Forecast Averaging with Panel Data Vector Autoregressions

Ryan Greenaway-McGrevy*
Department of Economics, The University of Auckland

October 2018

Abstract

In this paper we propose a new forecast model averaging method for panel data vector autoregressions that permit limited forms of parameterized heterogeneity (including fixed effects or incidental trends). Models are fitted using bias-corrected least squares in order to attenuate the effects of small sample bias of forecast loss. We begin by constructing a general estimator of the quadratic forecast risk of the averaged model that is asympotically unbiased as both n (cross sections) and T (time series) grow large. Armed with this result, we propose a specific weighting mechanism, in which weights are chosen to minimize the estimated quadratic risk of the averaged forecast error. The objective function in this minization problem is a version of the Mallows criterion proposed by Hansen (2008, J. Econometrics 146, 342-350) modified for application to the panel data setting. This averaging method performs well in a variety of pseudo-out-of-sample forecasting applications.

JEL classification codes: C23, C53

Keywords: Forecasting, Model Selection, Panel Data, Misspecification, Bias-Correction, Final Prediction Error, Mallows Criterion.

*Email: r.mcgrevy@auckland.ac.nz

1 Introduction

Model averaging is increasingly being used in applied econometrics and statistics as an alternative to model selection. Rather than base inference and prediction on a single model selected from an underlying set of candidate models, each model within the candidate set is assigned a relative weight prior to averaging. Inference and prediction then proceed based on the average model. Various methods have been developed for selecting the weights, including exponentiated AIC weights (Buckland, Burnham and Augustin, 1997; Burnham and Anderson, 2002), Bayesian methods and exponentiated BIC weights (Leamer, 1978; Hoeting et al., 1999), as well as cross-validation methods (Granger and Ramanathan, 1984). More recently, Hansen (2008) has proposed a method based on minimization of the Mallows (1973) criterion.

Model averaging has proven to be particularly effective in time series forecasting applications (where it is often referred to as *forecast combination*), yielding substantial reductions in forecast loss (for recent surveys, see Diebold and Lopez, 1996; Newbold and Harvey, 2001; or Timmerman, 2006). As yet, however, model averaging methods explicitly tailored to panel data models and applications have not been developed. In forecasting contexts, panel data models can offer substantial improvements in out-of-sample forecast accuracy compared to time series specifications (Baltagi, 2008), and so the development model averaging methods for panel data applications presents the prospect of further enhancements in forecast accuracy.

In this paper we develop new forecast averaging methods for a family of panel data models that are amenable to forecasting applications. We consider forecasting from a set of vector autoregressions (VARs) that permit limited cross section and time series heterogeneity (such as cross section and time period fixed effects) of the general form

$$y_{i,t} = \sum_{r=1}^{p} \beta_{i,r} \left(t-1\right)^{(r-1)} + \sum_{s=1}^{k} \alpha_s' \left(k\right) y_{i,t-s} + u_{i,t}\left(k\right); \quad k = 1, \dots, K, \tag{1}$$

where $y_{i,t}$ is an $m \times 1$ vector of variables observed for individual i in time period $t, k = 1, \dots, K$ indexes the lag order of the VARs in the candidate set of models, where K is the maximum lag order under consideration. The models are fitted using bias-corrected least squares (BCLS) in order to attenuate the Nickell bias of the OLS estimator, which inflates quadratic forecast loss (Greenaway-McGrevy 2013; 2015; 2018a). For each $k = 1, \dots, K$, we produce a forecast of $\{y_{i,T+1}\}_{i=1}^{n}$ based on the model fitted to data observed over $t = 1, \dots, T$ and $i = 1, \dots, n$. We then average these forecasts across $k = 1, \dots, K$ using weights $\{w_k\}_{k=1}^{K}$ restricted to the unit simplex (i.e., the scalar weights satisfy $w_k \in [0,1]$ and $\sum_{k=1}^{K} w_k = 1$.)

We provide an estimator of the quadratic risk of the averaged forecast, showing that the estimator is unbiased as n and T grow large under general conditions, and permitting the set of candidate models to grow at a restricted rate in the asymptotics (i.e., $K \to \infty$). Asymptotic theory is derived under the assumption that the true data generating process is an infinite order VAR, so that all fitted models within the candidate set are misspecified to some degree. Our theoretical framework therefore explicitly incorporates one of the primary motivations for adopting forecast combination in practice – model misspecification (see, e.g., p. 138-9 of Timmerman, 2006).[1] This setup also generates asymptotics that preserve the trade-off between specification error and model complexity that we face in practice when fitting models to data. Even in the limit, parsimonious models exhibit less model complexity (overfitting) and more misspecification, while larger models exhibit more model complexity but less misspecification.

Armed with this asymptotic theory, we turn our attention to analyzing specific averaging methods. The Mallows Model Averaging (MMA) method recently developed by Hansen (2007, 2008) selects weights in order to minimize the estimated quadratic risk of the weighted average forecast. MMA therefore penalizes the (positive) correlation between the forecast errors of different candidate models, meaning that models that with highly correlated forecast errors are collectively down-weighted. This feature of the weighting mechanism should, in theory, yield improvements in quadratic measure of forecast risk.[2]

We adapt the Hansen (2008) method to the panel data context by accounting for the effects of incidental parameters (such as cross section and period fixed effects), as well as weak cross sectional heterogeneity and dependence. To do so, we draw on the expressions of out-of-sample quadratic risk for panel data VARs developed in Greenaway-McGrevy (2015, 2018a), which in turn relies on much of the earlier research in time series forecasting and model selection (including Ing and Wei, 2003, 2005; and Findley and Wei, 1993).

We explore the performance of PMMA in a large Monte Carlo study. It outperforms other forecast averaging methods generalized to the panel context, such as equally-weighted (simple) averaging, BIC and AIC averaging, and constrained Granger and Ramanathan (1984) cross-validation.

We then showcase the panel data averaging methods in a couple of different empirical applications. First, we forecast US state GDP using a set of fixed effects autoregressive models. Second, we forecast US Metropolitan Statistical Area (MSA) employment and population using a set of bivariate vector autoregressions with fixed effects. In each of these applications, the PMMA forecasts tend to exhibit lower QFLs than extant averaging methods, including equally-weighted averages, BIC and AIC weights, and a constrained Granger and Ramanathan (1984) cross-validation.

Our work builds on a variety of work in forecast combination, panel data model selection and forecasting, and bias-correction. Forecast combination was first proposed by Bates and Granger (1969) and has subsequently remained an important topic within the forecasting literature. Detailed surveys include Clemen (1989), Granger (1989), Diebold and Lopez (1996), Newbold and Harvey (2001) and Timmerman (2006). In this paper we generalize several forecast averaging approaches to a panel data context, including the Mallows Model Averaging (MMA) proposed by Hansen (2007, 2008), a weighting based on the estimated QFR proposed in Greenaway-McGrevy (2018a, 2018b), the Granger and Ramanathan (1984) methods, and BIC and AIC averaging. Our work also relates

[1] More recently, Hansen (p., 342, 2008) states "[I]t is more appropriate to think of models as approximations, and that the "true" model is more complex than any of the models in our explicit class." Our framework permits a similar form of pervasive misspecification within the candidate set.

[2] In practice, simple averages that assign an equal weight to each model often outperform more complicated weighting designs (Timmerman, 2006), seemingly confounding this intuition. This finding does not hold in our panel applications, suggesting that the more complicated weighting techniques can work when applied to larger datasets.

to recent work on model selection in panel data settings, such as the KLIC and Schwarz criteria proposed by Lee and Phillips (2016), and the methods proposed by Han, Phillips and Sul (2017). Our work also builds on the panel data forecast model selection methods recently proposed by Greenaway-McGrevy (2018a, 2018b). We use the bias-corrected least squares estimator suggested by Hahn and Kuersteiner (2002) to construct forecasts. The paper therefore relates to a broad literature on panel data bias correction, including Kiviet (1995), Hahn and Kuersteiner (2011) and Phillips and Sul (2007).

The remainder of the paper is organized as follows. In the following section we introduce the family of models under consideration and introduce the method of estimation. In section three we produce asymptotic theory for the a general averaged forecast. A detailed Monte-Carlo study is presented in section four before we conclude.

2 Forecasting Models and Assumptions

The set of candidate forecasting models under consideration is given by (1) above, where $k = 1, 2, \ldots, K$ indexes the different models for some maximum lag order K. We outline the conditions imposed of the panel process $y_{i,t}$ in the following subsection before defining the bias-corrected least squares forecast.

2.1 Assumptions

We employ a set-up similar to that used in Greenaway-McGrevy (2018a). For completeness we repeat it here. The $m \times 1$ vector is generated by an infinite order VAR process of the form

$$y_{i,t} = \sum_{r=1}^{p} \beta_{i,r} (t-1)^{(r-1)} + \sum_{s=1}^{\infty} \alpha_s' y_{i,t-s} + e_{i,t},\ t = \ldots, -1, 0, 1, \ldots;\ i = 1, 2, \ldots \tag{2}$$

Under the assumptions outlined below, $e_{i,t}$ are independently distributed across time, but weak cross sectional dependence will be permitted.[3]

We impose the following assumptions on $\{\alpha_s\}_{s=1}^{\infty}$.

Assumption A

(*i*) $\det(I_m - \sum_{s=1}^{\infty} z^s \alpha_s) \neq 0$ *for* $|z| \leq 1$

(*ii*) $\sum_{s=1}^{\infty} s \|\alpha_s\| \leq C_\alpha < \infty$

(*iii*) *The diagonal elements of* α_s *are non-zero for infinitely many* s

Under Assumption A (i) the VAR process is invertible and can be expressed as a vector moving average process with moving average matrices $\{\theta_s\}_{s=0}^{\infty}$ recursively defined as $\theta_s := \sum_{r=0}^{s-1} \theta_r \alpha_{s-r}$, $s \geq 1$, with $\theta_0 = \alpha_0 = I_m$. Moreover, the absolute summability condition placed on $\{\alpha_s\}_{s=1}^{\infty}$ under

[3] Note that the order of the polynomial trend $(p-1)$ in the panel process coincides with that used in the candidate models (4).

(ii) implies that $\sum_{s=0}^{\infty} s\left\|\theta_s\right\|$ is also bounded (see Theorem 3.8.4 of Brillinger, 1981). Meanwhile (iii) ensures that the true DGP is an infinite order VAR.

We impose the following assumptions on the error process $e_{i,t}$.

Assumption B

(i) $\{e_{i,t}\}$ *is a sequence of independent random vectors with zero mean and variance* $\Sigma > 0$ *satisfying* $\sup_{-\infty<t<\infty} \mathrm{E}\left\|e_{i,t}\right\|^q < C_q$ *for* $q = 1, 2, \ldots$

(ii) For each $i = 1, 2 \ldots$ *and* $t = \ldots, -1, 0, 1, \ldots$, $\sup_{\underline{v}'\underline{v}=1} \mathrm{P}\left(b < \underline{v}' e_{i,t} < a\right) \leq C(a-b)^d$ *for scalars* $d > 0$, *a and b satisfying* $0 < a - b \leq c$, *and finite* C.

Under Assumption B (i) all moments of the errors are bounded, an assumption necessary to derive an asymptotic QFR expression under the relatively fast growth rate in model dimensionality permitted (see Assumption C below).[4] It also allows us to obtain asymptotic efficiency under the most general conditions on the relative rate of expansion in n and T. The sequence $\{e_{i,t}\}_{t=-\infty}^{\infty}$ is independently and heterogeneously distributed but with homogenous covariance $\Sigma_{i,i}$. Assumption (ii) characterizes the distribution of the error vector as *uniformly Lipschitz*, and it ensures that the expectation of the smallest eigenvalue of the regressor covariance matrix is bounded above zero for sufficiently large samples, even as k grows in the asymptotics (See, for example, Ing and Wei, 2003, 2005; Findley and Wei, 2002). Assumptions A, B (i) and (ii) apply standard conditions made in the time series forecasting literature to each time series in the panel (see, e.g., Ing and Wei, 2005).

It will be convenient to define

$$x_{i,t} := \sum_{s=1}^{\infty} \alpha_s' x_{i,t-s} + e_{i,t}, \tag{3}$$

such that $x_{i,t}$ is the stochastic component of the panel process $y_{i,t}$. In the following subsection we will consider fitting models of finite lag order k to the panel $y_{i,t}$ generated by (2). Because the panel process $y_{i,t}$ is a VAR(∞), these finite order fitted models are misspecified. For each $k = 1, 2, \ldots$ we define

$$s_{i,t}(k) := \sum_{s=1}^{\infty} \left(\alpha_s - \alpha_s(k)\right)' x_{i,t-s+1}, \quad \alpha_s(k) := 0 \text{ for } s \geq k+1,$$

where

$$\left(\{\alpha_s(k)\}_{s=1}^{k}\right) := \underset{a_s \in \mathbb{R}^{m\times m}}{\arg\min} \sum_{i=1}^{n} \mathrm{E}\left(\left\|x_{i,t+h} - \sum_{s=1}^{k} a_s' x_{i,t-s+1}\right\|^2\right),$$

such that $s_{i,t}(k)$ captures the specification error of the VAR(k) model. It will be convenient to decompose the panel process $\{x_{i,t}\}$ as

$$x_{i,t} = \sum_{s=1}^{k} \alpha_s(k)' x_{i,t-s} + u_{i,t}(k), \tag{4}$$

[4] Ing and Wei (2003) show that the bound can be relaxed at the expense of a slower rate of growth in the model dimension.

where $u_{i,t}(k) := e_{i,t} + s_{i,t}(k)$. We also define

$$\boldsymbol{\alpha}(k) := \left[\alpha_1'(k) : \cdots : \alpha_k'(k)\right]',$$

for each $k = 1, 2, \ldots, K$. The following captures the effect of specification error on quadratic measures of model loss.

$$\Lambda(k) = \textstyle\sum_{s,r=1}^{\infty} (\alpha_s - \alpha_s(k))' \, \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} x_{i,s}x_{i,r}'\right)(\alpha_r - \alpha_r(k))$$

such that $\Lambda(k) = \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} s_{i,t}(k)\, s_{i,t}(k)'\right)$. We signify the parameters with the number of lags k to indicate that the parameters that solve the minimization problem change with k in general.

The candidate set of regression models given in (4) are indexed by $k = 1, 2, \ldots, K$. We bound the maximum permissible lag order as follows.

Assumption C *The set of lags orders $k \leq K$, where K is a sequence of positive integers satisfying either*

$$C_1 \leq \lim_{n,T \longrightarrow \infty} \left(\frac{K^{1+\epsilon}}{\sqrt{n}} + \frac{K^{1+\epsilon}}{T}\right) \leq C_2 \tag{5}$$

or

$$C_1 \leq \lim_{n,T \longrightarrow \infty} \left(\frac{K^{1+\epsilon}}{\sqrt{T}}\right) \leq C_2 \tag{6}$$

for some $\epsilon > 0$ and finite constants C_1 and C_2.

The restriction imposed on the growth rate in K under either (5) or (6) ensure that the dependence between the sample used for estimation and the process to be forecast is asymptotically negligible, albeit under very different arguments. If $Kn^{-1/2} = o(1)$ as in (5), the dependence is shown to be negligible by exploiting the weak cross sectional dependence in the panel. This condition is less restrictive than (6) in "large n" panels ($n >> T$). By comparison, in the time series context, the maximum permissible rate of expansion in the lag order is restricted to be slower than $\sqrt{T}$ (Shibata, 1980; Ing and Wei, 2005). In "large n" panels, we can permit a faster rate of expansion in the maximum lag order, reflecting gains from pooling. If instead $KT^{-1/2} = o(1)$ as in (6), the dependence is shown to negligible by exploiting the weak time series dependence in the panel, as in Ing and Wei (2005). This condition is less restrictive than (5) in "large T" panels ($T >> n$).

2.2 Bias-Corrected Least Squares Forecasts

In this section we introduce the bias-corrected least squares (BCLS) estimator of the VAR(k) described in (1) and the associated forecast. It is well known that OLS exhibits an $O\left(T^{-1}\right)$ bias in the presence of parameterized heterogeneity (such as fixed effects), and that this bias inflates the quadratic forecast loss of the model (Greenaway-McGrevy, 2013; 2015; 2018a). Under this approach we use the analytic expression for the bias in order to make a first order correction to the OLS estimator before constructing the forecast. Unlike many other approaches to attenuate the bias, under standard conditions the bias-correction does not inflate the asymptotic variance of the OLS estimator (Hahn and Kuersteiner, 2002). Because quadratic measures of forecast loss are typically increasing in estimator variance, this feature makes bias-correction an attractive solution to the bias problem in a forecasting context.

The VAR(k) models are fitted to a common sample. That is, the forecasting models given in (1) are fitted to $\{y_{i,t}\}_{i=1,t=K+1}^{n,T}$ for each $k = 1, \ldots, K$. This means that for all lag orders $k < K$ we are omitting some data from the sample used to estimate the model. This will reduce forecast accuracy, especially since the cross-section specific parameters $\{\beta_i\}_{i=1}^{n}$ must be identified using only the time series information in the panel. In an extension in section 4.1 below we explore forecasts from VAR(k) models fitted to the full amount of available data.

For instructive clarity we will use matrix notation in order to describe the VAR(k) and BCLS estimation. Throughout we use $T_k := T - k$ for brevity. Let $\mathbf{Y}(k)$ be an $nT_K \times mk$ matrix of regressors of the VAR(k), ordered by consecutive time periods and then cross sectional units, as follows:

$$\mathbf{Y}(k) := \left[\mathbf{Y}_{1\cdot}'(k) : \cdots : \mathbf{Y}_{i\cdot}'(k) : \cdots : \mathbf{Y}_{n\cdot}'(k)\right]',$$

where $\mathbf{Y}_{i\cdot}(k)$ is a $T_K \times km$ matrix of the regressors for the ith time series in the panel, i.e.:

$$\mathbf{Y}_{i\cdot}(k) := \left[Y_{i,K}(k) : \cdots : Y_{i,t}(k) : \cdots : Y_{i,T-1}(k)\right]',$$

and

$$Y_{i,t}(k) := \left(y_{i,t}, y_{i,t-1}, \ldots, y_{i,t-k+1}\right)'.$$

Furthermore let $\mathbf{y}$ be an $nT_K \times m$ matrix of dependent variables ordered as follows:

$$\mathbf{y} := \left[\mathbf{y}_{1\cdot}' : \cdots : \mathbf{y}_{i\cdot}' : \cdots : \mathbf{y}_{n\cdot}'\right]', \quad \mathbf{y}_{i\cdot} := \left[y_{i,K+1} : \cdots : y_{i,t} : \cdots : y_{i,T}\right]'.$$

Next, let $\boldsymbol{\varsigma}_{T_K} := [\varsigma_1 : \cdots : \varsigma_{T-K}]'$ denote a $T_K \times p$ matrix of polynomial time trends, and define $\boldsymbol{\tau} := I_n \otimes \boldsymbol{\varsigma}_{T_K}$. We can then represent the VAR(k) in matrix form as

$$\mathbf{y} = \mathbf{Y}(k)\,\boldsymbol{\alpha}(k) + \boldsymbol{\tau}\boldsymbol{\beta} + \mathbf{u}(k), \quad \mathbf{u}(k) := \mathbf{s}(k) + \mathbf{e}, \tag{7}$$

where $\boldsymbol{\beta} := [\boldsymbol{\beta}_1 : \cdots : \boldsymbol{\beta}_i : \cdots : \boldsymbol{\beta}_n]'$,

$$\mathbf{s}(k) := \left[\mathbf{s}_{1\cdot}'(k) : \cdots : \mathbf{s}_{i\cdot}'(k) : \cdots : \mathbf{s}_{n\cdot}'(k)\right]', \ \mathbf{s}_{i\cdot}(k) := \left[s_{i,K}(k) : \cdots : s_{i,t}(k) : \cdots : s_{i,T-1}(k)\right]',$$

and

$$\mathbf{e} := \left[\mathbf{e}_{1\cdot}' : \cdots : \mathbf{e}_{i\cdot}' : \cdots : \mathbf{e}_{n\cdot}'\right]', \ \mathbf{e}_{i\cdot} := \left[e_{i,K+1} : \cdots : e_{i,t} : \cdots : e_{i,T}\right]'.$$

Next we define the BCLS estimator of the VAR(k) described in matrix notation in (7). We require some preliminary definitions. Let

$$\mathbf{Q}(k) := \tfrac{1}{nT_K}\mathbf{Y}(k)'\,\mathbf{M}_{\boldsymbol{\tau}}\mathbf{Y}(k), \quad \mathbf{v}(k) := \tfrac{1}{nT_K}\mathbf{Y}(k)'\,\mathbf{M}_{\boldsymbol{\tau}}\mathbf{y}, \tag{8}$$

and

$$\mathbf{P}_{\mathbf{M}_{\boldsymbol{\tau}}\mathbf{Y}(k)} := \mathbf{M}_{\boldsymbol{\tau}}\mathbf{Y}(k)\left(\mathbf{Y}(k)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{Y}(k)\right)^{-1}\mathbf{Y}(k)'\mathbf{M}_{\boldsymbol{\tau}}. \tag{9}$$

Then the bias-corrected least squares estimator of $\boldsymbol{\alpha}(k)$ is

$$\hat{\boldsymbol{\alpha}}(k) := \mathbf{Q}^{-1}(k)\mathbf{v}(k) + \tfrac{p}{T_K}\mathbf{Q}^{-1}(k)\hat{\boldsymbol{\xi}}(k), \tag{10}$$

where

$$\hat{\boldsymbol{\xi}}(k) := \left[\mathbf{1}_k \otimes \left(I_m - \mathbf{v}(k)'\mathbf{Q}^{-1}(k)\left(\mathbf{1}_k \otimes I_m\right)\right)^{-1}\right]\tfrac{1}{nT_K}\mathbf{y}'\left(\mathbf{M}_{\boldsymbol{\tau}} - \mathbf{P}_{\mathbf{M}_{\boldsymbol{\tau}}\mathbf{Y}(k)}\right)\mathbf{y}, \tag{11}$$

while the bias-corrected LS estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}(k) := \left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}'\left(\mathbf{y} - \mathbf{Y}(k)\hat{\boldsymbol{\alpha}}(k)\right). \tag{12}$$

The first term on the right hand side of (10) (i.e., $\mathbf{Q}^{-1}(k)\mathbf{v}(k)$) is the OLS estimator, while the second term is the Hahn and Kuersteiner (2002) bias correction. In the following remark we discuss the bias of the OLS estimator and the Hahn and Kuersteiner (2002) bias correction.

Remark 2.1 *It is well-known that the OLS estimator of fixed effects dynamic panel models exhibits $O\left(T^{-1}\right)$ bias (Nickell, 1981). Greenaway-McGrevy (2013) shows how this bias generalizes to correctly-specified VARs. The following equation characterizes the OLS bias for the multistep VAR(k) with heterogenous trends considered herein.*

$$\mathbf{Q}^{-1}(k)\mathbf{v}(k) - \boldsymbol{\alpha}(k) = \tfrac{p}{T_K}\boldsymbol{\Gamma}^{-1}(k)\boldsymbol{\xi}(k) + O_p\left(n^{-1/2}T^{-1}\right) + O_p\left(T^{-2}\right) + O_p\left(T^{-1}\left\|\Lambda(k)\right\|^{\frac{1}{2}}\right), \tag{13}$$

where

$$\boldsymbol{\Gamma}(k) := \mathrm{E}\left(\tfrac{1}{nT_K}\mathbf{Y}(k)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{Y}(k)\right),\ \boldsymbol{\xi}(k) := \left(1_k \otimes \left(I_m - \textstyle\sum_{s=1}^{\infty}\alpha_s'\right)^{-1}\right)\Sigma,$$

and where recall that the OLS estimator of $\boldsymbol{\alpha}(k)$ is given by $\mathbf{Q}^{-1}(k)\mathbf{v}(k)$. The first term on the right hand side of (13) is an analytic expression for the $O_p\left(T^{-1}\right)$ bias of the OLS estimator in correctly-specified models (i.e., $\boldsymbol{\alpha}_s = \mathbf{0}$ for all $s > k$), and coincides with the expression provided in Greenaway-McGrevy (2013) for the case of $p = 1$ and Hahn and Kuersteiner (2002) for the case of $p = h = 1$. The Hahn and Kuersteiner bias correction is based on this analytic expression. To see how, note that $\hat{\boldsymbol{\xi}}(k)$ serves as an estimator of $\boldsymbol{\xi}(k)$, and $\mathbf{Q}(k)$ serves as an estimator of $\boldsymbol{\Gamma}(k)$. The final $O_p\left(T^{-1}\left\|\Lambda(k)\right\|^{\frac{1}{2}}\right)$ term in (13) is proportional to $T^{-1}\left\|\boldsymbol{\Lambda}(k)\right\|^{\frac{1}{2}}$, where recall that $\boldsymbol{\Lambda}(k)$ is a quadratic measure of misspecification. In misspecified models – such as those considered herein – the final term in (13) will be also be manifest in the $O\left(T^{-1}\right)$ bias of the OLS estimator, and this component of the bias is ignored in the bias-correction. As discussed in Lee (2006), under misspecification the bias correction can in fact exacerbate – rather than attenuate – the bias of the estimator, which would inflate the risk of the forecast. However, as shown in Greenaway-McGrevy (2018a), the bias correction successfully eliminates the bias from the asymptotic QFR of the OLS forecast despite potential misspecification. This results from the fact that the ignored final term in (13) is proportional to the quadratic measure of misspecification, but it shrinks to zero at an

$O\left(T^{-1}\right)$ rate. Thus, in severely misspecified (i.e. parsimonious) models, the effect of bias on QFR is dominated by misspecification error, so the incorrect bias correction only has a second order effect on asymptotic forecast risk. Conversely, larger models exhibit less specification error, but the bias correction is more accurate.

In the following section we examine the weighted forecast obtained using the BCLS estimator.

3 Forecast Averaging and QFR Estimation

For each k, the $n \times m$ matrix of BCLS forecasts is

$$\hat{\mathbf{y}}_{\cdot T+1}(k) := \mathbf{Y}_{\cdot T}(k)\,\hat{\boldsymbol{\alpha}}(k) + \boldsymbol{\tau}_{T_K}\hat{\boldsymbol{\beta}}(k),$$

where $\mathbf{Y}_{\cdot T}(k) := [Y_{1,T}(k) : \cdots : Y_{i,T}(k) : \cdots : Y_{1,T}(k)]'$ and $\boldsymbol{\tau}_{T_K} := I_n \otimes \varsigma_{T_K}$ is $T_K n \times n$. The averaged BCLS forecasts are defined as

$$\hat{\mathbf{y}}_{\cdot T+1}(\mathbf{w}) = \textstyle\sum_{k=1}^{K} w_k \hat{\mathbf{y}}_{\cdot T+1}(k), \tag{14}$$

where $\mathbf{w} = (w_1, w_2, w_3, \ldots, w_K)'$ denotes a $K \times 1$ vector of weights with weights satisfying

$$w_k \in [0,1], \;\; \textstyle\sum_{k=1}^{K} w_k = 1.$$

Note that K will be permitted to grow at a restricted rate in what follows. The associated QFL of the weighted average forecast is

$$\mathcal{L}_{n,T}(\mathbf{w}) := \tfrac{1}{n}\left(\hat{\mathbf{y}}_{\cdot T+1}(\mathbf{w}) - \mathbf{y}_{\cdot T+1}\right)'\left(\hat{\mathbf{y}}_{\cdot T+1}(\mathbf{w}) - \mathbf{y}_{\cdot T+1}\right) \tag{15}$$

We will also use $\mathcal{L}_{n,T}(k)$ to denote the QFL in the case where $w_k = 1$ for some specific k (and thus $w_s = 0$ for all $s \neq k$). That is, only the kth model is used to make the forecast, i.e.

$$\mathcal{L}_{n,T}(k) := \tfrac{1}{n}\left(\hat{\mathbf{y}}_{\cdot T+1}(k) - \mathbf{y}_{\cdot T+1}\right)'\left(\hat{\mathbf{y}}_{\cdot T+1}(k) - \mathbf{y}_{\cdot T+1}\right).$$

Greenaway-McGrevy (2018a) provides an asymptotic expression of $\mathrm{E}\left(\mathcal{L}_{n,T}(k)\right)$ as $n \to \infty$ and $T \to \infty$. Using those results, it is straightforward to establish that

$$\mathrm{E}\left(\mathcal{L}_{n,T}(k)\right) = \Sigma\left(1 + p^2 T_K^{-1} + mk\,(nT_K)^{-1}\right) + \Lambda(k) + o\left(k\,(nT_K)^{-1}\right) + o\left(kT_K^{-2}\right).$$

In what follows, we normalize $\mathcal{L}_{n,T}(\mathbf{w})$ by subtracting $\Sigma\left(1 + p^2 T_K^{-1}\right)$, which is independent of the lag order k and hence the weights $\{w_k\}_{k=1}^{K}$. We thus refer to $\mathcal{L}_{n,T}(\mathbf{w}) - \Sigma\left(1 + p^2 T_K^{-1}\right)$ as the *second-order* QFL of the model.

Our goal is to find an estimator of the QFR of the averaged forecast (i.e., $\mathrm{E}\left(\mathcal{L}_{n,T}(\mathbf{w})\right)$). We consider estimators of the QFR defined as follows

$$\hat{L}_{n,T}(\mathbf{w}) := \mathcal{R}_{n,T}(\mathbf{w}) + \left(\tfrac{2m}{nT_K}\mathbf{k}'\mathbf{w} + \tfrac{p^2+p}{T_K}\right)\hat{\Sigma}(K), \tag{16}$$

where $\mathbf{k} = (1, 2, \ldots, k, \ldots, K)$ is a $K \times 1$ vector; $\mathcal{R}_{n,T}(\mathbf{w})$ is the in-sample quadratic loss of the weighted model, namely

$$\mathcal{R}_{n,T}(\mathbf{w}) := \tfrac{1}{nT_K}\hat{\mathbf{u}}(\mathbf{w})'\hat{\mathbf{u}}(\mathbf{w}), \tag{17}$$

where

$$\hat{\mathbf{u}}(\mathbf{w}) := \mathbf{y} - \hat{\mathbf{y}}(\mathbf{w}), \quad \hat{\mathbf{y}}(\mathbf{w}) := \textstyle\sum_{k=1}^{K} w_k \hat{\mathbf{y}}(k), \tag{18}$$

and

$$\hat{\mathbf{y}}(k) := \mathbf{Y}(k)\hat{\boldsymbol{\alpha}}(k) + \boldsymbol{\tau}\hat{\boldsymbol{\beta}}(k);$$

and $\hat{\Sigma}(K) := \frac{1}{nT_K - mK}\hat{\mathbf{u}}(K)'\hat{\mathbf{u}}(K)$ is used as an estimate of Σ, where $\hat{\mathbf{u}}(k) := \mathbf{y} - \hat{\mathbf{y}}(k)$ for $k = 1, \ldots, K$. The following result then establishes suitably-normalized $\hat{L}_{n,T}(\mathbf{w})$ as an unbiased estimator of the QFR, namely $\mathrm{E}(\mathcal{L}_{n,T}(\mathbf{w}))$, up to a second order approximation.

Theorem 3.1 *Under Assumptions A, B and C, as $n \to \infty$ and $T \to \infty$ subject to $n = O(T_K)$ we have*

$$\lim_{n,T\to\infty}\left|\frac{\mathrm{tr}\left(\Phi'\mathrm{E}\left[\hat{L}_{n,T}(\mathbf{w}) - \Sigma\left(1 + \frac{p^2}{T_K}\right)\right]\Phi\right)}{\mathrm{tr}\left(\Phi'\mathrm{E}\left[\mathcal{L}_{n,T}(\mathbf{w}) - \Sigma\left(1 + \frac{p^2}{T_K}\right)\right]\Phi\right)}\right| = 1,$$

for all $m \times m_0$ real matrices Φ satisfying $\|\Phi\| = 1$.

Here Φ reflects which elements of $y_{i,T+1}$ the researcher is interested in forecasting. The restriction on the spectral norm of Φ is a straightforward normalization. Note that we restrict n to grow at a rate no faster than T under Theorem 3.1. Later we relax this condition as an extension in Section 4.1.

3.1 Panel Mallows Model Averaging

Following Hansen (2008), Theorem 3.1 provides the basis for selection of the weights via minimization of the QFR estimator $\hat{L}_{n,T}(\mathbf{w})$. Therefore we choose $\mathbf{w}$ by minimizing $\mathrm{tr}\left(\Phi'\hat{L}_{n,T}(\mathbf{w})\Phi\right)$. In matrix notation we have

$$\mathrm{tr}\left(\Phi'\hat{L}_{n,T}(\mathbf{w})\Phi\right) = \mathbf{w}'\mathbf{D}(\boldsymbol{\Phi})\mathbf{w} + \left[\tfrac{2}{nT_K}\mathbf{k}'\mathbf{w} + \left(\tfrac{p^2+p}{T_K}\right)\right]\cdot \mathrm{tr}\left(\Phi'\hat{\Sigma}(K)\Phi\right),$$

where

$$\mathbf{D}(\Phi) := \tfrac{1}{n(T-K)}(\mathbf{I}_K \otimes \Phi\mathbf{1}_m)'\hat{\mathbf{u}}_{n,T}'\hat{\mathbf{u}}_{n,T}(\mathbf{I}_K \otimes \Phi\mathbf{1}_m),$$

and $\hat{\mathbf{u}}_{n,T} := [\hat{\mathbf{u}}(1) : \hat{\mathbf{u}}(2) : \cdots : \hat{\mathbf{u}}(K)]'$ is an $nT_K \times mK$ matrix of residuals. Finding a vector of weights that minimize $\mathrm{tr}\left(\Phi'\hat{L}_{n,T}(\mathbf{w})\Phi\right)$ can then be succinctly expressed as the solution quadratic programming problem:

$$\mathbf{w}^*(\Phi) = \arg\min_{\mathbf{w}} \left(\tfrac{1}{2}\mathbf{w}'\mathbf{D}(\mathbf{\Phi})\mathbf{w} + \tfrac{m}{nT_K}\mathbf{k}'\mathbf{w}\cdot \mathrm{tr}\left(\Phi'\hat{\Sigma}(K)\Phi\right)\right). \quad (19)$$

subject to[5]

$$\mathbf{1}_K'\mathbf{w} = 1, \quad w_k \in [0,1]$$

So-defined $\mathbf{w}^*(\Phi)$ is the panel VAR generalization of the MMA proposed by Hansen (2008).

3.2 Alternative Weighting Methods for Panel Data

In this subsection we consider how some well known weighting schemes can be generalized to our panel data context.

Bates and Granger (1969) originally proposed that forecast combination weights be based on estimates of the mean squared prediction error. This principle can be operationalized using a variety of estimators of quadratic forecast loss, such as final prediction error (Akaike, 1971; Shibata, 1980) or Mallows C_p (Mallows, 1973). In the panel data context, Greenaway-McGrevy (2018a, 2018b) generalizes model selection via minimization of estimated QFR, showing that it is asymptotically efficient. Following the principle first proposed by Bates and Granger (1969), these estimates of the QFR of the VAR(k) can also be used to construct weighted forecasts. Let

$$w_k^{**}(\Phi) := \frac{\mathrm{tr}\left(\Phi'\hat{L}_{n,T}(k)\Phi\right)}{\sum_{k=1}^{K}\mathrm{tr}\left(\Phi'\hat{L}_{n,T}(k)\Phi\right)}, \ \mathbf{w}^{**}(\Phi) := \left(w_1^{**}(\Phi), w_2^{**}(\Phi), \ldots, w_k^{**}(\Phi), \ldots, w_K^{**}(\Phi)\right)', \quad (20)$$

where

$$\hat{L}_{n,T}(k) := \tfrac{1}{nT_K}\hat{\mathbf{u}}(k)'\hat{\mathbf{u}}(k) + \left(\tfrac{k}{nT_K} + \tfrac{p^2+p}{T_K}\right)\hat{\Sigma}(K).$$

The above expression $\hat{L}_{n,T}(k)$ is the QFR estimator introduced by Greenaway-McGrevy (2018b) adapted to forecasting models fitted to data spanning $t = K+1,\ldots,T$. The primary difference between $\mathbf{w}^{**}(\Phi)$ and $\mathbf{w}^*(\Phi)$ is that the latter penalizes correlation between the forecasting models, whereas the former does not. Incidentally, the original model selection rule suggested by Greenaway-McGrevy (2018b) is

$$k^* = \arg\min_{1\le k\le K}\left(\mathrm{tr}\left(\Phi'\hat{L}_{n,T}(k)\Phi\right)\right). \quad (21)$$

We will explore the performance of this approach relative to the weighted forecasts in both the Monte Carlo and empirical studies.

Weights based on the relative size of information criteria proved popular in practice, with weights based on BIC (Leamer, 1978; Hoeting et al., 1999) and AIC (Buckland, Burnham and Augustin, 1997; Burnham and Anderson, 2002). Lee and Phillips (2016) generalize these criteria to panel models with incidental parameters, which yield a panel BIC and a panel Kullback-Leibler Information Criteria (KLIC). We use expressions derived by Lee and Phillips as the basis for a

[5] Note that we can omit the term $\left(\frac{p^2+p}{T_K}\right)\mathrm{tr}\left(\Phi\hat{\Sigma}(K)\right)$ from the minimization problem (19) because it is a constant independent of $\mathbf{w}$.

panel version of BIC and AIC averaging. However, because we use bias-corrected least squares, we omit the additional penalty term in the Lee and Phillips criteria included for the effect of incidental parameters on estimation. Our panel BIC weights are therefore

$$w_k^{\mathrm{BIC}} = \frac{e^{-\frac{1}{2}BIC(k)}}{\sum_{k=1}^{K} e^{-\frac{1}{2}BIC(k)}}, \quad BIC(k) = \ln\left(\hat{\sigma}^2(k)\right) + k\frac{\ln(nT_k)}{nT_k}, \tag{22}$$

where

$$\hat{\sigma}^2(k) := \frac{1}{nT_K}\mathrm{tr}\left(\Phi'\hat{\mathbf{u}}(k)'\,\hat{\mathbf{u}}(k)\,\Phi\right).$$

Meanwhile the panel AIC weights are given by

$$w_k^{\mathrm{AIC}} = \frac{e^{-\frac{1}{2}AIC(k)}}{\sum_{k=1}^{K} e^{-\frac{1}{2}AIC(k)}}, \quad AIC(k) = \ln\left(\hat{\sigma}^2(k)\right) + \frac{2k}{nT_k} \tag{23}$$

It is also straightforward to generalize the cross-validation approaches of Granger and Ramanathan (1984) to our panel data setting. Under this approach, weights are obtained by an in-sample BCLS fitting of the dependent variable on the predicted values from each VAR(k) (either with or without a constant). These weights can be constrained to satisfy $w_k^{\mathrm{GR}} \in (0,1)$ and $\sum_{k=1}^{K} w_k^{\mathrm{GR}} = 1$.

Finally, simple averaging, i.e., $w_k^{\mathrm{AVR}} = 1/K$ for all $k = 1, \ldots, K$, is also possible. In many applications, simple averaging (or slight variants thereof) perform well (Clemen, 1989; Chan, Stock and Watson, 1999; Timmerman, 2005).

4 Extensions

In this section we present several extensions of Theorem 3.1 and PMMA. First we present a refinement of the weighted QFR estimator for environments in which T is small relative to n. Next we consider augmenting the model with period fixed effects. Finally, we permit weak cross sectional heterogeneity and dependence in the panel.

4.1 A Refinement of the QFR Estimator for Small T Environments

In the previous section we considered a scenario in which each VAR(k) was fitted to a common sample spanning $T_K = T - K$ time series observations. There are, however, $T - k$ time series observations available to estimate the VAR(k). Because each cross-sectionally specific parameter vector β_i is identified from the time series dimension of the sample, there is likely to be a significant cost to limiting the estimation sample to T_K for panels in which T is small.

In this subsection we therefore construct forecasts using the full $T - k$ time series observations available when estimating the VAR(k) model. This is likely to be relevant for panels in which T is limited. Correspondingly, we permit n to grow faster than T in the asymptotics, which is more appropriate for panels in which n is much larger than T (such as micro-level panels).

In the interests of brevity, the BCLS estimators based on the full sample of data are defined in the Appendix (see (30) and (31)), and we use $\mathcal{L}_{n,T}^{(a)}(\mathbf{w})$ to denote the associated QFL of the forecast (see (32)).

Our constructed QFR estimators are slightly modified but still rely on the common sample of T_K time series observations in the sample. The estimators of the QFR are defined as follows

$$\hat{L}_{n,T}^{(a)}(\mathbf{w}) := \mathcal{R}_{n,T}(\mathbf{w}) + \left(\tfrac{2m}{nT_K}\mathbf{k}'\mathbf{w} + \tfrac{p^2}{T^2}\mathbf{w}'\mathbf{K}\mathbf{w} + \tfrac{p^2}{T} + \tfrac{p}{T_K}\right)\hat{\Sigma}(K) \tag{24}$$

where $\mathbf{K}$ is a $K \times K$ matrix in which the (r,s) element is $\min(r,s)$. The following result then establishes $\hat{L}_{n,T}^{(a)}(\mathbf{w})$ as an unbiased estimator of the QFR.

Theorem 4.1 *Under Assumptions A, B and C, as $n \to \infty$ and $T \to \infty$ we have*

$$\lim_{n,T\to\infty}\max_{1\le k\le K}\left|\frac{\operatorname{tr}\left(\Phi'\mathrm{E}\left[\hat{L}_{n,T}^{(a)}(\mathbf{w}) - \Sigma\left(1+\frac{p^2}{T}\right)\right]\Phi\right)}{\operatorname{tr}\left(\Phi'\mathrm{E}\left[\mathcal{L}_{n,T}^{(a)}(\mathbf{w}) - \Sigma\left(1+\frac{p^2}{T}\right)\right]\Phi\right)}\right| = 1,$$

for all $m \times m$ positive semi-definite matrices Φ satisfying $\|\Phi\| = 1$.

Panel MMA selection of the weights is as follows.

$$\mathbf{w}^*(\Phi) = \arg\min_{\mathbf{w}}\left(\tfrac{1}{2}\mathbf{w}'\left[\mathbf{D}(\boldsymbol{\Phi}) + \tfrac{p^2}{T^2}\mathbf{K}\cdot\operatorname{tr}\left(\Phi'\hat{\Sigma}(K)\Phi\right)\right]\mathbf{w} + \tfrac{m}{nT_K}\mathbf{k}'\mathbf{w}\cdot\operatorname{tr}\left(\Phi'\hat{\Sigma}(K)\Phi\right)\right),$$

subject to $\mathbf{1}_K'\mathbf{w} = 1$ and $w_k \in [0,1]$.

4.2 Period Fixed Effects

Period fixed effects present a simple method for permitting (and controlling for) cross-sectional dependence. To accommodate period fixed effects we introduce a composite error structure $e_{i,t} = \upsilon_t + \varepsilon_{i,t}$ and modify Assumption B as follows.

Assumption D *Let $e_{i,t} = \upsilon_t + \varepsilon_{i,t}$, where*

(i) *For each $i = 1, 2 \ldots$, $\{\varepsilon_{i,t}\}$ is a sequence of independent random vectors with zero mean and variance $\Sigma > 0$ satisfying $\sup_{-\infty<t<\infty}\mathrm{E}\|\varepsilon_{i,t}\|^q < C_q$ for $q = 1, 2, \ldots$*

(ii) *For each $i = 1, 2 \ldots$ and $t = \ldots, -1, 0, 1, \ldots$, $\sup_{\underline{v}'\underline{v}=1}\mathrm{P}(b < \underline{v}'\varepsilon_{i,t} < a) \le C(a-b)^d$ for scalars $d > 0$, a and b satisfying $0 < a - b \le c$, and finite C.*

(iii) *$\{\upsilon_t\}$ is a sequence of independent random vectors with zero mean and finite variance $\Sigma_\upsilon > 0$.*

Assumptions D (i) and (ii) apply Assumption B to the panel component of the composite error structure (i.e., $\varepsilon_{i,t}$), while Assumption D (iii) characterizes the common shock component of the composite error structure. Minimal restrictions are placed on the common shock υ_t since it is partialed out in estimation via the inclusion of period fixed effects in the model.

Estimation proceeds as described in Section 2.2 after first removing cross-sectional means from the relevant variables. We define $\ddot{\mathbf{Y}}(k) := \mathbf{M}_n \mathbf{Y}(k)$ and $\ddot{\mathbf{y}} := \mathbf{M}_n \mathbf{y}$, where $\mathbf{M}_n := \left[\left(I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\right) \otimes I_{T_K}\right]$. Then the BCLS estimators are defined as above in (10) and (12), but with $\ddot{\mathbf{Y}}(k)$ and $\ddot{\mathbf{y}}$ replacing $\mathbf{Y}(k)$ and $\mathbf{y}$ in the definitions of $\mathbf{Q}(k)$, $\mathbf{v}(k)$ and $\mathbf{P}_{\mathbf{M}_{\boldsymbol{\tau}}\mathbf{Y}(k)}$ given in (8) and (9). See (33) and (34) in the Appendix for the precise definitions. Letting $\hat{\hat{\boldsymbol{\alpha}}}(k)$ and $\hat{\hat{\boldsymbol{\beta}}}(k)$ denote the BCLS estimators of $\boldsymbol{\alpha}(k)$ and $\boldsymbol{\beta}(k)$, the averaged forecast is then defined as

$$\hat{\hat{\mathbf{y}}}_{\cdot T+1}(\mathbf{w}) := \sum_{k=1}^{K} w_k \hat{\hat{\mathbf{y}}}_{\cdot T+1}(k), \quad \hat{\hat{\mathbf{y}}}_{\cdot T+1}(k) := \ddot{\mathbf{Y}}_{\cdot T}(k)\hat{\hat{\boldsymbol{\alpha}}}(k) + \boldsymbol{\tau}_{T_K}\hat{\hat{\boldsymbol{\beta}}}(k),$$

where $\ddot{\mathbf{Y}}_{\cdot T}(k) := \left(I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'\right)\mathbf{Y}_{\cdot T}(k)$. The QFL is then

$$\mathcal{L}_{n,T}^{(b)}(\mathbf{w}) := \frac{1}{n}\left(\hat{\hat{\mathbf{y}}}_{\cdot T+1}(\mathbf{w}) - \ddot{\mathbf{y}}_{\cdot T+1}\right)'\left(\hat{\hat{\mathbf{y}}}_{\cdot T+1}(\mathbf{w}) - \ddot{\mathbf{y}}_{\cdot T+1}\right).$$

Next, we define $\hat{\hat{\mathbf{y}}}(k) := \ddot{\mathbf{Y}}(k)\hat{\hat{\boldsymbol{\alpha}}}(k) + \boldsymbol{\tau}\hat{\hat{\boldsymbol{\beta}}}(k)$, $\hat{\hat{\mathbf{u}}}(k) := \ddot{\mathbf{y}} - \hat{\hat{\mathbf{y}}}(k)$, and

$$\hat{\hat{\Sigma}}(K) := \frac{1}{nT_K - Km - n + 1}\hat{\hat{\mathbf{u}}}(k)'\hat{\hat{\mathbf{u}}}(k).$$

Then the QFR estimator is

$$\hat{L}_{n,T}^{(b)}(\mathbf{w}) := \mathcal{R}_{n,T}^{(b)}(\mathbf{w}) + \left(\frac{2m}{nT_K}\mathbf{k}'\mathbf{w} + \frac{p^2+p}{T_K} + \frac{2}{n}\right)\hat{\hat{\Sigma}}(K), \tag{25}$$

where $\mathcal{R}_{n,T}^{(b)}(\mathbf{w})$ is the in-sample QL of the fitted model with period fixed effects, namely

$$\mathcal{R}_{n,T}^{(b)}(\mathbf{w}) := \frac{1}{nT_K}\hat{\hat{\mathbf{u}}}(\mathbf{w})'\hat{\hat{\mathbf{u}}}(\mathbf{w}), \quad \hat{\hat{\mathbf{u}}}(\mathbf{w}) := \sum_{k=1}^{K} w_k \hat{\hat{\mathbf{u}}}(k).$$

The main difference compared to the estimator defined in (16) is the addition of the $\frac{2}{n}\hat{\hat{\Sigma}}(K)$ term. This accounts for the effect of the n period fixed effects on quadratic loss. Because this term is independent of $\mathbf{w}$, we can continue to select weights as defined in (19).

The following result then establishes $\hat{L}_{n,T}^{(b)}(\mathbf{w})$ as an unbiased estimator of the QFR.

Theorem 4.2 *Under Assumptions A, C and D, as $n \to \infty$ and $T \to \infty$ subject to $n = O(T_K)$ we have*

$$\lim_{n,T\to\infty}\left|\frac{\operatorname{tr}\left(\Phi'\mathrm{E}\left[\hat{L}_{n,T}^{(b)}(\mathbf{w}) - \Sigma\left(1 + \frac{p^2}{T_K} + \frac{1}{n}\right)\right]\Phi\right)}{\operatorname{tr}\left(\Phi'\mathrm{E}\left[\mathcal{L}_{n,T}^{(b)}(\mathbf{w}) - \Sigma\left(1 + \frac{p^2}{T_K} + \frac{1}{n}\right)\right]\Phi\right)}\right| = 1,$$

for all $m \times m$ positive semi-definite matrices Φ satisfying $\|\Phi\| = 1$.

Thus the

4.3 Panels with Weak Cross Sectional Heterogeneity and Dependence

Next we consider allowing weak cross section dependence and heterogeneity in the panel process. We impose the following assumptions on the error process $e_{i,t}$.

Assumption E

(*i*) *For each* $i = 1, 2\ldots$, $\{e_{i,t}\}$ *is a sequence of independent random vectors with zero mean and variance* $\Sigma_{i,i} > 0$ *satisfying* $\sup_{-\infty<t<\infty} \mathrm{E}\, \|e_{i,t}\|^q < C_q$ *for* $q = 1, 2, \ldots$

(*ii*) *For each* $i = 1, 2\ldots$ *and* $t = \ldots, -1, 0, 1, \ldots$, $\sup_{\underline{v}'\underline{v}=1} \mathrm{P}\left(b < \underline{v}' e_{i,t} < a\right) \leq C(a-b)^d$ *for scalars* $d > 0$, a *and* b *satisfying* $0 < a - b \leq c$, *and finite* C.

(*iii*) $\frac{1}{n} \sum_{i,j=1}^{n} \|\Sigma_{i,j}\| \leq C_\Sigma$ *for all* n, *where* $\Sigma_{i,j} := \mathrm{E}\left(e_{i,t} e_{j,t}'\right)$ *for each* $i, j = 1, 2, \ldots$

Assumption E (iii) characterizes the cross-sectional dependence in the panel. Although $e_{i,t}$ is independently distributed over time under Assumption E (i), it is not necessarily independently distributed between cross sections. Assumption E (iii) therefore limits the cross section correlation in $e_{i,t}$. Specifically, we only permit weak-form cross sectional dependence (also see (26) below). The assumption permits spatial autoregressive, spatial moving average, and spatial error components models (see Pesaran and Tosetti, 2011). Assumptions A and E (iii) ensure that the panel vectors $x_{i,t}$ obey *weak dependence* conditions of the form

$$\frac{1}{n} \sum_{i,j=1}^{n} \frac{1}{T} \sum_{s,t=1}^{T} \mathrm{E}\left(x_{i,t}' x_{j,s}\right) \leq \mathrm{tr}\left(\frac{1}{n} \sum_{i,j=1}^{n} \sum_{s,r=0}^{\infty} \theta_s' \Sigma_{i,j} \theta_r\right) \leq m \|C_\theta\|^2 \|C_\Sigma\| < \infty \tag{26}$$

(see Chudik, Pesaran and Tosetti, 2011), where $\sum_{s=0}^{\infty} \|\theta_s\|^2 \leq C_\theta$ for some finite C_θ.

Estimation and forecasting proceeds as above. The main difference is that our QFR estimator must account for the cross section dependence and heterogeneity. The QFR estimator is defined as follows

$$\hat{L}_{n,T}^{(c)}(\mathbf{w}) := \mathcal{R}_{n,T}(\mathbf{w}) + \frac{2}{nT_K} (\mathbf{k} \otimes I_m)' \mathbf{B} (\mathbf{w} \otimes I_m) + \left(\frac{p^2+p}{T_K}\right) \hat{\Sigma}(K) \tag{27}$$

where $\mathbf{B}$ is a $mK \times mK$ block diagonal matrix with the $m \times m$ matrices $\left\{\hat{\Pi}(k) \cdot \Phi\right\}_{k=1}^{K}$ along the diagonal, where

$$\hat{\Pi}(k) := \frac{1}{nk} \sum_{i,j=1}^{n} \frac{1}{T_K} \sum_{t=K}^{T-1} \mathrm{tr}\left(\hat{u}_{i,t+1}(K)\, \hat{u}_{j,t+1}(K)'\right) \cdot \mathrm{tr}\left[\tilde{Y}_{i,t}(k)\, \tilde{Y}_{j,t+s}(k)' \mathbf{Q}(k)^{-1}\right]$$

and $\hat{u}_{i,t+1}(k)'$ is the $[(i-1)T_K + t - K + 1]$th row of $\hat{\mathbf{u}}(k)$, while $\tilde{Y}_{i,t}(k)$ is the $[(i-1)T_K + t - K + 1]$th row of $\mathbf{M}_\tau \mathbf{Y}(k)$. Under independence and homoskedasticity, $\hat{\Pi}(k)$ simplifies to $\hat{\Sigma}(K)$, which yields $\hat{L}_{n,T}^{(c)}(\mathbf{w}) = \mathcal{R}_{n,T}(\mathbf{w}) + \left(\frac{2}{nT_K}\mathbf{k}'\mathbf{w} + \frac{p^2+p}{T_K}\right) \hat{\Sigma}(K) = \hat{L}_{n,T}(\mathbf{w})$.

The following result then establishes $\hat{L}_{n,T}^{(c)}(\mathbf{w})$ as an unbiased estimator of the QFR.

Theorem 4.3 *Under Assumptions A, E and C, as* $n \to \infty$ *and* $T \to \infty$ *subject to* $n = O(T_K)$ *we have*

$$\lim_{n,T\to\infty} \left| \frac{\mathrm{tr}\left(\Phi' \mathrm{E}\left[\hat{L}_{n,T}^{(c)}(\mathbf{w}) - \Sigma\left(1 + \frac{p^2}{T_K}\right)\right] \Phi\right)}{\mathrm{tr}\left(\Phi' \mathrm{E}\left[\mathcal{L}_{n,T}(\mathbf{w}) - \Sigma\left(1 + \frac{p^2}{T_K}\right)\right] \Phi\right)} \right| = 1,$$

for all $m \times m$ *positive semi-definite matrices* Φ *satisfying* $\|\Phi\| = 1$.

The panel MMA weights are then as follows.

$$\mathbf{w}^*(\Phi) = \arg\min_{\mathbf{w}} \left(\tfrac{1}{2}\mathbf{w}'\mathbf{D}(\boldsymbol{\Phi})\,\mathbf{w} + \tfrac{1}{nT_K}\mathbf{k}'\mathbf{B}(\Phi)\,\mathbf{w} \right),$$

subject to $\mathbf{1}'_K\mathbf{w} = 1$ and $w_k \in [0,1]$, where $\mathbf{D}(\Phi)$is a $K \times K$ diagonal matrix with $\left\{ \text{tr}\left(\hat{\Pi}(k)\,\Phi \right) \right\}_{k=1}^{K}$ along the principal diagonal.

5 Monte Carlo Studies

We conduct Monte Carlo studies in order to verify the theoretical results given above and to investigate the different averaging methods. We fit a set of AR(k) models to an ARMA$(1,1)$ process of the general form

$$y_{i,t} = \rho y_{i,t-1} + \theta e_{i,t-1} + e_{i,t},\ i = 1,\ldots,n;\ t = 2,\ldots,T+1, \tag{28}$$

and where $e_{i,t} \sim iidN(0,1)$.

For each simulated panel we fit AR(k) models for each $k = 1,\ldots,K$. We construct both the quadratic forecast loss (QFL) of the forecasts $\{\hat{y}_{i,T+1}(k)\}_{i=1}^{n}$ and the estimated QFR for each fitted AR(k) model as well as weighted average forecasts $\{\hat{y}_{i,T+1}(\mathbf{w}^*)\}_{i=1}^{n}$ and $\{\hat{y}_{i,T+1}(\mathbf{w}^{**})\}_{i=1}^{n}$. The maximum lag order K is determined by the sample dimensions of the simulated panel. We set K to the smallest integer greater than $\max\left(\sqrt{2T}, \min\left(\sqrt{2n}, 0.5T\right)\right)$, and we consider all combinations of $n \in \{10, 15, 25, 50, 100, 200\}$ and $T \in \{10, 15, 25, 50, 100, 200\}$.

The fitted models include cross section fixed effects ($p = 1$). In our finite sample investigation we also compute quantities which form the basis for other lag selection criteria, as we will outline below in more detail. Each simulation is replicated 20,000 times.

We consider all combinations of $\rho \in \{0.8, -0.8\}$ and $\theta \in \{0.5, 0.6, 0.7, 0.8\}$, which provides substantial variation in the degree of misspecification in the fitted models. For example, when $\rho = 0.8$ and $\theta = 0.8$ there is rather severe misspecification, reflected in that the lag order with the smallest QFL (on average) is $k = 17$ when $n = T = 200$ (i.e., the largest sample size considered). In contrast, when $\rho = -0.8$ and $\theta = 0.5$ we have a moderate degree of misspecification, which is reflected in that the lag order with the smallest QFL (on average) is $k = 6$ when $n = T = 200$.

We consider seven different forecast averaging methods: (i) the panel MMA described above in (19); (ii) estimated QFR weights described in (20) ('estimated QFR weights'),(iii) exponential BIC weights (22); exponential AIC weights (23); constrained Granger-Ramanathan averaging (with weights constrained to $w_k^{\text{GR}} \in (0,1)$ and $\sum_{k=1}^{K} w_k^{\text{GR}} = 1$); simple averaging, i.e., $w_k^{\text{AVR}} = 1/K$ for all $k = 1,\ldots,K$; and (vii) (single) model selection based on the minimization of the estimated QFR (21).

Tables 1 through 8 exhibit the QFLs of each forecast averaging approach, averaged over 20,000 replications. We make two adjustments to the QFLs in order to facilitate comparisons across the different averaging methods. First, we subtract $\left(1 + \frac{1}{T}\right)$, so that the QFLs are second order, in the sense that the first order terms independent of k have been removed. Second, we scale the second-order QFLs by that of PMMA, so that a ratio greater (smaller) than one means that the PMMA had a smaller (larger) QFL than the forecast method, on average.

The tables demonstrate that PMMA has the smallest (average) QFL in almost all simulation designs and sample sizes. Exceptions typically occur when $n = 10$. These results are consistent with the reasoning underlying the design of the PMMA, which is is based on minimizing the QFL of the weighted forecast.

Table 1: Scaled normalized QFLs

n	T	K	(A)	(B)	(C)	(D)	(E)	(F)	(G)
10	10	4	1	1.875	1.409	1.030	1.011	1.135	1.168
	15	5	1	1.724	1.323	1.148	1.161	1.196	1.142
	25	7	1	1.055	1.327	1.142	1.154	1.191	1.063
	50	10	1	1.052	1.477	1.209	1.224	1.278	1.107
	100	14	1	1.145	1.685	1.306	1.322	1.402	1.126
	200	20	1	1.324	1.765	1.328	1.342	1.437	1.128
15	10	5	1	1.450	2.248	0.999	0.922	1.352	1.160
	15	5	1	1.887	1.546	1.308	1.320	1.374	1.137
	25	7	1	1.067	1.523	1.279	1.290	1.344	1.035
	50	10	1	1.029	1.720	1.375	1.388	1.464	1.075
	100	14	1	1.104	2.068	1.557	1.573	1.688	1.112
	200	20	1	1.249	2.127	1.564	1.577	1.705	1.111
25	10	5	1	1.603	1.797	1.048	0.998	1.257	1.151
	15	7	1	2.811	1.156	1.026	1.033	1.062	1.268
	25	7	1	1.077	1.717	1.419	1.429	1.499	1.021
	50	10	1	1.020	1.963	1.550	1.561	1.658	1.045
	100	14	1	1.050	2.331	1.749	1.760	1.898	1.069
	200	20	1	1.124	2.314	1.709	1.719	1.861	1.075
50	10	5	1	1.824	1.622	1.116	1.089	1.258	1.161
	15	8	1	3.580	1.238	1.027	1.021	1.087	1.224
	25	10	1	1.538	1.837	1.493	1.503	1.582	1.039
	50	10	1	1.017	2.432	1.861	1.869	2.010	1.028
	100	14	1	1.044	3.292	2.358	2.368	2.598	1.056
	200	20	1	1.120	3.740	2.564	2.575	2.861	1.092
100	10	5	1	1.986	1.576	1.169	1.155	1.285	1.167
	15	8	1	3.911	1.255	1.069	1.067	1.124	1.222
	25	13	1	1.903	1.353	1.113	1.120	1.175	1.032
	50	14	1	1.057	1.835	1.417	1.422	1.523	1.138
	100	14	1	1.046	4.208	2.935	2.943	3.263	1.054
	200	20	1	1.118	5.877	3.851	3.861	4.363	1.093
200	10	5	1	2.064	1.546	1.192	1.184	1.294	1.148
	15	8	1	3.988	1.271	1.096	1.095	1.148	1.248
	25	13	1	1.833	1.442	1.171	1.175	1.241	0.989
	50	20	1	1.090	1.331	1.101	1.103	1.157	1.205
	100	20	1	1.229	2.528	1.809	1.812	1.989	1.210
	200	20	1	1.080	6.025	3.975	3.980	4.493	1.071

Data generated according to (28) with $\rho = 0.8$ and $\theta = 0.8$. Column (A) is Panel Mallows model averaging; (B) is estimated QFR weights; (C) is Granger-Ramanathan; (D) is BIC averaging; (E) is AIC averaging; (F) is simple averaging; (G) is Mallows minimization (i.e., choosing the model with the lowest estimated QFR; not model averaging). Table entries are the ratio of the second order QFL (averaged across 20,000 simulations) to that of the PMMA forecast.

Table 2: Scaled normalized QFLs

n	T	K	(A)	(B)	(C)	(D)	(E)	(F)	(G)
10	10	4	1	1.944	1.500	1.036	1.010	1.164	1.166
	15	5	1	1.963	1.227	1.079	1.091	1.120	1.179
	25	7	1	1.108	1.142	1.017	1.026	1.049	1.091
	50	10	1	1.214	1.193	1.029	1.039	1.071	1.149
	100	14	1	1.462	1.328	1.102	1.112	1.160	1.151
	200	20	1	1.760	1.420	1.169	1.177	1.233	1.125
15	10	5	1	1.432	2.755	1.076	0.978	1.542	1.139
	15	5	1	2.253	1.481	1.261	1.274	1.323	1.154
	25	7	1	1.118	1.312	1.134	1.143	1.181	1.072
	50	10	1	1.151	1.353	1.135	1.144	1.192	1.126
	100	14	1	1.362	1.550	1.245	1.255	1.324	1.139
	200	20	1	1.610	1.628	1.307	1.314	1.389	1.108
25	10	5	1	1.570	2.175	1.122	1.059	1.411	1.137
	15	7	1	3.351	1.118	0.997	1.000	1.032	1.224
	25	7	1	1.130	1.493	1.267	1.275	1.328	1.044
	50	10	1	1.079	1.506	1.245	1.252	1.314	1.111
	100	14	1	1.232	1.704	1.364	1.371	1.453	1.101
	200	20	1	1.332	1.672	1.351	1.357	1.434	1.079
50	10	5	1	1.775	1.938	1.188	1.154	1.395	1.192
	15	8	1	4.243	1.367	1.058	1.046	1.146	1.213
	25	10	1	1.439	1.398	1.196	1.202	1.249	1.049
	50	10	1	1.041	1.822	1.445	1.451	1.545	1.109
	100	14	1	1.199	2.271	1.707	1.714	1.854	1.125
	200	20	1	1.363	2.458	1.810	1.816	1.978	1.111
100	10	5	1	1.923	1.858	1.236	1.217	1.410	1.261
	15	8	1	4.648	1.320	1.063	1.056	1.139	1.201
	25	13	1	1.872	1.110	0.969	0.973	1.006	1.003
	50	14	1	1.089	1.359	1.127	1.130	1.186	1.275
	100	14	1	1.193	2.858	2.063	2.069	2.272	1.152
	200	20	1	1.407	3.737	2.567	2.573	2.870	1.160
200	10	5	1	1.987	1.801	1.249	1.239	1.405	1.320
	15	8	1	4.826	1.293	1.060	1.057	1.130	1.200
	25	13	1	1.815	1.150	0.991	0.993	1.033	0.997
	50	20	1	1.105	1.053	0.947	0.948	0.973	1.337
	100	20	1	1.451	1.740	1.357	1.359	1.455	1.393
	200	20	1	1.267	3.656	2.546	2.550	2.835	1.167

Data generated according to (28) with $\rho = 0.8$ and $\theta = 0.7$. Column (A) is Panel Mallows model averaging; (B) is estimated QFR weights; (C) is Granger-Ramanathan; (D) is BIC averaging; (E) is AIC averaging; (F) is simple averaging; (G) is Mallows minimization (i.e., choosing the model with the lowest estimated QFR; not model averaging). Table entries are the ratio of the second order QFL (averaged across 20,000 simulations) to that of the PMMA forecast.

Table 3: Scaled normalized QFLs

n	T	K	(A)	(B)	(C)	(D)	(E)	(F)	(G)
10	10	4	1	1.908	1.611	1.044	1.012	1.202	1.132
	15	5	1	2.147	1.075	0.981	0.991	1.007	1.200
	25	7	1	1.166	0.972	0.913	0.918	0.928	1.119
	50	10	1	1.417	1.014	0.939	0.944	0.959	1.158
	100	14	1	1.792	1.137	1.030	1.034	1.058	1.176
	200	20	1	2.210	1.277	1.154	1.157	1.187	1.159
15	10	5	1	1.361	3.465	1.198	1.072	1.818	1.104
	15	5	1	2.513	1.281	1.130	1.141	1.173	1.183
	25	7	1	1.153	1.089	0.996	1.002	1.021	1.097
	50	10	1	1.331	1.113	1.007	1.012	1.035	1.166
	100	14	1	1.620	1.237	1.090	1.095	1.129	1.169
	200	20	1	1.910	1.350	1.196	1.200	1.237	1.118
25	10	5	1	1.446	2.679	1.225	1.145	1.619	1.107
	15	7	1	3.795	1.104	0.989	0.987	1.024	1.204
	25	7	1	1.133	1.213	1.089	1.095	1.123	1.081
	50	10	1	1.183	1.193	1.064	1.069	1.099	1.142
	100	14	1	1.405	1.329	1.166	1.170	1.210	1.112
	200	20	1	1.494	1.345	1.196	1.199	1.236	1.067
50	10	5	1	1.571	2.303	1.253	1.210	1.539	1.160
	15	8	1	4.836	1.661	1.197	1.176	1.328	1.197
	25	10	1	1.365	1.139	1.042	1.045	1.068	1.088
	50	10	1	1.096	1.357	1.166	1.169	1.217	1.174
	100	14	1	1.380	1.638	1.355	1.359	1.432	1.168
	200	20	1	1.563	1.749	1.438	1.442	1.522	1.119
100	10	5	1	1.647	2.130	1.259	1.235	1.498	1.253
	15	8	1	5.264	1.541	1.167	1.156	1.276	1.195
	25	13	1	1.982	0.995	0.922	0.923	0.941	1.036
	50	14	1	1.151	1.089	0.984	0.985	1.011	1.400
	100	14	1	1.370	1.949	1.544	1.547	1.654	1.254
	200	20	1	1.656	2.465	1.880	1.883	2.037	1.206
200	10	5	1	1.675	2.026	1.251	1.238	1.466	1.346
	15	8	1	5.413	1.450	1.125	1.119	1.222	1.146
	25	13	1	1.993	1.015	0.932	0.933	0.955	1.025
	50	20	1	1.133	0.920	0.882	0.881	0.891	1.425
	100	20	1	1.597	1.309	1.133	1.134	1.180	1.541
	200	20	1	1.416	2.362	1.832	1.834	1.975	1.226

Data generated according to (28) with $\rho = 0.8$ and $\theta = 0.6$. Column (A) is Panel Mallows model averaging; (B) is estimated QFR weights; (C) is Granger-Ramanathan; (D) is BIC averaging; (E) is AIC averaging; (F) is simple averaging; (G) is Mallows minimization (i.e., choosing the model with the lowest estimated QFR; not model averaging). Table entries are the ratio of the second order QFL (averaged across 20,000 simulations) to that of the PMMA forecast.

Table 4: Scaled normalized QFLs

n	T	K	(A)	(B)	(C)	(D)	(E)	(F)	(G)
10	10	4	1	1.760	1.765	1.074	1.033	1.267	1.109
	15	5	1	2.304	0.972	0.927	0.933	0.939	1.192
	25	7	1	1.226	0.873	0.859	0.860	0.862	1.188
	50	10	1	1.583	0.915	0.895	0.895	0.900	1.177
	100	14	1	2.132	1.074	1.036	1.036	1.046	1.174
	200	20	1	2.714	1.272	1.224	1.223	1.238	1.146
15	10	5	1	1.266	4.462	1.400	1.236	2.231	1.077
	15	5	1	2.749	1.110	1.028	1.035	1.052	1.194
	25	7	1	1.178	0.942	0.908	0.911	0.917	1.146
	50	10	1	1.500	0.987	0.950	0.951	0.960	1.166
	100	14	1	1.890	1.103	1.046	1.047	1.061	1.198
	200	20	1	2.178	1.221	1.162	1.162	1.178	1.140
25	10	5	1	1.283	3.378	1.397	1.294	1.930	1.060
	15	7	1	4.178	1.200	1.070	1.061	1.109	1.170
	25	7	1	1.131	1.018	0.967	0.970	0.981	1.108
	50	10	1	1.312	1.038	0.988	0.990	1.002	1.165
	100	14	1	1.545	1.131	1.067	1.068	1.084	1.141
	200	20	1	1.602	1.172	1.114	1.115	1.131	1.070
50	10	5	1	1.315	2.769	1.359	1.305	1.739	1.087
	15	8	1	5.169	2.222	1.504	1.468	1.705	1.133
	25	10	1	1.331	1.004	0.968	0.970	0.978	1.117
	50	10	1	1.182	1.101	1.023	1.025	1.045	1.249
	100	14	1	1.527	1.273	1.156	1.158	1.189	1.210
	200	20	1	1.737	1.381	1.252	1.254	1.289	1.117
100	10	5	1	1.301	2.419	1.288	1.261	1.596	1.131
	15	8	1	5.658	2.092	1.496	1.476	1.667	1.185
	25	13	1	2.043	0.968	0.933	0.933	0.943	1.024
	50	14	1	1.215	0.951	0.915	0.916	0.925	1.514
	100	14	1	1.531	1.423	1.249	1.251	1.298	1.334
	200	20	1	1.900	1.766	1.510	1.512	1.582	1.265
200	10	5	1	1.270	2.207	1.228	1.213	1.496	1.173
	15	8	1	5.828	2.002	1.473	1.462	1.628	1.223
	25	13	1	2.061	0.972	0.933	0.933	0.944	1.033
	50	20	1	1.154	0.860	0.853	0.853	0.855	1.493
	100	20	1	1.713	1.088	1.020	1.020	1.039	1.665
	200	20	1	1.536	1.669	1.447	1.447	1.509	1.289

Data generated according to (28) with $\rho = 0.8$ and $\theta = 0.5$. Column (A) is Panel Mallows Model Averaging; (B) is estimated QFR weights; (C) is Granger-Ramanathan; (D) is BIC averaging; (E) is AIC averaging; (F) is simple averaging; (G) is Mallows minimization (i.e., choosing the model with the lowest estimated QFR; not model averaging). Table entries are the ratio of the second order QFL (averaged across 20,000 simulations) to that of the PMMA forecast.

Table 5: Scaled normalized QFLs

n	T	K	(A)	(B)	(C)	(D)	(E)	(F)	(G)
10	10	4	1	1.063	11.876	5.920	5.566	7.632	0.994
	15	5	1	1.239	3.898	3.304	3.217	3.493	1.030
	25	7	1	2.099	2.487	2.393	2.353	2.424	1.137
	50	10	1	3.908	2.062	2.043	2.021	2.049	1.249
	100	14	1	6.346	2.528	2.516	2.493	2.520	1.258
	200	20	1	7.243	2.809	2.800	2.781	2.803	1.194
15	10	5	1	1.014	92.647	14.858	12.026	33.139	0.987
	15	5	1	1.055	4.376	3.771	3.693	3.964	1.045
	25	7	1	1.562	2.760	2.664	2.629	2.695	1.080
	50	10	1	3.130	1.965	1.953	1.938	1.957	1.223
	100	14	1	4.522	2.056	2.049	2.038	2.052	1.160
	200	20	1	4.290	1.936	1.933	1.925	1.934	1.056
25	10	5	1	0.998	91.934	17.150	14.716	34.146	0.998
	15	7	1	0.994	37.368	13.992	12.950	19.889	1.006
	25	7	1	1.125	2.575	2.501	2.480	2.525	1.082
	50	10	1	1.982	1.663	1.656	1.650	1.658	1.099
	100	14	1	2.354	1.461	1.459	1.456	1.460	1.080
	200	20	1	2.148	1.328	1.327	1.325	1.327	1.027
50	10	5	1	1.000	102.683	21.179	19.315	39.174	1.000
	15	8	1	1.000	206.571	34.755	30.294	69.882	1.000
	25	10	1	0.998	12.371	9.799	9.619	10.589	1.056
	50	10	1	1.670	2.072	2.064	2.059	2.067	1.130
	100	14	1	2.577	1.702	1.700	1.698	1.701	1.114
	200	20	1	2.509	1.470	1.470	1.468	1.470	1.088
100	10	5	1	1.000	108.903	23.450	22.218	42.013	1.000
	15	8	1	1.000	229.375	40.589	37.473	78.422	1.000
	25	13	1	1.030	173.394	51.378	48.833	78.883	1.011
	50	14	1	0.991	4.812	4.745	4.727	4.768	1.158
	100	14	1	3.036	2.428	2.426	2.423	2.426	1.286
	200	20	1	3.801	2.080	2.080	2.078	2.080	1.232
200	10	5	1	1.000	110.697	24.390	23.646	42.984	1.000
	15	8	1	1.000	229.535	41.861	40.006	79.138	1.000
	25	13	1	1.001	159.021	48.365	47.005	73.287	1.002
	50	20	1	1.076	13.002	11.630	11.560	12.068	1.081
	100	20	1	1.263	2.231	2.228	2.226	2.229	1.126
	200	20	1	1.726	1.375	1.375	1.375	1.375	1.094

Data generated according to (28) with $\rho = -0.8$ and $\theta = 0.8$. Column (A) is Panel Mallows model averaging; (B) is estimated QFR weights; (C) is Granger-Ramanathan; (D) is BIC averaging; (E) is AIC averaging; (F) is simple averaging; (G) is Mallows minimization (i.e., choosing the model with the lowest estimated QFR; not model averaging). Table entries are the ratio of the second order QFL (averaged across 20,000 simulations) to that of the PMMA forecast.

Table 6: Scaled normalized QFLs

n	T	K	(A)	(B)	(C)	(D)	(E)	(F)	(G)
10	10	4	1	1.077	9.339	4.354	4.071	5.771	1.006
	15	5	1	1.169	2.830	2.329	2.265	2.487	1.114
	25	7	1	1.512	1.554	1.493	1.469	1.513	1.264
	50	10	1	2.032	1.056	1.051	1.042	1.053	1.339
	100	14	1	2.683	1.070	1.070	1.062	1.070	1.275
	200	20	1	2.913	1.126	1.126	1.120	1.126	1.186
15	10	5	1	1.014	68.709	9.742	7.674	23.460	0.995
	15	5	1	1.048	2.792	2.347	2.296	2.487	1.088
	25	7	1	1.205	1.484	1.430	1.413	1.448	1.245
	50	10	1	1.678	1.006	1.004	0.998	1.004	1.321
	100	14	1	2.201	0.995	0.995	0.991	0.995	1.218
	200	20	1	2.273	1.023	1.024	1.021	1.024	1.116
25	10	5	1	1.001	64.761	10.788	9.060	22.869	1.000
	15	7	1	1.008	21.845	7.350	6.734	10.914	1.015
	25	7	1	1.017	1.362	1.322	1.312	1.335	1.234
	50	10	1	1.314	0.986	0.985	0.982	0.985	1.206
	100	14	1	1.626	0.974	0.974	0.973	0.974	1.126
	200	20	1	1.625	0.997	0.997	0.996	0.997	1.070
50	10	5	1	1.000	65.905	12.224	11.001	23.839	1.000
	15	8	1	1.002	95.911	14.375	12.275	30.722	1.002
	25	10	1	1.115	4.235	3.210	3.147	3.519	1.173
	50	10	1	1.143	0.980	0.979	0.978	0.980	1.202
	100	14	1	1.578	0.975	0.975	0.975	0.975	1.155
	200	20	1	1.749	1.001	1.001	1.001	1.001	1.116
100	10	5	1	1.000	65.918	12.810	12.043	24.083	1.000
	15	8	1	1.001	95.594	15.133	13.804	30.887	1.001
	25	13	1	1.038	38.210	9.954	9.391	16.090	1.038
	50	14	1	0.997	1.441	1.421	1.416	1.428	1.147
	100	14	1	1.497	0.991	0.991	0.991	0.991	1.182
	200	20	1	1.952	1.042	1.041	1.041	1.042	1.163
200	10	5	1	1.000	65.257	13.000	12.548	23.985	1.000
	15	8	1	1.000	93.141	15.228	14.455	30.318	1.000
	25	13	1	1.030	36.010	9.645	9.335	15.357	1.030
	50	20	1	2.034	4.133	3.600	3.577	3.768	1.048
	100	20	1	1.096	1.177	1.177	1.176	1.177	1.106
	200	20	1	1.415	1.058	1.057	1.057	1.057	1.077

Data generated according to (28) with $\rho = -0.8$ and $\theta = 0.7$. Column (A) is Panel Mallows model averaging; (B) is estimated QFR weights; (C) is Granger-Ramanathan; (D) is BIC averaging; (E) is AIC averaging; (F) is simple averaging; (G) is Mallows minimization (i.e., choosing the model with the lowest estimated QFR; not model averaging). Table entries are the ratio of the second order QFL (averaged across 20,000 simulations) to that of the PMMA forecast.

Table 7: Scaled normalized QFLs

n	T	K	(A)	(B)	(C)	(D)	(E)	(F)	(G)
10	10	4	1	1.045	7.318	3.135	2.909	4.309	1.052
	15	5	1	1.031	2.067	1.643	1.595	1.775	1.227
	25	7	1	1.162	1.164	1.114	1.096	1.130	1.389
	50	10	1	1.714	0.922	0.923	0.915	0.922	1.193
	100	14	1	2.621	1.055	1.057	1.051	1.056	1.153
	200	20	1	2.974	1.155	1.157	1.151	1.156	1.158
15	10	5	1	1.007	50.516	6.261	4.777	16.470	1.010
	15	5	1	0.951	1.976	1.610	1.574	1.724	1.224
	25	7	1	1.017	1.148	1.103	1.089	1.117	1.409
	50	10	1	1.516	0.953	0.953	0.949	0.953	1.167
	100	14	1	2.285	1.036	1.038	1.035	1.037	1.164
	200	20	1	2.377	1.083	1.085	1.082	1.084	1.076
25	10	5	1	1.004	46.331	6.828	5.583	15.559	1.009
	15	7	1	1.024	14.924	4.465	4.039	6.975	1.075
	25	7	1	0.973	1.134	1.098	1.089	1.110	1.381
	50	10	1	1.264	0.979	0.979	0.977	0.979	1.141
	100	14	1	1.696	1.009	1.010	1.009	1.010	1.105
	200	20	1	1.694	1.041	1.041	1.040	1.041	1.042
50	10	5	1	1.005	44.972	7.444	6.587	15.423	1.006
	15	8	1	1.024	58.674	7.682	6.377	17.684	1.024
	25	10	1	1.128	3.645	2.604	2.549	2.912	1.502
	50	10	1	1.160	1.010	1.010	1.009	1.010	1.177
	100	14	1	1.720	1.065	1.063	1.063	1.064	1.114
	200	20	1	1.883	1.097	1.094	1.094	1.095	1.100
100	10	5	1	1.002	43.846	7.633	7.106	15.164	1.003
	15	8	1	1.024	57.234	7.959	7.140	17.358	1.024
	25	13	1	1.384	28.091	6.348	5.929	10.893	1.375
	50	14	1	0.957	1.563	1.538	1.533	1.547	1.219
	100	14	1	1.675	1.155	1.148	1.149	1.150	1.119
	200	20	1	2.165	1.247	1.237	1.238	1.241	1.160
200	10	5	1	1.001	42.995	7.687	7.378	14.947	1.001
	15	8	1	1.023	55.629	8.011	7.533	16.974	1.023
	25	13	1	1.459	28.381	6.604	6.357	11.137	1.457
	50	20	1	1.333	4.742	3.987	3.961	4.223	1.133
	100	20	1	1.195	1.325	1.323	1.323	1.324	1.073
	200	20	1	1.454	1.197	1.187	1.187	1.190	1.064

Data generated according to (28) with $\rho = -0.8$ and $\theta = 0.6$. Column (A) is Panel Mallows model averaging; (B) is estimated QFR weights; (C) is Granger-Ramanathan; (D) is BIC averaging; (E) is AIC averaging; (F) is simple averaging; (G) is Mallows minimization (i.e., choosing the model with the lowest estimated QFR; not model averaging). Table entries are the ratio of the second order QFL (averaged across 20,000 simulations) to that of the PMMA forecast.

Table 8: Scaled normalized QFLs

n	T	K	(A)	(B)	(C)	(D)	(E)	(F)	(G)
10	10	4	1	0.999	7.565	2.876	2.637	4.178	1.087
	15	5	1	0.942	2.102	1.569	1.515	1.732	1.326
	25	7	1	1.081	1.200	1.128	1.107	1.151	1.481
	50	10	1	1.829	1.005	1.005	0.996	1.005	1.110
	100	14	1	2.942	1.168	1.171	1.164	1.170	1.146
	200	20	1	3.374	1.294	1.296	1.290	1.296	1.129
15	10	5	1	0.987	50.229	5.280	3.860	15.576	1.020
	15	5	1	0.853	2.033	1.562	1.520	1.706	1.355
	25	7	1	0.991	1.233	1.165	1.148	1.187	1.488
	50	10	1	1.648	1.057	1.058	1.052	1.058	1.099
	100	14	1	2.646	1.162	1.164	1.160	1.164	1.147
	200	20	1	2.634	1.188	1.189	1.186	1.189	1.079
25	10	5	1	0.996	46.052	5.844	4.606	14.634	1.019
	15	7	1	1.012	16.001	4.160	3.694	6.938	1.148
	25	7	1	0.995	1.256	1.196	1.185	1.215	1.435
	50	10	1	1.375	1.077	1.077	1.074	1.077	1.084
	100	14	1	1.888	1.087	1.087	1.086	1.087	1.107
	200	20	1	1.782	1.092	1.091	1.090	1.091	1.028
50	10	5	1	1.010	44.431	6.402	5.535	14.362	1.016
	15	8	1	1.057	59.717	6.690	5.341	16.885	1.062
	25	10	1	1.029	4.571	3.007	2.935	3.459	1.706
	50	10	1	1.268	1.140	1.139	1.137	1.139	1.164
	100	14	1	1.945	1.160	1.157	1.156	1.158	1.138
	200	20	1	2.042	1.180	1.175	1.175	1.177	1.060
100	10	5	1	1.011	43.136	6.569	6.032	14.031	1.011
	15	8	1	1.087	59.554	7.142	6.262	16.895	1.087
	25	13	1	1.709	35.590	6.862	6.322	12.642	1.705
	50	14	1	0.876	1.979	1.932	1.924	1.948	1.134
	100	14	1	1.960	1.281	1.272	1.272	1.275	1.071
	200	20	1	2.427	1.408	1.392	1.392	1.397	1.121
200	10	5	1	1.007	42.163	6.609	6.294	13.770	1.007
	15	8	1	1.113	59.455	7.411	6.878	16.944	1.113
	25	13	1	1.819	36.145	7.189	6.867	12.977	1.815
	50	20	1	1.076	6.544	5.236	5.198	5.638	0.960
	100	20	1	1.385	1.540	1.537	1.536	1.538	0.982
	200	20	1	1.443	1.270	1.256	1.256	1.261	1.065

Data generated according to (28) with $\rho = -0.8$ and $\theta = 0.5$. Column (A) is Panel Mallows model averaging; (B) is estimated QFR weights; (C) is Granger-Ramanathan; (D) is BIC averaging; (E) is AIC averaging; (F) is simple averaging; (G) is Mallows minimization (i.e., choosing the model with the lowest estimated QFR; not model averaging). Table entries are the ratio of the second order QFL (averaged across 20,000 simulations) to that of the PMMA forecast.

6 Empirical Applications

In this section we explore the performance of various model averaging methods in two different empirical applications: (i) autoregressive forecasts of state-level GDP; and (ii) vector autoregressive forecasts of metropolitan statistical area (MSA) employment and population. In all our empirical applications, forecasts are based on VARs augmented with time and cross section effects:

$$y_{i,t} = \beta_i + \delta_t + \textstyle\sum_{s=1}^{k} \alpha_s y_{i,t-s} + u_{i,t}(k), \tag{29}$$

We use an identical psuedo-out-of-sample forecasting design in each of the empirical applications. Models and forecast combination weights are estimated using data observed over $t = 1, \ldots, s$ in order to make forecast(s) of the dependent variables at time $s + 1$. In all applications, we set the index $s = \text{int}\left(\frac{2}{3}T\right)$, $\text{int}\left(\frac{2}{3}T\right) + 1, \ldots, T - 1$, where T denotes the total number of time series observations in the sample, so that forecasts are generated for the final third of observed data. Note that the models and weights are recursively reestimated for each s. In all applications we set the maximum lag order $K = \text{int}\left(\frac{1}{3}T\right)$.

We consider the same seven different forecast averaging designs used in the Monte Carlo Study: (i) PMMA; (ii) estimated QFR weights described in (20); (iii) exponential BIC weights; (iv) exponential AIC weights; (v) constrained Granger-Ramanathan; (vi) simple averaging (i.e., $w_k = 1/K$); and (vii) the Mallows minimization criterion. In order to compare the relative accuracy of each method over time, we normalize the QFL of each forecast by the (ex post) minimum QFL among the VAR(k) models for that year. Thus, if the model with $k = 3$ lags has the smallest QFL for a given year, the QFLs of the weighted forecasts are expressed as a ratio of the QFL with $k = 3$ for that year.

We first present the forecasting performance of each of the averaging methods before making some observations of the relative performance of each method across the different applications.

State GDP growth. State level real GDP are obtained from the Bureau of Economic Analysis. The BEA reports state level GDP a Standard Industrial Classification (SIC) basis for 1987 to 1997, and on a North American Industrial Classification System (NAICS) from 1997 onwards (chained 2009 dollars). We use the growth rates in real GDP reported on the SIC basis to backcast each of the NAICS GDP time series to 1987.

Our sample includes all fifty states and the District of Columbia, and spans the years 1987 to 2017, inclusive. We consider simple autoregressive forecasts of GDP growth. The set of AR(k) models is defined as in (29) above, where $y_{i,t} := \ln(Y_{i,t}) - \ln(Y_{i,t-1})$, and $Y_{i,t}$ is real GDP of state i in year t. Our experimental design results in forecasts produced for each year between 2008 and 2017, inclusive. Table 9 below exhibits the scaled QFL of each forecast averaging method.

Table 9: Scaled QFL of weighting designs: MSA Employment Growth

	(A)	(B)	(C)	(D)	(E)	(F)	(G)
2008	1.085	1.143	1.109	1.146	1.147	1.146	1.103
2009	1.058	1.076	1.140	1.076	1.077	1.077	1.049
2010	1.018	1.036	1.038	1.036	1.036	1.036	1.031
2011	1.116	1.130	1.123	1.130	1.131	1.131	1.091
2012	1.046	1.051	1.125	1.050	1.050	1.050	1.044
2013	1.053	1.081	1.019	1.081	1.082	1.081	1.010
2014	1.030	1.030	1.106	1.030	1.029	1.030	1.031
2015	1.165	1.133	1.245	1.133	1.132	1.132	1.198
2016	1.073	1.086	1.124	1.085	1.086	1.086	1.062
2017	1.082	1.007	1.129	1.007	1.007	1.007	1.261
average	1.073	1.077	1.116	1.077	1.078	1.077	1.088

Normalized QFLs for various forecasting averaging methods. Column (A) is PMMA; (B) is estimated QFR weights; (C) is Granger-Ramanathan; (D) is BIC averaging; (E) is AIC averaging; (F) is simple averaging; (G) is Mallows minimization (i.e., choosing the model with the lowest estimated QFR; not model averaging). In each year the QFLs are scaled by the smallest QFL among the VAR(k) models.

Metropolitan Statistical Area Employment and Population. In this application we forecast Metropolitan Statistical Area (MSA) employment growth and population growth using bivariate VARs described in (29), where $y_{i,t} := (\ln(E_{i,t}) - \ln(E_{i,t-1}), \ln(P_{i,t}) - \ln(P_{i,t-1}))'$. This model specification is motivated in part by the regional adjustment literature, in which region-specific labor supply and demand shocks are equilibrated over the long-run through inter-regional migration (see, among others, Blanchard and Katz, 1992). Regional labor demand shocks are highly persistent (Greenaway-McGrevy and Hood, 2016; Amior and Manning, 2018), meaning that local booms and bust can generate persistent changes in employment and population through the migration channel. Because these regional adjustment models are used to describe relative changes in population and other labor market variables, we include period fixed effects in the estimated VARs.

We obtain employment and population data for 383 MSAs for 1969 to 2016 from Table CA04 of the Bureau of Economic Analysis.[6] Our experimental design results in pseudo-out-of-sample forecasts for each year between 2001 to 2016, inclusive. Reported QFLs are based on an equally-weighted average of the two forecasted variables in the bivariate VAR. Table 10 below exhibits the scaled QFLs of each method.

Remarks.

(*i*) In both applications, the PMMA forecast combination method (column (A)) exhibits the lowest QFL, on average. This result is consistent with the underlying motivation of the method, which is to select the weights to minimize quadratic forecast loss. This finding is of particular interest because weighting schemes that are designed to penalize the correlation in forecast errors between different models tend to perform poorly in the time series context. It is possible that the correlations between the forecast errors of the different models are more accurately estimated by pooling between cross sections, giving rise to superior forecast performance.

(*ii*) The PMMA does not, however, exhibit the lowest QFL in each year. For the state GDP forecasts, it has the smallest QFL among averaging methods in 8 of the 10 years. For the MSA forecasts of population and employment, PMMA has the lowest QFL among forecast averaging methods for 8 of the 15 years. Interestingly, in this latter experiment, the superior performance of PMMA occurs at the end of the sample (2009 to 2016). This may indicate that the PMMA method requires large samples in order to accurately estimate optimal weights for large dimensional models.

(*iii*) The estimated QFR weighting scheme (column (B)), the BIC and AIC averages (columns (D) and (E)), and the simple average (column (F)) exhibit very similar QFLs – both on average

[6] Regional adjustment models usually include regional unemployment rates. We do not include this variable because county (and thus MSA) unemployment is only available from 1990 onwards, and has been subject to methodological revisions that make it difficut to use in a forecasting exercise (see Greenaway-McGrevy and Hood, 2016, for a discussion of the change in method).

Table 10: Scaled QFL of weighting designs: MSA Employment Growth

	(A)	(B)	(C)	(D)	(E)	(F)	(G)
2002	1.010	1.004	1.004	1.008	1.008	1.007	1.026
2003	1.194	1.192	1.450	1.193	1.193	1.192	1.459
2004	1.080	1.074	1.234	1.073	1.073	1.073	1.226
2005	1.127	1.119	1.264	1.117	1.117	1.118	1.259
2006	1.134	1.137	1.365	1.131	1.131	1.133	1.261
2007	1.135	1.097	1.014	1.102	1.102	1.100	1.087
2008	1.141	1.144	1.477	1.135	1.135	1.138	1.245
2009	1.151	1.197	1.769	1.180	1.180	1.185	1.279
2010	1.013	1.059	1.387	1.049	1.049	1.052	1.046
2011	1.055	1.088	1.371	1.076	1.076	1.080	1.034
2012	1.111	1.140	1.283	1.131	1.131	1.134	1.107
2013	1.057	1.104	1.374	1.091	1.091	1.094	1.053
2014	1.045	1.106	1.337	1.093	1.093	1.097	1.052
2015	0.994	1.036	1.232	1.026	1.025	1.028	1.067
2016	1.089	1.132	1.435	1.116	1.115	1.120	1.144
average	1.089	1.109	1.333	1.101	1.101	1.103	1.156

Normalized QFLs for various forecasting averaging methods. Column (A) is PMMA; (B) is estimated QFR weights; (C) is Granger-Ramanathan; (D) is BIC averaging; (E) is AIC averaging; (F) is simple averaging; (G) is Mallows minimization (i.e., choosing the model with the lowest estimated QFR; not model averaging). In each year the QFLs are scaled by the smallest QFL among the VAR(k) models.

and in each year. This is due to the fact that the difference in the magnitude of the QFR estimator for different lags k is often small, leading to relatively uniform weights. The same is true of the BIC and AIC.

(*iv*) In time series applications it is quite common for simple averages (column (F)) tend to outperform more sophisticated model averaging methods (see Winkler and Makridakis, 1983; Clemen, 1989; Stock and Watson, 1999; and Fildes and Ord, 2002). This result does not generalize to these panel data examples, wherein the simple average often exhibits a larger QFL, on average, than the panel MMA forecast averaging method. However, we caution that further empirical work in panel data settings is required before establishing any such stylized fact.

(*v*) Selection of a single forecasting model via minimization of the Mallows criterion (column (G)) performs relatively poorly when evaluated by the average QFL across all time periods. This is consistent with empirical findings in the time series forecasting literature, in which a simple averaging scheme often outperforms selection of a single model using an informed selection criterion. However, looking at the individual years, it often outperforms the averaging methods. For the state GDP forecasts, it has a smaller QFL than all of the averaging methods in 5 of the 10 years. However, for the MSA forecasts of population and employment, PMMA has the lowest QFL among forecast averaging methods for only 3 of the 15 years.

7 Concluding Remarks

In this paper we design new forecast averaging methods for panel data. In particular, we generalize the Mallows Model Averaging method developed by Hansen (2007, 2008) to a panel data setting, which is based on the principle of selecting weights to minimize the estimated QFR of the average model. Consistent with the underlying principle of panel MMA, the method is shown to outperform other common weighting mechanisms adapted to the panel setting in both simulated and empirical out of sample forecasting experiments. We show how PMMA can be simplified under certain assumptions on the data and can accommodate commonly-used panel data period model specifications, such as fixed effects and incidental (or heterogenous) linear trends.

8 Appendix

8.1 Additional Notation and Definitions

8.1.1 Small T Environments

We define the $n(T-k) \times mk$ matrix $\underline{\mathbf{Y}}(k)$ as

$$\underline{\mathbf{Y}}(k) := \left[\underline{\mathbf{Y}}'_{1\cdot}(k) : \cdots : \underline{\mathbf{Y}}'_{i\cdot}(k) : \cdots : \underline{\mathbf{Y}}'_{n\cdot}(k)\right]', \ \underline{\mathbf{Y}}_{i\cdot}(k) := \left[Y_{i,k}(k) : \cdots : Y_{i,t}(k) : \cdots : Y_{i,T-1}(k)\right]',$$

so that the underscore "$\underline{\cdot}$" indicates that the dimension spans $T-k$ time series observations rather than $T-K$. $\underline{\mathbf{X}}(k)$ is similarly defined. We also define the $n(T-k) \times m$ matrix $\underline{\mathbf{y}}(k)$ as

$$\underline{\mathbf{y}}(k) := \left[\underline{\mathbf{y}}_{1\cdot} : \cdots : \underline{\mathbf{y}}_{i\cdot} : \cdots : \underline{\mathbf{y}}_{n\cdot}\right]', \ \ \underline{\mathbf{y}}_{i\cdot} := \left[y_{i,k+1} : \cdots : y_{i,t} : \cdots : y_{i,T}\right]'.$$

$\underline{\mathbf{u}}(k)$ and $\underline{\mathbf{e}}(k)$ are similarly defined. We also define $\mathbf{M}_{\boldsymbol{\tau}}(k) := I_{nT_k} - \boldsymbol{\tau}_{T-k}\left(\boldsymbol{\tau}'_{T-k}\boldsymbol{\tau}_{T-k}\right)^{-1}\boldsymbol{\tau}'_{T-k}$, $\mathbf{M}_{\mathbf{M}_{\boldsymbol{\tau}}\underline{\mathbf{Y}}(k)} := I_{nT_K} - \mathbf{P}_{\mathbf{M}_{\boldsymbol{\tau}}\underline{\mathbf{Y}}(k)}$, and $\mathbf{P}_{\mathbf{M}_{\boldsymbol{\tau}}\underline{\mathbf{Y}}(k)} := \mathbf{M}_{\boldsymbol{\tau}}\underline{\mathbf{Y}}(k)\left(\underline{\mathbf{Y}}(k)'\mathbf{M}_{\boldsymbol{\tau}}(k)\underline{\mathbf{Y}}(k)\right)^{-1}\underline{\mathbf{Y}}(k)'\mathbf{M}_{\boldsymbol{\tau}}(k)$. $\check{\boldsymbol{\alpha}}(k)$ and $\check{\boldsymbol{\beta}}(k)$ are defined as the the bias-corrected LS estimates of $\boldsymbol{\alpha}(k)$ and $\boldsymbol{\beta}(k)$ based on samples spanning $t = k, \ldots, T$ as follows:

$$\check{\boldsymbol{\alpha}}(k) := \left(\underline{\mathbf{Y}}(k)'\mathbf{M}_{\boldsymbol{\tau}}(k)\underline{\mathbf{Y}}(k)\right)^{-1}\left(\underline{\mathbf{Y}}(k)'\mathbf{M}_{\boldsymbol{\tau}}(k)\underline{\mathbf{y}}(k) + n\check{\boldsymbol{\xi}}(k)\right) \tag{30}$$

and

$$\check{\boldsymbol{\beta}}(k) := \left(\boldsymbol{\tau}'_{T-k}\boldsymbol{\tau}_{T-k}\right)^{-1}\boldsymbol{\tau}'_{T-k}\mathbf{M}_{\underline{\mathbf{Y}}(k)}\underline{\mathbf{y}}(k) + \left(\boldsymbol{\tau}'_{T-k}\boldsymbol{\tau}_{T-k}\right)^{-1}\boldsymbol{\tau}'_{T-k}\mathbf{M}_{\boldsymbol{\tau}}\underline{\mathbf{Y}}(k)\left(\underline{\mathbf{Y}}(k)'\mathbf{M}_{\boldsymbol{\tau}}(k)\underline{\mathbf{Y}}(k)\right)^{-1}n\check{\boldsymbol{\xi}}(k), \tag{31}$$

where

$$\begin{aligned}\check{\boldsymbol{\xi}}(k) \ :&= \left[\mathbf{1}_{mk} \otimes \left(I_m - \underline{\mathbf{y}}(k)'\mathbf{M}_{\boldsymbol{\tau}}(k)\underline{\mathbf{Y}}(k)\left(\underline{\mathbf{Y}}(k)'\mathbf{M}_{\boldsymbol{\tau}}(k)\underline{\mathbf{Y}}(k)\right)^{-1}(\mathbf{1}_k \otimes I_m)\right)^{-1}\right] \times \\ &\frac{1}{nT_k}\underline{\mathbf{y}}(k)'\left[\mathbf{M}_{\boldsymbol{\tau}}(k) - \mathbf{M}_{\underline{\mathbf{Y}}(k)}\right]\underline{\mathbf{y}}(k).\end{aligned}$$

The first terms on the right hand side of (30) and (31) are the OLS estimates, while the second terms denote the bias corrections. The $n \times m$ forecast matrix is then

$$\check{\mathbf{y}}_{\cdot T+1}(k) := \mathbf{y}_{\cdot T}(k)\check{\boldsymbol{\alpha}}(k) + \tau_{T-k}\check{\boldsymbol{\beta}}(k)$$

with associated forecast error $\check{\mathbf{u}}_{\cdot T+1}(k) := \mathbf{y}_{\cdot T+1} - \check{\mathbf{y}}_{\cdot T+1}(k)$. Then the averaged forecast is $\check{\mathbf{y}}_{\cdot T+1}(\mathbf{w}) := \sum_{s=1}^{K} w_s \check{\mathbf{y}}_{\cdot T+1}(s)$ and averaged forecast error is $\check{\mathbf{u}}_{\cdot T+1}(\mathbf{w}) := \sum_{s=1}^{K} w_s \check{\mathbf{u}}_{\cdot T+1}(s)$. The associated QFL is

$$\underline{\mathcal{L}}_{n,T}(\mathbf{w}) := \frac{1}{n}\left(\mathbf{y}_{\cdot T+1} - \check{\mathbf{y}}_{\cdot T+1}(\mathbf{w})\right)'\left(\mathbf{y}_{\cdot T+1} - \check{\mathbf{y}}_{\cdot T+1}(\mathbf{w})\right) = \frac{1}{n}\check{\mathbf{u}}_{\cdot T+1}(\mathbf{w})'\check{\mathbf{u}}_{\cdot T+1}(\mathbf{w}) \tag{32}$$

8.1.2 Period Fixed Effects

Let

$$\ddot{\mathbf{Q}}(k) := \frac{1}{nT_K}\ddot{\mathbf{Y}}(k)'\mathbf{M}_{\boldsymbol{\tau}}\ddot{\mathbf{Y}}(k), \quad \ddot{\mathbf{v}}(k) := \frac{1}{nT_K}\ddot{\mathbf{Y}}(k)'\mathbf{M}_{\boldsymbol{\tau}}\ddot{\mathbf{y}},$$

and

$$\mathbf{P}_{\mathbf{M}_{\boldsymbol{\tau}}\ddot{\mathbf{Y}}(k)} := \mathbf{M}_{\boldsymbol{\tau}}\ddot{\mathbf{Y}}(k)\left(\ddot{\mathbf{Y}}(k)'\mathbf{M}_{\boldsymbol{\tau}}\ddot{\mathbf{Y}}(k)\right)^{-1}\ddot{\mathbf{Y}}(k)'\mathbf{M}_{\boldsymbol{\tau}}.$$

Then the bias-corrected least squares estimator of $\boldsymbol{\alpha}(k)$ is

$$\hat{\hat{\boldsymbol{\alpha}}}(k) := \ddot{\mathbf{Q}}^{-1}(k)\ddot{\mathbf{v}}(k)+\frac{p}{T_K}\ddot{\mathbf{Q}}^{-1}(k)\left[\mathbf{1}_k\otimes\left(I_m-\ddot{\mathbf{v}}(k)'\ddot{\mathbf{Q}}^{-1}(k)(\mathbf{1}_k\otimes I_m)\right)^{-1}\right]\frac{1}{nT_K}\ddot{\mathbf{y}}'\left(\mathbf{M}_{\boldsymbol{\tau}}-\mathbf{P}_{\mathbf{M}_{\boldsymbol{\tau}}\ddot{\mathbf{Y}}(k)}\right)\ddot{\mathbf{y}}, \tag{33}$$

while the bias-corrected LS estimator of $\boldsymbol{\beta}$ is

$$\hat{\hat{\boldsymbol{\beta}}}(k) := \left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}'\left(\ddot{\mathbf{y}}-\ddot{\mathbf{Y}}(k)\hat{\hat{\boldsymbol{\alpha}}}(k)\right). \tag{34}$$

8.2 Proofs

Throughout the proofs we focus on the scalar case ($m=1$) without loss of generality. This simplifies notation considerably.

8.2.1 Proof of Theorem 3.1

We make extensive use of the following lemma, the proof of which is Greenaway-McGrevy (2018a).

Lemma 8.1 *Under Assumptions A and E, for all s satisfying $1\le s\le K$,*

(i) $\mathrm{E}\left\|\frac{1}{nT_K}\mathbf{X}(s)'\mathbf{e}\right\|^q \le C_q\left(\left(\frac{s}{nT_K}\right)^{\frac{q}{2}}\right)$

(ii) $\mathrm{E}\left\|\frac{1}{nT_K}\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{e}-\frac{p}{T-K}\boldsymbol{\xi}(s)\right\|^q \le C_q\left(\frac{s}{nT_K^2}\right)^{\frac{q}{2}}$

(iii) $\mathrm{E}\left\|\frac{1}{nT_K}\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{e}\right\|^q \le C_q\left(\left(\frac{s}{nT_K}\right)^{\frac{q}{2}}+\left(\frac{s}{T_K^2}\right)^{\frac{q}{2}}\right)$

(iv) $\mathrm{E}\left\|\frac{1}{nT_K}\mathbf{e}'\mathbf{e}-\Sigma\right\|^q \le C_q n^{-\frac{q}{2}}T_K^{-\frac{q}{2}}$

(v) $\mathrm{E}\left\|\frac{1}{n}\mathbf{e}'\mathbf{P}_{\boldsymbol{\tau}}\mathbf{e}-p\Sigma\right\|^q \le C_q n^{-\frac{q}{2}}T_K^{-\frac{q}{2}}$

(vi) $\mathrm{E}\left\|\frac{1}{nT_K}\mathbf{s}(s)'\mathbf{s}(s)-\Lambda(s)\right\|^q \le C_q n^{-\frac{q}{2}}T_K^{-\frac{q}{2}}\left\|\Lambda(s)\right\|^q$

(vii) $\mathrm{E}\left\|\frac{1}{nT_K}\mathbf{s}(s)'\mathbf{P}_{\boldsymbol{\tau}}\mathbf{s}(s)\right\|^q \le C_q T_K^{-q}\left\|\Lambda(s)\right\|^q$

(viii) $\mathrm{E}\left\|\frac{1}{nT_K}\mathbf{X}(s)'\mathbf{s}(s)\right\|^q \le C_q\left(\left(\frac{s}{nT_K}\right)^{\frac{q}{2}}\left\|\Lambda(s)\right\|^{\frac{q}{2}}\right)$

(ix) $\mathrm{E}\left\|\frac{1}{nT_K}\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{s}(s)\right\|^q \le C_q\left(\left(\frac{s}{nT_K}\right)^{\frac{q}{2}}+\left(\frac{s}{T_K^2}\right)^{\frac{q}{2}}\right)\|\Lambda(s)\|^{\frac{q}{2}}$

(x) $\mathrm{E}\left\|\frac{1}{nT_K}\mathbf{e}'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{s}(s)\right\|_{\Phi}^q \le C_q\left((nT_K)^{-\frac{q}{2}}+T_K^{-q}\right)\|\Lambda(s)\|^{\frac{q}{2}}$

(xi) $\mathrm{E}\left\|\frac{1}{nT_K}\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)-\mathbf{\Gamma}(s)\right\|^q \le C_q\left(\left(\frac{s}{nT_K}\right)^{\frac{q}{2}}+\left(\frac{s}{T_K^2}\right)^{\frac{q}{2}}\right)$

(xii) $\mathrm{E}\left\|\frac{1}{nT_K}\mathbf{X}(s)'\mathbf{P}_{\boldsymbol{\tau}}\mathbf{X}(s)\right\|^q \le C_q\left(\left(\frac{s}{nT_K^3}\right)^{\frac{q}{2}}+\left(\frac{s}{T_K^2}\right)^{\frac{q}{2}}\right)$

$(xiii)$ $\mathrm{E}\left\|\left(\frac{1}{nT_K}\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)\right)^{-1}\right\|^q = O(1)$

(xiv) $\mathrm{E}\left\|\frac{p}{T_K}\hat{\boldsymbol{\xi}}(s)-\frac{p}{T_K}\boldsymbol{\xi}(s)\right\|^q \le C_q\left(\left(\frac{s}{nT_K^3}\right)^{\frac{q}{2}}+\left(\frac{s}{T_K^4}\right)^{\frac{q}{2}}\right)\left(1+\|\Lambda(s)\|^{\frac{q}{2}}\right)$

Let $\mathbf{P}(\mathbf{w}) := \sum_{s=1}^K w_s\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)\left(\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)\right)^{-1}\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}$ and $\zeta(\mathbf{w}) := \sum_{s=1}^K w_s\boldsymbol{\xi}(s)'\mathbf{\Gamma}^{-1}(s)\boldsymbol{\xi}(s)$. We can then express

$$\begin{aligned}\mathcal{R}_{n,T}(\mathbf{w}) &= \mathcal{L}_{n,T}(\mathbf{w})-\frac{2}{nT_K}\mathbf{e}'\mathbf{P}(\mathbf{w})\mathbf{e}+\frac{2p^2}{T_K^2}\zeta(\mathbf{w})-\frac{p^2+p}{T_K}\Sigma+\\ &\quad\underbrace{\mathcal{R}_{n,T}(\mathbf{w})-\mathcal{L}_{n,T}(\mathbf{w})+\frac{2}{nT_K}\mathbf{e}'\mathbf{P}(\mathbf{w})\mathbf{e}-\frac{2p^2}{T_K^2}\zeta(\mathbf{w})+\frac{p^2+p}{T_K}\Sigma}_{=:b_{n,T}(\mathbf{w})}\end{aligned} \tag{35}$$

Meanwhile from (16) we have $\mathcal{R}_{n,T}(\mathbf{w}) = \hat{L}_{n,T}(\mathbf{w})-\left(\frac{2}{nT_K}\mathbf{k}'\mathbf{w}+\frac{p^2+p}{T_K}\right)\hat{\Sigma}(K)$. Substituting this into the expression above and rearranging yields

$$\hat{L}_{n,T}(\mathbf{w}) = \mathcal{L}_{n,T}(\mathbf{w})+\frac{2}{nT_K}\left[\mathbf{k}'\mathbf{w}\hat{\Sigma}(K)-\mathbf{e}'\mathbf{P}(\mathbf{w})\mathbf{e}+\frac{p^2n}{T_K}\zeta(\mathbf{w})\right]+\frac{p^2+p}{T_K}\left(\hat{\Sigma}(K)-\Sigma\right)+b_{n,T}(\mathbf{w}) \tag{36}$$

Now by the Hölder's and triangle inequalities, lemma (xiv), and lemmas 8.1 (i), (ii), (ix), (xi), (vii), and $(xiii)$, it follows that

$$\mathrm{E}\left(\hat{\Sigma}(K)\right) = \Sigma+\Lambda(K)+o\left(T_K^{-1}\right). \tag{37}$$

Next, by the Hölder's and triangle inequalities, lemmas 8.1 (i), (ii), (xi), and $(xiii)$,

$$\frac{1}{nT_K}\mathrm{E}\left(\mathbf{e}'\mathbf{P}(\mathbf{w})\mathbf{e}\right) = \sum_{s=1}^K w_s\frac{2s}{nT_K}\Sigma+\frac{p^2}{T_K^2}\zeta(\mathbf{w})+o\left(\frac{1}{T_K^2}\right). \tag{38}$$

Meanwhile lemma 8.2 below shows that $\|\mathrm{E}(b_{n,T}(\mathbf{w}))\| = o\left(K(nT_K)^{-1}\right)+o(\|\Lambda(\mathbf{w})\|)+O\left(T_K^{-2}\right)$. Thus, noting that $\frac{2}{nT_K}\mathbf{k}'\mathbf{w} = \sum_{s=1}^K w_s\frac{2s}{nT_K}$, taking expectations of (36), we have

$$\mathrm{E}\left(\hat{L}_{n,T}(\mathbf{w})\right) = \mathrm{E}(\mathcal{L}_{n,T}(\mathbf{w}))+\left(\sum_{s=1}^K w_s\frac{2s}{nT_K}+\frac{p^2+p}{T_K}\right)\left\{\Lambda(K)+o\left(T_K^{-1}\right)\right\}+o\left(\frac{K}{nT_K}\right)+o(\|\Lambda(\mathbf{w})\|)+O\left(T_K^{-2}\right).$$

The stated result then follows, noting that $\left[K\left(nT_K\right)^{-1}+\|\Lambda\left(\mathbf{w}\right)\|\right]^{-1}T_K^{-2}\leq K^{-1}nT_K^{-1}\rightarrow 0$ as $K\rightarrow\infty$ and $n=O\left(T_K\right)$, and

$$\left[K\left(nT_K\right)^{-1}\Sigma+\Lambda\left(\mathbf{w}\right)\right]^{-1}\Lambda\left(K\right)\leq\left[K\left(nT_K\right)^{-1}\Sigma+\Lambda\left(\mathbf{w}\right)\right]^{-1}\Lambda\left(\mathbf{w}\right)\leq I_m$$

in the matrix sense.

Lemma 8.2 *$b_{n,T}\left(\mathbf{w}\right)$ defined in (35) satisfies* $\|\mathrm{E}\left(b_{n,T}\left(\mathbf{w}\right)\right)\|=o\left(K\left(nT_K\right)^{-1}\right)+o\left(\|\Lambda\left(\mathbf{w}\right)\|\right)+O\left(T_K^{-2}\right).$

Proof. We begin by decomposing

$$\begin{aligned}\hat{\mathbf{u}}\left(\mathbf{w}\right) &= \textstyle\sum_{s=1}^{K}w_s\left\{\mathbf{u}\left(s\right)-\mathbf{X}\left(s\right)\left(\hat{\boldsymbol{\alpha}}\left(s\right)-\boldsymbol{\alpha}\left(s\right)\right)\right\}\\ &= \textstyle\sum_{s=1}^{K}w_s\left\{\mathbf{u}\left(s\right)-\mathbf{X}\left(s\right)\left[\mathbf{X}\left(s\right)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}\left(s\right)\right]^{-1}\left(\mathbf{X}\left(s\right)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}\left(s\right)+pn\hat{\boldsymbol{\xi}}\left(s\right)\right)\right\},\end{aligned}$$

such that the in-sample quadratic loss $\mathcal{R}_{n,T}\left(\mathbf{w}\right)=\frac{1}{nT_K}\hat{\mathbf{u}}\left(\mathbf{w}\right)'\hat{\mathbf{u}}\left(\mathbf{w}\right)$ can be decomposed into four terms as follows:

$$\begin{aligned}\mathcal{R}_{n,T}\left(\mathbf{w}\right) &= \sum_{s,r=1}^{K}w_sw_r\left\{\tfrac{1}{nT_K}\mathbf{u}\left(r\right)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}\left(s\right)\right\}- \qquad (39)\\ &\quad \sum_{s,r=1}^{K}w_sw_r\left\{\tfrac{1}{nT_K}\mathbf{u}\left(r\right)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}\left(s\right)\left(\mathbf{X}\left(s\right)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}\left(s\right)\right)^{-1}\left[\mathbf{X}\left(s\right)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}\left(s\right)+pn\boldsymbol{\xi}\left(s\right)\right]\right\}-\\ &\quad \sum_{s,r=1}^{K}w_sw_r\left\{\tfrac{1}{nT_K}\left[\mathbf{X}\left(s\right)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}\left(s\right)+pn\boldsymbol{\xi}\left(s\right)\right]'\left(\mathbf{X}\left(s\right)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}\left(s\right)\right)^{-1}\mathbf{X}\left(s\right)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}\left(r\right)\right\}+\\ &\quad \sum_{s,r=1}^{K}w_sw_r\left\{\tfrac{1}{nT_K}\left[\mathbf{X}\left(s\right)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}\left(s\right)+pn\boldsymbol{\xi}\left(s\right)\right]'\left(\mathbf{X}\left(s\right)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}\left(s\right)\right)^{-1}\times\right.\\ &\qquad\left.\mathbf{X}\left(s\right)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}\left(r\right)\left(\mathbf{X}\left(r\right)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}\left(r\right)\right)^{-1}\left[\mathbf{X}\left(r\right)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}\left(r\right)+pn\boldsymbol{\xi}\left(r\right)\right]\right\}\end{aligned}$$

We similarly decompose the forecast error

$$\begin{aligned}\hat{\mathbf{u}}_{\cdot T+1}\left(\mathbf{w}\right) &= \textstyle\sum_{s=1}^{K}w_s\left\{\ddot{\mathbf{u}}_{\cdot T+1}\left(s\right)-\ddot{\mathbf{x}}_{\cdot T}\left(s\right)\left(\hat{\boldsymbol{\alpha}}\left(s\right)-\boldsymbol{\alpha}\left(s\right)\right)\right\} \qquad (40)\\ &= \textstyle\sum_{s=1}^{K}w_s\left\{\ddot{\mathbf{u}}_{\cdot T+1}\left(s\right)-\ddot{\mathbf{x}}_{\cdot T}\left(s\right)\left[\mathbf{X}\left(s\right)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}\left(s\right)\right]^{-1}\left(\mathbf{X}\left(s\right)\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}\left(s\right)+pn\hat{\boldsymbol{\xi}}\left(s\right)\right)\right\},\end{aligned}$$

so that out-of-sample quadratic loss $\mathcal{L}_{n,T}(\mathbf{w}) = \frac{1}{n}\hat{\mathbf{u}}_{\cdot T+1}(\mathbf{w})' \hat{\mathbf{u}}_{\cdot T+1}(\mathbf{w})$ can be decomposed into four terms as follows:

$$\begin{aligned}
\mathcal{L}_{n,T}(\mathbf{w}) \;=\; & \sum_{s,r=1}^{K} w_s w_r \left\{\tfrac{1}{n}\ddot{\mathbf{u}}_{\cdot T+1}(r)' \ddot{\mathbf{u}}_{\cdot T+1}(s)\right\} - \\
& \sum_{s,r=1}^{K} w_s w_r \left\{\tfrac{1}{n}\ddot{\mathbf{u}}_{\cdot T+1}(r)' \ddot{\mathbf{x}}_{\cdot T}(s)\left(\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)\right)^{-1}\left[\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}(s) + pn\boldsymbol{\xi}(s)\right]\right\} - \\
& \sum_{s,r=1}^{K} w_s w_r \left\{\left[\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}(s) + pn\boldsymbol{\xi}(s)\right]'\left(\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)\right)^{-1}\tfrac{1}{n}\ddot{\mathbf{x}}_{\cdot T}(s)' \ddot{\mathbf{u}}_{\cdot T+1}(r)\right\} + \\
& \sum_{s,r=1}^{K} w_s w_r \left\{\tfrac{1}{nT_K}\left[\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}(s) + pn\boldsymbol{\xi}(s)\right]'\left(\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)\right)^{-1} \times \right. \\
& \qquad \left.\left(\tfrac{1}{n}\ddot{\mathbf{x}}_{\cdot T}(s)'\ddot{\mathbf{x}}_{\cdot T}(r)\right)\left(\mathbf{X}(r)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(r)\right)^{-1}\left[\mathbf{X}(r)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}(r) + pn\boldsymbol{\xi}(r)\right]'\right\}
\end{aligned} \tag{41}$$

Then using the decompositions for $\mathcal{R}_{n,T}(\mathbf{w})$ and $\mathcal{L}_{n,T}(\mathbf{w})$ given in (39) and (41), we have

$$\begin{aligned}
& b_{n,T}(\mathbf{w}) = \mathcal{R}_{n,T}(\mathbf{w}) - \mathcal{L}_{n,T}(\mathbf{w}) + \tfrac{2}{nT_K}\mathbf{e}'\mathbf{P}(\mathbf{w})\mathbf{e} - \tfrac{2}{T_K^2}\zeta(\mathbf{w}) + \tfrac{p^2+p}{T_K}\Sigma \\
= & \sum_{s,r=1}^{K} w_s w_r \left\{\left[\tfrac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}(s)}{nT_K} + \tfrac{p}{T_K}\hat{\boldsymbol{\xi}}(s)\right]'\left(\tfrac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)}{nT_K}\right)^{-1} \times \right. \\
& \qquad \left.\left[\tfrac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(r)}{nT_K} - \tfrac{\ddot{\mathbf{x}}_{\cdot T}(s)'\ddot{\mathbf{x}}_{\cdot T}(r)}{n}\right]\left(\tfrac{\mathbf{X}(r)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(r)}{nT_K}\right)^{-1}\left[\tfrac{\mathbf{X}(r)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}(r)}{nT_K} + \tfrac{p}{T_K}\hat{\boldsymbol{\xi}}(r)\right]\right\} + \\
& \sum_{s,r=1}^{K} w_s w_r \left\{\left[\tfrac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}(s)}{nT_K} + \tfrac{p}{T_K}\hat{\boldsymbol{\xi}}(s)\right]'\left(\tfrac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)}{nT_K}\right)^{-1}\left[\tfrac{\ddot{\mathbf{x}}_{\cdot T}(s)'\ddot{\mathbf{u}}_{\cdot T+1}(r)}{n} - \tfrac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}(r)}{nT_K}\right]\right\} + \\
& \qquad \left[\tfrac{\mathbf{e}'\mathbf{P}(\mathbf{w})\mathbf{e}}{T_K} - \tfrac{p^2}{T_K}\zeta(\mathbf{w})\right] + \\
& \sum_{s,r=1}^{K} w_s w_r \left\{\left[\tfrac{\ddot{\mathbf{x}}_{\cdot T}(s)'\ddot{\mathbf{u}}_{\cdot T+1}(r)}{n} - \tfrac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}(r)}{nT_K}\right]'\left(\tfrac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)}{nT_K}\right)^{-1}\left[\tfrac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}(s)}{nT_K} + \tfrac{p}{T_K}\hat{\boldsymbol{\xi}}(s)\right]\right\} + \\
& \qquad \left[\tfrac{\mathbf{e}'\mathbf{P}(\mathbf{w})\mathbf{e}}{T_K} - \tfrac{p^2}{T_K}\zeta(\mathbf{w})\right] + \\
& \sum_{s,r=1}^{K} w_s w_r \left\{\tfrac{1}{nT_K}\mathbf{u}'(r)\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}(s) - \tfrac{1}{n}\ddot{\mathbf{u}}'_{\cdot T+1}(r)\ddot{\mathbf{u}}_{\cdot T+1}(s)\right\} + \tfrac{p^2+p}{T_K}\Sigma \\
= & \; I + II + (II)' + III
\end{aligned}$$

We provide a bound for the expectation of these three terms. Beginning with I, by lemmas 8.1 (i), (ii), (ix), (xi), and $(xiii)$, we have

$$\mathrm{E}\left\|\tfrac{1}{nT_K}\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}(s) + \tfrac{p}{T_K}\hat{\boldsymbol{\xi}}(s)\right\|^q \leq C_q K^{\frac{q}{2}}\left((nT_K)^{-\frac{q}{2}} + T_K^{-2q}\right), \tag{42}$$

$$\mathrm{E}\left\|\left(\tfrac{1}{nT_K}\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)\right)^{-1}\right\|^q = O(1), \tag{43}$$

$$\mathrm{E}\left\|\tfrac{1}{nT_K}\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s) - \boldsymbol{\Gamma}(s)\right\|^q \leq C_q K^{\frac{q}{2}}\left((nT_K)^{-\frac{q}{2}} + T_K^{-q}\right), \tag{44}$$

for each $s = 1, \ldots, K$, and

$$\mathrm{E}\left\| \frac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(r)}{nT_K} - \frac{\ddot{\mathbf{x}}_{\cdot T}(s)'\ddot{\mathbf{x}}_{\cdot T}(r)}{n} \right\|^{2q} \leq C_q \left(\frac{K}{\sqrt{n}}\right)^q,$$

for all $q = 1, 2, \ldots$, we can establish that $\mathrm{E}(I) \leq C\left((nT_K)^{-1} + T_K^{-2}\right)\frac{K}{\sqrt{n}}$ via Hölder's inequality. Next we decompose II as follows.

$$\begin{aligned}
II &= \sum_{s,r=1}^{K} w_s w_r \left\{ \left[\frac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}(s)}{nT_K} + \frac{p}{T_K}\hat{\boldsymbol{\xi}}(s)\right]' \left(\frac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)}{nT_K}\right)^{-1} \left[\frac{\ddot{\mathbf{x}}_{\cdot T}(s)'\ddot{\mathbf{u}}_{\cdot T+1}(r)}{n} - \frac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}(r)}{nT_K}\right]\right\} + \\
&\quad \frac{\mathbf{e}'\mathbf{P}(\mathbf{w})\mathbf{e}}{nT_K} - \frac{p^2}{T_K}\zeta(\mathbf{w}) \\
&= \sum_{s,r=1}^{K} w_s w_r \left\{ \left[\frac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}(s)}{nT_K} + \frac{p}{T_K}\hat{\boldsymbol{\xi}}(s)\right]' \left(\frac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)}{nT_K}\right)^{-1} \left[\frac{\ddot{\mathbf{x}}_{\cdot T}(s)'\ddot{\mathbf{s}}_{\cdot T}(r)}{n} - \frac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{s}(r)}{nT_K}\right]\right\} + \\
&\quad \sum_{s=1}^{K} w_s \left\{ \left[\frac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}(s)}{nT_K} + \frac{p}{T_K}\hat{\boldsymbol{\xi}}(s)\right]' \left(\frac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)}{nT_K}\right)^{-1} \frac{1}{n}\ddot{\mathbf{x}}_{\cdot T}(s)'\ddot{\mathbf{e}}_{\cdot T+1}\right\} - \\
&\quad \sum_{s=1}^{K} w_s \left\{ \mathbf{s}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)\left(\frac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)}{nT_K}\right)^{-1} \frac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{e}}{nT_K}\right\} - \\
&\quad \sum_{s=1}^{K} w_s \left\{ \frac{p}{T_K}\hat{\boldsymbol{\xi}}(s)'\left(\frac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)}{nT_K}\right)^{-1} \frac{\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{e}}{nT_K} + \frac{p^2}{T_K}\zeta(s)\right\},
\end{aligned}$$

noting that $\sum_{s=1}^{K} w_s \mathbf{e}'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)\left(\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)\right)^{-1}\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{e} = \mathbf{e}'\mathbf{P}(\mathbf{w})\mathbf{e}$. Now since $\mathrm{E}\left(\frac{1}{n}\mathbf{x}_{\cdot T}(s)'\mathbf{s}_{\cdot T}(r)\right) = \mathrm{E}\left(\frac{1}{nT_K}\mathbf{X}(s)'\mathbf{s}(r)\right)$, we can establish that

$$\mathrm{E}\left\| \frac{\mathbf{x}_{\cdot T}(s)'\mathbf{s}_{\cdot T}(r)}{n} - \frac{\mathbf{X}(s)'\mathbf{s}(r)}{nT_K} \right\|^{q} \leq C_q \left(\frac{K}{n}\right)^{\frac{q}{2}} \|\Lambda(\mathbf{w})\|^{\frac{q}{2}}$$

using the arguments as those used in the proof of lemma 3 in Ing and Wei (2003). Meanwhile, by Hölder's inequality and lemmas 8.1 (xii) and (vii),

$$\begin{aligned}
\mathrm{E}\left[\frac{1}{nT_K}\mathbf{X}(s)'\mathbf{P}_{\boldsymbol{\tau}}\mathbf{s}(r)\right] &\leq C_q K^{\frac{q}{2}} T_K^{-q} \|\Lambda(\mathbf{w})\|^{\frac{q}{2}}, \\
\mathrm{E}\left[\frac{1}{n}\mathbf{X}(s)'\boldsymbol{\tau}\left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}_T'\mathbf{s}_{\cdot T}(r)\right] &\leq C_q K^{\frac{q}{2}} T_K^{-\frac{q}{2}} \|\Lambda(\mathbf{w})\|^{\frac{q}{2}}, \\
\mathrm{E}\left[\frac{1}{n}\mathbf{s}(r)'\boldsymbol{\tau}\left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}_T'\mathbf{x}_{\cdot T}(r)\right] &\leq C_q K^{\frac{q}{2}} T_K^{-\frac{q}{2}} \|\Lambda(\mathbf{w})\|^{\frac{q}{2}}, \\
\mathrm{E}\left[\frac{1}{n}\mathbf{s}(r)'\boldsymbol{\tau}\left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}_T'\boldsymbol{\tau}_T\left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}'\mathbf{X}(s)\right] &\leq C_q K^{\frac{q}{2}} T_K^{-q} \|\Lambda(\mathbf{w})\|^{\frac{q}{2}},
\end{aligned}$$

so that by Hölder's inequality and (42), (43) and (44), the first term in the decomposition for II has expectation bounded by $CK\left[(nT_K)^{-\frac{1}{2}} + T_K^{-1}\right]\left(T_K^{-\frac{1}{2}} + n^{-\frac{1}{2}}\right)\|\Lambda(\mathbf{w})\|^{\frac{1}{2}}$. Next, noting that $\mathrm{E}\left[\frac{1}{n}\mathbf{x}_{\cdot T}(s)'\boldsymbol{\tau}\left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}_T'\boldsymbol{\tau}_T\left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}'\mathbf{e}\right] = \frac{p}{T_K}\boldsymbol{\xi}(s) + O\left(T_K^{-2}\right)$,

$$\mathrm{E}\left\|\frac{1}{nT_K}\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{s}(s)\right\|^q \leq C_q K^{\frac{q}{2}}\left((nT_K)^{-\frac{q}{2}} + T_K^{-q}\right)\|\Lambda(\mathbf{w})\|^{\frac{q}{2}}, \tag{45}$$

$$\mathrm{E}\left\|\tfrac{1}{nT_K}\mathbf{X}(s)'\mathbf{P}_{\boldsymbol{\tau}}\mathbf{e}-\tfrac{p}{T_K}\hat{\boldsymbol{\xi}}(s)\right\|^q \le C_q K^{\frac{q}{2}}\left(n^{-\frac{q}{2}}T_K^{-\frac{3q}{2}}+T_K^{-2q}\right), \tag{46}$$

$$\mathrm{E}\left\|\tfrac{1}{n}\mathbf{x}_{\cdot T}(s)'\boldsymbol{\tau}\left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}_T'\boldsymbol{\tau}_T\left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}'\mathbf{e}-\tfrac{p}{T_K}\boldsymbol{\xi}(s)\right\|^q \le C_q K^{\frac{q}{2}}\left(n^{-\frac{q}{2}}T_K^{-\frac{3q}{2}}+T_K^{-2q}\right),$$

and

$$\mathrm{E}\left[\tfrac{\mathbf{e}'\mathbf{X}(s)}{nT_K}\boldsymbol{\Gamma}^{-1}(s)\tfrac{\mathbf{x}_{\cdot T}(s)'\boldsymbol{\tau}_T(\boldsymbol{\tau}'\boldsymbol{\tau})^{-1}\boldsymbol{\tau}'\mathbf{e}}{n}\right]=O\left(Kn^{-1}T_K^{-2}\right),$$

using (42), (43) and (44) together with Hölder's inequality, we can show that the second term in the decomposition for II has expectation bounded by

$$CK\left[\left(n^{-1}T_K^{-\frac{3}{2}}+n^{-\frac{1}{2}}T_K^{-2}\right)+\tfrac{1}{\sqrt{n}}\left((nT_K)^{-\frac{1}{2}}+T_K^{-1}\right)\|\Lambda(\mathbf{w})\|^{\frac{1}{2}}\right]$$

Next, by Hölder's inequality and (42), (43), (44) and (45) we can show that the third term in the decomposition for II has expectation bounded by

$$CK\left(\left[n^{-1}T_K^{-\frac{3}{2}}+T_K^{-3}\right]+\left(\tfrac{1}{\sqrt{n}}+\tfrac{1}{T_K^2}\right)\left[(nT_K)^{-\frac{1}{2}}+T_K^{-1}\right]\|\Lambda(\mathbf{w})\|^{\frac{1}{2}}\right)$$

For the final term, by (46) and similar arguments we can establish that its expectation is bounded by

$$CK\left[n^{-1}T_K^{-\frac{1}{2}}+T_K^{-1}\right]\left(n^{-\frac{1}{2}}T_K^{-\frac{3}{2}}+T_K^{-2}\right)=CK\left[n^{-\frac{3}{2}}T_K^{-2}+n^{-\frac{1}{2}}T_K^{-\frac{5}{2}}+n^{-1}T_K^{-\frac{5}{2}}+T_K^{-3}\right]$$

Lastly, for III we have

$$\begin{aligned}
III &= \sum_{s,r=1}^{K} w_s w_r\left\{\tfrac{1}{nT_K}\mathbf{u}(r)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}(s)-\tfrac{1}{n}\ddot{\mathbf{u}}_{\cdot T+1}(r)'\ddot{\mathbf{u}}_{\cdot T+1}(s)\right\}+\tfrac{p^2+p}{T_K}\Sigma\\
&= \left(\tfrac{1}{nT_K}\mathbf{e}'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{e}-\tfrac{1}{n}\ddot{\mathbf{e}}_{\cdot T+1}'\ddot{\mathbf{e}}_{\cdot T+1}+\tfrac{p^2+p}{T_K}\Sigma\right)+\\
&\quad \left(\tfrac{1}{nT_K}\mathbf{e}'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{s}(\mathbf{w})-\tfrac{1}{n}\ddot{\mathbf{e}}_{\cdot T+1}'\ddot{\mathbf{s}}_{\cdot T}(\mathbf{w})\right)+\left(\tfrac{1}{nT_K}\mathbf{s}(\mathbf{w})'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{e}-\tfrac{1}{n}\ddot{\mathbf{s}}_{\cdot T}(\mathbf{w})'\ddot{\mathbf{e}}_{\cdot T+1}\right)+\\
&\quad \left(\tfrac{1}{nT_K}\mathbf{s}(\mathbf{w})'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{s}(\mathbf{w})-\tfrac{1}{n}\ddot{\mathbf{s}}_{\cdot T}(\mathbf{w})'\ddot{\mathbf{s}}_{\cdot T}(\mathbf{w})\right)
\end{aligned}$$

Note that $\mathrm{E}\left(\tfrac{1}{nT_K}\mathbf{e}'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{e}-\tfrac{1}{n}\ddot{\mathbf{e}}_{\cdot T+1}'\ddot{\mathbf{e}}_{\cdot T+1}\right) = -\tfrac{p^2+p}{T_K}\Sigma$, so that the expectation of the first by he Cauchy-Schwarz inequality we have

$$\begin{aligned}
&\mathrm{E}\left[\tfrac{1}{nT_K}\mathbf{e}'\mathbf{P}_{\boldsymbol{\tau}}\mathbf{s}(\mathbf{w})-\tfrac{1}{n}\mathbf{e}'\boldsymbol{\tau}\left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}_T'\boldsymbol{\tau}_T\left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}'\mathbf{s}(\mathbf{w})\right]\\
=\;&\mathrm{E}\left[\tfrac{1}{n}\mathbf{e}'\boldsymbol{\tau}\left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\left(\tfrac{1}{T_K}\boldsymbol{\tau}'\boldsymbol{\tau}-\boldsymbol{\tau}_T'\boldsymbol{\tau}_T\right)\left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}'\mathbf{s}(\mathbf{w})\right]\le CT_K^{-1}\|\Lambda(\mathbf{w})\|^{\frac{1}{2}},
\end{aligned}$$

and using the same arguments as used in the proof of Theorem 3.1 in Greenaway-McGrevy (2015, p. 123), we have

$$\mathrm{E}\left[\tfrac{1}{n}\mathbf{e}'\boldsymbol{\tau}\left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}_T'\mathbf{s}_{\cdot T}(\mathbf{w})\right]\le CT_K^{-2}.$$

Then for the final term, note that $\mathrm{E}\left(\frac{1}{nT_K}\mathbf{s}(\mathbf{w})'\mathbf{s}(\mathbf{w})\right) = \mathrm{E}\left(\frac{1}{n}\mathbf{s}_{\cdot T}(\mathbf{w})'\mathbf{s}_{\cdot T}(\mathbf{w})\right) = \Lambda(\mathbf{w})$, and

$$\begin{aligned}
\mathrm{E}\left[\tfrac{1}{nT_K}\mathbf{s}(\mathbf{w})'\mathbf{P}_{\boldsymbol{\tau}}\mathbf{s}(\mathbf{w})\right] &\leq CT_K^{-1}\|\Lambda(\mathbf{w})\|,\\
\mathrm{E}\left[\tfrac{1}{n}\mathbf{s}(\mathbf{w})'\boldsymbol{\tau}\left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}_T'\mathbf{s}_{\cdot T}(\mathbf{w})\right] &\leq CT_K^{-\frac{1}{2}}\|\Lambda(\mathbf{w})\|,\\
\mathrm{E}\left[\tfrac{1}{n}\mathbf{s}(\mathbf{w})'\boldsymbol{\tau}\left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}_T'\boldsymbol{\tau}_T\left(\boldsymbol{\tau}'\boldsymbol{\tau}\right)^{-1}\boldsymbol{\tau}'\mathbf{s}(\mathbf{w})\right] &\leq CT_K^{-1}\|\Lambda(\mathbf{w})\|.
\end{aligned}$$

Thus by application of Holders' inequality, the expectation of III is bounded by $C\left[T_K^{-2}+T_K^{-\frac{1}{2}}\|\Lambda(\mathbf{w})\|\right]$. Thus, putting all the bounds together, we obtain

$$\begin{aligned}
b_{n,T}(\mathbf{w}) \leq\ & C\left[T_K^{-2}+T_K^{-\frac{1}{2}}\|\Lambda(\mathbf{w})\|\right]+CK\left[(nT_K)^{-\frac{1}{2}}+T_K^{-1}\right]\left(T_K^{-\frac{1}{2}}+n^{\frac{1}{2}}\right)\|\Lambda(\mathbf{w})\|^{\frac{1}{2}}+\\
& C\left[(nT_K)^{-1}+T_K^{-2}\right]\left(\tfrac{K}{\sqrt{n}}+\|\Lambda(\mathbf{w})\|^{\frac{1}{2}}\right)=o\left(K(nT_K)^{-1}\right)+o(\|\Lambda(\mathbf{w})\|)+O\left(T_K^{-2}\right),
\end{aligned}$$

which is the stated result. ∎

8.2.2 Theorem 4.1

Refer to section 8.1.1 above for notation employed in this subsection. We begin by decomposing

$$\begin{aligned}
\mathcal{R}_{n,T}(\mathbf{w}) =\ & \mathcal{L}_{n,T}^{(a)}(\mathbf{w})-\tfrac{2}{nT_K}\mathbf{e}'\mathbf{P}(\mathbf{w})\mathbf{e}+\tfrac{2p^2}{T_K^2}\zeta(\mathbf{w})-\left(\tfrac{p}{T_K}+\tfrac{p^2}{T}+\tfrac{p^2}{T^2}\mathbf{w}'\mathbf{K}\mathbf{w}\right)\Sigma+\\
& \underbrace{\mathcal{R}_{n,T}(\mathbf{w})-\mathcal{L}_{n,T}^{(a)}(\mathbf{w})+\tfrac{2}{nT_K}\mathbf{e}'\mathbf{P}(\mathbf{w})\mathbf{e}-\tfrac{2p^2}{T_K^2}\zeta(\mathbf{w})+\left(\tfrac{p}{T_K}+\tfrac{p^2}{T}+\tfrac{p^2}{T^2}\mathbf{w}'\mathbf{K}\mathbf{w}\right)\Sigma}_{=:b_{n,T}^{(a)}(\mathbf{w})}
\end{aligned}$$

Meanwhile from (24) we have $\mathcal{R}_{n,T}(\mathbf{w}) = \hat{L}_{n,T}^{(a)}(\mathbf{w})-\left(\frac{p}{T_K}+\frac{p^2}{T}+\frac{p^2}{T^2}\mathbf{w}'\mathbf{K}\mathbf{w}\right)\hat{\Sigma}(K)$. Substituting this into the expression above and rearranging yields

$$\hat{L}_{n,T}^{(a)}(\mathbf{w})=\mathcal{L}_{n,T}^{(a)}(\mathbf{w})+\tfrac{2}{nT_K}\left[\mathbf{k}'\mathbf{w}\hat{\Sigma}(K)-\mathbf{e}'\mathbf{P}(\mathbf{w})\mathbf{e}+\tfrac{p^2n}{T_K}\zeta(\mathbf{w})\right]+\tfrac{p^2+p}{T_K}\left(\hat{\Sigma}(K)-\Sigma\right)+b_{n,T}^{(a)}(\mathbf{w})$$

where lemma 8.3 below shows that $\left\|\mathrm{E}\left(b_{n,T}^{(a)}(\mathbf{w})\right)\right\| = o\left(K(nT_K)^{-1}\right)+o(\|\Lambda(\mathbf{w})\|)+O\left(T_K^{-2}\right)$. Thus, taking expectations incorporating (37) and (38) yields

$$\mathrm{E}\left(\hat{L}_{n,T}^{(a)}(\mathbf{w})\right)=\mathrm{E}\left(\mathcal{L}_{n,T}^{(a)}(\mathbf{w})\right)+\left(\sum_{s=1}^{K}w_s\tfrac{2s}{nT_K}+\tfrac{p^2+p}{T_K}\right)\left\{\Lambda(K)+o\left(T_K^{-1}\right)\right\}+o\left(\tfrac{K}{nT_K}\right)+o(\|\Lambda(\mathbf{w})\|)+O\left(T_K^{-2}\right).$$

The stated result then follows, noting that $\left[K(nT_K)^{-1}+KT_K^{-2}+\Lambda(\mathbf{w})\right]^{-1}T_K^{-2}\leq K^{-1}\to 0$ as $K\to\infty$ without restriction on n relative to T.

Lemma 8.3 $\left\|\mathrm{E}\left(b_{n,T}^{(a)}(\mathbf{w})\right)\right\| = o\left(K(nT_K)^{-1}\right)+o(\|\Lambda(\mathbf{w})\|)+O\left(T_K^{-2}\right)$.

Proof. Throughout the proof we rely on the fact that lemma 8.1 applies with $\underline{\mathbf{X}}(s)$, $\underline{\mathbf{s}}(s)$, $\underline{\mathbf{e}}$ and T_s replacing $\mathbf{X}(s)$, $\mathbf{s}(s)$, $\mathbf{e}$, and T_K, respectively. We decompose the forecast error as follows

$$\check{\mathbf{u}}_{\cdot T+1}(s) = \underbrace{\underline{\ddot{\mathbf{u}}}_{\cdot T+1}(s) - \ddot{\mathbf{x}}_{\cdot T}(s)(\hat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}(s))}_{=:\check{\mathbf{u}}^{\circ}_{\cdot T+1}(s)} + \underbrace{\ddot{\mathbf{x}}_{\cdot T}(s)(\hat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}(s)) - \underline{\ddot{\mathbf{x}}}_{\cdot T}(s)(\check{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}(s))}_{=:\mathbf{a}_T(s)},$$

where

$$\underline{\ddot{\mathbf{u}}}_{\cdot T+1}(s) := \mathbf{u}_{\cdot T+1}(s) - \tau_{T-s}\left(\boldsymbol{\tau}'_{T-s}\boldsymbol{\tau}_{T-s}\right)^{-1}\boldsymbol{\tau}'_{T-s}\underline{\mathbf{u}}(s).$$

Our strategy is to show that $\mathbf{a}_T(k)$ is asymptotically negligible, so that $\mathrm{E}\left(\mathcal{L}^{(a)}_{n,T}(\mathbf{w})\right)$ is dominated by $\mathrm{E}\left(\frac{1}{n}\check{\mathbf{u}}^{\circ}_{\cdot T+1}(\mathbf{w})'\check{\mathbf{u}}^{\circ}_{\cdot T+1}(\mathbf{w})\right)$. This requires demonstrating that $\mathrm{E}\left(\frac{1}{n}\mathbf{a}_T(\mathbf{w})'\mathbf{a}_T(\mathbf{w})\right) = o\left(K\left[n^{-1}T_K^{-1} + T_K^{-2}\right]\right)$ and $\mathrm{E}\left(\frac{1}{n}\mathbf{a}_T(\mathbf{w})'\check{\mathbf{u}}^{\circ}_{\cdot T+1}(\mathbf{w})\right) = o\left(K\left[n^{-1}T_K^{-1} + T_K^{-2}\right]\right)$. First, we decompose $\mathbf{a}_T(k)$ as follows:

$$\begin{aligned}
\mathbf{a}_T(\mathbf{w}) \;=\; & \sum_{s=1}^{K} w_s\left\{\ddot{\mathbf{x}}_{\cdot T}(s)\left(\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)\right)^{-1}\left[\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}(s) + pn\hat{\boldsymbol{\xi}}(s) - \underline{\mathbf{X}}(s)'\mathbf{M}_{\boldsymbol{\tau}}(s)\underline{\mathbf{u}}(s) - pn\check{\boldsymbol{\xi}}(s)\right]\right\} + \\
& \sum_{s=1}^{K} w_s\left\{\ddot{\mathbf{x}}_{\cdot T}(s)\left[\left(\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)\right)^{-1} - \left(\underline{\mathbf{X}}(s)'\mathbf{M}_{\boldsymbol{\tau}}(s)\underline{\mathbf{X}}(s)\right)^{-1}\right]\left[\underline{\mathbf{X}}(s)'\mathbf{M}_{\boldsymbol{\tau}}(s)\underline{\mathbf{u}}(s) + pn\check{\boldsymbol{\xi}}(s)\right]\right\} + \\
& \sum_{s=1}^{K} w_s\left\{\left(\ddot{\mathbf{x}}_{\cdot T}(s) - \underline{\ddot{\mathbf{x}}}_{\cdot T}(s)\right)\left(\underline{\mathbf{X}}(s)'\mathbf{M}_{\boldsymbol{\tau}}(s)\underline{\mathbf{X}}(s)\right)^{-1}\left[\underline{\mathbf{X}}(s)'\mathbf{M}_{\boldsymbol{\tau}}(s)\underline{\mathbf{u}}(s) + pn\check{\boldsymbol{\xi}}(s)\right]\right\},
\end{aligned}$$

where

$$\underline{\ddot{\mathbf{x}}}_{\cdot T}(s) := \mathbf{x}_{\cdot T}(s) - \tau_{T-s}\left(\boldsymbol{\tau}'_{T-s}\boldsymbol{\tau}_{T-s}\right)^{-1}\boldsymbol{\tau}'_{T-s}\underline{\mathbf{X}}(s).$$

Now it follows that

$$\begin{aligned}
\mathrm{E}\left\|\tfrac{1}{nT_s}\left[\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{u}(s) + pn\hat{\boldsymbol{\xi}}(s) - \underline{\mathbf{X}}(s)'\mathbf{M}_{\boldsymbol{\tau}}(s)\underline{\mathbf{u}}(s) - pn\check{\boldsymbol{\xi}}(s)\right]\right\|^q &\leq C_q\left(\tfrac{\sqrt{K}}{nT_s} + \tfrac{\sqrt{K}+\sqrt{s}}{\sqrt{n}T_s}\right)^q && (47)\\
\mathrm{E}\left\|\left[\left(\tfrac{1}{nT_s}\mathbf{X}(s)'\mathbf{M}_{\boldsymbol{\tau}}\mathbf{X}(s)\right)^{-1} - \left(\tfrac{1}{nT_s}\underline{\mathbf{X}}(s)'\mathbf{M}_{\boldsymbol{\tau}}(s)\underline{\mathbf{X}}(s)\right)^{-1}\right]\right\|^q &\leq C_q\left(\tfrac{K}{T_s} + \tfrac{\sqrt{K}+\sqrt{s}}{\sqrt{n}T_s}\right)^q && (48)\\
\mathrm{E}\left\|\left[\tfrac{1}{nT_s}\underline{\mathbf{X}}(s)'\mathbf{M}_{\boldsymbol{\tau}}(s)\underline{\mathbf{u}}(s) + \tfrac{p}{T_s}\check{\boldsymbol{\xi}}(s)\right]\right\|^q &\leq C_q\left(\tfrac{\sqrt{s}}{\sqrt{nT_s}} + \tfrac{\sqrt{s}}{\sqrt{n}T_s}\right)^q && (49)\\
\mathrm{E}\left\|\tfrac{1}{\sqrt{n}}\left(\ddot{\mathbf{x}}_{\cdot T}(s) - \underline{\ddot{\mathbf{x}}}_{\cdot T}(s)\right)\right\|^q &\leq C_q\left(\tfrac{\sqrt{K}}{T_s}\right)^q, && (50)
\end{aligned}$$

where we assume without loss of generality that $KT_K^{-1} \leq 1$, so straightforwardly it follows that $\mathbf{a}_T(\mathbf{w})$ satisfies $\mathrm{E}\left(\frac{1}{n}\mathbf{a}_T(\mathbf{w})'\mathbf{a}_T(\mathbf{w})\right) = o\left(K\left[n^{-1}T_K^{-1} + T_K^{-2}\right]\right)$ by applications of Hölder's inequality. Next we decompose

$$\check{\mathbf{u}}^{\circ}_{\cdot T+1}(s) := \textstyle\sum_{s=1}^{K} w_s\left\{\underline{\ddot{\mathbf{u}}}_{\cdot T+1}(s) - \ddot{\mathbf{x}}_{\cdot T}(s)(\hat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}(s))\right\},$$

Applying (47) through (50) together with Hölder's inequality, we have

$$\mathrm{E}\left(\tfrac{1}{n}\mathbf{a}_T(r)'\ddot{\mathbf{x}}_{\cdot T}(s)(\hat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}(s))\right) = o\left(K\left[n^{-1}T_K^{-1} + T_K^{-2}\right]\right) \quad (51)$$

Meanwhile, decomposing

$$\underline{\ddot{\mathbf{u}}}_{\cdot T+1}(s) = \underbrace{\mathbf{s}_{\cdot T}(s) - \tau_{T-s}\left(\boldsymbol{\tau}'_{T-s}\boldsymbol{\tau}_{T-s}\right)^{-1}\boldsymbol{\tau}'_{T-s}\underline{\mathbf{s}}(s)}_{=:\underline{\ddot{\mathbf{s}}}_{\cdot T}(s)} + \underbrace{\mathbf{e}_{\cdot T+1} - \tau_{T-s}\left(\boldsymbol{\tau}'_{T-s}\boldsymbol{\tau}_{T-s}\right)^{-1}\boldsymbol{\tau}'_{T-s}\underline{\mathbf{e}}(s)}_{=:\underline{\ddot{\mathbf{e}}}_{\cdot T+1}(s)}$$

we have $\mathrm{E}\left(\frac{1}{n}\mathbf{a}_T(r)'\mathbf{e}_{\cdot T+1}\right) = 0$; by Hölder's inequality, $\mathrm{E}\left\|\frac{1}{\sqrt{n}}\mathbf{a}_T(s)\right\|^q = o(1)$ and $\mathrm{E}\left(\frac{1}{n}\sum_{r,s=1}^K w_s w_r \underline{\ddot{\mathbf{s}}}_{\cdot T}(s)'\underline{\ddot{\mathbf{s}}}_{\cdot T}(r)\right.$ $\boldsymbol{\Lambda}(\mathbf{w})$, we have $\left\|\mathrm{E}\left(\frac{1}{n}\sum_{r,s=1}^K w_s w_r \mathbf{a}_T(s)'\underline{\ddot{\mathbf{s}}}_{\cdot T}(r)\right)\right\| = o(\boldsymbol{\Lambda}(\mathbf{w}))$; and by Hölder's inequality, $\mathrm{E}\left\|\frac{1}{\sqrt{n}}\mathbf{a}_T(s)\right\|^q \leq$ $C_q\left(K^{\frac{q}{2}}n^{-\frac{q}{2}}T_K^{-\frac{3q}{2}}\right)\left(K^q + K^{\frac{q}{2}}\right)$ under (47) through (50), and $\mathrm{E}\left\|\frac{1}{\sqrt{n}}\tau_{T-s}\left(\boldsymbol{\tau}'_{T-s}\boldsymbol{\tau}_{T-s}\right)^{-1}\boldsymbol{\tau}'_{T-s}\underline{\mathbf{e}}(s)\right\|^q \leq$ $C_q T_K^{-\frac{q}{2}}$, so that

$$\mathrm{E}\left(\frac{1}{n}\mathbf{a}_T(\mathbf{w})'\check{\mathbf{u}}^{\circ}_{\cdot T+1}(\mathbf{w})\right) = o\left(K\left[(nT_K)^{-1} + T_K^{-2}\right]\right) + o\left(\|\Lambda(\mathbf{w})\|\right). \tag{52}$$

Having established that $\mathbf{a}_T(\mathbf{w})$ is asymptotically negligible, we can proceed using $\mathrm{E}\left(\mathcal{L}^{(a)}_{n,T}(\mathbf{w})\right) \simeq$ $\mathrm{E}\left(\frac{1}{n}\check{\mathbf{u}}^{\circ}_{\cdot T+1}(\mathbf{w})'\check{\mathbf{u}}^{\circ}_{\cdot T+1}(\mathbf{w})\right)$. Following analogous arguments to those in the proof of lemma 8.2, we can establish that

$$\begin{aligned}
&\mathrm{E}\left(\mathcal{R}^{(a)}_{n,T}(\mathbf{w})\right) - \mathrm{E}\left(\frac{\check{\mathbf{u}}^{\circ}_{\cdot T+1}(\mathbf{w})'\check{\mathbf{u}}^{\circ}_{\cdot T+1}(\mathbf{w})}{n}\right) + \frac{2}{nT_K}\mathrm{E}\left(\mathbf{e}'\mathbf{P}(\mathbf{w})\mathbf{e}\right) - \frac{2p^2}{T_K^2}\zeta(\mathbf{w}) + \left(\frac{p}{T_K} + \frac{p^2}{T} + \frac{p^2}{T^2}\mathbf{w}'\mathbf{K}\mathbf{w}\right)\Sigma \\
= \;& \left\|\mathrm{E}\left(b^{(a)}_{n,T}(\mathbf{w})\right)\right\| = o\left(K\left[(nT)^{-1} + T^{-2}\right] + \|\Lambda(\mathbf{w})\|\right)
\end{aligned}$$

Note that the second term in the expression for $\check{\mathbf{u}}^{\circ}_{\cdot T+1}(s) = \underline{\ddot{\mathbf{u}}}_{\cdot T+1}(s) - \ddot{\mathbf{x}}_{\cdot T}(s)(\hat{\boldsymbol{\alpha}}(s) - \boldsymbol{\alpha}(s))$ is the same as that given for $\hat{\mathbf{u}}_{\cdot T+1}(s)$ stated in (40), which means we can use many of the results above in order to solve for $\mathrm{E}\left(\frac{1}{n}\check{\mathbf{u}}^{\circ}_{\cdot T+1}(\mathbf{w})'\check{\mathbf{u}}^{\circ}_{\cdot T+1}(\mathbf{w})\right)$. The critical difference is that

$$\begin{aligned}
\mathrm{E}\left(\frac{1}{n}\underline{\ddot{\mathbf{e}}}_{\cdot T+1}(\mathbf{w})'\underline{\ddot{\mathbf{e}}}_{\cdot T+1}(\mathbf{w})\right) &= \mathrm{E}\left(\frac{1}{n}\mathbf{e}'_{\cdot T+1}\mathbf{e}_{\cdot T+1}\right) + \\
&\quad \sum_{s,r=1}^{K} w_s w_r \mathrm{E}\left(\frac{1}{n}\tau_{T-s}\left(\boldsymbol{\tau}'_{T-s}\boldsymbol{\tau}_{T-s}\right)^{-1}\boldsymbol{\tau}'_{T-s}\underline{\mathbf{e}}(s)\underline{\mathbf{e}}(r)'\boldsymbol{\tau}_{T-r}\left(\boldsymbol{\tau}'_{T-r}\boldsymbol{\tau}_{T-r}\right)^{-1}\tau_{T-r}\right) \\
&= \Sigma + \Sigma\left(\frac{p^2}{T} + \frac{p^2}{T^2}\mathbf{w}'\mathbf{K}\mathbf{w}\right) + o\left(T_K^{-3}\right).
\end{aligned}$$

∎

8.2.3 Theorem 4.2

Refer to section 8.1.2 above for notation employed in this subsection. The proof of Theorem 4.2 follows the same strategy as the proof to Theorem 3.1. First, we note that analogous versions of the results given in lemma 8.1 apply under Assumption D, but with $\ddot{\mathbf{X}}(s)$, $\ddot{\mathbf{s}}(s)$ and $\ddot{\mathbf{e}}$ replacing

$\mathbf{X}(s)$, $\mathbf{s}(s)$ and $\mathbf{e}$, respectively. Next, we use a similar decomposition as that given in (36) above

$$\begin{aligned}\mathcal{R}_{n,T}^{(b)}(\mathbf{w}) &= \mathcal{L}_{n,T}^{(b)}(\mathbf{w}) - \tfrac{2}{nT_K}\ddot{\mathbf{e}}'\mathbf{P}(\mathbf{w})\ddot{\mathbf{e}} + \tfrac{2}{T_K^2}\zeta(\mathbf{w}) - \left(\tfrac{p^2+p}{T_K} + \tfrac{2}{n}\right)\Sigma + \\ &\quad \underbrace{\mathcal{R}_{n,T}^{(b)}(\mathbf{w}) - \mathcal{L}_{n,T}^{(b)}(\mathbf{w}) + \tfrac{2}{nT_K}\ddot{\mathbf{e}}'\mathbf{P}(\mathbf{w})\ddot{\mathbf{e}} - \tfrac{2}{T_K^2}\zeta(\mathbf{w}) + \left(\tfrac{p^2+p}{T_K} + \tfrac{2}{n}\right)\Sigma}_{=:b_{n,T}^{(b)}(\mathbf{w})},\end{aligned}$$

Meanwhile, from (25) we have $\mathcal{R}_{n,T}^{(b)}(\mathbf{w}) = \hat{L}_{n,T}^{(b)}(\mathbf{w}) - \left(\frac{2}{nT_K}\mathbf{k}'\mathbf{w} + \frac{p^2+p}{T_K} + \frac{2}{n}\right)\hat{\hat{\Sigma}}(K)$. Substituting this into the expression above and rearranging yields

$$\hat{L}_{n,T}^{(b)}(\mathbf{w}) = \mathcal{L}_{n,T}^{(b)}(\mathbf{w}) + \tfrac{2}{nT_K}\left[\mathbf{k}'\mathbf{w}\hat{\hat{\Sigma}}(K) - \ddot{\mathbf{e}}'\mathbf{P}(\mathbf{w})\ddot{\mathbf{e}} + \tfrac{n}{T_K}\zeta(\mathbf{w})\right] + \left(\tfrac{p^2+p}{T_K} + \tfrac{2}{n}\right)\left(\hat{\hat{\Sigma}}(K) - \Sigma\right) + b_{n,T}^{(b)}(\mathbf{w}). \tag{53}$$

Now we have

$$\mathrm{E}\left(\hat{\hat{\Sigma}}(K)\right) = \Sigma + \Lambda(K) + o\left(T_K^{-1}\right) + o\left(n^{-1}\right)$$

and

$$\tfrac{1}{nT_K}\mathrm{E}\left(\ddot{\mathbf{e}}'\mathbf{P}(\mathbf{w})\ddot{\mathbf{e}}\right) = \textstyle\sum_{s=1}^{K} w_s\tfrac{2s}{nT_K}\Sigma + \tfrac{1}{T_K^2}\zeta(\mathbf{w}) + o\left(\tfrac{1}{T_K^2}\right)$$

Meanwhile, using similar arguments to those used in the proof of lemma 8.2, we have $\left\|\mathrm{E}\left(b_{n,T}^{(b)}(\mathbf{w})\right)\right\| = o\left(K(nT_K)^{-1}\right) + o(\|\Lambda(\mathbf{w})\|) + O\left(T_K^{-2}\right)$. Thus, noting that $\frac{2}{nT_K}\mathbf{k}'\mathbf{w} = \sum_{s=1}^{K} w_s\frac{2s}{nT_K}$, taking expectations of (36) we have

$$\begin{aligned}\mathrm{E}\left(\hat{L}_{n,T}^{(b)}(\mathbf{w})\right) &= \mathrm{E}\left(\mathcal{L}_{n,T}^{(b)}(\mathbf{w})\right) + \left(\textstyle\sum_{s=1}^{K} w_s\tfrac{2s}{nT_K} + \tfrac{p^2+p}{T_K} + \tfrac{2}{n}\right)\left\{\Lambda(K) + o\left(T_K^{-1}\right) + o\left(n^{-1}\right)\right\} + \\ &\quad o\left(\tfrac{K}{nT_K}\right) + o(\|\Lambda(\mathbf{w})\|) + O\left(T_K^{-2}\right).\end{aligned}$$

The stated result then follows using the same arguments as given in the proof of Theorem 3.1.

8.2.4 Theorem 4.3

The proof follows the same steps as that of the proof of Theorem 3.1. We have a similar decomposition as that given in (36) above, except

$$\hat{L}_{n,T}^{(c)}(\mathbf{w}) = \mathcal{L}_{n,T}^{(c)}(\mathbf{w}) + \tfrac{2}{nT_K}\left[(\mathbf{k}\otimes I_m)'\mathbf{D}(\Phi)(\mathbf{w}\otimes I_m) - \mathbf{e}'\mathbf{P}(\mathbf{w})\mathbf{e} + \tfrac{n}{T_K}\zeta(\mathbf{w})\right] + \tfrac{p^2+p}{T_K}\left(\hat{\Sigma}(K) - \Sigma\right) + b_{n,T}(\mathbf{w})$$

where $b_{n,T}(\mathbf{w})$ is defined as above and continues to satisfy $\mathrm{E}\|b_{n,T}(\mathbf{w})\| = o\left(K(nT_K)^{-1}\right) + O\left(T_K^{-2}\right) + o(\|\Lambda(\mathbf{w})\|)$ under weak dependence. Now using lemma 7.6 from Greenaway-McGrevy (2018a) we have

$$\mathrm{E}\left[(\mathbf{k}\otimes I_m)'\mathbf{D}(\Phi)(\mathbf{w}\otimes I_m)\right] - \mathrm{E}\left(\mathbf{e}'\textstyle\sum_{s=1}^{K} w_s\mathbf{X}(s)\left(\mathbf{X}(s)'\mathbf{X}(s)\right)^{-1}\mathbf{X}(s)'\mathbf{e}\right) \le C\left(T_K^{-1/2} + \|\Lambda(K)\|\right),$$

so that

$$\mathrm{E}\left(\hat{L}_{n,T}^{(c)}(\mathbf{w})\right) = \mathrm{E}\left(\mathcal{L}_{n,T}^{(c)}(\mathbf{w})\right) + \frac{p^2+p}{T_K}\left\{\Lambda(K) + o\left(T_K^{-1}\right)\right\} + o\left(\frac{K}{nT_K}\right) + o\left(\|\Lambda(\mathbf{w})\|\right) + O\left(T_K^{-2}\right)$$

and thus the stated result follows using the same arguments as given in the proof of Theorem 3.1.

References

[1] Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* 23, 163–180.

[2] Akaike, H. (1970). Statistical predictor identification. *Annals of the Institute of Statistical Mathematics* 22, 203–419.(chained 2009 dollars)

[3] Amior, M. and A. Manning (2018). The Persistence of Local Joblessness. *The American Economic Review* 107, 1942–70.

[4] Anselin, L. (2001). Spatial econometrics, Ch.14 in A companion to theoretical econometrics, ed., B.H. Baltagi, Blackwell Publishers Ltd., Massachusetts.

[5] Baltagi, B.H. (2008) Forecasting with Panel Data. *Journal of Forecasting* 27, 153–173.

[6] Baltagi, B. H., and Q. Li (1992) Prediction in the One-way Error Component Model with Serial Correlation. *Journal of Forecasting* 11, 561–567.

[7] Bates, J.M., Granger, C.M.W. (1969). The combination of forecasts. Operations Research Quarterly 20, 451 468.

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

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

[10] Blanchard, O. J. and L. F. Katz (1992). Regional Evolutions. *Brookings Papers on Economic Activity*, Vol. 23 (1), pp. 1–76.

[11] Brillinger, D. R. (1981). Time series: data analysis and theory. San Francisco, CA: Holden-Day.

[12] Chevillon, G. (2007). Direct Multi-step Estimation and Forecasting. *Journal of Economic Surveys* 21, 746–785. doi:10.1111/j.1467-6419.2007.00518.x

[13] Chudik, A., Pesaran, M. H., and Tosetti, E. (2011). Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal*, 1, 45–90.

[14] Clemen, R.T. (1989). "Combining forecasts: A review and annotated bibliography". *International Journal of Forecasting* 5, 559–581.

[15] Diebold, F.X., Lopez, J.A. (1996). "Forecast evaluation and combination". In: Maddala, G.S., Rao, C.R. (Eds.), Statistical Methods in Finance, Handbook of Statistics, vol. 14. Elsevier, Amsterdam, pp. 241– 268.

[16] Eisinberg, A., and G. Fedele (2007) Discrete Orthogonal Polynomials on Equidistant Nodes. *International Mathematical Forum* 21, 1007–1020

[17] Findley, D.F. (1984) On some ambiguities associated with the fitting of ARMA models to time series. *Journal of Time Series Analysis* 5, 213–225.

[18] Findley, D.F., and C.Z.,Wei (1993) Moment bounds for deriving time series CLT's and model selection procedures. *Statistica Sinica* 3, 453–480.

[19] Findley, D.F., and 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.

[20] Fuller, W.A. and D.P. Hasza (1981). Properties of predictors for autoregressive time series. *Journal of American Statistical Association* 76, 155–161.

[21] Granger, C.W.J., 1989. Combining Forecasts Twenty years later. *Journal of Forecasting* 8, 167 173.

[22] Granger, C.W.J., and R. Ramanathan (1984). Improved methods of combining forecast accuracy. *Journal of Forecasting* 19, 197 204.

[23] Greenaway-McGrevy, R. (2013). Multistep Prediction of Panel Autoregressive Processes. *Econometric Theory* 29, 699-734

[24] Greenaway-McGrevy, R. (2015). Evaluating Panel Data Forecasts under Independent Realization. *Journal of Multivariate Analysis*, forthcoming.

[25] Greenaway-McGrevy, R. (2018a). Asymptotically Efficient Model selection for Panel Data Forecasting *Econometric Theory,* forthcoming.

[26] Greenaway-McGrevy, R. (2018b). Multistep Forecast Selection for Panel Data. SSRN

[27] Greenaway-McGrevy, R., and K. K. Hood (2016). Worker Migration or Job Creation? Persistent Shocks and Regional Recoveries. *Journal of Urban Economics*, 96, 1-16.

[28] Hahn, J., and G. Kuersteiner (2002). Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects. *Econometrica* 70, 1639–1657.

[29] Han, C., Phillips, P. C. B., and D. Sul (2013). Lag length selection in panel autoregression. Forthcoming in *Econometric Reviews*

[30] Hansen, B. E. (2008). Least Squares Forecast Averaging, *Journal of Econometrics* 146, 342–350.

[31] Hansen, C. B. (2007). Asymptotic properties of a robust variance matrix when T is large. *Journal of Econometrics,* 141, 597–620.

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

[33] Ing, C.K. and C.Z. Wei (2003). On same-realization prediction in an infinite-order autoregressive process. *Journal of Multivariate Analysis* 85, 130–155.

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

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

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

[37] Lee, Y. (2006). A General Approach to Bias Correction in Dynamic Panels under Time Series Misspecification. Ph.D. Dissertation, Yale University.

[38] Lee, Y. (2012). Bias in Dynamic Panel Models under Time Series Misspecification. *Journal of Econometrics* 169, 54–60.

[39] Lee, Y. and P.C.B. Phillips (2015). Model Selection in the Presence of Incidental Parameters. *Journal of Econometrics,* forthcoming. Retrieved from: http://www.sciencedirect.com/science/article/pii/S0304407615000810

[40] Mallows, C. L. (1973). Some Comments on C_p *Technometrics* 15, 661–675.

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

[42] Newbold, P., and Harvey, D.I. (2001). "Forecast combination and encompassing". In: Clements, M.P., Hendry, D.F. (Eds.), A Companion to Economic Forecasting. Blackwell, Oxford

[43] Pesaran, M. H., and E. Tosetti (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics* 161, 182–202.

[44] Phillips, P. C. B., and Sul, D. (2003). Dynamic panel estimation and homogeneity testing under cross section dependence. *Econometrics Journal*, 6(1), 217–259.

[45] Phillips, P.C.B. and D. Sul (2007). Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and Cross Section Dependence. *Journal of Econometrics* 137, 162–188.

[46] Schorfheide, F. (2005). VAR Forecasting under Misspecification *Journal of Econometrics* 128, 99–136.

[47] 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.

[48] Speed, T.P. and B. Yu (1993) Model selection and prediction: normal regression. *Annals of the Institute of Statistical Mathematics* 45, 35–54.

[49] Taub, A.J. (1979). Prediction in the Context of the Variance Components Model. *Journal of Econometrics* 10, 103–107.

[50] Timmermann, Allan, (2006) Forecast combinations. In: Elliott, G., Granger, C.W.J., Timmermann, A. (Eds.), Handbook of Economic Forecasting, vol. 1. Elsevier, Amsterdam, pp. 135 196.