Conditional Rotation Between Forecasting Models

Yinchu Zhu[a] Allan Timmermann[b]

[a] Department of Economics, Brandeis University, 415 South Street Waltham, MA 02453, U.S.A.; yinchuzhu@brandeis.edu (corresponding author)

[b] Rady School of Management, University of California, San Diego, 9500 Gilman Dr, La Jolla, CA 92093, U.S.A.; atimmermann@ucsd.edu

March 14, 2021

Abstract

We establish conditions under which forecasting performance can be improved by rotating between a set of underlying forecasts whose predictive accuracy is tracked using a set of time-varying monitoring instruments. We characterize the properties that the monitoring instruments must possess to be useful for identifying, at each point in time, the best forecast and show that these reflect both the accuracy of the predictors used by the underlying forecasting models and the strength of the monitoring instruments. Finite-sample bounds on forecasting performance that account for estimation error are used to compute the expected loss of the competing forecasts as well as for the dynamic rotation strategy. Finally, using Monte Carlo simulations and empirical applications to forecasting inflation and stock returns, we demonstrate the potential gains from using conditioning information to rotate between forecasts.

Keywords: Forecasting performance; real time monitoring; finite sample bounds

JEL classifications: C53, C32, C18

1 Introduction

Decision makers frequently observe two or more forecasts of the same outcome. A common strategy in this situation is to compare the predictive accuracy of the different forecasts and choose that which minimizes expected loss. Indeed, a large academic literature–including important contributions by Granger and Newbold (1977), Chong and Hendry (1986), Diebold and Mariano (1995), West (1996), and Clark and McCracken (2001)–has developed ways to formally compare the accuracy of different forecasts. Such comparisons typically consider whether one forecast on average (unconditionally) is more accurate than other, competing, forecasts.

However, past forecasting performance is frequently found to be a poor predictor of future performance (Aiolfi and Timmermann (2006)) and it is possible for forecasts to be poor on average, yet still be relatively accurate in some states of the world. Provided that such periods can be identified ex-ante by means of a set of time-varying monitoring instruments, a forecast that is poor on average could be the preferred forecast at a given point in time, conditional on information contained in the monitoring instruments. This point is important given the widespread empirical evidence of model instability in economics, (e.g., Rossi (2013) and Stock and Watson (1996)) which shows that it is rare to find a single forecasting approach that dominates other forecasts uniformly through time.

This paper presents a theoretical framework for understanding the factors that determine the feasibility of improving predictive accuracy by monitoring the performance of individual forecasts and choosing, at each point in time and conditional on information in a set of monitoring instruments, the forecast with the best expected performance–a strategy we label dynamic rotation and which was first proposed by Diebold and Mariano (1995) and Giacomini and White (2006). Our setup allows us to rank expected forecasting performance, assuming that the underlying forecasts are generated by a set of linear models whose parameters are updated through time. We compare the predictive accuracy of models that use predictors with different strength and provide conditions under which bounds can be established on differences in the forecasts' mean squared error (MSE) loss. Our approximate uniform bounds hold in finite samples and account for estimation error both in the underlying forecasting model and in the monitoring (rotation) regression. Our theoretical bounds are derived under a set of restrictive assumptions, including the existence of a large number of moments for the underlying data. These assumptions may be valid for some empirical applications but not for others, which should be borne in mind when interpreting the practical implications of our results.

In practice, estimation error tends to be important for understanding the out-of-sample forecasting performance of both individual models and of alternative strategies such as forecast combination. Our analysis is the first to establish precise conditions under which a conditional rotation strategy which uses additional information to select among competing forecasts can dominate common alternatives, including in the frequently encountered situation with highly correlated forecasts. Specifically, we characterize the properties that monitoring instruments must possess in order to contain valuable information about time variation in competing models' forecasting performance. In the case with non-nested forecasting models, a monitoring instrument can be used to track time variation in the models' relative squared error forecasting performance if the instrument is sufficiently strongly correlated with the cross-product of the forecast error and predictors included by one model but excluded by others. The strength of both the predictor and the monitoring instrument turn out to matter for our ability to rank different models' conditional forecasting performance. In the nested case, rotating between a "small" and a "big" model which contains an additional predictor can lead to improvements provided that the additional predictor in the big model contains "weak" information, i.e., its coefficient is at most local-to-zero.

Moreover, we show that using monitoring instruments to rotate between competing forecasts can be a better strategy than alternatives such as forecast combination or adding the monitoring instruments directly to the underlying forecasting models. It is possible for a monitoring instrument to be correlated with the difference in two forecasts' MSE loss while at the same time being uncorrelated with the individual models' forecast errors. In this case, the monitoring instrument may be useful for selecting a forecast, even if it is not useful if added directly to the forecasting models. Further, adding a monitoring instrument to the underlying prediction models introduces estimation error which can dominate the signal in the instrument.

We explore the empirical relevance of our theoretical analysis through Monte Carlo simulations and in two empirical applications. The simulations demonstrate instances with sizable gains from rotating between forecasts as well as instances with little or no gains in predictive accuracy. Our empirical applications cover survey and central bank forecasts of US inflation and model-based forecasts of US stock returns. For the inflation application, we find that monitoring instruments can be used to dynamically rotate between survey and Federal Reserve forecasts to significantly reduce the mean squared errors of the least accurate forecast while simultaneously either improving on or preserving the predictive accuracy of the best forecast. For the stock return application–a case with both weak predictors and monitoring instruments–the dynamic rotation scheme performs a little worse than a prevailing mean (small) model but better than a set of larger (univariate) forecasting models.

Our paper proceeds as follows. Section 2 introduces the forecasting environment. Sections 3 (non-nested case) and 4 (nested case) contain our theoretical analysis which establishes how competing forecasting models, as well as the dynamic rotation rule, can be ranked. Section 5 uses Monte Carlo simulations to illustrate the theoretical analysis, while Section 6 presents our empirical applications and Section 7 concludes.

2 Forecast Environment

We first introduce the forecast environment and describe a dynamic rotation strategy that exploits the conditional information in the monitoring instruments to choose between different forecasts.

2.1 Estimates of Relative Forecasting Performance

Pairwise comparisons are now routinely carried out in studies of forecasting performance, see, e.g., Clark and McCracken (2013). Typically it is assumed that forecasts are generated from a set of underlying linear models whose parameters are updated as new information arrives. Moreover, it is common practice to use out-of-sample tests to avoid using the same data sample to estimate model parameters and evaluate the resulting forecasts and to mimic "real-time" forecasting, see Diebold and Rudebusch (1991), Inoue and Kilian (2008), and Pesaran and Timmermann (2009).[1]

In common with previous studies, we focus on univariate forecasting problems and squared error loss. Let $\hat{y}_{1,t+1|t}$ and $\hat{y}_{2,t+1|t}$ be a pair of one-step-ahead forecasts of the outcome y_{t+1} generated using information known at time t. Following Diebold and Mariano (1995), we evaluate the accuracy of the forecasts using a loss function

[1] As pointed out by Inoue and Kilian (2005) and Hansen and Timmermann (2015), splitting the sample in this manner entails a loss in the power of tests of predictive accuracy. We do not address this issue here as our objective is to evaluate the performance of procedures in common use.

$L(\hat{y}_{t+1|t}, y_{t+1})$, where $\hat{y}_{t+1|t} \in \{\hat{y}_{1,t+1|t}, \hat{y}_{2,t+1|t}\}$. Under squared error loss

$$L(\hat{y}_{t+1|t}, y_{t+1}) = (y_{t+1} - \hat{y}_{t+1|t})^2. \tag{2.1}$$

The loss differential, $\Delta L_{t+1} \equiv L(\hat{y}_{1,t+1|t}, y_{t+1}) - L(\hat{y}_{2,t+1|t}, y_{t+1})$, is then given by

$$\Delta L_{t+1} = e^2_{1,t+1} - e^2_{2,t+1}, \tag{2.2}$$

where $e_{j,t+1} = y_{t+1} - \hat{y}_{j,t+1|t}$ for $j = 1, 2$ are the individual forecast errors.

Even if one forecast is worse *on average* than another forecast, it might perform better in certain states of the world. This suggests choosing the model with the best conditionally expected performance given information available to the forecaster at time t, $\mathcal{G}_t$, see Giacomini and White (2006) (GW, henceforth). Choice of forecast is thus naturally based on the sign of $\mathbb{E}[\Delta L_{t+1}|\mathcal{G}_t]$. In practice, this conditional expectation is unknown but, given a set of monitoring instruments $Z_t \in \mathbb{R}^{dz}$ in $\mathcal{G}_t$, we can approximate $\mathbb{E}\left[\Delta L_{t+1}|Z_t\right]$ through a linear regression:

$$\Delta L_{t+1} = (\theta_0, \theta_1)\begin{pmatrix} 1 \\ z_{Mt} \end{pmatrix} + u_{t+1} \equiv \theta' z_t + u_{t+1}, \tag{2.3}$$

where $z_t = (1,\, z_{Mt})'$, $\mathbb{E}[u_{t+1} \mid z_{Mt}] = 0$, and $z_{Mt} \in Z_t$. Instead of the linear regression in (2.3), one could alternatively adopt a probit or logit model. However, while such a specification may be more appealing from an empirical standpoint, the inherent nonlinearity makes it difficult to obtain analytical results under these alternative specifications. In practice, it is oftentimes found that differences between linear probability models and probit or logit models tends to be quite small, especially when one needs to make a binary decision.[2]

2.2 Dynamic Rotation

Using the monitoring regression, (2.3), we can compute the expected future loss differential $\mathbb{E}\left(\Delta L_{t+1}|Z_t\right)$ by $\theta' Z_t$ and use this to choose the preferred forecast for period $t+1$. Following GW, we consider a conditional decision (dynamic rotation,

[2]Under a probit or logit model, the decision would be $\mathbf{1}\{f(\pi' z_t) > c\}$ for a probit or logit link function f and a cutoff point $c \in (0, 1)$. This is equivalent to $\mathbf{1}\{\pi' z_t > f^{-1}(c)\}$ and thus reduces to a decision based on a linear index; the linear probability model avoids issues related to the choice of c.

DR) rule that chooses forecast 1 if $\mathbb{E}(\Delta L_{t+1} \mid Z_t) \leq 0$, otherwise chooses forecast 2:

$$\hat{y}_{DR,t+1|t} = \hat{y}_{1,t+1|t}\mathbf{1}\left\{\mathbb{E}[\Delta L_{t+1}|z_t] \leq 0\right\} + \hat{y}_{2,t+1|t}\mathbf{1}\left\{\mathbb{E}[\Delta L_{t+1}|z_t] > 0\right\}, \tag{2.4}$$

where $\mathbf{1}\left\{\mathbb{E}[\Delta L_{t+1}|z_t] > 0\right\}$ is an indicator variable that equals one if the first forecasting model has the highest expected loss conditional on $Z_t = z_t$, otherwise is zero.

Compared to a strategy of always using forecasts from a particular model, we expect gains from the dynamic rotation in (2.4) provided that neither of the underlying forecasting models is too dominant since $\mathbb{E}\left[\Delta L_{t+1}|z_t\right]$ is required to change sign for different values of z_t. However, this argument does not account for estimation error, nor does it quantify the potential gains from monitoring forecasting performance and, at each point in time, selecting the forecast with the lowest conditionally expected loss. This is the topic of the next two sections.

Model instability is often considered to be a likely source of time-variation in the relative accuracy of a pair of forecasts. For example, discrete breaks to the data generating process can lead to breaks in the average loss differential of alternative forecasting models (Pesaran and Timmermann (2007)). Our analysis focuses on a different mechanism, however. We model time variation in the conditional mean of loss differences given a set of instruments used to monitor forecasting performance. These instruments can be used to track time variation in the relative forecasting performance (loss differentials) across models even in the absence of model instability. Equivalently, even if the sequence of realized losses is a stationary process with no breaks, it can still be predictable; we exploit such predictable patterns in loss differentials and construct a monitoring procedure.

Forecast combination offers a popular alternative to the rotation scheme in (2.4). As we will show in our analysis, forecast combination has some important limitations, however. First, the two forecasts under consideration are often highly correlated. In practice, high correlations often give rise to large and volatile estimates of the combination weights and poor out-of-sample forecasting performance. Second, forecast combinations are typically based on unconditional moments such as the MSE. Conversely, our goal is to incorporate additional information that can be useful in determining the performance of different forecasts. Alternative forecast combination schemes might be possible by explicitly modeling variation in the conditional MSE and the conditional covariance between forecast errors, but it is far from obvious that such an approach leads to a simple and practical methodology.

3 Comparing Forecasts from Non-nested Models

This section studies pair-wise comparisons of forecasts generated by a set of non-nested models, both of which may only partially capture the information in the data generating process for the outcome, y_{t+1}. We establish approximate finite sample bounds on the relative performance of the forecasting models in the presence of estimation error. Further, we also characterize conditions under which the dynamic rotation strategy can be expected to produce better forecasts than either model. To simplify the analysis, we use a stylized setting with only two predictors and a one-period forecast horizon, but note that our analysis can be generalized to a setting with multiple predictors and longer horizons. Consistent with the common finding in empirical analysis that forecasts tend to be strongly correlated, our analysis assumes that the predictive signal is quite weak so that the two forecasts have a correlation tending to one with the same asymptotic MSE. For example, forecasts of individual survey participants tend to be strongly correlated, helping to explain why it is difficult to outperform a simple equal-weighted combination, see, e.g., Chapter 14 in Elliott and Timmermann (2016).

3.1 Pairwise Comparisons

Consider the data generating process (DGP)

$$y_{t+1} = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \varepsilon_{t+1}, \tag{3.1}$$

where x_{1t} and x_{2t} are a set of predictor variables known at time t. To capture the non-nested case, we assume that model 1 takes the form $y_{t+1} = \beta_1 x_{1,t} + \varepsilon_{1t+1}$, while model 2 takes the form $y_{t+1} = \beta_2 x_{2,t} + \varepsilon_{2t+1}$. For simplicity, we assume that x_{1t} and x_{2t} are univariate processes. We do not rule out that β_1 or β_2 equal zero.

Suppose we observe a sample of T data points $\{(x_{j,t}, y_t)\}_{t=1}^{T}$, $j = 1, 2$. We assume that the parameters of the forecasting models ($\hat{\beta}_{j,n,t}$) are estimated using a window of the most recent n observations

$$\hat{\beta}_{j,n,t} = \left(\sum_{s=t-n}^{t-1} x_{j,s}^2 \right)^{-1} \left(\sum_{s=t-n}^{t-1} x_{j,s} y_{,s+1} \right), \ j = 1, 2. \tag{3.2}$$

The resulting forecasts are generated as $\hat{y}_{j,t+1|t} = \hat{\beta}_{j,n,t} x_{j,t}$ for $j = 1, 2$. Under this setup, the sample size, T, is split into a window of length n used to estimate $\hat{\beta}_{j,n,t}$, the parameters of the jth forecasting model, and a test sample containing the remaining $p = T - n$ observations. Both n and p can be functions of T and may or may not tend to infinity but, for simplicity, we write n and p instead of n_T and p_T.

Using (3.1), the squared error loss differential ΔL_{t+1} becomes

$$\begin{aligned} \Delta L_{t+1} &= (y_{t+1} - \hat{y}_{1,t+1|t})^2 - (y_{t+1} - \hat{y}_{2,t+1|t})^2 \\ &= 2\varepsilon_{t+1} \left(\beta_2 x_{2,t} - \beta_1 x_{1,t} - \delta_{1,n,t} x_{1,t} + \delta_{2,n,t} x_{2,t} \right) \\ &\quad + \left(\beta_2 x_{2,t} - \beta_1 x_{1,t} - \delta_{1,n,t} x_{1,t} + \delta_{2,n,t} x_{2,t} \right) \left(\beta_1 x_{1,t} + \beta_2 x_{2,t} - \delta_{1,n,t} x_{1,t} - \delta_{2,n,t} x_{2,t} \right), \end{aligned} \tag{3.3}$$

where $\delta_{j,n,t} = \hat{\beta}_{j,n,t} - \beta_j$ denotes the estimation error for model j.

For a fixed t, one can use a CLT to show that $\delta_{j,n,t} = O_P(n^{-1/2})$ and thus $\Delta L_{t+1} = \beta_2^2 x_{2,t}^2 - \beta_1^2 x_{1,t}^2 + 2(\beta_2 x_{2,t} - \beta_1 x_{1,t})\varepsilon_{t+1} + O_P(n^{-1/2})$. The potential lack of uniformity in the $O_P(n^{-1/2})$ terms presents challenges, however. To overcome these difficulties, we next derive bounds that are valid (approximately) uniformly across t, hold in finite samples, and thus do not require an asymptotic framework.

Remark 1. We now explain the purpose of the non-asymptotic analysis. When several quantities (e.g., n, m and T) tend to infinity, the asymptotic analysis becomes complicated. One has to clarify whether limits are taken sequentially or jointly; if the limit is taken jointly, one often has to make assumptions on the relative rate of the quantities that tend to infinity and the asymptotic results can be sensitive to these assumptions. To provide a more transparent way of deriving asymptotic results, a recent literature in statistics and econometrics adopts non-asymptotic analysis with the aim of deriving bounds that hold in any sample sizes. This ensures that the asymptotic implications will be immediate and transparent.[3]

[3] A well-known example is the analysis of Lasso in which the model dimensionality and the sample size both tend to infinity. Without a non-asymptotic analysis, one could question whether the rate at which the model dimensionality grows impacts the theoretical analysis. In contrast, asymptotic implications are immediately clear from non-asymptotic bounds since taking the (joint) limit is the last step and usually a one-line argument; see , e.g., Candes and Tao (2007); Bickel et al. (2009); Koltchinskii et al. (2011); Belloni et al. (2014)).

3.2 Determinants of Average Forecasting Performance

To ensure that our analysis of forecasting performance allows for a broad set of time-series dependencies, we adopt the beta mixing condition in Chen et al. (2016), adapted to an array setting similarly to Andrews (1988):

Definition 1. The array $\{W_{T,t}\}_{t=-\infty}^{\infty}$ is said to be beta-mixing with coefficient $\mathcal{B}_{mix}(\cdot)$ if

$$\mathcal{B}_{mix}(t) = \sup_{-\infty<i<\infty,\, T\geq 1} \mathbb{E}\left(\sup_{A\in\mathcal{F}_{i+t,\infty}^{T}} \left|\mathbb{P}\left(A|\mathcal{F}_{-\infty,i}^{T}\right) - \mathbb{P}(A)\right| \right),$$

where $\mathcal{F}_{-\infty,i}^{T} = \sigma(\cdots, W_{T,i-1}, W_{T,i})$ and $\mathcal{F}_{i+t,\infty}^{T} = \sigma(W_{T,i+t}, W_{T,i+t+1}, \cdots)$.

The array structure in this definition is general enough to allow for many types of nonstationary data and provides a convenient way of analyzing DGPs indexed by the sample size. The beta-mixing condition can be satisfied by many time series models, including ARMA, GARCH and general hidden Markov models and is commonly used in the literature, see, e.g., Tong (1990); Fan and Yao (2003); Carrasco and Chen (2002); Meyn and Tweedie (2012); Douc et al. (2011). While it is ultimately difficult to verify empirically whether the beta mixing property holds for a given data set, to the extent that it embodies a limited memory feature it is plausible that the property holds for many economic applications spanning long sample periods.

Our analysis makes use of the following list of assumptions:

Assumption 1. *The following hold for $j \in \{1,2\}$:*

(i) There exist constants $r > 8$ and $D > 0$ such that $\mathbb{E}|x_{j,t}|^r$ and $\mathbb{E}|\varepsilon_{t+1}|^r$ are bounded above by D. Moreover, $\mathbb{E}x_{j,t} = \mathbb{E}x_{1,t}x_{2,t} = \mathbb{E}x_{j,t}\varepsilon_{t+1} = 0$.

(ii) $\{x_{j,t}, \varepsilon_{t+1}\}_{t=-\infty}^{\infty}$ is a beta mixing array with coefficient $\mathcal{B}_{mix}(\cdot)$ such that $\forall t > 0$, $\mathcal{B}_{mix}(t) \leq b\exp(-t^c)$, for constants $b, c > 0$. Moreover, for some constants $Q_1, Q_2 > 0$, $\mathbb{E}(k^{-1/2}\sum_{s=t-k}^{t-1} x_{1,s}(x_{2,s}\beta_2 + \varepsilon_{s+1}))^2$, $\mathbb{E}(k^{-1/2}\sum_{s=t-k}^{t-1} x_{2,s}(x_{1,s}\beta_1 + \varepsilon_{s+1}))^2$, $\mathbb{E}x_{1,t}^2$ and $\mathbb{E}x_{2,t}^2$ lie in $[Q_1, Q_2]$ for all k, t.

(iii) There exist constants $\alpha_{x,j} \in [0,\infty]$, $c_{\beta,j} > 0$ such that $\beta_j = c_{\beta,j}n^{-\alpha_{x,j}}$, where

- *(a) $\alpha_{x,2} < \alpha_{x,1}$,*
- *(b) $\alpha_{x,2} < 1/2$.*

(iv) $n/T > \kappa$ *for some constant* $\kappa > 0$.

Some of these assumptions are quite restrictive. For example, the assumption in part (i) that the x-variables have zero mean is restrictive in the non-nested case. In the nested case we can include an intercept and so the zero mean condition is without loss of generality provided $E[x_t]$ does not depend on t. This holds because transforming the X matrix to XA for any invertible matrix A (where A demeans x if $E[x_t]$ does not depend on t) will not change the predictions, even out-of-sample. Relaxing the second part of Assumption 1(i) by allowing the x-variables to be correlated would make it less clear what it means that one signal (x_2) is stronger than the other (x_1) since even the weaker x-variable could be significant through its correlation with the stronger one. The analysis might then depend on whether the two x-variables are strongly or weakly correlated.[4]

The mixing condition in Assumption 1(ii) ensures weak dependence in the data. Importantly, we do not impose stationarity and allow for heteroskedasticity. Assumption 1(iii) characterizes the strength of the predictors through the order of magnitude of their coefficients in the forecasting model, $\beta_j = c_{\beta,j} n^{-\alpha_{x,j}}$. Here, β_j is close to zero which is consistent with a setting in which the forecast errors are highly correlated with one another. The smaller the value $\alpha_{x,j}$, the stronger the predictor, with $\alpha_{x,j} = 0$ representing the conventional case with a strong predictor whose presence can be detected with certainty as the sample size increases, while $\alpha_{x,j} = 1/2$ represents the local-to-zero case with a weaker predictor whose importance is much harder to detect. Without loss of generality, we assume that the predictor in the second model is stronger than the predictor in the first model ($\alpha_{x,2} < \alpha_{x,1}$). Further, we assume that the dominant predictor is stronger than local-to-zero ($\alpha_{x,2} < 1/2$). Studying sequences of parameter values whose magnitude declines as the sample size grows bigger ensures that parameter uncertainty is preserved asymptotically. In contrast, with a fixed alternative ($\alpha_{x,2} = 0$), uncertainty about the parameter estimates disappears asymptotically. Note that we do not consider the knife edge case of $\alpha_{x,1} = \alpha_{x,2}$ and

[4]The forecasting models considered in our analysis do not include a core set of common predictors which is often relevant in empirical applications. However, one can consider extending the analysis in the following way. Suppose there is a common predictor, x_0, that is orthogonal to both x_1 and x_2. Then one can first project out x_0 and treat the residual as y and apply the current framework. Conversely, if x_0 is not orthogonal to x_1 and x_2, one can rotate the x-variables: project x_1 and x_2 onto x_0 and treat the residuals as the new x_1 and x_2; since the new x_1 and x_2 are now orthogonal to x_0, we can again project y onto x_0 and take the residual as y and proceed within the current setup.

$c_{\beta,1} \neq c_{\beta,2}$. Estimation error has a significant impact on the forecasting performance of our dynamic rotation scheme in this case which leads to a much more complicated technical analysis. Finally, note that we do not require exponential-type tails, which are routinely imposed in papers that handle uniformly valid bounds; see e.g., Fan et al. (2011) and Bonhomme and Manresa (2015).

Our results use an estimation window that contains the most recent n observations, and we assume that n/T (where T is the total sample size) is bounded above zero. Hence, while our results hold for particular values of n, m, and T, in an asymptotic setting Assumption 1(iv) requires that the length of the estimation window grows in line with (or faster than) T. Our estimation scheme is therefore, in an asymptotic setting, equivalent to the rolling estimator used in McCracken (2000) while it differs from the expanding estimation scheme adopted in studies such as McCracken (2000) and Inoue and Kilian (2005) which uses all available data.

The following result provides finite-sample properties of the (approximate) expected squared error performance of the two prediction models:

Proposition 1. *Consider the data generating process* $y_{t+1} = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \varepsilon_{t+1}$. *Moreover, assume that the parameters of the forecasting models,* $\hat{\beta}_{j,n,t}$, *are estimated using* n *observations. Then, under squared error loss and Assumption 1,*

(1) there exist constants $G_1, G_2 > 0$ *and an array of random variables* $\{\Delta L_{t+1,*}\}_{t=n}^{T-1}$ *such that for* $T \geq G_1$,

$$\mathbb{P}\left(\bigcap_{t=n}^{T-1} \{\Delta L_{t+1,*} = \Delta L_{t+1}\}\right) \geq 1 - G_2 T^{\max\{1-r/8,\ 1+(\alpha_{x,2}-1/2)(2+r/4)\}}.$$

(2) there exist constants $G_3, G_4 > 0$ *such that, for* $n \leq t \leq T-1$,

$$G_3 T^{-2\alpha_{x,2}} \leq \mathbb{E}\Delta L_{t+1,*} \leq G_4 T^{-2\alpha_{x,2}}.$$

Part 1 of Proposition 1 is a coupling result that allows us to study the behavior of ΔL_{t+1}. Since ΔL_{t+1} might not have bounded moments for finite n, we consider $\{\Delta L_{t+1,*}\}$, which coincides with $\{\Delta L_{t+1}\}$ with high probability. If $T^{\max\{1-r/8,\ 1+(\alpha_{x,2}-1/2)(2+r/4)\}}$ vanishes, i.e., $r > \max\{8,\ 16\alpha_{x,2}/(1-2\alpha_{x,2})\}$, tests computed based on $\{\Delta L_{t+1}\}_{t=n+m}^{n+m+p-1}$ have the same asymptotic properties as those computed using $\{\Delta L_{t+1,*}\}_{t=n+m}^{n+m+p-1}$.

If the data have more than $\max\{8,\ 16\alpha_{x,2}/(1-2\alpha_{x,2})\}$ moments, the bounds in Proposition 1 are sharp in the sense that they establish the equivalence of ΔL_{t+1} and a random variable whose expectation exists. Conversely, if the data have fewer moments, such bounds do not hold, and results are much harder to obtain.

Proposition 1 is stated in a technical way for the following reason. The ideal result would be to directly state $G_3 T^{-2\alpha_{x,2}} \leq \mathbb{E}\Delta L_{t+1} \leq G_4 T^{-2\alpha_{x,2}}$ without reference to $\Delta L_{t+1,*}$. Unfortunately, $\mathbb{E}\Delta L_{t+1}$ might not exist due to the effect of estimation error. For example, the estimate $\hat{\beta}_{j,t}$ contains $\left(\sum_{s=t-n}^{t-1} x_{j,s}^2\right)^{-1}$, which might not be integrable; $\sum_{s=t-n}^{t-1} x_{j,s}^2$ can be arbitrarily close to zero with a small probability such that the expectation of $\left(\sum_{s=t-n}^{t-1} x_{j,s}^2\right)^{-1}$ is infinity. Therefore, in order to make the argument mathematically correct, we compute $\mathbb{E}\Delta L_{t+1,*}$, rather than $\mathbb{E}\Delta L_{t+1}$, where $\Delta L_{t+1,*}$ is almost always equal to ΔL_{t+1}.

Proposition 1 establishes bounds on the amount by which Model 2 is expected to dominate Model 1 provided that Assumption 1 holds so that the predictor of Model 2 is stronger than local to zero ($\alpha_{x,2} < 1/2$) and more powerful than the predictor of Model 1 ($\alpha_{x,2} < \alpha_{x,1}$). The expected MSE difference depends on the strength of the predictive signals, as a stronger predictor (a smaller $\alpha_{x,2}$) is associated with a larger expected gain from using forecasts from model 2 rather than forecasts from model 1.

Before turning to the performance of the dynamic rotation strategy, we show that, in the present setting, a simple forecast combination strategy $\bar{y}_{t+1|t}^C = \lambda \hat{y}_{1,t+1|t} + (1-\lambda)\hat{y}_{2,t+1|t}$ is dominated in MSE terms by forecasts from the second model.

Proposition 2. *Consider the setting in Proposition 1. Let* $\Delta L_{t+1}^C = (y_{t+1} - \bar{y}_{t+1|t}^C)^2 - (y_{t+1} - \hat{y}_{2,t+1|t})^2$, *where* $\bar{y}_{t+1|t}^C = \lambda \hat{y}_{1,t+1|t} + (1-\lambda)\hat{y}_{2,t+1|t}$ *for* $\lambda \in (0,1)$. *Then*

(1) there exist constants $G_1, G_2 > 0$ *and an array of random variables* $\{\Delta L_{t+1,*}^C\}_{t=n}^{T-1}$ *such that for* $T \geq G_1$,

$$\mathbb{P}\left(\bigcap_{t=n}^{T-1} \left\{\Delta L_{t+1,*}^C = \Delta L_{t+1}^C\right\}\right) \geq 1 - G_2 T^{\max\{1-r/8,\ 1+(\alpha_{x,2}-1/2)(2+r/4)\}}.$$

(2) there exist constants $G_3, G_4 > 0$ *such that, for* $n \leq t \leq T-1$,

$$G_3 T^{-2\alpha_{x,2}} \leq \mathbb{E}\Delta L_{t+1,*}^C \leq G_4 T^{-2\alpha_{x,2}}.$$

Proposition 2 shows that a combination of the two forecasts can be expected to produce higher expected squared errors than model 2. It follows that, to the extent that dynamic rotation is capable of performing better than model 2, it is also expected to produce more accurate forecasts than the combination.

3.3 Expected Gains from Dynamic Rotation

We next analyze the expected gains from dynamic rotation between forecasts. We assume that the parameters of the forecast monitoring regression in (2.3) are estimated using a monitoring window based on the most recent m observations

$$\hat{\theta}_{m,t} = \left(m^{-1} \sum_{s=t-m}^{t-1} z_s z_s' \right)^{-1} \left(m^{-1} \sum_{s=t-m}^{t-1} z_s \Delta L_{s+1} \right), \tag{3.4}$$

where $z_s = (1, z_{Ms})'$. The dynamic rotation rule chooses model 2 if and only if $z_t'\hat{\theta}_{m,t} > 0$. Using (2.4), the dynamic rotation forecasts $\{\hat{y}_{DR,t+1}\}_{t=n+m}^{T-1}$ take the form

$$\hat{y}_{DR,t+1|t} = \hat{y}_{1,t+1|t} \mathbf{1}\{z_t'\hat{\theta}_{m,t} \leq 0\} + \hat{y}_{2,t+1|t} \mathbf{1}\{z_t'\hat{\theta}_{m,t} > 0\}. \tag{3.5}$$

Under dynamic rotation, the sample size, $T = n + m + p$, is split into estimation windows of length n (used to estimate $\hat{\beta}_{j,n,t}$) and m (used for $\hat{\theta}_{m,t}$), respectively, and a test sample with the remaining p observations.

To establish results on the expected gains from monitoring forecasting performance and using the dynamic rotation rule (3.5), we need to make assumptions about the correlation between the monitoring instrument, z_{Mt}, and the $x_{i,t}\varepsilon_{t+1}$ terms in the forecast errors. We collect these in Assumption 2:

Assumption 2. *The following hold for* $j \in \{1, 2\}$*:*

(i) $\{x_{j,t}, z_{Mt}, \varepsilon_{t+1}\}_{t=-\infty}^{\infty}$ *is a beta-mixing array with coefficient* $\mathcal{B}_{mix}(\cdot)$ *such that* $\forall t > 0$, $\mathcal{B}_{mix}(t) \leq b \exp(-t^c)$*, for constants* $b, c > 0$*.*

(ii) There exist constants $\alpha_{z,1}, \alpha_{z,2} \in [0, \infty]$, $c_{\rho,1}, c_{\rho,2} > 0$ *such that* $corr(x_{j,t}\varepsilon_{t+1}, z_{Mt}) = c_{\rho,j} m^{-\alpha_{z,j}}$*, where*

(a) $2r\alpha_{z,2}/(r-2) < \alpha_{x,2}$*,*

(b) $\alpha_{x,2} + \alpha_{z,2} < \alpha_{x,1} + \alpha_{z,1}$*.*

(iii) For some constants $\kappa_1, \kappa_2 > 0$, $\kappa_1 \mathbb{E} x_{2,t}\varepsilon_{t+1} z_{Mt} \leq \mathbb{E} x_{2,t}\varepsilon_{t+1}\mathbf{1}\{z_{Mt} > 0\} \leq \kappa_2 \mathbb{E} x_{2,t}\varepsilon_{t+1} z_{Mt}$.

(iv) On some fixed neighborhood of zero, the p.d.f. of z_{Mt} *is uniformly bounded.*

(v) $\mathbb{E} z_{Mt} = 0$. *Moreover,* $\mathbb{E}|z_{Mt}|^r \leq D$ *for some* $D > 0$.

(vi) T/m *is bounded.*

The mixing condition in Assumption 2(i) ensures weak dependence in the predictors, monitoring instrument, and outcomes and naturally extends Assumption 1(i). Assumption 2(ii) ensures that the monitoring instrument, z_{Mt}, is not too weak for the second model (part a) and that the "combined" strength of the predictor and monitoring instrument ($\alpha_{x,2} + \alpha_{z,2}$) is stronger for model 2 than for model 1 (part b). Assumption 2(iii) links the selection rule and the correlation between $x_{2,t}\varepsilon_{t+1}$ and z_{Mt}. The condition says that the correlation between $x_{2,t}\varepsilon_{t+1}$ and z_{Mt} is of the same order of magnitude as the correlation between $x_{2,t}\varepsilon_{t+1}$ and $\mathbf{1}\{z_{Mt} > 0\}$. This means that the dependence between $x_{2,t}\varepsilon_{t+1}$ and z_{Mt} can be measured in approximately equivalent ways, either by the correlation between $x_{2,t}\varepsilon_{t+1}$ and z_{Mt} or by the correlation between $x_{2,t}\varepsilon_{t+1}$ and $\mathbf{1}\{z_{Mt} > 0\}$. Assumptions 2(iv)-(v) impose mild assumptions on the distribution and moments of z_{Mt}. By a fixed neighborhood in part (iv), we mean that this does not depend on n, m, and T, so that asymptotically this neighborhood does not collapse to a point or something with measure zero. Finally, we assume in part (vi) that, in an asymptotic setting, the length of the monitoring window, m, grows in proportion with the sample size, T.

In practice, one might wish to test the validity of these conditions and only apply the proposed rotation method when the conditions can be shown to hold. Constructing such tests is not an easy task, however, and fortunately may not be necessary in practice. Suppose we simply always apply dynamic rotation without testing if the conditions hold. If the conditions are valid, conditional rotation has guaranteed theoretical benefits as we show below. Conversely, if the conditions do not hold, then either the two forecasts are not very different or the instrument z_{Mt} is not sufficiently informative to differentiate between the two forecasts. In the latter case the dynamic rotation is not expected to vastly underperform the benchmark of simply using one of the forecasts.

Turning to the existence of instruments that are useful for monitoring, it is entirely possible that some variables possess predictive power themselves while also, at the same time, serving as valuable monitoring instruments. As an example, suppose that $y_{t+1} = z_t'\beta + \exp(z_t'\gamma)\varepsilon_{t+1}$, where ε_{t+1} is independent of z_t with $E\varepsilon_{t+1} = 0$ and $E\varepsilon_{t+1}^2 = 1$. Clearly, the instrument z_t can be used for forecasting. However, because the squared-error loss for the forecast $z_t'\beta$ is $(y_{t+1} - z_t'\beta)^2 = \exp(2z_t'\gamma)\varepsilon_{t+1}^2$, z_t is also useful in monitoring the performance of the forecast. This example illustrates that being a valid monitoring instrument does not preclude a variable from possessing predictive power over outcomes, thus widening the set of candidate variables that can be used as instruments.

We next characterize the expected gain from monitoring forecasting performance and using the dynamic rotation rule (3.5) rather than either models 1 or 2:

Proposition 3. *Consider the data generating process* $y_{t+1} = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \varepsilon_{t+1}$. *Moreover, assume that the parameters of the forecasting models,* $\hat{\beta}_{j,n,t}$*,* $(j = 1, 2)$ *and of the monitoring rule,* $\hat{\theta}_{m,t}$*, are estimated using* n *and* m *observations, respectively. Then, assuming squared error loss, under Assumptions 1 and 2, the following hold:*

(1) There exist constants $G_1, G_2, G_3 > 0$ *and an array* $\{S_{t+1}\}_{t=n+m}^{T-1}$ *such that for* $T \geq G_1$ *and* $n + m \leq t \leq T - 1$

$$\mathbb{P}\left(\bigcap_{t=n+m}^{T-1} \left\{S_{t+1} = \Delta L_{t+1}\mathbf{1}\{z_t'\hat{\theta}_{m,t} > 0\}\right\}\right) \geq 1 - G_2 T^{\max\{1-r/8,\ 1+(\alpha_{x,2}-1/2)(2+r/4)\}}$$

and

$$\mathbb{E}S_{t+1} \geq G_3 T^{-(\alpha_{x,2}+\alpha_{z,2})}.$$

(2) There exist constants $G_4, G_5, G_6 > 0$ *and an array* $\{\tilde{S}_{t+1}\}_{t=n+m}^{T-1}$ *such that for* $T \geq G_4$ *and* $n + m \leq t \leq T - 1$

$$\mathbb{P}\left(\bigcap_{t=n+m}^{T-1} \left\{\tilde{S}_{t+1} = -\Delta L_{t+1}\mathbf{1}\{z_t'\hat{\theta}_{m,t} < 0\}\right\}\right) \geq 1 - G_5 T^{\max\{1-r/8,\ 1+(\alpha_{x,2}-1/2)(2+r/4)\}}$$

and

$$\mathbb{E}\tilde{S}_{t+1} \geq G_6 T^{-(\alpha_{x,2}+\alpha_{z,2})}.$$

To interpret part 1 of Proposition 3, notice that $(y_{t+1} - \hat{y}_{1,t+1|t})^2 - (y_{t+1} -$

$\hat{y}_{DR,t+1|t})^2 = \Delta L_{t+1}\mathbf{1}\{z_t'\hat{\theta}_{m,t} > 0\}$ so S_{t+1} captures the expected squared error loss of model 1 relative to the dynamic rotation rule, provided that $T^{\max\{1-r/8,\ 1+(\alpha_{x,2}-1/2)(2+r/4)\}}$ is small. When this holds, tests computed based on $\{\Delta L_{t+1}\mathbf{1}\{z_t'\hat{\theta}_{m,t} > 0\}\}$ have the same asymptotic properties as those computed using S_{t+1}.

Part 1 of Proposition 3 shows that the expected benefit from dynamic rotation rather than using the first model is bounded below by a positive sequence of order $T^{-(\alpha_{x,2}+\alpha_{z,2})}$. Hence, the more accurate the predictor variable of model 2 (smaller $\alpha_{x,2}$) and the better the monitoring instrument (smaller $\alpha_{z,2}$), the bigger the expected gain from dynamic rotation relative to model 1.

Part 2 of Proposition 3 computes the expected gain from dynamic rotation versus always using model 2. To see this, notice that $(y_{t+1} - \hat{y}_{2,t+1|t})^2 - (y_{t+1} - \hat{y}_{DR,t+1|t})^2 = -\Delta L_{t+1}\mathbf{1}\{z_t'\hat{\theta}_{m,t} < 0\}$, so $\tilde{S}_{t+1}$ captures the squared error loss of model 2 relative to the rotation rule with a high probability if $G_5 T^{\max\{1-r/8,\ 1+(\alpha_{x,2}-1/2)(2+r/4)\}}$ is small. Hence, to a good approximation, the expected loss of dynamic rotation relative to the second model is also bounded below by a positive sequence of order $T^{-(\alpha_{x,2}+\alpha_{z,2})}$.

An important condition for dynamic rotation (3.5) to work is that $\mathbb{E}[x_{jt}\varepsilon_{t+1}z_{Mt}] \neq 0$, so the monitoring instrument, z_{Mt}, is capable of picking up predictable forecast errors. This condition can hold even if $\mathbb{E}[\varepsilon_{t+1}z_{Mt}] = 0$. Hence, the instrument need not have any predictive power if added to the forecasting model on its own. Monitoring instruments can therefore be useful for tracking the (relative) performance of a particular forecast even though they need not have predictive power over the outcome when added as predictors. Of course, adding the cross-product term $x_{jt}z_{Mt}$ as a predictor to the original forecasting model might produce better results. However, this strategy is often not a feasible option since x_{jt} might not be observed, as in the case of survey data or any third-party forecasts that are not generated by the forecast user. Moreover, this strategy tends to increase the effect of estimation error which can lead to reduced accuracy of the forecasts, as we show in the empirical analysis.

A cost to implementing the rotation scheme is the requirement of an estimation window with m observations to estimate the parameters $\hat{\theta}_{m,t}$ of the rotation rule.[5]

[5] This is not an issue if the forecasts are given, as in the case of survey data. However, if the forecasts are model-generated and require estimation of parameters, the requirement of a monitoring window of length m can effectively reduce the number of observations available for estimation of $\hat{\beta}_{j,t}$. Of course, this does not matter if $m + n$ is smaller than the number of data points available at the desired starting point of the test sample.

In practice, of course, the monitoring procedure is not the only option and when the number of data points available is too small to reasonably implement the monitoring procedure, one can alternatively stick with one model or use forecast combination. Once the number of data points becomes large enough, one can then start the monitoring procedure.

3.4 Weak Predictor with a Strong Monitoring Instrument

Proposition 3 establishes results for dynamic rotation under conditions that the (joint) signal in the predictor and monitoring instrument is more powerful for model 2 than for model 1 through the assumptions $\alpha_{x,2} < \alpha_{x,1}$ and $\alpha_{x,2} + \alpha_{z,2} < \alpha_{x,1} + \alpha_{z,1}$. In this case, the dominant term in the monitoring rule is the correlation between z_{Mt} and $2\beta_2\varepsilon_{t+1}x_{2,t}$.

However, suppose that model 2 uses the strongest predictor ($\alpha_{x,2} < \alpha_{x,1}$) but that the monitoring instrument is stronger for model 1 and that $\alpha_{x,2}+\alpha_{z,2} > \alpha_{x,1}+\alpha_{z,1}$. In this case, the dominant term in the monitoring rule becomes the correlation between z_{Mt} and $2\beta_1\varepsilon_{t+1}x_{1,t}$. We next show that it is possible to generate gains from dynamic rotation also in this case. We capture the case with a weak predictor and a strong monitoring instrument through the following assumption:

Assumption 3. *Let Assumption 2 (i), (iii)-(vi) hold for some* $r \geq 10$*, but replace Assumption 2(ii) with the assumption that there exist constants* $\alpha_{z,1}, \alpha_{z,2} \in [0, \infty]$, $c_{\rho,1}, c_{\rho,2} > 0$ *such that* $corr(x_{j,t}\varepsilon_{t+1}, z_{Mt}) = c_{\rho,j}m^{-\alpha_{z,j}}$ *and*

1. $\alpha_{z,1} < \alpha_{x,1}$,

2. $\alpha_{x,1} + \alpha_{z,1} < \min\{1/2,\ (3r-2)\alpha_{x,2}/(2r-2)\}$, *and*

3. $\alpha_{x,2} + \alpha_{z,2} > \alpha_{x,1} + \alpha_{z,1}$.

Note the parameter restrictions for this case. We require that the monitoring instrument be more strongly correlated with the cross-product $x_{1,t}\varepsilon_{t+1}$ than the correlation between the "weak" predictor and the outcome ($\alpha_{z,1} < \alpha_{x,1}$), at least for large T. We also require that $\alpha_{x,1} + \alpha_{z,1} < 1/2$, although this bound could be tighter, depending on the values of r and $\alpha_{x,2}$. The last part of Assumption 3 captures that the combined strength of the predictor and monitoring instrument for model 1 exceeds that for model 2. Assumption 3 slightly tightens the restrictions on r by requiring

$r \geq 10$ instead of $r > 8$. We view Assumption 3 as a further restriction on $\alpha_{x,j}$ and $\alpha_{z,j}$.

Proposition 4. *Suppose that Assumptions 1 and 3 are satisfied. Then, the following hold:*

(1) There exist constants $G_1, G_2, G_3 > 0$ and an array $\{S_{t+1}\}_{t=n+m}^{T-1}$ such that for $T \geq G_1$ and $n + m \leq t \leq T - 1$

$$\mathbb{P}\left(\bigcap_{t=n+m}^{T-1} \left\{ S_{t+1} = \Delta L_{t+1} \mathbf{1}\{z_t' \hat{\theta}_{m,t} > 0\} \right\} \right) \geq 1 - G_2 T^{\max\{1-r/8,\ 1+(\alpha_{x,2}-1/2)(2+r/4)\}}$$

and

$$\mathbb{E} S_{t+1} \geq G_3 T^{-\alpha_{x,1}-\alpha_{z,1}}.$$

(2) There exist constants $G_4, G_5, G_6 > 0$ and an array $\{\tilde{S}_{t+1}\}_{t=n+m}^{T-1}$ such that for $T \geq G_4$ and $n + m \leq t \leq T - 1$

$$\mathbb{P}\left(\bigcap_{t=n+m}^{T-1} \left\{ \tilde{S}_{t+1} = -\Delta L_{t+1} \mathbf{1}\{z_t' \hat{\theta}_{m,t} < 0\} \right\} \right) \geq 1 - G_5 T^{\max\{1-r/8,\ 1+(\alpha_{x,2}-1/2)(2+r/4)\}}$$

and

$$\mathbb{E} \tilde{S}_{t+1} \geq G_6 T^{-(\alpha_{x,1}+\alpha_{z,1})}.$$

Proposition 4 shows that the expected gain from dynamic rotation, relative to either always using the forecasts from model 1 or always using the forecasts from model 2, is bounded from below by terms that are of order $T^{-(\alpha_{x,1}+\alpha_{z,1})}$, which, by assumption, is bigger than $T^{-1/2}$. Together with the result in Proposition 3, this shows that there can be expected gains from dynamic rotation in cases with (i) a strong predictor and a strong monitoring instrument (Proposition 3) or (ii) a weak predictor but a monitoring instrument that is strongly correlated with the cross-product of the weak predictor and the residual from the forecasting model (Proposition 4).

4 Nested Models

Comparisons of forecasts from nested models arise in a number of applications in economics and finance and this case can be addressed by modifying the analysis in

Section 3. Suppose the DGP includes an intercept and a time-varying regressor:

$$y_{t+1} = \mu + \beta x_t + \varepsilon_{t+1}, \tag{4.1}$$

Moreover, suppose that model 2 (the "big" model) coincides with the DGP in (4.1), while model 1 is a (nested) small model that only includes an intercept:

$$y_{t+1} = \mu + \varepsilon_{t+1}. \tag{4.2}$$

Both models are estimated using OLS so that, for $n \leq t \leq T$,

$$\begin{aligned} \tilde{\mu}_{n,t} &= n^{-1} \sum_{s=t-n+1}^{t} y_s, \\ \begin{pmatrix} \hat{\mu}_{n,t} \\ \hat{\beta}_{n,t} \end{pmatrix} &= \left[n^{-1} \sum_{s=t-n}^{t-1} \begin{pmatrix} 1 \\ x_s \end{pmatrix} \begin{pmatrix} 1 & x_s \end{pmatrix} \right]^{-1} \left[n^{-1} \sum_{s=t-n}^{t-1} \begin{pmatrix} 1 \\ x_s \end{pmatrix} y_{s+1} \right], \end{aligned} \tag{4.3}$$

and $\hat{y}_{1,t+1|t} = \tilde{\mu}_{n,t}$, while $\hat{y}_{2,t+1|t} = \hat{\mu}_{n,t} + \hat{\beta}_{n,t} x_t$.

Using these notations, the difference in squared error loss is

$$\begin{aligned} \Delta L_{t+1} &= (y_{t+1} - \tilde{\mu}_{n,t})^2 - (y_{t+1} - \hat{\mu}_{n,t} - \hat{\beta}_{n,t} x_t)^2 \\ &= (-\delta_{small,n,t} + \beta x_t + \varepsilon_{t+1})^2 - (-\delta_{big,n,t} + \varepsilon_{t+1})^2 \\ &= \beta^2 x_t^2 + 2\beta x_t \varepsilon_{t+1} + \delta_{small,n,t}^2 - \delta_{big,n,t}^2 + 2\delta_{big,n,t} \varepsilon_{t+1} - 2\delta_{small,n,t}(\beta x_t + \varepsilon_{t+1}), \end{aligned} \tag{4.4}$$

where $\delta_{small,n,t} = \tilde{\mu}_{n,t} - \mu$ and $\delta_{big,n,t} = \hat{\mu}_{n,t} - \mu + (\hat{\beta}_{n,t} - \beta) x_t$.

Using the earlier notations, we can capture this case by setting $\alpha_{x,1} = \alpha_{z,1} = \infty$, so that $\beta_1 = \delta_{1,n,t} = 0$. This allows us to simplify the notations by setting $\beta_2 = \beta = cn^{-\alpha_x}$ and $corr(x_t \varepsilon_{t+1}, z_{Mt}) = c_\rho n^{-\alpha_z}$, where $c, c_\rho > 0$ and $\alpha_x, \alpha_z \geq 0$ are constants. Moreover, $\delta_{2,n,t} = \delta_{n,t}$, where $\delta_{n,t} = (\sum_{s=t-n}^{t-1} x_s^2)^{-1} (\sum_{s=t-n}^{t-1} x_s \varepsilon_{s+1})$.

Note a subtle difference between the nested and non-nested case: In the nested case, we impose on the small model that the parameter of the additional predictor that is only included in the big model takes a value of zero so that fewer parameters are estimated by the small model. Conversely, in the non-nested case, no such constraint is imposed and so we do not have a "big" and a "small" model.

4.1 Performance of Big and Small Forecasting Models

We summarize our assumptions for the case with nested models in Assumption 4:

Assumption 4. *Assume that the following hold*

(i) The r-th moments of x_t, z_{Mt} and ε_{t+1} are uniformly bounded for some constant $r > 8$.

(ii) $\{x_t, z_{Mt}, \varepsilon_t\}_{t=-\infty}^{\infty}$ is a beta-mixing array with coefficient $\mathcal{B}_{mix}(\cdot)$ such that $\forall t > 0$, $\mathcal{B}_{mix}(t) \leq b\exp(-t^c)$, for constants $b, c > 0$.

(iii) $\mathbb{E}(\varepsilon_{t+1} \mid \{(x_s, \varepsilon_s)\}_{s=-\infty}^{t}) = 0$ and $\mathbb{E}x_t = \mathbb{E}z_{Mt} = 0$.

(iv) $\beta = cn^{-\alpha_x}$ for some constants $\alpha_x \in [0, \infty)$, $c > 0$.

(iv) $M_1 \leq \mathbb{E}x_t^2 \leq M_2$ for some constants $M_1, M_2 > 0$.

(v) T/n and T/m are bounded.

Using this assumption, we can characterize the expected squared error loss performance of the small versus the big models for the nested case:

Proposition 5. *Consider the data generating process $y_{t+1} = \mu + x_t\beta + \varepsilon_{t+1}$ and suppose that Assumption 4 holds.*

(1) *Suppose that $\alpha_x < 1/2$. Then there exist constants $C_1, \cdots, C_4 > 0$ and an array $\{\Delta L_{t+1,*}\}_{t=n}^{T-1}$ such that for $T \geq C_1$*

$$\mathbb{P}\left(\bigcap_{t=n+m}^{T-1} \{\Delta L_{t+1} = \Delta L_{t+1,*}\}\right) \geq 1 - C_2 T^{\max\{1-r/8,\ 1+(\alpha_x-1/2)(2+r/4)\}}$$

and

$$C_3 T^{-2\alpha_x} \leq \mathbb{E}\Delta L_{t+1,*} \leq C_4 T^{-2\alpha_x} \quad \forall n \leq t \leq T-1.$$

(2) *Suppose that $\alpha_x > 1/2$. Then there exist constants $C_5, \cdots, C_8 > 0$ and an array of random variables $\{\Delta L_{t+1,*}\}_{t=n}^{T-1}$ such that for $T \geq C_5$*

$$\mathbb{P}\left(\bigcap_{t=n+m}^{T-1} \{\Delta L_{t+1} = \Delta L_{t+1,*}\}\right) \geq 1 - C_6 T^{\max\{1-r/8,\ 1+(\alpha_x-1/2)(2+r/4)\}}$$

and

$$-C_7 T^{-1} \leq \mathbb{E}\Delta L_{t+1,*} \leq -C_8 T^{-1} \quad \forall n \leq t \leq T-1.$$

Part 1 of Proposition 5 shows that the expected squared error loss of the big model is smaller than that of the small model provided that the strength of the extra predictor included in the big model is sufficiently large to overcome the effect of estimation error ($\alpha_x < 1/2$). Moreover, the amount by which the big model is expected to outperform the small model gets bigger, the stronger the predictive signal, i.e., the smaller is α_x. Conversely, part 2 of Proposition 5 says that if the predictive signal underlying the big model is too weak ($\alpha_x > 1/2$), the estimation error of the big model dominates the signal, leading us to expect that the big model will underperform the small model, although the expected underperformance is only of order $O(T^{-1})$.

4.2 Expected Gains from Dynamic Rotation

We next characterize the expected gains from dynamic rotation for the nested case. For this analysis we make use of the following assumption:

Assumption 5. *The following hold*

(i) There exist constants $\alpha_z \in [0, \infty]$, $c_\rho > 0$ *such that* $corr(x_t\varepsilon_{t+1}, z_{Mt}) = c_\rho m^{-\alpha_z}$, *where* $2r\alpha_z/(r-2) < \alpha_x$.

(ii) For some constants $\kappa_1, \kappa_2 > 0$, $\kappa_1 \mathbb{E}x_t\varepsilon_{t+1}z_{Mt} \leq \mathbb{E}x_t\varepsilon_{t+1}\mathbf{1}\{z_{Mt} > 0\} \leq \kappa_2 \mathbb{E}x_t\varepsilon_{t+1}z_{Mt}$.

(iii) On some fixed neighborhood of zero, the p.d.f. of z_{Mt} *is uniformly bounded.*

Using Assumption 5, we have the following result for dynamic rotation:

Proposition 6. *Consider the data generating process* $y_{t+1} = \mu + x_t\beta + \varepsilon_{t+1}$. *Suppose Assumptions 4 and 5 hold. Then*

(1) there exist constants $M_1, M_2, M_3 > 0$ *and an array* $\{S_{t+1}\}_{t=n+m}^{T-1}$ *such that for* $T \geq M_1$ *and* $\forall n+m \leq t \leq T-1$,

$$\mathbb{P}\left(\bigcap_{t=n+m}^{n+m+p-1}\left\{S_{t+1} = \Delta L_{t+1}\mathbf{1}\{z_t'\hat{\theta}_{m,t} > 0\}\right\}\right) \geq 1 - M_2 T^{\max\{1-r/8,\ 1+(\alpha_x-1/2)(2+r/4)\}}$$

and

$$\mathbb{E}S_{t+1} \geq M_3 T^{-(\alpha_x+\alpha_z)}.$$

(2) there exist constants $M_4, M_5, M_6 > 0$ *and an array* $\{\tilde{S}_{t+1}\}_{t=n+m}^{T-1}$ *such that for* $T \geq M_4$ *and* $\forall n+m \leq t \leq T-1$,

$$\mathbb{P}\left(\bigcap_{t=n+m}^{n+m+p-1}\left\{\tilde{S}_{t+1} = -\Delta L_{t+1}\mathbf{1}\{z_t'\hat{\theta}_{m,t} < 0\}\right\}\right) \geq 1 - M_5 T^{\max\{1-r/8,\ 1+(\alpha_x-1/2)(2+r/4)\}}$$

and

$$\mathbb{E}\tilde{S}_{t+1} \geq M_6 T^{-(\alpha_x+\alpha_z)}.$$

Part 1 of Proposition 6 shows that dynamic rotation is expected to perform better than the small model by an amount bounded by a factor of order $T^{-(\alpha_x+\alpha_z)}$. A similar result holds for the amount by which dynamic rotation is expected to outperform the big model (Part 2).

5 Simulation Results

This section presents results from a set of Monte Carlo simulations which quantify the gains from dynamic rotation. For the nested case we show the joint effects of varying the strength of the predictor and the monitoring instrument on the predictive performance of (i) a small forecasting model; (ii) a big forecasting model; and (iii) dynamic rotation. We also consider alternative forecasting methods based on augmenting the forecasting model with the monitoring instrument, a pre-test for determining whether to include a predictor, and an equal-weighted forecast combination.

For each point in time $t \geq m+n+1$, define the estimator for a model that includes x_{it} as a predictor

$$\hat{\beta}_{i,n,t} = \left[\sum_{s=t-n}^{t-1} x_{is}x_{is}\right]^{-1}\left[\sum_{s=t-n}^{t-1} x_{is}y_{s+1}\right],$$

with resulting forecast $\hat{y}_{i,t+1|t} = \hat{\beta}_{i,n,t}x_t$. The squared error loss differential of models 1 versus 2 is given by

$$\Delta L_{t+1} = (y_{t+1} - \hat{\beta}_{1,n,t}x_{1t})^2 - (y_{t+1} - \hat{\beta}_{2,n,t}x_{2t})^2. \quad (5.1)$$

To evaluate the dynamic rotation rule, for $t \geq m+n+1$, define the estimates from regressing ΔL_{t+1} on $(1, z_{Mt})$:

$$(\hat{\theta}_{0,t}, \hat{\theta}_{1,t})' = \left[\sum_{s=t-m}^{t-1} (1, z_{1s})(1, z_{1s})' \right]^{-1} \left[\sum_{s=t-m}^{t-1} (1, z_{1s})' \Delta L_{s+1} \right]$$

and the associated conditional forecast of the loss differential

$$\widehat{\Delta L}_{t+1|t} = \hat{\theta}_{0,t} + \hat{\theta}'_{1,t} z_t. \tag{5.2}$$

Forecasts from dynamic rotation take the form

$$\hat{y}_{DR,t+1|t} = \mathbf{1}\{\widehat{\Delta L}_{t+1|t} \leq 0\} \hat{\beta}_{1,n,t} x_{1t} + \mathbf{1}\{\widehat{\Delta L}_{t+1|t} > 0\} \hat{\beta}_{2,n,t} x_{2t}.$$

For each simulated sample we compute the mean squared errors of the two forecasts and for dynamic rotation as $MSE^j = p^{-1} \sum_{t=m+n+1}^{T-1} (y_{t+1} - x_{jt} \hat{\beta}_{j,n,t})^2$, and $MS\mathbb{E}^{DR} = p^{-1} \sum_{t=m+n+1}^{T-1} (y_{t+1} - \hat{y}_{DR,t+1|t})^2$, where $T = n + m + p$. We report results in the form of MSE ratios.

In principle, our bounds on out-of-sample MSE values can be explored to conduct hypothesis testing using either the finite-sample framework of Giacomini and White (2006) or the asymptotic approach of West (1996) and Clark and McCracken (2001). However, the MSE values map directly to our bounds whereas hypothesis tests and formal inference results depend on choices of test statistics, bandwidths used to compute standard errors, and the power of the tests. Our analysis has little to say about these issues so we choose to report the simulation results in the form of ratios of MSE values.

5.1 Nested case

In the nested case, data are generated from a simple linear regression model

$$y_{t+1} = \beta x_t + \varepsilon_{t+1}, \tag{5.3}$$

where $x_t \sim i.i.d. U(-1, 1)$. The residual ε_{t+1} is generated as follows. Let $s_{t+1} \in \{0, 1\}$ be a binary random variable such that $\mathbb{P}(s_{t+1} = 1 \mid x_t > 0) = \mu + \delta$ and $\mathbb{P}(s_{t+1} = 1 \mid x_t \leq 0) = \mu - \delta$, where $\mu = 1/2$. The two forecasting models are given by (4.1) and (4.2) with μ set to zero. Hence, the small model predicts zero while the big model coincides with the data generating process in (5.3) and so includes the regressor x_t. Define

$$\varepsilon_{t+1} = s_{t+1}Q_{1,t} + (1 - s_{t+1})Q_{2,t}, \tag{5.4}$$

where $Q_{1,t}$ and $Q_{2,t}$ are $N(0,1)$ random variables that are mutually independent and independent of s_{t+1} and x_t. To control the correlation between the residual in (5.3) and the monitoring instrument, z_{Mt}, we generate the latter as

$$z_{Mt} = Q_{1,t} - Q_{2,t}. \tag{5.5}$$

It is now easy to see that $\mathbb{E}x_t\varepsilon_{t+1} = \mathbb{E}z_{Mt}\varepsilon_{t+1} = \mathbb{E}\varepsilon_{t+1} = \mathbb{E}z_{Mt} = \mathbb{E}x_t = 0$ and

$$Corr(x_t\varepsilon_{t+1}, z_{Mt}) = \sqrt{\frac{3}{2}}\delta.$$

Our simulations set $\beta = 3n^{-\alpha_x}$ and we choose δ such that $Corr(x_t\varepsilon_{t+1}, z_{Mt}) = 0.6n^{-\alpha_z}$. We report the outcome of 5,000 simulations based on a sample size $(n, m, p) = (100, 100, 200)$, so that $T = 400$. For each simulation we compute MSE values for alternative forecasting schemes and report ratios of MSE values, averaged across simulations. MSE ratios equal to one suggest that two forecasting methods are equally accurate on average while ratios below (above) unity indicate that the model in the numerator (denominator) produces lower MSE values.

Table 1 presents results from our simulations. First consider the performance of the big versus the small forecasting model (top row) in Panel A. When $\alpha_x = 0$, the big model produces MSE-values that, on average, are about one-quarter the size of those produced by the small model. This is a very large difference in predictive accuracy and happens because the predictor is very strong in this case. However, as α_x increases, the bigger model's performance rapidly declines and the big and small models are broadly equally accurate for $\alpha_x = 0.5$. For larger values of α_x, the small model is marginally more accurate than the big model, consistent with Proposition 5.

Turning to the comparison of the dynamic rotation and the small forecasting model (Panel B), the dynamic rotation rule strongly dominates the small model if the predictor is reasonably strong, i.e., $\alpha_x \leq 0.40$. Moreover, having a strong monitoring instrument can reduce the MSE ratio by about 10% when the predictor is moderately strong ($\alpha_x = 0.25$ or $\alpha_x = 0.40$). Conversely, the benefits from rotating forecasts based on a strong instrument are smaller when the predictor is very strong or very weak. When the predictor is very strong ($\alpha_x = 0$), the benefits from rotating between the large and small models are very small since the large model almost always dominates. In fact, the better average performance of the big model relative to the small model is mostly picked up by the intercept θ_0 in the monitoring regression and so holds independently of the value of θ_1. Conversely, if the predictor is weak ($\alpha_x = 1$), there is very little signal in the forecasts from the big model which reduces the value from having an instrument even when this is highly accurate.

The MSE values of the dynamic rotation scheme are closer to those of the big forecasting model (Panel C) than those from the small forecasting model (Panel B). Relative to the big model, dynamic rotation produces the best performance when the predictor falls in the range $\alpha_x \in [0.1, 0.5]$ and the instrument is strong, i.e., $\alpha_z = 0$ or $\alpha_z = 0.1$. For these scenarios, the predictor is quite accurate and the instrument is sufficiently strong to identify periods where it is beneficial to switch to the small forecasting model. Conversely, if the monitoring instrument is weak and the predictor is quite strong, rotating between the small and big models leads to a loss of 1-2% in MSE performance relative to the big model. In this setting, the big model is far more accurate than the small model, so sticking to the forecasts from the big model becomes very difficult to beat. Even in this case, the loss in predictive accuracy from using the dynamic rotation strategy is very modest, however, compared to the benefits obtainable in the case with a strong monitoring instrument.

Table 2 reports the performance of dynamic rotation measured relative to four widely used forecasting methods. Panel A shows that dynamic rotation dominates an equal-weighted average of the two forecasting models–often by a sizable margin–if either (i) the big model is very good (i.e., α_x is small), regardless of the accuracy of the monitoring instrument; or (ii) if the monitoring instrument is very accurate (small α_z), regardless of the strength of the predictor. Dynamic rotation and equal-weighted forecast combination are equally accurate if both the monitoring instrument and the predictor are quite poor, i.e., if α_x and α_z both exceed 0.5.

Panel B compares the performance of dynamic rotation to that of a pre-test approach that includes x in the forecasting model only if its slope coefficient is statistically significant using a t-test. Dynamic rotation performs better than pre-testing provided that the monitoring instrument is fairly accurate (small α_z), particularly if the predictor is moderately accurate. Moreover, when dynamic rotation loses out to pre-testing, i.e., when α_x is small and α_z is large, it does so only by 1-2% in MSE terms.

Panel C shows that dynamic rotation generally produces more accurate forecasts than those from an augmented model that includes both x_t and the monitoring instrument, z_{Mt}, as predictors and that the gains from rotation tend to be particularly large when the monitoring instrument is accurate (α_z is low) and the predictor variable is moderately strong.

Finally, Panel D considers a forecast combination scheme which lets the combination weights be proportional to the inverse MSE estimates and computes the forecast as $\hat{y}_{1,t+1|t}w_t + \hat{y}_{2,t+1|t}(1 - w_t)$, where

$$w_t = \frac{\widehat{MSE}_{1,t}}{\widehat{MSE}_{1,t} + \widehat{MSE}_{2,t}},$$

with $\widehat{MSE}_{1,t} = m^{-1}\sum_{s=t-m+1}^{t}(y_s - \hat{y}_{1,s|s-1})^2$ and $\widehat{MSE}_{2,t} = m^{-1}\sum_{s=t-m+1}^{t}(y_s - \hat{y}_{2,s|s-1})^2$. The simulation results lead to similar conclusions as those obtained for the equal-weighted combination scheme in Panel A, although differences in relative MSE-performance are smaller for the combination scheme in Panel D. This happens because the weighted forecast combination assigns lower weight to the small model than the equal-weighted combination in scenarios where this model performs particularly poorly.

5.2 Non-nested case

For the non-nested case, let $\{(x_{1,t}, z_{1t,1}, s_{1,t}, \varepsilon_{1,t+1})\}_{t=1}^{T}$ and $\{(x_{2,t}, z_{1t,2}, s_{2,t}, \varepsilon_{2,t+1})\}_{t=1}^{T}$ be independent copies of the process $\{(x_t, z_{1t}, s_t, \varepsilon_{t+1})\}_{t=1}^{T}$ in (5.3) - (5.5) such that in generating $\{(x_{j,t}, z_{1t,j}, s_{j,t}, \varepsilon_{j,t+1})\}_{t=1}^{T}$, we use $\delta_j = \sqrt{2/3} \times 0.6n^{-\alpha_{z,j}}$ for $j \in \{1, 2\}$. Then we set $z_{Mt} = (z_{1t,1} + z_{1t,2})/2$, $\varepsilon_{t+1} = (\varepsilon_{1,t+1} + \varepsilon_{2,t+1})/2$ and $\beta_j = 3n^{-\alpha_{x,j}}$ in

$$y_{t+1} = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \varepsilon_{t+1}. \tag{5.6}$$

Table 3 shows the outcome of three comparisons of predictive accuracy for model 1 versus model 2 (first four columns), model 2 against dynamic rotation (columns 5-8), and model 1 versus dynamic rotation (columns 9-12). We let $\alpha_{x,1}$ and $\alpha_{x,2}$ take values of $\{0, 0.25, 0.5, 1\}$. The four panels in the table correspond to different combinations of the accuracy of the monitoring instrument for models 1 and 2, with $\{\alpha_{z,1}, \alpha_{z,2}\} = \{0, 0\}$ (panel A), $\{\alpha_{z,1}, \alpha_{z,2}\} = \{0, 1\}$ (panel B), $\{\alpha_{z,1}, \alpha_{z,2}\} = \{0.5, 0.5\}$ (panel C), and $\{\alpha_{z,1}, \alpha_{z,2}\} = \{1, 1\}$ (panel D).

As expected, the first model produces more accurate forecasts than the second model (high MSE ratios exceeding one) provided that $\alpha_{x,2} > \alpha_{x,1}$ with the ratio growing in $\alpha_{x,2} - \alpha_{x,1}$. Conversely, model 2 is more accurate than model 1 (small MSE ratios below one) when $\alpha_{x,1} > \alpha_{x,2}$ so that the second predictor is stronger than the first one, and the two models are equally accurate (MSE ratios equal to one) on the diagonal where $\alpha_{x,1} = \alpha_{x,2}$.

The dynamic rotation rule performs as well as the underlying forecasting models when these are equally accurate, i.e., on the diagonals of the middle and right-most sub panels in Panel A. When one predictor is a little stronger than the other, dynamic rotation typically performs notably better than the forecasting model that uses the weakest predictor while performing on a par with the model based on the stronger predictor.

Panel B investigates the case with asymmetric power of the monitoring instruments by letting the first instrument be strong ($\alpha_{z_1} = 0$) while the second instrument is weak ($\alpha_{z_2} = 1$). For the case where both predictors are equally strong ($\alpha_{x_1} = \alpha_{x_2}$), the MSE value of the dynamic rotation scheme is now always lower than that of the underlying models –in contrast to our other simulations which generate MSE ratios of unity when the forecasting models are equally accurate. To explain this finding note that, ignoring estimation error, the main term in equation (3.3) is $2\varepsilon_{t+1}(\beta_2 x_{2t} - \beta_1 x_{1t})$. In our data generating process, $z_{Mt} = (z_{1t} + z_{2t})/2$ and $\beta_1 = \beta_2$, so when $\alpha_{x_1} = \alpha_{x_2}$ and $\alpha_{z_1} = \alpha_{z_2}$, the correlation between z_{Mt} and $2\varepsilon_{t+1}(\beta_2 x_{2t} - \beta_1 x_{1t})$ equals zero. If instead, $\alpha_{x_1} = \alpha_{x_2}$ but $\alpha_{z_1} = 0$ and $\alpha_{z_2} = 1$, the correlation between z_{Mt} and $2\varepsilon_{t+1}(\beta_2 x_{2t} - \beta_1 x_{1t})$ is no longer equal to zero and the rotation scheme will do better.

While panels A and B assume that at least one monitoring instrument is highly accurate, Panels C and D instead let both instruments be relatively poor. Although the absence of accurate monitoring instruments means that the probability that dynamic rotation outperforms both of the underlying forecasting models deteriorates, overall our results are quite robust with respect to the strength of the monitoring instruments. When one of the predictors is powerful ($\alpha_{x_1} = 0$ or $\alpha_{x_2} = 0$), the strength of the monitoring instruments does not matter much to the MSE ratios because one forecasting model is dominant and so the best forecast gets picked up by the intercept

(θ_0) in the monitoring equation. The strength of the monitoring instruments matters more in cases with weak predictors (α_x values of 0.5 or 1) where it is not obvious which forecasting model is best, leaving greater room for improvement relative to the individual models.

Reducing the length of the monitoring window (m) is likely to lead to a deterioration in the performance of the dynamic rotation rule relative to that of the individual forecasting models while increasing this window should improve its (relative) performance. This is indeed what we find in a set of Monte Carlo simulations which are reported in Tables A1-A6 in the online supplement. Relative to the baseline case with $m = 100$, setting $m = 50$ generates worse performance, while setting $m = 200$ leads to better performance for the dynamic rotation scheme, ceteris paribus. However, changes in the (relative) performance of the dynamic rotation rule due to these shifts in m are relatively modest.

5.3 Inclusion of interaction term $x_t z_{Mt}$ in the forecasting model

Our dynamic rotation rule exploits predictability arising from the non-zero expected value of $x_t z_{Mt} \varepsilon_{t+1}$. In cases with generated forecasts both x_t and z_{Mt} are observed and so it is an option to consider an augmented forecasting model that includes this interaction term:

$$y_{t+1} = \beta_x x_t + \beta_z z_{Mt} + \beta_{xz} x_t z_{Mt} + \varepsilon_{t+1}. \tag{5.7}$$

Table 4 reports average MSE ratios of forecasts from the dynamic rotation scheme compared to forecasts from the augmented model in (5.7) that includes the interaction term. For the data generating process in (5.3) and (5.4) (Panel A), when the monitoring instrument is very strong ($\alpha_z = 0$ or 0.1), the augmented forecasting model tends to do better, producing average MSE values that are 5-10 percent lower than those from the dynamic rotation scheme. For instruments of medium strength ($\alpha_z = 0.25, 0.5$) the predictive accuracy of the dynamic rotation and augmented model is very similar, and for weak instruments ($\alpha_z \geq 0.75$) the dynamic rotation scheme performs a little better than the augmented model in (5.7).

We also consider a second data generating process which is similar to the first one, except that we now generate innovations as

$$\varepsilon_{t+1} = s_{t+1} Q_{1,t} + (1 - s_{t+1}) Q_{2,t} + 2(Q_{1,t}^2 + Q_{2,t}^2 - 2) X_t, \tag{5.8}$$

rather than $\varepsilon_{t+1} = s_{t+1}Q_{1,t} + (1 - s_{t+1})Q_{2,t}$.

Panel B in Table 4 shows results for (5.8). For this data generating process, the dynamic rotation scheme performs better than the augmented model for a wider range of parameter values α_x, α_z and does so by a larger margin compared to the first one. To see why, note that for the forecasting model augmented with the interaction term $x_t z_{Mt}$ to work, z_{Mt} needs to be correlated with $x_t \varepsilon_{t+1}$. In (5.8), this correlation is not zero and so the augmented forecasting model does not fully incorporate the information from z_{Mt}.

Overall, these simulations demonstrate that there are settings in which the dynamic rotation scheme can generate sizable improvements in predictive accuracy but also that there are many parameter configurations for which such improvements are either very small or non-existent.

6 Empirical Applications

We finally illustrate our analysis through two empirical applications to inflation forecasts and predictability of stock market returns.

6.1 Inflation Forecasts

We first compare the accuracy of the Federal Reserve's quarterly Greenbook forecasts of the GDP price deflator to the mean forecast of the same variable from the Survey of Professional Forecasters (SPF)–a case of non-nested forecasts–over the sample period 1968Q4-2014Q4. We consider forecast horizons of one through four quarters and use two different ways to measure the actual outcome, namely the value of the GDP deflator series available at the end of December 2020 (final, revised in Panel A) or the value that becomes available two quarters after the predicted quarter (vintage, in Panel C). Data on the forecasts are obtained from the Federal Reserve Bank of Philadelphia, while data on the final GDP deflator series are taken from the St Louis Federal Reserve Bank's Fred data base.

The first column in Table 5 reports MSE ratios for the Greenbook forecasts relative to the SPF forecasts with ratios below unity suggesting that the Greenbook forecasts were more accurate. Using the final, revised data to measure outcomes (Panel A), the Greenbook forecasts were 5% more accurate than the SPF forecasts at the shortest and longest ($h = 1, 4$) horizons and 10-20% more accurate at the two intermediate horizons. This difference grows larger when the vintage data is used to measure outcomes (Panel B); Greenbook forecasts are now between 25% ($h = 1$) and 20% ($h = 2, 4$) more accurate than the mean SPF forecasts.

Our dynamic rotation rule uses a 10-year rolling estimation window for the monitoring regression ($m = 40$) and is implemented for a variety of instruments.[6] First, we use the lagged loss differential measured over the most recent four quarters, $\Delta\bar{L}_{t-3:t} = (1/4)\sum_{\tau=1}^{4}\Delta L_{t+1-\tau}$. The rationale for this choice is that lags of ΔL_t will be correlated with ΔL_{t+1} provided that the regressors used by the underlying forecasts are persistent and do not completely overlap. Second, we use the squared difference in forecasts, again averaged over the most recent four quarters, $\Delta\hat{y}^2_{t-3:t} = (1/4)\sum_{\tau=0}^{3}(\hat{y}^2_{1t+1-\tau|t-\tau} - \hat{y}^2_{2t+1-\tau|t-\tau})$. Using $z_t = (\hat{y}^2_{1,t+1|t}, \hat{y}^2_{2,t+1|t})'$ as an instrument makes sense because, from (3.3), $\mathbb{E}(\Delta L_{t+1} \mid z_t) = \hat{y}^2_{2,t+1|t} - \hat{y}^2_{1,t+1|t}$, provided that $\mathbb{E}(\varepsilon_{t+1} \mid x_{1,t}, x_{2,t}) = 0$. Hence, we can regress ΔL_{t+1} on $\hat{y}^2_{1,t+1|t} - \hat{y}^2_{2,t+1|t}$ to compute $\mathbb{E}(\Delta L_{t+1} \mid z_t)$. As a third instrument we use the unemployment gap (UG) from Stock and Watson (2010). While the first two instruments are readily available in real time, the third instrument might involve a short delay.

Using the final, revised data to measure outcomes and the unemployment gap as the monitoring instrument, the dynamic rotation scheme reduces the MSE of the Greenbook and SPF forecasts by 10-15% at the one-quarter horizon and by 10-30% at the longer horizons. Using instead the vintage data to measure outcomes, the MSE values for the dynamic rotation scheme and the Greenbook forecasts are close to unity across most horizons, while dynamic rotation reduces the MSE values of the SPF forecasts by 17-23% across the four horizons.

Comparing results across the three sets of monitoring instruments, the best results for the dynamic rotation scheme are obtained using the four-quarter moving average of the lagged loss differential, $\Delta\bar{L}_{t-3:t}$, followed by the unemployment gap and the squared difference in forecasts, $\Delta\hat{y}^2_{t-3:t}$. The dynamic rotation rule that uses the lagged loss differential reduces the MSE of the Greenbook forecasts by 7 percent at the two longest horizons ($h = 3, 4$) and outperforms the SPF forecasts by a considerable margin across all horizons.

In this application, one forecast (Greenbook) is notably more accurate than the other (SPF). Encouragingly, the dynamic rotation rule still manages to either out-

[6] Similar results are obtained if instead we use a 15-year estimation window ($m = 60$).

perform or at least perform on a par with the best of the underlying forecasts, and it performs substantially better than the least accurate forecast. This suggests that the dynamic rotation scheme has an attractive risk profile, i.e., it does not underperform any of the underlying forecasts by a sizable margin and it produces the most accurate forecasts in many cases.

6.2 Forecasts of Stock Returns

Our second application covers the nested case and considers predictability of quarterly returns on the S&P500 stock index, measured net of a short T-bill rate. Following Welch and Goyal (2008), we use a constant mean as our small model:

$$y_{t+1} = \mu + \varepsilon_{t+1}, \tag{6.1}$$

while, again consistent with Welch and Goyal (2008), the big model is a univariate regression model with a single predictor variable:

$$y_{t+1} = \mu + \beta x_t + \varepsilon_{Bt+1}. \tag{6.2}$$

Our data cover the sample 1927-2019. We use a 20-year rolling window to estimate the parameters of the underlying forecasting models ($n = 80$) and also use 20 years of out-of-sample forecasts to run the monitoring regressions ($m = 80$) and compute the expected loss differential in (2.3). Our analysis of the small and big models' out-of-sample forecasting performance thus runs from 1967 through 2019, a total of 212 quarterly observations. We implement the dynamic rotation rule using the same set of monitoring instruments as in our inflation application, namely the lagged loss differential measured over the most recent four quarters, $\Delta \bar{L}_{t-3:t}$, the four-quarter moving average of squared differences in forecasts, $\Delta \hat{y}^2_{t-3:t|t-4:t-1}$, and the unemployment gap.

Table 6 reports the forecasting performance for 17 different predictor variables taken from the Goyal-Welch data set. Due to the very low signal-to-noise ratio of the return prediction models (large α_x), the MSE value of the big forecasting model is higher than that of the simple prevailing mean model (constant only) for 16 of 17 predictors. This is consistent with the results in Welch and Goyal (2008).

Using the unemployment gap as our instrument, the dynamic rotation scheme does a little better, reducing the MSE value of the prevailing mean model for three predictors, while improving on the big model for 10 of 17 predictors. Moreover, when the MSE value of the dynamic rotation scheme exceeds that of the big model, it only does so by a small margin that does not exceed one percent. Conversely, we see improvements of up to 23 percent (for the stock variance (svar) predictor) for the dynamic rotation scheme relative to the big model.

Similar results are obtained for the other instruments with a few notable exceptions. For example, the dynamic rotation scheme performs somewhat worse relative to the small model for the svar predictor when using $\Delta\bar{L}_{t-3:t}$ or $\Delta\hat{y}^2_{t-3:t|t-4:t-1}$ rather than the unemployment gap as instruments. This happens because the svar predictor is affected by large outliers which also introduce outliers in the two loss and forecast-based instruments.

Univariate return prediction models are clearly a case with weak predictors and weak instruments (large α_x and α_z). Hence, our finding that the dynamic rotation scheme performs a little better than the large model but is worse than the small forecasting model is consistent with the Monte Carlo simulations in Table 1.

The final column in Table 6 reports results when the forecasts from the large model are replaced with an equal-weighted (EW) average of the 17 underlying univariate forecasts. The MSE ratio of this EW average relative to the prevailing mean is now much lower (0.977), so using an EW average improves forecasting performance by a substantial amount. Similarly, the dynamic rotation scheme substantially reduces the MSE value of the prevailing mean model and its performance is on a par with the EW forecast.

The final row in each panel in Table 6 compares the augmented model that includes x_t,z_{Mt},and $x_t z_{Mt}$ as predictors with the dynamic rotation scheme. Across all 17 predictors, the dynamic rotation scheme produces lower MSE values than this augmented model. This shows that in applications such as this one with weak predictors and weak instruments (large α_x and α_z), augmenting the forecasting model with the interaction term can produce less accurate forecasts. This finding is consistent with the Monte Carlo simulation results in Table 4.

7 Conclusion

We derive finite-sample bounds on differences in the expected mean squared error performance of competing forecasting models conditional on a set of monitoring instruments. Our analysis covers both the case with nested and non-nested forecasting models and accounts for parameter estimation error. We show that the possibility of establishing gains from monitoring the performance of competing forecasts and selecting, at each point in time, the forecast with the smallest expected loss requires conditions on the accuracy of both the predictors used by the underlying forecasting models as well as the strength of the monitoring instruments. For this dynamic rotation to work, at least one of the models must use predictors that are not too weak. Moreover, none of the underlying forecasting models can be too dominant as, otherwise, there is little space for improvements by alternating between the two forecasting models.

Monte Carlo simulations and two empirical applications to forecasts of inflation and stock returns illustrate that there are conditions under which the dynamic rotation scheme works well and produces sizable gains in predictive accuracy, but also that there are cases in which such gains are negligible or reversed. Even when it does not produce the most accurate forecasts, the dynamic rotation scheme typically performs better than the worst of the underlying forecasts and only underperforms the best forecast by a relatively modest margin. This limits the downside from applying dynamic rotation and makes it an attractive option relative to always sticking with a single model whose performance may deteriorate over time.

Acknowledgements

We are grateful to two anonymous referees for many constructive comments on the paper. We also thank Neil Ericsson, Frank Diebold, Hashem Pesaran, Dacheng Xiu, and seminar participants at University of Chicago (Booth), UIUC, USC, the Tinbergen Institute at Erasmus University, Rotterdam, the 2016 Duke conference on New Developments in Measuring & Forecasting Financial Volatility, the 2017 SoFiE conference in New York City, and the 2017 ISF conference in Cairns for comments on the paper. An earlier version of the paper was circulated under the title “monitoring forecasting performance”.

References

Aiolfi, M. and Timmermann, A. (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics*, 135(1):31–53.

Andrews, D. W. (1988). Laws of large numbers for dependent non-identically distributed random variables. *Econometric theory*, 4(03):458–467.

Athreya, K. B. and Lahiri, S. N. (2006). *Measure Theory and Probability Theory*. Springer Science & Business Media.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650.

Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009). Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, pages 1705–1732.

Bonhomme, S. and Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, 83(3):1147–1184.

Bradley, R. C. (2007). *Introduction to strong mixing conditions*, volume 1. Kendrick Press Heber City.

Candes, E. and Tao, T. (2007). The dantzig selector: statistical estimation when p is much larger than n. *The Annals of Statistics*, pages 2313–2351.

Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various garch and stochastic volatility models. *Econometric Theory*, 18(1):17–39.

Chen, X., Shao, Q.-M., and Wu, W. B. (2016). Self-normalized cramér-type moderate deviations under dependence. *The Annals of Statistics*, 44(4):1593–1617.

Chong, Y. Y. and Hendry, D. F. (1986). Econometric evaluation of linear macroeconomic models. *The Review of Economic Studies*, 53(4):671–690.

Clark, T. and McCracken, M. (2013). Advances in forecast evaluation. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2, part B, pages 1107–1201. Elsevier.

Clark, T. E. and McCracken, M. W. (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of econometrics*, 105(1):85–110.

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

Diebold, F. X. and Rudebusch, G. D. (1991). Forecasting output with the composite leading index: A real-time analysis. *Journal of the American Statistical Association*, 86(415):603–610.

Douc, R., Moulines, E., Olsson, J., and Van Handel, R. (2011). Consistency of the maximum likelihood estimator for general hidden markov models. *The Annals of Statistics*, 39(1):474–513.

Elliott, G. and Timmermann, A. (2016). *Economic Forecasting*. Princeton University Press.

Fan, J., Liao, Y., and Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor models. *Annals of statistics*, 39(6):3320.

Fan, J. and Yao, Q. (2003). *Nonlinear time series*, volume 2. Springer.

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578.

Granger, C. W. J. and Newbold, P. (1977). *Forecasting economic time series*. Academic Press.

Hansen, P. R. and Timmermann, A. (2015). Equivalence between out-of-sample forecast comparisons and wald statistics. *Econometrica*, 83(6):2485–2505.

Inoue, A. and Kilian, L. (2005). In-sample or out-of-sample tests of predictability: Which one should we use? *Econometric Reviews*, 23(4):371–402.

Inoue, A. and Kilian, L. (2008). How useful is bagging in forecasting economic time series? a case study of us consumer price inflation. *Journal of the American Statistical Association*, 103(482):511–522.

Koltchinskii, V., Lounici, K., and Tsybakov, A. B. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics*, 39(5):2302–2329.

McCracken, M. W. (2000). Robust out-of-sample inference. *Journal of Econometrics*, 99(2):195–223.

Meyn, S. P. and Tweedie, R. L. (2012). *Markov chains and stochastic stability*. Springer Science & Business Media.

Peña, V. H., Lai, T. L., and Shao, Q.-M. (2008). *Self-normalized processes: Limit theory and Statistical Applications*. Springer Science & Business Media.

Pesaran, M. H. and Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137(1):134–161.

Pesaran, M. H. and Timmermann, A. (2009). Testing dependence among serially correlated multicategory variables. *Journal of the American Statistical Association*, 104(485):325–337.

Rossi, B. (2013). Advances in forecasting under instability. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2, part B, chapter 21, pages 1203–1324. Elsevier.

Stock, J. H. and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, 14(1):11–30.

Stock, J. H. and Watson, M. W. (2010). Modeling inflation after the crisis. Technical report, National Bureau of Economic Research.

Tong, H. (1990). *Non-linear time series: A dynamical system approach*. Oxford University Press.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica: Journal of the Econometric Society*, pages 1067–1084.

Table 1: Predictive performance of nested models and dynamic rotation

$\alpha_z \backslash \alpha_x$	0	0.1	0.25	0.4	0.5	0.75	1
A: Big vs. small model (MSE_{big}/MSE_{small})							
	0.253	0.460	0.777	0.939	0.981	1.007	1.010
B: DR vs. small model (MSE_{DR}/MSE_{small})							
0.0	0.242	0.408	0.671	0.849	0.912	0.973	0.981
0.1	0.258	0.456	0.741	0.893	0.943	0.985	0.991
0.25	0.258	0.468	0.781	0.932	0.971	0.997	1.000
0.5	0.257	0.466	0.788	0.948	0.986	1.003	1.004
0.75	0.256	0.466	0.789	0.951	0.988	1.004	1.005
1.0	0.256	0.465	0.789	0.951	0.987	1.004	1.005
C: DR vs. big model (MSE_{DR}/MSE_{big})							
0.0	0.955	0.883	0.865	0.903	0.930	0.966	0.972
0.1	1.021	0.990	0.953	0.952	0.961	0.978	0.981
0.25	1.022	1.017	1.004	0.994	0.991	0.990	0.990
0.5	1.018	1.013	1.015	1.011	1.005	0.996	0.995
0.75	1.017	1.013	1.015	1.013	1.007	0.997	0.995
1.0	1.017	1.013	1.015	1.013	1.007	0.997	0.995

This table reports the relative predictive accuracy in the form of mean squared error (MSE) ratios associated with forecasts generated using big or small forecasting models or a dynamic rotation (DR) scheme. All results are based on 5,000 Monte Carlo simulations and use a sample size of $(n, m, p) = (100, 100, 200)$.

Table 2: Predictive performance of dynamic rotation versus forecast combinations, pretesting and a model augmented with the monitoring instrument

$\alpha_z \backslash \alpha_x$	0	0.1	0.25	0.4	0.5	0.75	1
			A: DR vs. equal-weighted combination (MSE_{DR}/MSE_{EW})				
0.0	0.551	0.687	0.811	0.893	0.930	0.973	0.979
0.1	0.589	0.768	0.893	0.941	0.962	0.985	0.989
0.25	0.589	0.789	0.941	0.982	0.991	0.997	0.997
0.5	0.586	0.786	0.951	0.999	1.006	1.003	1.002
0.75	0.585	0.786	0.952	1.002	1.007	1.004	1.003
1.0	0.585	0.785	0.951	1.001	1.007	1.004	1.003
			B: DR vs. pre-test forecast ($MSE_{DR}/MSE_{pretest}$)				
0.0	0.955	0.883	0.865	0.893	0.918	0.970	0.978
0.1	1.021	0.990	0.953	0.942	0.950	0.982	0.988
0.25	1.022	1.017	1.004	0.982	0.979	0.994	0.997
0.5	1.018	1.013	1.015	1.000	0.993	1.000	1.001
0.75	1.017	1.013	1.015	1.001	0.995	1.001	1.002
1.0	1.017	1.013	1.015	1.001	0.995	1.001	1.002
			C: DR vs. augmented forecast ($MSE_{DR}/MSE_{augmented}$)				
0.0	0.936	0.866	0.848	0.885	0.912	0.948	0.953
0.1	1.002	0.971	0.934	0.934	0.943	0.959	0.962
0.25	1.002	0.997	0.984	0.974	0.972	0.971	0.970
0.5	0.998	0.993	0.995	0.991	0.986	0.977	0.976
0.75	0.997	0.993	0.995	0.993	0.987	0.978	0.976
1.0	0.997	0.993	0.996	0.994	0.987	0.978	0.976
			D: DR vs. weighted forecast comb ($MSE_{DR}/MSE_{weight-comb}$)				
0.0	0.851	0.793	0.823	0.894	0.930	0.973	0.979
0.1	0.910	0.887	0.907	0.942	0.962	0.985	0.989
0.25	0.911	0.911	0.956	0.983	0.991	0.997	0.997
0.5	0.906	0.908	0.966	1.000	1.006	1.003	1.002
0.75	0.906	0.908	0.967	1.003	1.007	1.004	1.003
1.0	0.906	0.908	0.966	1.002	1.007	1.004	1.003

This table compares the MSE performance of the dynamic rotation (DR) scheme to that of an equal-weighted (EW) forecast combination (Panel A), a pre-test approach that includes a predictor in the forecasting model if its regression coefficient is statistically significant (Panel B), forecasts from an augmented forecasting model that includes both the predictor, x_t, and the monitoring instrument, z_{Mt}, (Panel C) and forecasts from a combination scheme with weights proportional to the inverse of the MSE of the individual forecasts (Panel D).

Table 3: Pairwise comparisons of predictive performance for the non-nested case

	MSE_{j_2}/MSE_{j_1}											
	$(j_1, j_2) = (1, 2)$				$(j_1, j_2) = (2, DR)$				$(j_1, j_2) = (1, DR)$			
					A: $(\alpha_{z,1}, \alpha_{z,2}) = (0, 0)$							
$\alpha_{x,1} \backslash \alpha_{x,2}$	0	0.25	0.5	1	0	0.25	0.5	1	0	0.25	0.5	1
0.0	0.999	4.371	6.611	6.992	1.000	0.229	0.152	0.144	0.999	1.003	1.004	1.005
0.25	0.228	1.003	1.512	1.597	1.020	0.998	0.665	0.621	0.233	1.001	1.005	0.992
0.5	0.151	0.663	1.000	1.059	1.033	1.008	0.995	0.926	0.156	0.668	0.995	0.980
1.0	0.143	0.626	0.944	1.000	1.035	0.995	0.980	0.988	0.148	0.623	0.925	0.988
					B: $(\alpha_{z,1}, \alpha_{z,2}) = (0, 1)$							
$\alpha_{x,1} \backslash \alpha_{x,2}$	0	0.25	0.5	1	0	0.25	0.5	1	0	0.25	0.5	1
0.0	0.999	4.367	6.597	6.988	0.947	0.229	0.152	0.143	0.946	1.002	1.001	1.001
0.25	0.228	0.998	1.510	1.597	1.024	0.913	0.670	0.631	0.234	0.911	1.011	1.008
0.5	0.151	0.662	1.001	1.059	1.034	0.989	0.958	0.951	0.156	0.654	0.959	1.007
1.0	0.143	0.625	0.944	1.000	1.036	0.993	0.974	0.992	0.148	0.621	0.920	0.992
					C:$(\alpha_{z,1}, \alpha_{z,2}) = (0.5, 0.5)$							
$\alpha_{x,1} \backslash \alpha_{x,2}$	0	0.25	0.5	1	0	0.25	0.5	1	0	0.25	0.5	1
0.0	1.000	4.374	6.602	6.990	1.000	0.229	0.152	0.143	1.000	1.001	1.001	1.001
0.25	0.229	1.002	1.512	1.600	1.018	1.000	0.669	0.630	0.233	1.002	1.011	1.007
0.5	0.152	0.663	1.001	1.059	1.029	1.014	1.000	0.955	0.156	0.672	1.001	1.011
1.0	0.143	0.626	0.944	1.000	1.031	1.011	1.012	1.002	0.148	0.633	0.955	1.002
					D: $(\alpha_{z,1}, \alpha_{z,2}) = (1, 1)$							
$\alpha_{x,1} \backslash \alpha_{x,2}$	0	0.25	0.5	1	0	0.25	0.5	1	0	0.25	0.5	1
0.0	1.002	4.382	6.607	7.002	0.999	0.229	0.151	0.143	1.001	1.001	1.001	1.001
0.25	0.229	0.999	1.511	1.599	1.017	0.999	0.670	0.631	0.233	0.998	1.012	1.008
0.5	0.151	0.662	1.000	1.059	1.028	1.014	1.000	0.955	0.156	0.671	1.000	1.012
1.0	0.143	0.626	0.944	1.000	1.031	1.011	1.013	1.002	0.148	0.632	0.956	1.003

This table reports the ratio of MSE values for method j_2 versus method j_1. Data are generated according to the non-nested model

$$y_{t+1} = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \varepsilon_{t+1},$$

where x_{1t} and x_{2t} are predictor variables that are known at time t. Model 1 takes the form $y_{t+1} = \beta_1 x_{1,t} + \varepsilon_{1t+1}$, while model 2 takes the form $y_{t+1} = \beta_2 x_{2,t} + \varepsilon_{2t+1}$. The strength of the predictors in models 1 and 2 is parameterized as $\beta_j = c_{\beta,j} n^{-\alpha_{x,j}}$, while the accuracy of the monitoring instrument is captured as $corr(x_{j,t}\varepsilon_{t+1}, z_{1t}) = c_{\rho,j} m^{-\alpha_{z,j}}$. All results are based on 5,000 MC simulations and use a sample size of $(n, m, p) = (100, 100, 200)$.

Table 4: Predictive performance of dynamic rotation versus a model augmented with the monitoring instrument and interaction terms

$\alpha_z \backslash \alpha_x$	0	0.1	0.25	0.4	0.5	0.75	1
	A: DGP I (MSE_{DR}/MSE_{aug})						
0.0	1.102	1.058	1.078	1.114	1.110	1.095	1.094
0.1	1.076	1.051	1.028	1.031	1.038	1.045	1.046
0.25	1.023	1.015	1.004	0.998	0.997	0.996	0.998
0.5	0.999	0.993	0.990	0.986	0.983	0.978	0.977
0.75	0.995	0.989	0.988	0.986	0.982	0.976	0.975
1.0	0.995	0.990	0.988	0.986	0.983	0.976	0.975
	B: DGP II (MSE_{DR}/MSE_{aug})						
0.0	0.999	0.956	0.912	0.901	0.899	0.904	0.904
0.1	1.007	0.975	0.936	0.921	0.921	0.921	0.921
0.25	1.010	0.987	0.951	0.934	0.932	0.931	0.931
0.5	1.010	0.990	0.954	0.937	0.935	0.934	0.935
0.75	1.010	0.990	0.955	0.936	0.935	0.935	0.935
1.0	1.010	0.989	0.955	0.938	0.935	0.935	0.936

This table reports mean squared error ratios for a dynamic rotation (DR) scheme that rotates between forecasts generated by a small and a big forecasting model versus forecasts from an augmented model that includes x_t, z_{Mt} and $x_t z_{Mt}$ as regressors. All simulations use a nested model setup with $y_{t+1} = \beta x_t + \varepsilon_{t+1}$, where under DGP I, $\varepsilon_{t+1} = s_{t+1}Q_{1,t} + (1 - s_{t+1})Q_{2,t}$, while under DGP II, $\varepsilon_{t+1} = s_{t+1}Q_{1,t} + (1 - s_{t+1})Q_{2,t} + 2(Q_{1,t}^2 + Q_{2,t}^2 - 2)X_t$. $Q_{1,t}$ and $Q_{2,t}$ are $N(0,1)$ random variables that are mutually independent and independent of s_{t+1} and x_t. All results are based on 5,000 MC simulations and use a sample size of $(n, m, p) = (100, 100, 200)$.

Table 5: Forecasting performance: Greenbook, Survey of Professional Forecasters, and Dynamic Rotation Scheme

$h \backslash Z_t$	Uncond	UG_t		$\overline{\Delta L}_{t-3:t}$		$\overline{\Delta \hat{y}^2_{t-3:t}}$	
	$\frac{MSE_{GB}}{MSE_{SPF}}$	$\frac{MSE_{DR}}{MSE_{GB}}$	$\frac{MSE_{DR}}{MSE_{SPF}}$	$\frac{MSE_{DR}}{MSE_{GB}}$	$\frac{MSE_{DR}}{MSE_{SPF}}$	$\frac{MSE_{DR}}{MSE_{GB}}$	$\frac{MSE_{DR}}{MSE_{SPF}}$
A: Final, revised data							
1	0.948	0.884	0.838	0.874	0.829	0.910	0.863
2	0.797	0.856	0.682	0.800	0.638	0.861	0.686
3	0.890	0.901	0.801	0.826	0.735	0.930	0.827
4	0.944	0.870	0.822	0.815	0.769	0.932	0.880
B: Vintage data							
1	0.740	1.086	0.804	1.043	0.772	1.018	0.753
2	0.799	1.010	0.807	1.008	0.805	1.000	0.799
3	0.771	0.994	0.767	0.935	0.721	0.969	0.747
4	0.812	1.019	0.827	0.933	0.758	1.012	0.821

This table reports MSE ratios for the quarterly GDP deflator using Greenbook (GB) forecasts, forecasts from the Survey of Professional Forecasters (SPF), or forecasts from a dynamic rotation (DR) scheme. The dynamic rotation scheme is based on a 10-year estimation window ($m = 40$) and uses the monitoring instruments Z_t listed above the rows. The forecast horizon runs from $h = 1$ through $h = 4$ quarters. The sample period is 1968Q4-2014Q4.

Table 6: Predictive performance of nested models and dynamic rotation

	d/p	d/y	e/p	d/e	b/m	tbl	dfy	lty	cay	ntis	infl	ltr	tms	svar	csp	ik	dfr	avg
									$Z_t = UG_t$									
$\frac{MSE_{big}}{MSE_{small}}$	1.004	1.003	1.065	1.043	1.055	1.014	1.064	1.022	1.030	1.018	1.019	1.015	1.011	1.281	0.990	1.015	1.039	0.977
$\frac{MSE_{DR}}{MSE_{small}}$	0.999	1.007	1.035	1.042	1.013	1.021	1.067	0.990	1.035	1.021	0.989	1.029	1.009	0.988	1.000	1.008	1.008	0.976
$\frac{MSE_{DR}}{MSE_{big}}$	0.995	1.004	0.972	0.999	0.961	1.007	1.004	0.969	1.006	1.003	0.971	1.014	0.998	0.771	1.010	0.993	0.970	0.999
$\frac{MSE_{DR}}{MSE_{aug}}$	0.943	0.980	0.938	0.907	0.946	0.901	0.655	0.842	0.858	0.885	0.954	0.919	0.961	0.406	0.920	1.049	0.784	0.759
									$Z_t = \overline{\Delta L_{t-3:t}}$									
$\frac{MSE_{big}}{MSE_{small}}$	1.004	1.003	1.064	1.043	1.053	1.019	1.064	1.025	1.031	1.018	1.020	1.013	1.010	1.278	0.990	1.004	1.040	0.977
$\frac{MSE_{DR}}{MSE_{small}}$	0.997	1.033	1.057	0.990	1.019	1.068	1.065	1.052	1.026	1.006	1.004	1.014	1.031	1.283	1.000	0.999	1.021	0.983
$\frac{MSE_{DR}}{MSE_{big}}$	0.993	1.030	0.993	0.950	0.968	1.048	1.001	1.026	0.995	0.988	0.984	1.001	1.020	1.003	1.010	0.995	0.982	1.006
$\frac{MSE_{DR}}{MSE_{aug}}$	0.875	0.933	0.473	0.541	0.761	0.945	0.566	0.830	0.933	0.897	0.788	0.942	0.934	0.154	0.976	0.987	0.856	0.824
									$Z_t = \overline{\Delta \hat{y}^2_{t-3:t}}$									
$\frac{MSE_{big}}{MSE_{small}}$	1.004	1.003	1.064	1.043	1.053	1.019	1.064	1.025	1.031	1.018	1.020	1.013	1.010	1.278	0.990	1.004	1.040	0.977
$\frac{MSE_{DR}}{MSE_{small}}$	1.019	1.043	0.981	1.009	1.030	1.039	1.005	1.021	1.014	1.015	1.020	1.008	1.015	1.287	0.999	1.016	1.009	0.978
$\frac{MSE_{DR}}{MSE_{big}}$	1.015	1.040	0.922	0.968	0.978	1.020	0.945	0.996	0.983	0.996	1.000	0.995	1.005	1.007	1.009	1.012	0.970	1.001
$\frac{MSE_{DR}}{MSE_{aug}}$	0.846	0.885	0.479	0.835	0.872	0.936	0.022	0.910	0.925	0.885	0.842	0.966	0.854	0.060	0.900	0.934	0.771	0.748

This table reports ratios of mean squared errors (MSEs) for forecasts of quarterly stock market returns based on a small prediction model that only includes an intercept, a big model that includes an intercept plus one of the predictors listed in each column, and a dynamic rotation (DR) scheme that uses the instrument listed above each panel. Data on quarterly stock returns along with the 17 predictor variables are obtained from Welch and Goyal (2008) which provides details on the construction of the individual predictors. The variables are: dividend-price ratio (d/p), dividend yield (d/y), earnings-price ratio (e/p), dividend payout ratio (d/e), book-to-market ratio (b/m), T-bill rate (tbl), default yield (dfy), long term yield (lty), consumption, wealth, income ratio (cay), net equity expansion (ntis), inflation (infl), long-term rate of return (ltr), term spread (tms), stock variance (svar), cross-sectional premium (csp), investment to capital ratio (ik) and default return spread (dfr). The final column (avg) uses an equal-weighted average of the 17 univariate forecasts as the forecast from the large model. Z_t refers to the monitoring instrument used by the dynamic rotation rule. All forecasts are generated out-of-sample for the period 1967-2019.

Online Appendix for "Conditional Rotation Between Forecasting Models"

Yinchu Zhu[a] and Allan Timmermann[b]

[a] Department of Economics, Brandeis University, 415 South Street Waltham, MA 02453, U.S.A.; yinchuzhu@brandeis.edu

[b] Rady School of Management, University of California, San Diego, 9500 Gilman Dr, La Jolla, CA 92093, U.S.A.; atimmermann@ucsd.edu

This appendix contains proofs of theoretical results in the paper as well as additional tables of Monte Carlo simulations discussed in the paper.

Appendix A Technical results used in the proofs

This appendix provides proofs of the theoretical results in the paper. The appendix is structured as follows. Appendix A provides technical tools used in the proofs while the theoretical results in the main text are proved in Appendix B.

First, some comments on notation. Throughout the appendix, the constants do not depend on T, n or t. For a vector $x = (x_1, \cdots, x_p)' \in \mathbb{R}^p$, $\|x\|_r = (\sum_{i=1}^p |x_i|^r)^{1/r}$. For a random variable or vector X, let $\|X\|_{L^r(\mathbb{P})} = (\mathbb{E}\|X\|_r^r)^{1/r}$. For two sequences $a_T, b_T > 0$, we say that $a_T \asymp b_T$ if $a_T = O(b_T)$ and $b_T = O(a_T)$. For any real number $x \geq 0$, we define $\lfloor x \rfloor$ to be the largest integer no larger than x.

Lemma 1. *Let $\mathcal{F}$ and $\mathcal{G}$ be σ-algebras with strong mixing coefficient α. Let $X \in \mathcal{F}$ and $Y \in \mathcal{G}$ be random variables with $\mathbb{E}X = 0$. Suppose that $\|X\|_{L^p(\mathbb{P})} \leq C_1$, $\|Y\|_{L^q(\mathbb{P})} \leq C_2$ for some constants $C_1, C_2 > 0$ and $p, q \in (1, \infty]$ satisfying $1/p + 1/q < 1$. Then $\mathbb{E}\,|\mathbb{E}(X|\mathcal{G})Y| \leq 8\alpha^{1-1/p-1/q}C_1C_2$.*

Proof. Let $Z = \mathbb{E}(X \mid \mathcal{G})$. Define h to be the sign of YZ, i.e., $h = \mathbf{1}\{YZ > 0\} - \mathbf{1}\{YZ < 0\}$. Therefore, $Z, h \in \mathcal{G}$. We notice that

$$\mathbb{E}|ZY| \overset{(i)}{=} \mathbb{E}(ZYh) = \mathbb{E}\left[\mathbb{E}(X \mid \mathcal{G})Yh\right] \overset{(ii)}{=} \mathbb{E}(XYh) \overset{(iii)}{\leq} 8\alpha^{1-1/p-1/q}\|X\|_{L^p(\mathbb{P})}\|hY\|_{L^q(\mathbb{P})},$$

where (i) holds by $|ZY| = ZYh$, (ii) follows by $Y, h \in \mathcal{G}$ and the law of iterated expectations and (iii) holds by $\mathbb{E}X = 0$ and Davydov's Theorem (Theorem 3.7 of Bradley (2007)). By the above display and $\mathbb{P}(|h| \leq 1) = 1$, the desired result follows. □

Lemma 2. *Let X and Y satisfy that $\mathbb{E}|X|^{c_1}, \mathbb{E}|Y|^{c_2} \leq D$. Then $\mathbb{E}|XY|^v \leq D$, where $v = c_1c_2/(c_1 + c_2)$.*

Proof. Let $p = c_1/v$ and $q = c_2/v$. Then $p^{-1}+q^{-1} = 1$. The result follows by Holder's inequality:

$$\mathbb{E}|XY|^v \leq (\mathbb{E}|X|^{vp})^{1/p} (\mathbb{E}|Y|^{vq})^{1/q} = (\mathbb{E}|X|^{c_1})^{1/p} (\mathbb{E}|Y|^{c_2})^{1/q} \leq D.$$

□

Lemma 3. *Let $\{X_i\}_{i=1}^n$ be independent random variables. Suppose that $\mathbb{E}X_i = 0$ and $\max_{1\leq i\leq n} \mathbb{E}|X_i|^p < K$ for some constants $p > 2$ and $K < \infty$. Then $\forall a, t > 0$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i \mathbf{1}\{|X_i| \leq a\}\right| \geq \sqrt{n}t + na^{1-p}K\right) \leq 2\exp\left[-\frac{t^2}{2\left(a^{2-p}K + atn^{-1/2} + n^{-1}\sum_{i=1}^n \mathbb{E}X_i^2\right)}\right].$$

Proof. Let $\tilde{X}_i = X_i\mathbf{1}\{|X_i| \leq a\}$, $Z_i = \tilde{X}_i - \mathbb{E}\tilde{X}_i$ and $B_n^2 = \sum_{i=1}^n \mathbb{E}Z_i^2$. Thus, $\mathbb{E}Z_i = 0$ and $\mathbb{P}(Z_i \leq 2a) = 1$. It follows by Theorem 2.17 of Peña et al. (2008) that $\forall z > 0$

$$\mathbb{P}\left(\sum_{i=1}^n Z_i \geq z\right) \leq \exp\left(-\frac{z^2}{2(B_n^2 + 2az)}\right).$$

Applying the same result to $\{-Z_i\}_{i=1}^n$, we obtain

$$\mathbb{P}\left(-\sum_{i=1}^n Z_i \geq z\right) \leq \exp\left(-\frac{z^2}{2(B_n^2 + 2az)}\right).$$

The above two inequalities imply that

$$\mathbb{P}\left(\left|\sum_{i=1}^n Z_i\right| \geq z\right) \leq 2\exp\left(-\frac{z^2}{2(B_n^2 + 2az)}\right). \tag{A.1}$$

We now bound $\mathbb{E}\tilde{X}_i$. Notice that

$$|\mathbb{E}\tilde{X}_i| = |\mathbb{E}(X_i - X_i\mathbf{1}\{|X_i|>a\})| \overset{(i)}{=} |\mathbb{E}X_i\mathbf{1}\{|X_i|>a\}| \leq \mathbb{E}|X_i|\mathbf{1}\{|X_i|>a\} \\ \leq \mathbb{E}\left|\frac{|X_i|^{p-1}}{a^{p-1}}\left(|X_i|\mathbf{1}\{|X_i|>a\}\right)\right| \leq \mathbb{E}|X_i|^p a^{1-p},$$

where (i) holds by $\mathbb{E}X_i = 0$. Therefore,

$$\left|\sum_{i=1}^n \mathbb{E}\tilde{X}_i\right| \leq \sum_{i=1}^n |\mathbb{E}\tilde{X}_i| \leq nKa^{1-p}. \tag{A.2}$$

Moreover, $\forall t > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{i=1}^n X_i\mathbf{1}\{|X_i|\leq a\}\right| \geq \sqrt{n}t + na^{1-p}K\right) \\
&= \mathbb{P}\left(\left|\left(\sum_{i=1}^n Z_i\right) + \left(\sum_{i=1}^n \mathbb{E}\tilde{X}_i\right)\right| \geq \sqrt{n}t + na^{1-p}K\right) \\
&\leq \mathbb{P}\left(\left|\sum_{i=1}^n Z_i\right| + \left|\sum_{i=1}^n \mathbb{E}\tilde{X}_i\right| \geq \sqrt{n}t + na^{1-p}K\right) \\
&\overset{(i)}{\leq} \mathbb{P}\left(\left|\sum_{i=1}^n Z_i\right| \geq \sqrt{n}t\right) \overset{(ii)}{\leq} 2\exp\left(-\frac{nt^2}{2(B_n^2 + 2a\sqrt{n}t)}\right),
\end{aligned} \tag{A.3}$$

where (i) follows by (A.2) and (ii) follows by (A.1) with $z = \sqrt{n}t$.

It remains to bound B_n^2. Notice that

$$\begin{aligned}
B_n^2 - \sum_{i=1}^n \mathbb{E}X_i^2 &= \sum_{i=1}^n \left[\mathbb{E}\tilde{X}_i^2 - \mathbb{E}X_i^2 - (\mathbb{E}\tilde{X}_i)^2\right] \\
&= \sum_{i=1}^n \left[\mathbb{E}X_i^2\mathbf{1}\{|X_i|>a\} - (\mathbb{E}\tilde{X}_i)^2\right] \\
&\leq \sum_{i=1}^n \mathbb{E}X_i^2\mathbf{1}\{|X_i|>a\} \\
&\leq \sum_{i=1}^n \mathbb{E}\left[\frac{|X_i|^{p-2}}{a^{p-2}}\left(X_i^2\mathbf{1}\{|X_i|>a\}\right)\right] \leq \sum_{i=1}^n \mathbb{E}|X_i|^p a^{2-p} \leq na^{2-p}K.
\end{aligned}$$

The desired result follows from the above inequality and equation (A.3). □

Lemma 4. *Let $\{Y_t\}_{t=1}^T$ be random variables with beta-mixing coefficient satisfying $\beta(i) \leq \tau_1\exp(-\tau_2 i^{\tau_3})$ for some constants $\tau_1, \tau_2, \tau_3 > 0$. Suppose that $\mathbb{E}Y_t = 0$ and*

$\max_{1\le t\le T}\mathbb{E}|Y_t|^p \le D$ *for some constants* $p>2$ *and* $D>0$. *Then for any* $p_0\in(2,p)$, *there exist constants* $K_1,...,K_5>0$ *such that* $\forall w\ge 1$,

$$\mathbb{P}\left(\left|\sum_{t=s+1}^{s+n} Y_t\right| \ge \sqrt{n}K_1 w\right) \le 2\exp\left(-K_2w^2\right) + K_3 n^{1-p_0/2}\log^{-K_4} n$$

and

$$\mathbb{P}\left(\max_{1\le s\le T-n}\left|\sum_{t=s+1}^{s+n} Y_t\right| \ge \sqrt{n\log T}K_5\right) \le 2K_3 T n^{1-p_0/2}\log^{-K_4} n.$$

Proof. Fix $1\le s\le T-n$. Let $m_1>m_2$ and $k=\lfloor n/m\rfloor$, where $m=m_1+m_2$. For $1\le j\le k$, define $H_{j,1}=\{(j-1)m+i:1\le i\le m_1\}$ and $H_{j,2}=\{(j-1)m+i:\le m_1+1\le i\le m\}$. Also define $H_*=\{km+1,...,n\}$. Let $W_{j,1}=m_1^{-1/2}\sum_{t\in H_{j,1}}Y_t$, $W_{j,2}=m_2^{-1/2}\sum_{t\in H_{j,2}}Y_t$ and $W_*=\sum_{t\in H_*}Y_t$.

Step 1: bound $\sum_{j=1}^k W_{j,1}$.

We apply Lemma 3 together with a Berbee-type coupling result. By Lemma 7.1 of Chen et al. (2016), there exist independent random variables $\{Z_j\}_{j=1}^k$ (possibly on an extended probability space) such that Z_j and $W_{j,1}$ have the same distribution and

$$\mathbb{P}\left(\bigcup_{j=1}^k\{Z_j\ne W_{j,1}\}\right) \le k\beta(m_2) \le k\tau_1\exp\left(-\tau_2 m_2^{\tau_3}\right). \tag{A.4}$$

Lemma 7.2 of Chen et al. (2016) also implies that there exist constants $M_0,M_1,M_2>0$ such that

$$\mathbb{E}|W_{j,1}|^{p_0}\le M_1M_2^{p_0} \quad\text{and}\quad \mathbb{E}|W_{j,1}|^2\le M_0. \tag{A.5}$$

Let $Q_{m_1,k,p_0}:=\max_{1\le j\le k}\mathbb{E}|W_{j,1}|^{p_0}$. Let $a_T\to\infty$ be a sequence to be chosen later. Applying Lemma 3, we obtain that $\forall t>0$,

$$\mathbb{P}\left(\left|\sum_{j=1}^k Z_j\mathbf{1}\{|Z_j|\le a_T\}\right| \ge \sqrt{k}t + ka_T^{1-p_0}Q_{m_1,k,p_0}\right)$$
$$\le 2\exp\left[-\frac{t^2}{2\left(a_T^{2-p_0}Q_{m_1,k,p_0}+a_Ttk^{-1/2}+k^{-1}\sum_{j=1}^k\mathbb{E}Z_j^2\right)}\right].$$

Hence, by (A.5), we have

$$\mathbb{P}\left(\left|\sum_{j=1}^{k} Z_j \mathbf{1}\{|Z_j| \le a_T\}\right| \ge \sqrt{k}t + k a_T^{1-p_0} M_1 M_2^{p_0}\right) \le 2\exp\left[-\frac{t^2}{2\left(a_T^{2-p_0} M_1 M_2^{p_0} + a_T t k^{-1/2} + M_0\right)}\right].$$

The above display and (A.4) imply that

$$\mathbb{P}\left(\left|\sum_{j=1}^{k} W_{j,1} \mathbf{1}\{|W_{j,1}| \le a_T\}\right| \ge \sqrt{k}t + k a_T^{1-p_0} M_1 M_2^{p_0}\right) \le 2\exp\left[-\frac{t^2}{2\left(a_T^{2-p_0} M_1 M_2^{p_0} + a_T t k^{-1/2} + M_0\right)}\right] + k\tau_1 \exp\left(-\tau_2 m_2^{\tau_3}\right). \quad \text{(A.6)}$$

By (A.5), we have

$$\mathbb{P}\left(\bigcup_{j=1}^{k}\{|W_{j,1}| \ge a_T\}\right) \le \sum_{j=1}^{k} \mathbb{P}\left(|W_{j,1}|^{p_0} \ge a_T^{p_0}\right) \le \sum_{j=1}^{k} \mathbb{E}|W_{j,1}|^{p_0} a_T^{-p_0} \le k a_T^{-p_0} M_1 M_2^{p_0}.$$

By the above two displays,

$$\mathbb{P}\left(\left|\sum_{j=1}^{k} W_{j,1}\right| \ge \sqrt{k}t + k a_T^{1-p_0} M_1 M_2^{p_0}\right) \le 2\exp\left[-\frac{t^2}{2\left(a_T^{2-p_0} M_1 M_2^{p_0} + a_T t k^{-1/2} + M_0\right)}\right] + k\tau_1 \exp\left(-\tau_2 m_2^{\tau_3}\right) + k a_T^{-p_0} M_1 M_2^{p_0}. \quad \text{(A.7)}$$

Step 2: bound $\sum_{j=1}^{k} W_{j,2}$ and W_*

Similar to Step 1, we can show that for any $t > 0$,

$$\mathbb{P}\left(\left|\sum_{j=1}^{k} W_{j,2}\right| \ge \sqrt{k}t + k a_T^{1-p_0} M_1 M_2^{p_0}\right)$$

$$\leq 2\exp\left[-\frac{t^2}{2\left(a_T^{2-p_0}M_1M_2^{p_0}+a_Ttk^{-1/2}+M_0\right)}\right]+k\tau_1\exp\left(-\tau_2 m_1^{\tau_3}\right)+ka_T^{-p_0}M_1M_2^{p_0}. \tag{A.8}$$

Notice that there are fewer than m elements in H_*. Hence, $\forall t>0$,

$$\mathbb{P}\left(|W_*|\geq t\right)\leq \frac{\mathbb{E}|W_*|^{p_0}}{t^{p_1}}\overset{(i)}{\leq} M_1M_2^{p_0}(\sqrt{m}/t)^{p_0}, \tag{A.9}$$

where (i) follows by Lemma 7.2 of Chen et al. (2016) (with the same constants M_1 and M_2 as in (A.5)).

Step 3: derive the final result.

Now we choose $m_2=1+\left\lfloor[(p_0/\tau_2)\log n]^{4+1/\tau_3}\right\rfloor$, $m_1=m_2^2$, $m=m_1+m_2$, $k=\lfloor n/m\rfloor$ and $a_T=\sqrt{k}$. Let

$$\begin{aligned}g_n=&k\tau_1\exp\left(-\tau_2 m_2^{\tau_3}\right)+ka_T^{-p_0}M_1M_2^{p_0}\\&+2\exp\left[-\frac{n/(m_2k)}{2\left(a_T^{2-p_0}M_1M_2^{p_0}+a_T\sqrt{n/(m_2k)}k^{-1/2}+M_0\right)}\right]\\&+k\tau_1\exp\left(-\tau_2 m_1^{\tau_3}\right)+ka_T^{-p_0}M_1M_2^{p_0}+M_1M_2^{p_0}(m/n)^{p_0/2},\end{aligned}$$

For large n and $z\geq 1$, we have that

$$\begin{aligned}&\mathbb{P}\left(\left|\sum_{t=s+1}^{s+n}Y_t\right|\geq 5\sqrt{n}z\right)\\
=&\mathbb{P}\left(\left|\sqrt{m_1}\sum_{j=1}^{k}W_{j,1}+\sqrt{m_2}\sum_{j=1}^{k}W_{j,2}+W_*\right|\geq 5\sqrt{n}z\right)\\
\leq&\mathbb{P}\left(\left|\sqrt{m_1}\sum_{j=1}^{k}W_{j,1}\right|\geq 2\sqrt{n}z\right)+\mathbb{P}\left(\left|\sqrt{m_2}\sum_{j=1}^{k}W_{j,2}\right|\geq 2\sqrt{n}\right)+\mathbb{P}\left(|W_*|\geq\sqrt{n}\right)\\
\overset{(i)}{\leq}&\mathbb{P}\left(\left|\sum_{j=1}^{k}W_{j,1}\right|\geq\sqrt{k}z+ka_T^{1-p_0}M_1M_2^{p_0}\right)\\
&+\mathbb{P}\left(\left|\sum_{j=1}^{k}W_{j,2}\right|\geq\sqrt{k}\sqrt{n/(m_2k)}+ka_T^{1-p_0}M_1M_2^{p_0}\right)+\mathbb{P}\left(|W_*|\geq\sqrt{n}\right)\end{aligned}$$

$$\underbrace{\overset{(ii)}{\leq} 2\exp\left[-\frac{z^2}{2\left(a_T^{2-p_0}M_1M_2^{p_0}+a_Tzk^{-1/2}+M_0\right)}\right]}_{\Psi_n(z)}+g_n,$$

where (i) holds by $z \geq 1 \geq \sqrt{k}a_T^{1-p_0}M_1M_2^{p_0}$ and $n/m_1 \geq k$ and (ii) follows by (A.7), (A.8) and (A.9). By straight-forward computations, we have that $\Psi_n(z) \leq 2\exp(-M_3z^2)$ and $g_n \leq n^{1-p_0/2}\log^{-M_4} n$, where M_3 and M_4 are positive constants. This proves the first claim. The second claim follows by the union bound. □

To study properties of estimation errors of the form $(\sum_{s=1}^n X_s\varepsilon_{s+1})/(\sum_{s=1}^n X_s^2)$, we consider the following condition.

Condition 1. Let $\{X_s, e_{s+1}\}_{s=1}^n$ be random variables with *beta*-mixing coefficient satisfying $\mathcal{B}_{mix}(i) \leq \tau_1\exp(-\tau_2 i^{\tau_3})$ for some constants $\tau_1,\tau_2,\tau_3 > 0$. Suppose that $\mathbb{E}X_se_{s+1} = 0$, $\max_{1\leq s\leq n}\mathbb{E}|X_s|^p \leq D$ and $\max_{1\leq s\leq n}\mathbb{E}|e_{s+1}|^p \leq D$ for some constants $p > 4$ and $D > 0$. Moreover, $D_0 \leq \mathbb{E}(n^{-1/2}\sum_{s=1}^n X_se_{s+1})^2 \leq D_1$ and $D_2 \leq \mathbb{E}X_s^2 \leq D_3$ for some constants $D_0, .., D_3 > 0$.

Lemma 5. *Let Condition 1 hold. Define*

$$\delta = \frac{\sum_{s=1}^n X_se_{s+1}}{\sum_{s=1}^n X_s^2}.$$

Then for any $p_0 \in (2, p/2)$, there exist $\tilde{\delta} \in \sigma(\{X_s, e_{s+1}\}_{s=1}^n)$ and constants $C_1, ..., C_5 > 0$ such that $\mathbb{P}(\tilde{\delta} \neq \delta) \leq C_1n^{1-p_0/2}\log^{-C_2} n$, $|\mathbb{E}\tilde{\delta}| \leq n^{-1}\sqrt{\log n}C_3$ and $n^{-1}C_4 \leq \mathbb{E}\tilde{\delta}^2 \leq n^{-1}C_5$. Moreover, $|\mathbb{E}\tilde{\delta}^2 - \mathbb{E}(\sum_{s=1}^n X_se_{s+1})^2/\mathbb{E}(\sum_{s=1}^n X_s^2)^2| \leq C_6\sqrt{n^{-3}\log n}$ for some constant $C_6 > 0$.

Proof. Let $Z_{n,1} = n^{-1}\sum_{s=1}^n X_se_{s+1}$ and $Z_{n,2} = n^{-1}\sum_{s=1}^n X_s^2$. Hence, $\delta = Z_{n,1}/Z_{n,2}$. The proof proceeds in two steps.

Step 1: bound $\mathbb{E}|Z_{n,1}|$.

By Davydov's inequality (Corollary 16.2.4 of Athreya and Lahiri (2006)) and the uniform boundedness of $\mathbb{E}|X_se_{s+1}|^{2+c}$ for some $c > 0$, we have that for $|s_1 - s_2| > 1$, $|\mathbb{E}X_{s_1}X_{s_2}e_{s_1+1}e_{s_2+1}| \leq M_1[\beta(|s_1 - s_2 - 1|)]^{M_2}$ for some constants $M_1, M_2 > 0$. The exponential-decay of the beta-mixing coefficient implies that for $|s_1 - s_2| > 1$,

$$|\mathbb{E}X_{s_1}X_{s_2}e_{s_1+1}e_{s_2+1}| \leq M_3\exp\left(-M_4|s_1 - s_2 - 1|^{M_5}\right), \tag{A.10}$$

where $M_3, M_4, M_5 > 0$ are constants. Let $S = \{(s_1, s_2) : 1 \leq s_1, s_2 \leq n,\ |s_1 - s_2| > 1\}$. Let $M_6 > 0$ be a constant such that $|\mathbb{E}X_{s_1}X_{s_2}e_{s_1+1}e_{s_2+1}| \leq M_6$; such a constant exists since $\mathbb{E}|X_s e_{s+1}|^2 \leq (\mathbb{E}|X_s|^4 \mathbb{E}|e_{s+1}|^4)^{1/2}$ is uniformly bounded. Notice that

$$
\begin{aligned}
\mathbb{E}|Z_{n,1}|^2 =& n^{-2}\sum_{s_1=1}^{n}\sum_{s_2=1}^{n}\mathbb{E}X_{s_1}X_{s_2}e_{s_1+1}e_{s_2+1} \\
=& n^{-2}\sum_{s=1}^{n}\mathbb{E}X_s^2 e_{s+1}^2 + n^{-2}\sum_{s=1}^{n-1}\mathbb{E}X_s X_{s+1}e_{s+1}e_{s+2} \\
&+ n^{-2}\sum_{s=2}^{n-1}\mathbb{E}X_s X_{s-1}e_{s+1}e_s + n^{-2}\sum_{(s_1,s_2)\in S}\mathbb{E}X_{s_1}X_{s_2}e_{s_1+1}e_{s_2+1} \\
\leq& n^{-1}M_6 + 2n^{-2}(n-1)M_6 + n^{-2}\sum_{(s_1,s_2)\in S}\mathbb{E}X_{s_1}X_{s_2}e_{s_1+1}e_{s_2+1} \\
\overset{(i)}{\leq}& n^{-1}M_6 + 2n^{-2}(n-1)M_6 + n^{-2}\sum_{(s_1,s_2)\in S} M_3\exp\left(-M_4|s_1-s_2-1|^{M_5}\right) \\
\leq& n^{-1}M_6 + 2n^{-2}(n-1)M_6 + n^{-2}\sum_{s_1=1}^{n}\sum_{s_2=1}^{\infty} M_3\exp\left(-M_4|s_1-s_2-1|^{M_5}\right) \\
\overset{(ii)}{\leq}& n^{-1}M_6 + 2n^{-2}(n-1)M_6 + n^{-2}nM_7
\end{aligned}
\tag{A.11}
$$

for some constant $M_7 > 0$, where (i) follows by (A.10) and (ii) follows by the fact that $\sum_{s_2=1}^{\infty} M_3\exp(-M_4|s_1-s_2-1|^{M_5})$ is uniformly bounded for any $1 \leq s_1 \leq n$. Hence, there exists a constant $M_8 > 0$ such that

$$
\mathbb{E}|Z_{n,1}| \leq \sqrt{\mathbb{E}|Z_{n,1}|^2} \leq M_8 n^{-1/2}. \tag{A.12}
$$

Step 2: derive the desired result.

Notice that $X_s^2 - \mathbb{E}X_s^2$ has uniformly bounded $0.5p$-th moments. By Lemma 4 (applied with $Y_s = X_s^2 - \mathbb{E}X_s^2$), we have that

$$
\mathbb{P}\left(|Z_{n,2} - \mathbb{E}Z_{n,2}| \geq K_1\sqrt{n^{-1}\log n}\right) \leq K_2 n^{1-p_0/2}\log^{-K_3} n,
$$

where $K_1, K_2, K_3 > 0$ are constants. Let $\bar{\delta} = Z_{n,1}/\mathbb{E}Z_{n,2}$ and $\tilde{\delta} = Z_{n,1}/\tilde{Z}_{n,2}$ with

$$\tilde{Z}_{n,2} = \begin{cases} \mathbb{E}Z_{n,2} + K_1\sqrt{n^{-1}\log n} & \text{if } Z_{n,2} \geq \mathbb{E}Z_{n,2} + K_1\sqrt{n^{-1}\log n} \\ \mathbb{E}Z_{n,2} - K_1\sqrt{n^{-1}\log n} & \text{if } Z_{n,2} \leq \mathbb{E}Z_{n,2} - K_1\sqrt{n^{-1}\log n} \\ Z_{n,2} & \text{otherwise} \end{cases}.$$

Clearly, $\tilde{\delta} \in \sigma\left(\{X_s, e_{s+1}\}_{s=1}^n\right)$. Moreover,

$$\mathbb{P}\left(\delta \neq \tilde{\delta}\right) = \mathbb{P}\left(Z_{n,2} \neq \tilde{Z}_{n,2}\right) \leq K_2 n^{1-p_0/2}\log^{-K_3} n. \tag{A.13}$$

Notice that

$$\begin{aligned}
\mathbb{E}|\tilde{\delta} - \bar{\delta}| = \mathbb{E}\left|\frac{Z_{n,1}}{\tilde{Z}_{n,2}} - \frac{Z_{n,1}}{\mathbb{E}Z_{n,2}}\right| &= \mathbb{E}\left|\frac{Z_{n,1}(\tilde{Z}_{n,2} - \mathbb{E}Z_{n,2})}{\tilde{Z}_{n,2}\mathbb{E}Z_{n,2}}\right| \\
&\overset{(i)}{\leq} \frac{K_1\sqrt{n^{-1}\log n}}{(\mathbb{E}Z_{n,2})\left(\mathbb{E}Z_{n,2} + K_1\sqrt{n^{-1}\log n}\right)}\mathbb{E}\left|Z_{n,1}\right| \\
&\overset{(ii)}{\leq} K_4 n^{-1}\sqrt{\log n} \qquad \text{for some constant } K_4 > 0,
\end{aligned}$$

where (i) holds by $|\tilde{Z}_{n,2} - \mathbb{E}Z_{n,2}| \leq K_1\sqrt{n^{-1}\log n}$ (by the definition of $\tilde{Z}_{n,2}$) and (ii) follows by (A.12) and $\mathbb{E}Z_{n,2} \geq D_2$. Since $\mathbb{E}X_s e_{s+1} = 0$, we have $\mathbb{E}Z_{n,1} = 0$ and $\mathbb{E}\bar{\delta} = 0$. Hence, the above display implies that

$$|\mathbb{E}\tilde{\delta}| \leq |\mathbb{E}\bar{\delta}| + \mathbb{E}|\tilde{\delta} - \bar{\delta}| \leq K_4 n^{-1}\sqrt{\log n}. \tag{A.14}$$

Lastly, notice that

$$\begin{aligned}
\left|\mathbb{E}(\tilde{\delta}^2 - \bar{\delta}^2)\right| &= \mathbb{E}\left[\frac{Z_{n,1}^2|\tilde{Z}_{n,2} - \mathbb{E}Z_{n,2}| \cdot |\tilde{Z}_{n,2} + \mathbb{E}Z_{n,2}|}{\tilde{Z}_{n,2}^2\left(\mathbb{E}Z_{n,2}\right)^2}\right] \\
&\overset{(i)}{\leq} \left\{\frac{K_1\sqrt{n^{-1}\log n}(2\mathbb{E}Z_{n,2} + K_1\sqrt{n^{-1}\log n})}{\left(\mathbb{E}Z_{n,2}\right)^2\left(\mathbb{E}Z_{n,2} - K_1\sqrt{n^{-1}\log n}\right)^2}\right\}\mathbb{E}Z_{n,1}^2 \overset{(ii)}{\leq} K_5\sqrt{n^{-3}\log n}
\end{aligned} \tag{A.15}$$

for some constant $K_5 > 0$, where (i) follows by $|\tilde{Z}_{n,2} - \mathbb{E}Z_{n,2}| \leq K_1\sqrt{n^{-1}\log n}$ (by the definition of $\tilde{Z}_{n,2}$) and (ii) follows by (A.11) and $\mathbb{E}Z_{n,2} \geq D_2$. Since $\mathbb{E}\bar{\delta}^2 =$

$\mathbb{E}Z_{n,1}^2/(\mathbb{E}Z_{n,2})^2$, $D_2 \le \mathbb{E}Z_{n,2} \le D_3$ and $n^{-1}D_0 \le \mathbb{E}Z_{n,1}^2 \le n^{-1}D_1$, it follows, by (A.15), that there exist constants $K_6, K_7 > 0$ such that for large n,

$$n^{-1}K_6 \le \mathbb{E}\bar{\delta}^2 - \left|\mathbb{E}(\tilde{\delta}^2 - \bar{\delta}^2)\right| \le \mathbb{E}\tilde{\delta}^2 \le \mathbb{E}\bar{\delta}^2 + \left|\mathbb{E}(\tilde{\delta}^2 - \bar{\delta}^2)\right| \le n^{-1}K_7.$$

The desired result follows by (A.13), (A.14) and the above display. □

Lemma 6. *Let Condition 1 hold. Define*

$$\delta = \frac{\sum_{s=1}^{n} X_s e_{s+1}}{\sum_{s=1}^{n} X_s^2} \qquad \text{and} \qquad \bar{\delta} = \frac{\sum_{s=1}^{n-a_n} X_s e_{s+1}}{\sum_{s=1}^{n-a_n} X_s^2},$$

where $a_n \le cn$ for some constant $c \in (0,1)$. Then for any $p_0 \in (2, p/2)$, there exists a constant $M > 0$ such that $\forall x > 0$,

$$\mathbb{P}\left(|\bar{\delta} - \delta| \ge x\right) \le M \max\left\{n^{1-p_0/2}\log^{-K_3} n,\ (nx/\sqrt{a_n})^{-p_0},\ \left(a_n^{-1}n^{3/2}x/\sqrt{\log n}\right)^{-p/2}\right\}.$$

Proof. After straightforward computations, we have that

$$\delta - \bar{\delta} = \underbrace{\frac{\sum_{s=n-a_n+1}^{n} X_s e_{s+1}}{\sum_{s=1}^{n} X_s^2}}_{J_1} - \underbrace{\frac{\left(\sum_{s=1}^{n-a_n} X_s e_{s+1}\right)\left(\sum_{s=n-a_n+1}^{n} X_s^2\right)}{\left(\sum_{s=1}^{n} X_s^2\right)\left(\sum_{s=1}^{n-a_n} X_s^2\right)}}_{J_2}. \tag{A.16}$$

Notice that by Lemma 2, both $X_s^2 - \mathbb{E}X_s^2$ and $X_s e_{s+1}$ has uniformly bounded $0.5p$-th moments. Applying Lemma 4 (with $Y_s = X_s^2 - \mathbb{E}X_s^2$ and $Y_s = X_s e_{s+1}$) and using $(1-c)n \le n - a_n \le n$, we have that for some constants $K_1, K_2, K_3 > 0$,

$$\begin{cases} \mathbb{P}\left(|\sum_{s=1}^{n}(X_s^2 - \mathbb{E}X_s^2)| \ge K_1\sqrt{n\log n}\right) \le K_2 n^{1-p_0/2}\log^{-K_3} n \\ \mathbb{P}\left(\left|\sum_{s=1}^{n-a_n}(X_s^2 - \mathbb{E}X_s^2)\right| \ge K_1\sqrt{n\log n}\right) \le K_2 n^{1-p_0/2}\log^{-K_3} n \\ \mathbb{P}\left(\left|\sum_{s=1}^{n-a_n}(X_s^2 - \mathbb{E}X_s^2)\right| \ge K_1\sqrt{n\log n}\right) \le K_2 n^{1-p_0/2}\log^{-K_3} n \\ \mathbb{P}\left(\left|\sum_{s=1}^{n-a_n} X_s e_{s+1}\right| \ge K_1\sqrt{n\log n}\right) \le K_2 n^{1-p_0/2}\log^{-K_3} n. \end{cases}$$

By Condition 1, $\mathbb{E}X_s^2 \ge D_2$. Since $K_1\sqrt{n^{-1}\log n} < D_2/2$ for large n, we have $\max\left\{\mathbb{P}\left(\sum_{s=1}^{n} X_s^2 \le nD_2/2\right), \mathbb{P}\left(\sum_{s=1}^{n-a_n} X_s^2 \le nD_2/2\right)\right\} \le K_4 n^{1-p_0/2}\log^{-K_3} n$ for some constant $K_4 \geq K_2$. Notice that

$$\begin{aligned}
\mathbb{P}\left(J_1 \geq x/2\right) &\leq \mathbb{P}\left(\sum_{s=1}^{n} X_s^2 \leq nD_2/2\right) + \mathbb{P}\left(\left|\sum_{s=n-a_n+1}^{n} X_s e_{s+1}\right| \geq D_2 nx/4\right) \\
&\leq \mathbb{P}\left(\sum_{s=1}^{n} X_s^2 \leq nD_2/2\right) + \left(D_2 nx/4\right)^{-p_0} \mathbb{E}\left|\sum_{s=n-a_n+1}^{n} X_s e_{s+1}\right|^{p_0} \\
&\overset{(i)}{\leq} \mathbb{P}\left(\sum_{s=1}^{n} X_s^2 \leq nD_2/2\right) + \left(D_2 nx/4\right)^{-p_0} K_5 a_n^{p_0/2} \text{ for a constant } K_5 > 0 \\
&\leq K_4 n^{1-p_0/2} \log^{-K_3} n + K_5 \left(D_2 nx/4\right)^{-p_0} a_n^{p_0/2},
\end{aligned}$$

where (i) holds by Lemma 7.2 of Chen et al. (2016). Also notice that

$$\begin{aligned}
&\mathbb{P}\left(J_2 \geq x/2\right) \\
&\leq \mathbb{P}\left(\sum_{s=1}^{n-a_n} X_s^2 \leq nD_2/2\right) + \mathbb{P}\left(\sum_{s=1}^{n} X_s^2 \leq nD_2/2\right) \\
&\quad + \mathbb{P}\left(\left|\sum_{s=1}^{n-a_n} X_s e_{s+1}\right| \geq K_1\sqrt{n\log n}\right) + \mathbb{P}\left(\left|\sum_{s=n-a_n+1}^{n} X_s^2\right| \geq \frac{D_2^2 n^2 x}{8K_1\sqrt{n\log n}}\right) \\
&\leq 3K_4 n^{1-p_0/2} \log^{-K_3} n + \mathbb{P}\left(\left|\sum_{s=n-a_n+1}^{n} X_s^2\right| \geq \frac{D_2^2 n^2 x}{8K_1\sqrt{n\log n}}\right) \\
&\leq 3K_4 n^{1-p_0/2} \log^{-K_3} n + \left(\frac{D_2^2 n^2 x}{8K_1\sqrt{n\log n}}\right)^{-p/2} \mathbb{E}\left|\sum_{s=n-a_n+1}^{n} X_s^2\right|^{p/2} \\
&\overset{(i)}{\leq} 3K_4 n^{1-p_0/2} \log^{-K_3} n + \left(\frac{D_2^2 n^2 x}{8K_1\sqrt{n\log n}}\right)^{-p/2} (a_n D)^{p/2},
\end{aligned}$$

where (i) follows by $\mathbb{E}|\sum_{s=n-a_n+1}^{n} X_s^2|^{p/2} = \|\sum_{s=n-a_n+1}^{n} X_s^2\|_{L^{p/2}(\mathbb{P})}^{p/2} \leq (\sum_{s=n-a_n+1}^{n} \|X_s^2\|_{L^{p/2}(\mathbb{P})})^{p/2} \leq (a_n D)^{p/2}$. The desired result follows by combining (A.16) with the above two displays. □

Appendix B Proofs of main results

For notational simplicity, we omit the n subscript. For example, we write $\delta_{j,t}$ rather than $\delta_{j,n,t}$ for $j \in \{1, 2\}$.

B.1 Proof of Proposition 1

Our proof of Proposition 1 relies on two lemmas, Lemma 7 and 8. We first state and prove these lemmas before proving the proposition.

Lemma 7. *Let Assumption 1 hold. For any constants $K > 0$, $h \in (0,1)$ and $p_1 \in (2, r/2)$, we can enlarge the probability space and construct random variables $\delta_{1,t,*}$, $\delta_{2,t,*}$, $\bar{\delta}_{1,t}$ and $\bar{\delta}_{2,t}$ such that for $j \in \{1,2\}$,*

$$\begin{cases} \mathbb{P}\left(\delta_{j,t,*} \neq \delta_{j,t}\right) \leq C_1 T^{-\min\{p_1/2-1,\ (1-h)p_1\}} \\ \mathbb{P}\left(|\delta_{j,t,*} - \bar{\delta}_{j,t}| \leq KT^{-h}\right) = 1 \\ \bar{\delta}_{1,t} \text{ and } \bar{\delta}_{2,t} \text{ are independent of } \{x_{1,s}, x_{2,s}, \varepsilon_{s+1}\}_{s\geq t-1} \\ |\mathbb{E}\bar{\delta}_{j,t}| \leq T^{-1}\sqrt{\log T} C_2 \text{ and } T^{-1}C_3 \leq \mathbb{E}\bar{\delta}_{j,t}^2 \leq T^{-1}C_4 \\ \left|\mathbb{E}\bar{\delta}_{j,t}^2 - \mathbb{E}[\sum_{s=t-n}^{t-1} x_{1,t}(\varepsilon_{t+1} + \beta_2 x_{2,t})]^2 / \mathbb{E}(\sum_{s=t-n}^{t-1} x_{1,t}^2)^2\right| \leq C_5\sqrt{n^{-3}\log n}, \end{cases}$$

where $C_1, ..., C_5 > 0$ are constants depending only on the constants in Assumption 1.

Proof. We construct $\delta_{j,t,*}$ for $j = 1$; the case for $j = 2$ is analogous. Notice that

$$\delta_{1,t} = \frac{\sum_{s=t-n}^{t-1} x_{1,t}(\varepsilon_{t+1} + \beta_2 x_{2,t})}{\sum_{s=t-n}^{t-1} x_{1,t}^2}.$$

Recall the constants in Assumption 1. Let $a_n = \min\{a \in \mathbb{N} \mid a \geq (r/2 - 1)^{1/c} \log^{1/c} n\}$.

By Theorem 16.2.1 of Athreya and Lahiri (2006), we can extend the probability space with random variables $\{\dot{x}_{1,s}, \dot{x}_{2,s}, \dot{\varepsilon}_{s+1}\}_{s=t-n}^{t-a_n-1}$ such that

$$\begin{cases} \{\dot{x}_{1,s}, \dot{x}_{2,s}, \dot{\varepsilon}_{s+1}\}_{s=t-n}^{t-a_n-1} \text{ has the same distribution as } \{x_{1,s}, x_{2,s}, \varepsilon_{s+1}\}_{s=t-n}^{t-a_n-1} \\ \{\dot{x}_{1,s}, \dot{x}_{2,s}, \dot{\varepsilon}_{s+1}\}_{s=t-n}^{t-a_n-1} \text{ is independent of } \{x_{1,s}, x_{2,s}, \varepsilon_{s+1}\}_{s\geq t-1} \\ \mathbb{P}\left(\{\dot{x}_{1,s}, \dot{x}_{2,s}, \dot{\varepsilon}_{s+1}\}_{s=t-n}^{t-a_n-1} \neq \{x_{1,s}, x_{2,s}, \varepsilon_{s+1}\}_{s=t-n}^{t-a_n-1}\right) = \beta(a_n) \leq b\exp(-a_n^c). \end{cases} \quad \text{(B.1)}$$

Let $\mathcal{F}_n$ be the σ-algebra generated by $\{\dot{x}_{1,s}, \dot{x}_{2,s}, \dot{\varepsilon}_{s+1}\}_{s=t-n}^{t-a_n-1}$. Hence, $\dot{\delta}_{1,t} \in \mathcal{F}_n$ and

$$\mathbb{P}\left(\dot{\delta}_{1,t} \neq \ddot{\delta}_{1,t}\right) \leq b\exp(-a_n^c), \quad \text{(B.2)}$$

where

$$\dot{\delta}_{1,t} = \frac{\sum_{s=t-n}^{t-a_n-1} \dot{x}_{1,t}(\dot{\varepsilon}_{t+1} + \beta_2 \dot{x}_{2,t})}{\sum_{s=t-n}^{t-a_n-1} \dot{x}_{1,t}^2} \text{ and } \ddot{\delta}_{1,t} = \frac{\sum_{s=t-n}^{t-a_n-1} x_{1,t}(\varepsilon_{t+1} + \beta_2 x_{2,t})}{\sum_{s=t-n}^{t-a_n-1} x_{1,t}^2}.$$

Now we apply Lemma 5 (with $X_s = \dot{x}_{1,s}$ and $e_{s+1} = \beta_2 \dot{x}_{2,s} + \dot{\varepsilon}_{s+1}$) and obtain that there exist $\bar{\delta}_{1,t} \in \mathcal{F}_n$ satisfying

$$\begin{cases} \mathbb{P}\left(\dot{\delta}_{1,t} \neq \bar{\delta}_{1,t}\right) \leq M_0 n^{1-p_1/2} \log^{-M_1} n \\ |\mathbb{E}\bar{\delta}_{1,t}| \leq n^{-1}\sqrt{\log n} M_2 \\ n^{-1} M_3 \leq \mathbb{E}\bar{\delta}_{1,t}^2 \leq n^{-1} M_4 \end{cases} \tag{B.3}$$

where $M_0, ..., M_4 > 0$ are constants. Lemma 5 also implies that $|\mathbb{E}\bar{\delta}_{1,t}^2 - \mathbb{E}[\sum_{s=t-n}^{t-a_n-1} \dot{x}_{1,t}(\dot{\varepsilon}_{t+1} + \beta_2 \dot{x}_{2,t})]^2 / \mathbb{E}(\sum_{s=t-n}^{t-a_n-1} \dot{x}_{1,t}^2)^2| \leq G\sqrt{(n-a_n)^{-3}\log(n-a_n)}$ for some constant $G > 0$. Notice that $a_n \asymp \log^{1/c} n$. In computing this expectation, we can replace $\{\dot{x}_{1,s}, \dot{x}_{2,s}, \dot{\varepsilon}_{s+1}\}_{s=t-n}^{t-a_n-1}$ with $\{x_{1,s}, x_{2,s}, \varepsilon_{s+1}\}_{s=t-n}^{t-a_n-1}$ since they have the same distribution. It is not hard to verify that $|\mathbb{E}\bar{\delta}_{1,t}^2 - \mathbb{E}[\sum_{s=t-n}^{t-1} x_{1,t}(\varepsilon_{t+1} + \beta_2 x_{2,t})]^2 / \mathbb{E}(\sum_{s=t-n}^{t-1} x_{1,t}^2)^2| \leq G\sqrt{n^{-3}\log n}$ for some constant $G' > 0$.

Since $\bar{\delta}_{1,t} \in \mathcal{F}_n$, (B.1) implies that

$$\bar{\delta}_{1,t} \text{ is independent of } \{x_{1,s}, x_{2,s}, \varepsilon_{s+1}\}_{s \geq t-1}. \tag{B.4}$$

By Lemma 6 (applied with $X_s = x_{1,s}$ and $e_{s+1} = \beta_2 x_{2,s} + \varepsilon_{s+1}$), we have that for $x = Kn^{-h}$,

$$\begin{aligned} &\mathbb{P}\left(|\ddot{\delta}_{1,t} - \delta_{1,t}| \geq x\right) \\ &\quad \leq M_5 \max\left\{n^{1-p_1/2}\log^{-K_3} n,\ (nx/\sqrt{a_n})^{-p_1},\ \left(a_n^{-1} n^{3/2} x/\sqrt{\log n}\right)^{-r/2}\right\}, \end{aligned} \tag{B.5}$$

where $M_5 > 0$ is a constant. Define

$$\delta_{1,t,*} = \begin{cases} \bar{\delta}_{1,t} + x & \text{if } \delta_{1,t} \geq \bar{\delta}_{1,t} + x \\ \bar{\delta}_{1,t} - x & \text{if } \delta_{1,t} \leq \bar{\delta}_{1,t} - x \\ \delta_{1,t} & \text{otherwise} \end{cases}.$$

Hence,

$$|\delta_{1,t,*} - \bar{\delta}_{1,t}| \leq x. \tag{B.6}$$

Notice that, for some constant $M_6 > 0$, we have

$$\begin{aligned}
&\mathbb{P}\left(\delta_{1,t,*} \neq \delta_{1,t}\right) \\
&\leq \mathbb{P}\left(|\delta_{1,t} - \ddot{\delta}_{1,t}| \geq x\right) + \mathbb{P}\left(\ddot{\delta}_{1,t} \neq \dot{\delta}_{1,t}\right) + \mathbb{P}\left(\dot{\delta}_{1,t} \neq \bar{\delta}_{1,t}\right) \\
&\overset{(i)}{\leq} M_6 \max\left\{n^{1-p_1/2}\log^{-K_3} n,\ (nx/\sqrt{a_n})^{-p_1},\ \left(a_n^{-1} n^{3/2} x/\sqrt{\log n}\right)^{-r/2},\ \exp(-a_n^c)\right\} \\
&\overset{(ii)}{\leq} M_6 \max\left\{n^{1-p_1/2},\ (nx)^{-p_1},\ \left(n^{3/2}x\right)^{-r/2}\right\} \\
&\overset{(iii)}{\leq} M_6 \max\left\{n^{1-p_1/2},\ (nx)^{-p_1}\right\} \leq M_7 n^{-\min\{p_1/2-1,\ (1-h)p_1\}} \text{ for some constant } M_7 > 0,
\end{aligned}$$

where (i) holds by (B.2), (B.3) and (B.5), (ii) follows by the fact that $\exp(-a_n^c) \leq n^{1-r/2} < n^{1-p_1/2}$ and (iii) follows by $(n^{3/2}x)^{-r/2} < (n^{3/2}x)^{-p_1} < (nx)^{-p_1}$ (due to $p_1 < r/2$).

Since $n \asymp T$, the desired result follows by the above display, (B.6), (B.4) and (B.3). □

Lemma 8. *Let Assumption 1 hold. Then $\forall p_1 \in (2, r/2)$ and $\forall h \in (2\alpha_{x,2}, 1)$, there exist constants $G_1, ..., G_4 > 0$ and an array of random variables $\{\Delta L_{t+1,*}\}_{t=n}^{T-n}$ such that for $T \geq G_1$,*

$$\mathbb{P}\left(\bigcap_{t=n}^{T-1} \{\Delta L_{t+1,*} = \Delta L_{t+1}\}\right) \geq 1 - G_2 T^{1-\min\{p_1/2-1,\ (1-h)p_1\}}$$

and

$$G_3 T^{-2\alpha_{x,2}} \leq \mathbb{E}\Delta L_{t+1,*} \leq G_4 T^{-2\alpha_{x,2}}.$$

Proof. Let $\delta_{j,t} = \hat{\beta}_{j,t} - \beta_{j,t}$. Recall from (3.3) that

$$\begin{aligned}
\Delta L_{t+1} &= 2\varepsilon_{t+1}\left(\beta_2 x_{2,t} - \beta_1 x_{1,t} - \delta_{1,t}x_{1,t} + \delta_{2,t}x_{2,t}\right) \\
&\quad + \left(\beta_2 x_{2,t} - \beta_1 x_{1,t} - \delta_{1,t}x_{1,t} + \delta_{2,t}x_{2,t}\right)\left(\beta_1 x_{1,t} + \beta_2 x_{2,t} - \delta_{1,t}x_{1,t} - \delta_{2,t}x_{2,t}\right).
\end{aligned} \tag{B.7}$$

Let $\delta_{j,t,*}$ and $\bar{\delta}_{j,t}$ be defined as in the statement of Lemma 7 (with $K = 1$):

$$\begin{cases} \mathbb{P}\left(\delta_{j,t,*} \neq \delta_{j,t}\right) \leq C_1 T^{-\min\{p_1/2-1,\ (1-h)p_1\}} \\ \mathbb{P}\left(|\delta_{j,t,*} - \bar{\delta}_{j,t}| \leq T^{-h}\right) = 1 \\ \bar{\delta}_{1,t} \text{ and } \bar{\delta}_{2,t} \text{ are independent of } \{x_{1,s}, x_{2,s}, \varepsilon_{s+1}\}_{s \geq t-1} \\ |\mathbb{E}\bar{\delta}_{j,t}| \leq T^{-1}\sqrt{\log T} C_2 \text{ and } T^{-1}C_3 \leq \mathbb{E}\bar{\delta}_{j,t}^2 \leq T^{-1}C_4, \end{cases} \tag{B.8}$$

where $C_1, C_2, C_3, C_4 > 0$ are constants. Define

$$\Delta L_{t+1,*} = 2\varepsilon_{t+1}\left(\beta_2 x_{2,t} - \beta_1 x_{1,t} - \delta_{1,t,*}x_{1,t} + \delta_{2,t,*}x_{2,t}\right) \\ + \left(\beta_2 x_{2,t} - \beta_1 x_{1,t} - \delta_{1,t,*}x_{1,t} + \delta_{2,t,*}x_{2,t}\right)\left(\beta_1 x_{1,t} + \beta_2 x_{2,t} - \delta_{1,t,*}x_{1,t} - \delta_{2,t,*}x_{2,t}\right). \tag{B.9}$$

The first statement in (B.8) implies that

$$\mathbb{P}\left(\bigcap_{t=n}^{T-1}\{\Delta L_{t+1,*} = \Delta L_{t+1}\}\right) \geq 1 - \sum_{t=n}^{T-1}\mathbb{P}\left(\delta_{1,t} \neq \delta_{1,t,*} \text{ or } \delta_{2,t} \neq \delta_{2,t,*}\right) \\ \geq 1 - 2C_1(T-n)T^{-\min\{p_1/2-1,\ (1-h)p_1\}}. \tag{B.10}$$

Since $T - n < T$, the first claim follows.

Now we compute $\mathbb{E}\Delta L_{t+1,*}$. Notice that there exist a constant $K_1 > 0$ such that for $j_1, j_2 \in \{1, 2\}$,

$$|\mathbb{E}\delta_{j_1,t,*}x_{j_1,t}\varepsilon_{t+1}| \leq |\mathbb{E}\bar{\delta}_{j_1,t}x_{j_1,t}\varepsilon_{t+1}| + \mathbb{E}|(\delta_{j_1,t,*} - \bar{\delta}_{j_1,t})x_{j_1,t}\varepsilon_{t+1}| \\ \overset{(i)}{=} \mathbb{E}|(\delta_{j_1,t,*} - \bar{\delta}_{j_1,t})x_{j_1,t}\varepsilon_{t+1}| \overset{(ii)}{\leq} T^{-h}\mathbb{E}|x_{j_1,t}\varepsilon_{t+1}| \leq K_1 T^{-h}, \tag{B.11}$$

where (i) follows by the fact that $\bar{\delta}_{j_1,t}$ is independent of $x_{j_1,t}\varepsilon_{t+1}$ and $\mathbb{E}x_{j_1,t}\varepsilon_{t+1} = 0$ and (ii) follows by second statement in (B.8). Similarly, we have that for some constants $K_2, K_3, K_4 > 0$,

$$|\mathbb{E}\delta_{j_1,t,*}x_{j_1,t}x_{j_2,t}| \leq |\mathbb{E}\bar{\delta}_{j_1,t}x_{j_1,t}x_{j_2,t}| + \mathbb{E}|(\delta_{j_1,t,*} - \bar{\delta}_{j_1,t})x_{j_1,t}x_{j_2,t}| \\ \overset{(i)}{=} |\mathbb{E}\bar{\delta}_{j_1,t}| \cdot |\mathbb{E}x_{j_1,t}x_{j_2,t}| + \mathbb{E}|(\delta_{j_1,t,*} - \bar{\delta}_{j_1,t})x_{j_1,t}x_{j_2,t}| \\ \overset{(ii)}{\leq} T^{-1}\sqrt{\log T}C_2K_2 + T^{-h}K_3 \overset{(iii)}{\leq} T^{-h}K_4, \tag{B.12}$$

where (i) follows by the independence between $\bar{\delta}_{j_1,t}$ and $\{x_{j_1}, x_{j_2}\}$, (ii) follows by the second and third statements in (B.8) and (iii) follows by $h < 1$. Moreover, we have that for constants $K_5 \geq \mathbb{E}|x_{j_1,t}x_{j_2,t}|$ and K_6 large enough,

$$
\begin{aligned}
&|\mathbb{E}\delta_{j_1,t,*}\delta_{j_2,t,*}x_{j_1,t}x_{j_2,t}| \\
&\leq |\mathbb{E}\bar{\delta}_{j_1,t}\bar{\delta}_{j_2,t}x_{j_1,t}x_{j_2,t}| + \mathbb{E}|(\delta_{j_1,t,*} - \bar{\delta}_{j_1,t})\delta_{j_2,t,*}x_{j_1,t}x_{j_2,t}| + \mathbb{E}|(\delta_{j_2,t,*} - \bar{\delta}_{j_2,t})\bar{\delta}_{j_1,t}x_{j_1,t}x_{j_2,t}| \\
&\overset{(i)}{\leq} |\mathbb{E}\bar{\delta}_{j_1,t}\bar{\delta}_{j_2,t}| \cdot \mathbb{E}|x_{j_1,t}x_{j_2,t}| + T^{-h}\mathbb{E}|\delta_{j_2,t,*}x_{j_1,t}x_{j_2,t}| + T^{-h}\mathbb{E}|\delta_{j_1,t,*}x_{j_1,t}x_{j_2,t}| \\
&\overset{(ii)}{\leq} |\mathbb{E}\bar{\delta}_{j_1,t}\bar{\delta}_{j_2,t}| \cdot \mathbb{E}|x_{j_1,t}x_{j_2,t}| + T^{-h}\mathbb{E}\left[(|\bar{\delta}_{j_2,t}| + T^{-h})|x_{j_1,t}x_{j_2,t}|\right] \\
&\quad + T^{-h}\mathbb{E}\left[(|\bar{\delta}_{j_1,t}| + T^{-h})|x_{j_1,t}x_{j_2,t}|\right] \\
&\overset{(iii)}{\leq} |\mathbb{E}\bar{\delta}_{j_1,t}\bar{\delta}_{j_2,t}| \cdot \mathbb{E}|x_{j_1,t}x_{j_2,t}| + T^{-h}\mathbb{E}(|\bar{\delta}_{j_2,t}| + T^{-h}) \cdot \mathbb{E}|x_{j_1,t}x_{j_2,t}| \\
&\quad + T^{-h}\mathbb{E}(|\bar{\delta}_{j_1,t}| + T^{-h}) \cdot \mathbb{E}|x_{j_1,t}x_{j_2,t}| \\
&\overset{(iv)}{\leq} T^{-1}C_4K_5 + 2T^{-h}(T^{-1/2}C_4^{1/2} + T^{-h})K_5 \\
&\leq K_6 \max\left\{T^{-h-1/2},\ T^{-2h}\right\},
\end{aligned} \tag{B.13}
$$

where (i), (ii) and (iii) follow by computations based on the independence between $\{\bar{\delta}_{j_1,t}, \bar{\delta}_{j_2,t}\}$ and $\{x_{j_1,t}, x_{j_2,t}\}$ and $|\delta_{j,t,*} - \bar{\delta}_{j,t}| \leq T^{-h}$, while (iv) follows by (B.8), i.e., $|\mathbb{E}\bar{\delta}_{j_1,t}\bar{\delta}_{j_2,t}| \leq [(\mathbb{E}\bar{\delta}^2_{j_1,t})(\mathbb{E}\bar{\delta}^2_{j_2,t})]^{1/2} \leq T^{-1}C_4$ and $\mathbb{E}|\bar{\delta}_{j,t}| \leq (\mathbb{E}\bar{\delta}^2_{j,t})^{1/2} \leq T^{-1/2}C_4^{1/2}$.

By straight-forward computations based on the previous three displays and (B.9), we have that for some constant $K_7 > 0$,

$$
\mathbb{E}\left|\Delta L_{t+1,*} - \left[2\varepsilon_{t+1}\left(\beta_2 x_{2,t} - \beta_1 x_{1,t}\right) + \beta_2^2 x_{2,t}^2 - \beta_{1,t}^2 x_{1,t}^2\right]\right| \leq K_7 T^{-h}. \tag{B.14}
$$

Since $\mathbb{E}\varepsilon_{t+1}\beta_j x_{j,t} = 0$, we have that for some constant $K_8 > 0$,

$$
\left|\mathbb{E}\Delta L_{t+1,*} - \left(\beta_1^2\mathbb{E}x_{1,t}^2 - \beta_2^2\mathbb{E}x_{2,t}^2\right)\right| \leq K_8 T^{-h}.
$$

Since $\alpha_{x,2} < \alpha_{x,1}$ (by Assumption 1), $\beta_2^2\mathbb{E}x_{2,t}^2 - \beta_{1,t}^2\mathbb{E}x_{1,t}^2 \geq K_9 T^{-2\alpha_{x,2}}$ for some constant $K_9 > 0$. Since $h > 2\alpha_{x,2}$, we have that for large T,

$$
K_{10}T^{-2\alpha_{x,2}} \leq \mathbb{E}\Delta L_{t+1,*} \leq K_{11}T^{-2\alpha_{x,2}},
$$

where $K_{10}, K_{11} > 0$ are constants. This proves the second claim. □

Proof of Proposition 1. Let $p_1 = 2 + r/4$ and $h = \alpha_{x,2} + 1/2$. Since $r > 8$ and $\alpha_{x,2} \in [0, 1/2)$, we have that $p_1 \in (2, r/2)$ and $h \in (2\alpha_{x,2}, 1)$. Applying Lemma 8, we obtain that there exist constants $G_1, ..., G_4 > 0$ and an array of random variables $\{\Delta L_{t+1,*}\}_{t=n}^{T-n}$ such that for $T \geq G_1$,

$$\mathbb{P}\left(\bigcap_{t=n}^{T-1} \{\Delta L_{t+1,*} = \Delta L_{t+1}\}\right) \geq 1 - G_2 T^{1-\min\{p_1/2-1,\ (1-h)p_1\}}$$

and

$$G_3 T^{-2\alpha_{x,2}} \leq \mathbb{E}\Delta L_{t+1,*} \leq G_4 T^{-2\alpha_{x,2}}.$$

Notice that

$$\begin{aligned} 1 - \min\{p_1/2 - 1,\ (1-h)p_1\} &= \max\{2 - p_1/2,\ 1 + (h-1)p_1\} \\ &= \max\{2 - p_1/2,\ 1 + (\alpha_{x,2} - 1/2)p_1\} \\ &= \max\{1 - r/8,\ 1 + (\alpha_{x,2} - 1/2)(2 + r/4)\} \end{aligned}$$

The proof is complete. □

B.2 Proof of Proposition 2

Proof of Proposition 2. We recall all the notations in the proof of Lemma 8. By straight-forward algebra, it is not hard to see that

$$\begin{aligned} \Delta L_{t+1}^C &= \lambda^2 \Delta L_{t+1} + 2\lambda(1-\lambda)(y_{t+1} - \hat{y}_{2,t+1|t})(\hat{y}_{2,t+1|t} - \hat{y}_{1,t+1|t}) \\ &= \lambda^2 \Delta L_{t+1} + 2\lambda(1-\lambda)(\varepsilon_{t+1} + \beta_1 x_{1,t} - \delta_{2,t} x_{2,t})\left((\beta_2 + \delta_{2,t})x_{2,t} - (\beta_1 + \delta_{1,t})x_{1,t}\right). \end{aligned}$$

Recall $\Delta L_{t+1,*}$ as defined in (B.9). We also define

$$Q_t = (\varepsilon_{t+1} + \beta_1 x_{1,t} - \delta_{2,t,*} x_{2,t})\left((\beta_2 + \delta_{2,t,*})x_{2,t} - (\beta_1 + \delta_{1,t,*})x_{1,t}\right).$$

As argued in (B.10),

$$\begin{aligned} \mathbb{P}\left(\bigcap_{t=n}^{T-1} \left\{\Delta L_{t+1}^C = \lambda^2 \Delta L_{t+1,*} + 2\lambda(1-\lambda)Q_t\right\}\right) &\geq 1 - \sum_{t=n}^{T-1} \mathbb{P}\left(\delta_{1,t} \neq \delta_{1,t,*} \text{ or } \delta_{2,t} \neq \delta_{2,t,*}\right) \\ &\geq 1 - 2C_1(T-n)T^{-\min\{p_1/2-1,\ (1-h)p_1\}}. \quad \text{(B.15)} \end{aligned}$$

Notice that

$$
\begin{aligned}
\mathbb{E}Q_t &= \mathbb{E}(\varepsilon_{t+1} + \beta_1 x_{1,t} - \delta_{2,t,*}x_{2,t})\left((\beta_2 + \delta_{2,t,*})x_{2,t} - (\beta_1 + \delta_{1,t,*})x_{1,t}\right) \\
&= \mathbb{E}\varepsilon_{t+1}\left((\beta_2 + \delta_{2,t,*})x_{2,t} - (\beta_1 + \delta_{1,t,*})x_{1,t}\right) \\
&\quad + \mathbb{E}\beta_1 x_{1,t}\left((\beta_2 + \delta_{2,t,*})x_{2,t} - (\beta_1 + \delta_{1,t,*})x_{1,t}\right) \\
&\quad - \mathbb{E}\delta_{2,t,*}x_{2,t}\left((\beta_2 + \delta_{2,t,*})x_{2,t} - (\beta_1 + \delta_{1,t,*})x_{1,t}\right) \\
&\overset{(i)}{=} \mathbb{E}\delta_{2,t,*}x_{2,t}\varepsilon_{t+1} - \mathbb{E}\delta_{1,t,*}x_{1,t}\varepsilon_{t+1} + \beta_1\mathbb{E}x_{1,t}x_{2,t}\delta_{2,t,*} - \beta_1\mathbb{E}x_{1,t}^2\delta_{1,t,*} + \beta_1^2\mathbb{E}x_{1,t}^2 \\
&\quad - \beta_2\mathbb{E}\delta_{2,t,*}x_{2,t}^2 - \mathbb{E}\delta_{2,t,*}^2x_{2,t}^2 + \beta_1\mathbb{E}\delta_{2,t,*}x_{1,t}x_{2,t} + \mathbb{E}\delta_{1,t,*}\delta_{2,t,*}x_{1,t}x_{2,t},
\end{aligned}
$$

where (i) follows by the assumption of $\mathbb{E}x_{1,t}x_{2,t} = \mathbb{E}x_{1,t}\varepsilon_{t+1} = \mathbb{E}x_{2,t}\varepsilon_{t+1} = 0$. By (B.11), we have

$$
\max\left\{|\mathbb{E}\delta_{2,t,*}x_{2,t}\varepsilon_{t+1}|,\ |\mathbb{E}\delta_{1,t,*}x_{1,t}\varepsilon_{t+1}|\right\} \le K_1 T^{-h}.
$$

By (B.12), we have that for $j \in \{1,2\}$,

$$
\max\left\{|\mathbb{E}\delta_{j,t,*}x_{1,t}x_{2,t}|,\ |\mathbb{E}\delta_{j,t,*}x_{j,t}^2|\right\} \le T^{-h}K_4,
$$

By (B.13),

$$
\max\left\{|\mathbb{E}\delta_{1,t,*}\delta_{2,t,*}x_{1,t}x_{2,t}|,\ |\mathbb{E}\delta_{2,t,*}^2x_{2,t}^2|\right\} \le K_6\max\left\{T^{-h-1/2},\ T^{-2h}\right\}.
$$

It follows that

$$
\begin{aligned}
|\mathbb{E}Q_t| \le 2K_1T^{-h} + 2|\beta_1|K_4T^{-h} + \beta_1^2\mathbb{E}x_{1,t}^2 + |\beta_2|K_4T^{-h} \\
+ 2K_6\max\left\{T^{-h-1/2},\ T^{-2h}\right\} + |\beta_1|K_4T^{-h}.
\end{aligned}
$$

Recall that $h > 2\alpha_{x,1}$. By the assumption of $\alpha_{x,1} > \alpha_{x,2}$ and $\beta_j = c_{\beta,j}n^{-\alpha_{x,j}}$, it follows that $|\mathbb{E}Q_t| = o(n^{-2\alpha_{x,1}})$. By Proposition 1, $G_3T^{-2\alpha_{x,2}} \le \mathbb{E}\Delta L_{t+1,*} \le G_4T^{-2\alpha_{x,2}}$. Since $\alpha_{x,1} > \alpha_{x,2}$, we have that $|\mathbb{E}Q_t| = o(\mathbb{E}\Delta L_{t+1,*})$. Hence, there exist constants $D_1, D_2 > 0$ such that

$$
D_1T^{-2\alpha_{x,2}} \le \mathbb{E}\left(\lambda^2\Delta L_{t+1,*} + 2\lambda(1-\lambda)Q_t\right) \le D_2T^{-2\alpha_{x,2}}.
$$

Now we recall (B.15). The desired result follows by taking $\Delta L_{t+1,*}^C = \lambda^2\Delta L_{t+1,*} +$

$2\lambda(1-\lambda)Q_t$. $\square$

B.3 Proof of Proposition 3

Our proof of Proposition 3 relies on two lemmas, Lemma 9 and 10. We first state and prove these lemmas before proving Proposition 3.

Lemma 9. *Suppose that Assumptions 1 and 2 hold. Let $\Delta L_{t+1,*}$ be defined as in (B.9) in the proof of Lemma 8. Let $\tilde{\theta}_t = (\tilde{\theta}_{1,t}, \tilde{\theta}_{2,t})' = (\sum_{s=t-m}^{t-1} z_s z_s')^{-1}(\sum_{s=t-m}^{t-1} z_s \Delta L_{t+1,*})$, where $z_s = (1, z_{Ms})'$. Fix $p \in (2, r/3)$. Then there exist constants $G_0, G_1, G_2, G_3 > 0$ such that for $T \geq G_0$,*

$$\mathbb{P}\left(\left|\frac{\tilde{\theta}_{1,t}}{\tilde{\theta}_{2,t}}\right| \geq G_1 T^{\alpha_{z,2}-\alpha_{x,2}}\right) \leq G_2 T^{1-p/2} \log^{-G_3}$$

and

$$\mathbb{P}\left(\tilde{\theta}_{2,t} \leq 0\right) \leq G_2 T^{1-p/2} \log^{-G_3}.$$

Proof. Let $\Psi_{t+1} = m^{\alpha_{x,2}} \Delta L_{t+1,*}$, $\pi_t = (\sum_{s=t-m}^{t-1} \mathbb{E} z_s z_s')^{-1}(\sum_{s=t-m}^{t-1} \mathbb{E} z_s \Psi_{s+1})$, $\hat{\pi}_t = (\sum_{s=t-m}^{t-1} z_s z_s')^{-1}(\sum_{s=t-m}^{t-1} z_s \Psi_{s+1})$ and $\{\xi_{s+1}\}_{s=t-m}^{t-1}$ with $\xi_{s+1} = \Psi_{s+1} - z_s' \pi_t$. Clearly, $\hat{\pi}_t = m^{\alpha_{x,2}} \tilde{\theta}_t$. Let $\gamma_t = \hat{\pi}_t - \pi_t$. The proof proceeds in two steps. We first bound γ_t and then show the desired results.

Step 1: bound γ_t

By simple computation, we have

$$\gamma_t = \underbrace{\left[m^{-1} \sum_{s=t-m}^{t-1} z_s z_s'\right]^{-1}}_{J_t} \cdot \underbrace{m^{-1} \sum_{s=t-m}^{t-1} z_s \xi_{s+1}}_{B_t}.$$

Since entries of $z_s z_s' - \mathbb{E} z_s z_s'$ has uniformly bounded $0.5r$-th moments, it follows, by Lemma 4, that for some constants $K_1, K_2, K_3 > 0$,

$$\mathbb{P}\left(\left\|m^{-1} \sum_{s=t-m}^{t-1} (z_s z_s' - \mathbb{E} z_s z_s')\right\|_\infty \geq K_1 \sqrt{m^{-1} \log m}\right) \leq K_2 m^{1-p/2} \log^{-K_3} m.$$

Since $m^{-1} \sum_{s=t-m}^{t-1} \mathbb{E} z_s z_s' = \text{diag}(1, m^{-1} \sum_{s=t-m}^{t-1} \mathbb{E} z_{Ms}^2)$ and $\mathbb{E} z_{Ms}^2$ is bounded away from zero and infinity, the eigenvalues of $m^{-1} \sum_{s=t-m}^{t-1} \mathbb{E} z_s z_s'$ lie in $[K_4, K_5]$ for some constants $K_4, K_5 > 0$. By the above display, the eigenvalues of $m^{-1}\sum_{s=t-m}^{t-1} z_s z_s'$ lie in $[K_4/2, 2K_5]$ for large n with probability at least $1 - K_2 m^{1-p/2}\log^{-K_3} m$. Hence, for some constants $K_6, K_7 > 0$, we have that for $n \geq K_6$,

$$\mathbb{P}\left(\|J_t\|_\infty \geq K_7\right) \leq K_2 m^{1-p/2}\log^{-K_3} m.$$

Recall the definition of $\Delta L_{t+1,*}$ (in (B.9) in the proof of Lemma 8). We apply Lemma 2. By straight-forward computations, we have that Ψ_{t+1} has uniformly bounded $0.5r$-th moments. Hence, $\xi_{s+1} = \Psi_{s+1} - z_s'\pi_s$ has uniformly bounded $0.5r$-th moments. Again by Lemma 2, entries of $z_s\xi_{s+1}$ has uniformly bounded $\frac{r}{3}$-th moments. Notice that $p \in (2, r/3)$. It follows by Lemma 4 (applied to each entry of $z_s\xi_{s+1}$) that for some constants $K_8, K_9, K_{10} > 0$, we have

$$\mathbb{P}\left(\|B_t\|_\infty \geq K_8\sqrt{m^{-1}\log m}\right) \leq K_9 m^{1-p/2}\log^{-K_{10}} m.$$

It follows by the above two displays that for large m,

$$\begin{aligned}
\mathbb{P}\left(\|\gamma_t\|_\infty \geq 2K_7K_8\sqrt{m^{-1}\log m}\right) &\leq \mathbb{P}\left(2\|J_t\|_\infty\|B_t\|_\infty \geq 2K_7K_8\sqrt{m^{-1}\log m}\right)\\
&\leq \mathbb{P}\left(\|J_t\|_\infty \geq K_7\right) + \mathbb{P}\left(\|B_t\|_\infty \geq K_8\sqrt{m^{-1}\log m}\right)\\
&\leq K_2 m^{1-p/2}\log^{-K_3} m + K_9 m^{1-p/2}\log^{-K_{10}} m.
\end{aligned}$$

Since $T \asymp m$, there are constants $K_{11}, ..., K_{14} > 0$ such that for $T \geq K_{11}$,

$$\mathbb{P}\left(\|\gamma_t\|_\infty \geq K_{12}\sqrt{T^{-1}\log T}\right) \leq K_{13}T^{1-p/2}\log^{-K_{14}}. \tag{B.16}$$

Step 2: show the desired results.

Partition $\pi_t = (\pi_{1,t}, \pi_{2,t})'$. By $\mathbb{E} = 0$, it follows that $\pi_{1,t} = m^{-1}\sum_{s=t-m}^{t-1}\mathbb{E}m^{\alpha_{x,2}}\Delta L_{s+1,*}$ and $\pi_{2,t} = (\sum_{s=t-m}^{t-1}\mathbb{E}m^{\alpha_{x,2}}\Delta L_{s+1,*})/(m^{-1}\sum_{s=t-m}^{t-1}\mathbb{E}z_{Ms}^2)$. By Lemma 8, there are constants $K_{15}, K_{16} > 0$ such that

$$K_{15}T^{-\alpha_{x,2}} \leq \pi_{1,t} \leq K_{16}T^{-\alpha_{x,2}}. \tag{B.17}$$

By computations similar to (B.14) in the proof of Lemma 8, one can show that

$$\mathbb{E}\left|z_{Mt}\Delta L_{t+1,*} - z_{Mt}\left[2\varepsilon_{t+1}\left(\beta_2 x_{2,t} - \beta_1 x_{1,t}\right) + \beta_2^2 x_{2,t}^2 - \beta_{1,t}^2 x_{1,t}^2\right]\right| \leq K_{17} T^{-1/2}\sqrt{\log T},$$

where $K_{17} > 0$ is a constant. By Assumption 1,

$$\mathbb{E}\left\{z_{Mt}\left[2\varepsilon_{t+1}\left(\beta_2 x_{2,t} - \beta_1 x_{1,t}\right) + \beta_2^2 x_{2,t}^2 - \beta_{1,t}^2 x_{1,t}^2\right]\right\} \geq K_{18} T^{-\alpha_{x,2}-\alpha_{z,2}}.$$

It follows that for large T, $\mathbb{E} z_{Mt}\Delta L_{t+1,*} \geq K_{18} T^{-\alpha_{x,2}-\alpha_{z,2}}/2$. Hence, for some constant $K_{19} > 0$,

$$\pi_{2,t} \geq K_{19} T^{-\alpha_{z,2}}. \tag{B.18}$$

Let $x = 2K_{16} T^{-\alpha_{x,2}}$ and $M = 2x/(K_{19} T^{-\alpha_{z,2}})$. Then

$$\begin{aligned}
&\mathbb{P}\left(|\hat{\pi}_{1,t}| \geq M|\hat{\pi}_{2,t}|\right)\\
&\leq \mathbb{P}\left(|\hat{\pi}_{1,t}| \geq x\right) + \mathbb{P}\left(|\hat{\pi}_{2,t}| \leq x/M\right)\\
&\leq \mathbb{P}\left(|\hat{\pi}_{1,t} - \pi_{1,t}| \geq x - |\pi_1|\right) + \mathbb{P}\left(|\hat{\pi}_{2,t} - \pi_{2,t}| \geq |\pi_{2,t}| - x/M\right)\\
&\overset{(i)}{\leq} \mathbb{P}\left(|\hat{\pi}_{1,t} - \pi_{1,t}| \geq x - K_{16} T^{-\alpha_{x,2}}\right) + \mathbb{P}\left(|\hat{\pi}_{2,t} - \pi_{2,t}| \geq K_{19} T^{-\alpha_{z,2}} - x/M\right)\\
&\leq \mathbb{P}\left(\|\gamma_t\|_\infty \geq x - K_{16} T^{-\alpha_{x,2}}\right) + \mathbb{P}\left(\|\gamma_t\|_\infty \geq K_{19} T^{-\alpha_{z,2}} - x/M\right)\\
&= \mathbb{P}\left(\|\gamma_t\|_\infty \geq K_{16} T^{-\alpha_{x,2}}\right) + \mathbb{P}\left(\|\gamma_t\|_\infty \geq K_{19} T^{-\alpha_{z,2}}/2\right)\\
&\overset{(ii)}{\leq} 2K_{13} T^{1-p/2}\log^{-K_{14}},
\end{aligned}$$

where (i) holds by (B.17) and (B.18) and (ii) follows by (B.16) together with $T^{-\alpha_{x,2}} \gg \sqrt{T^{-1}\log T}$ and $T^{-\alpha_{z,2}} \gg \sqrt{T^{-1}\log T}$. The first claim follows by $|\hat{\pi}_{1,t}/\hat{\pi}_{2,t}| = |\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}|$ and $M = 2x/(K_{19} T^{-\alpha_{z,2}}) = (4K_{16}/K_{19}) T^{\alpha_{z,2}-\alpha_{x,2}}$.

To see the second claim, notice that

$$\begin{aligned}
\mathbb{P}\left(\tilde{\theta}_{2,t} \leq 0\right) \overset{(i)}{=} \mathbb{P}\left(\hat{\pi}_{2,t} \leq 0\right) = \mathbb{P}\left(\gamma_{2,t} \leq -\pi_{2,t}\right) &\overset{(ii)}{\leq} \mathbb{P}\left(\gamma_{2,t} \leq -K_{19} T^{-\alpha_{z,2}}\right)\\
&\leq \mathbb{P}\left(\|\gamma_t\|_\infty \geq K_{19} T^{-\alpha_{z,2}}\right)\\
&\overset{(iii)}{\leq} K_{13} T^{1-p/2}\log^{-K_{14}},
\end{aligned}$$

where (i) holds by $\tilde{\theta}_t = m^{-\alpha_{x,2}}\hat{\pi}_t$, (ii) follows by (B.18) and (iii) holds by (B.16) and $T^{-\alpha_{z,2}} \gg \sqrt{T^{-1}\log T}$. □

Lemma 10. *Let Assumptions 1 and 2 hold. Fix any $p_1 \in (2, r/2)$ and $h \in (2\alpha_{x,2}, 1)$. Then there exist constants $G_0, G_1, ..., G_5 > 0$ and an array $\{S_{t+1}\}_{t=n+m}^{T-1}$ such that $\forall T \geq G_0$,*

$$\mathbb{P}\left(\bigcap_{t=n+m}^{T-1}\left\{S_{t+1} = \Delta L_{t+1}\mathbf{1}\{z_t'\hat{\theta}_{m,t} > 0\}\right\}\right) \geq 1 - G_2 T^{1-\min\{p_1/2-1,\ (1-h)p_1\}}$$

and

$$\mathbb{E}S_{t+1} \geq K_5 T^{-\alpha_{x,2}-\alpha_{z,2}}.$$

Proof. Let $\tilde{\theta}_t = (\sum_{s=t-m}^{t-1} z_s z_s')^{-1}(\sum_{s=t-m}^{t-1} z_s \Delta L_{t+1,*})$, where $\Delta L_{t+1,*}$ is defined as in (B.9) in the proof of Lemma 8. Define $S_{t+1} = \Delta L_{t+1,*}\mathbf{1}\{z_t'\tilde{\theta}_t > 0\}$. Notice that

$$\bigcap_{t=n}^{T-1}\{\Delta L_{t+1,*} = \Delta L_{t+1}\} \subseteq \bigcap_{t=n+m}^{T-1}\left\{S_{t+1} = \Delta L_{t+1}\mathbf{1}\{z_t'\hat{\theta}_{m,t} > 0\}\right\}.$$

Hence, the first claim follows by Lemma 8.

It remains to bound $\mathbb{E}S_{t+1}$. To this end, let $q = r/2$ and $\nu = r/(r-2)$. Hence, $q, \nu > 1$ and $q^{-1} + \nu^{-1} = 1$. Notice that

$$\begin{aligned}
&\mathbb{E}\left(|\Delta L_{t+1,*}|\left|\mathbf{1}\{z_t'\tilde{\theta}_t > 0\} - \mathbf{1}\{z_{Mt} > 0\}\right|\right)\\
&= \mathbb{E}\Bigg(|\Delta L_{t+1,*}| \cdot \Bigg|\mathbf{1}\{z_{Mt} > -\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t} \text{ and } \tilde{\theta}_{2,t} > 0\}\\
&\qquad\qquad + \mathbf{1}\{z_{Mt} < -\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t} \text{ and } \tilde{\theta}_{2,t} \leq 0\} - \mathbf{1}\{z_{Mt} > 0\}\Bigg|\Bigg)\\
&\leq \mathbb{E}\left(|\Delta L_{t+1,*}|\left[\mathbf{1}\left\{0 < z_{Mt} \leq -\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}\right\} + \mathbf{1}\left\{\tilde{\theta}_{2,t} \leq 0\right\}\right]\right)\\
&\overset{(i)}{\leq} \|\Delta L_{t+1,*}\|_{L^q(\mathbb{P})}\left\|\mathbf{1}\left\{0 < z_{Mt} \leq -\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}\right\} + \mathbf{1}\left\{\tilde{\theta}_{2,t} \leq 0\right\}\right\|_{L^\nu(\mathbb{P})}\\
&\leq \|\Delta L_{t+1,*}\|_{L^q(\mathbb{P})}\left[\left\|\mathbf{1}\left\{0 < z_{Mt} \leq -\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}\right\}\right\|_{L^\nu(\mathbb{P})} + \left\|\mathbf{1}\left\{\tilde{\theta}_{2,t} \leq 0\right\}\right\|_{L^\nu(\mathbb{P})}\right]\\
&= \|\Delta L_{t+1,*}\|_{L^q(\mathbb{P})}\left(\left[\mathbb{P}\left(0 < z_{Mt} \leq -\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}\right)\right]^{1/\nu} + \left[\mathbb{P}\left(\tilde{\theta}_{2,t} \leq 0\right)\right]^{1/\nu}\right) \qquad \text{(B.19)}
\end{aligned}$$

where (i) follows by Holder's inequality. By Assumption 2, the p.d.f of z_{Mt} in a fixed neighborhood of zero is bounded above by some constant $K_0 > 0$. Recall constants

$G_1, ..., G_4 > 0$ in the statement of Lemma 9. Hence,

$$\begin{aligned}\mathbb{P}\left(0 < z_{Mt} \leq -\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}\right) &\leq \mathbb{P}\left(0 < z_{Mt} \leq \left|\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}\right|\right) \\ &\leq \mathbb{P}\left(0 < z_{Mt} \leq G_1 T^{\alpha_{z,2}-\alpha_{x,2}}\right) + \mathbb{P}\left(\left|\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}\right| \geq G_1 T^{\alpha_{z,2}-\alpha_{x,2}}\right) \\ &\overset{(i)}{\leq} K_0 G_1 T^{\alpha_{z,2}-\alpha_{x,2}} + G_2 T^{1-p/2} \log^{-G_3},\end{aligned} \tag{B.20}$$

where (i) follows by the bounded p.d.f of z_{Mt} near zero and $T^{\alpha_{z,2}-\alpha_{x,2}} = o(1)$, as well as by Lemma 9.

Since $r > 8$, it is not hard to show that $r/3 > 2 + r/(2r - 4) = 2 + \nu/2$. By Assumptions 1 and 2, $2\nu\alpha_{z,2} < \alpha_{x,2} < 1/2$. Since $r > 8$, we have that $r/3 > 8/3 > 2 + 1/2 > 2 + 2\nu\alpha_{z,2}$. Fix $p \in (2 + 2\nu\alpha_{z,2}, r/3)$. Now (B.19), (B.20) and Lemma 9 imply that for some constants $K_1, K_2 > 0$

$$\begin{aligned}&\mathbb{E}\left(|\Delta L_{t+1,*}| \left|\mathbf{1}\{z_t'\tilde{\theta}_t > 0\} - \mathbf{1}\{z_{Mt} > 0\}\right|\right) \\ &\qquad \leq K_1 \|\Delta L_{t+1,*}\|_{L^q(\mathbb{P})} \left[T^{(\alpha_{z,2}-\alpha_{x,2})/\nu} + \left(T^{1-p/2}\log^{-K_2}\right)^{1/\nu}\right].\end{aligned} \tag{B.21}$$

By (B.9), we have that

$$\mathbb{E}\Delta L_{t+1,*}\mathbf{1}\{z_{Mt} > 0\} \geq 2\mathbb{E}\left[\varepsilon_{t+1}\left(\beta_2 x_{2,t} - \beta_1 x_{1,t}\right)\mathbf{1}\{z_{Mt} > 0\}\right] - A_t, \tag{B.22}$$

where

$$\begin{aligned}A_t &= 2\mathbb{E}\left|\varepsilon_{t+1}\left(\delta_{2,t,*}x_{2,t} - \delta_{1,t,*}x_{1,t}\right)\right| \\ &+ \mathbb{E}\left|\left(\beta_2 x_{2,t} - \beta_1 x_{1,t} - \delta_{1,t,*}x_{1,t} + \delta_{2,t,*}x_{2,t}\right)\left(\beta_1 x_{1,t} + \beta_2 x_{2,t} - \delta_{1,t,*}x_{1,t} - \delta_{2,t,*}x_{2,t}\right)\right|.\end{aligned}$$

After computations similar to (B.14) in the proof of Lemma 8, we can use the rate conditions in Assumption 1 and show that for some constant $K_3 > 0$,

$$A_t \leq K_3 T^{-2\alpha_{x,2}}. \tag{B.23}$$

(B.22) and (B.23) imply that for some constants $K_4, K_5, K_6 > 0$, we have that for $T \geq K_4$,

$$\mathbb{E}\Delta L_{t+1,*}\mathbf{1}\{z_{Mt} > 0\} \geq 2\mathbb{E}\left[\varepsilon_{t+1}\left(\beta_2 x_{2,t} - \beta_1 x_{1,t}\right)\mathbf{1}\{z_{Mt} > 0\}\right] - K_3 T^{-1/2} - K_4 T^{-2\alpha_{x,2}}$$

$$\overset{(i)}{\geq} K_5 T^{-\alpha_{x,2}-\alpha_{z,2}} - K_3 T^{-2\alpha_{x,2}},$$

where (i) holds by Assumption 1. By the above display and (B.21), we have that for large T,

$$\mathbb{E}S_{t+1} \geq K_5 T^{-\alpha_{x,2}-\alpha_{z,2}} - K_3 T^{-2\alpha_{x,2}} \\ - K_1 \|\Delta L_{t+1,*}\|_{L^q(\mathbb{P})} \left[T^{(\alpha_{z,2}-\alpha_{x,2})/\nu} + \left(T^{1-p/2} \log^{-K_2}\right)^{1/\nu} \right].$$

Recall that in Step 1 of the proof of Lemma 9, we have that $m^{\alpha_{x,2}} \Delta L_{t+1,*}$ has uniformly bounded $0.5r$-th moments. Since $q = 0.5r$, we have that $\|m^{\alpha_{x,2}} \Delta L_{t+1,*}\|_{L^q(\mathbb{P})}$ is bounded above by a constant. Hence, for some constant $K_7 > 0$,

$$\mathbb{E}S_{t+1} \geq K_5 T^{-\alpha_{x,2}-\alpha_{z,2}} - K_3 T^{-2\alpha_{x,2}} \\ - K_1 K_7 T^{-\alpha_{x,2}} \left[T^{(\alpha_{z,2}-\alpha_{x,2})/\nu} + \left(T^{1-p/2} \log^{-K_2}\right)^{1/\nu} \right]. \quad \text{(B.24)}$$

Since $p > 2 + 2\nu\alpha_{z,2}$ and $\nu = r/(r-2)$, it is not hard to show that $-\alpha_{x,2} + (1-p/2)/\nu < -\alpha_{x,2} - \alpha_{z,2}$. By Assumption 1, it is straight-forward to verify that $-2\alpha_{x,2} < -\alpha_{x,2} - \alpha_{z,2}$ and $-\alpha_{x,2} + (\alpha_{z,2} - \alpha_{x,2})/\nu < -\alpha_{x,2} - \alpha_{z,2}$. The desired result follows by (B.24). □

Proof of Proposition 3. We choose p_1 and h as in the proof of Proposition 1. Then Part (1) follows by Lemma 10 and the computations in the proof of Proposition 1. Part (2) follows by Part (1) and Proposition 1. □

B.4 Proof of Proposition 4

Our proof of Proposition 4 relies on two lemmas, Lemma 11 and 12. We first state and prove these lemmas before proving Proposition 4.

Lemma 11. *Suppose that the assumptions of Proposition 4 hold. Let $\Delta L_{t+1,*}$ be defined as in (B.9) in the proof of Lemma 8. Let $\tilde{\theta}_t = (\tilde{\theta}_{1,t}, \tilde{\theta}_{2,t})' = (\sum_{s=t-m}^{t-1} z_s z_s')^{-1} (\sum_{s=t-m}^{t-1} z_s \Delta L_{t+1,*})$, where $z_s = (1,)'$. Fix $p \in (2, r/3)$. Then there exist some constants $G_0, G_1, G_2, G_3 > 0$ such that for $T \geq G_0$,*

$$\mathbb{P}\left(\left| \frac{\tilde{\theta}_{1,t}}{\tilde{\theta}_{2,t}} \right| \geq G_1 T^{\alpha_{x,1}+\alpha_{z,1}-2\alpha_{x,2}} \right) \leq G_2 T^{1-p/2} \log^{-G_3}$$

and

$$\mathbb{P}\left(\tilde{\theta}_{2,t} \leq 0\right) \leq G_2 T^{1-p/2} \log^{-G_3}.$$

Proof. Let $\Psi_{t+1} = m^{\alpha_{x,2}} \Delta L_{t+1,*}$, $\pi_t = (\sum_{s=t-m}^{t-1} \mathbb{E} z_s z_s')^{-1}(\sum_{s=t-m}^{t-1} \mathbb{E} z_s \Psi_{s+1})$, $\hat{\pi}_t = (\sum_{s=t-m}^{t-1} z_s z_s')^{-1}(\sum_{s=t-m}^{t-1} z_s \Psi_{s+1})$ and $\{\xi_{s+1}\}_{s=t-m}^{t-1}$ with $\xi_{s+1} = \Psi_{s+1} - z_s' \pi_t$. Clearly, $\hat{\pi}_t = m^{\alpha_{x,2}} \tilde{\theta}_t$. Let $\gamma_t = \hat{\pi}_t - \pi_t$. The proof proceeds in two steps. We first bound γ_t and then show the desired results.

Step 1: bound γ_t

By simple computation, we have

$$\gamma_t = \underbrace{\left[m^{-1} \sum_{s=t-m}^{t-1} z_s z_s'\right]^{-1}}_{J_t} \cdot \underbrace{m^{-1} \sum_{s=t-m}^{t-1} z_s \xi_{s+1}}_{B_t}.$$

Notice that $p \in (2, r/3)$. Since entries of $z_s z_s' - \mathbb{E} z_s z_s'$ has uniformly bounded $0.5r$-th moments, it follows, by Lemma 4, that for some constants $K_1, K_2, K_3 > 0$,

$$\mathbb{P}\left(\left\|m^{-1} \sum_{s=t-m}^{t-1} (z_s z_s' - \mathbb{E} z_s z_s')\right\|_\infty \geq K_1 \sqrt{m^{-1} \log m}\right) \leq K_2 m^{1-p/2} \log^{-K_3} m.$$

Since $m^{-1} \sum_{s=t-m}^{t-1} \mathbb{E} z_s z_s' = \operatorname{diag}(1, m^{-1} \sum_{s=t-m}^{t-1} \mathbb{E} z_{Ms}^2)$ and $\mathbb{E} z_{Ms}^2$ is bounded away from zero and infinity, the eigenvalues of $m^{-1} \sum_{s=t-m}^{t-1} \mathbb{E} z_s z_s'$ lie in $[K_4, K_5]$ for some constants $K_4, K_5 > 0$. By the above display, the eigenvalues of $m^{-1} \sum_{s=t-m}^{t-1} z_s z_s'$ lie in $[K_4/2, 2K_5]$ for large n with probability at least $1 - K_2 m^{1-p/2} \log^{-K_3} m$. Hence, for some constants $K_6, K_7 > 0$, we have that for $n \geq K_6$,

$$\mathbb{P}\left(\|J_t\|_\infty \geq K_7\right) \leq K_2 m^{1-p/2} \log^{-K_3} m.$$

Recall the definition of $\Delta L_{t+1,*}$ (in (B.9) in the proof of Lemma 8). We apply Lemma 2. By straight-forward computations, we have that Ψ_{t+1} has uniformly bounded $0.5r$-th moments. Hence, $\xi_{s+1} = \Psi_{s+1} - z_s' \pi_s$ has uniformly bounded $0.5r$-th moments. Again by Lemma 2, entries of $z_s \xi_{s+1}$ has uniformly bounded $\frac{r}{3}$-th moments. Notice that $p \in (2, r/3)$. It follows by 4 (applied to each entry of $z_s \xi_{s+1}$) that for some constants $K_8, K_9, K_{10} > 0$, we have

$$\mathbb{P}\left(\|B_t\|_\infty \geq K_8 \sqrt{m^{-1} \log m}\right) \leq K_9 m^{1-p/2} \log^{-K_{10}} m.$$

It follows by the above two displays that for large m,

$$\begin{aligned}\mathbb{P}\left(\|\gamma_t\|_\infty \geq 2K_7K_8\sqrt{m^{-1}\log m}\right) &\leq \mathbb{P}\left(2\|J_t\|_\infty\|B_t\|_\infty \geq 2K_7K_8\sqrt{m^{-1}\log m}\right)\\ &\leq \mathbb{P}\left(\|J_t\|_\infty \geq K_7\right) + \mathbb{P}\left(\|B_t\|_\infty \geq K_8\sqrt{m^{-1}\log m}\right)\\ &\leq K_2 m^{1-p/2}\log^{-K_3} m + K_9 m^{1-p/2}\log^{-K_{10}} m.\end{aligned}$$

Since $T \asymp m$, there are constants $K_{11}, ..., K_{14} > 0$ such that for $T \geq K_{11}$,

$$\mathbb{P}\left(\|\gamma_t\|_\infty \geq K_{12}\sqrt{T^{-1}\log T}\right) \leq K_{13}T^{1-p/2}\log^{-K_{14}}. \tag{B.25}$$

Step 2: show the desired results.

Partition $\pi_t = (\pi_{1,t}, \pi_{2,t})'$. By $\mathbb{E} = 0$, it follows that $\pi_{1,t} = m^{-1}\sum_{s=t-m}^{t-1}\mathbb{E}m^{\alpha_{x,2}}\Delta L_{s+1,*}$ and $\pi_{2,t} = (\sum_{s=t-m}^{t-1}\mathbb{E}m^{\alpha_{x,2}}\Delta L_{s+1,*})/(m^{-1}\sum_{s=t-m}^{t-1}\mathbb{E}z_{Ms}^2)$. By Lemma 8, there are constants $K_{15}, K_{16} > 0$ such that

$$K_{15}T^{-\alpha_{x,2}} \leq \pi_{1,t} \leq K_{16}T^{-\alpha_{x,2}}. \tag{B.26}$$

By computations similar to (B.14) in the proof of Lemma 8, one can show that

$$\mathbb{E}\left|z_{Mt}\Delta L_{t+1,*} - z_{Mt}\left[2\varepsilon_{t+1}\left(\beta_2 x_{2,t} - \beta_1 x_{1,t}\right) + \beta_2^2 x_{2,t}^2 - \beta_{1,t}^2 x_{1,t}^2\right]\right| \leq K_{17}T^{-1/2}\sqrt{\log T},$$

where $K_{17} > 0$ is a constant. By the assumptions of Proposition 4,

$$\mathbb{E}\left\{z_{Mt}\left[2\varepsilon_{t+1}\left(\beta_2 x_{2,t} - \beta_1 x_{1,t}\right) + \beta_2^2 x_{2,t}^2 - \beta_{1,t}^2 x_{1,t}^2\right]\right\} \geq K_{18}T^{-\alpha_{x,1}-\alpha_{z,1}}.$$

It follows that for large T, $\mathbb{E}z_{Mt}\Delta L_{t+1,*} \geq K_{18}T^{-\alpha_{x,1}-\alpha_{z,1}}/2$. Hence, for some constant $K_{19} > 0$,

$$\pi_{2,t} \geq K_{19}T^{\alpha_{x,2}-\alpha_{x,1}-\alpha_{z,1}}. \tag{B.27}$$

Let $x = 2K_{16}T^{-\alpha_{x,2}}$ and $M = 2x/(K_{19}T^{\alpha_{x,2}-\alpha_{x,1}-\alpha_{z,1}})$. Then

$$\begin{aligned}&\mathbb{P}\left(|\hat{\pi}_{1,t}| \geq M|\hat{\pi}_{2,t}|\right)\\ &\leq \mathbb{P}\left(|\hat{\pi}_{1,t}| \geq x\right) + \mathbb{P}\left(|\hat{\pi}_{2,t}| \leq x/M\right)\\ &\leq \mathbb{P}\left(|\hat{\pi}_{1,t} - \pi_{1,t}| \geq x - |\pi_1|\right) + \mathbb{P}\left(|\hat{\pi}_{2,t} - \pi_{2,t}| \geq |\pi_{2,t}| - x/M\right)\end{aligned}$$

$$
\begin{aligned}
&\overset{(i)}{\leq} \mathbb{P}\left(|\hat{\pi}_{1,t}-\pi_{1,t}| \geq x - K_{16}T^{-\alpha_{x,2}}\right) + \mathbb{P}\left(|\hat{\pi}_{2,t}-\pi_{2,t}| \geq K_{19}T^{\alpha_{x,2}-\alpha_{x,1}-\alpha_{z,1}} - x/M\right)\\
&\leq \mathbb{P}\left(\|\gamma_t\|_\infty \geq x - K_{16}T^{-\alpha_{x,2}}\right) + \mathbb{P}\left(\|\gamma_t\|_\infty \geq K_{19}T^{\alpha_{x,2}-\alpha_{x,1}-\alpha_{z,1}} - x/M\right)\\
&= \mathbb{P}\left(\|\gamma_t\|_\infty \geq K_{16}T^{-\alpha_{x,2}}\right) + \mathbb{P}\left(\|\gamma_t\|_\infty \geq K_{19}T^{\alpha_{x,2}-\alpha_{x,1}-\alpha_{z,1}}/2\right)\\
&\overset{(ii)}{\leq} 2K_{13}T^{1-p/2}\log^{-K_{14}},
\end{aligned}
$$

where (i) holds by (B.26) and (B.27) and (ii) follows by (B.25) together with $T^{-\alpha_{x,2}} \gg \sqrt{T^{-1}\log T}$ and $T^{\alpha_{x,2}-\alpha_{x,1}-\alpha_{z,1}} \gg \sqrt{T^{-1}\log T}$. The first claim follows by $|\hat{\pi}_{1,t}/\hat{\pi}_{2,t}| = |\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}|$ and $M = 2x/(K_{19}T^{\alpha_{x,2}-\alpha_{x,1}-\alpha_{z,1}}) = (4K_{16}/K_{19})T^{\alpha_{x,1}+\alpha_{z,1}-2\alpha_{x,2}}$.

To see the second claim, notice that

$$
\begin{aligned}
\mathbb{P}\left(\tilde{\theta}_{2,t} \leq 0\right) \overset{(i)}{=} \mathbb{P}\left(\hat{\pi}_{2,t} \leq 0\right) = \mathbb{P}\left(\gamma_{2,t} \leq -\pi_{2,t}\right) &\overset{(ii)}{\leq} \mathbb{P}\left(\gamma_{2,t} \leq -K_{19}T^{\alpha_{x,2}-\alpha_{x,1}-\alpha_{z,1}}\right)\\
&\leq \mathbb{P}\left(\|\gamma_t\|_\infty \geq K_{19}T^{\alpha_{x,2}-\alpha_{x,1}-\alpha_{z,1}}\right)\\
&\overset{(iii)}{\leq} K_{13}T^{1-p/2}\log^{-K_{14}},
\end{aligned}
$$

where (i) holds by $\tilde{\theta}_t = m^{-\alpha_{x,2}}\hat{\pi}_t$, (ii) follows by (B.27) and (iii) holds by (B.25) and $T^{\alpha_{x,2}-\alpha_{x,1}-\alpha_{z,1}} \gg \sqrt{T^{-1}\log T}$. □

Lemma 12. *Let the assumptions of Proposition 4 hold. Then $\forall p_1 \in (2, r/3)$, there exist constants $G_0, G_1, ..., G_5 > 0$ and an array $\{S_{t+1}\}_{t=n+m}^{T-1}$ such that $\forall T \geq G_0$,*

$$
\mathbb{P}\left(\bigcap_{t=n+m}^{T-1}\left\{S_{t+1} = \Delta L_{t+1}\mathbf{1}\{z_t'\hat{\theta}_{m,t} > 0\}\right\}\right) \geq 1 - G_1T^{2-p_1/2}\log^{-G_2}T
$$

and

$$
\mathbb{E}S_{t+1} \geq K_5T^{-\alpha_{x,1}-\alpha_{z,1}}.
$$

Proof. Let $\tilde{\theta}_t = (\sum_{s=t-m}^{t-1} z_sz_s')^{-1}(\sum_{s=t-m}^{t-1} z_s\Delta L_{t+1,*})$, where $\Delta L_{t+1,*}$ is defined as in (B.9) in the proof of Lemma 8. Define $S_{t+1} = \Delta L_{t+1,*}\mathbf{1}\{z_t'\tilde{\theta}_t > 0\}$. Notice that

$$
\bigcap_{t=n}^{T-1}\{\Delta L_{t+1,*} = \Delta L_{t+1}\} \subseteq \bigcap_{t=n+m}^{T-1}\left\{S_{t+1} = \Delta L_{t+1}\mathbf{1}\{z_t'\hat{\theta}_{m,t} > 0\}\right\}.
$$

Hence, the first claim follows by Lemma 8.

It remains to bound $\mathbb{E}S_{t+1}$. To this end, let $q = r/2$ and $\nu = r/(r-2)$. Hence,

$q, \nu > 1$ and $q^{-1} + \nu^{-1} = 1$. Notice that

$$\begin{aligned}
&\mathbb{E}\left(|\Delta L_{t+1,*}| \left|\mathbf{1}\{z_t'\tilde{\theta}_t > 0\} - \mathbf{1}\{z_{Mt} > 0\}\right|\right) \\
&= \mathbb{E}\Bigg(|\Delta L_{t+1,*}| \cdot \Bigg|\mathbf{1}\{z_{Mt} > -\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t} \text{ and } \tilde{\theta}_{2,t} > 0\} \\
&\qquad\qquad + \mathbf{1}\{z_{Mt} < -\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t} \text{ and } \tilde{\theta}_{2,t} \leq 0\} - \mathbf{1}\{z_{Mt} > 0\}\Bigg|\Bigg) \\
&\leq \mathbb{E}\left(|\Delta L_{t+1,*}| \left[\mathbf{1}\left\{0 < z_{Mt} \leq -\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}\right\} + \mathbf{1}\left\{\tilde{\theta}_{2,t} \leq 0\right\}\right]\right) \\
&\overset{(i)}{\leq} \|\Delta L_{t+1,*}\|_{L^q(\mathbb{P})} \left\|\mathbf{1}\left\{0 < z_{Mt} \leq -\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}\right\} + \mathbf{1}\left\{\tilde{\theta}_{2,t} \leq 0\right\}\right\|_{L^\nu(\mathbb{P})} \\
&\leq \|\Delta L_{t+1,*}\|_{L^q(\mathbb{P})} \left[\left\|\mathbf{1}\left\{0 < z_{Mt} \leq -\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}\right\}\right\|_{L^\nu(\mathbb{P})} + \left\|\mathbf{1}\left\{\tilde{\theta}_{2,t} \leq 0\right\}\right\|_{L^\nu(\mathbb{P})}\right] \\
&= \|\Delta L_{t+1,*}\|_{L^q(\mathbb{P})} \left(\left[\mathbb{P}\left(0 < z_{Mt} \leq -\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}\right)\right]^{1/\nu} + \left[\mathbb{P}\left(\tilde{\theta}_{2,t} \leq 0\right)\right]^{1/\nu}\right) \qquad \text{(B.28)}
\end{aligned}$$

where (i) follows by Holder's inequality. By the assumptions of Proposition 4, the p.d.f of z_{Mt} in a fixed neighborhood of zero is bounded above by some constant $K_0 > 0$. Recall constants $G_1, ..., G_4 > 0$ in the statement of Lemma 11. Hence,

$$\begin{aligned}
&\mathbb{P}\left(0 < z_{Mt} \leq -\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}\right) \\
&\leq \mathbb{P}\left(0 < z_{Mt} \leq \left|\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}\right|\right) \\
&\leq \mathbb{P}\left(0 < z_{Mt} \leq G_1 T^{\alpha_{x,1}+\alpha_{z,1}-2\alpha_{x,2}}\right) + \mathbb{P}\left(\left|\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}\right| \geq G_1 T^{\alpha_{x,1}+\alpha_{z,1}-2\alpha_{x,2}}\right) \\
&\overset{(i)}{\leq} K_0 G_1 T^{\alpha_{x,1}+\alpha_{z,1}-2\alpha_{x,2}} + G_2 T^{1-p/2} \log^{-G_3}, \qquad \text{(B.29)}
\end{aligned}$$

where (i) follows by the bounded p.d.f of z_{Mt} near zero and $T^{\alpha_{x,1}+\alpha_{z,1}-2\alpha_{x,2}} = o(1)$, as well as by Lemma 9.

It is not hard to show that $r/3 > 2 + r/(r-2) = 2 + \nu$ for $r \geq 10$. Fix $p \in (2+\nu, r/3)$. Now (B.28), (B.29) and Lemma 11 imply that for some constants $K_1, K_2 > 0$

$$\begin{aligned}
&\mathbb{E}\left(|\Delta L_{t+1,*}| \left|\mathbf{1}\{z_t'\tilde{\theta}_t > 0\} - \mathbf{1}\{z_{Mt} > 0\}\right|\right) \\
&\qquad\leq K_1 \|\Delta L_{t+1,*}\|_{L^q(\mathbb{P})} \left[T^{(\alpha_{x,1}+\alpha_{z,1}-2\alpha_{x,2})/\nu} + \left(T^{1-p/2} \log^{-K_2}\right)^{1/\nu}\right]. \qquad \text{(B.30)}
\end{aligned}$$

By (B.9), we have that

$$\mathbb{E}\Delta L_{t+1,*}\mathbf{1}\{z_{Mt}>0\} \geq 2\mathbb{E}\left[\varepsilon_{t+1}\left(\beta_2 x_{2,t}-\beta_1 x_{1,t}\right)\mathbf{1}\{z_{Mt}>0\}\right]-A_t, \tag{B.31}$$

where

$$\begin{aligned} A_t &= 2\mathbb{E}\left|\varepsilon_{t+1}\left(\delta_{2,t,*}x_{2,t}-\delta_{1,t,*}x_{1,t}\right)\right| \\ &+\mathbb{E}\left|\left(\beta_2 x_{2,t}-\beta_1 x_{1,t}-\delta_{1,t,*}x_{1,t}+\delta_{2,t,*}x_{2,t}\right)\left(\beta_1 x_{1,t}+\beta_2 x_{2,t}-\delta_{1,t,*}x_{1,t}-\delta_{2,t,*}x_{2,t}\right)\right|. \end{aligned}$$

After computations similar to (B.14) in the proof of Lemma 8, we can use the rate conditions in the assumptions of Proposition 4 and show that for some constant $K_3>0$,

$$A_t \leq K_3 T^{-2\alpha_{x,2}}. \tag{B.32}$$

(B.31) and (B.32) imply that for some constants $K_4, K_5, K_6>0$, we have that for $T\geq K_4$,

$$\begin{aligned} \mathbb{E}\Delta L_{t+1,*}\mathbf{1}\{z_{Mt}>0\} &\geq 2\mathbb{E}\left[\varepsilon_{t+1}\left(\beta_2 x_{2,t}-\beta_1 x_{1,t}\right)\mathbf{1}\{z_{Mt}>0\}\right]-K_3T^{-1/2}-K_4T^{-2\alpha_{x,2}} \\ &\overset{(i)}{\geq} K_5 T^{-\alpha_{x,1}-\alpha_{z,1}}-K_3T^{-2\alpha_{x,2}}, \end{aligned}$$

where (i) holds by the assumptions of Proposition 4. By the above display and (B.30), we have that for large T,

$$\begin{aligned} \mathbb{E}S_{t+1} \geq\ & K_5 T^{-\alpha_{x,1}-\alpha_{z,1}}-K_3T^{-2\alpha_{x,2}} \\ & -K_1\|\Delta L_{t+1,*}\|_{L^q(\mathbb{P})}\left[T^{(\alpha_{x,1}+\alpha_{z,1}-2\alpha_{x,2})/\nu}+\left(T^{1-p/2}\log^{-K_2}\right)^{1/\nu}\right]. \end{aligned}$$

Recall that in Step 1 of the proof of Lemma 9, we have that $m^{\alpha_{x,2}}\Delta L_{t+1,*}$ has uniformly bounded $0.5r$-th moments. Since $q=0.5r$, we have that $\|m^{\alpha_{x,2}}\Delta L_{t+1,*}\|_{L^q(\mathbb{P})}$ is bounded above by some constant $K_7>0$. Hence,

$$\begin{aligned} \mathbb{E}S_{t+1} \geq\ & K_5 T^{-\alpha_{x,1}-\alpha_{z,1}}-K_3T^{-2\alpha_{x,2}} \\ & -K_1K_7T^{-\alpha_{x,2}}\left[T^{(\alpha_{x,1}+\alpha_{z,1}-2\alpha_{x,2})/\nu}+\left(T^{1-p/2}\log^{-K_2}\right)^{1/\nu}\right]. \end{aligned} \tag{B.33}$$

Since $p>\nu+2$ and $\nu=r/(r-2)$, it is not hard to show that $-\alpha_{x,2}+(1-$

$p/2)/\nu < (1-p/2)/\nu < -1/2 < -\alpha_{x,1} - \alpha_{z,1}$. By the assumptions of Proposition 4, $-2\alpha_{x,2} < -\alpha_{x,1} - \alpha_{z,1}$ and $-\alpha_{x,2} + (\alpha_{x,1} + \alpha_{z,1} - 2\alpha_{x,2})/\nu < -\alpha_{x,1} - \alpha_{z,1}$. The desired result follows by (B.33). □

Proof of Proposition 4. Part (1) follows by Lemma 12 and the arguments in the proof of Proposition 1. Part (2) follows by Part (1) and Proposition 1. □

B.5 Proof of Proposition 5

Our proof of Proposition 5 relies on three lemmas, lemmas 13-15. We first state and prove these lemmas before proving Proposition 5. For notational simplicity, we write $(\hat{\mu}_t, \hat{\beta}_t, \tilde{\mu}_t)$ instead of $(\hat{\mu}_{n,t}, \hat{\beta}_{n,t}, \tilde{\mu}_{n,t})$.

Lemma 13. *Let Assumption 4 hold. Define $\Delta_{t,big} = (\hat{\mu}_t - \mu, \hat{\beta}_t - \beta)'$ and $\delta_{t,small} = \tilde{\mu}_t - \mu$. For any constants $h \in (0,1)$ and $p_1 \in (2, r/2)$, we can enlarge the probability space and construct random variables $\delta_{t,small,*}$, $\Delta_{t,big,*}$, $\bar{\delta}_{t,small}$ and $\bar{\Delta}_{t,big}$ such that*

$$\begin{cases} \mathbb{P}\left(\delta_{t,small,*} \neq \delta_{t,small}\right) \leq C_1 T^{-\min\{p_1/2-1,\ (1-h)p_1\}} \\ \mathbb{P}\left(|\delta_{t,small,*} - \bar{\delta}_{t,small}| \leq T^{-h}\right) = 1 \\ \bar{\delta}_{t,small} \text{ is independent of } \{x_s, \varepsilon_{s+1}\}_{s \geq t-1} \\ |\mathbb{E}\bar{\delta}_{t,small}| \leq T^{-1}\sqrt{\log T} C_2 \text{ and } T^{-1}C_3 \leq \mathbb{E}\bar{\delta}^2_{t,small} \leq T^{-1}C_4 \\ \left|\mathbb{E}\bar{\delta}^2_{t,small} - n^{-2}\mathbb{E}(\sum_{s=t-n}^{t-1} x_s\beta + \varepsilon_{s+1})^2\right| \leq C_5\sqrt{T^{-3}\log T} \end{cases}$$

and

$$\begin{cases} \mathbb{P}\left(\Delta_{t,big,*} \neq \Delta_{t,big}\right) \leq C_1 T^{-\min\{p_1/2-1,\ (1-h)p_1\}} \\ \mathbb{P}\left(\|\Delta_{t,big,*} - \bar{\Delta}_{t,big}\|_\infty \leq T^{-h}\right) = 1 \\ \bar{\Delta}_{t,big} \text{ is independent of } \{x_s, \varepsilon_{s+1}\}_{s \geq t-1} \\ \|\mathbb{E}\bar{\Delta}_{t,big}\|_\infty \leq T^{-1}\sqrt{\log T} C_2 \text{ and } T^{-1}C_3 \leq \mathbb{E}\|\bar{\Delta}_{t,big}\|^2_\infty \leq T^{-1}C_4 \\ \left\|\mathbb{E}\bar{\Delta}_{t,big}\bar{\Delta}'_{t,big} - \Sigma^{-1}_{X,t}\mathbb{E}(n^{-2}\sum_{s=t-n}^{t-1} \bar{x}_s\bar{x}'_s\varepsilon^2_{s+1})\Sigma^{-1}_{X,t}\right\|_\infty \leq C_5\sqrt{T^{-3}\log T}, \end{cases}$$

where $\bar{x}_t = (1, x_t)'$, $\Sigma_{X,t} = n^{-1}\sum_{s=t-n}^{t-1} \bar{x}_s\bar{x}'_s$ and $C_1, ..., C_4 > 0$ are constants depending only on the constants in Assumption 4.

Proof. The result follows by essentially the same argument as in the proof of Lemma

7. For results on $\Delta_{t,big}$, adjustments to allow multivariate $\bar{x}_s$ are needed but the arguments are essentially identical. $\square$

Lemma 14. *Let Assumption 4 hold. Let $\alpha_x < 1/2$. Then $\forall p_1 \in (2, r/2)$ and $\forall h \in (0,1)$, there exist constants $G_1, ..., G_4 > 0$ and an array of random variables $\{\Delta L_{t+1,*}\}_{t=n}^{T-n}$ such that for $T \geq G_1$,*

$$\mathbb{P}\left(\bigcap_{t=n}^{T-1} \{\Delta L_{t+1,*} = \Delta L_{t+1}\}\right) \geq 1 - G_2 T^{1-\min\{p_1/2-1,\ (1-h)p_1\}}$$

and

$$G_3 T^{-2\alpha_x} \leq \mathbb{E}\Delta L_{t+1,*} \leq G_4 T^{-2\alpha_x}.$$

Proof. Recall from (4.4) that

$$\Delta L_{t+1} = \beta^2 x_t^2 + 2\beta x_t \varepsilon_{t+1} + \delta_{t,small}^2 - \delta_{t,big}^2 + 2\delta_{t,big}\varepsilon_{t+1} - 2\delta_{t,small}(\beta x_t + \varepsilon_{t+1}). \quad \text{(B.34)}$$

Recall $\delta_{t,small} = \tilde{\mu}_t - \mu$. Let $\Delta_{t,big} = (\hat{\mu}_t - \mu, \hat{\beta}_t - \beta)'$. Let $\delta_{t,small,*}$, $\Delta_{t,big,*}$, $\bar{\delta}_{t,small}$ and $\bar{\Delta}_{t,big}$ be defined as in Lemma 13. We define $\delta_{t,big,*} = \Delta'_{t,big,*}\bar{x}_t$ with $\bar{x}_t = (1, x_t)'$ and

$$\Delta L_{t+1,*} = \beta^2 x_t^2 + 2\beta x_t \varepsilon_{t+1} + \delta_{t,small,*}^2 - \delta_{t,big,*}^2 + 2\delta_{t,big,*}\varepsilon_{t+1} - 2\delta_{t,small,*}(\beta x_t + \varepsilon_{t+1}). \quad \text{(B.35)}$$

The first claim follows by Lemma 13.

By computations similar to (B.14) using Lemma 13, we have that

$$\mathbb{E}|\Delta L_{t+1,*} - \beta^2 x_t^2| \leq K T^{-h_1}$$

for some constants $K > 0$ and $h_1 \in (2\alpha_x, 1)$. The second claim follows. $\square$

Lemma 15. *Let Assumption 4 hold. Let $\alpha_x > 1/2$. Then $\forall p_1 \in (2, r/2)$ and $\forall h \in (0,1)$, there exist constants $G_1, ..., G_4 > 0$ and an array of random variables $\{\Delta L_{t+1,*}\}_{t=n}^{T-n}$ such that for $T \geq G_1$,*

$$\mathbb{P}\left(\bigcap_{t=n}^{T-1} \{\Delta L_{t+1,*} = \Delta L_{t+1}\}\right) \geq 1 - G_2 T^{1-\min\{p_1/2-1,\ (1-h)p_1\}}$$

and

$$G_3 T^{-1} \le \mathbb{E}\Delta L_{t+1,*} \le G_4 T^{-1}.$$

Proof. Consider $\Delta L_{t+1,*}$ defined in (B.35). Recall $\delta_{t,small,*}$, $\Delta_{t,big,*}$, $\bar{\delta}_{t,small}$ and $\bar{\Delta}_{t,big}$ be defined as in Lemma 13. Define

$$\overline{\Delta L}_{t+1} = \beta^2 x_t^2 + 2\beta x_t \varepsilon_{t+1} + \bar{\delta}_{t,small}^2 - \bar{\delta}_{t,big}^2 + 2\bar{\delta}_{t,big}\varepsilon_{t+1} - 2\bar{\delta}_{t,small}(\beta x_t + \varepsilon_{t+1}).$$

By computations similar to (B.14) using Lemma 13, we have that

$$\mathbb{E}|\Delta L_{t+1,*} - \overline{\Delta L}_{t+1}| \le K T^{-1/2-\alpha_x} \tag{B.36}$$

for some constant $K > 0$. Let $\bar{x}_t = (1, x_t)'$. Now we compute

$$\begin{aligned}\mathbb{E}\overline{\Delta L}_{t+1} &= \beta^2 \mathbb{E}x_t^2 + \mathbb{E}\bar{\delta}_{t,small}^2 - \mathbb{E}\left(\bar{x}_t'\bar{\Delta}_{t,big}\right)^2 + 2\mathbb{E}\bar{\Delta}_{t,big}'\bar{x}_t\varepsilon_{t+1} - 2\mathbb{E}\bar{\delta}_{t,small}(\beta x_t + \varepsilon_{t+1}) \\ &\overset{(i)}{=} \beta^2 \mathbb{E}x_t^2 + \mathbb{E}\bar{\delta}_{t,small}^2 - \mathbb{E}\left(\bar{x}_t'\bar{\Delta}_{t,big}\bar{\Delta}_{t,big}'\bar{x}_t\right),\end{aligned}$$

where (i) follows by the fact that $\bar{\Delta}_{t,big}$ and $\bar{\delta}_{t,small}$ are independent of $\bar{x}_t$ and ε_{t+1}. By Lemma 13,

$$\left|\left[\mathbb{E}\bar{\delta}_{t,small}^2 - \mathbb{E}\left(\bar{x}_t'\bar{\Delta}_{t,big}\right)^2\right] - \left[n^{-2}\mathbb{E}(\sum_{s=t-n}^{t-1} x_s\beta + \varepsilon_{s+1})^2 - \mathbb{E}(\bar{x}_t'\Sigma_{X,t}^{-1}\Omega_t\Sigma_{X,t}^{-1}\bar{x}_t)\right]\right| \le K_1\sqrt{T^{-3}\log T}$$

for some constant $K_1 > 0$, where $\Sigma_{X,t} = n^{-1}\sum_{s=t-n}^{t-1}\mathbb{E}\bar{x}_s\bar{x}_s'$ and $\Omega_t = \mathbb{E}(n^{-2}\sum_{s=t-n}^{t-1}\bar{x}_s\bar{x}_s'\varepsilon_{s+1}^2)$. Notice that $\Sigma_{X,t} = \mathrm{diag}(1, \sigma_{x,t}^2)$ with $\sigma_{x,t}^2 = n^{-1}\sum_{s=t-n}^{t-1}\mathbb{E}x_s^2$. Hence,

$$\begin{aligned}&\mathbb{E}(\bar{x}_t'\Sigma_{X,t}^{-1}\Omega_t\Sigma_{X,t}^{-1}\bar{x}_t) \\ &\qquad = n^{-2}\sum_{s=t-n}^{t-1}\mathbb{E}\varepsilon_{s+1}^2 + n^{-2}\sum_{s=t-n}^{t-1}\sigma_{x,t}^{-4}\mathbb{E}(x_t^2 x_s^2\varepsilon_{s+1}^2) + 2n^{-2}\sum_{s=t-n}^{t-1}\sigma_{x,t}^{-2}\mathbb{E}x_t x_s\varepsilon_{s+1}^2.\end{aligned}$$

It follows that

$$n^{-2}\mathbb{E}\left(\sum_{s=t-n}^{t-1} x_s\beta + \varepsilon_{s+1}\right)^2 - \mathbb{E}(\bar{x}_t'\Sigma_{X,t}^{-1}\Omega_t\Sigma_{X,t}^{-1}\bar{x}_t)$$

$$= n^{-1}\beta^2\sigma_{x,t}^2 - n^{-2}\sum_{s=t-n}^{t-1}\sigma_{x,t}^{-4}\mathbb{E}(x_t^2x_s^2\varepsilon_{s+1}^2) - 2n^{-2}\sum_{s=t-n}^{t-1}\sigma_{x,t}^{-2}\mathbb{E}x_tx_s\varepsilon_{s+1}^2.$$

By computations based on the exponential decay of beta-mixing coefficients similar to (A.10) and (A.11), it is not hard to show that $\sum_{s=t-n}^{t-1}\sigma_{x,t}^{-2}\mathbb{E}x_tx_s\varepsilon_{s+1}^2$ is uniformly bounded by a constant $K_2 > 0$. Hence,

$$\left|\mathbb{E}\overline{\Delta L}_{t+1} - \beta^2\mathbb{E}x_t^2 - n^{-1}\beta^2\sigma_{x,t}^2 + n^{-2}\sum_{s=t-n}^{t-1}\sigma_{x,t}^{-4}\mathbb{E}(x_t^2x_s^2\varepsilon_{s+1}^2)\right| \le 2n^{-2}K_2 + K_1\sqrt{T^{-3}\log T}.$$

Since $\sigma_{x,t}^{-4}\mathbb{E}(x_t^2x_s^2\varepsilon_{s+1}^2)$ and $\sigma_{x,t}^2$ are bounded away from zero and infinity, $\beta \asymp T^{-\alpha_x}$ with $\alpha_x > 1/2$, it follows that $-\mathbb{E}\overline{\Delta L}_{t+1} \asymp T^{-1}$. The desired result follows by (B.36) and $\alpha_x > 1/2$. □

Proof of Proposition 5. Part (1) follows by Lemma 14 and the arguments in the proof of Proposition 1. Part (2) follows by Lemma 15 and the arguments in the proof of Proposition 1. □

B.6 Proof of Proposition 6

Our proof of Proposition 6 relies on two lemmas, Lemmas 16 and 17. We first state and prove these lemmas before proving Proposition 6.

Lemma 16. *Suppose that the assumptions of Proposition 6 hold. Let $\Delta L_{t+1,*}$ be defined as in (B.9) in the proof of Lemma 8. Let $\tilde{\theta}_t = (\tilde{\theta}_{1,t}, \tilde{\theta}_{2,t})' = (\sum_{s=t-m}^{t-1} z_sz_s')^{-1}(\sum_{s=t-m}^{t-1} z_s\Delta L_{t+1,*})$, where $z_s = (1, z_{Ms})'$. Fix $p \in (2, r/3)$. Then there exist some constants $G_0, G_1, G_2, G_3 > 0$ such that for $T \ge G_0$,*

$$\mathbb{P}\left(\left|\frac{\tilde{\theta}_{1,t}}{\tilde{\theta}_{2,t}}\right| \ge G_1T^{\alpha_z-\alpha_x}\right) \le G_2T^{1-p/2}\log^{-G_3}$$

and

$$\mathbb{P}\left(\tilde{\theta}_{2,t} \le 0\right) \le G_2T^{1-p/2}\log^{-G_3}.$$

Proof. The proof is similar to the proof of Lemma 11. Let $\Psi_{t+1} = m^{\alpha_x}\Delta L_{t+1,*}$, $\pi_t = (\sum_{s=t-m}^{t-1}\mathbb{E}z_sz_s')^{-1}(\sum_{s=t-m}^{t-1}\mathbb{E}z_s\Psi_{s+1})$, $\hat{\pi}_t = (\sum_{s=t-m}^{t-1}z_sz_s')^{-1}(\sum_{s=t-m}^{t-1}z_s\Psi_{s+1})$ and $\{\xi_{s+1}\}_{s=t-m}^{t-1}$ with $\xi_{s+1} = \Psi_{s+1} - z_s'\pi_t$. Clearly, $\hat{\pi}_t = m^{\alpha_x}\tilde{\theta}_t$. Let $\gamma_t = \hat{\pi}_t - \pi_t$. By the same argument as Step 1 in the proof of Lemma 11, we can show that there exist constants $M_1, M_2, M_3 > 0$ such that

$$\mathbb{P}\left(\|\gamma_t\|_\infty \geq M_1\sqrt{T^{-1}\log T}\right) \leq M_2 T^{1-p/2}\log^{-M_3}. \tag{B.37}$$

Now we characterize π_t. By Lemma 8, there are constants $M_4, M_5 > 0$ such that

$$M_4 T^{-\alpha_x} \leq \pi_{1,t} \leq M_5 T^{-\alpha_x}. \tag{B.38}$$

By computations similar to (B.14) in the proof of Lemma 8, one can show that

$$\mathbb{E}\left|z_{Mt}\Delta L_{t+1,*} - z_{Mt}(2\beta x_t\varepsilon_{t+1} + \beta^2 x_t^2)\right| \leq M_6 T^{-1/2}\sqrt{\log T},$$

where $M_6 > 0$ is a constant. By the assumptions of Proposition 6, there exists a constant $M_7 > 0$ with

$$\mathbb{E}\left[z_{Mt}\left(2\beta x_t\varepsilon_{t+1} + \beta^2 x_t^2\right)\right] \geq M_7 T^{-\alpha_x-\alpha_z}.$$

Hence, for some constant $M_8 > 0$ we have

$$\pi_{2,t} \geq M_8 T^{-\alpha_z}. \tag{B.39}$$

Let $x = 2M_5 T^{-\alpha_x}$ and $G = 2x/(M_8 T^{-\alpha_z})$. Then

$$\begin{aligned}
&\mathbb{P}\left(|\hat{\pi}_{1,t}| \geq G|\hat{\pi}_{2,t}|\right)\\
&\leq \mathbb{P}\left(|\hat{\pi}_{1,t}| \geq x\right) + \mathbb{P}\left(|\hat{\pi}_{2,t}| \leq x/G\right)\\
&\leq \mathbb{P}\left(|\hat{\pi}_{1,t} - \pi_{1,t}| \geq x - |\pi_1|\right) + \mathbb{P}\left(|\hat{\pi}_{2,t} - \pi_{2,t}| \geq |\pi_{2,t}| - x/G\right)\\
&\overset{(i)}{\leq} \mathbb{P}\left(|\hat{\pi}_{1,t} - \pi_{1,t}| \geq x - M_5 T^{-\alpha_x}\right) + \mathbb{P}\left(|\hat{\pi}_{2,t} - \pi_{2,t}| \geq M_8 T^{-\alpha_z} - x/G\right)\\
&\leq \mathbb{P}\left(\|\gamma_t\|_\infty \geq x - M_5 T^{-\alpha_x}\right) + \mathbb{P}\left(\|\gamma_t\|_\infty \geq M_8 T^{-\alpha_z} - x/G\right)\\
&= \mathbb{P}\left(\|\gamma_t\|_\infty \geq M_5 T^{-\alpha_{x,2}}\right) + \mathbb{P}\left(\|\gamma_t\|_\infty \geq M_8 T^{-\alpha_z}/2\right)\\
&\overset{(ii)}{\leq} 2M_2 T^{1-p/2}\log^{-M_3},
\end{aligned}$$

where (i) holds by (B.38) and (B.39) and (ii) follows by (B.37) together with $T^{-\alpha_x} \gg \sqrt{T^{-1}\log T}$ and $T^{-\alpha_z} \gg \sqrt{T^{-1}\log T}$. The first claim follows by $|\hat{\pi}_{1,t}/\hat{\pi}_{2,t}| = |\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}|$

and $G = 2x/(M_8 T^{-\alpha_z}) = (4M_5/M_8)T^{\alpha_z - \alpha_x}$.

To see the second claim, notice that

$$\mathbb{P}\left(\tilde{\theta}_{2,t} \leq 0\right) \overset{(i)}{=} \mathbb{P}\left(\hat{\pi}_{2,t} \leq 0\right) = \mathbb{P}\left(\gamma_{2,t} \leq -\pi_{2,t}\right)$$
$$\overset{(ii)}{\leq} \mathbb{P}\left(\gamma_{2,t} \leq -M_8 T^{-\alpha_z}\right) \leq \mathbb{P}\left(\|\gamma_t\|_\infty \geq M_8 T^{-\alpha_z}\right)$$
$$\overset{(iii)}{\leq} M_2 T^{1-p/2} \log^{-M_3},$$

where (i) holds by $\tilde{\theta}_t = m^{-\alpha_x}\hat{\pi}_t$, (ii) follows by (B.27) and (iii) holds by (B.37) and $T^{-\alpha_z} \gg \sqrt{T^{-1}\log T}$. □

Lemma 17. *Let the assumptions of Proposition 6 hold. Then $\forall p_1 \in (2, r/3)$, there exist constants $G_0, G_1, ..., G_5 > 0$ and an array $\{S_{t+1}\}_{t=n+m}^{T-1}$ such that $\forall T \geq G_0$,*

$$\mathbb{P}\left(\bigcap_{t=n+m}^{T-1} \left\{S_{t+1} = \Delta L_{t+1}\mathbf{1}\{z_t'\hat{\theta}_{m,t} > 0\}\right\}\right) \geq 1 - G_1 T^{2-p_1/2}\log^{-G_2} T$$

and

$$\mathbb{E}S_{t+1} \geq K_5 T^{-\alpha_x - \alpha_z}.$$

Proof. The proof is similar to the proof of Lemma 12. Let $\tilde{\theta}_t = (\sum_{s=t-m}^{t-1} z_s z_s')^{-1}(\sum_{s=t-m}^{t-1} z_s \Delta L_{t+1,*})$, where $\Delta L_{t+1,*}$ is defined as in (B.35). Define $S_{t+1} = \Delta L_{t+1,*}\mathbf{1}\{z_t'\tilde{\theta}_t > 0\}$. Notice that

$$\bigcap_{t=n}^{T-1} \{\Delta L_{t+1,*} = \Delta L_{t+1}\} \subseteq \bigcap_{t=n+m}^{T-1} \left\{S_{t+1} = \Delta L_{t+1}\mathbf{1}\{z_t'\hat{\theta}_{m,t} > 0\}\right\}.$$

Hence, the first claim follows by Lemma 14.

To show the second claim, let $q = (r+2)/4$ and $\nu = (r+2)/(r-2)$. Hence, $q^{-1} + \nu^{-1} = 1$. Notice that by the same argument as (B.28) in the proof of Lemma 12, we have that

$$\mathbb{E}\left(|\Delta L_{t+1,*}|\left|\mathbf{1}\{z_t'\tilde{\theta}_t > 0\} - \mathbf{1}\{z_{Mt} > 0\}\right|\right)$$
$$\leq \|\Delta L_{t+1,*}\|_{L^q(\mathbb{P})}\left(\left[\mathbb{P}\left(0 < z_{Mt} \leq -\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}\right)\right]^{1/\nu} + \left[\mathbb{P}\left(\tilde{\theta}_{2,t} \leq 0\right)\right]^{1/\nu}\right).$$

Similar to the argument in (B.29), we have

$$\begin{aligned}\mathbb{P}\left(0 < z_{Mt} \le -\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}\right) &\le \mathbb{P}\left(0 < z_{Mt} \le \left|\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}\right|\right) \\ &\le \mathbb{P}\left(0 < z_{Mt} \le G_1 T^{\alpha_z - \alpha_x}\right) + \mathbb{P}\left(\left|\tilde{\theta}_{1,t}/\tilde{\theta}_{2,t}\right| \ge G_1 T^{\alpha_z - \alpha_x}\right) \\ &\overset{(i)}{\le} K_1 G_1 T^{\alpha_z - \alpha_x} + G_2 T^{1-p/2} \log^{-G_3}\end{aligned}$$

for some constant $K_1 > 0$, where (i) holds by the bounded p.d.f of z_{Mt} around zero and Lemma 16. The above two displays and Lemma 16 imply that for some constant $K_2 > 0$,

$$\begin{aligned}&\mathbb{E}\left(|\Delta L_{t+1,*}| \left|\mathbf{1}\{z_t'\tilde{\theta}_t > 0\} - \mathbf{1}\{z_{Mt} > 0\}\right|\right) \\ &\qquad \le K_2 \|\Delta L_{t+1,*}\|_{L^q(\mathbb{P})} \left(T^{(\alpha_z - \alpha_x)/\nu} + (T^{1-p/2} \log^{-G_3})^{1/\nu}\right).\end{aligned}$$

By (B.9), we have that

$$\mathbb{E}\Delta L_{t+1,*}\mathbf{1}\{z_{Mt} > 0\} \ge 2\mathbb{E}\left[\beta x_t \varepsilon_{t+1} \mathbf{1}\{z_{Mt} > 0\}\right] - A_t,$$

where

$$A_t = \mathbb{E}\left|\beta^2 x_t^2 + \delta_{t,small,*}^2 - \delta_{t,big,*}^2 + 2\delta_{t,big}\varepsilon_{t+1} - 2\delta_{t,small,*}(\beta x_t + \varepsilon_{t+1})\right|.$$

After computations similar to (B.14) in the proof of Lemma 8, we can use the rate conditions in the assumptions of Proposition 6 and show that for some constant $K_3 > 0$,

$$A_t \le K_3(T^{-2\alpha_x} + T^{-1/2}). \tag{B.40}$$

(B.40) implies that for some constants $K_4, K_5 > 0$, we have that for $T \ge K_4$,

$$\begin{aligned}\mathbb{E}\Delta L_{t+1,*}\mathbf{1}\{z_{Mt} > 0\} &\ge 2\mathbb{E}\left[\beta x_t \varepsilon_{t+1} \mathbf{1}\{z_{Mt} > 0\}\right] - K_3(T^{-2\alpha_x} + T^{-1/2}) \\ &\overset{(i)}{\ge} K_5 T^{-\alpha_x - \alpha_z} - K_3(T^{-2\alpha_x} + T^{-1/2}),\end{aligned}$$

where (i) holds by the assumptions of Proposition 6. By the above display and (B.30), we have that for $T \ge K_4$,

$$\mathbb{E}S_{t+1} \geq K_5 T^{-\alpha_x-\alpha_z} - K_3(T^{-2\alpha_x} + T^{-1/2}) \\ - K_2 \|\Delta L_{t+1,*}\|_{L^q(\mathbb{P})} \left(T^{(\alpha_z-\alpha_x)/\nu} + (T^{1-p/2}\log^{-G_3})^{1/\nu}\right).$$

Recall that in Step 1 of the proof of Lemma 9, we have that $T^{\alpha_x}\Delta L_{t+1,*}$ has uniformly bounded $0.5r$-th moments. Since $q = (r+2)/4 < 0.5r$, we have that $\|T^{\alpha_x}\Delta L_{t+1,*}\|_{L^q(\mathbb{P})}$ is bounded above by some constant $K_6 > 0$. Hence,

$$\mathbb{E}S_{t+1} \geq K_5 T^{-\alpha_x-\alpha_z} - K_3(T^{-2\alpha_x} + T^{-1/2}) \\ - K_2 K_6 T^{-\alpha_x} \left(T^{(\alpha_z-\alpha_x)/\nu} + (T^{1-p/2}\log^{-G_3})^{1/\nu}\right).$$

It is not hard to show that $-\alpha_x - \alpha_z > -2\alpha_x$, $-\alpha_x - \alpha_z > -1/2$, $-\alpha_x - \alpha_z > -\alpha_x + (\alpha_z - \alpha_x)/\nu$ and $-\alpha_x - \alpha_z > -\alpha_x + (1-p/2)/\nu$. The desired result follows. □

Proof of Proposition 6. Part (1) follows by Lemma 17 and the arguments in the proof of Proposition 1. Part (2) follows by Part (1) and Proposition 5. □

Appendix C Appendix Tables

Table A.1: Predictive performance of nested models and dynamic rotation ($m = 50$)

	A: Big vs. small model (MSE_{big}/MSE_{small})						
$\alpha_z \backslash \alpha_x$	0	0.1	0.25	0.4	0.5	0.75	1
	0.253	0.460	0.777	0.939	0.981	1.007	1.010
	B: Dynamic rotation vs. small model (MSE_{DR}/MSE_{small})						
0.0	0.243	0.409	0.674	0.849	0.912	0.967	0.974
0.1	0.260	0.458	0.746	0.899	0.946	0.983	0.988
0.25	0.261	0.473	0.788	0.939	0.975	0.997	1.000
0.5	0.260	0.474	0.800	0.954	0.987	1.004	1.005
0.75	0.260	0.474	0.803	0.956	0.988	1.004	1.005
1.0	0.259	0.475	0.802	0.956	0.989	1.004	1.005
	C: Dynamic rotation vs. big model (MSE_{DR}/MSE_{big})						
0.0	0.964	0.890	0.868	0.904	0.930	0.961	0.965
0.1	1.030	0.998	0.960	0.957	0.965	0.977	0.979
0.25	1.033	1.028	1.014	1.000	0.995	0.991	0.990
0.5	1.030	1.031	1.030	1.016	1.007	0.996	0.995
0.75	1.029	1.030	1.032	1.018	1.008	0.997	0.996
1.0	1.028	1.030	1.032	1.018	1.008	0.997	0.996

This table reports result from 5,000 Monte Carlo simulations using a sample size of $(n, m, p) = (100, 50, 200)$.

Table A.2: Predictive performance of dynamic rotation versus model combination, pretesting and a model augmented with the monitoring instrument ($m = 50$)

$\alpha_z \backslash \alpha_x$	0	0.1	0.25	0.4	0.5	0.75	1
	A: Dynamic rotation vs. equal-weighted forecast combination (MSE_{DR}/MSE_{EW})						
0.0	0.555	0.690	0.814	0.894	0.930	0.967	0.972
0.1	0.594	0.774	0.900	0.946	0.965	0.983	0.986
0.25	0.596	0.798	0.950	0.989	0.994	0.997	0.997
0.5	0.593	0.800	0.965	1.005	1.007	1.003	1.003
0.75	0.594	0.799	0.968	1.007	1.008	1.004	1.003
1.0	0.592	0.800	0.967	1.007	1.008	1.004	1.003
	B: Dynamic rotation vs. pre-test forecast ($MSE_{DR}/MSE_{pretest}$)						
0.0	0.964	0.890	0.868	0.894	0.919	0.964	0.971
0.1	1.030	0.998	0.960	0.946	0.953	0.980	0.985
0.25	1.033	1.028	1.014	0.988	0.982	0.994	0.997
0.5	1.030	1.031	1.030	1.004	0.994	1.000	1.002
0.75	1.029	1.030	1.032	1.006	0.996	1.001	1.002
1.0	1.028	1.030	1.032	1.007	0.996	1.001	1.002
	C: Dynamic rotation vs. augmented ($x_t + z_{Mt}$) forecast (MSE_{DR}/MSE_{aug})						
0.0	0.946	0.873	0.851	0.887	0.912	0.942	0.946
0.1	1.010	0.979	0.941	0.939	0.946	0.958	0.959
0.25	1.013	1.008	0.994	0.980	0.975	0.971	0.971
0.5	1.010	1.010	1.010	0.996	0.987	0.977	0.976
0.75	1.009	1.010	1.012	0.998	0.988	0.978	0.977
1.0	1.009	1.010	1.012	0.999	0.989	0.978	0.977
	D: Dynamic rotation vs. weighted combination ($MSE_{DR}/MSE_{weight-comb}$)						
0.0	0.857	0.797	0.826	0.895	0.930	0.967	0.972
0.1	0.916	0.893	0.914	0.947	0.965	0.983	0.986
0.25	0.919	0.921	0.965	0.990	0.995	0.997	0.997
0.5	0.915	0.923	0.980	1.006	1.007	1.003	1.003
0.75	0.915	0.922	0.982	1.008	1.008	1.004	1.003
1.0	0.914	0.923	0.982	1.008	1.008	1.004	1.003

In each panel, we compare the MSE performance of the switching approach to that of an equal-weighted combination (Panel A), an approach that includes a predictor in the forecasting model if its regression coefficient is statistically significant (pretest, in Panel B), forecasts from an augmented model that includes both the predictor, x_t, and the monitoring instrument, z_{Mt}, in the forecasting model (Panel C), and forecasts from a combination scheme with weights that are proportional to the inverse of the MSE of the individual forecasts (Panel D).

Table A.3: Pairwise comparisons of predictive performance for the non-nested case ($m = 50$)

	$(j_1, j_2) = (1, 2)$				MSE_{j_2}/MSE_{j_1} $(j_1, j_2) = (2, DR)$				$(j_1, j_2) = (1, DR)$			
					A: $(\alpha_{z,1}, \alpha_{z,2}) = (0, 0)$							
$\alpha_{x,1} \backslash \alpha_{x,2}$	0	0.25	0.5	1	0	0.25	0.5	1	0	0.25	0.5	1
0.0	1.001	4.377	6.609	7.008	0.998	0.231	0.153	0.145	0.999	1.010	1.013	1.016
0.25	0.229	1.000	1.511	1.597	1.027	0.997	0.672	0.625	0.235	0.998	1.015	0.998
0.5	0.152	0.662	0.999	1.059	1.041	1.017	0.992	0.929	0.158	0.674	0.991	0.984
1.0	0.143	0.626	0.943	1.000	1.045	1.001	0.984	0.984	0.150	0.626	0.928	0.984
					B: $(\alpha_{z,1}, \alpha_{z,2}) = (0, 1)$							
$\alpha_{x,1} \backslash \alpha_{x,2}$	0	0.25	0.5	1	0	0.25	0.5	1	0	0.25	0.5	1
0.0	1.000	4.367	6.604	7.017	0.960	0.231	0.152	0.144	0.960	1.009	1.007	1.007
0.25	0.229	1.003	1.511	1.595	1.032	0.927	0.680	0.642	0.236	0.930	1.027	1.025
0.5	0.151	0.662	1.001	1.060	1.043	0.997	0.965	0.954	0.158	0.660	0.966	1.011
1.0	0.143	0.626	0.944	1.000	1.047	0.999	0.980	0.992	0.150	0.626	0.925	0.992
					C:$(\alpha_{z,1}, \alpha_{z,2}) = (0.5, 0.5)$							
$\alpha_{x,1} \backslash \alpha_{x,2}$	0	0.25	0.5	1	0	0.25	0.5	1	0	0.25	0.5	1
0.0	1.001	4.372	6.607	6.996	0.998	0.231	0.152	0.144	0.999	1.008	1.007	1.007
0.25	0.229	1.000	1.513	1.598	1.025	1.000	0.681	0.641	0.235	0.999	1.030	1.024
0.5	0.152	0.662	1.000	1.059	1.036	1.033	1.000	0.960	0.157	0.684	1.000	1.017
1.0	0.143	0.626	0.944	1.000	1.036	1.027	1.017	1.002	0.148	0.642	0.960	1.002
					D: $(\alpha_{z,1}, \alpha_{z,2}) = (1, 1)$							
$\alpha_{x,1} \backslash \alpha_{x,2}$	0	0.25	0.5	1	0	0.25	0.5	1	0	0.25	0.5	1
0.0	0.999	4.376	6.603	6.994	0.999	0.230	0.153	0.144	0.999	1.008	1.007	1.007
0.25	0.228	0.998	1.509	1.600	1.026	1.001	0.684	0.640	0.234	0.999	1.031	1.024
0.5	0.152	0.663	0.999	1.060	1.034	1.034	1.001	0.960	0.157	0.685	0.999	1.017
1.0	0.143	0.625	0.944	1.000	1.038	1.027	1.018	1.002	0.148	0.642	0.961	1.002

This table reports MSE of methods j_1 and j_2. Data are generated according to the non-nested model

$$y_{t+1} = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \varepsilon_{t+1},$$

where x_{1t} and x_{2t} are a set of predictor variables that are known at time t. Model 1 takes the form $y_{t+1} = \beta_1 x_{1,t} + \varepsilon_{1t+1}$, while model 2 takes the form $y_{t+1} = \beta_2 x_{2,t} + \varepsilon_{2t+1}$. The strength of the predictors in models 1 and 2 is parameterized as $\beta_j = c_{\beta,j} n^{-\alpha_{x,j}}$, while the accuracy of the monitoring instrument is captured as $corr(x_{j,t}\varepsilon_{t+1}, z_{1t}) = c_{\rho,j} m^{-\alpha_{z,j}}$. All results are based on 5,000 MC simulations and use a sample size of $(n, m, p) = (100, 50, 200)$.

Table A.4: Predictive performance of nested models and dynamic rotation ($m = 200$)

	A: Big model vs. small model (MSE_{big}/MSE_{small})						
$\alpha_z \backslash \alpha_x$	0	0.1	0.25	0.4	0.5	0.75	1
	0.253	0.461	0.777	0.939	0.981	1.007	1.010
	B: Dynamic rotation vs. small model (MSE_{DR}/MSE_{small})						
0.0	0.240	0.405	0.670	0.846	0.912	0.978	0.989
0.1	0.258	0.454	0.736	0.892	0.941	0.987	0.994
0.25	0.257	0.465	0.777	0.929	0.969	0.997	1.000
0.5	0.257	0.464	0.783	0.945	0.985	1.002	1.003
0.75	0.257	0.464	0.782	0.947	0.987	1.003	1.004
1.0	0.256	0.464	0.783	0.948	0.987	1.003	1.004
	C: Dynamic rotation vs. big model (MSE_{DR}/MSE_{big})						
0.0	0.952	0.880	0.863	0.901	0.929	0.971	0.979
0.1	1.016	0.985	0.947	0.949	0.960	0.981	0.985
0.25	1.017	1.011	1.000	0.990	0.988	0.990	0.991
0.5	1.015	1.008	1.007	1.006	1.004	0.996	0.994
0.75	1.015	1.007	1.007	1.008	1.006	0.996	0.994
1.0	1.015	1.007	1.007	1.009	1.006	0.997	0.994

This table reports results from 5,000 Monte Carlo simulations using a sample size of $(n, m, p) = (100, 200, 200)$.

Table A.5: Predictive performance of dynamic rotation versus model combination, pretesting and a model augmented with the monitoring instrument ($m = 200$)

$\alpha_z \backslash \alpha_x$	0	0.1	0.25	0.4	0.5	0.75	1
	A: Dynamic rotation vs. equal-weighted combination (MSE_{DR}/MSE_{EW})						
0.0	0.548	0.683	0.809	0.891	0.930	0.978	0.987
0.1	0.587	0.765	0.888	0.939	0.960	0.987	0.992
0.25	0.587	0.785	0.937	0.978	0.988	0.997	0.998
0.5	0.585	0.783	0.944	0.995	1.004	1.002	1.001
0.75	0.586	0.782	0.944	0.997	1.006	1.003	1.002
1.0	0.585	0.782	0.944	0.998	1.006	1.003	1.001
	B: Dynamic rotation vs. pre-test forecast ($MSE_{DR}/MSE_{pretest}$)						
0.0	0.952	0.880	0.863	0.891	0.918	0.975	0.986
0.1	1.016	0.985	0.947	0.938	0.948	0.984	0.991
0.25	1.017	1.011	1.000	0.978	0.976	0.994	0.997
0.5	1.015	1.008	1.007	0.995	0.992	0.999	1.000
0.75	1.015	1.007	1.007	0.997	0.994	1.000	1.001
1.0	1.015	1.007	1.007	0.997	0.994	1.000	1.001
	C: Dynamic rotation vs. augmented ($x_t + z_{Mt}$) forecast (MSE_{DR}/MSE_{aug})						
0.0	0.934	0.863	0.846	0.884	0.911	0.952	0.961
0.1	0.996	0.966	0.929	0.931	0.941	0.962	0.966
0.25	0.998	0.992	0.981	0.970	0.969	0.971	0.971
0.5	0.996	0.988	0.987	0.987	0.985	0.976	0.974
0.75	0.996	0.988	0.987	0.988	0.987	0.977	0.975
1.0	0.995	0.988	0.987	0.989	0.986	0.977	0.975
	D: Dynamic rotation vs. weighted combination ($MSE_{DR}/MSE_{weight-comb}$)						
0.0	0.850	0.790	0.821	0.892	0.930	0.978	0.987
0.1	0.907	0.885	0.902	0.939	0.960	0.987	0.992
0.25	0.908	0.908	0.952	0.979	0.988	0.997	0.998
0.5	0.907	0.905	0.959	0.996	1.004	1.002	1.001
0.75	0.907	0.904	0.959	0.998	1.006	1.003	1.002
1.0	0.906	0.905	0.959	0.999	1.006	1.003	1.001

In each panel, we compare the MSE performance of the dynamic rotation approach to that of an equal-weighted combination (Panel A), an approach that includes a predictor in the forecasting model if its regression coefficient is statistically significant (pretest, in Panel B), forecasts from an augmented model that includes both the predictor, x_t, and the monitoring instrument, z_{Mt}, in the forecasting model (Panel C), and forecasts from a combination scheme with weights proportional to the inverse of the MSE of the individual forecasts (Panel D).

Table A.6: Pairwise comparisons of predictive performance for the non-nested case ($m = 200$)

	MSE_{j_2}/MSE_{j_1}											
	$(j_1, j_2) = (1, 2)$				$(j_1, j_2) = (2, DR)$				$(j_1, j_2) = (1, DR)$			
	A: $(\alpha_{z,1}, \alpha_{z,2}) = (0, 0)$											
$\alpha_{x,1} \backslash \alpha_{x,2}$	0	0.25	0.5	1	0	0.25	0.5	1	0	0.25	0.5	1
0.0	1.000	4.375	6.596	6.994	1.000	0.229	0.152	0.143	1.000	1.001	1.001	1.002
0.25	0.228	1.000	1.510	1.601	1.017	0.998	0.663	0.617	0.232	0.998	1.001	0.988
0.5	0.152	0.662	1.000	1.059	1.030	1.003	0.997	0.923	0.156	0.664	0.997	0.977
1.0	0.143	0.626	0.943	1.000	1.031	0.991	0.977	0.993	0.147	0.620	0.922	0.993
	B: $(\alpha_{z,1}, \alpha_{z,2}) = (0, 1)$											
$\alpha_{x,1} \backslash \alpha_{x,2}$	0	0.25	0.5	1	0	0.25	0.5	1	0	0.25	0.5	1
0.0	1.002	4.369	6.623	7.001	0.933	0.229	0.151	0.143	0.935	1.000	1.000	1.000
0.25	0.228	1.001	1.510	1.598	1.019	0.899	0.665	0.627	0.233	0.900	1.005	1.002
0.5	0.151	0.663	1.001	1.060	1.030	0.984	0.952	0.948	0.156	0.653	0.953	1.005
1.0	0.143	0.625	0.944	1.000	1.032	0.988	0.971	0.995	0.147	0.617	0.917	0.995
	C: $(\alpha_{z,1}, \alpha_{z,2}) = (0.5, 0.5)$											
$\alpha_{x,1} \backslash \alpha_{x,2}$	0	0.25	0.5	1	0	0.25	0.5	1	0	0.25	0.5	1
0.0	0.997	4.380	6.620	7.019	1.002	0.228	0.151	0.143	0.999	1.000	1.000	1.000
0.25	0.228	0.999	1.511	1.602	1.017	0.999	0.664	0.626	0.232	0.999	1.003	1.002
0.5	0.152	0.664	1.001	1.060	1.029	1.006	1.000	0.950	0.156	0.668	1.000	1.007
1.0	0.143	0.624	0.943	1.000	1.030	1.005	1.008	1.002	0.147	0.627	0.950	1.002
	D: $(\alpha_{z,1}, \alpha_{z,2}) = (1, 1)$											
$\alpha_{x,1} \backslash \alpha_{x,2}$	0	0.25	0.5	1	0	0.25	0.5	1	0	0.25	0.5	1
0.0	1.001	4.366	6.617	7.010	0.999	0.229	0.151	0.143	1.000	1.000	1.000	1.000
0.25	0.229	1.000	1.510	1.600	1.017	1.001	0.664	0.626	0.233	1.001	1.003	1.002
0.5	0.151	0.664	1.000	1.059	1.028	1.006	1.001	0.951	0.156	0.667	1.000	1.007
1.0	0.143	0.625	0.944	1.000	1.030	1.005	1.007	1.002	0.147	0.628	0.951	1.002

This table reports MSE of methods j_1 and j_2. Data are generated according to the non-nested model

$$y_{t+1} = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \varepsilon_{t+1},$$

where x_{1t} and x_{2t} are a set of predictor variables that are known at time t. Model 1 takes the form $y_{t+1} = \beta_1 x_{1,t} + \varepsilon_{1t+1}$, while model 2 takes the form $y_{t+1} = \beta_2 x_{2,t} + \varepsilon_{2t+1}$. The strength of the predictors in models 1 and 2 is parameterized as $\beta_j = c_{\beta,j} n^{-\alpha_{x,j}}$, while the accuracy of the monitoring instrument is captured as $corr(x_{j,t}\varepsilon_{t+1}, z_{1t}) = c_{\rho,j} m^{-\alpha_{z,j}}$. All results are based on 5,000 MC simulations and use a sample size of $(n, m, p) = (100, 200, 200)$.