(III) \le M \left(\left\langle \frac{\bar{\mathbf{X}}_1' \bar{\mathbf{e}}_1}{N} \right\rangle_2 + \left\langle \frac{\bar{\mathbf{X}}_1' \bar{\mathbf{X}}_1}{N} \right\rangle_2 + \left\langle \frac{\bar{\mathbf{e}}_1' \bar{\mathbf{e}}_1}{N} \right\rangle_2 + \left\langle \frac{\mathbf{X}_T' \mathbf{X}_T}{N} - \mathbf{R}_q \right\rangle_2\right)$$
$$\times \left(\mathbb{E}\left\|\delta_{NT} \mathrm{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1}\right)\right\|^4\right)^{1/2}.$$

By (A.2) and (A.3), Lemmas A.3(ii)(a) and A.2(vii), $\mathbb{E}(III) = O(T^{-1}) + O(N^{-1/2})$. ■

LEMMA A.8. *$lim_{N,T\to\infty}\delta_{NT}^2\mathbb{E}\left(\mathrm{tr}\left(Q_{dd}(q,h) - \frac{1}{T}a(h)\right)\right) = \kappa_T q a(h)$, for $a(h)$ is defined in (21)*

Proof. $Q_{dd}(q,h) = \mathrm{tr}(W_m \Lambda_h' \frac{1}{N} \sum_{i=1}^N \bar{e}_{i,1}(q) \bar{e}_{i,1}'(q) \Lambda_h)$, where $\mathbb{E}(\bar{e}_{i,1}(q) \bar{e}_{i,1}'(q)) = \frac{1}{T-q} \Sigma_u = \frac{1}{T} \Sigma_u + \frac{q}{T^2} \Sigma_u + O(\frac{1}{T^3})$ by Lemma A.1(i). ■

LEMMA A.9. *$lim_{N,T\to\infty}\delta_{NT}^2\mathbb{E}(\mathrm{tr}(Q_{dc}(q,h)))=0$.*

Proof. Notation defined herein applies only within this proof. We have

$$\begin{aligned}\delta_{NT}^2\,\mathrm{tr}(Q_{dc}(q,h))=\delta_{NT}^2\Bigg[&\frac{\mathrm{vec}(\bar{\mathbf{e}}_1)'(I_m\otimes\mathbf{X}_T)}{N}\left(W_m\Lambda_h'\otimes\mathbf{I}_{mq}\right)\mathbf{L}_{q,h}\\ &-\frac{\mathrm{vec}(\bar{\mathbf{e}}_1)'\left(I_m\otimes\bar{\mathbf{X}}_1\right)}{N}\left(W_m\Lambda_h'\otimes\mathbf{I}_{mq}\right)\\ &-\frac{\mathrm{vec}(\bar{\mathbf{e}}_1)'(I_m\otimes\bar{\mathbf{e}}_1)}{N}\left(W_m\Lambda_h'\otimes\mathbf{J}_{q,1}'\right)\mathbf{M}_{q,h}\Bigg]\mathrm{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,1}-\boldsymbol{\alpha}_{q,1}\right).\end{aligned}$$

Using $\mathrm{vec}(\tilde{\boldsymbol{\alpha}}_{q,1}-\boldsymbol{\alpha}_{q,1})=(I_m\otimes\mathbf{R}_q^{-1})\mathbf{s}_{q,1}-\frac{1}{T}\mathrm{vec}(\mathbf{B}_q)+\rho_{q,1}$, where $\langle\rho_{q,1}\rangle_2=O(\delta_{NT}^{-2})$ by Lemma A.3(i)(a), we have $\delta_{NT}^2\mathrm{tr}(Q_{dc}(q,h))=I-II-III-IV+V$, where

$$I=\delta_{NT}^2\,\mathrm{tr}\left[\left(W_m\Lambda_h'\otimes\mathbf{I}_{mq}\right)\mathbf{L}_{q,h}\left(I_m\otimes\mathbf{R}_q^{-1}\right)\mathbf{s}_{q,1}\left(\frac{\mathrm{vec}(\bar{\mathbf{e}}_1)'(I_m\otimes\mathbf{X}_T)}{N}\right)\right],$$

$$II=\frac{\delta_{NT}^2}{N}\mathrm{vec}(\bar{\mathbf{e}}_1)'\left(I_m\otimes\bar{\mathbf{X}}_1\right)\left(\Lambda_hW_m\Lambda_h'\otimes\mathbf{I}_{mq}\right)\left(I_m\otimes\mathbf{R}_q^{-1}\right)\mathbf{s}_{q,1},$$

$$III=\frac{\delta_{NT}^2}{N}\mathrm{vec}(\bar{\mathbf{e}}_1)'(I_m\otimes\bar{\mathbf{e}}_1)\left(I_m\otimes\mathbf{J}_{q,1}'\right)\left(W_m\Lambda_h'\otimes\mathbf{I}_{mq}\right)\mathbf{M}_{q,h}\left(I_m\otimes\mathbf{R}_q^{-1}\right)\mathbf{s}_{q,1},$$

$$\begin{aligned}IV=\delta_{NT}^2\Bigg[&\left(\frac{\mathrm{vec}(\bar{\mathbf{e}}_1)'(I_m\otimes\mathbf{X}_T)}{N}\right)\left(\Lambda_hW_m\otimes\mathbf{I}_{mq}\right)\mathbf{L}_{q,h}\\ &-\left(\frac{\mathrm{vec}(\bar{\mathbf{e}}_1)'\left(I_m\otimes\bar{\mathbf{X}}_1\right)}{N}\right)\left(\Lambda_hW_m\Lambda_h'\otimes\mathbf{I}_{mq}\right)\\ &-\left(\frac{\mathrm{vec}(\bar{\mathbf{e}}_1)'(I_m\otimes\bar{\mathbf{e}}_1)}{N}\right)\left(I_m\otimes\mathbf{J}_{q,1}'\right)\left(\Lambda_hW_m\otimes\mathbf{I}_{mq}\right)\mathbf{M}_{q,h}\Bigg]\frac{1}{T-q}\mathrm{vec}(\mathbf{B}_q),\end{aligned}$$

and where $\mathbb{E}(V)\le M\left(\langle\frac{1}{N}\mathbf{X}_T'\bar{\mathbf{e}}_1\rangle_2+\langle\frac{1}{N}\bar{\mathbf{X}}_1'\bar{\mathbf{e}}_1\rangle_2+\langle\frac{1}{N}\bar{\mathbf{e}}_1'\bar{\mathbf{e}}_1\rangle_2\right)\delta_{NT}^2\langle\rho_{q,1}\rangle_2=O(T^{-1})$ by (A.3) and (A.7) above, Lemma A.2(i)(a), and Hölder's inequality. Now $\mathbb{E}(I)=o(1)$ by the same arguments used in the proof of Lemma A.6. Next decompose

$$\begin{aligned}II=\delta_{NT}^2&\left[\frac{(I_m\otimes\bar{\mathbf{X}}_1')\mathrm{vec}(\bar{\mathbf{e}}_1)}{N}-\frac{\mathrm{vec}\left(\mathbf{R}_q\mathbf{B}_q\right)}{T}\right]'\left(\Lambda_hW_m\Lambda_h'\otimes\mathbf{R}_q^{-1}\right)\mathbf{s}_{q,1}\\ &+\frac{\delta_{NT}^2}{T}\mathrm{vec}\left(\mathbf{R}_q\mathbf{B}_q\right)'\left(\Lambda_hW_m\Lambda_h'\otimes\mathbf{R}_q^{-1}\right)\mathbf{s}_{q,1}.\end{aligned}$$

The expectation of the first term is $O(\delta_{NT}^2N^{-1}T^{-2})=o(1)$ by Lemmas A.2(iii),(vi) and Hölder's inequality, while the second term has zero expectation since $\mathbb{E}(\mathbf{s}_{q,1})=\mathbf{0}_{m^2q\times1}$, so that $\mathbb{E}(II)=o(1)$. We can use a similar argument to show $\mathbb{E}(III)=o(1)$. Last, rearrange IV as

$$IV=\mathrm{tr}\left[W_m\Lambda_h'\sum_{j=1}^h\left(\frac{\bar{\mathbf{e}}_1'\mathbf{X}_T}{N}\mathbf{A}_q^{h-j}\mathbf{B}_q\theta_{j-1}-\frac{\bar{\mathbf{e}}_1'\bar{\mathbf{X}}_1}{N}\mathbf{B}_q\Lambda_h-\frac{\bar{\mathbf{e}}_1'\bar{\mathbf{e}}_1}{N}\mathbf{J}_{q,1}\sum_{k=1}^j\mathbf{A}_q^{j-k}\mathbf{B}_q\theta_{k-1}\right)\right].$$

We then use Lemmas A.1(i), (iii), and (v) to take the expectation of IV, noting that

$$\begin{aligned}\Lambda_h'\mathbf{B}_q'\mathbf{R}_q\left(\mathbf{B}_q\Lambda_h-\mathbf{K}_{q,h}\right)&=\sum_{j=1}^h\Lambda_h'\Sigma_u\mathbf{J}_{q,1}'\left(\mathbf{I}_{mq}-\mathbf{A}_q\right)^{-1}\left(\mathbf{I}_{mq}-\mathbf{A}_q^{h-j}\right)\mathbf{B}_q\theta_{j-1}\\&=\Lambda_h'\Sigma_u\mathbf{J}_{q,1}'\sum_{j=1}^h\mathbf{K}_{q,j},\end{aligned}$$

where $\mathbf{K}_{q,j}:=\sum_{k=1}^j\mathbf{A}_q^{j-k}\mathbf{B}_q\theta_{k-1}$. It then follows that $\mathbb{E}(IV)=o(\delta_{NT}^{-2})$. ■

Theorem 3.4(ii). Consider (A.4) when $q<p$. Because $\mathbb{E}\left\|\operatorname{vec}\left(\tilde{\boldsymbol{\alpha}}_{q,1}-\boldsymbol{\alpha}_{q,1}\right)\right\|^{2j}=o(1)$ for $j=1,\ldots,h+1$ and $\mathbb{E}\|\tilde{\mu}_i(q,1)-\mu_i\|^2=o(1)$, then, by using (A.5), Hölder's inequality, and Lemma A.2(vii), the MSPE is dominated by $\mathbb{E}(N^{-1}\sum_{i=1}^N\omega_{i,b}(q,h)\omega_{i,b}'(q,h))$. The result follows by solving for $\mathbb{E}(\omega_{i,b}(q,h)\omega_{i,b}'(q,h))$.

A.4. Proof of Theorem 3.5 (Direct bias-corrected least squares predictor). For any finite $q\geq 1$ the forecast error can be expressed as

$$\ddot{y}_{i,T+h}^{\mathrm{D}}(q)-\mu_i-\boldsymbol{\alpha}_{p,h}'\mathbf{x}_{i,t}(p)=\left(\ddot{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}\right)'\mathbf{x}_{i,T}(q)+(\ddot{\mu}_i(q,h)-\mu_i)+\omega_{i,a}(q,h).$$

Theorem 3.5(i). For $q\geq p$, we have $\ddot{\mathrm{L}}_D(q,h)=\sum_{r,s\in\{a,b\}}\mathbb{E}(\ddot{Q}_{rs}(q,h))$, where we define $\ddot{Q}_{rs}(q,h):=\frac{1}{N}\sum_{i=1}^N\ddot{\varepsilon}_{i,r}(q,h)\ddot{\varepsilon}_{i,s}(q,h)$, $\ddot{\varepsilon}_{i,a}(q,h):=(\ddot{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h})'(\mathbf{x}_{i,T}(q)-\bar{\mathbf{x}}_{i,h}(q))$, and $\ddot{\varepsilon}_{i,b}(q,h):=\varepsilon_{i,b}(q,h)$. Theorem 3.5(i) follows from solving for $\lim_{N,T\to\infty}\delta_{NT}^2\mathbb{E}(\ddot{Q}_{rs}(q,h))$, $r,s\in\{a,b\}$. First, $\lim_{N,T\to\infty}\delta_{NT}^2\mathbb{E}(\ddot{Q}_{bb}(q,h))=\kappa_T(a(h)q+c(h))$ as in Lemma A.5 above. The following lemmas solve for the two remaining expectations.

LEMMA A.10. $\lim_{N,T\to\infty}\delta_{NT}^2\mathbb{E}(\ddot{Q}_{aa}(q,h))=\kappa_N f_D(q,h)$.

Proof. Notation defined herein applies only within this proof. Decompose $\delta_{NT}^2\ddot{Q}_{aa}(q,h)=I+II$,

$$\begin{aligned}I&=\delta_{NT}^2\operatorname{tr}\left(\mathbf{R}_q\left(\ddot{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}\right)\left(\ddot{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}\right)'\right),\\II&=\delta_{NT}^2\operatorname{tr}\left(\left(\frac{1}{N}(\mathbf{X}_T-\bar{\mathbf{X}}_h)'(\mathbf{X}_T-\bar{\mathbf{X}}_h)-\mathbf{R}_q\right)\left(\ddot{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}\right)\left(\ddot{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}\right)'\right).\end{aligned}$$

First, by Hölder's inequality,

$$\begin{aligned}\mathbb{E}(II)\leq M&\left(\left\langle\frac{\mathbf{X}_T'\mathbf{X}_T}{N}-\mathbf{R}_q\right\rangle_{h+1}+\left\langle\frac{2\mathbf{X}_T'\bar{\mathbf{X}}_h}{N}\right\rangle_{h+1}+\left\langle\frac{\bar{\mathbf{X}}_h'\bar{\mathbf{X}}_h}{N}\right\rangle_{h+1}\right)\\&\times\left(\mathbb{E}\left\|\delta_{NT}\left(\ddot{\boldsymbol{\alpha}}_{q,h}-\boldsymbol{\alpha}_{q,h}\right)\right\|^{(2h+2)/h}\right)^{h/(h+1)}.\end{aligned}$$

We have $\mathbb{E}(II)=O(N^{-1/2})+O\left(T^{-1}\right)$ by Lemmas A.2(vii), A.3(ii)(b), and (A.2). Next, $\mathbb{E}(I)=\kappa_N f_D(q,h)+o(1)$ follows from Lemmas A.2(vii) and A.3(i)(b), using the same arguments presented in the proof of Lemma A.4 above. ■

LEMMA A.11. $\lim_{N,T\to\infty}\delta_{NT}^2\mathbb{E}(\ddot{Q}_{ba}(q,h))=0$.

Proof. Notation defined herein applies only within this proof. Decompose $\delta_{NT}^2 \ddot{Q}_{ba}(q,h) = I - II$,

$$I = \delta_{NT}^2 \Lambda_h \left(\frac{1}{N} \mathbf{X}_T' \bar{\mathbf{e}}_h - \frac{1}{N} \bar{\mathbf{X}}_h' \bar{\mathbf{e}}_h + \frac{1}{T} \sum_{j=1}^{h-1} \mathbf{A}_q^{j-1\prime} \mathbf{J}_{q,1} \Sigma_u \left(\Lambda_h - \Lambda_j \right) \right)' \left(\ddot{\boldsymbol{\alpha}}_{q,h} - \boldsymbol{\alpha}_{q,h} \right),$$

$$II = \frac{\delta_{NT}}{T} \Lambda_h \left(\sum_{j=1}^{h-1} \mathbf{A}_q^{j-1\prime} \mathbf{J}_{q,1} \Sigma_u \left(\Lambda_h - \Lambda_j \right) \right)' \cdot \delta_{NT} \left(\ddot{\boldsymbol{\alpha}}_{q,h} - \boldsymbol{\alpha}_{q,h} \right).$$

Now since $\mathbb{E}(\mathbf{s}_{q,h}) = \mathbf{0}_{m^2 q \times 1}$, by Lemma A.3(i)(b) and Hölder's inequality we have $\mathbb{E}(II) = o(1)$. For I, using Hölder's inequality,

$$\mathbb{E}(I) \le \delta_{NT} \left\langle \frac{\mathbf{X}_T' \bar{\mathbf{e}}_h}{N} - \frac{\bar{\mathbf{X}}_h' \bar{\mathbf{e}}_h}{N} + \frac{1}{T} \sum_{j=1}^{h-1} \mathbf{A}_q^{j-1\prime} \mathbf{J}_{q,1} \Sigma_u \left(\Lambda_h - \Lambda_j \right) \right\rangle_{h+1} \times \left\langle \delta_{NT} \left(\ddot{\boldsymbol{\alpha}}_{q,h} - \boldsymbol{\alpha}_{q,h} \right) \right\rangle_{h/h+1},$$

such that $\mathbb{E}(I) = O(\delta_{NT} T^{-3/2} N^{-1/2}) = O(\delta_{NT}^{-1})$ by Lemmas A.2(iii), (v), and Lemma A.3(ii)(b). ∎

Theorem 3.5(ii). The proof of Theorem 3.5(ii) follows the same arguments as those for Theorem 3.3(ii).

A.5. Proof of Theorem 3.6 (Recursive bias-corrected LS predictor). For general $q \ge 1$, the prediction error can be expressed as in (A.4) and (A.5), with $\ddot{\boldsymbol{\alpha}}_{q,1}$ replacing $\tilde{\boldsymbol{\alpha}}_{q,1}$ and $\ddot{\mu}_i(q,1)$ replacing $\tilde{\mu}_i(q,1)$, and for a $q \times q$ matrix $\ddot{\mathbf{g}}_{a,j}$ that satisfies $\|\mathrm{vec}(\ddot{\mathbf{g}}_{a,j})\| \le M \sum_{k=2}^{j} \|\ddot{\boldsymbol{\alpha}}_{q,1} - \boldsymbol{\alpha}_{q,1}\|^k$.

Theorem 3.6(i). As in the proof of Theorem 3.4(i), we can express the prediction error as

$$\ddot{y}_{i,T+h}^{\mathrm{R}}(q) - \mu_i - \boldsymbol{\alpha}_{p,h}' \mathbf{x}_{i,t}(p) = \sum_{r \in \{c,d\}} \ddot{\varepsilon}_{i,r}(q,h) + \mathbf{x}_{i,T}'(q) \ddot{\mathbf{g}}_{a,h} + \ddot{g}_{b,i,h},$$

where $\ddot{\varepsilon}_{i,c}(q,h)$ is defined analogously to $\varepsilon_{i,c}(q,h)$ in Section A.3 but with $\ddot{\boldsymbol{\alpha}}_{q,1}$ replacing $\tilde{\boldsymbol{\alpha}}_{q,1}$, $\ddot{\varepsilon}_{i,d}(q,h) := \varepsilon_{i,d}(q,h)$, and for some scalars $\mathbf{x}_{i,T}'(q) \ddot{\mathbf{g}}_{a,h}$ and $\ddot{g}_{b,i,h}$ that can be shown to satisfy

$$\frac{1}{N} \sum_{i=1}^{N} \mathbb{E} \left\| \mathbf{x}_{i,T}'(q) \ddot{\mathbf{g}}_{a,h} \right\|^2 = O\left(\delta_{NT}^{-4}\right), \quad \frac{1}{N} \sum_{i=1}^{N} \mathbb{E} \left\| g_{b,i,h} \right\|^2 = O\left(\delta_{NT}^{-4}\right),$$

by using Lemma A.3(ii)(b). Thus we can express $\ddot{\mathrm{L}}_R(q,h) = \sum_{r,s \in \{c,d\}} \mathbb{E}(\ddot{Q}_{rs}(q,h)) + \ddot{v}_{NT}$, where $\ddot{v}_{NT}$ denotes the smaller $o(\delta_{NT}^{-2})$ terms associated with $\mathbf{x}_{i,T}'(q) \ddot{\mathbf{g}}_{a,h}$ and $\ddot{g}_{b,i,h}$, and $\ddot{Q}_{rs}(q,h) := \frac{1}{N} \sum_{i=1}^{N} \ddot{\varepsilon}_{i,r}(q,h) \ddot{\varepsilon}_{i,s}(q,h)$ for $r,s \in \{c,d\}$. The theorem follows by the same arguments as those used in the proof of Theorems 3.5(i) and 3.4(i) to solve for $\lim_{N,T \to \infty} \delta_{NT}^2 \mathbb{E}(\ddot{Q}_{rs}(q,h))$, $r,s \in \{c,d\}$.

Theorem 3.6(ii). The proof of Theorem 3.6(ii) follows the same arguments as those for Theorem 3.4(ii).