Testing for Structural Stability of Factor Augmented Forecasting Models*

Valentina Corradi[1] and Norman R. Swanson[2]

[1]University of Warwick and [2]Rutgers University

November 2011

Abstract

In recent years, there has been increasing interest in the problem of testing for the constancy of factor loadings. Nevertheless, to the best of our knowledge, there is no consistent test for the structural stability of forecasting models estimated using a vector of factors (i.e. diffusion indexes). The aim of this paper is to fill this gap, by introducing a test for the null hypothesis of equality of expected forecast error loss based on (i) full sample estimation of factors and associated factor augmented forecasting model; and (ii) analogous expected forecast error loss based on rolling estimation. In certain cases, when parameter estimation error vanishes, the limiting distribution of the suggested statistic may be degenerate. We overcome this problem via the use of *m out of n* (*moon*) bootstrap critical values. The use of this bootstrap approach ensures that in the degenerate case the bootstrap statistic approaches zero at a slower rate than the actual statistic. We provide an empirical illustration by testing for the structural stability of factor augmented forecasting models for 11 U.S. macroeconomic indicators.

Keywords: diffusion index, factors, forecast failure, forecast stability, *m out of n* bootstrap
JEL classification: C12, C22, C53.

* Valentina Corradi, Department of Economics, University of Warwick, Coventry, CV4 7AL, UK, v.corradi@warwick.ac.uk. Norman Swanson, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA, nswanson@econ.rutgers.edu. For useful discussions and comments on the paper, we would like to thank Silvia Goncalves, David Hendry, Atsushi Inoue, Barbara Rossi, Moto Shintani, seminar participants at the Bank of Canada, and conference participants at the Festschrift conference on Causality, Prediction, and Specification Analysis: Recent Advances and Future Directions in Honor of Halbert White at the University of California, San Diego on May 6-7, 2011, the fifth CIREQ Conference on Time Series.at the University of Montreal on May 27-28, 2011, and the Netherlands Econometric Study Group at Erasmus University on June 10, 2011. Excellent research assistance was provided by Hyun Hak Kim.

1 Introduction

The issue of forecast instability arising because of structural instability has received considerable attention in recent years (see e.g. Clements and Hendry 2002, Hendry and Mizon 2005, and Castle, Doornik and Hendry 2010). Among the main causes of instability, Hendry and Clements (2002) point out the importance of intercept shifts, mainly arising because of shifts in the means of omitted variables. Several ways to cope with forecast failure in regression models have been suggested in the forecasting literature (see e.g. Clements and Hendry, 2006, and the references cited therein). Moreover, among the different remedies proposed, there is some consensus that forecast pooling is one of the most effective, as discussed in Stock and Watson (2004), who provide empirical evidence supporting this view. The intuition behind pooling is that, if the intercept shifts are sufficiently uncorrelated across different regressions, then by averaging forecasts we are also averaging out intercept shifts. Following this intuition, Stock and Watson (2009) argue that a similar logic should also apply to diffusion index models. If factor loading coefficient instability is sufficiently independent across the different series, then the use of a large numbers of series in factor estimation can average out such instability. In this sense, estimated factors can be quite robust to time varying factor loadings. Indeed, Stock and Watson (2002) formally proved that estimated factors are consistent even in the presence of moderate time variation in factor loading coefficients. On the other hand, in the presence of substantial factor loading instability, estimated factors are in general no longer consistent for the "true" unobservable factors. Breitung and Eickmeier (2011) propose tests for the null hypothesis of factor loading coefficient stability. A limitation of their approach is that it requires cross sectional independence among the idiosyncratic shocks. Tests for constancy of the factor loadings, which allow for some spatial correlation, have been recently suggested by Chen, Dolado and Gonzalo (2011), and by Han and Inoue (2011).

In this paper, we go one step further by noting that, even if estimated factors are robust to time varying loading coefficients, there still remains the issue of possible structural instability in the relation between the diffusion indexes and the variable to be forecasted. Namely, instability may also appear in the factor augmented forecasting model used to construct predictions. Hence, the need for a test for the null hypothesis of structural stability.

In related literature, Banerjee, Marcellino and Marsten (2009) provide an extensive Monte Carlo study of the forecast performance of diffusion index models in the presence of structural instability, and find evidence in favor of the use of diffusion indexes for forecasting in unstable environments. Stock and Watson (2009) disentangle instability into three different components, factor loading coefficient instability, factor dynamics instability, and factor model idiosyncratic component induced instability. They suggest using the full sample for factor estimation and instead using susbamples, or time-varying parameter techniques, for estimating regression coefficients in subsequent diffusion index models. The use of recursive and rolling techniques for both factor estimation and factor augmented forecasting model estimation is analyzed in a series of prediction experiments by Kim and Swanson (2011a).

To the best of our knowledge, there is no consistent test for the null hypothesis of factor augmented forecasting model structural stability. The aim of this paper is thus to fill this gap in the literature. For a given loss function, our approach involves testing the equality of expect forecast error loss based on (i) full sample estimation of factors and associated diffusion index type forecasting model; and (ii) analogous expected forecast error loss based on rolling estimation. In this way, we take into account both instability between a set of potential predictors and factors, as well as instability between the variable to be predicted and the factors. The limiting distribution of the suggested statistic is degenerate in the case where parameter estimation error vanishes and both factor loadings and regression coefficients are structurally stable. To circumvent this problem, we use critical values based on the *m out of n (moon)* bootstrap, for which we establish first order asymptotic validity. In particular, use of *moon* bootstrap critical values ensures correct asymptotic size in non degenerate cases and an asymptotic size of zero in the degenerate case. Unitary asymptotic power is ensured in all cases.

It is worth noting that all of our asymptotic results assume only that $\sqrt{T}/N \to 0$, where T is the number of time series observations, and N is the number of variables used to construct factors. In fact, while in financial applications N is generally larger than T, in macroeconomic applications we typically have $N < T$, see e.g. the well known Stock and Watson (2002a,b).

If the null of structural stability is rejected, one can further investigate the cause of forecast model instability. For example, one may be able to disentangle between factor loading coefficient instability and instability of the structural relation between the factors and the target variable being predicted. One would then remain with the issue of selecting the estimation window for either factor loading coefficient estimation or for factor augmented model regression coefficient estimation, or for both, along the lines of Pesaran and Timmermann (2007).

In an empirical illustration we test for the structural stability of factor augmented forecasting models for 11 U.S. macroeconomic variables, including: the unemployment rate, personal income less transfer payments, the 10 year Treasury-bond yield, the consumer price index, the producer price index, non-farm payroll employment, housing starts, industrial production, M2, the S&P 500 index, and gross domestic product, using an extended version of the Stock and Watson macroeconomic dataset first examined in Kim and Swanson (2011a). Our findings suggest that the null of structural stability is rejected for 2 or 3 of 11 variables, depending upon the forecast horizon, when an ex ante prediction period of the last 15 years is specified. This result is shown to be robust across a variety of bootstrap sampling setups, as well as across different loss functions.

The rest of this paper is organized as follows. Section 2 defines the set-up and introduces the test for diffusion index model structural stability. Section 3 establishes the asymptotic properties of the suggested statistic. Section 4 establishes the asymptotic first order validity of *moon* bootstrap critical values in our context. Finally, Section 5 reports the findings of an empirical illustration based on the use of a largescale macroeconomic dataset. All proofs are gathered in an Appendix.

2 Set-Up

We begin by outlining the diffusion index model used in the sequel. Let

$$X_t = \mu_{0,t} + \Lambda_{0,t} F_{0,t} + u_t, \tag{1}$$

where X_t is a $N \times 1$ vector, $\Lambda_{0,t}$ is a $N \times r$ factor loading matrix, $\mu_{0,t}$ is a (possibly time varying) $N \times 1$ intercept vector, $F_{0,t}$ is the unobserved $r \times 1$ factor vector, and u_t is an error term.[1]

Our objective is to predict a scalar target variable, y_{t+h}, where h denotes the forecast horizon. For sake of simplicity, we develop our methodology in the context of predictive models based on only factors. Generalization to factor augmented autoregression models follows straightforwardly. Namely, consider the following forecasting model based on the use of diffusion indexes.

$$\begin{aligned} y_{t+h} &= \alpha_{0,t} + \beta_{0,1,t} F_{0,1,t} + \ldots + \beta_{0,r,t} F_{0,r,t} + \epsilon_{t+h} \\ &= \alpha_{0,t} + F_{0,t}' \beta_{0,t} + \epsilon_{t+h}, \end{aligned} \tag{2}$$

where $\alpha_{0,t}$ is a (possibly time varying) intercept, and ϵ_t is an error term. Needless to say, we can augment the model in (2) with both additional regressors and lagged factors. As such generalizations do not change any of our results, we focus our discussion on this simpler model. For a complete discussion of the usefulness of factor augmented models for forecasting, see e.g. Banerjee, Marcellino and Masten (2010), Dufour and Stevanovic (2011).

[1] Note that (1) implies that $X_{i,t} = \lambda_i F_t + u_{i,t}$.

There are two sources of potential index model structural instability.[2] The first potential source of instability is in the structural relation between the covariates X_t and the factors $F_{0,t}$, and is captured by the loading factor matrix $\Lambda_{0,t}$ and the associated intercept vector, $\mu_{0,t}$. The second source is in the structural relation between the factors and the variable to be predicted, and it is captured by $\alpha_{0,t}$ and $\beta_{0,t}$.

Turning our attention to testing for forecast model stability, our approach involves comparing the expected forecast error of a prediction based on full sample estimation of factors and forecast model regression coefficients and the analogous expected forecast error based on rolling estimation. To this end, we construct predictions of y_{t+h} using factors and parameters estimated in two different ways. Namely, we construct factors using both the full sample, and using a sequence of rolling data windows of length R. Let $P = T - R - h$, be the forecast period for which ex-ante $h-$step ahead predictions are constructed. The forecasting models are specified as follows. First, using full sample estimation, define:

$$\widetilde{y}_{t+h} = \widetilde{\alpha}_{0,T} + \widetilde{\beta}_{1,T}\widetilde{F}_{1,t,T,N} + ... + \widetilde{\beta}_{r,T}\widetilde{F}_{r,t,T,N}, \quad t = R, ..., T-h, \tag{3}$$

where $\widetilde{\alpha}_{0,T} = T^{-1}\sum_{t=1}^{T} y_t$, and where $\widetilde{F}_{t,T,N}$ is the $r \times 1$ factor vector at time t, estimated using the entire sample. Namely,

$$\begin{aligned} &\left(\widetilde{F}_{t,T,N}, \widetilde{\Lambda}_{T,N}\right) \\ &= \arg\min_{\Lambda,F} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\left(X_{i,t} - \frac{1}{T}\sum_{t=1}^{T} X_t\right) - \lambda_i' F_t\right)^2, \end{aligned} \tag{4}$$

and

$$\widetilde{\beta}_T = \left(\sum_{t=1}^{T-h} \widetilde{F}_{t,T,N}\widetilde{F}'_{t,T,N}\right)^{-1} \times \sum_{t=1}^{T-h}\left(\widetilde{F}_{t,T,N}\left(y_{t+h} - \frac{1}{T}\sum_{t=1}^{T} y_t\right)\right). \tag{5}$$

Second, using rolling estimation, define:

$$\widehat{y}_{t+h} = \widehat{\alpha}_{0,t,R} + \widehat{\beta}_{1,t}\widehat{F}_{1,t,t-R+1,N} + ... + \widehat{\beta}_{r,t}\widehat{F}_{r,t,t-R+1,N}, \quad t = R, ..., T-h, \tag{6}$$

where $\widehat{\alpha}_{0,t,R} = R^{-1}\sum_{j=t-R+1}^{t} y_j$, and $\widehat{F}_{t,t-R+1}$ is the $r \times 1$ factor vector at time t, estimated using observations from $t - R$ to t, where R is the rolling window length Namely,

$$\left(\widehat{F}_{j,t-R,N}, \widehat{\Lambda}_{j,t-R,N}\right) \tag{7}$$

[2] In principle one could also explicitly model factor dynamics, and this could be an additional source of instability (see e.g. Stock and Watson (2009)).

$$= \arg\min_{\Lambda,F} \frac{1}{NR} \sum_{i=1}^{N} \sum_{j=t-R+1}^{t} \left(\left(X_{i,j} - \frac{1}{R} \sum_{j=t-R+1}^{t} X_j \right) - \lambda_j' F_j \right)^2 ,$$

for $t = R, ..., R+P-h$, and

$$\widehat{\beta}_t = \left(\sum_{j=t-R+1}^{t} \widehat{F}_{j,t-R,N} \widehat{F}'_{j,t-R,N} \right)^{-1} \quad (8)$$

$$\times \sum_{j=t-R+1}^{t} \left(\widehat{F}_{j,t-R,N} \left(y_{t+h} - \frac{1}{R} \sum_{j=t-R+1}^{t} y_j \right) \right), \; t = R, ..., R+P-h.$$

More precisely, $\widehat{F}_{t,t-R+1,N}$ denotes the last observation on the factor vector, where the factors are constructed using observations from $t-R+1$ to t. Moreover, forecasts are constructed for time periods $R+h$ to T, yielding $P-h$ $h-$step ahead predictions. This means that the factor predictors used in the construction of $\widehat{y}_{t+h}$ are taken from a newly estimated rolling vector of factors, constructed at each point in time. When constructing $\widehat{y}_{t+h}$, parameters are re-estimated at time t, using data available at time t, and for the rolling sample period running from $t-R+1$ to t, prior to the construction of each new forecast. When constructing $\widetilde{y}_{t+h}$, parameters are constructed using the full sample.

Under mild conditions, outlined in Assumption A below, $\widetilde{\alpha}_{0,T}$, $\widehat{\alpha}_{0,t,R}$, $\widetilde{F}_{t,T,N}$, $\widehat{F}_{t,t-R+1,N}$, $\widetilde{\beta}_T$ and $\widehat{\beta}_t$ have a well defined probability limits, as follows:

$$\alpha^{\dagger} = \text{plim}_T \widetilde{\alpha}_{0,T},$$

$$\text{plim}_{T,R\to\infty} \left(\sup_{t\geq R} \left(\widehat{\alpha}_{0,t,R} - \alpha_t^{\ddagger} \right) \right) = 0,$$

$$Q^{\dagger}\beta^{\dagger} = \text{plim}_{N,T\to\infty} \widetilde{\beta}_T,$$

$$\text{plim}_{T,R,N\to\infty} \left(\sup_{t\geq R} \left(\widehat{\beta}_t - Q_t^{\ddagger}\beta_t^{\ddagger} \right) \right) = 0,$$

$$\text{plim}_{T,N\to\infty} \left(\widetilde{F}_{t,T,N} - H^{\dagger\prime} F_t^{\dagger} \right) = 0, \text{ for all } t,$$

and

$$\text{plim}_{T,R,N\to\infty} \left(\widehat{F}_{j,t-R+1,N} - H_t^{\ddagger\prime} F_j^{\ddagger} \right) = 0, \text{ for all } j = t-R+1, ..., t, \; t \geq R,$$

where $Q^{\dagger}, Q_t^{\ddagger}, H^{\dagger}, H_t^{\ddagger}$ are defined in the Appendix, in the proof of Theorem 1. Further, note that, as shown in the Appendix, $H^{\dagger\prime}Q^{\dagger} = I_r$ and $H_t^{\ddagger\prime}Q_t^{\ddagger} = I_r$.

The estimated parameters in (6) change over time, not only because rolling windows of data are used to construct the parameters estimators, but also because the factor vectors are estimated using a different rolling subset of the original data, at each point in time. This is opposed to the estimated parameters in (3), which are based on parameter and factor estimates calculated using the full sample.

Note that in the case of structural factor stability, $\alpha_t^{\ddagger} = \alpha^{\dagger} = \alpha_0$, $H_t^{\ddagger\prime} F_t^{\ddagger} = H^{\dagger\prime} F_t^{\dagger} = H_0' F_{0,t}$ and $Q_t^{\ddagger} \beta_t^{\ddagger} = Q^{\dagger} \beta^{\dagger} = Q_0 \beta_0$, for all t. By comparing (4) and (7), it is immediate to see that $H_t^{\ddagger\prime} F_t^{\ddagger} = H^{\dagger\prime} F_t^{\dagger} = H_0' F_{0,t}$ holds only if $\mu_{0,t} = \mu_0$ for all t. In this sense, shifts in the intercept terms are detected as causes of structural instability.

Thus far, we have remained silent on how to choose the number of factors, r. In principle, one can use either the full sample or rolling samples for the implementation of the information criteria suggested by Bai and Ng (2002), in order to estimate r. In our empirical illustration, we use the full sample to determine r. In the case of factor loading instability, this may lead to a possible overestimate of r, as documented by Breitung and Eickmeier (2011). However, Han and Inoue (2011) show that the Bai and Ng IC criterion works properly in the case of a single break. For additional discussion of testing for the number of factors, see Onatski (2009).

We now outline our test for forecast model stability. Let,

$$\epsilon_{1,t+h} = y_{t+h} - \alpha^{\dagger} - \sum_{i=1}^{r} \beta_i^{\dagger} F_{i,t}^{\dagger}$$

and

$$\epsilon_{2,t+h} = y_{t+h} - \alpha_t^{\ddagger} - \sum_{i=1}^{r} \beta_{i,t}^{\ddagger} F_{i,t}^{\ddagger}.$$

We test the following hypotheses:

$$H_0 : \mathrm{E}\left(g\left(\epsilon_{1,t+h}\right) - g\left(\epsilon_{2,t+h}\right)\right) = 0, \text{ for all } t \geq R \tag{9}$$

versus

$$H_A : \mathrm{E}\left(g\left(\epsilon_{1,t+h}\right) - g\left(\epsilon_{2,t+h}\right)\right) \neq 0, \text{ for all } t \in \mathcal{T},\ t \geq R,\ \mathcal{T}/P \to \tau \neq 0, \tag{10}$$

where g is a given loss function. Under the null hypothesis, the expected prediction loss from a model allowing for possible time variation in the loadings and in the β_s, and one allowing no variation, is the same. It is immediate to see that when $\alpha^{\dagger} = \alpha_t^{\ddagger}$ and $\beta_i^{\dagger} F_{i,t}^{\dagger} = \beta_{i,t}^{\ddagger} F_{i,t}^{\ddagger}$, a.s. for all t, i, then $\epsilon_{1,t+h} = \epsilon_{2,t+h}$, a.s. for all t, i. This is exactly the same situation arising in the context of forecast evaluation, when one compares the predictive accuracy of two or more nested models (see e.g. Diebold and Mariano (1995), White (2000), and Corradi and Swanson (2007)).
In order to test the hypotheses in (9) and (10), we thus suggest using the following statistic:[3]

[3]For notational simplicity, herafter we omit the subscript N.

$$S_P = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-h} \left(g\left(\widetilde{\epsilon}_{t+h}\right) - g\left(\widehat{\epsilon}_{t+h}\right)\right), \tag{11}$$

where $\widetilde{\varepsilon}_{t+h} = y_{t+h} - \widetilde{\alpha}_T - \widetilde{\beta}_T' \widetilde{F}_{t,T}$ and $\widehat{\varepsilon}_{t+h} = y_{t+h} - \widehat{\alpha}_{t,R} - \widehat{\beta}_{t,R}' \widehat{F}_{t,t-R+1}$.

Giacomini and Rossi (2009) have recently suggested tests for forecast failure in standard regression contexts. Their notion of forecast failure is that expected out of sample loss is larger than expected in sample loss. In this sense, our test is not a test for predictive failure, but is instead a test for structural instability. In fact, the statistic in (11), for given loss function, compares average prediction errors for the out of sample period only. Our test can detect various forms of predictive failure, however, such as shifts in the mean of omitted variables. This because we are using different approaches to recentering in (4)-(5) and in (7)-(8). On the other hand, our test may not be able to detect shifts in the slope of omitted variables.

3 Asymptotics

In order to derive the limiting distribution of S_P, we require the following Assumption. Hereafter, for a matrix B, $\|B\| = (tr\,(B'B))^{1/2}$, and C denotes a a generic constant

Assumption A:

A1: (**i**) For $i = 1, ..., N$, $\left(F_t^\dagger, u_{it}^\dagger\right)$ and $\left(F_t^\ddagger, u_{it}^\ddagger\right)$ are $\alpha-$mixing with size $-4\,(4+\psi)\,/\psi$, $\psi > 0$. (**ii**) for $i = 1, ..., N$ and $j = 1, ..., r$, $\sup_t \mathrm{E}\left(\left|F_{jt}^\dagger\right|^{2k}\right) \leq C$, $\sup_t \mathrm{E}\left(\left|F_{jt}^\ddagger\right|^{2k}\right) \leq C$, $\sup_t \mathrm{E}\left(\left|u_{it}^\dagger\right|^{2k}\right) \leq C$, and $\sup_t \mathrm{E}\left(\left|u_{it}^\ddagger\right|^{2k}\right) \leq C$, with $k > 2\,(2+\psi)$. (**iii**) For $i = 1, ..., N$, $j = 1, ..., r$, $\sup_{i,j,t} \left|\lambda_{ij,t}^\dagger\right| \leq C$ and $\sup_{i,j,t} \left|\lambda_{ij,t}^\ddagger\right| \leq C$. (**iv**) $\mathrm{E}\left(F_t^\dagger u_{it}^\dagger\right) = \mathrm{E}\left(F_t^\ddagger u_{it}^\ddagger\right) = 0$.

A2: Let $\sigma_{ij,ts}^\dagger = \mathrm{E}\left(u_{it}^\dagger u_{js}^\dagger\right)$, $\sigma_{ij,ts}^\ddagger = \mathrm{E}\left(u_{it}^\ddagger u_{js}^\ddagger\right)$, $\sup_{t,s}\left|\sigma_{ij,ts}^\dagger\right| = \tau_{ij}^\dagger$, and $\sup_{t,s}\left|\sigma_{ij,ts}^\ddagger\right| = \tau_{ij}^\ddagger$. (**i**) $\frac{1}{N}\sum_{i=1}^N \sum_{j=1}^N \tau_{ij}^\dagger \leq C$ and $\frac{1}{N}\sum_{i=1}^N \sum_{j=1}^N \tau_{ij}^\ddagger \leq C$. (**ii**) $\sup_{t,s} \mathrm{E}\left(N^{-1/2}\sum_{i=1}^N \left|u_{it}^\dagger u_{is}^\dagger - \mathrm{E}\left(u_{it}^\dagger u_{is}^\dagger\right)\right|^4\right) \leq C$; and the same holds with $u_{it}^\dagger u_{is}^\dagger$ replaced by $u_{it}^\ddagger u_{is}^\ddagger$. (**iii**) For all t, $\frac{1}{\sqrt{N}}\sum_{i=1}^N \lambda_{it}^\dagger u_{it}^\dagger$ and $\frac{1}{\sqrt{N}}\sum_{i=1}^N \lambda_{it}^\ddagger u_{it}^\ddagger$ satisfy a central limit theorem.

A3: (**i**) y_t is $\alpha-$mixing with size $-4\,(4+\psi)\,/\psi$, $\psi > 0$. (**ii**) $\sup_t \mathrm{E}\left(|y_t|^{2k}\right) \leq C$, with $k > 2\,(2+\psi)$ and $\mathrm{E}\left(F_t^\dagger \epsilon_{1t}\right) = \mathrm{E}\left(F_t^\ddagger \epsilon_{2t}\right) = 0$. (**iii**) For $j = 1, ..., r$, $\iota = 1, 2$, $\sup_t \mathrm{E}\left(\left|\nabla g_{F_j}\left(\epsilon_{\iota t}\right)\right|^{2k}\right) \leq C$, with $k > 2\,(2+\psi)$, where ∇g_{F_j} denotes the derivative of g with respect to factor j. (**iv**) For $i = 1, ..., N$, $\mathrm{E}\left(\nabla g_{F_j}\left(\epsilon_{1t}\right) u_{it}^\dagger\right) = \mathrm{E}\left(\nabla g_{F_j}\left(\epsilon_{2t}\right) u_{it}^\ddagger\right) = 0$.

Assumption **A1(i)** requires that the probability limits of the factors estimated either via the use of the full sample or via rolling windows, are strongly mixing; and likewise, for the idiosyncratic errors. Note that if $X_{i,t}$ is strong mixing for all i, and **A1(iii)** holds, then $F_t^{\dagger}$ and $F_t^{\ddagger}$ are strong mixing by construction. It is worthwhile to also notice that most of the papers on the estimation of the diffusion index models and factor augmented regressions (see e.g. Stock and Watson (2002a,b) and Bai and Ng (2006)) do not impose direct assumptions on the memory of the factors. On the other hand, they assume that the error term in the factor augmented regression is a martingale difference sequence. We do not require that either $\epsilon_{1,t+h}$ or $\epsilon_{2,t+h}$, are martingale difference sequences, as we want to allow for possible dynamic misspecification.[4] Assumptions **A1(ii)** and **A1(iv)** are rather standard assumptions, and shall thus not be discussed here. Assumption **A2** controls the degree of cross correlation among the idiosyncratic errors. The degree of time serial correlation among idiosyncratic errors is already controlled by the strong mixing assumption. Assumption **A3** provides primitive sufficient conditions under which Bai (2003) and Bai and Ng (2006) central limit theorems apply to averages of rolling estimators, based on dependent and heterogeneous series (see e.g. Corradi and Swanson (2006a)).

Theorem 1: *Let Assumption A hold. Also, as* $N, T, P, R \to \infty$, $N/\sqrt{T} \to \infty$, *and* $P/R \to \pi$, $0 < \pi < \infty$.

Then, under H_0, *we distinguish four cases.*

Case I: $\beta^{\dagger\prime} F_t^{\dagger} \neq \beta_t^{\ddagger\prime} F_t^{\ddagger}$, *for all* $t \in \mathcal{T}$, $\mathcal{T}/T \to \tau \neq 0$, $D_\delta^{\dagger} = \mathrm{E}(\nabla g_\delta(\epsilon_{1,t+h})) \neq 0$, $D_\delta^{\ddagger} = \mathrm{E}(\nabla g_\delta(\epsilon_{2,t+h})) \neq 0$, *and* $\delta = (\alpha, \beta)$. *Then, regardless of whether* $\alpha^{\dagger} = \alpha_t^{\ddagger}$,

$$S_p \xrightarrow{d} N(0, \Omega_1),$$

with

$$\begin{aligned}
\Omega_1 = & \lim_{P\to\infty} \mathrm{var}\left(\frac{1}{\sqrt{P}} \sum_{t=R}^{T-h} \left(g(\epsilon_{1,t+h}) - g(\epsilon_{2,t+h})\right)\right) \\
& + D_\delta^{\dagger\prime} \lim_{P\to\infty} \mathrm{var}\left(\sqrt{P}\left(\widetilde{\delta}_T - Q_\delta^{\dagger}\delta^{\dagger}\right)\right) D_\delta^{\dagger} + D_\delta^{\ddagger\prime} \lim_{P\to\infty} \mathrm{var}\left(\frac{1}{\sqrt{P}} \sum_{t=R}^{T-h} \left(\widehat{\delta}_t - Q_{\delta,t}^{\ddagger}\delta_t^{\ddagger}\right)\right) D_\delta^{\ddagger} \\
& - 2D_\delta^{\dagger\prime} \lim_{P\to\infty} \mathrm{cov}\left(\sqrt{P}\left(\widetilde{\delta}_T - Q_\delta^{\dagger}\eta^{\dagger}\right), \frac{1}{\sqrt{P}} \sum_{t=R}^{T-h} \left(g(\epsilon_{1,t+h}) - g(\epsilon_{2,t+h})\right)\right) \\
& + 2D_\delta^{\ddagger\prime} \lim_{P\to\infty} \mathrm{cov}\left(\frac{1}{\sqrt{P}} \sum_{t=R}^{T-h} \left(\widehat{\delta}_t - Q_{\delta,t}^{\ddagger}\delta_t^{\ddagger}\right), \frac{1}{\sqrt{P}} \sum_{t=R}^{T-h} \left(g(\epsilon_{1,t+h}) - g(\epsilon_{2,t+h})\right)\right) \\
& - 2D_\delta^{\ddagger\prime} \lim_{P\to\infty} \mathrm{cov}\left(\frac{1}{\sqrt{P}} \sum_{t=R}^{T-h} \left(\widehat{\delta}_t - Q_{\delta,t}^{\ddagger}\delta_t^{\ddagger}\right), \sqrt{P}\left(\widetilde{\delta}_T - Q_\delta^{\dagger}\delta^{\dagger}\right)\right) D_\delta^{\dagger\prime},
\end{aligned}$$

[4] A mixing assumption on factors and idiosyncratic errors is instead used by Fan, Liao and Mincheva (2010).

where $\widetilde{\delta}_T = \left(\widetilde{\alpha}_T, \widetilde{\beta}_T\right)'$, $\widehat{\delta}_t = \left(\widehat{\alpha}_t, \widehat{\beta}_t\right)'$, $Q_\delta^\dagger \delta^\dagger = \left(Q^\dagger \beta^\dagger, \alpha^\dagger\right)'$, *and* $Q_{\delta,t}^\ddagger \delta_t^\ddagger = \left(Q_t^\ddagger \beta_t^\ddagger, \alpha_t^\ddagger\right)$.

Case II: $\beta^{\dagger\prime} F_t^\dagger \neq \beta_t^{\ddagger\prime} F_t^\ddagger$ *for all* $t \in \mathcal{T}$, $\mathcal{T}/T \to \tau \neq 0$, $D_\delta^\dagger = \mathrm{E}\left(\nabla g_\delta\left(\epsilon_{1,t+h}\right)\right) = D_\delta^\ddagger = \mathrm{E}\left(\nabla g_\delta\left(\epsilon_{2,t+h}\right)\right) = 0$. *Then, regardless of whether* $\alpha^\dagger = \alpha_t^\ddagger$ *or not,*

$$S_p \xrightarrow{d} N\left(0, \Omega_2\right),$$

where

$$\Omega_2 = \lim_{P \to \infty} \operatorname{var}\left(\frac{1}{\sqrt{P}} \sum_{t=R}^{T-h} \left(g\left(\epsilon_{1,t+h}\right) - g\left(\epsilon_{2,t+h}\right)\right)\right)$$

Case III: $\alpha^\dagger = \alpha_t^\ddagger$, $\beta^{\dagger\prime} F_t^\dagger = \beta_t^{\ddagger\prime} F_t^\ddagger$ *for all* $t \in T$, $D_\delta^\dagger = \mathrm{E}(\nabla g_\delta\left(\epsilon_{1,t+h}\right)) \neq 0$ $D_\delta^\ddagger = \mathrm{E}(\nabla g_\delta\left(\epsilon_{1,t+h}\right)) \neq 0$. *Then*[5]*,*

$$S_p \xrightarrow{d} N\left(0, \Omega_3\right),$$

where

$$\begin{aligned}
\Omega_3 = {} & D_\delta^{\dagger\prime} \lim_{P \to \infty} \operatorname{var}\left(\sqrt{P}\left(\widetilde{\delta}_T - Q_\delta^\dagger \delta^\dagger\right)\right) D_\delta^\dagger + D_\delta^{\ddagger\prime} \lim_{P \to \infty} \operatorname{var}\left(\frac{1}{\sqrt{P}} \sum_{t=R}^{T-h} \left(\widehat{\delta}_t - Q_{\delta,t}^\ddagger \delta_t^\ddagger\right)\right) D_\delta^\ddagger \\
& - 2 D_\delta^{\ddagger\prime} \lim_{P \to \infty} \operatorname{cov}\left(\frac{1}{\sqrt{P}} \sum_{t=R}^{T-h} \left(\widehat{\delta}_t - Q_{\delta,t}^\ddagger \delta_t^\ddagger\right), \sqrt{P}\left(\widetilde{\delta}_T - Q_\delta^\dagger \delta^\dagger\right)\right) D_\delta^{\dagger\prime},
\end{aligned}$$

Case IV: $\alpha^\dagger = \alpha_t^\ddagger$, $\beta^{\dagger\prime} F_t^\dagger = \beta_t^{\ddagger\prime} F_t^\ddagger$ *for all* $t \in T$, *and* $D_\delta^\dagger = \mathrm{E}\left(\nabla g_\delta\left(\epsilon_{1,t+h}\right)\right) = D_\delta^\ddagger = \mathrm{E}\left(\nabla g_\delta\left(\epsilon_{2,t+h}\right)\right) = 0$. *Then,*

$$S_p = O_p\left(\frac{1}{\sqrt{P}}\right) + O_p\left(\max\left\{\frac{\sqrt{P}}{N}, \frac{\sqrt{P}}{R}\right\}\right).$$

Under H_A, *there exists* $\varepsilon > 0$, *such that*

$$\lim_{P \to \infty} \Pr\left(P^{-1/2}\left|S_p\right| \geq \varepsilon\right) = 1.$$

Note that in all cases, the estimation error due to factor estimation vanishes. This is due to the assumption that $N/\sqrt{T} \to \infty$, as shown by Bai and Ng (2006).

Also, as $\widetilde{\alpha}_T$, $\widehat{\alpha}_{t,R}$, $\widetilde{\beta}_T$ and $\widehat{\beta}_{t,R}$ are OLS estimators, the cases where $D_\delta^\dagger = D_\delta^\ddagger = 0$ are the only relevant ones when g is a quadratic loss function (i.e., Cases II and IV). This is because, in these cases the same loss is used for both estimation and prediction, and hence the contribution of parameter estimation error is negligible.

Finally, when we have structural stability (i.e., $\alpha^\dagger = \alpha_t^\ddagger$ and $\beta_i^\dagger F_{i,t}^\dagger = \beta_{i,t}^\ddagger F_{i,t}^\ddagger$, for all $t \in T$), and we specify a quadratic loss function, the statistic is degenerate. Also, note that the case of

[5] If $\alpha^\dagger \neq \alpha_t^\ddagger$ for all $t \in \mathcal{T}$, $\mathcal{T}/T \to \tau \neq 0$, then $\Omega_2 \neq 0$, and so the statement in cases III and IV should be modified accordingly.

$\pi = 0$ is equivalent to the case of $D_\delta^\dagger = D_\delta^\ddagger = 0$, as in both situations the contribution of parameter estimation error becomes negligible.

In Cases I-III, the limiting covariance matrix has closed form, which is given in the Appendix, and can be estimated. Moreover, and as discussed above, whenever g is a quadratic loss function, there is a possibility that the statistic approaches zero in probability. If g is quadratic and the null is true, then we are either in Case II or in Case IV. Suppose, we estimate the asymptotic variance as in Case II, but we are instead in Case IV. The estimator of the standard deviation approaches zero at rate $\frac{1}{\sqrt{P}} + \max\left\{\frac{\sqrt{P}}{N}, \frac{\sqrt{P}}{R}\right\}$, but S_P also approaches zero at rate $\frac{1}{\sqrt{P}} + \max\left\{\frac{\sqrt{P}}{N}, \frac{\sqrt{P}}{R}\right\}$. As a direct consequence, the statistic, scaled by its standard error, is bounded but does not have a well defined limiting distribution. Hence, we may not able to distinguish between Case IV and the alternative. Moreover, even though Case IV resembles the case of forecast comparison of nested models, because of the interplay between N and P, the techniques used to deal with Diebold-Mariano tests in this context (e.g. using the methods of McCracken (2007)), are not immediately available. Further, while one knows whether two models are nested or not, here we do not know whether $\alpha^\dagger = \alpha_t^\ddagger$ and $\beta^{\dagger\prime} F_t^\dagger = \beta_t^{\ddagger\prime} F_t^\ddagger$, for all $t \in T$. The next section establishes the validity of *moon* bootstrap critical values in the current context.

4 Moon Bootstrap Critical Values

In order to circumvent the problem of the degeneracy of the statistic in Case IV, we rely on the m *out of* n (*moon*) bootstrap. The idea underlying the *moon* bootstrap is to resample m observations out of a sample of n observations, with $m/n \to 0$. The key difference between the *moon* bootstrap and subsampling is that in the former we resample with replacement, while in the latter we resample without replacement (see e.g. Bickel, Götze and van Zwet, (1997)). One of the advantages of the m *out of* n bootstrap over the subsampling is that m can be chosen in a data-driven manner (see e.g. Bickel and Sakov, (2008)).

In the sequel, we show that *moon* bootstrap critical values are asymptotically valid for all cases in Theorem 1.

Let $T^* = P^* + R^*$, where $T^*/T \to 0$, and $P^*/R^* \to \pi$ (i.e. $(P^*/R^* - P/R) \to 0$). We require three layers of resampling.

(i) Resample for the construction of full sample estimators:

Resample b_{T^*} blocks of length l_{T^*}, $b_{T^*} \times l_{T^*} = T^* - h$ from $\left(y_t, \widetilde{F}_{t-h,T}\right)$, $t > h$, to obtain

$\left(y_t^*, \widetilde{F}_{t-h,T}^*\right)$ and construct the bootstrap analog of $\widetilde{\beta}_T$ as

$$\widetilde{\beta}_{T^*}^* = \left(\sum_{t=1}^{T^*-h} \widetilde{F}_{t,T}^* \widetilde{F}_{t,T}^{\prime *}\right)^{-1} \sum_{t=1}^{T-h} \widetilde{F}_{t,T}^* \left(y_{t+h}^* - \frac{1}{T^*}\sum_{t=1}^{T^*} y_t^*\right).$$

The bootstrap analog of $\widetilde{\alpha}_T$ is constructed in like fashion.

(ii) Resample for the construction of rolling estimators:

Let $\left(\widehat{F}_{1,1}, \widehat{F}_{2,1}, ..., \widehat{F}_{R,1}\right)$ be the factor estimates obtained using a window of data, $X_1, ..., X_R$ (i.e., using the first R observations). Further, let $\left(\widehat{F}_{k,k}, \widehat{F}_{k+1,k}, ..., \widehat{F}_{k+R-1,k}\right)$ be the factor estimates obtained using a window of data, $X_k, ..., X_{k+R-1}$ (i.e. using observations from $t = k$ to $t = k+R-1$), and so on.

Resample b_{R^*} blocks of length l_{R^*}, $b_{R^*} \times l_{R^*} = R^* - h$ from $\left(y_{1+h}, ..., y_R, \widehat{F}_{1,1}, \widehat{F}_{2,1}, ..., \widehat{F}_{R-h,1}\right)$ to obtain $\left(y_{1+h}^*, ..., y_{1,R^*}^*, \widehat{F}_{1,1}^*, ..., \widehat{F}_{R^*-h,1}^*\right)$, and construct

$$\widehat{\beta}_{R^*-h}^* = \left(\sum_{j=1}^{R^*-h} \widehat{F}_{j,1}^* \widehat{F}_{j,1}^{\prime *}\right)^{-1} \sum_{j=1}^{R^*-h} \widehat{F}_{j,1}^* \left(y_{j+1+h}^* - \frac{1}{R}\sum_{j=1}^{R^*-h} y_{j+h}^*\right).$$

The bootstrap analog of $\widehat{\alpha}_t$ is constructed in like fashion.

Analogously, resample b_{R^*} blocks of length l_{R^*}, $b_{R^*} \times l_{R^*} = R^* - h$ from $\left(y_{k+1+h}, ..., y_{R+k}, \widehat{F}_{k+1,k+1}, ..., \widehat{F}_{R+k,k+1}\right)$ to obtain $\left(y_{k+1+h}^*, ..., y_{R+k}^*, \widehat{F}_{k+1,k+1}^*, ..., \widehat{F}_{R^*+k,k+1}^*\right)$, and construct

$$\widehat{\beta}_{R^*+k-h}^* = \left(\sum_{j=1+k}^{R^*+k-h} \widehat{F}_{j,k+1}^* \widehat{F}_{j,k+1}^{*\prime}\right)^{-1} \sum_{j=1+k}^{R^*+k-h} \widehat{F}_{j,k+1}^* \left(y_{j+h}^* - \frac{1}{R}\sum_{j=1+k}^{R^*+k-h} y_{j+h}^*\right),$$

and so on, obtaining $\widehat{\beta}_{R^*-h}^*, ..., \widehat{\beta}_{R^*+1}^*, ..., \widehat{\beta}_{R^*+P^*-h}^*$, and analogous estimators, $\widehat{\alpha}_{R^*}^*, \widehat{\alpha}_{R^*+1}^*, ..., \widehat{\alpha}_{R^*+P^*-h}^*$.

(iii) Resample for the construction of the statistics:

Resample b_{P^*} blocks of length l_{P^*}, $b_{P^*} \times l_{P^*} = P^* - h$ from $\left(y_{t+h}, \widetilde{F}_{t,T}, \widehat{F}_{t,t-R+1}\right)$, $t = R, ..., T-h$, to obtain $\left(y_{t+h}^*, \widetilde{F}_{t,T}^*, \widehat{F}_{t,t-R+1}^*\right)$.

We can now define the bootstrap statistic,

$$S_{P^*}^* = \frac{1}{\sqrt{P^*}} \sum_{t=R^*}^{T^*-h} \left(g\left(\widetilde{\epsilon}_{t+h}^*\right) - g\left(\widehat{\epsilon}_{t+h}^*\right)\right),$$

where

$$\widetilde{\epsilon}_{t+h}^* = y_{t+h}^* - \widetilde{\alpha}_{T^*}^* - \widetilde{\beta}_{T^*}^{*\prime} \widetilde{F}_{t,T}^*$$

and

$$\widehat{\epsilon}_{t+h}^* = y_{t+h}^* - \widehat{\alpha}_t^* - \widehat{\beta}_t^{*\prime} \widehat{F}_{t,t-R+1}^*.$$

Note that in order to capture the contribution of parameter estimation error, we resample the estimated factors using a different resampling scheme for full and rolling samples. Nevertheless, we do not resample the N variables X_i (i.e., we do not construct factor estimators based on resampled observations). This is because, as we assume $\sqrt{T}/N \to 0$, the contribution of factor estimation error is asymptotically negligible. However, factor estimation error may matter in finite samples. For this reason, Goncalves and Perron (2010) suggest a residual-based approach which properly mimics the contribution of factor estimation error. Their objective is to provide valid bootstrap standard errors for estimated regression coefficients involving estimated factors.

Importantly, if g is a quadratic function, then we know that the contribution of parameter estimation error is asymptotically negligible. In this case, it suffices to perform only step (iii) (i.e., only perform resampling in order to construct the statistic using forecast models parameters estimated as described in the previous section).

As stated above and in Theorem 1, whenever g is a quadratic loss function and $\alpha^{\dagger} = \alpha_t^{\ddagger}$, $\beta^{\dagger\prime} F_t^{\dagger} = \beta_t^{\ddagger\prime} F_t^{\ddagger}$ for all $t \in T$, then the statistic approaches zero in probability. This occurs because we have assumed that $\sqrt{T}/N \to 0$. If, for example, $\sqrt{T}/N \to c \neq 0$, then the contribution of the estimated factors to the asymptotic covariance matrix will never vanish, and the statistic will never be degenerate on zero. Thus, we would not need to use the *m out of n* bootstrap, but could rely on "usual" block bootstrap. However, in this case we would also need to resample the $X_{i,t}$, in order to capture factor estimation error. The difficulty would then center around how to resample in order to capture cross correlation among the $X_{i,t}$. In this context, whether the residual-based bootstrap of Goncalves and Perron (2010) can be extended to the rolling estimation scheme, and thus applied in our framework when $\sqrt{T}/N \nrightarrow 0$, in the presence of cross correlation among idiosyncratic errors, is left to future research.

We now establish the first order validity of the *moon* bootstrap procedure outlined above.

Theorem 2: *Let Assumption A hold. Also, as $N, T, P, R \to \infty$, $N/\sqrt{T} \to \infty$, and $P/R \to \pi$, $0 < \pi < \infty$. Additionally, as $P^*, R^*, T^* \to \infty$, assume that $T^*/T \to 0$, $P^*/P \to 0$, $R^*/R \to 0$, $P^*/R^* \to \pi$. Finally, as $l_{P^*}, l_{R^*}, l_{T^*} \to \infty$, $l_{T^*}/\sqrt{T^*} \to 0, l_{P^*}/\sqrt{P^*} \to 0$, and $l_{R^*}/\sqrt{R^*} \to 0$.*
Then, under H_0, in Cases I-III,

$$P\left(\omega : \sup_{v \in R} \left| \overset{*}{\Pr}\left(S_{P^*}^* \leq v\right) - \Pr\left(S_p \leq v\right) \right| < \varepsilon \right) \to 0,$$

and in Case IV, provided that $P^* = o\left(N^2/P\right)$,

$$\frac{S_P}{S_{P^*}^*} = o_{p^*}\left(1\right),$$

conditional on the sample, and for all samples except a subset with probability measure approaching zero.

Under H_A, $S^*_{P^*}$ diverges at rate $\sqrt{P^*}$.

Let $c^*_{(1-\alpha)}$ be the $(1-\alpha)-$percentile of the empirical distribution of $S^*_{P^*}$. If we do not reject H_0, whenever $S_P \leq c^*_{(1-\alpha)}$ and we reject otherwise, we have a test with asymptotic size equal to α in Cases I-III, and asymptotic size equal to zero in Case IV. Asymptotic power is equal to 1. Indeed, from the theorem, we see that under H_0, in Cases I-III, S_P and $S^*_{P^*}$ have the same limiting distribution, conditional on sample, while in Case IV, provided that $P^* = o\left(N^2/P\right)$, $S^*_{P^*}$ approaches zero at a slower rate than S_P. Finally, under the alternative, S_P diverges at a faster rate than $S^*_{P^*}$, thus ensuring unit asymptotic power. The reason why we construct a one-sided test is that in the degenerate case, S_P goes to zero at a faster rate than S_{P^*}, but it may occur that S_P has negative sign, while $S^*_{P^*}$ has a positive sign.

5 Empirical Illustration

We illustrate the implementation of the proposed test statistic, S_P by constructing predictions of the same 11 macroeconomic variables examined in Armah and Swanson (2010), as summarized in Table 1. Prediction models are constructed according to the generic specification given in equations (1) and (2). Namely, we implement forecasting models of the form: $y_{t+h} = \alpha_{0,t} + \beta_{0,1,t}F_{0,1,t} + ... + \beta_{0,r,t}F_{0,r,t} + \epsilon_{t+h}$, as defined in Section 2 above, and factors are estimated as in (4) and (7). The number of factors r is selected using the approach of Bai and Ng (2002). Factor are based on macroeconomic dataset first introduced of Stock and Watson (2002a,b), and extended by and Kim and Swanson (2011a,b). We have 155 monthly variables for the period 1960:1 - 2009:5, so that $N = 155$ and $T = 560$. As outlined in Section, 2 and 4, values for T, R, P, T^*, R^*, P^*, m, b_{T^*}, b_{R^*}, and b_{P^*} are required for implementation of the test. Various values for these parameters were tried, as outlined in Table 2.

A selected subset of our empirical findings are collected in Tables 3-6, based on predictions constructed 1- and 3-months ahead. Results across all other parameter permutations were qualitatively similar, and are available upon request from the authors. In the tables, entries are given for (i) the test statistic; (ii) the 95th, 90th, and 50th percentiles of the empirical bootstrap distribution, for given values of b_{T^*}, b_{R^*}; and b_{P^*}; and (iii) the probability of rejection (p-value) under the null, based on the empirical bootstrap distribution. Tables 3-4 correspond to experiments run using a quadratic loss function, while Tables 5-6 repeat the same set of experiments, but using the following linex loss function: $g(u) = e^{au} - au - 1$, with $a = 1$. The reason why we compare outcomes based on quadratic and linex loss functions is that we want to control for the possible conservativeness of the *moon* critical values. In the case a quadratic loss function parameter estimation error may vanish under the null, while in linex loss it cannot vanish (at least given our choices of P, R, P^*, R^*). Broadly speaking, we want to see whether the failure to reject the null is consistent across the two different loss function. If this is the case, we do not have to be worried about the possible conservativeness of *moon* critical values.

A number of conclusions emerge from examination of the results in the tables. First, comparing results in any individual table, we see that inference is robust across different values of b_{T^*}, b_{R^*}; and b_{P^*}. Second, when comparing results across tables (e.g. compare Tables 3 and 4), empirical findings are somewhat dependent upon forecast horizon. Namely, at a 10% level, the null of stability rejected for TB10Y and PPI, when $h = 1$. On the other hand, when $h = 3$, the null hypothesis is rejected for CPI, PPI, and HS. One of the reasons for this finding may be that, in our simple empirical illustration, we include only factors as regressors, and do not include lags. In order to properly explore the stability of our variables, it is evident that a much more exhaustive and detailed empirical analysis is needed. Finally, notice that our inference is robust to choice of loss function. Namely, the variables for which the null is rejected does not change if a linex loss function rather than a quadratic loss function is specified.

6 Concluding Remarks

We have developed a simple to implement test for the structural stability of factor augmented forecasting models. Our null hypothesis involves jointly testing stability of factor loading and forecast model coefficients via examination on prediction errors. Implementation of the test involves use of a Diebold-Mariano (1995) type test where sequences of forecast errors are constructed using both full sample and rolling estimation schemes. Asymptotically valid critical values are constructed using the *m out of n* (*moon*) bootstrap. In an empirical illustration, we show that the test is convenient to implement, and offers inference that is robust across various parameters of interest, such as moon bootstraps sample size, block lengths and ex-ante prediction periods, and loss function choice.

7 Appendix

Hereafter, for the sake of simplicity and clarity, we provide all proofs for the case where $\mu_{0,t} = \alpha_{0,t} = 0$ (i.e. assuming that both X_t and y_t have zero mean). As a consequence, we do not estimate α_s, and we estimate factors and β_s without recentering.

Proof of Theorem 1: We begin by considering the full sample estimation scheme. Now,

$$\begin{aligned}\widetilde{\epsilon}_{1,t+h} &= \epsilon_{1,t+h} - \left(\widetilde{F}_t - H'_{N,T}F_t^{\dagger}\right)' Q^{\dagger}\beta^{\dagger} + F_t^{\dagger\prime} H_{N,T}\left(\widetilde{\beta}_T - Q^{\dagger}\beta^{\dagger}\right) \\ &\quad + \left(\widetilde{F}_t - H'_{N,T}F_t^{\dagger}\right)'\left(\widetilde{\beta}_T - Q^{\dagger}\beta^{\dagger}\right),\end{aligned}$$

where $Q^{\dagger} = V^{\dagger 1/2}\Upsilon^{\dagger\prime}\Sigma_{\lambda}^{\dagger -1/2}$, with $\Sigma_{\lambda}^{\dagger} = p\lim_{N\to\infty}\frac{\Lambda^{\dagger\prime}\Lambda^{\dagger}}{N}$, $V^{\dagger} = \operatorname{diag}\left(v_1^{\dagger},...,v_r^{\dagger}\right)$, where $v_1^{\dagger} > v_2^{\dagger} > ... > v_r^{\dagger} > 0$ are the eigenvalues of $\Sigma_{\lambda}^{\dagger -1/2}\Sigma_F^{\dagger}\Sigma_{\lambda}^{\dagger -1/2}$, $\Sigma_F^{\dagger} = p\lim_{T\to\infty}\frac{F'F^{\dagger}}{T}$. Also, $\Upsilon^{\dagger}$ is the matrix of the eigenvectors associated with $\left(v_1^{\dagger},...,v_r^{\dagger}\right)$, such that $\Upsilon^{\dagger\prime}\Upsilon^{\dagger} = I_r$, $H_{N,T}^{\dagger} = \frac{\Lambda^{\dagger\prime}\Lambda^{\dagger}}{N}\frac{F^{\dagger\prime}F}{T}V_{N,T}^{-1}$, with $V_{N,T}$ an $r\times r$ diagonal matrix whose elements are the largest r eigenvalues of $\frac{XX'}{NT}$. As $\Sigma_F^{\dagger} = p\lim_{T\to\infty}\frac{F'F^{\dagger}}{T} = \Sigma_{\lambda}^{\dagger -1/2}\Upsilon^{\dagger}V^{\dagger 1/2}$ (see e.g. Bai (2003), p.162), it is immediate to see that $p\lim_{N,T\to\infty} H'_{N,T}Q^{\dagger} = I_r$.

By taking a Taylor expansion around $H'_{N,T}F_t^{\dagger}$ and $Q^{\dagger}\beta^{\dagger}$, we have that

$$\begin{aligned}&\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T-h} g\left(\widetilde{\epsilon}_{1,t+h}\right) \\ &= \left(\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T-h} g\left(\epsilon_{1,t+h}\right) - \frac{1}{\sqrt{P}}\sum_{t=R+1}^{T-h}\left(\widetilde{F}_t - H'_{N,T}F_t^{\dagger}\right)'\nabla g_F\left(\epsilon_{1,t+h}\right) - \frac{1}{\sqrt{P}}\sum_{t=R+1}^{T-h}\nabla g_{\beta}\left(\epsilon_{1,t+h}\right)'\left(\widetilde{\beta}_T - Q^{\dagger}\beta^{\dagger}\right)\right. \\ &\left. + \frac{1}{2\sqrt{P}}\sum_{t=R+1}^{T-h}\left(\widetilde{F}_t - H'_{N,T}F_t^{\dagger}\right)'\nabla^2 g_{F,\beta}\left(\epsilon_{1,t+h}\right)\left(\widetilde{\beta}_T - Q^{\dagger}\beta^{\dagger}\right)\right)(1+o_p(1)). \qquad (12)\end{aligned}$$

Since we have, from A3(iv), that $\mathrm{E}\left(\nabla_F g\left(\epsilon_{1,t+h}\right)u_{i,t+h}^{\dagger}\right) = 0$ for all i,t, then by Lemma A.1(ii) in Bai and Ng (2006), it follows that

$$\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T-h}\left(\widetilde{F}_t - H'_{N,T}F_t^{\dagger}\right)'\nabla g_F\left(\epsilon_{1,t+h}\right) = O_p\left(\max\left\{\frac{\sqrt{P}}{N},\frac{\sqrt{P}}{T}\right\}\right).$$

The last term in the brackets in (12) is of smaller order than the second. As for the second term in the brackets in (12), since by construction $\frac{1}{T}\sum_{t=1}^{T}\widetilde{F}_t\widetilde{F}_t' = I_r$,

$$\begin{aligned}\widetilde{\beta}_T &= \frac{1}{T}\sum_{t=1}^{T-h}\widetilde{F}_t y_{t+h} = \frac{1}{T}\sum_{t=1}^{T-h}\widetilde{F}_t F_t^{\dagger\prime}\beta^{\dagger} + \frac{1}{T}\sum_{t=1}^{T}\widetilde{F}_t\epsilon_{1,t+h} \\ &= \left(Q^{\dagger}\beta^{\dagger} + \frac{1}{T}\sum_{t=1}^{T}\widetilde{F}_t\epsilon_{1,t+h}\right)(1+o_p(1)),\end{aligned}$$

as $\frac{1}{T}\sum_{t=1}^{T-h}\widetilde{F}_t F_t^{\dagger\prime} \xrightarrow{p} Q^{\dagger}$, by Proposition 1 in Bai (2003). Hence,

$$\begin{aligned}
&\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T-h}\nabla g_\beta\left(\epsilon_{1,t+h}\right)'\left(\widetilde{\beta}_T - Q^{\dagger}\beta^{\dagger}\right)\\
&= D^{\dagger\prime}\sqrt{P}\left(\widetilde{\beta}_T - Q^{\dagger}\beta^{\dagger}\right) + \frac{1}{P}\sum_{t=R+1}^{T-h}\left(\nabla g_\beta\left(\epsilon_{1,t+h}\right)' - D^{\dagger\prime}\right)\sqrt{P}\left(\widetilde{\beta}_T - Q^{\dagger}\beta^{\dagger}\right)\\
&= D^{\dagger\prime}\sqrt{P}\left(\widetilde{\beta}_T - Q^{\dagger}\beta^{\dagger}\right) + O_p\left(\frac{1}{\sqrt{P}}\right)
\end{aligned}$$

as $\left(\frac{1}{P}\sum_{t=R+1}^{T-h}\nabla g_\beta\left(\epsilon_{1,t+h}\right) - D^{\dagger}\right) = O_p\left(\frac{1}{\sqrt{P}}\right)$. Now,

$$\begin{aligned}
&\sqrt{P}\left(\widetilde{\beta}_T - Q^{\dagger}\beta^{\dagger}\right)\\
&= \sqrt{\frac{\pi}{1+\pi}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{F}_t\epsilon_{1,t+1}\\
&= \sqrt{\frac{\pi}{1+\pi}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}H_{N,T}'F_t^{\dagger}\epsilon_{1,t+1} + \sqrt{\frac{\pi}{1+\pi}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\widetilde{F}_t - H_{N,T}'F_t^{\dagger}\right)\epsilon_{1,t+1}\\
&= \sqrt{\frac{\pi}{1+\pi}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}H_{N,T}'F_t^{\dagger}\epsilon_{1,t+1} + O_p\left(\max\left\{\frac{\sqrt{P}}{N},\frac{\sqrt{P}}{T}\right\}\right),
\end{aligned}$$

as $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\widetilde{F}_t - H_{N,T}'F_t^{\dagger}\right)\epsilon_{1,t+1} = O_p\left(\max\left\{\frac{\sqrt{T}}{N},\frac{1}{\sqrt{T}}\right\}\right)$, by Lemma A1(iv) in Bai and Ng (2006), with $\pi = P/R$. Thus,

$$\begin{aligned}
&\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T-h}g\left(\widetilde{\epsilon}_{1,t+h}\right)\\
&= \frac{1}{\sqrt{P}}\sum_{t=R+1}^{T-h}g\left(\epsilon_{1,t+h}\right) - \sqrt{\frac{\pi}{1+\pi}}D^{\dagger\prime}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}H_{N,T}'F_t^{\dagger}\epsilon_{1,t+1} + O_p\left(\frac{1}{\sqrt{P}}\right)\\
&+ O_p\left(\max\left\{\frac{\sqrt{P}}{N},\frac{\sqrt{P}}{T}\right\}\right).
\end{aligned} \tag{13}$$

We now turn to the rolling estimation scheme. Notice that

$$\begin{aligned}
\widehat{\epsilon}_{2,t+h} = \epsilon_{2,t+h} &- \left(\widehat{F}_{t,t-R} - H_{N,R,t}'F_t^{\ddagger}\right)'Q_t^{\ddagger}\beta_t^{\ddagger} - F_t^{\ddagger\prime}H_{N,R,t}\left(\widehat{\beta}_{t,R} - Q_t^{\ddagger}\beta_t^{\ddagger}\right)\\
&+ \left(\widehat{F}_{t,t-R} - H_{N,R,t}'F_t^{\ddagger}\right)'\left(\widehat{\beta}_{t,R} - Q_t^{\ddagger}\beta_t^{\ddagger}\right),
\end{aligned}$$

where $Q_t^{\ddagger} = V_t^{\ddagger 1/2}\Upsilon_t^{\ddagger\prime}\Sigma_{\lambda,t}^{\ddagger -1/2}$, with $\Sigma_{\lambda,t}^{\ddagger} = p\lim_{N\to\infty}\frac{\Lambda_t^{\ddagger\prime}\Lambda_t^{\ddagger}}{N}$, $\widetilde{\Lambda}_t^{\ddagger} = \frac{F^{(t)\prime}X^{(t)}}{R}$, $X^{(t)}$ is an $R\times N$ matrix whose columns are given by $\left(x_{i,t-R+1},...,x_{i,t}\right)$, $t = R+1,..,T-h$, for $i = 1,...,N$, and $\widehat{F}^{(t)}$ is the $r\times R$ collection of vectors of factors estimated using observations from $t-R+1$ to t. Also, $V_t^{\ddagger} = \mathrm{diag}\left(v_{1,t}^{\ddagger},...,v_r^{\ddagger}\right)$, where $v_{1,t}^{\ddagger} > v_{2,t}^{\ddagger} > ... > v_{r,t}^{\ddagger} > 0$ are the eigenvalues of $\Sigma_{\lambda,t}^{\ddagger -1/2}\Sigma_{F,t}^{\ddagger}\Sigma_{\lambda,t}^{\ddagger -1/2}$,

$\Sigma_{F,t}^{\ddagger} = p\lim_{N,R\to\infty} \frac{F^{(t)\prime}F^{\ddagger(t)}}{R}$, $F^{\ddagger(t)} = \left(F_{t-R+1}^{\ddagger}, ..., F_t^{\ddagger}\right)$, and $\Upsilon_t^{\ddagger}$ is the matrix of the eigenvectors associated with $\left(v_{1,t}^{\ddagger}, ..., v_{r,t}^{\ddagger}\right)$, such that $\Upsilon_t^{\ddagger\prime}\Upsilon_t^{\ddagger} = I_r$, and $H_{N,R,t}^{\ddagger} = \frac{\Lambda_t^{\ddagger\prime}\Lambda_t^{\ddagger}}{N}\frac{F^{\ddagger\prime}F^{(t)}}{T}V_{N,R,t}^{-1}$, with $V_{N,R,t}$ a $r\times r$ diagonal matrix whose elements are the largest r eigenvalues of $\frac{X^{(t)}X^{(t)\prime}}{NR}$. Since, for all $t = R+1, ..., T$, $\Sigma_{F,t}^{\ddagger} = p\lim_{N,R\to\infty}\frac{F^{(t)\prime}F^{\ddagger(t)}}{R} = \Sigma_{\lambda,t}^{\ddagger-1/2}\Upsilon_t^{\ddagger}V_t^{\ddagger 1/2}$ (see e.g. Bai (2003), p.162), it is immediate to see that $p\lim_{N,R\to\infty} H_{N,R,t}^{\prime}Q_t^{\ddagger} = I_r$.

By taking a Taylor expansion around $H_{N,T}^{\prime}F_t^{\dagger}$ and $Q^{\dagger}\beta^{\dagger}$, we have that

$$
\begin{aligned}
&\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T-h} g\left(\widehat{\epsilon}_{2,t+h}\right)\\
&= \left(\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T-h} g\left(\epsilon_{2,t+h}\right) - \frac{1}{\sqrt{P}}\sum_{t=R+1}^{T-h}\left(\widehat{F}_{t,t-R} - H_{N,R,t}^{\prime}F_t^{\ddagger}\right)^{\prime}\nabla g_F\left(\epsilon_{2,t+h}\right)\right.\\
&- \frac{1}{\sqrt{P}}\sum_{t=R+1}^{T-h}\nabla g_{\beta}\left(\epsilon_{2,t+h}\right)^{\prime}\left(\widehat{\beta}_{t,R} - Q_t^{\ddagger}\beta_t^{\ddagger}\right)\\
&\left.+\frac{1}{2\sqrt{P}}\sum_{t=R+1}^{T-h}\left(\widehat{F}_{t,t-R} - H_{N,R,t}^{\prime}F_t^{\ddagger}\right)^{\prime}\nabla^2 g_{F,\beta}\left(\epsilon_{2,t+h}\right)\left(\widehat{\beta}_{t,R} - Q_t^{\ddagger}\beta_t^{\ddagger}\right)\right)\left(1 + o_p(1)\right).
\end{aligned}
$$

We first need to show that $\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T-h}\left(\widehat{F}_{t,t-R} - H_{N,R,t}^{\prime}F_t^{\ddagger}\right)^{\prime}\nabla g_F\left(\epsilon_{2,t+h}\right) = O_p\left(\max\left\{\frac{\sqrt{P}}{N}, \frac{\sqrt{P}}{R}\right\}\right)$, in the case in which $\widehat{F}_{t,t-R}$, is the factor at time t, and is estimated using observations from $t-R+1$ up to t. This is accomplished by showing that under assumptions A1-A3, Lemma A1 in Bai and Ng (2006) holds also for the rolling estimation case. Now, in this case

$$
\begin{aligned}
\widehat{F}_{t,t-R} - H_{N,R,t}^{\prime}F_t^{\ddagger} = \widehat{V}_{t,R}^{-1}&\left(\frac{1}{R}\sum_{j=t-R+1}^{t}\widehat{F}_{j,t-R}\gamma_{j,t} + \frac{1}{R}\sum_{j=t-R+1}^{t}\widehat{F}_{j,t-R}\zeta_{j,t}\right.\\
&\left.+\frac{1}{R}\sum_{j=t-R+1}^{t}\widehat{F}_{j,t-R}\eta_{j,t} + \frac{1}{R}\sum_{j=t-R+1}^{t}\widehat{F}_{j,t-R}\xi_{j,t}\right), \qquad (14)
\end{aligned}
$$

where $\widehat{V}_{t,R}$ is an $r\times r$ diagonal matrix containing the largest r eigenvalues of $\left(XX^{\prime}\right)^{(t-R,t)}/NR$, with the superscript denoting the subset of observations used, $\gamma_{j,t} = \mathrm{E}\left(\frac{1}{N}\sum_{i=1}^{N}u_{i,j}^{\ddagger}u_{i,t}^{\ddagger}\right)$, $\zeta_{j,t} = \frac{1}{N}\sum_{i=1}^{N}\left(u_{i,j}^{\ddagger}u_{i,t}^{\ddagger} - \gamma_{j,t}\right)$, $\eta_{j,t} = \frac{1}{N}\sum_{i=1}^{N}\lambda_i^{\prime\ddagger}F_{j,t-R}^{\ddagger}u_{i,t}^{\ddagger}$ and $\xi_{j,t} = \frac{1}{N}\sum_{i=1}^{N}\lambda_i^{\prime\ddagger}F_{t,t-R}^{\ddagger}u_{j,t}^{\ddagger}$. Hence, the only difference with respect to the full sample case is that the summation inside the brackets is taken from $t-R$ to t, for $t > R$, instead of over the full sample. Now, for $P/R \to \pi$, $0 \le \pi < \infty$, Assumption A1(i) ensures that for all i, l $\frac{1}{\sqrt{P}}\frac{1}{R}\sum_{t=R+1}^{T}\sum_{j=t-R+1}^{t}\mathrm{E}\left(u_{i,t}^{\ddagger}u_{l,j}^{\ddagger}\right) = O\left(1/\sqrt{P}\right)$. Hence, Lemma A1(ii) in Bai and Ng (2006) applies also for the rolling estimation case. Now,

$$
\widehat{\beta}_{t,R} = \left(\frac{1}{R}\sum_{j=t+1-R}^{t-h}\widehat{F}_{j,t-R}\widehat{F}_{j,t-R}^{\prime}\right)^{-1}\frac{1}{R}\sum_{j=t+1-R}^{t-h}\widehat{F}_{j,t-R}y_{j+h}
$$

$$= \frac{1}{R} \sum_{j=t+1-R}^{t-h} \widehat{F}_{j,t-R} y_{j+h} = \frac{1}{R} \sum_{j=t+1-R}^{t-h} \widehat{F}_{j,t-R} F_j^{\ddagger\prime} \beta_t^{\ddagger} + \frac{1}{R} \sum_{j=t+1-R}^{t-h} \widehat{F}_{j,t-R} \epsilon_{2,j+h}$$

$$= Q_t^{\ddagger} \beta_t^{\ddagger} + \frac{1}{R} \sum_{j=t+1-R}^{t-h} \widehat{F}_{j,t-R} \epsilon_{2,j+h}.$$

Thus,

$$\frac{1}{\sqrt{P}} \sum_{t=R+1}^{T-h} \nabla g_{\beta} \left(\epsilon_{2,t+h}\right)' \left(\widehat{\beta}_{t,R} - Q_t^{\ddagger} \beta_t^{\ddagger}\right)$$

$$= D^{\ddagger\prime} \frac{1}{\sqrt{P} R} \sum_{t=R+1}^{T-h} \sum_{j=t+1-R}^{t-h} \left(H'_{N,R,t} F_j^{\ddagger} \epsilon_{2,j+h} + \left(\widehat{F}_{j,t-R} - H'_{N,R,t} F_j^{\ddagger}\right) \epsilon_{2,j+h}\right) + O_p\left(\frac{1}{\sqrt{P}}\right),$$

as $\frac{1}{P} \sum_{t=R+1}^{T-h} \nabla g_{\beta} \left(\epsilon_{2,t+h}\right)' - D^{\ddagger} = O_p\left(\frac{1}{\sqrt{P}}\right)$. By Lemma A1(ii) in Bai and Ng (2006),

$$\frac{1}{\sqrt{P} R} \sum_{t=R+1}^{T-h} \sum_{j=t+1-R}^{t-h} \left(\widehat{F}_{j,t-R} - H'_{N,R,t} F_j^{\ddagger}\right) \epsilon_{2,j+h} = O_p\left(\max\left\{\frac{\sqrt{P}}{N}, \frac{\sqrt{P}}{R}\right\}\right).$$

Hence,

$$\frac{1}{\sqrt{P}} \sum_{t=R+1}^{T-h} g\left(\widehat{\epsilon}_{2,t+h}\right)$$

$$= \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T-h} g\left(\epsilon_{2,t+h}\right) + D^{\ddagger\prime} \frac{1}{\sqrt{P} R} \sum_{t=R+1}^{T-h} \sum_{j=t+1-R}^{t-h} H'_{N,R,t} F_j^{\ddagger} \epsilon_{2,j+h} + O_p\left(\frac{1}{\sqrt{P}}\right) + O_p\left(\max\left\{\frac{\sqrt{P}}{N}, \frac{\sqrt{P}}{R}\right\}\right). \tag{15}$$

Given (13)-(15), and recalling that R and T grow at the same rate, we have that

$$\frac{1}{\sqrt{P}} \sum_{t=R+1}^{T-h} \left(g\left(\widetilde{\epsilon}_{1,t+h}\right) - g\left(\widehat{\epsilon}_{2,t+h}\right)\right)$$

$$= \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T-h} \left(g\left(\epsilon_{1,t+h}\right) - g\left(\epsilon_{2,t+h}\right)\right) - \sqrt{\frac{\pi}{1+\pi}} D^{\dagger\prime} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} H'_{N,T} F_t^{\dagger} \epsilon_{1,t+1}$$

$$+ D^{\ddagger\prime} \frac{1}{\sqrt{P} R} \sum_{t=R+1}^{T-h} \sum_{j=t+1-R}^{t-h} H'_{N,R,t} F_j^{\ddagger} \epsilon_{2,j+h} + O_p\left(\frac{1}{\sqrt{P}}\right) + O_p\left(\max\left\{\frac{\sqrt{P}}{N}, \frac{\sqrt{P}}{R}\right\}\right). \tag{16}$$

Let

$$V_{\epsilon} = \sum_{j=-\infty}^{\infty} \mathrm{E}\left(\left(g\left(\epsilon_{1,1}\right) - g\left(\epsilon_{2,1}\right)\right) \times \left(g\left(\epsilon_{1,1+j}\right) - g\left(\epsilon_{2,1+j}\right)\right)\right), \tag{17}$$

$$V_{F^{\dagger}} = \sum_{j=-\infty}^{\infty} \mathrm{E}\left(H^{\dagger\prime} F_1^{\dagger} \epsilon_{1,1} \epsilon_{1,1+j} F_{1+j}^{\dagger\prime} H^{\dagger}\right),$$

and

$$V_{F^{\ddagger}} = \sum_{j=-\infty}^{\infty} \mathrm{E}\left(H^{\ddagger\prime} F_1^{\ddagger} \epsilon_{2,1} \epsilon_{2,1+j} F_{1+j}^{\ddagger\prime} H^{\ddagger}\right),$$

where $H^{\dagger} = p\lim_{T,N} \frac{1}{P}\sum_{t=R+1}^{T} H'_{N,T}$, $H^{\ddagger} = p\lim_{T,N} \frac{1}{P}\sum_{t=R+1}^{T} H'_{N,R,t}$,

$$C_{\epsilon,F^{\dagger}} = \sum_{j=-\infty}^{\infty} \mathrm{E}\left(\left(\left(g\left(\epsilon_{1,1}\right) - g\left(\epsilon_{2,1}\right)\right) \times H^{\dagger\prime} F_{1}^{\dagger} \epsilon_{1,1}\right)\right.$$
$$\left.\times\left(\left(g\left(\epsilon_{1,1+j}\right) - g\left(\epsilon_{2,1+j}\right)\right) \times H^{\dagger\prime} F_{1+j}^{\dagger} \epsilon_{1,1}\right)\right),$$

$$C_{\epsilon,F^{\ddagger}} = \sum_{j=-\infty}^{\infty} \mathrm{E}\left(\left(\left(g\left(\epsilon_{1,1}\right) - g\left(\epsilon_{2,1}\right)\right) \times H^{\ddagger\prime} F_{1}^{\ddagger} \epsilon_{1,1}\right)\right.$$
$$\left.\times\left(\left(g\left(\epsilon_{1,1+j}\right) - g\left(\epsilon_{2,1+j}\right)\right) \times H^{\ddagger\prime} F_{1+j}^{\ddagger} \epsilon_{1,1}\right)\right),$$

and

$$C_{F^{\dagger},F^{\ddagger}} = \sum_{j=-\infty}^{\infty} \mathrm{E}\left(H^{\dagger\prime} F_{1}^{\dagger} \epsilon_{1,1} \epsilon_{2,1+j} F_{1+j}^{\ddagger\prime} H^{\ddagger}\right).$$

Note that, for notational simplicity, we have written the expressions for the long-run covariances under the assumption of covariance stationarity. Nevertheless, even under the null of factor structural stability, $F_t^{\dagger}, F_t^{\ddagger}$ may display some time heterogeneity.
In this case, $C_{F^{\dagger},F^{\ddagger}} = \lim_{T,l_T\to\infty} \sum_{j=1+l_T}^{T-l_T} \sum_{\tau=-l_T}^{l_T} \mathrm{E}\left(H^{\dagger\prime} F_j^{\dagger} \epsilon_{1,j} \epsilon_{2,j+\tau} F_{j+\tau}^{\ddagger\prime} H^{\ddagger}\right)$, and the same applies for the definition of the other covariances.

By Lemma 4.1 in West and McCracken (1998), and along the same lines as in the proof of Proposition 1(a) in Corradi and Swanson (2006b), for $P \leq R$,

$$\lim_{T,N\to\infty} \operatorname{var}\left(\frac{1}{\sqrt{P}R} \sum_{t=R+1}^{T-h} \sum_{j=t+1-R}^{t-h} H^{\ddagger\prime} F_j^{\ddagger} \epsilon_{2,j+h}\right) = \left(\pi - \frac{\pi^2}{3}\right) V_{F^{\ddagger}},$$

$$\lim_{T,N\to\infty} \operatorname{cov}\left(\frac{1}{\sqrt{P}R} \sum_{t=R+1}^{T-h} \sum_{j=t+1-R}^{t-h} H^{\ddagger\prime} F_j^{\ddagger} \epsilon_{2,j+h}, \frac{1}{\sqrt{P}} \sum_{t=R+1}^{T-h} \left(g\left(\epsilon_{1,t+h}\right) - g\left(\epsilon_{2,t+h}\right)\right)\right)$$
$$= \frac{\pi}{2} C_{\epsilon,F^{\ddagger}},$$

and

$$\lim_{T,N\to\infty} \operatorname{cov}\left(\frac{1}{\sqrt{P}R} \sum_{t=R+1}^{T-h} \sum_{j=t+1-R}^{t-h} H'_{N,R,t} F_j^{\ddagger} \epsilon_{2,j+h}, \frac{1}{\sqrt{T}} \sum_{t=1}^{T} H'_{N,T} F_t^{\dagger} \epsilon_{1,t+1}\right)$$
$$= \frac{\pi}{2} C_{F^{\dagger},F^{\ddagger}}.$$

Hence, as a straightforward application of the central limit theorem for possibly heterogeneous mixing processes, (see e.g. Wooldridge and White (1988)), under the null of $\mathrm{E}\left(g\left(\epsilon_{1,t+h}\right) - g\left(\epsilon_{2,t+h}\right)\right) = 0$,

$$\frac{1}{\sqrt{P}} \sum_{t=R+1}^{T-h} \left(g\left(\widetilde{\epsilon}_{1,t+h}\right) - g\left(\widehat{\epsilon}_{2,t+h}\right)\right) \xrightarrow{d} N\left(0, \Omega\right),$$

where

$$\Omega = V_\epsilon + \frac{\pi}{1+\pi} D^{\dagger\prime} V_{F^\dagger} D^\dagger + \left(\pi - \frac{\pi^2}{3}\right) D^{\ddagger\prime} V_{F^\ddagger} D^\ddagger - \sqrt{\frac{\pi}{1+\pi}} D^{\dagger\prime} C_{\epsilon,F^\dagger}$$
$$+ \frac{\pi}{2}\sqrt{\frac{\pi}{1+\pi}} D^{\ddagger\prime} C_{\epsilon,F^\ddagger} - \frac{\pi}{2}\sqrt{\frac{\pi}{1+\pi}} D^{\dagger\prime} C_{F^\dagger,F^\ddagger} D^\ddagger, \tag{18}$$

and, in the case where $P > R$, the terms $\left(\pi - \frac{\pi^2}{3}\right)$ and $\frac{\pi}{2}$ in (18) should be replaced by $\left(1 - \frac{1}{3\pi}\right)$ and $\left(1 - \frac{1}{2\pi}\right)$, respectively.

The statements for Case II and Case III follow in straightforward fashion, given (17) and by inspection of the asymptotic covariance matrix in (18).

Turning to the statement in Case IV, it follows immediately from (16) that $\frac{1}{\sqrt{P}} \sum_{t=R+1}^{T-h} \left(g\left(\widetilde{\epsilon}_{1,t+h}\right) - g\left(\widehat{\epsilon}_{2,t+h}\right)\right) = O_p\left(\frac{1}{\sqrt{P}}\right) + O_p\left(\max\left\{\frac{\sqrt{P}}{N}, \frac{\sqrt{P}}{R}\right\}\right)$. Hence, for $N/R \to \infty$, this expression is of probability order $1/\sqrt{P}$, while if $R/N^2 \to 0$ but $R/N \to \infty$, it is at most of probability order $\sqrt{P}/N$. Note that $O_p\left(\max\left\{\frac{\sqrt{P}}{N}, \frac{\sqrt{P}}{R}\right\}\right)$ is an upper bound, rather than an "exact" order. This is because Lemma A1 in Bai and Ng (2006) follows via a sequence of majorizations.

Finally, the statement under the alternative hypothesis follows immediately, as $\mathrm{E}((g(\epsilon_{1,t+h}) - g(\epsilon_{2,t+h}))) \neq 0$.□

Hereafter, $\Pr^*$ denotes the probability law of the bootstrap samples, conditional on the sample, and , E^* and var^* denote the mean and the variance under $\Pr^*$. Also, $O_{p^*}(1)$ and $o_{p^*}(1)$ denote terms bounded and converging to zero under $\Pr^*$. Let $H_{N,T}' F_t^{\dagger *}$ and $H_{N,R,t}' F_t^{\ddagger *}$ $t = R^*, ..., R^* + P^* - h$, be the series of factors resampled from $H_{N,T}' F_t^\dagger$ and from $H_{N,R,t}' F_t^\ddagger$, respectively, for $t = R, ..., R+P-h$. In the sequel, we rely on the following Lemma.

Lemma 1: *Let Assumption A hold. Also, as* $N, T, P, R \to \infty$, $N/\sqrt{T} \to \infty$, *and* $P/R \to \pi$, *with* $0 \leq \pi < \infty$. *Additionally, as* $P^*, R^*, T^* \to \infty$, *assume that* $T^*/T \to 0$, $P^*/P \to 0$, $R^*/R \to 0$, *and* $P^*/R^* \to \pi$. *Finally, as* $l_{P^*}, l_{R^*}, l_{T^*} \to \infty$, $l_{T^*}/\sqrt{T^*} \to 0, l_{P^*}/\sqrt{P^*} \to 0$, *and* $l_{R^*}/\sqrt{R^*} \to 0$. *Then:*
(i)

$$\frac{1}{\sqrt{P^*}} \sum_{t=R^*+1}^{T^*-h} \left(\widetilde{F}_t^* - H_{N,T}' F_t^{\dagger *}\right)' \nabla g_F\left(\epsilon_{1,t+h}^*\right)$$
$$= O_{p^*}\left(\max\left\{\frac{\sqrt{P^*}}{N}, \frac{\sqrt{P^*}}{T}\right\} + \sqrt{l^* P^{1/k} \max\left\{\frac{1}{T}, \frac{1}{N}\right\}} + l^*/\sqrt{P^*}\right) = O_{p^*}\left(d_{N,T,P^*}^{(1)}\right).$$

and **(ii)**

$$\frac{1}{\sqrt{P^*}} \sum_{t=R^*+1}^{T^*-h} \left(\widehat{F}_{t,t-R}^* - H_{N,R,t}' F_t^{\ddagger *}\right)' \nabla g_F\left(\epsilon_{2,t+h}^*\right)$$

$$= O_{p^*}\left(\max\left\{\frac{\sqrt{P^*}}{N},\frac{\sqrt{P^*}}{R}\right\}+\sqrt{l^* P^{1/k}\max\left\{\frac{1}{R},\frac{1}{N}\right\}}+l^*/\sqrt{P^*}\right)=O_{p^*}\left(d^{(1)}_{N,R,P^*}\right),$$

where k is as defined in A1(ii)-(iii).

Proof of Theorem 2: We begin with the case of the full sample estimation scheme. Let $\widetilde{\epsilon}^*_{1,t+h}=y^*_{t+h}-\widetilde{F}^{*\prime}_t\widetilde{\beta}^*_T$, and $\epsilon^*_{1,t+h}=y^*_{t+h}-F^{\dagger*\prime}_t H_{N,T}\widetilde{\beta}_T$, so that

$$\begin{aligned}\widetilde{\epsilon}^*_{1,t+h}&=\epsilon^*_{1,t+h}-\left(\widetilde{F}^*_t-H'_{N,T}F^{\dagger*}_t\right)'\widetilde{\beta}_T-F^{\dagger*\prime}_t H_{N,T}\left(\widetilde{\beta}^*_{T^*}-\widetilde{\beta}_T\right)\\&\quad+\left(\widetilde{F}^*_t-H'_{N,T}F^{\dagger*}_t\right)'\left(\widetilde{\beta}^*_{T^*}-\widetilde{\beta}_T\right).\end{aligned}$$

By taking a Taylor expansion of $g\left(\widetilde{\epsilon}^*_{1,t+h}\right)$ around $\widetilde{\beta}_T$ and $H'_{N,T}F^{\dagger*}_t$, note that, by Lemma 1(i),

$$\begin{aligned}&\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\left(\widetilde{F}^*_t-H'_{N,T}F^{\dagger*}_t\right)'\nabla g_F\left(\epsilon^*_{1,t+h}\right)\\&=O_{p^*}\left(\max\left\{\frac{\sqrt{P^*}}{N},\frac{\sqrt{P^*}}{T}\right\}+\sqrt{l^* P^{1/k}\max\left\{\frac{1}{T},\frac{1}{N}\right\}}+l^*/\sqrt{P^*}\right)=O_{p^*}\left(d^{(1)}_{N,T,P^*}\right).\end{aligned}$$

Recalling the definition of $\epsilon^*_{1,t+h}$, $\widetilde{\beta}^*_{T^*}$ and $\widetilde{\beta}_T$, note that

$$\begin{aligned}&\sqrt{P^*}\left(\widetilde{\beta}^*_{T^*}-\widetilde{\beta}_T\right)\\&=\frac{\sqrt{P^*}}{T^*}\sum_{t=1}^{T^*-h}H'_{N,T}F^{\dagger*}_t\epsilon^*_{1,t+h}+\frac{\sqrt{P^*}}{T^*}\sum_{t=1}^{T^*-h}H'_{N,T}F^{\dagger*}_t\epsilon^*_{1,t+h}\left(\left(\frac{1}{T^*}\sum_{t=R^*+1}^{T^*-h}\widetilde{F}^*_t\widetilde{F}^{*\prime}_t\right)^{-1}-I_r\right)\\&+\left(\frac{1}{T^*}\sum_{t=1}^{T^*-h}\widetilde{F}^*_t\widetilde{F}^{*\prime}_t\right)^{-1}\frac{\sqrt{P^*}}{T^*}\sum_{t=1}^{T^*-h}\left(\widetilde{F}^*_t-H'_{N,T}F^{\dagger*}_t\right)\epsilon^*_{1,t+h}\\&=\frac{\sqrt{P^*}}{T^*}\sum_{t=R^*+1}^{T^*-h}H'_{N,T}F^{\dagger*}_t\epsilon^*_{1,t+h}+O_{p^*}\left(d^{(1)}_{N,T,P^*}\right).\end{aligned}\tag{19}$$

Thus,

$$\begin{aligned}&\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}g\left(\widetilde{\epsilon}^*_{1,t+h}\right)\\&=\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}g\left(\epsilon^*_{1,t+h}\right)+\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}D^{\dagger\prime}H'_{N,T}F^{\dagger*}_t\epsilon^*_{1,t+h}+O_{p^*}\left(\frac{1}{\sqrt{P^*}}\right)\\&+O_p\left(\frac{1}{\sqrt{P}}\right)+O_{p^*}\left(d^{(1)}_{N,P,P^*}\right),\end{aligned}\tag{20}$$

as $\frac{1}{P^*}\sum_{t=R^*+1}^{T^*-h}\left(\nabla g_\beta\left(\epsilon^*_{1,t+h}\right)'-\mathrm{E}^*\left(\nabla g_\beta\left(\epsilon^*_{1,t+h}\right)'\right)\right)=O_{p^*}\left(\frac{1}{\sqrt{P^*}}\right)$, and $\mathrm{E}^*\left(\nabla g_\beta\left(\epsilon^*_{1,t+h}\right)'\right)-D^{\dagger\prime}=O_p\left(\frac{1}{\sqrt{P}}\right)$.

We now turn to the case of a rolling estimation scheme. Let $\widehat{\epsilon}^*_{2,t+h} = y^*_{t+h} - \widehat{F}^{*\prime}_{t,t-R}\widehat{\beta}^*_{t,R^*}$, and $\epsilon^*_{2,t+h} = y^*_{t+h} - F^{\ddagger*\prime}_t H_{N,R,t}\widehat{\beta}_{t,R}$, so that

$$\widehat{\epsilon}^*_{2,t+h} = \epsilon^*_{2,t+h} - \left(\widehat{F}^*_{t,t-R} - H'_{N,R,t}F^{\ddagger*}_t\right)'\widehat{\beta}^*_{t,R^*} - F^{\ddagger*\prime}_t H_{N,R,t}\left(\widehat{\beta}^*_{t,R^*} - \widehat{\beta}_{t,R}\right)$$
$$+ \left(\widehat{F}^*_{t,t-R} - H'_{N,R,t}F^{\ddagger*}_t\right)'\left(\widehat{\beta}^*_{t,R^*} - \widehat{\beta}_{t,R}\right).$$

By taking a Taylor expansion of $g\left(\widehat{\epsilon}^*_{2,t+h}\right)$ around $\widehat{\beta}_{t,R}$ and $H'_{N,R,t}F^{\ddagger*}_t$, note that

$$\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h} g\left(\widehat{\epsilon}^*_{2,t+h}\right)$$
$$= \left(\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h} g\left(\epsilon^*_{2,t+h}\right) - \frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\left(\widehat{F}^*_{t,t-R} - H'_{N,R,t}F^{\ddagger*}_t\right)'\nabla g_F\left(\epsilon^*_{2,t+h}\right)\right.$$
$$- \frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\nabla g_\beta\left(\epsilon^*_{2,t+h}\right)'\left(\widehat{\beta}^*_{t,R^*} - \widehat{\beta}_{t,R}\right)$$
$$\left.+\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\left(\widehat{F}^*_{t,t-R} - H'_{N,R,t}F^{\ddagger*}_t\right)'\nabla^2 g_{F,\beta}\left(\epsilon^*_{2,t+h}\right)\left(\widehat{\beta}^*_{t,R^*} - \widehat{\beta}_{t,R}\right)\right)(1 + o_{p^*}(1)).$$

Given Lemma 1(ii),

$$\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\left(\widehat{F}^*_{t,t-R} - H'_{N,R,t}F^{\ddagger*}_t\right)'\nabla g_F\left(\epsilon^*_{2,t+h}\right)$$
$$= O_{p^*}\left(\max\left\{\frac{\sqrt{P^*}}{N}, \frac{\sqrt{P^*}}{R}\right\} + \sqrt{l^*P^{1/k}\max\left\{\frac{1}{R}, \frac{1}{N}\right\} + l^*/\sqrt{P^*}}\right) = O_{p^*}\left(d^{(1)}_{N,R,P^*}\right).$$

Now,

$$\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\left(\widehat{\beta}^*_{t,R^*} - \widehat{\beta}_{t,R}\right)$$
$$= \frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\frac{1}{R^*}\sum_{j=t-R^*+1}^{t} H'_{N,t}F^{\ddagger*}_j\epsilon^*_{1,j+h}$$
$$+ \frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\left(\left(\frac{1}{R^*}\sum_{j=t-R^*+1}^{t}\widehat{F}^*_{j,t}\widehat{F}^{*\prime}_{j,t}\right)^{-1} - I_r\right)\frac{1}{R^*}\sum_{j=t-R^*+1}^{t} H'_{N,t}F^{\ddagger*}_j\epsilon^*_{1,j+h}$$
$$+ \frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\left(\frac{1}{R^*}\sum_{j=t-R^*+1}^{t}\widehat{F}^*_{j,t}\widehat{F}^{*\prime}_{j,t}\right)^{-1}\frac{1}{R^*}\sum_{j=t-R^*+1}^{t}\left(\widehat{F}^*_{j,t} - H'_{N,t}F^{\ddagger*}_j\right)\epsilon^*_{1,j+h}$$
$$= \frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\frac{1}{R^*}\sum_{j=t-R^*+1}^{t} H'_{N,t}F^{\ddagger*}_j\epsilon^*_{1,j+h} + O_{p^*}\left(d^{(1)}_{N,R,P^*}\right). \tag{21}$$

Thus,

$$\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h} g\left(\widehat{\epsilon}_{2,t+h}^*\right)$$
$$= \frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h} g\left(\epsilon_{2,t+h}^*\right) + D^{\ddagger\prime}\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\frac{1}{R^*}\sum_{j=t-R^*+1}^{t} H_{N,t}^{\prime}F_j^{\ddagger *}\epsilon_{1,j+h}^* +$$
$$+ O_p\left(\frac{1}{\sqrt{P}}\right) + O_{p^*}\left(d_{N,P,P^*}^{(1)}\right) + O_{p^*}\left(\frac{1}{\sqrt{P^*}}\right), \tag{22}$$

as $\frac{1}{P^*}\sum_{t=R^*+1}^{T^*-h}\left(\nabla g_\beta\left(\epsilon_{2,t+h}^*\right)^{\prime} - \mathrm{E}^*\left(\nabla g_\beta\left(\epsilon_{2,t+h}^*\right)^{\prime}\right)\right) = O_{p^*}\left(\frac{1}{\sqrt{P^*}}\right)$ and $\mathrm{E}^*\left(\nabla g_\beta\left(\epsilon_{2,t+h}^*\right)^{\prime}\right) - D^{\ddagger\prime} = O_p\left(\frac{1}{\sqrt{P}}\right).$

Given (20) and (22), and recalling that R and T grow at the same rate,

$$\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\left(g\left(\widehat{\epsilon}_{1,t+h}^*\right) - g\left(\widehat{\epsilon}_{2,t+h}^*\right)\right)$$
$$= \frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\left(g\left(\epsilon_{1,t+h}^*\right) - g\left(\epsilon_{2,t+h}^*\right)\right) - \frac{\sqrt{P^*}}{T^*}\sum_{t=R^*+1}^{T^*-h} D^{\dagger\prime}H_{N,T}^{\prime}F_t^{\dagger *}\epsilon_{1,t+h}^*$$
$$+ D^{\ddagger\prime}\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\frac{1}{R^*}\sum_{j=t-R^*+1}^{t} H_{N,t}^{\prime}F_j^{\ddagger *}\epsilon_{1,j+h}^*$$
$$+ O_p\left(\frac{1}{\sqrt{P}}\right) + O_{p^*}\left(d_{N,R,P^*}^{(1)}\right) + O_{p^*}\left(\frac{1}{\sqrt{P^*}}\right). \tag{23}$$

We first need to show that in Cases I-III, the left hand terms in (16) and (23) have the same limiting distribution, conditional on the sample, and for all samples except a set with probability measure approaching to zero. Now, note that

$$\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\left(g\left(\epsilon_{1,t+h}^*\right) - g\left(\epsilon_{2,t+h}^*\right)\right)$$
$$= \frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\left(g\left(y_{t+h}^* - F_t^{\dagger *\prime}\beta^{\dagger}\right) - g\left(y_{t+h}^* - F_t^{\ddagger *\prime}\beta_t^{\ddagger}\right)\right) + O_p\left(\frac{\sqrt{P^*}}{\sqrt{P}}\right), \tag{24}$$

so that in Cases I-III the first term on the RHS of (24) has the same limiting distribution, conditional on the sample, as $\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T-h}\left(g\left(\epsilon_{1,t+h}\right) - g\left(\epsilon_{2,t+h}\right)\right).$ Furthermore, by the same argument as in the proof of Proposition 2 of Corradi and Swanson (2006b), whenever $D^{\dagger}$ and $D^{\ddagger}$ are different from zero, $\frac{\sqrt{P^*}}{T^*}\sum_{t=R^*+1}^{T^*-h} D^{\dagger\prime}H_{N,T}^{\prime}F_t^{\dagger *}\epsilon_{1,t+h}^*$ and $D^{\ddagger\prime}\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\frac{1}{R^*}\sum_{j=t-R^*+1}^{t} H_{N,t}^{\prime}F_j^{\ddagger *}\epsilon_{1,j+h}^*$ have the same limiting distribution as $\sqrt{\frac{\pi}{1+\pi}}D^{\dagger\prime}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} H_{N,T}^{\prime}F_t^{\dagger}\epsilon_{1,t+1}$ and $D^{\ddagger\prime}\frac{1}{\sqrt{P}R}\sum_{t=R+1}^{T-h}\sum_{j=t+1-R}^{t-h} H_{N,R,t}^{\prime}F_j^{\ddagger}\epsilon_{2,j+h}$, respectively, conditional on the sample and for all samples except a set with probability measure approaching zero.

We now need to show that in Case IV, $\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\left(g\left(\widetilde{\epsilon}_{1,t+h}^*\right)-g\left(\widehat{\epsilon}_{2,t+h}^*\right)\right)$ approaches zero at a slower rate than $\frac{1}{\sqrt{P}}\sum_{t=R^*+1}^{T^*-h}\left(g\left(\widetilde{\epsilon}_{1,t+h}\right)-g\left(\widehat{\epsilon}_{2,t+h}\right)\right)$. By looking at (16) and (23), whenever $N/R\to\infty$ or $N/R\to c>0$, the statistic is $O_p\left(1/\sqrt{P}\right)$, while the bootstrap statistic cannot go to zero at a rate faster than $1/\sqrt{P^*}$. As $P^*/P\to 0$, the bootstrap statistic will always approach zero at a slower rate. On the other hand, when $N/R\to 0$, the statistic in (16) is at most of order $O_p\left(\frac{\sqrt{P}}{N}\right)$. Moreover, the statistic in (23) cannot go to zero at a rate faster than $\frac{1}{\sqrt{P^*}}$. Hence, if $P^*<\frac{N^2}{P}$, then the bootstrap statistic will approach zero at a slower rate than the actual statistic. This ensures a test with an asymptotic zero size. Finally, under the alternative, S_P diverges at rate $\sqrt{P}$, while $S_{P^*}^*$ diverges at rate $\sqrt{P}^*$, thus ensuring unit asymptotic power.

Proof of Lemma 1: (i)

$$
\begin{aligned}
&\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\left(\widetilde{F}_t^*-H_{N,T}'F_t^{\dagger*}\right)'\nabla g_F\left(\epsilon_{1,t+h}^*\right)\\
&=\mathrm{E}^*\left(\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\left(\widetilde{F}_t^*-H_{N,T}'F_t^{\dagger*}\right)'\nabla g_F\left(\epsilon_{1,t+h}^*\right)\right)\\
&+\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\left(\left(\widetilde{F}_t^*-H_{N,T}'F_t^{\dagger*}\right)'\nabla g_F\left(\epsilon_{1,t+h}^*\right)-\mathrm{E}^*\left(\left(\widetilde{F}_t^*-H_{N,T}'F_t^{\dagger*}\right)'\nabla g_F\left(\epsilon_{1,t+h}^*\right)\right)\right). \quad (25)
\end{aligned}
$$

Now,

$$
\begin{aligned}
&\mathrm{E}^*\left(\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\left(\widetilde{F}_t^*-H_{N,T}'F_t^{\dagger*}\right)'\nabla g_F\left(\epsilon_{1,t+h}^*\right)\right)\\
&=\sqrt{P^*}\frac{1}{P}\sum_{t=R}^{T}\left(\widetilde{F}_t-H_{N,T}'F_t^{\dagger}\right)'\nabla g_F\left(\epsilon_{1,t+h}\right)+O_p\left(l^*/P^*\right)\\
&=O_p\left(\max\left\{\frac{\sqrt{P^*}}{N},\frac{\sqrt{P^*}}{\sqrt{P}}\right\}\right)+O_p\left(l^*/P^*\right) \qquad (26)
\end{aligned}
$$

by the same argument used in the proof of Theorem 1. Now, by a similar argument as in the proof of Theorem 3 in Corradi and Swanson (2006c), up to a $O_p\left(l^*/\sqrt{P^*}\right)$ term,

$$
\begin{aligned}
&\mathrm{var}^*\left(\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\left(\widetilde{F}_t^*-H_{N,T}'F_t^{\dagger*}\right)'\nabla g_F\left(\epsilon_{1,t+h}^*\right)\right)\\
&=\frac{1}{P}\sum_{t=R-l^*}^{T-l^*}\sum_{j=-l^*}^{l^*}\left(\left(\widetilde{F}_t-H_{N,T}'F_t^{\dagger}\right)'\nabla g_F\left(\epsilon_{1,t+h}\right)\right)\left(\left(\widetilde{F}_{t-j}-H_{N,T}'F_{t-j}^{\dagger}\right)'\nabla g_F\left(\epsilon_{1,t+h-j}\right)\right)+o_p(1)\\
&\le l^*\sup_{t\ge R}\left|\left(\widetilde{F}_t-H_{N,T}'F_t^{\dagger}\right)'\nabla g_F\left(\epsilon_{1,t+h}\right)\right|\frac{1}{P}\sum_{t=R-l^*}^{T-l^*}\left|\left(\widetilde{F}_t-H_{N,T}'F_t^{\dagger}\right)'\nabla g_F\left(\epsilon_{1,t+h}\right)\right|
\end{aligned}
$$

As in the proof of Theorem 1,

$$\frac{1}{P}\sum_{t=R-l^*}^{T-l^*}\left|\left(\widetilde{F}_t-H_{N,T}'F_t^{\dagger}\right)'\nabla g_F\left(\epsilon_{1,t+h}\right)\right|=O_p\left(\max\left\{\frac{1}{P},\frac{1}{N}\right\}\right). \tag{27}$$

Now, from Eq.(A.1) in Bai and Ng (2006),

$$\left(\widetilde{F}_t-H_{N,T}'F_t^{\dagger}\right)'\nabla g_F\left(\epsilon_{1,t+h}\right)=\widehat{V}_T^{-1}\left(\frac{1}{T}\sum_{j=1}^{T}\widetilde{F}_j'\nabla g_F\left(\epsilon_{1,t+h}\right)\gamma_{j,t}+\frac{1}{T}\sum_{j=1}^{T}\widetilde{F}_j'\nabla g_F\left(\epsilon_{1,t+h}\right)\zeta_{j,t}\right.$$
$$\left.\frac{1}{T}\sum_{j=1}^{T}\widetilde{F}_j'\nabla g_F\left(\epsilon_{1,t+h}\right)\eta_{j,t}+\frac{1}{T}\sum_{j=1}^{T}\widetilde{F}_j'\nabla g_F\left(\epsilon_{1,t+h}\right)\xi_{j,t}\right),$$

where $\gamma_{j,t},\zeta_{j,t},\eta_{j,t}$ and $\xi_{j,t}$ are defined as below Eq.(14). Now, $\sup_{t,j}|\gamma_{j,t}|<M$, by A2(i), for $\sqrt{T}/N\to 0$, $\sup_t|\zeta_{j,t}|=o_p(1)$ given A2(ii), given A1(ii), for $\sqrt{T}/N\to 0$, $\sup_t|\zeta_{j,t}|=o_p(1)$ and

$$\sup_{t,j\geq R}|\xi_{j,t}|=\left|\frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{r}\sum_{\iota=1}^{r}\lambda_{i,l}F_{l,t}^{\dagger}\nabla g_{F_\iota}\left(\epsilon_{1,t+h}\right)u_{i,j}\right|$$
$$\leq\max_{l,h=1,\dots,r}\sup_{t\geq R}\left|F_{l,t}^{\dagger}\nabla g_{F_h}\left(\epsilon_{1,t+h}\right)\right|\sup_{t\geq R}\frac{1}{N}\sum_{i=1}^{N}\sum_{l=1}^{r}|\lambda_{i,l}u_{i,j}|\leq\max_{l,h=1,\dots,r}\sup_{t\geq R}\left|F_{l,t}^{\dagger}\nabla g_{F_h}\left(\epsilon_{1,t+h}\right)\right|O_p(1). \tag{28}$$

Hence, given (28) and noting that $\frac{1}{T}\sum_{j=1}^{T}\left|\widetilde{F}_j'\nabla g_F\left(\epsilon_{1,t+h}\right)\right|=O_p(1)$ uniformly in t,

$$\begin{aligned}&\sup_{t\geq R}\left|\left(\widetilde{F}_t-H_{N,T}'F_t^{\dagger}\right)'\nabla g_F\left(\epsilon_{1,t+h}\right)\right|\\&\leq\sup_{t,j\geq R}\left(|\xi_{j,t}|\frac{1}{T}\sum_{j=1}^{T}\left|\widetilde{F}_j'\nabla g_F\left(\epsilon_{1,t+h}\right)\right|\right)\\&\leq O_p(1)\max_{\iota,h=1,\dots,r}\sup_{t\geq R}\left|F_{\iota,t}^{\dagger}\nabla g_{F_h}\left(\epsilon_{1,t+h}\right)\right|.\end{aligned}$$

Now,

$$\begin{aligned}&\Pr\left(\max_{\iota,h=1,\dots,r}\sup_{t\geq R}P^{-1/k}\left|F_{\iota,t}^{\dagger}\nabla g_{F_h}\left(\epsilon_{1,t+h}\right)\right|>\varepsilon\right)\\&\leq C\sum_{t=R}^{T}\Pr\left(P^{-1/k}\left|F_{l,t}^{\dagger}\nabla g_{F_h}\left(\epsilon_{1,t+h}\right)\right|>\varepsilon\right)\leq\frac{C}{\varepsilon^{k+1}}P^{\frac{k}{k+1}-1}\sup_{t\geq R}\mathrm{E}\left(\left|F_{l,t}^{\dagger}\nabla g_{F_h}\left(\epsilon_{1,t+h}\right)\right|^{k+1}\right)=o(1),\end{aligned}$$

given A3(iii). Thus, recalling (27), up to a $O_p\left(l^*/\sqrt{P^*}\right)$ term,

$$\mathrm{var}^*\left(\frac{1}{\sqrt{P^*}}\sum_{t=R^*+1}^{T^*-h}\left(\widetilde{F}_t^*-H_{N,T}'F_t^{\dagger*}\right)'\nabla g_F\left(\epsilon_{1,t+h}^*\right)\right)$$

$$= O_p\left(l^* P^{1/k}\right) O_p\left(\max\left\{\frac{1}{P}, \frac{1}{N}\right\}\right). \tag{29}$$

Given (26) and (29), it follows that

$$\frac{1}{\sqrt{P^*}} \sum_{t=R^*+1}^{T^*-h} \left(\widetilde{F}_t^* - H_{N,T}' F_t^{\dagger *}\right)' \nabla g_F\left(\epsilon_{1,t+h}^*\right)$$
$$= O_{p^*}\left(\max\left\{\frac{\sqrt{P^*}}{N}, \frac{\sqrt{P^*}}{T}\right\} + \sqrt{l^* P^{1/k} \max\left\{\frac{1}{T}, \frac{1}{N}\right\} + l^*/\sqrt{P^*}}\right) = O_{p^*}\left(d_{N,T,P^*}^{(1)}\right). \tag{30}$$

(**ii**) Recalling (14), by a similar argument as that used in the full sample estimation scheme,

$$\frac{1}{\sqrt{P^*}} \sum_{t=R^*+1}^{T^*-h} \left(\widehat{F}_{t,t-R}^* - H_{N,R,t}' F_t^{\ddagger *}\right)' \nabla g_F\left(\epsilon_{2,t+h}^*\right)$$
$$= O_{p^*}\left(\max\left\{\frac{\sqrt{P^*}}{N}, \frac{\sqrt{P^*}}{R}\right\} + \sqrt{l^* P^{1/k} \max\left\{\frac{1}{R}, \frac{1}{N}\right\} + l^*/\sqrt{P^*}}\right) = O_{p^*}\left(d_{N,R,P^*}^{(1)}\right) \tag{31}$$

8 References

Armah, Nii Ayi and Norman R. Swanson, (2010). Seeing Inside the Black Box: Using Diffusion Index Methodology to Construct Factor Proxies in Largescale Macroeconomic Time Series Environments, *Econometric Reviews*, 29, 476-510.

Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions, *Econometrica*, 71, 135-171.

Bai, J. and S. Ng (2002). Determining the Number of Factors in Approximate Factor Models, *Econometrica*, 70, 191-221.

Bai, J. and S. Ng (2006). Confidence Intervals for Diffusion Index Forecasts and Inference with Factor Augmented Regressions. *Econometrica*, 74, 1133-1150.

Banerjee, A., M. Marcellino and I. Marsten (2009). Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Changes. forthcoming in *Forecasting in the Presence of Structural Breaks and Model Uncertainty,* edited by D. Rapach and M. Wohar, Elsevier.

Banerjee, A., M. Marcellino and I. Marsten (2010). Forecasting with Factor-Augmented Error Correction Models. Working Paper, European University Institute.

Bickel, P.J., F. Götze and W.R. van Zwet (1997). Resampling Fewer than n Observations: Gains, Losses and Remedies for Losses. *Statistica Sinica,* 7, 1-31.

Bickel, P.J. and A. Sakov (2008). On the Choice of m in the m out n Bootstrap and Confidence Bounds for Extrema. *Statistica Sinica,* 18, 967-985

Breitung, J. and S. Eickmeier (2011). Testing for Structural Breaks in Dynamic Factor Models. *Journal of Econometrics,* 163, 71-84.

Castle, J.L., J.A. Doornik and D.F. Hendry (2010). Model Selection when there are Multiple Breaks. Working Paper, University of Oxford.

Chen, L., J.J. Dolado and J. Gonzalo (2011). Detecting Big Structural Breaks in Large Factor Models. MPRA W.P. 31344.

Clements, M.P. and D.F. Hendry (2002). Modeling Methodology and Forecast Failure. *Econometrics Journal,* 5, 319-344.

Clements, M.P. and D.F. Hendry (2006). Forecasting with Breaks, in *Handbook of Economic Forecasting,* edited by G. Elliott, C.W.J. Granger and A. Timmermann, North-Holland Elsevier.

Corradi, V. and N.R. Swanson (2006a). Predictive Density Evaluation, in *Handbook of Economic Forecasting,* edited by G. Elliott, C.W.J. Granger and A. Timmermann, North-Holland Elsevier.

Corradi, V. and N.R. Swanson (2006b). Predictive Density and Conditional Confidence Interval Accuracy Tests. *Journal of Econometrics,* 135-228.

Corradi, V. and N.R. Swanson (2006c). Bootstrap Conditional Distribution in the Presence of Dynamic Misspecification. *Journal of Econometrics,* 133, 779-806.

Corradi, V. and N.R. Swanson (2007). Nonparametric Bootstrap Procedures for Predictive Inference Based on Recursive Estimation Schemes, *International Economic Review*, 48, 67-109.

Diebold F.X., and R.S. Mariano (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics,* 13, 253-263.

Dufour, J.-M. and D. Stevanovic (2011). Factor Augmented VARMA Models: Identification, Estimation, Forecasting and Impulse Responses. Working Paper, McGill University.

Fan, J., Y. Liao and M. Mincheva (2011). High Dimensional Covariance Matrix Estimation in Approximate Factor Models. Manuscript, Princeton University.

Giacomini R. and B. Rossi (2009). Detecting and Predicting Forecast Breakdowns. *Review of Economic Studies,* 76, 669-705.

Goncalves S. and B. Perron (2011). Bootstrapping Factor-Augmented Regression Models. Working Paper, University of Montreal.

Hendry D.F. and M.P. Clements (2002). Pooling of Forecasts. *Econometrics Journal.* 5, 1-26.

Han, X. and A. Inoue (2011). Tests for Parameter Instability in Dynamic Factor Models. Working Paper, North Carolina State Univesity.

Hendry D.F. and G. Mizon (2005). Forecasting in the Presence of Structural Breaks and Policy Regime Shifts in *Identification and Inference for Econometric Models: Essays in Honour of Thomas Rothemberg,* edited by D.W.K. Andrews and J.H. Stock, Cambridge University Press.

Kim H.H. and N.R. Swanson (2011a). Forecasting Financial and Macroeconomic Variables Using Data Reduction Methods: New Empirical Evidence. Working Paper, Rutgers University.

Kim H.H. and N.R. Swanson (2011b). Diffusion Indices Using Nonlinear Factor Methods. Working Paper, Rutgers University.

McCracken, M.W. (2007) Asymptotics for Out-of-Sample Tests for Granger Causality. *Journal of Econometrics,* 140, 719-752.

Onatski, Alexei (2009). Testing Hypotheses About the Number of Factors in Large Factor Models. *Econometrica,* 77, 1447-1479.

Pesaran, M.H. and A. Timmermann (2007). Selection in Estimation Window in the Presence of Breaks. *Journal of Econometrics,* 137, 134-161.

Stock, J.H. and M.W. Watson (2002a). Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business and Economic Statistics,* 20, 147-162.

Stock, J.H. and M.W. Watson (2002b). Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association,* 97, 1167-1179.

Stock, J.H. and M.W. Watson (2004). Combination Forecasts of Output Growth in a Seven-Countries Data-Set. *Journal of Forecasting,* 23, 405-430.

Stock, J.H. and M.W. Watson (2009). Forecasting in Dynamic Factor Models Subject to Structural Instability. *The Methodology and Practice of Econometrics: Festschrift in Honour of D.F. Hendry,* edited by J. Castle and N. Shephard. Oxford University Press.

West K.D. and M.W. McCracken (1998). Regression-Based Tests for Predictive Ability. *International Economic Review,* 39, 817-840.

White, H. (2000). A Reality for Data Snooping. *Econometrica,* 68, 1097-1126.

Wooldridge J.M. and H. White (1988). Some Invariance Principle and Central Limit Theorems for Dependent and Heterogeneous Processes. *Econometric Theory,* 4, 210-230.

Table 1: Target Forecasting Variables *

Series	Abbreviation	Y_{t+h}
Unemployment Rate	UR	$Z_{t+1} - Z_t$
Personal Income less transfer payments	PILT	$\ln(Z_{t+1}/Z_t)$
10-Year Treasury Bond	TB10Y	$Z_{t+1} - Z_t$
Consumer Price Index	CPI	$\ln(Z_{t+1}/Z_t)$
Producer Price Index	PPI	$\ln(Z_{t+1}/Z_t)$
Nonfarm Payroll Employment	NPE	$\ln(Z_{t+1}/Z_t)$
Housing Starts	HS	$\ln(Z_t)$
Industrial Production	IPX	$\ln(Z_{t+1}/Z_t)$
M2	M2	$\ln(Z_{t+1}/Z_t)$
S&P 500 Index	SNP	$\ln(Z_{t+1}/Z_t)$
Gross Domestic Product	GNP	$\ln(Z_{t+1}/Z_t)$

* Notes: The data used in our empirical illustration are monthly U.S. figures for the period 1960:1-2009:5. The transformation used in forecast model specification and forecast construction is given in the last column of the table.

Table 2: Empirical Setup - Samples Sizes and Various Parameter Settings *

a. Size of Rolling Windows - Statistic Calculations

	Case1	Case2
T	560	560
R	260	360
P	300	200

b. Size of Rolling Windows - Critical Value Calculations

Case1	m=0.5	m=0.8	m=0.9
T*	290	460	520
R*	140	220	250
P*	150	240	270

Case2	m=0.5	m=0.8	m=0.9
T*	260	450	500
R*	180	290	320
P*	80	160	180

c. Boostrap Parameter Settings

Permutation	b_{T^*}	b_{R^*}	b_{P^*}
1	10	5	5
2	10	2	2
3	10	10	10
4	5	2	2
5	5	5	5
6	2	2	2

* Notes: This table lists various parameter settings used in our empirical illustration. In some cases, sample sizes are rounded in order that bootstrap blocks not be truncated when forming bootstrap samples. For complete details, see Section 5.

Table 3: Stability Test Results for 11 Macroeconomic Variables - Quadratic Loss *

Results Are Tabulated for the Following Case: R = 360, P=200 $m = 0.5$ and $h = 1$

		UR	PI	TB10Y	CPI	PPI	NPE	HS	IPX	M2	SNP	GDP
Test Statistic		0.0002	0.0005	1.0958	0.1690	0.8130	0.2580	0.0247	0.0030	0.2200	0.2500	0.0271
$b_{T*} = 10$, $b_{R*} = b_{P*} = 5$	95%	0.8034	0.7262	1.3299	0.6669	0.4109*	2.7717	0.1261	2.5246	4.1167	0.5212	1.4919
	90%	0.5290	0.5341	0.7161*	0.4556	0.2811	1.9496	0.0909	1.8358	3.5638	0.4244	1.1106
	50%	0.1116	0.1048	0.1435	0.0803	0.0741	0.3051	0.0133	0.3991	1.8603	0.1112	0.2060
	P-value	0.9760	0.9580	0.0620	0.2940	0.0060	0.5360	0.3540	0.9560	0.9900	0.2620	0.8140
$b_{T*} = 10$, $b_{R*} = b_{P*} = 2$	95%	0.8871	0.7295	1.3702	0.5299	0.4237*	3.0189	0.1329	2.3651	3.6833	0.5687	1.4351
	90%	0.6741	0.5516	0.7196*	0.3887	0.3067	1.9173	0.0925	1.8296	3.2013	0.4415	0.9678
	50%	0.1168	0.1192	0.1346	0.0753	0.0753	0.3736	0.0164	0.3418	1.8157	0.0991	0.1830
	P-value	0.9780	0.9740	0.0600	0.3040	0.0040	0.5660	0.3980	0.9680	0.9880	0.2500	0.7980
$b_{T*} = 10$, $b_{R*} = b_{P*} = 10$	95%	0.6984	0.8243	2.0356	0.5676	0.3686*	2.7881	0.1324	2.7272	4.1229	0.5676	1.6297
	90%	0.5026	0.5836	0.8774*	0.3797	0.2944	2.0546	0.0897	1.8669	3.6062	0.4427	0.9770
	50%	0.0819	0.1250	0.1246	0.0685	0.0791	0.3877	0.0126	0.3747	2.0349	0.1108	0.2013
	P-value	0.9700	0.9600	0.0840	0.3020	0.0020	0.5620	0.3680	0.9680	1.0000	0.2860	0.8080
$b_{T*} = 5$, $b_{R*} = b_{P*} = 2$	95%	0.7253	0.7955	1.7473	0.5327	0.3931*	3.3354	0.1238	2.3776	3.9955	0.5448	1.5213
	90%	0.5220	0.5578	0.8543*	0.3526	0.3029	2.3553	0.0897	1.7434	3.5529	0.4271	1.0805
	50%	0.0785	0.1058	0.1453	0.0772	0.0828	0.2988	0.0111	0.3026	1.7921	0.1061	0.2268
	P-value	0.9820	0.9600	0.0800	0.3000	0.0020	0.5400	0.3620	0.9420	0.9860	0.2600	0.8240
$b_{T*} = 5$, $b_{R*} = b_{P*} = 5$	95%	0.6847	0.6964	1.5341	0.5818	0.4099*	2.4038	0.1364	2.3252	3.9578	0.5570	1.5522
	90%	0.5315	0.4994	0.7991*	0.4332	0.3298	1.8779	0.0932	1.7667	3.4933	0.4469	1.0452
	50%	0.1020	0.1266	0.1561	0.0897	0.0921	0.2943	0.0137	0.3755	1.9059	0.1176	0.2158
	P-value	0.9820	0.9720	0.0620	0.3240	0.0000	0.5260	0.3600	0.9340	0.9880	0.2700	0.8100
$b_{T*} = 2$, $b_{R*} = b_{P*} = 2$	95%	0.8117	0.7206	1.3097	0.6117	0.4274*	3.1398	0.1406	2.6367	4.0293	0.5428	1.6265
	90%	0.5783	0.5125	0.6698*	0.4451	0.2936	2.0869	0.0973	1.9786	3.4715	0.4073	1.0895
	50%	0.1020	0.1106	0.1286	0.0925	0.0771	0.3336	0.0176	0.3620	1.8662	0.1036	0.2079
	P-value	0.9760	0.9620	0.0600	0.3540	0.0040	0.5540	0.4320	0.9480	0.9920	0.2520	0.8360

* Entries in this table are given for (i) the test statistic (first row of numerical entries); (ii) the 95th, 90th, and 50th percentiles of the empirical bootstrap distribution (rows denoted by 95%, 90%, and 50%), for given values of b_{T^*}, b_{R^*}; and b_{P^*}; and (iii) the probability of rejection (p-value) under the null of forecast model stability, based on the empirical bootstrap distribution. For complete details, see Section 5.

Table 4: Stability Test Results for 11 Macroeconomic Variables - Quadratic Loss *

Results Are Tabulated for the Following Case: R = 360, P=200 $m = 0.5$ and $h = 3$

		UR	PI	TB10Y	CPI	PPI	NPE	HS	IPX	M2	SNP	GDP
Test Statistic		0.0003	0.6030	0.0603	0.4730	0.4900	0.4670	0.0967	0.3270	0.3020	0.0479	0.1440
$b_{T*} = 10$, $b_{R*} = b_{P*} = 5$	95%	0.8128	2.3840	0.2226	0.7575	0.2251*	2.1745	0.1187	2.8616	5.2609	2.4868	3.9066
	90%	0.5470	1.0546	0.1666	0.4215*	0.1540	1.3016	0.0653*	1.6194	4.4388	1.1576	2.6222
	50%	0.0920	0.1700	0.0379	0.0698	0.0317	0.2347	0.0087	0.2586	1.8612	0.0793	0.3194
	P-value	0.9780	0.1940	0.3620	0.0860	0.0100	0.3520	0.0640	0.4480	0.8980	0.5960	0.6540
$b_{T*} = 10$, $b_{R*} = b_{P*} = 2$	95%	0.9370	2.5989	0.2494	0.5668	0.2306*	2.0640	0.1692	2.8054	5.1962	2.3060	2.8934
	90%	0.6702	1.0468	0.1646	0.3650*	0.1672	1.5475	0.0841*	1.7164	4.2748	1.1078	1.6962
	50%	0.1303	0.1743	0.0309	0.0727	0.0290	0.2460	0.0099	0.3077	1.6455	0.0787	0.2490
	P-value	0.9700	0.1760	0.3460	0.0680	0.0120	0.3560	0.0880	0.4900	0.8860	0.5940	0.6080
$b_{T*} = 10$, $b_{R*} = b_{P*} = 10$	95%	0.9025	1.1364	0.2126	0.6903	0.2463*	2.1654	0.1246	2.8938	5.4660	2.1186	3.4521
	90%	0.5959	0.6513	0.1554	0.3440*	0.1502	1.5676	0.0643*	1.6593	4.3775	0.8113	2.0482
	50%	0.0940	0.1274	0.0243	0.0519	0.0252	0.2591	0.0098	0.2874	2.0113	0.0653	0.2822
	P-value	0.9740	0.1180	0.3060	0.0780	0.0080	0.3680	0.0680	0.4800	0.9340	0.5540	0.6320
$b_{T*} = 5$, $b_{R*} = b_{P*} = 2$	95%	0.7943	1.5141	0.2003	0.5538	0.2943*	2.0816	0.1354	3.0554	5.1235	2.7141	3.7543
	90%	0.5960	0.9177	0.1439	0.3620*	0.1771	1.6099	0.0837*	2.0773	4.4331	1.6926	2.2328
	50%	0.1035	0.1361	0.0288	0.0647	0.0283	0.2905	0.0087	0.2741	1.8168	0.0799	0.3165
	P-value	0.9800	0.1540	0.3120	0.0700	0.0260	0.3820	0.0860	0.4720	0.8940	0.6060	0.6420
$b_{T*} = 5$, $b_{R*} = b_{P*} = 5$	95%	0.9270	1.6421	0.2453	0.6327	0.2290*	2.0237	0.1164	2.4935	4.9852	2.1119	3.5494
	90%	0.6300	0.9869	0.1758	0.3762*	0.1419	1.3333	0.0683*	1.7219	4.1573	0.9152	2.0072
	50%	0.1058	0.1339	0.0322	0.0580	0.0275	0.2572	0.0081	0.2391	1.7096	0.0783	0.3187
	P-value	0.9820	0.1780	0.3460	0.0740	0.0160	0.3400	0.0680	0.4420	0.8920	0.5960	0.6560
$b_{T*} = 2$, $b_{R*} = b_{P*} = 2$	95%	0.7992	1.3753	0.2239	0.8336	0.2473*	1.8222	0.1281	2.6364	5.2709	1.9217	3.8909
	90%	0.6186	0.8650	0.1634	0.4373*	0.1543	1.4144	0.0883*	1.7826	4.3729	0.7872	2.3220
	50%	0.0801	0.1448	0.0286	0.0604	0.0302	0.2595	0.0109	0.2849	1.8905	0.1037	0.2601
	P-value	0.9700	0.1460	0.3060	0.0920	0.0100	0.3500	0.0880	0.4620	0.9000	0.6580	0.5900

* See notes to Table 3.

Table 5: Stability Test Results for 11 Macroeconomic Variables - Linex Loss *

Results Are Tabulated for the Following Case: R = 360, P=200 m = 0.5 and h = 1

		UR	PI	TB10Y	CPI	PPI	NPE	HS	IPX	M2	SNP	GDP
Test Statistic		0.0007	0.0001	0.5583	0.0431	0.2090	0.0647	0.0036	0.0007	0.0551	0.0694	0.0070
$b_{T*} = 10$, $b_{R*} = b_{P*} = 5$	95%	0.2328	0.1800	0.4931*	0.1678	0.1022*	0.6929	0.0404	0.6299	1.0218	0.1266	0.3729
	90%	0.1544	0.1327	0.2781	0.1137	0.0699	0.4889	0.0276	0.4655	0.8862	0.1034	0.2766
	50%	0.0292	0.0260	0.0520	0.0201	0.0184	0.0764	0.0039	0.0998	0.4630	0.0289	0.0514
	P-value	0.9240	0.9600	0.0460	0.2920	0.0040	0.5360	0.5180	0.9580	0.9900	0.2140	0.8140
$b_{T*} = 10$, $b_{R*} = b_{P*} = 2$	95%	0.2781	0.1814	0.5376*	0.1331	0.1055*	0.7572	0.0359	0.5934	0.9149	0.1402	0.3583
	90%	0.2040	0.1374	0.2589	0.0968	0.0762	0.4770	0.0273	0.4646	0.7946	0.1077	0.2409
	50%	0.0285	0.0297	0.0487	0.0190	0.0186	0.0934	0.0044	0.0859	0.4515	0.0250	0.0461
	P-value	0.9040	0.9760	0.0460	0.3020	0.0040	0.5680	0.5360	0.9680	0.9860	0.2120	0.7900
$b_{T*} = 10$, $b_{R*} = b_{P*} = 10$	95%	0.2143	0.2047	0.8738	0.1430	0.0918*	0.6928	0.0383	0.6858	1.0260	0.1386	0.4057
	90%	0.1490	0.1455	0.3149*	0.0950	0.0730	0.5160	0.0240	0.4685	0.8963	0.1069	0.2439
	50%	0.0222	0.0309	0.0448	0.0171	0.0196	0.0967	0.0036	0.0937	0.5065	0.0281	0.0502
	P-value	0.9220	0.9640	0.0680	0.2980	0.0020	0.5600	0.5040	0.9680	1.0000	0.2300	0.8020
$b_{T*} = 5$, $b_{R*} = b_{P*} = 2$	95%	0.2369	0.1982	0.6974	0.1339	0.0977*	0.8333	0.0397	0.5918	0.9941	0.1345	0.3798
	90%	0.1551	0.1388	0.3373*	0.0884	0.0754	0.5895	0.0264	0.4394	0.8807	0.1033	0.2705
	50%	0.0214	0.0262	0.0491	0.0190	0.0204	0.0750	0.0034	0.0755	0.4454	0.0266	0.0572
	P-value	0.9120	0.9640	0.0580	0.3000	0.0020	0.5400	0.4900	0.9480	0.9860	0.2220	0.8200
$b_{T*} = 5$, $b_{R*} = b_{P*} = 5$	95%	0.2197	0.1729	0.5386*	0.1465	0.1021*	0.6014	0.0413	0.5845	0.9845	0.1395	0.3858
	90%	0.1688	0.1239	0.2847	0.1085	0.0820	0.4670	0.0275	0.4433	0.8680	0.1153	0.2604
	50%	0.0270	0.0313	0.0547	0.0225	0.0230	0.0737	0.0042	0.0917	0.4737	0.0283	0.0541
	P-value	0.9160	0.9740	0.0500	0.3100	0.0000	0.5280	0.5300	0.9340	0.9880	0.2340	0.8080
$b_{T*} = 2$, $b_{R*} = b_{P*} = 2$	95%	0.2665	0.1790	0.4782*	0.1539	0.1064*	0.7877	0.0416	0.6600	1.0023	0.1335	0.4062
	90%	0.1703	0.1272	0.2619	0.1120	0.0733	0.5183	0.0261	0.4880	0.8631	0.1005	0.2713
	50%	0.0282	0.0274	0.0438	0.0233	0.0192	0.0831	0.0047	0.0909	0.4645	0.0264	0.0518
	P-value	0.9300	0.9660	0.0440	0.3460	0.0020	0.5540	0.5620	0.9500	0.9920	0.2160	0.8360

* See notes to Table 3.

Table 6: Stability Test Results for 11 Macroeconomic Variables - Linex Loss *

Results Are Tabulated for the Following Case: R = 360, P=200 m = 0.5 and h = 3

		UR	PI	TB10Y	CPI	PPI	NPE	HS	IPX	M2	SNP	GDP
Test Statistic		0.0005	0.1520	0.0259	0.1190	0.1250	0.1170	0.0147	0.0810	0.0754	0.0108	0.0361
$b_{T*} = 10$, $b_{R*} = b_{P*} = 5$	95%	0.2779	0.5898	0.0546	0.1900	0.0563*	0.5421	0.0204	0.7186	1.3091	0.5728	0.9664
	90%	0.1800	0.2609	0.0380	0.1062*	0.0386	0.3254	0.0124*	0.4046	1.1042	0.2660	0.6509
	50%	0.0270	0.0428	0.0081	0.0173	0.0080	0.0594	0.0019	0.0653	0.4641	0.0191	0.0804
	P-value	0.9180	0.1900	0.1800	0.0860	0.0080	0.3520	0.0820	0.4500	0.8980	0.6060	0.6540
$b_{T*} = 10$, $b_{R*} = b_{P*} = 2$	95%	0.3160	0.6442	0.0554	0.1425	0.0581*	0.5192	0.0315	0.6936	1.2944	0.5416	0.7161
	90%	0.2131	0.2603	0.0369	0.0907*	0.0419	0.3883	0.0158	0.4313	1.0650	0.2514	0.4204
	50%	0.0405	0.0433	0.0068	0.0183	0.0071	0.0785	0.0021	0.0787	0.4102	0.0172	0.0613
	P-value	0.9400	0.1700	0.1700	0.0680	0.0100	0.3840	0.1120	0.4900	0.8860	0.6040	0.6080
$b_{T*} = 10$, $b_{R*} = b_{P*} = 10$	95%	0.3020	0.2807	0.0453	0.1737	0.0621*	0.5884	0.0245	0.7139	1.3612	0.4899	0.8575
	90%	0.1986	0.1612	0.0351	0.0866*	0.0377	0.3864	0.0139*	0.4162	1.0896	0.1792	0.5073
	50%	0.0312	0.0314	0.0059	0.0131	0.0063	0.0789	0.0021	0.0716	0.5008	0.0157	0.0710
	P-value	0.9220	0.1120	0.1480	0.0780	0.0060	0.3800	0.0940	0.4760	0.9340	0.5700	0.6320
$b_{T*} = 5$, $b_{R*} = b_{P*} = 2$	95%	0.2872	0.3743	0.0414	0.1393	0.0739*	0.4862	0.0246	0.7454	1.2763	0.6122	0.9261
	90%	0.2041	0.2284	0.0312	0.0904*	0.0442	0.3616	0.0160	0.5231	1.1041	0.3796	0.5537
	50%	0.0313	0.0341	0.0062	0.0161	0.0071	0.0540	0.0016	0.0700	0.4524	0.0205	0.0784
	P-value	0.9300	0.1540	0.1380	0.0700	0.0260	0.3340	0.1120	0.4720	0.8940	0.6040	0.6400
$b_{T*} = 5$, $b_{R*} = b_{P*} = 5$	95%	0.3047	0.4062	0.0558	0.1589	0.0575*	0.5053	0.0236	0.6067	1.2417	0.4874	0.8765
	90%	0.2178	0.2450	0.0413	0.0942*	0.0356	0.3330	0.0129*	0.4295	1.0352	0.2035	0.4956
	50%	0.0326	0.0333	0.0072	0.0146	0.0070	0.0641	0.0016	0.0619	0.4248	0.0190	0.0791
	P-value	0.9380	0.1760	0.1780	0.0740	0.0160	0.3380	0.0860	0.4460	0.8920	0.5940	0.6540
$b_{T*} = 2$, $b_{R*} = b_{P*} = 2$	95%	0.2761	0.3395	0.0530	0.2097	0.0622*	0.4546	0.0232	0.6549	1.3131	0.4395	0.9659
	90%	0.1954	0.2147	0.0353	0.1094*	0.0388	0.3539	0.0160	0.4385	1.0885	0.1759	0.5736
	50%	0.0261	0.0359	0.0064	0.0151	0.0075	0.0653	0.0022	0.0699	0.4712	0.0250	0.0639
	P-value	0.9180	0.1440	0.1580	0.0920	0.0100	0.3500	0.1160	0.4640	0.9000	0.6480	0.5900

* See notes to Table 3.