Inference for Factor-MIDAS Regression Models

Yookyung Julia Koh
McGill University

November 10, 2023

Latest version

Abstract

Factor-MIDAS regression models are often used to forecast a target variable using common factors extracted from a large panel of predictors observed at higher frequencies. In the paper, we derive the asymptotic distribution of the factor-MIDAS regression estimator coefficients. We show that there exists an asymptotic bias because the factors are estimated. However, the fact that factors and their lags are aggregated in a MIDAS regression model implies that the asymptotic bias depends on both serial and cross-sectional dependence in the idiosyncratic errors of the factor model. Thus, bias correction is more complicated in this setting. Our second contribution is to propose a bias correction method based on a plug-in version of the analytical formula we derive. This bias correction can be used in conjunction with asymptotic normal critical values to produce asymptotically valid inference. Alternatively, we can use a bootstrap method, which is our third contribution. We show that correcting for bias is important in simulations and in an empirical application to forecasting quarterly U.S. real GDP growth rates using monthly factors.

1 Introduction

MIDAS (Mixed-Data Sampling) regressions are popular tools in forecasting. Originally proposed by Ghysels et al. (2004; 2005; 2006; 2007), these models combine predictors observed at high frequencies by relying on a parametric temporal aggregation function to forecast a target variable sampled at a lower frequency. Originally proposed to handle financial variables, they have become standard tools in macroeconomic forecasting (see e.g., Clements and Galvão (2008; 2009), which relies on MIDAS autoregressions for nowcasting U.S. real output growth).

More recently, standard MIDAS regressions have been generalized to "factor-MIDAS regressions" (or "factor-augmented MIDAS regression models") by including as predictors common factors extracted from a large panel of time series sampled at a higher frequency than the target variable. By combining with the dimension reduction properties of factor models, factor-MIDAS regressions are powerful tools for forecasting and they are often used in empirical applications (see for instance Marcellino and Schumacher (2010), Monteforte and Moretti (2013), Kim and Swanson (2018), and Ferrara and Marsilli (2019)). Estimation of factor-MIDAS regressions is complicated by the fact that some of the predictors are latent common factors. It typically proceeds in two steps: we first extract the common factors using principal component analysis, and then estimate the model using nonlinear least squares, where the estimated factors are aggregated by a temporal aggregation scheme.

Although factor-MIDAS regressions are empirically popular, no formal inference methods have been proposed in the literature. The paper proposes inference methods for factor-MIDAS regression models and provides the theoretical justification for these methods. The main contributions of this paper are as follows. Firstly, the asymptotic distribution of the factor-MIDAS regression estimators is derived. We show that there is an asymptotic bias in the second step due to the estimation of the factors in the first step. Secondly, we propose two inference methods accounting for this bias: a bias correction method based on the bias formula we derive and a bootstrap method.

Our work is related to the existing literature on factor-augmented regression models (without mixed frequencies). Bai and Ng (2006) first studied the "generated regressor" problem in standard factor-augmented regression models. They showed that inference for the regression coefficients could proceed as if the estimated factors were observed if the cross-sectional dimension N was sufficiently large relative to the time dimension T, more precisely if $\sqrt{T}/N \to 0$. More recently, Gonçalves and Perron (2014) (henceforth, GP (2014)) showed that an asymptotic bias may appear under more relaxed assumption (i.e. if $\sqrt{T}/N \to c,\ 0 < c < \infty$). We extend these results to factor-MIDAS regression models. This is not a trivial extension for two main reasons. First, the estimation problem in a factor-MIDAS regression model is more complicated because the predictors include latent factors (and their lags) sampled at a different frequency than a variable of interest. In addition, the second step is based on nonlinear least squares (rather than OLS) because of a temporal aggregation, and this complicates the asymptotic analysis. In particular, whereas the bias derived in Gonçalves and Perron (2014) depends only on the cross-sectional dependence, the asymptotic bias of a factor-MIDAS regression model depends on both serial and cross-sectional dependence in the idiosyncratic errors. Consequently, different methods of inference are required for factor-MIDAS regressions.

We consider two different methods of inference in this context. The first is an analytical bias correction that can be used along with asymptotic normal critical values. Our plug-in bias correction is robust to both serial and cross-sectional dependence of unknown form in the idiosyncratic errors. It is based on the asymptotic formula of the bias we derive, replacing unknown parameters with consistent estimators. As in Ludvigson and Ng (2009), who also proposed a bias correction formula for the standard factor-augmented regression model without mixed frequencies, we rely on the CS-HAC estimator of Bai and Ng (2006) to correct for cross-sectional dependence. However, our estimator is more complex since it also requires robustness to serial dependence.

Our second method of inference is based on the bootstrap. The bootstrap has two significant advantages: it can perform better in finite samples, and it avoids the explicit estimation of the bias term which can be complicated in this context. We propose a bootstrap procedure inspired by Gonçalves and Perron (2014), which is a residual-based bootstrap. Although the method is inspired by Gonçalves and Perron (2014), the asymptotic justification is substantially more complicated. More importantly, the need to mimic the asymptotic bias requires the bootstrap to be robust to both serial and cross-sectional dependence. Since none of the existing bootstrap methods in the literature allows for both forms of dependence, we propose a new bootstrap method for factor models that has these properties. Our method is based on an application of the sieve bootstrap to the idiosyncratic residuals of each time series in the panel data model, where the corresponding innovations are resampled using the cross-sectional dependent bootstrap proposed by Gonçalves and

Perron (2020). We show that this bootstrap method is asymptotically valid when each idiosyncratic error in the factor model is generated by an $AR(\infty)$ process with innovations that are potentially cross-sectionally correlated across the panel. A special case of this new bootstrap method is considered by Gonçalves, Koh, and Perron (2023) when testing for the number of common factors in group factor models (as proposed by Andreou, Gagliardini, Ghysels, and Rubin (2019)).

We illustrate the good finite sample performance of the plug-in bias estimator and the bootstrap using Monte Carlo simulations. In particular, the results show that it is important to correct the bias due to the estimation of the factors in the first step. Although both the plug-in bias correction and the bootstrap methods replicate the bias well, the bootstrap outperforms the plug-in bias estimator by further reducing the coverage rate distortions. Finally, we apply our new inference methods to an empirical application where we nowcast quarterly U.S. real GDP growth rate using monthly macroeconomic factors. The results show that there is a significant bias, thereby indicating the importance of correcting it.

The rest of this paper is organized as follows. In Section 2, we derive the asymptotic distribution of the factor-augmented MIDAS regression model and propose a plug-in bias estimator. In Section 3, we propose and theoretically justify the bootstrap. The simulation results are shown in Section 4, and the empirical application is discussed in Section 5. Section 6 concludes the paper. Additionally, we include three mathematical appendices: Appendix A delivers the primitive assumptions necessary for proving the results in the paper and Appendices B - C shows the proof of the results in Sections 2 -3, respectively.

2 Asymptotic Theory

2.1 Factor-augmented MIDAS regression models

The MIDAS regression model projects high-frequency variables onto a target variable, which is denoted as y_t. The regressors are observed at most m times between t and $t-1$. To handle variables sampled at mixed frequency, a MIDAS regression aggregates the high-frequency variables with a lag polynomial function. The basic MIDAS regression model with a single observed regressor x_t can be written as follows:

$$y_t = \beta_0 + \beta_1 W(L^{1/m};\theta)x_t + \varepsilon_t, \quad t = 1, \ldots, T, \tag{1}$$

where $W(L^{1/m};\theta) = \sum_{k=1}^{K} w_k(\theta) L^{k/m}$ and $L^{k/m} x_t = x_{t-k/m}$. Here, $w_k(\theta)$ is a weighting function that temporally aggregates the regressor and its lags, and θ is a $p \times 1$ vector of weighting parameters. To identify β_1, we assume that $w_k(\theta) \in (0,1)$ and $\sum_{k=1}^{K} w_k(\theta) = 1$. A common weighting scheme in the MIDAS regression model is the exponential Almon lag with two parameters such that

$$w_k(\theta) = \frac{\exp(\theta_1 k + \theta_2 k^2)}{\sum_{k=1}^{K} \exp(\theta_1 k + \theta_2 k^2)}. \tag{2}$$

Other weighting schemes include the beta function and the linear function. Details can be found in Ghysels, Valkanov, and Serrano (2009). Although the high-frequency variable is used in the regression in a linear manner, the estimation of the parameters is done through a nonlinear estimation method as the MIDAS regression itself is a nonlinear function of the parameters.

In this paper, we consider the factor-MIDAS regression model, which employs unobserved high-frequency factors as regressors. In particular, letting the regressor x_t in (1) be replaced by a latent factor, we write the model as follows:

$$y_t = \beta_0 + \beta_1 W(L^{1/m};\theta) f_t + \varepsilon_t = \beta_0 + \beta_1 \sum_{k=1}^{K} w_k(\theta) f_{t-k/m} + \varepsilon_t, \quad t = 1, \ldots, T,$$

where $f_{t-k/m}$ is a (single) factor in the following panel factor model,

$$X_{t-k/m} = \Lambda f_{t-k/m} + e_{t-k/m}, \quad k = m-1, \ldots, 0, \text{ and } t = 1, \ldots, T. \tag{3}$$

The factor model includes factor loadings denoted by Λ and an idiosyncratic error term, $e_{t-k/m}$. If there are r unobserved factors, represented by a $r \times 1$ vector of common factors denoted by $f_{t-k/m}$ in the factor model (3), then the model can be generalized as follows:

$$y_t = \beta_0 + \beta_1' W(L^{1/m};\theta) f_t + \varepsilon_t = \beta_0 + \beta_1' F_t(\theta) + \varepsilon_t, \quad t = 1, \ldots, T, \tag{4}$$

where $\beta_1 = (\beta_{1,1}, \ldots, \beta_{1,r})'$, and $\theta = (\theta_1', \ldots, \theta_r')'$ with $\theta_j = (\theta_{j,1}, \ldots, \theta_{j,p})'$, a $p \times 1$ weighting parameter for j-th factor, for $j = 1, \ldots, r$. We define $F_t(\theta) \equiv W(L^{1/m};\theta) f_t$ in the second equality. In fact, the temporal aggregation in this generalized model applies on a vector as

$$F_t(\theta) = \sum_{k=1}^{K} w_k(\theta) L^{k/m} f_t = \sum_{k=1}^{K} w_k(\theta) f_{t-k/m},$$

where $w_k(\theta)$ is a $r \times r$ diagonal matrix such that $w_k(\theta) \equiv \text{diag}(w_{k,1}(\theta_1), \ldots, w_{k,r}(\theta_r))$, where $w_{k,j}(\theta_j)$

is the weight for the k-th lag of the j-th factor.[1] To derive the distribution in the next section, we further simplify the general factor-MIDAS regression model (4) to

$$y_t = g(F_t, \alpha) + \varepsilon_t, \quad t = 1, \dots, T, \tag{5}$$

where $g(F_t, \alpha) = \beta_0 + \beta_1' F_t(\theta)$, $\alpha = (\beta', \theta')'$ with $\beta = (\beta_0, \beta_1')'$, and $F_t = (1, f_t', f_{t-1/m}', \dots, f_{t-K/m}')'$. For convenience, we use the high frequency time index denoted by $t_h = 1, \dots, T_H$, where $T_H = mT$. We derive this by noting that $t_h = m((t-1) + i/m)$ for $i = 1, \dots, m$, and $t = 1, \dots, T$.[2] Using this notation, we can write the factor model as $X_{t_h} = \Lambda f_{t_h} + e_{t_h}$, for $t_h = 1, \dots, T_H$. Using the matrix notation, we write the factor model as $X = f\Lambda' + e$, where X is a $T_H \times N$ matrix of high-frequency time series, $f = (f_1, \dots, f_{T_H})'$ is a $T_H \times r$ matrix of common factors, and e is a $T_H \times N$ matrix of idiosyncratic errors.

2.2 Asymptotic Theory

We denote NLS estimators by $\hat{\alpha}$ when the factors are observed. Then, we could show that the limiting distribution of $\hat{\alpha}$ is:

$$\sqrt{T}(\hat{\alpha} - \alpha_0) \xrightarrow{d} N(0, \Sigma^{-1}\Omega\Sigma^{-1}), \tag{6}$$

where $\alpha_0 = (\beta', \theta')'$, $\Sigma = E[g_{\alpha,t} g_{\alpha,t}']$, and $\Omega = E[\varepsilon_t^2 g_{\alpha,t} g_{\alpha,t}']$ with $g_{\alpha,t} = \partial g(F_t, \alpha)/\partial \alpha$. When the true factors are observed, the estimators are normally distributed with mean zero and a sandwich variance.

In factor-MIDAS models, however, the factors are latent, and we have to estimate them. Accordingly, the estimation in the factor-MIDAS regression model proceeds in two steps. First, we estimate the common factors from a panel dataset of high-frequency indicators by principal component analysis (PCA). The estimated factors, $\tilde{f}$, are equivalent to $\sqrt{T_H}$ times the eigenvectors of $XX'/T_H N$ corresponding to the r largest eigenvalues (in decreasing order). The estimated factor loadings are $\tilde{\Lambda} = X'\tilde{f}/T_H$.[3] Second, we estimate the parameters, β and θ using nonlinear least

[1] Note that when $m = 1$ and $K = 0$, the factor-MIDAS regression model is equivalent to the standard factor-augmented regression model in GP (2014).

[2] With this notation, a high-frequency observation at t_h is equivalent to observing it at the i-th intra-period between $t-1$ and t. Note that the time notation in the factor model (3) can be written as $(t-1) + (m-k)/m$.

[3] When $T_H > N$, we use normalization such that $\Lambda'\Lambda/N = I_r$ and $f'f$ is a diagonal matrix, which is computationally easier. In this case, $\tilde{\Lambda}$ is the matrix of $\sqrt{N}$ times the eigenvectors of $X'X/T_H N$ corresponding to the r largest eigenvalues and the estimated factors are $\tilde{f} = X\tilde{\Lambda}/N$.

squares (NLS) by regressing the low frequency variable on the temporally aggregated estimated factors at high-frequency. In the factor model, the estimated factors $\tilde{f}_t$ are only consistent for Hf_t, where the rotation matrix H is defined as $H = \tilde{V}^{-1}\frac{\tilde{f}'f}{T_H}\frac{\Lambda'\Lambda}{N}$, and $\tilde{V}$ is a $r \times r$ diagonal matrix of eigenvalues of XX'/T_HN in a descending order (for more details, see Bai (2003)). By incorporating the estimated factors in the regression and noting the rotation of the factors, we can rewrite (4) as follows:

$$y_t = \beta_0 + \beta_1' H^{-1}\tilde{F}_t(\theta) + \beta_1' H^{-1}(HF_t(\theta) - \tilde{F}_t(\theta)) + \varepsilon_t = g(\tilde{F}_t, \alpha) + \xi_t, \tag{7}$$

where $g(\tilde{F}_t, \alpha) = \beta_0 + \beta_1' H^{-1}\tilde{F}_t(\theta)$, $\alpha = (\beta_0, \beta_1' H^{-1}, \theta')'$, and $\tilde{F}_t(\theta) = \sum_{k=1}^{K} w_k(\theta)\tilde{f}_{t-k/m}$. The coefficient on the aggregated factors estimates $\beta_1' H^{-1}$. Moreover, the estimation error of the factors implies that the regression error term is $\xi_t = \beta_1' H^{-1}(HF_t(\theta) - \tilde{F}_t(\theta)) + \varepsilon_t$. We denote the NLS estimators of α in (7) by $\tilde{\alpha} = (\tilde{\beta}', \tilde{\theta}')'$ to distinguish from $\hat{\alpha} = (\hat{\beta}', \hat{\theta}')'$, which are the estimators from the regression of y_t on the true factors f_t. Next, we derive the limiting distribution of $\sqrt{T}(\tilde{\alpha} - \alpha)$ under the assumption that $\sqrt{T}/N \to c$, where $0 \leq c < \infty$.

The asymptotic distribution of the estimators is derived under the Assumptions A.1 - A.6 in Appendix A. We also introduce the following notations: $V \equiv \text{plim}\, \tilde{V}$, $Q \equiv \text{plim}\left(\frac{\tilde{f}'f}{T_H}\right)$, $Q_k \equiv \text{plim} \sum_{t_h=k+1}^{T_H} \tilde{f}_{t_h}' f_{t_h-k}$, and $\Sigma_{\tilde{f}} \equiv V^{-1}Q\Gamma Q'V^{-1}$, which is the asymptotic variance of $\sqrt{N}(\tilde{f}_{t_h} - Hf_{t_h})$.[4] The asymptotic variance of the factor estimation error is a function of Γ, which is defined by $\Gamma \equiv \lim_{N\to\infty} Var\left(\frac{\Lambda' e_{t_h}}{\sqrt{N}}\right)$. We assume that the idiosyncratic errors in the factor model, e_{t_h} is stationary in Assumption A.2-(d). Under the stationarity of the idiosyncratic errors, we also denote $\Gamma_k \equiv \lim_{N\to\infty} Cov\left(\frac{\Lambda' e_{t_h-k}}{\sqrt{N}}, \frac{\Lambda' e_{t_h}}{\sqrt{N}}\right)$. Note that by the identification assumption, Assumption A.1-(d) in Appendix A, we have $Q = H_0$, where $H_0 = \text{plim}\, H$, and H_0 is a diagonal matrix of ± 1, where the sign is determined by the sign of $\tilde{f}'f/T_H$ (for the detail of the proof, see the proof of (2) in Bai and Ng (2013)). Therefore, the asymptotic variance can be also written as $\Sigma_{\tilde{f}} = V^{-1}H_0\Gamma H_0' V^{-1}$.

Theorem 2.1 (Asymptotic distribution of the estimators in the factor-MIDAS models) *If $\sqrt{T}/N \to c$, where $0 \leq c < \infty$, and the Assumptions A.1 - A.6 in Appendix A hold,*

$$\sqrt{T}(\tilde{\alpha} - \alpha) \xrightarrow{d} N(-c\Delta_\alpha, \Sigma_\alpha), \tag{8}$$

[4] For the details, see Bai (2003).

where $\Sigma_\alpha \equiv \Phi_0'^{-1}\Sigma^{-1}\Omega\Sigma^{-1}\Phi_0^{-1}$ *with* $\Phi_0 = diag(1, H_0, I_p)$, *and*

$$\Delta_\alpha = \begin{bmatrix} \Delta_\beta \\ \Delta_\theta \end{bmatrix} = (\Phi_0 \Sigma \Phi_0')^{-1} \begin{bmatrix} B_\beta \\ B_\theta \end{bmatrix}. \tag{9}$$

$B_\beta = (B_{\beta_0}, B_{\beta_1}')'$ *and* B_θ *are such that* $B_{\beta_0} = 0$,

$$\begin{aligned} B_{\beta_1} = \Big[& \sum_{k=1}^{K} w_k(\theta) \Big\{ \Sigma_{\tilde{f}} + V \Sigma_{\tilde{f}} V^{-1} \Big\} w_k(\theta) \\ & + \sum_{k=1}^{K} \sum_{l \neq k}^{K} w_k(\theta) \Big\{ V^{-1} H_0 \Gamma_{k-l} H_0' V^{-1} + Q_{k-l} \Gamma H_0' V^{-2} \Big\} w_l(\theta) \Big] \operatorname{plim}(\tilde{\beta}_1), \end{aligned} \tag{10}$$

and

$$\begin{aligned} B_\theta = \operatorname{plim}(\tilde{\beta}_1) \circ \Big[& \sum_{k=1}^{K} \frac{\partial w_k(\theta)}{\partial \theta} \Big\{ \Sigma_{\tilde{f}} + V \Sigma_{\tilde{f}} V^{-1} \Big\} w_k(\theta) \\ & + \sum_{k=1}^{K} \sum_{l \neq k}^{K} \frac{\partial w_k(\theta)}{\partial \theta} \Big\{ V^{-1} H_0 \Gamma_{k-l} H_0' V^{-1} + Q_{k-l} \Gamma H_0' V^{-2} \Big\} w_l(\theta) \Big] \operatorname{plim}(\tilde{\beta}_1), \end{aligned} \tag{11}$$

where $\frac{\partial w_k(\theta)}{\partial \theta} \equiv \operatorname{diag}\left(\frac{\partial w_{k,1}(\theta_1)}{\partial \theta_1}, \ldots, \frac{\partial w_{k,r}(\theta_r)}{\partial \theta_r} \right)$ *is a block diagonal matrix and the* j*-th diagonal block is a* $p \times 1$ *vector given by* $\frac{\partial w_{k,j}(\theta_j)}{\partial \theta_j}$ *for* $j = 1, \ldots, r$.

In (11) in Theorem 2.1, we use the Hadamard product which is equivalent to $(A \circ B)_{ij} = A_{ij} B_{ij}$. More specifically, $\beta \circ \frac{\partial w_k(\theta)}{\partial \theta}$ is a block diagonal matrix where the j-th diagonal block contains $\beta_j \frac{\partial w_{j,k}(\theta_j)}{\partial \theta_j}$ for $j = 1, \ldots, r$. Based on Theorem 2.1, the bias of the estimators is proportional to c, the limiting value of $\sqrt{T}/N$, and also to $\operatorname{plim}(\tilde{\beta}_1) = (H^{-1})'\beta_1$. This implies that the estimates are biased unless $\beta_1 = 0$ or $c = 0$. Additionally, the asymptotic variance of the estimated factors, $\Sigma_{\tilde{f}}$, affects the bias. Since the variance of the factor estimation error depends on Γ, which is a variance of the scaled average of the factor loadings and the idiosyncratic errors in the factor model, the cross-sectional dependence of factor errors matters. These findings are similar to the bias in the context of GP (2014).

It is important to highlight some differences between our results and GP (2014). Firstly, the bias in the MIDAS regression model depends on the weighting scheme, $w_k(\theta)$, due to a temporal aggregation. Secondly, there exists a bias in the weighting parameters, θ. The bias in θ is similar to the bias in the slope coefficient, β_1. However, it differs in that the bias in the weighting parameters depends on the derivative of the weighting scheme and quadratic form of the slope coefficient. This occurs because MIDAS regressions are a nonlinear function of the weighting parameters.

Finally, in (10) and (11), both biases depend on the covariance of the cross-sectional average of factor loadings and the idiosyncratic error terms between two distinct periods, represented as Γ_{k-l}. This term arises due to the presence of the lags of the estimated factors. More specifically, as we include the lags of the estimated factors, we have an extra term such that $\frac{1}{T_H-k}\sum_{t_h=k+1}^{T_H} Cov(\sqrt{N}(\tilde{f}_{t_h}-H\tilde{f}_{t_h}), \sqrt{N}(\tilde{f}_{t_h-k}-H\tilde{f}_{t_h-k}))$ for $k \neq 0$. Since this is a function of Γ_k, the bias in our context relies on the serial dependence as well as cross-sectional dependence of the idiosyncratic error term in the factor model. This implies that the bias will depend on serial and cross-sectional dependence in the factor-MIDAS regression models without temporal aggregation.[5]

2.3 Plug-in Bias

In this section, we propose an analytical estimator to account for the bias identified in Theorem 2.1. In the context of the factor-augmented regression model, Ludvigson and Ng (2009) proposed a plug-in bias estimator by replacing the unknown quantities with their consistent estimators and correcting the bias. Similarly, we propose a bias-corrected estimator for factor-augmented MIDAS regression models.

In order to do that, we need a consistent estimator for the term Γ_k. This term has not been explored in previous literature and it depends on the cross-sectional and the serial dependence of the idiosyncratic error term. When the idiosyncratic error term is serially but not cross-sectionally correlated, we can estimate this term as $\tilde{\Gamma}_k = \frac{1}{N(T_H-k)}\sum_{t_h=k+1}^{T_H}\sum_{i=1}^{N}\tilde{\lambda}_i\tilde{\lambda}_i'\tilde{e}_{i,t_h}\tilde{e}_{i,t_h-k}$, where $\tilde{\Gamma}_k$ denotes the estimator of Γ_k. However, when the idiosyncratic error term is cross-sectionally and serially dependent, estimating this term is no longer straightforward, as discussed in Bai and Ng (2006). To address this issue, Bai and Ng (2006) propose an estimator for the variance-covariance matrix of the cross-sectional average of factor loadings and the idiosyncratic error term, denoted by Γ. They use the time series observations and truncation with $n < N$ under the covariance stationarity such that $\tilde{\Gamma}_{\text{CS-HAC}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{\lambda}_i\tilde{\lambda}_j\frac{1}{T_H}\sum_{t_h=1}^{T_H}\tilde{e}_{i,t_h}\tilde{e}_{j,t_h}$.

To propose a method to estimate Γ_k that takes into account cross-sectional and serial dependence, we take an approach, similar to the one used in Bai and Ng (2006). We use the time series

[5]When there is no temporal aggregation, the MIDAS regression becomes unrestricted MIDAS (U-MIDAS) proposed by Foroni, Marcellino, and Schumacher (2015).

observations and a truncation method, that limits $n < N$ observations. We denote the estimator for Γ_k by $\tilde{\Gamma}_k$, which is defined as follows:

$$\tilde{\Gamma}_{k,\text{CS-HAC}} = \frac{1}{T_H - k} \sum_{t_h=k+1}^{T_H} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \tilde{\lambda}_i \tilde{\lambda}_j' \tilde{e}_{i,t_h} \tilde{e}_{j,t_h-k}, \tag{12}$$

where $n = \min(\sqrt{N}, \sqrt{T_H})$. Note that by Assumption A.2-(d), Γ_k does not depend on time.

Theorem 2.2 *Suppose the Assumptions A.1-A.4 in Appendix A hold. Then, for any fixed* $k = 0, 1, 2, \dots,$

$$\|\tilde{\Gamma}_k - H_0^{-1\prime} \Gamma_k H_0^{-1}\| \xrightarrow{p} 0 \quad if \quad \frac{n}{\min(N, T_H)} \to 0,$$

Here, in Theorem 2.2, $\tilde{\Gamma}_k$ depends on the assumption on the serial and cross-sectional dependence in the idiosyncratic errors of the factor model. If there is only serial dependence, $\tilde{\Gamma}_k = \frac{1}{N} \sum_{i=1}^{N} \tilde{\lambda}_i \tilde{\lambda}_i' \frac{1}{T_H - k} \sum_{t_h=k+1}^{T_H} \tilde{e}_{i,t_h} \tilde{e}_{i,t_h-k}$. If we allow for cross-sectionally dependence additionally, $\tilde{\Gamma}_k = \tilde{\Gamma}_{k,\text{CS-HAC}}$ defined in (12). Note that if $k = 0$, our estimators are equivalent to the estimators proposed in Bai and Ng (2006). Theorem 2.2 enables us to construct consistent estimators for (10) and (11) as follows:

$$\tilde{B}_{\beta_1} = \Big[2 \sum_{k=1}^{K} w_k(\tilde{\theta}) \tilde{\Sigma}_{\tilde{f}} w_k(\tilde{\theta}) + \sum_{k=1}^{K} \sum_{l \neq k}^{K} w_k(\tilde{\theta}) \Big\{ \tilde{V}^{-1} \tilde{\Gamma}_{k-l,\text{CS-HAC}} \tilde{V}'^{-1} + \tilde{Q}_{k-l} \tilde{\Gamma}_{\text{CS-HAC}} \tilde{V}^{-2} \Big\} w_l(\tilde{\theta}) \Big] \tilde{\beta}_1, \text{ and}$$

$$\tilde{B}_{\theta} = \tilde{\beta}_1 \circ \Big[2 \sum_{k=1}^{K} \frac{\partial w_k(\tilde{\theta})}{\partial \theta} \tilde{\Sigma}_{\tilde{f}} w_k(\tilde{\theta}) + \sum_{k=1}^{K} \sum_{l \neq k}^{K} \frac{\partial w_k(\tilde{\theta})}{\partial \theta} \Big\{ \tilde{V}^{-1} \tilde{\Gamma}_{k-l,\text{CS-HAC}} \tilde{V}^{-1} + \tilde{Q}_{k-l} \tilde{\Gamma}_{\text{CS-HAC}} \tilde{V}^{-2} \Big\} w_l(\tilde{\theta}) \Big] \tilde{\beta}_1,$$

where $\tilde{\Sigma}_{\tilde{f}} = \tilde{V}^{-1} \tilde{Q} \tilde{\Gamma}_{\text{CS-HAC}} \tilde{Q} \tilde{V}^{-1}$ with $\tilde{Q} = \tilde{f}' \tilde{f} / T_H$, and $\tilde{Q}_{k-l} = \sum_{t_h=k+1}^{T_H} \tilde{f}_{t_h}' \tilde{f}_{t_h-k}$. Note that the bias estimates can be simpler under the restriction on either cross-sectional or serial dependence, or both. We denote the bias-corrected estimator by $\tilde{\alpha}_{\text{BC}}$ such that $\tilde{\alpha}_{\text{BC}} \equiv \tilde{\alpha} - (-\frac{1}{N} \tilde{\Delta}_{\alpha})$. Here, $-\tilde{\Delta}_{\alpha}$ is the estimate of the bias in $\tilde{\alpha}$, where $\tilde{\Delta}_{\alpha} = \tilde{\Sigma}^{-1} (\tilde{B}_{\beta}', \tilde{B}_{\theta}')'$ with $\tilde{B}_{\beta} = (\tilde{B}_{\beta_0}, \tilde{B}_{\beta_1}')'$ and $\tilde{B}_{\beta_0} = 0$.

Proposition 2.1 *Suppose the Assumptions A.1 - A.6 in Appendix A hold and* $\sqrt{T}/N \to c$, *where* $0 \leq c < \infty$, *then*

$$\sqrt{T} (\tilde{\alpha}_{BC} - \alpha) \xrightarrow{d} N(0, \Sigma_{\alpha}). \tag{13}$$

Based on Proposition 2.1, the bias corrected estimator no longer contains an asymptotic bias. However, it is well known that an approach based on asymptotic theory does not perform well in finite samples. Additionally, the bias takes a very complicated form in our context, which makes it difficult to implement. Therefore, we discuss an alternative approach, a bootstrap method in the next section.

3 Bootstrap method: Autoregressive-sieve + CSD bootstrap

In this section, we propose a bootstrap method and show its validity by proving that our method satisfies bootstrap high level conditions under which any general residual-based bootstrap is satisfied. The bootstrap high level conditions are similar to those of GP (2014), hence we leave them in the appendix (see Appendix C).

In particular, we propose a bootstrap procedure, where we resample the factor model and the MIDAS regression model, and then obtain the bootstrap estimates. For resampling the idiosyncratic errors in the factor model, GP (2014) proposed a wild bootstrap and proved its validity in the context of the factor-augmented regression models under no cross-sectional dependence. To allow for cross-sectional dependence, Gonçalves and Perron (2020) proposed a bootstrap method that utilizes a thresholding technique to allow for the cross-sectional dependence, so-called CSD (cross-sectional dependent) bootstrap. However, these methods cannot be used in our context as it destroys the serial dependence in the idiosyncratic error terms.

Therefore, we propose a new method that combines autoregressive sieve bootstrap, which was originally proposed in Bühlmann (1997) and has been further discussed in Kreiss, Paparoditis, and Politis (2011) and the CSD bootstrap. We refer our bootstrap method to AR-sieve + CSD bootstrap. Our bootstrap method is recently considered by Gonçalves et al. (2023), where they replace the autoregressive sieve bootstrap with an autoregressive parametric bootstrap. The AR-sieve + CSD bootstrap method resamples each time series residual in the factor model with an autoregressive sieve process and the corresponding innovations by the CSD bootstrap method. Therefore, the cross-sectional dependence is captured in the innovation terms and the serial dependence is captured by an autoregressive process.

In order to prove our bootstrap method is valid, we assume that $\{e_{i,t_h}\}_{t_h=1}^{T_H}$ is a causal process that can be represented as an AR(∞) process such that $e_{i,t_h} = \sum_{j=1}^{\infty} a_{i,j} e_{i,t_h-j} + u_{i,t_h}$, for $t_h = 1, \ldots, T_H$ and $i = 1, \ldots, N$. We assume the autoregressive process is stationary such that the coefficients are absolutely summable, $\sum_{j=1}^{\infty} |a_{i,j}| < \infty$, for each $i = 1, \ldots, N$. The innovation terms in AR(∞) process, $u_{t_h} = (u_{1,t_h}, \ldots, u_{N,t_h})'$, are identically and independently distributed from a distribution with mean zero and finite variance, Σ_u. Here, Σ_u is assumed to be non-diagonal to account for cross-sectional dependence in the idiosyncratic error term. Our bootstrap algorithm is as follows.

Bootstrap Algorithm

1. For each $i = 1, \ldots, N$, select an order $p_i = p_i(T_H)$, $p_i << T_H$, and fit a p_i-th order autoregressive model to $\tilde{e}_{i,1}, \ldots, \tilde{e}_{i,T_H}$, where $\tilde{e}_{i,t_h} = X_{i,t_h} - \tilde{\lambda}_i \tilde{f}_{t_h}$. We denote $\tilde{a}_i(p_i) = (\tilde{a}_{i,j}(p_i), j = 1, \ldots, p_i)$, the Yule-Walker autoregressive parameter estimators, such that $\tilde{a}_i(p_i) = \tilde{\Gamma}(p_i)^{-1}\tilde{\gamma}_{p_i}$, with $\tilde{\gamma}_{p_i} = (\tilde{\gamma}_e(1), \tilde{\gamma}_e(2), \ldots, \tilde{\gamma}_e(p_i))'$ and $\tilde{\Gamma}(p_i) = (\tilde{\gamma}_e(r-s))_{r,s=1,2,\ldots,p_i}$ such that

$$\tilde{\gamma}_e(\tau) = \frac{1}{T_H} \sum_{t_h=1}^{T_H - |\tau|} (\tilde{e}_{i,t_h} - \bar{e}_i)(\tilde{e}_{i,t_h+|\tau|} - \bar{e}_i), \tag{14}$$

   for $\tau = 0, \ldots, p_i$ and $\bar{e}_i = T_H^{-1} \sum_{t_h=1}^{T_H} \tilde{e}_{i,t_h}$.

   With chosen lag length $p_i = p_i(T_H)$,

$$e^*_{i,t_h} = \sum_{j=1}^{p_i} \tilde{a}_{i,j}(p_i) e^*_{i,t_h-j} + u^*_{i,t_h}, \text{ for } t_h = 1, \ldots, T_H, \tag{15}$$

   where $u^*_{t_h} = (u^*_{1,t_h}, \ldots, u^*_{N,t_h}) = \tilde{\Sigma}_u^{1/2} \eta_{t_h}$ with $\eta_{t_h} \sim$ i.i.d $(0, I_N)$. The initial conditions are $e^*_{i,0}, \ldots, e^*_{i,1-p_i} = 0$, for $i = 1, \ldots, N$, which is equivalent to the stationary mean of e^*_{i,t_h} in the bootstrap world. Following Gonçalves and Perron (2020), we choose $\tilde{\Sigma}_u$ by a thresholding technique such that

$$\tilde{\Sigma}_u = (\hat{\sigma}_{u,ij})_{i,j=1,\ldots,N},$$

   with

$$\hat{\sigma}_{u,ij} = \begin{cases} \tilde{\sigma}_{u,ij} & i = j \\ \tilde{\sigma}_{u,ij} 1\left(|\tilde{\sigma}_{u,ij}| > \omega\right) & i \neq j, \end{cases} \quad \text{with } \tilde{\sigma}_{u,ij} = \frac{1}{T_H} \sum_{t_h=1}^{T_H} \tilde{u}_{i,t_h} \tilde{u}_{j,t_h},$$

   where ω is a threshold and $\tilde{u}_{i,t_h} = \tilde{e}_{i,t_h} - \sum_{j=1}^{p_i} \tilde{a}_{i,j}(p_i) \tilde{e}_{i,t_h-j}$ for $i = 1, \ldots, N$ and $t_h = 1, \ldots, T_H$.

2. For $t = 1, \ldots, T$,

$$y_t^* = \tilde{\beta}_0 + \tilde{\beta}_1' \tilde{F}_t(\tilde{\theta}) + \varepsilon_t^*,$$

where $\varepsilon_t^* = \nu_t \hat{\varepsilon}_t$, $\hat{\varepsilon}_t = y_t - \tilde{\beta}_0 - \tilde{\beta}_1' \tilde{F}_t(\tilde{\theta})$ and ν_t is randomly generated from a standard normal distribution i.i.d. across t.

3. We obtain the estimated factors, $\tilde{f}^*$ and factor loadings, $\tilde{\Lambda}^*$ by principal component analysis on bootstrap panel, X_t^*.

4. By regressing y_t^* on 1 and temporally aggregated $(f_{t-1/m}^{*'}, \ldots, f_{t-K/m}^{*'})'$, we obtain the estimates in the bootstrap world, $\tilde{\beta}^*$ and $\tilde{\theta}^*$.

In step 1, we resample the residuals of the factor model by AR sieve + CSD bootstrap. The way we resample the residuals in the factor model is similar to the bootstrap procedure in Kreiss et al. (2011) and Bühlmann (1997). The difference is that we resample the innovation terms in the autoregressive process for each series using CSD bootstrap proposed by Gonçalves and Perron (2020). In the second step, we resample the regression errors by a simple wild bootstrap, which is the same bootstrap method used in GP (2014) and in Gonçalves and Perron (2020) in their second step. Finally, we estimate the factors and factor loadings from a bootstrap panel dataset, $X_{t_h}^*$, for $t_h = 1, \ldots, T_H$, and estimate the parameters by regressing the bootstrap samples, y_t^* on 1 and $F_t^*(\tilde{\theta})$. To prove the validity of AR sieve + CSD bootstrap, we introduce the following additional assumptions.

Assumption 1 λ_i *are either deterministic such that* $\|\lambda_i\| \leq M \leq \infty$, *or stochastic such that* $E\|\lambda_i\|^{12} \leq M < \infty$ *for all* i*:* $E\|f_{t_h}\|^{12} \leq M < \infty$*;* $E|e_{i,t_h}|^{12} \leq M < \infty$, *for all* (i, t_h)*; and for some* $q > 1$, $E|\varepsilon_t|^{4q} \leq M < \infty$, *for all* t.

Assumption 2 $E(\varepsilon_t | y_t, F_t, y_{t-1}, F_{t-1}, \ldots) = 0$, *and* $F_t = (f_t, \ldots, f_{t-k/m})'$ *and* ε_t *are independent of the idiosyncratic errors* e_{i,s_h} *for all* (i, s_h, t).

Assumption 3 $e_{i,t_h} = \sum_{j=1}^{\infty} a_{i,j} e_{i,t_h-j} + u_{i,t_h}$, *with* $\sum_{j=1}^{\infty} |a_{i,j}| < \infty$, *for* $t_h = 1, \ldots, T_H$ *and* $i = 1, \ldots, N$.

Assumption 4 $\Sigma_u \equiv E(u_{t_h} u_{t_h}') = (\sigma_{u,ij})_{i,j=1,\ldots,N}$, *with* $u_{t_h} = (u_{1,t_h}, \ldots, u_{N,t_h})'$, *for all* t_h, i, j *and is such that* $\lambda_{\min}(\Sigma_u) > c_1$ *and* $\lambda_{\max}(\Sigma_u) < c_2$ *for some positive constants* c_1 *and* c_2.

Assumptions 1 and 2 are equivalent to the Assumptions 6 and 7 in GP (2014). In Assumption 1, we strengthen the moment conditions for the factors and factor loadings in Assumption A.1 in Appendix A. Assumption 2 justifies the wild bootstrap in the second step as the regression error term is a martingale difference sequence. Furthermore, we assume that each time series idiosyncratic error term is a stationary autoregressive process of infinite order in Assumption 3. Finally, Assumption 4 is similar to the CS assumption in Gonçalves and Perron (2020) (on the idiosyncratic error terms) and Gonçalves et al. (2023) (on the innovations of the idiosyncratic error terms). We assume that the variance-covariance matrix of the innovation terms is time-invariant and the innovation terms are weakly dependent in cross-sectional dimension. Under these additional assumptions, we show the validity of the AR-sieve +CSD bootstrap method in the following theorem.

Theorem 3.1 *Suppose that autoregressive sieve with CSD (AR-sieve + CSD) bootstrap and wild bootstrap are used to generate* $\{e^*_{i,t_h}\}$ *and* $\{\varepsilon^*_t\}$*, respectively with* $E^*|\eta_{i,t_h}|^4 < C$ *for all* (i, t_h) *and* $E^*|\nu_t|^{4q} < C$ *for all* t*, for some* $q > 1$*. If Assumptions A.1 - A.6 in Appendix A and Assumptions 1 - 4 hold,*

$$\sup_{x\in\mathbb{R}^{r+p}} |P^*(\sqrt{T}(\Phi^*\tilde{\alpha}^* - \tilde{\alpha}) \leq x) - P(\sqrt{T}(\tilde{\alpha} - \alpha) \leq x)| \xrightarrow{p} 0.$$

4 Monte Carlo Simulation

In this section, we confirm the presence of bias in the factor-MIDAS regression models, and show the finite sample performance of both inference methods we propose. The data generating process (DGP) is similar to GP (2014) and Aastveit, Foroni, and Ravazzolo (2017). We consider the factor-MIDAS regression model with a single factor model as follows:

$$y_t = \beta_0 + \beta_1 \sum_{k=1}^{K} w_k(\theta) f_{t-k/m} + \varepsilon_t, \tag{16}$$

$$X_{i,t-k/m} = \lambda_i f_{t-k/m} + e_{i,t-k/m}, \; k = m-1, \dots, 0. \tag{17}$$

For a weighting function, $w_k(\theta)$, for $k = 1, \dots, K$, we use the exponential Almon lag with two parameters, (2).

The factors and factor loadings are generated similarly to GP (2014). The single factor f_t is randomly drawn from a standard normal distribution independently over time. The factor loading, λ_i is randomly drawn from a uniform distribution of the interval $[0, 1]$ independently across indicators, i. We consider that the high-frequency variable is observed at most 3 times between $t-1$ and t (equivalent to low-frequency data being quarterly and high-frequency data being monthly). The parameters are $\beta_0 = 0$, $\beta_1 = 2.5$, $\theta_1 = 0.007$ and $\theta_2 = -0.01$. We choose the weighting parameters similar to Aastveit et al. (2017) to induce fast-decaying weights.

Table 1: Data generating process

DGP	ε_t	e_{i,t_h}
1	$N(0,1)$	$N(0,1)$
2	$\varepsilon_t = \sqrt{h_t} v_t$	$N(0,1)$
3	$\varepsilon_t = \sqrt{h_t} v_t$	$N(0,\sigma_i^2)$
4	$\varepsilon_t = \sqrt{h_t} v_t$	AR + $N(0,\sigma_i^2)$
5	$\varepsilon_t = \sqrt{h_t} v_t$	CS + $N(0,1)$
6	$\varepsilon_t = \sqrt{h_t} v_t$	CS + AR

where $h_t = 0.1 + 0.3\varepsilon_{t-1}^2 + 0.6h_{t-1}$ and $\nu_t \sim$ i.i.d.$N(0,1)$ for $t = 1, \ldots, T$ and $t_h = 1, \ldots, T_H$.

Table 1 shows six different scenarios to generate the idiosyncratic error terms and MIDAS regression error terms. We consider the error term in the regression model to be either homoskedastic or heteroskedastic. In DGP 1, we consider homoskedastic error term and in the rest of the DGPs, the error terms are conditionally heteroskedastic. When they are homoskedastic, the errors are drawn independently and identically from a standard normal distribution. To allow for heteroskedasticity, we assume that the error terms follow a GARCH model, which implies that they are conditionally heteroskedastic but unconditionally homoskedastic. Particularly, we use the same process in Aastveit et al. (2017): $\varepsilon_t = \sqrt{h_t} v_t$ where $h_t = 0.1 + 0.3\varepsilon_{t-1}^2 + 0.6h_{t-1}$ and $v_t \sim$ i.i.d.$N(0,1)$.

For the idiosyncratic term in the factor model, we use the same data-generating process in GP (2014). In DGP 1 and DGP 2, the idiosyncratic error terms are homoskedastic by randomly generating them from a standard normal distribution. DGP 3 induces heteroskedasticity in the idiosyncratic term, where the variance for each indicator is drawn from $U[0.5, 1.5]$. DGP 4 introduces the serial correlation by generating the idiosyncratic term from an autoregressive model of order one such that $e_{i,t_h} = \rho_i e_{i,t_h-1} + u_{i,t_h}$, where $u_{i,t_h} \sim$ i.i.d.$N(0,1)$. For simplicity, we let $\rho_i = \rho$ for all $i = 1, \ldots, N$, and $\rho = 0.5$. The idiosyncratic terms are re-scaled by $(1-\rho^2)^{1/2}$ so that the variance of the idiosyncratic error terms is 1. DGP 5 allows for cross-sectional dependence in the homoskedastic idiosyncratic terms as in GP (2014) and Bai and Ng (2006). Precisely, we let the correlation between e_{i,t_h} and e_{j,t_h} be $0.5^{|i-j|}$ for $|i-j| \leq 5$ and 0 for otherwise. In DGP 6, the idiosyncratic error terms have both serial and cross-sectional dependence. The idiosyncratic error terms follow the autoregressive process of order 1 with the innovation term being cross-sectionally correlated. The idiosyncratic terms in DGP 5 and 6 are also re-scaled to have the variance 1, the same as in other designs.

We report the size of the bias in a slope coefficient for the single factor, β_1. Mainly, we report two sets of results: based on asymptotic theory and based on the bootstrap method. The bias based on asymptotic theory is reported when we use the true factor, the estimated factor, and the plug-in bias estimator. We also impose that we know $Cov(e_{i,t_h}, e_{i,t_h-k}) = 0$ for $k > 1$, and therefore we only compute the bias term up to the first degree covariance term. The other set of results includes the bias based on two different bootstrap methods: wild bootstrap and AR sieve + CSD bootstrap. The wild bootstrap is only valid when the idiosyncratic error terms do not have the serial and cross-sectional dependence, DGP 1 - 3. For the rest of the designs, the wild bootstrap is not valid. Therefore, under the general settings (DGP 4 - 6), we can quantify the cost of not accounting for either time series or cross-sectional dependence or both in the idiosyncratic error term by comparing two bootstrap methods.

To compute the size of bias, we use the approach described in GP (2014). The bias in the original sample is calculated as the average of $H\tilde{\beta}_1 - \beta_1$. This guarantees each estimator in the replication to be consistent for β_1. In the bootstrap world, similarly, we compute the bias of the bootstrap estimator as the average of $HH^*\tilde{\beta}_1^* - H\tilde{\beta}_1$. We also report the 95% coverage rate for the associated estimators: estimated factors, plug-in bias and two bootstrap methods. The coverage rates associated with the bootstrap methods are reported by using the bootstrap equal-tailed percentile-t method.

All our simulation results are based on 5000 replications and 399 bootstraps. We consider $T = 50, 100, 200$ and $N = 50, 100, 200$. Since the high frequency is observed $m = 3$ times more, the time series dimensions in the factor model as 150, 300, and 600, respectively. We choose $K = 11$, which implies that a low-frequency variable can be explained by 11 lagged monthly factors.

Table 2 shows the results of DGP 1 and 2 in each panel. The first panel shows the results of

Table 2: DGP 1 & DGP 2 - Bias and coverage rate of 95% CIs for β

		$N=50$			$N=100$			$N=200$		
		$T=50$	100	200	50	100	200	50	100	200
		$T_H=150$	300	600	150	300	600	150	300	600
	bias									
DGP 1: homo & homo	True Factor	-0.01	-0.01	0.00	-0.02	-0.01	0.00	0.00	0.00	0.00
	Estimated Factor	-0.32	-0.31	-0.29	-0.20	-0.17	-0.16	-0.12	-0.10	-0.08
	Plug-in	-0.38	-0.34	-0.32	-0.21	-0.19	-0.18	-0.10	-0.10	-0.09
	WB	-0.25	-0.24	-0.23	-0.16	-0.15	-0.14	-0.11	-0.09	-0.08
	AR-sieve+CSD	-0.24	-0.24	-0.23	-0.16	-0.15	-0.14	-0.10	-0.09	-0.08
	95% coverage rate									
	Estimated Factor	84.8	82.0	73.9	89.6	90.5	88.3	91.7	92.7	93.4
	Plug-in	87.6	89.1	89.3	90.4	92.1	92.4	91.2	92.7	93.6
	WB	94.1	94.7	93.3	95.0	95.6	94.5	92.7	95.4	94.9
	AR-sieve+CSD	95.8	94.9	92.4	95.8	96.1	95.0	96.0	96.3	95.3
	bias									
DGP 2: hetero & homo	True Factor	-0.01	0.00	0.00	0.00	0.01	-0.01	0.01	-0.01	0.00
	Estimated Factor	-0.34	-0.30	-0.29	-0.19	-0.16	-0.16	-0.10	-0.10	-0.09
	Plug-in	-0.37	-0.34	-0.32	-0.20	-0.19	-0.18	-0.10	-0.10	-0.09
	WB	-0.24	-0.24	-0.23	-0.16	-0.15	-0.14	-0.10	-0.09	-0.08
	AR-sieve+CSD	-0.24	-0.24	-0.23	-0.16	-0.15	-0.14	-0.10	-0.09	-0.08
	95% coverage rate									
	Estimated Factor	78.1	76.2	68.4	85.9	88.1	86.2	88.7	91.5	91.6
	Plug-in	82.7	86.8	88.3	86.6	89.8	92.5	88.9	92.3	92.5
	WB	91.7	93.0	93.1	92.6	93.3	94.2	91.0	94.4	94.0
	AR-sieve+CSD	92.5	92.9	92.2	94.0	95.2	93.8	93.5	94.8	94.8

In DGP 1, both error terms are homoskedastic. In DGP 2, MIDAS regression error terms are heteroskedastic and idiosyncratic error terms are homoskedastic. The results of coverage rates, when we use the estimated factors and plug-in bias, are based on asymptotic theory. The bootstrap coverage rates use the bootstrap equal-tailed percentile t method.

DGP 1, where both error terms are randomly generated from an i.i.d. standard normal distribution. Below the row "bias", we have a size of bias for each case: true factor, estimated factor, plug-in bias and two bootstrap methods. The fourth and fifth rows contain the bias when we use the bootstrap methods, wild bootstrap, and autoregressive sieve with CSD bootstrap, respectively. The results indicate that there is no bias when using the true factor, however, a bias does exist when using the estimated factor as a regressor. Increasing the sample size in both cross-sectional and time series dimensions results in a decrease in bias. If the cross-sectional dimension is small (50 and 100), the plug-in bias tends to overestimate the bias size. Both bootstrap methods perform similarly and replicate bias size well. When no method is used to correct the bias, size distortion occurs in terms of coverage rates. The plug-in bias somewhat recovers the size distortion, but bootstrap methods outperform the plug-in bias method. The results of DGP 1 and DGP 2 are similar, and both bootstrap methods are valid for these scenarios since the idiosyncratic error terms are randomly selected from a standard normal distribution.

The results of DGP 3 and 4 are presented in Table 3. In both scenarios, the MIDAS regression error terms are now heteroskedastic for both DGPs. The idiosyncratic error terms are heteroskedastic. The results of DGP 3 are similar to those of DGP 1 and 2. We have a bias when we use the estimated factor and the plug-in estimator overestimates the magnitude of the bias, especially in small samples. Both bootstrap methods outperform the plug-in estimator in terms of replicating the bias size and correcting the distortion. In DGP 4, the idiosyncratic error terms exhibit not only heteroskedasticity but also display serially dependence. In contrast to DGP 3, the bias size increases as we introduce serial dependence in the error term of the factor model, and it is about twice as large as that in DGP 3. This is consistent with the asymptotic bias result in Theorem 2.1, where time-series dependence contributes to the bias. The plug-in bias is no longer overestimating the bias size.[6]

Comparing the two bootstrap methods, it is evident that the autoregressive sieve with the CSD bootstrap method performs better than the wild bootstrap method. Note that the wild bootstrap is no longer valid under serial dependence. In fact, for some sample sizes, the wild bootstrap even performs worse than the plug-in bias. We can also confirm that the autoregressive sieve + CSD bootstrap procedure outperforms the plug-in bias and wild bootstrap procedure by comparing the results of coverage rates. The coverage rates from AR sieve + CSD bootstrap outperform the plug-in and wild bootstrap methods in all sample sizes.

Finally, we present the results of DGP 5 and 6, which are shown in Table 4. In DGP 5, the idiosyncratic error term is only cross-sectionally correlated. The AR-sieve + CSD bootstrap performs better than the wild bootstrap method but worse than the plug-in bias method. However,

[6] It is important to note that since the bias depends on the serial dependence, the persistence in the idiosyncratic error term may also have an impact. We have observed that with an increase in persistence, the bias also increases (refer to the additional table in Appendix D).

Table 3: DGP 3 & DGP 4 - Bias and coverage rate of 95% CIs for β

		$N=50$			$N=100$			$N=200$		
		$T=50$	100	200	50	100	200	50	100	200
		$T_H=150$	300	600	150	300	600	150	300	600
	bias									
DGP 3: hetero & hetero	True Factor	0.00	-0.01	0.00	-0.01	0.00	0.00	0.01	0.00	0.00
	Estimated Factor	-0.37	-0.34	-0.32	-0.22	-0.19	-0.17	-0.12	-0.11	-0.10
	Plug-in	-0.41	-0.36	-0.35	-0.22	-0.20	-0.19	-0.11	-0.11	-0.10
	WB	-0.27	-0.26	-0.26	-0.17	-0.16	-0.15	-0.11	-0.10	-0.09
	AR-sieve+CSD	-0.26	-0.26	-0.25	-0.17	-0.16	-0.15	-0.11	-0.10	-0.09
	95% coverage rate									
	Estimated Factor	75.0	72.6	63.9	85.0	85.5	84.4	88.5	90.3	91.0
	Plug-in	80.9	87.9	88.9	86.8	89.3	92.1	88.9	91.1	92.5
	WB	91.7	94.2	92.7	92.6	93.5	94.1	91.3	93.9	93.8
	AR-sieve+CSD	93.7	92.1	90.4	93.6	94.3	94.1	94.1	95.1	93.6
	bias									
DGP 4: hetero & AR	True Factor	0.00	0.00	0.00	-0.01	0.00	0.00	-0.01	0.00	0.00
	Estimated Factor	-0.64	-0.57	-0.54	-0.41	-0.35	-0.31	-0.28	-0.21	-0.18
	Plug-in	-0.45	-0.42	-0.41	-0.26	-0.26	-0.25	-0.14	-0.14	-0.14
	WB	-0.22	-0.22	-0.22	-0.15	-0.14	-0.14	-0.10	-0.09	-0.08
	AR-sieve+CSD	-0.38	-0.37	-0.36	-0.29	-0.26	-0.25	-0.22	-0.18	-0.16
	95% coverage rate									
	Estimated Factor	52.2	44.5	29.2	72.3	71.8	67.3	81.5	85.0	84.1
	Plug-in	72.0	77.1	77.1	81.1	86.0	87.9	85.0	90.1	91.3
	WB	82.8	79.4	68.7	89.0	88.8	86.1	89.6	92.4	91.3
	AR-sieve+CSD	88.7	87.4	81.4	91.9	91.9	91.3	93.6	94.9	93.5

In DGP 3, both error terms are heteroskedastic. In DGP 4, the idiosyncratic error term is generated as the autoregressive process of lag 1 for each variable and with heteroskedastic. For coverage rates, the results for estimated factors and plug-ins are based on asymptotic theory. The bootstrap coverage rates use the bootstrap equal-tailed percentile t method.

AR-sieve + CSD recovers the size distortion better than the plug-in method in all sample sizes when it comes to coverage rates. This is because there must be some variance effect when the bootstrap method is used. In DGP 6, we allow for cross-sectional dependence as well as serial dependence in the idiosyncratic error terms. The results follow a similar pattern to the findings of DGP 5. The plug-in bias method replicates the bias better than bootstrap methods. However, it does worse than AR-sieve+CSD bootstrap in terms of recovering the size distortion in the coverage rates. When

Table 4: DGP 5 & DGP 6 - Bias and coverage rate of 95% CIs for β

		N = 50			N = 100			N = 200		
		$T = 50$	100	200	50	100	200	50	100	200
		$T_H = 150$	300	600	150	300	600	150	300	600
		bias								
DGP 5: hetero & CSD	True Factor	0.00	-0.01	0.00	-0.01	0.00	0.00	0.01	0.00	0.00
	Estimated Factor	-0.37	-0.34	-0.32	-0.22	-0.19	-0.17	-0.12	-0.11	-0.10
	Plug-in	-0.41	-0.36	-0.35	-0.22	-0.20	-0.19	-0.11	-0.11	-0.10
	WB	-0.10	-0.10	-0.10	-0.06	-0.06	-0.04	-0.04	-0.04	-0.03
	AR-sieve+CSD	-0.16	-0.16	-0.16	-0.10	-0.10	-0.10	-0.06	-0.06	-0.06
		95% coverage rate								
	Estimated Factor	75.0	72.6	63.9	85.0	85.5	84.4	88.5	90.3	91.0
	Plug-in	80.9	87.9	88.9	86.8	89.3	92.1	88.9	91.1	92.5
	WB	88.7	86.2	79.5	92.7	92.6	90.0	94.2	93.5	93.5
	AR-sieve+CSD	90.9	90.0	87.0	93.3	94.1	92.3	94.3	93.9	93.7
		bias								
DGP 6: hetero & CSD+AR	True Factor	0.00	0.00	0.00	-0.01	0.00	0.00	-0.01	0.00	0.00
	Estimated Factor	-0.64	-0.57	-0.54	-0.41	-0.35	-0.31	-0.28	-0.21	-0.18
	Plug-in	-0.45	-0.42	-0.41	-0.26	-0.26	-0.25	-0.14	-0.14	-0.14
	WB	-0.08	-0.09	-0.08	-0.06	-0.06	-0.05	-0.04	-0.03	-0.03
	AR-sieve+CSD	-0.23	-0.23	-0.24	-0.17	-0.16	-0.16	-0.12	-0.10	-0.10
		95% coverage rate								
	Estimated Factor	52.2	44.5	29.2	72.3	71.8	67.3	81.5	85.0	84.1
	Plug-in	72.0	77.1	77.1	81.1	86.0	87.9	85.0	90.1	91.3
	WB	76.5	66.2	47.4	87.5	84.2	77.6	91.1	91.5	89.3
	AR-sieve+CSD	86.3	80.0	73.5	91.0	89.8	87.1	93.2	93.2	92.6

In DGP 5 and 6, both error terms are heteroskedastic. In DGP 5, the idiosyncratic error term contains the cross-sectional dependence. In DGP 6, we impose the dependence in both dimensions for the idiosyncratic error terms. For coverage rates, the results for estimated factors and plug-in are based on asymptotic theory. The bootstrap coverage rates use the bootstrap equal-tailed percentile t method.

the time series dimension is as small as 50, the plug-in bias method performs even worse than the wild bootstrap method, which is not valid in this design. Overall, the AR-sieve+CSD bootstrap works well in correcting the distortion.

5 Empirical Application

In this section, we apply the factor-MIDAS regression model to validate the presence of bias in an empirical example. It is well documented that incorporating high-frequency indicators to forecast a quarterly variable using the MIDAS regression model improves the forecast performance (e.g., see Clements and Galvão (2008; 2009), Aastveit et al. (2017), Marcellino and Schumacher (2010), Andreou, Ghysels, and Kourtellos (2013), and Beyhum and Striaukas (2023)).

In this paper, we focus on nowcasting quarterly U.S. real GDP growth using monthly macroeconomic factors from 1984 Q1 to 2022 Q4 including great moderation period. We have divided this period into two: the long period (1984 Q1 - 2022 Q4), which includes the COVID pandemic period, and the short period (1984 Q1 to 2019 Q4). Our nowcasting model is similar to the model in Beyhum and Striaukas (2023). Given the number of leading months, $l = 1, 2, 3$, we write our model as follows:

$$y_t = \beta_0 + \sum_{i=1}^{p_y} \rho_i y_{t-i} + \beta_1' \sum_{k=1-l}^{K-l} w_{(k-1)+l}(\theta) f_{t-1-(j-1)/m} + \varepsilon_t, \tag{18}$$

where y_t is quarterly U.S. GDP growth rate. We denote common factors containing timely information about monthly macroeconomic predictors by $f_{t-k/m}$. The number of leading months represents a nowcasting horizon, denoted by h. For instance, $l = 1$ indicates that we exploit information of one leading month; hence, we nowcast two months away ($h = 2$). We use the exponential Almon lag with two parameters defined in (2) for the lag polynomial function. The quarterly U.S. output is obtained from a FRED-QD dataset (for detail, see M. McCracken and Ng (2020)). As U.S. real output is available in level in the dataset, we compute the growth rate in percentage, by $\{\ln(\text{GDP})_t - \ln(\text{GDP})_{t-1})\} \times 100$. We also include the lags of the growth rate in the regression. The number of lags of the dependent variable is chosen by BIC, before the MIDAS regression. BIC selects one lag in the long and three lags in the short periods.

To estimate the monthly factors, we utilize the FRED-MD dataset[7] (for detail, see M. W. McCracken and Ng (2016)). We only consider the 74 macroeconomic variables available for the entire period and exclude all financial variables. Using PCA, we extract two common factors in both periods. The information criterion proposed by Bai and Ng (2002) (particularly, IC_p) chooses eight factors in the long period and five factors in the short period. Although the information criterion

[7] We use the 'current' version downloaded on October 3rd, 2023.

chooses more than 2 factors, the two factors we extract explain more than 60% of the variability explained by all the factors chosen by the information criterion proposed by Bai and Ng (2002).

Our primary goal is to verify the existence of bias in the estimators. Instead of focusing solely on the forecasting performance of the factor-MIDAS regression model, we aim to examine the behaviour of the estimators, particularly their 90% confidence interval. We present three sets of confidence intervals, one based on asymptotic theory and the other two based on the bootstrap method. We use two different bootstrap methods for resampling the idiosyncratic error terms in the factor model: wild bootstrap and AR-sieve + CSD bootstrap, described in Section 3. We also rotate the bootstrap estimators, $\tilde{\beta}_1^*$, with the rotation matrix H^* as in GP (2014) and Gonçalves and Perron (2020).

Table 5: Estimates in the long period (1984 Q1 - 2023 Q4)

		$h = 2$		$h = 1$		$h = 0$	
constant		0.90		0.83		0.99	
	Asymptotic	0.67	1.01	0.67	0.99	0.78	1.21
	WB	0.71	0.98	0.69	0.95	0.73	1.28
	AR sieve+CSD	0.71	0.98	0.69	0.94	0.75	1.26
first factor		2.54		3.79		1.87	
	Asymptotic	1.64	3.44	2.97	4.61	0.31	3.44
	WB	2.01	3.56	3.29	4.72	0.91	3.93
	AR sieve+CSD	2.13	3.54	3.34	4.80	0.90	3.39
second factor		0.04		0.36		-0.95	
	Asymptotic	-0.22	0.30	0.08	0.65	-1.47	-0.43
	WB	-0.17	0.37	0.14	0.75	-1.62	-0.01
	AR sieve+CSD	-0.12	0.38	0.16	0.77	-1.63	-0.21
y_{t-1}		-0.30		-0.30		-0.58	
	Asymptotic	-0.54	-0.06	-0.52	-0.09	-0.87	-0.28
	WB	-0.49	-0.12	-0.44	-0.14	-1.25	-0.26
	AR sieve+CSD	-0.49	-0.12	-0.43	-0.14	-1.22	-0.25

The interval based on the asymptotic theory is obtained by adding and subtracting 1.645 times the heteroskedasticity robust standard errors. The confidence intervals based on bootstrap methods are obtained with equal-tailed bootstrap intervals with a bootstrap number, of 4999. WB indicates that we use wild bootstrap and AR sieve + CSD indicates that we use the bootstrap algorithm described in Section 3.

In Table 5, we present the confidence interval for the point estimates in the long period, 1984 Q1 - 2022 Q4 for each nowcasting horizon, $h = 2, 1$, and 0. We also report the estimate associated with each parameter on the top of the three confidence intervals. The confidence intervals of the intercept coefficient are similar, implying that there is no bias for the intercept estimator. However, a bias does exist in the estimators associated with the factors. For example, the point estimate associated with the first factor for horizon $h = 2$ is 2.54. The confidence interval of this estimate is centered around 2.54, but the bootstrap interval shifts to the right, suggesting a negative bias. The results are similar for the other horizons, $h = 1$ and 0. The second factor is not significant in nowcasting the GDP growth rate when we are two months ahead. However, it is significant if we are one month ahead ($h = 1$), or we are at the end of the quarter ($h = 0$). We can also confirm that there exists a bias in the estimator associated with the second factor. When $h = 1$, the result implies a negative bias, whereas when $h = 0$, there exists a positive bias, shifting the interval to the left. Comparing the two bootstrap methods, there is a small change in the bootstrap confidence intervals of the estimators associated with the two factors. However, the difference is not huge, indicating that the serial and cross-sectional dependence in this example may be small.

In Table 6, we present the results after excluding the COVID pandemic period. The results are similar to those shown in Table 5. When using the bootstrap method, the confidence intervals associated with the factors shift. However, the bias does not have a significant impact on the estimates for the lags of the dependent variable. Additionally, it is worth noting that as we exclude the COVID period, the sign of the estimates associated with the two factors is reversed. Previously, the slope coefficient for the aggregated factors was positive, whereas it becomes negative without the COVID period. This suggests that monthly information during the COVID period has a considerable influence on nowcasting the GDP growth rate.

6 Conclusion

In this paper, we derive the asymptotic distribution of the estimators in the factor-augmented MIDAS regression models. We find that there exists an asymptotic bias arising from the fact that the factors are latent and must be estimated. We show that the bias depends on the serial dependence as well as the cross-sectional dependence of the idiosyncratic error term in the factor model, because MIDAS temporally aggregates the factors and their lags. We propose two inference

Table 6: Estimation result of long period (1984 Q1 - 2019 Q4)

		$h=2$		$h=1$		$h=0$	
constant		0.87		0.92		0.88	
	Asymptotic	0.70	1.03	0.79	1.06	0.75	1.02
	WB	0.76	1.03	0.84	1.09	0.77	1.02
	AR sieve+CSD	0.79	1.05	0.86	1.11	0.79	1.04
first factor		-1.10		-1.34		-1.27	
	Asymptotic	-1.48	-0.73	-1.67	-1.01	-1.53	-1.00
	WB	-1.52	-0.92	-1.78	-1.20	-1.61	-1.12
	AR sieve+CSD	-1.56	-0.98	-1.83	-1.27	-1.66	-1.16
second factor		0.09		-0.14		-0.01	
	Asymptotic	-0.67	0.84	-0.35	0.07	-0.58	0.56
	WB	-0.13	0.26	-0.40	0.03	-0.23	0.14
	AR sieve+CSD	-0.17	0.24	-0.48	0.02	-0.28	0.13
y_{t-1}		-0.11		-0.19		-0.17	
	Asymptotic	-0.24	0.03	-0.31	-0.06	-0.30	-0.04
	WB	-0.26	0.00	-0.33	-0.10	-0.31	-0.06
	AR sieve+CSD	-0.26	-0.01	-0.35	-0.11	-0.31	-0.06
y_{t-2}		-0.06		-0.09		-0.04	
	Asymptotic	-0.24	0.12	-0.24	0.05	-0.17	0.09
	WB	-0.24	0.08	-0.27	0.03	-0.17	0.08
	AR sieve+CSD	-0.24	0.08	-0.27	0.02	-0.18	0.07
ρ_3		-0.16		-0.14		-0.15	
	Asymptotic	-0.29	-0.02	-0.26	-0.03	-0.26	-0.03
	WB	-0.28	-0.04	-0.26	-0.04	-0.26	-0.04
	AR sieve+CSD	-0.29	-0.04	-0.27	-0.05	-0.26	-0.04

The interval based on the asymptotic theory is obtained by adding and subtracting 1.645 times the heteroskedasticity robust standard errors. The confidence intervals based on bootstrap methods are obtained with equal-tailed bootstrap intervals with a bootstrap number, of 4999. WB indicates that we use wild bootstrap and AR sieve + CSD indicates that we use the bootstrap algorithm described in Section 3.

methods that account for this bias: an analytical bias estimator based on the formula derived and a bootstrap method. Both inference methods are robust to serial and cross-sectional dependence.

Although our simulation results supports the theoretical results, the bootstrap method performs better in terms of correcting the size distortion in the coverage rates. We also apply the factor-

MIDAS regression model to nowcasting quarterly U.S. GDP growth rate using monthly macroeconomic factors. Our empirical results imply that there exists a bias in the estimates associated with the estimated factors.

Our results can be extended to construct forecast intervals, similar to Gonçalves, Perron, and Djogbenou (2017), where they construct it in the context of the factor-augmented regression models without mixing the frequencies. By denoting by $\hat{y}_{T+1} = g(\tilde{F}_T, \tilde{\alpha})$ the forecast of y_{T+1} based on information up to time T, we can decompose the forecast error as:

$$\hat{y}_{T+1} - y_{T+1} = -\varepsilon_{T+1} + \frac{1}{\sqrt{T}} \frac{\partial g(\tilde{F}_t, \alpha)}{\partial \alpha'} \sqrt{T}(\tilde{\alpha} - \alpha) + \frac{1}{\sqrt{N}} \beta' H^{-1} \sqrt{N}(\tilde{F}_t(\theta) - HF_t(\theta)) + o_p(1).$$

Acknowledgement

Computations were made on the supercomputer Beluga from McGill University, managed by Calcul Québec and Compute Canada. The operation of this supercomputer is funded by the Canada Foundation for Innovation (CFI), Ministère de l'Économie, des Sciences et de l'Innovation du Québec (MESI) and le Fonds de recherche du Québec – Nature et technologies (FRQ-NT).

A Primitive assumptions

This section delivers the primitive assumption for asymptotic theory. The factor-augmented MIDAS regression involves two frequencies, thus we use two time indices: $t_h = 1, \ldots, T_H$ denotes the high-frequency time index and $t = 1, \ldots, T$ denotes the low-frequency time index. Particularly, we use a subscript h to denote high-frequency time index (e.g. s_h also denotes the high-frequency time index).

Assumption A.1 (Factors and Factor Loadings)

(a) f_{t_h} *are stationary with* $E\|f_{t_h}\|^4 \leq M$ *and* $\frac{1}{T_H}\sum_{t_h=1}^{T_H} f_{t_h} f_{t_h}' \xrightarrow{p} \Sigma_F > 0$*, where* Σ_f *is a non-random* $r \times r$ *matrix.*

(b) The factor loadings λ_i *are either deterministic such that* $\|\lambda_i\| \leq M$*, or stochastic such that* $E\|\lambda_i\|^4 \leq M$*. In either case,* $\Lambda'\Lambda/N \xrightarrow{p} \Sigma_\Lambda > 0$*, where* Σ_Λ *is a non-random matrix.*

(c) The eigenvalues of the $r \times r$ *matrix* $(\Sigma_\Lambda \Sigma_f)$ *are distinct.*

(d) $f'f/T_H = I_r$ *and* $\Lambda'\Lambda$ *is a diagonal matrix with distinct entries.*

Assumption A.2 (Time and Cross Section Dependence and Heteroskedasticity)

(a) $E(e_{i,t_h}) = 0, E|e_{i,t_h}|^8 \leq M.$

(b) $E(e_{i,t_h}e_{j,s_h}) = \sigma_{ij,t_hs_h}, |\sigma_{ij,t_hs_h}| \leq \bar{\sigma}_{ij}$ *for all* (t_h, s_h) *and* $|\sigma_{ij,t_hs_h}| \leq \tau_{t_hs_h}$ *for all* (i,j) *such that* $\frac{1}{N}\sum_{i,j=1}^{N}\bar{\sigma}_{ij} \leq M$*,* $\frac{1}{T_H}\sum_{t_h,s_h=1}^{T_H}\tau_{t_hs_h} \leq M$*, and* $\frac{1}{NT_H}\sum_{t_h,s_h,i,j}|\sigma_{ij,t_hs_h}| \leq M.$

(c) For every $(t_h, s_h), E\big|N^{-1/2}\sum_{i=1}^{N}(e_{i,t_h}e_{i,s_h} - E(e_{i,t_h}e_{i,s_h}))\big|^4 \leq M.$

(d) $E(e_{i,t_h}e_{j,t_h}) = \sigma_{ij}$ *and* $E(e_{i,t_h}e_{j,t_h-k}) = \sigma_{ij,k}$ *for all* t *and* k*.*

Assumption A.3 (Moments and Weak Dependence Among $\{f_{t_h}\}, \{\lambda_i\}$ and $\{e_{i,t_h}\}$)

(a) $E\big(\frac{1}{N}\sum_{i=1}^{N}\|\frac{1}{\sqrt{T_H}}\sum_{t_h=1}^{T_H}f_{t_h}e_{i,t_h}\|^2\big) \leq M$*, where* $E(f_{t_h}e_{i,t_h}) = 0$ *for all* (i, t_h)*.*

(b) For each t_h*,* $E\|\frac{1}{\sqrt{T_HN}}\sum_{s_h=1}^{T_H}\sum_{i=1}^{N}f_{s_h}(e_{i,t_h}e_{i,s_h} - E(e_{i,t_h}e_{i,s_h}))\|^2 \leq M.$

(c) $E\|\frac{1}{\sqrt{T_HN}}\sum_{t_h=1}^{T_H}f_{t_h}e'_{t_h}\Lambda\|^2 \leq M$*, where* $E(f_{t_h}\lambda'_ie_{i,t_h}) = 0$ *for all* (i, t_h)*.*

(d) $E\big(\frac{1}{T_H}\sum_{t_h=1}^{T_H}\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_ie_{i,t_h}\|^2\big) \leq M$*, where* $E(\lambda_ie_{i,t_h}) = 0$ *for all* (i, t_h)*.*

(e) As $N \to \infty$*,* $\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda'_je_{i,t_h}e_{j,t_h} - \Gamma \xrightarrow{p} 0$ *and* $\Gamma \equiv \lim_{N\to\infty}Var\big(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_ie_{i,t_h}\big)$*.*

Assumption A.4 (Serial Dependence between $\{f_{t_h}\}$, $\{\lambda_i\}$ and $\{e_{i,t_h}\}$)

(a) $\frac{1}{T_H}\sum_{t_h=1}^{T_H}f_{t_h}f'_{t_h-k} \xrightarrow{p} \Sigma_{f,k}$*, where* $\Sigma_{f,k}$ *is a non-random* $r \times r$ *matrix.*

(b) For each t_h *and all* k*,* $E\|\frac{1}{\sqrt{T_HN}}\sum_{s_h=1}^{T_H}\sum_{i=1}^{N}f_{s_h}(e_{i,t_h}e_{i,s_h-k} - E(e_{i,t_h}e_{i,s_h-k}))\|^2 \leq M.$

(c) $E\|\frac{1}{\sqrt{NT_H}}\sum_{t_h=1}^{T_H}f_{t_h}e'_{t_h-k}\Lambda\|^2 \leq M$*, where* $E(f_{t_h}\lambda'_ie_{i,t_h-k}) = 0$ *for all* (i, t_h) *and all* k*.*

(d) As $N \to \infty$*,* $\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda'_je_{i,t_h}e_{j,t_h-k} - \Gamma_k \xrightarrow{p} 0$ *and* $\Gamma_k \equiv \lim_{N\to\infty}Cov\big(\frac{\Lambda'e_{t_h}}{\sqrt{N}}, \frac{\Lambda'e_{t_h-k}}{\sqrt{N}}\big)$*.*

Assumption A.5 (Weak Dependence Between Idiosyncratic Errors and Regression Errors)

(a) For each t*,* $E\big|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T}\sum_{i=1}^{N}\varepsilon_s(e_{i,t-j/m}e_{i,s-j/m} - E(e_{i,t-j/m}e_{i,s-j/m}))\big|^2 \leq M$ *for* $j = 0, \ldots, m-1$*.*

(b) $E\|\frac{1}{\sqrt{TN}}\sum_{t=1}^{T}\sum_{i=1}^{N}\lambda_i e_{i,t-j/m}\varepsilon_t\|^2 \leq M$*, where* $E(\lambda_i e_{i,t-j/m}\varepsilon_t) = 0$ *for all* (i,t) *and* $j = 0,\ldots,m-1$*.*

Assumption A.6 (Moments and CLT for the Score Vector)

(a) $E(\varepsilon_t) = 0$ *and* $E|\varepsilon_t|^2 < M$*.*

(b) $E\|g_{\alpha,t}\|^4 \leq M$ *and* $\frac{1}{T}\sum_{t=1}^{T} g_{\alpha,t}g'_{\alpha,t} \xrightarrow{p} \Sigma > 0$ *where* $g_{\alpha,t} = \partial g(F_t;\alpha)/\partial\alpha$*.*

(c) As $T \to \infty$*,* $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_{\alpha,t}\varepsilon_t \xrightarrow{d} N(0,\Omega)$*, where* $E\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_{\alpha,t}\varepsilon_t\|^2 < M$ *and* $\Omega \equiv \lim_{T\to\infty} Var\big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_{\alpha,t}\varepsilon_t\big) > 0$*.*

Assumption A.1 are standard assumptions on the factors and the factor loadings in the factor analysis. Additionally, we assume that the factors are stationary. This is to allow $\Sigma_f = \text{plim}\,\frac{1}{T_H}\sum_{t_h=1}^{T_H} f_{t_h}f'_{t_h} = \text{plim}\,\frac{1}{T}\sum_{t=1}^{T} f_{t-j/m}f'_{t-j/m}$, for all j. Assumption A.1-(d) is one of the identifying restrictions from Bai and Ng (2013). By imposing this assumption, the rotation matrix H_0 is a diagonal matrix of ± 1, where the sign is determined by $\tilde{f}'f/T_H$. However, since the true factors are unknown, we still do not know the sign of the rotation matrix.

Assumption A.2 and Assumption A.3 can be found equivalently in GP (2014) (their Assumption 2 and 3, respectively). In Assumption A.2, we allow weak cross-sectional and serial dependence in the idiosyncratic error terms. In Assumption A.3, we impose some moment condition between the factors, idiosyncratic error terms, and the factor loadings. We also allow some weak dependence among them. Due to the MIDAS structure, we also allow some serial dependence between them in Assumption A.4. This assumption is new in the context of the factor-augmented regression models. Without MIDAS structure, Assumption A.3 is sufficient. However, as the factors are temporally aggregated with MIDAS structure, we introduce Assumption A.4.

We impose some weak dependence between idiosyncratic error terms and the regression errors in Assumption A.5. This Assumption is equivalent to the Assumption 4 in GP (2014). Assumption A.6 imposes some moment condition on $\{\varepsilon_t\}$ and the score vector, $g_{\alpha,t}$. Assumption A.6-(b) requires that we can apply a law of large numbers on $\{g_{\alpha,t}g'_{\alpha,t}\}$. By introducing Assumption A.6-(c), we can apply a central limit theorem on $\{g_{\alpha,t}\varepsilon_t\}$. Assumption A.5 and A.6 are same assumptions in GP (2014).

B Proof of results in Section 2

In this section, we prove the asymptotic distribution of NLS estimators in Theorem 2.1 and Theorem 2.2, the consistency of the variance-covariance of the cross-sectional average of the factor loadings and idiosyncratic error term across time for the plug-in bias estimator. To prove the asymptotic distribution, we use the following lemmas.

Lemma B.1 $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(\tilde{F}_t(\theta) - HF_t(\theta)) = o_p(1)$.

Lemma B.2 *If* $\sqrt{T}/N \to c$, *where* $0 \leq c < \infty$,

(a) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{f}_{t-j/m} - HF_{t-j/m})(\tilde{f}_{t-j/m} - Hf_{t-j/m})' = cV^{-1}H\Gamma HV^{-1} + o_p(1)$,

(b) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{f}_{t-j/m} - Hf_{t-j/m})(\tilde{f}_{t-l/m} - Hf_{t-l/m})' = cV^{-1}H\Gamma_{j-l}HV^{-1} + o_p(1)$,

(c) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Hf_{t-j/m}(\tilde{f}_{t-j/m} - Hf_{t-j/m})' = cH\Gamma Q'V^{-2} + o_p(1)$,

(d) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Hf_{t-l/m}(\tilde{f}_{t-j/m} - Hf^{(m)}_{t-j/m})' = cQ_{j-l}\Gamma Q'V^{-2} + o_p(1)$.

Lemma B.3 *If* $\sqrt{T}/N \to c$, *where* $0 \leq c < \infty$,

(a) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{F}_t(\theta) - HF_t(\theta))(\tilde{F}_t(\theta) - HF_t(\theta))'$
$= cV^{-1}Q\Big\{\sum_{k=1}^{K}w_k(\theta)\Gamma w_k(\theta) + \sum_{k=1}^{K}\sum_{l\neq k}^{K}w_k(\theta)\Gamma_{k-l}w_l(\theta)\Big\}Q'V^{-1} + o_p(1)$,

(b) $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{F}_t(\theta) - HF_t(\theta))(HF_t(\theta))'$
$= c\Big\{\sum_{k=1}^{K}w_k^2(\theta)H + \sum_{k=1}^{K}\sum_{l\neq k}^{K}w_k(\theta)Q_{k-l}w_l(\theta)\Big\}\Gamma Q'V^{-2} + o_p(1)$.

Note that we write $F_t(\theta) = \sum_{k=1}^{K}w_k(\theta)f_{t-k/m}$, where $w_k(\theta) \equiv \text{diag}(w_{k,1}(\theta_1), \ldots, w_{k,r}(\theta_r))$ is a $r \times r$ diagonal matrix. We also define $\delta_{NT_H} = \min(\sqrt{N}, \sqrt{T_H})$. We first prove Theorem 2.1 and then we prove Lemmas B.1 - B.3.

Proof of Theorem 2.1. As the NLS estimators $\tilde{\alpha}$ maximizes the objective function $\tilde{Q}_T(\alpha) = -\frac{1}{T}\sum_{t=1}^{T}[y_t - g(\tilde{F}_t, \alpha)]^2$, we have

$$\sqrt{T}(\tilde{\alpha} - \alpha) = -\Big[\frac{1}{T}\sum_{t=1}^{T}H(\tilde{F}_t, \alpha_T)\Big]^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s(\tilde{F}_t, \alpha), \tag{19}$$

where α_T is the intermediate between $\tilde{\alpha}$ and α and $H(\tilde{F}_t, \alpha)$ is a hessian matrix and $s(\tilde{F}_t, \alpha)$ is a score vector. For deriving the asymptotic distribution, we analyse the convergence of each term. We write the term with a score vector as follows.

$$
\begin{aligned}
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} s(\tilde{F}_t, \alpha) &= 2 \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [\varepsilon_t + \beta' H^{-1}(HF_t(\theta) - \tilde{F}_t(\theta))] g_\alpha(\tilde{F}_t, \alpha) \\
&= 2 \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [\varepsilon_t + \beta' H^{-1}(HF_t(\theta) - \tilde{F}_t(\theta))] (\Phi_0 g_\alpha(F_t, \alpha) + P_t),
\end{aligned}
$$

where where $\Phi_0 = \mathrm{diag}(H_0, I_p)$ and $H_0 = \mathrm{plim}\, H$ and P_t is a $(r+p) \times 1$ vector such that

$$
P_t = \begin{bmatrix} \tilde{F}_t(\theta) - HF_t(\theta) \\ \left(\frac{\partial \tilde{F}_t(\theta)}{\partial \theta} H^{-1} - \frac{\partial F_t(\theta)}{\partial \theta} \right)' \beta \end{bmatrix},
$$

with $\frac{\partial \tilde{F}_t(\theta)'}{\partial \theta} = \mathrm{diag}\left(\frac{\partial \tilde{F}_{1,t}(\theta_1)}{\partial \theta_1}, \ldots, \frac{\partial \tilde{F}_{r,t}(\theta_r)}{\partial \theta_r} \right)$ is a $r \times r$ block-diagonal matrix. k-th block is $\partial \tilde{F}_{j,t}(\theta_j)/\partial \theta_k$, which is a $p_j \times 1$ column vector, for $j = 1, \ldots, r$. Under Assumption A.6 and Lemma B.1, we have $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t g_\alpha(\tilde{F}_t, \alpha) \xrightarrow{d} N(0, \Phi_0 \Omega \Phi_0')$. The remaining term drives the bias in Theorem 2.1. As the parameters α contain the slope coefficients, β and the weighting parameters, θ, we take a look into each term. With respect to β, the remaining term is as follows:

$$
\begin{aligned}
&\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{F}_t(\theta) [HF_t(\theta) - \tilde{F}_t(\theta)]' H^{-1'} \beta \\
&= -\Big[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\tilde{F}_t(\theta) - HF_t(\theta))(\tilde{F}_t(\theta) - HF_t(\theta))' + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} HF_t(\theta)(\tilde{F}_t(\theta) - HF_t(\theta))' \Big] H^{-1'} \beta \\
&= -c \Big[V^{-1} H \Big\{ \sum_{k=1}^{K} w_k(\theta) \Gamma w_k(\theta) + \sum_{k=1}^{K} \sum_{l \neq k}^{K} w_k(\theta) \Gamma_{k-l} w_l(\theta) \Big\} H V^{-1} \\
&\qquad + \Big\{ \sum_{k=1}^{K} w_k(\theta) H w_k(\theta) + \sum_{k=1}^{K} \sum_{l \neq k}^{K} w_k(\theta) Q_{k-l} w_l(\theta) \Big\} \Gamma Q' V^{-2} \Big] \, \mathrm{plim}(\tilde{\beta}) \\
&= -c B_\beta + o_p(1), \qquad (20)
\end{aligned}
$$

where $\mathrm{plim}(\tilde{\beta}) = H^{-1'} \beta$. The second equality follows by applying Lemma B.3. Similarly, with respect to θ, we have

$$
\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial \tilde{F}_t(\theta)'}{\partial\theta}H^{-1'}\beta\beta' H^{-1}[HF_t(\theta)-\tilde{F}_t(\theta)] \\
&= -H^{-1'}\beta \circ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{F}_{t,\theta}(\theta)[\tilde{F}_t(\theta)-HF_t(\theta)]' H^{-1'}\beta \\
&= -c\,\mathrm{plim}(\tilde{\beta})\circ\Big[V^{-1}H\Big\{\sum_{k=1}^{K}\frac{\partial w_k(\theta)}{\partial\theta}\Gamma w_k(\theta)+\sum_{k=1}^{K}\sum_{l\neq k}^{K}\frac{\partial w_k(\theta)}{\partial\theta}\Gamma_{k-l}w_l(\theta)\Big\}HV^{-1} \\
&\qquad +\Big\{\sum_{k=1}^{K}\frac{\partial w_k(\theta)}{\partial\theta}Hw_k(\theta)+\sum_{k=1}^{K}\sum_{l\neq k}^{K}\frac{\partial w_k(\theta)}{\partial\theta}Q_{k-l}w_l(\theta)\Big\}\Gamma Q' V^{-2}\Big]\,\mathrm{plim}(\tilde{\beta}) \\
&= -cB_\theta + o_p(1), \qquad (21)
\end{aligned}
$$

where $\tilde{F}_{t,\theta}(\theta)=\Big(\frac{\partial \tilde{F}_{1,t}(\theta_1)}{\partial\theta_1},\ldots,\frac{\partial \tilde{F}_{r,t}(\theta_r)}{\theta_r}\Big)'$. To apply the lemmas, we use the Hadamard product such that $(A\circ B)_{ij}=A_{ij}B_{ij}$. By applying Hadamard product, we have $\frac{\partial \tilde{F}_t(\theta)'}{\partial\theta}H^{-1'}\beta = H^{-1'}\beta\circ\tilde{F}_{t,\theta}(\theta)$ to obtain the first equality. Then, we apply Lemma B.3 for the second equality. Finally, we have $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s(\tilde{F}_t,\alpha)\xrightarrow{d}N(-cB_\alpha,\Phi_0\Omega\Phi_0')$. Next, we derive the term with Hessian matrix. First, we rewrite the first term in (19) as follows:

$$
\frac{1}{T}\sum_{t=1}^{T}H(\tilde{F}_t,\alpha)=\frac{1}{T}\sum_{t=1}^{T}\Big[\varepsilon_t+\beta' H^{-1}(HF_t(\theta)-\tilde{F}_t(\theta))\Big]\frac{\partial^2 g(\tilde{F}_t,\alpha)}{\partial\alpha\partial\alpha'}+\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g(\tilde{F}_t,\alpha)}{\partial\alpha}\frac{\partial g(\tilde{F}_t,\alpha)}{\partial\alpha'}.
$$

Under Assumption Assumption A.6 and Lemma B.1, $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\frac{\partial^2 g(\tilde{F}_t,\alpha)}{\partial\alpha\partial\alpha'}=o_p(1)$. We can also show that $-\frac{1}{T}\sum_{t=1}^{T}\beta' H^{-1}(\tilde{F}_t(\theta)-HF_t(\theta))\frac{\partial^2 g(\tilde{F}_t,\alpha)}{\partial\alpha\partial\alpha'}=o_p(1)$. Finally, for the second term, we have

$$
\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g(\tilde{F}_t,\alpha)}{\partial\alpha}\frac{\partial g(\tilde{F}_t,\alpha)}{\partial\alpha'}=\Phi_0\Sigma\Phi_0'+o_p(1) \qquad (22)
$$

where $\Sigma\equiv E\big[\frac{\partial g(F_t,\alpha)}{\partial\alpha}\frac{\partial g(F_t,\alpha)}{\partial\alpha'}\big]$ by replacing $\frac{\partial g(\tilde{F}_t,\alpha)}{\partial\alpha}$ with $\Phi\frac{\partial g(F_t,\alpha)}{\partial\alpha}+P_t$. Then, by Lemma B.2, we have $\frac{1}{T}\sum_{t=1}^{T}g_\alpha(F_t,\alpha)P_t'=o_p(1)$ and $\frac{1}{T}\sum_{t=1}^{T}P_tP_t'=o_p(1)$. By plugging the terms, (20), (11), and (22) into (19), we have $\sqrt{T}(\tilde{\alpha}-\alpha)\xrightarrow{d}N(-c(\Phi_0\Sigma\Phi_0')^{-1}B_\alpha,\Phi_0'^{-1}\Sigma^{-1}\Omega\Sigma^{-1}\Phi_0^{-1})$. ■

Next, we prove Lemmas we used to prove Theorem 2.1. We can obtain the Lemma B.1 by directly applying the proof of Lemma 1.1 in GP (2014) (the only difference is that we use the high-frequency and low-frequency time indices in our context). The proofs for (a) and (c) in Lemma B.2 are also similar to the proof of Lemma A.2, (a) and (b) in GP (2014). Therefore, we only show the proof for (b) and (d) in Lemma B.2.

Proof of Lemma B.2 - (b).

First, we use the identity for the factor estimation error in GP (2014) such that $\tilde{f}_{t_h} - Hf_{t_h} = \tilde{V}^{-1}\Big(A_{1,t_h} + A_{2,t_h} + A_{3,t_h} + A_{4,t_h}\Big)$, where $A_{1,t_h} = \frac{1}{T_H}\sum_{s_h=1}^{T_H} \tilde{f}_{s_h}\gamma_{s_h t_h}$, $A_{2,t_h} = \frac{1}{T_H}\sum_{s_h}^{T_H} \tilde{f}_{s_h}\zeta_{s_h t_h}$, $A_{3,t_h} = \frac{1}{T_H}\sum_{s_h}^{T_H} \tilde{f}_{s_h}\eta_{s_h t_h}$, and $A_{4,t_h} = \frac{1}{T_H}\sum_{s_h}^{T_H} \tilde{f}_{s_h}\xi_{s_h t_h}$. Each term in A_{i,t_h} for $i = 1,2,3,4$ denotes the following: $\gamma_{s_h t_h} = E\big(\frac{1}{N}\sum_{i=1}^{N} e_{i,s_h}e_{i,t_h}\big)$, $\zeta_{s_h t_h} = \frac{1}{N}\sum_{i=1}^{N}(e_{i,s_h}e_{i,t_h} - E(e_{i,s_h}e_{i,t_h}))$, $\eta_{s_h t_h} = f'_{s_h}\frac{\Lambda' e_{t_h}}{N}$, and $\xi_{s_h t_h} = f'_{t_h}\frac{\Lambda' e_{s_h}}{N} = \eta_{t_h s_h}$. Under this identity and using the low-frequency notation, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{f}_{t-j/m} - Hf_{t-j/m})(\tilde{f}_{t-l/m} - Hf_{t-l/m})' = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big[\tilde{V}^{-1}(A_{1,t-j/m} + A_{2,t-j/m} + A_{3,t-j/m} + A_{4,t-j/m})$$
$$\times (A_{1,t-l/m} + A_{2,t-l/m} + A_{3,t-l/m} + A_{4,t-l/m})'\tilde{V}^{-1}\Big],$$

for $j = 1,\ldots,m-1$. We analyse the convergence limit of each term, respectively. The proof is similar to the proof of Lemma A.2 - (a) in GP (2014). By applying the Cauchy-Schwarz inequality, we have $\|\frac{1}{T}\sum_{t=1}^{T} A_{1,t-j/m}A'_{1,t-l/m}\| \le \big(\frac{1}{T}\sum_{t=1}^{T}\|A_{1,t-j/m}\|^2\big)^{1/2}\big(\frac{1}{T}\sum_{t=1}^{T}\|A_{1,t-l/m}\|^2\big)^{1/2} = O_p(1/T)$, by Assumptions A.1 and A.2. This implies $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} A_{1,t-j/m}A'_{1,t-l/m} = o_p(1)$. We can also show that $\|\frac{1}{T}\sum_{t=1}^{T} A_{2,t-j/m}A'_{2,t-l/m}\| \le \big(\frac{1}{T}\sum_{t=1}^{T}\|A_{2,t-j/m}\|^2\big)^{1/2}\big(\frac{1}{T}\sum_{t=1}^{T}\|A_{2,t-l/m}\|^2\big)^{1/2} = O_p(N^{-1}\delta_{NT_H}^{-2})$ by Cauchy-Schwarz. We also use $\frac{1}{T}\sum_{t=1}^{T}\|A_{2,t-j/m}\|^2 = O_p(N^{-1}\delta_{NT_H}^{-2})$ by Assumption A.2 and $\frac{1}{T_H}\sum_{s_h=1}^{T_H}\|\tilde{f}_s - Hf_s\|^2 = O_p(\delta_{NT_H}^{-2})$ in Bai and Ng (2006). Again, this implies $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} A_{2,t-j/m}A'_{2,t-l/m} = o_p(1)$. Similarly, we can show all the terms are negligible, except the term $\frac{1}{T}\sum_{t=1}^{T} A_{3,t-j/m}A'_{3,t-l/m}$. In fact, this term is $O_p(1/N)$, which is non-negligible when it is multiplied by $\sqrt{T}$ under our assumption, $\sqrt{T}/N \to c$. To see this, we first rewrite the term as follows:

$$\frac{1}{T}\sum_{t=1}^{T} A_{3,t-j/m}A'_{3,t-l/m} = \frac{1}{T}\sum_{t=1}^{T}\Big(\frac{1}{T_H}\sum_{t=1}^{T_H}(\tilde{f}_s - Hf_s + Hf_s)\eta_{s,t-j/m}\Big)\Big(\frac{1}{T_H}\sum_{s=1}^{T_H}(\tilde{f}_s - Hf_s + Hf_s)\eta_{s,t-l/m}\Big)'$$
$$= b_{33.1} + b_{33.2} + b'_{33.2} + b_{33.3}$$

The first term $b_{33.1}$ is bounded by $\big(\frac{1}{T_H}\sum_{s=1}^{T_H}\|\tilde{f}_s - Hf_s\|^2\big)\big(\frac{1}{TT_H}\sum_{t=1}^{T}\sum_{s=1}^{T_H}|\eta_{s,t-j/m}\eta_{s,t-l/m}|\big)$ by applying Cauchy-Schwarz inequality. This is $O_p(N^{-1}\delta_{NT_H}^{-2})$ by $\frac{1}{TT_H}\sum_{t=1}^{T}\sum_{s_h=1}^{T_H}|\eta_{s_h,t-j/m}|^2 = O_p(N^{-1})$ under Assumption A.3. Similarly, the second term is bounded by Cauchy-Schwarz such that $b_{33.2} \le \big(\frac{1}{T_H}\sum_{s=1}^{T_H}\|Hf_s(\tilde{f}_s - Hf_s)\|\big)\big(\frac{1}{TT_H}\sum_{t=1}^{T}\sum_{s=1}^{T_H}|\eta_{s,t-j/m}\eta_{s,t-l/m}|\big) = O_p(N^{-1}\delta_{NT_H}^{-1})$.

Then, the final term is $b_{33.3} = H\Big(\frac{f'f}{T_H}\Big)\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{\Lambda' e_{t-j/m}}{N}\Big)\Big(\frac{e'_{t-l/m}\Lambda}{N}\Big)\Big]\Big(\frac{f'f}{T_H}\Big)H' = O_p\big(\frac{1}{N}\big)$ by Assumption A.3. Thus,

$$\sqrt{T}b_{33.3} = \frac{\sqrt{T}}{N}H\Big[\frac{1}{T}\sum_{t=1}^{T}\Big(\frac{\Lambda' e_{t-j/m}}{\sqrt{N}}\Big)\Big(\frac{e'_{t-l/m}\Lambda}{\sqrt{N}}\Big)\Big]H = cH\Gamma_{j-l}H + o_p(1),$$

where we use $\frac{f'f}{T_H} = I_r$ by Assumptions A.1-(d) and A.4-(d). Finally, we have $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{f}^{(m)}_{t-j/m} - Hf^{(m)}_{t-j/m})(\tilde{f}^{(m)}_{t-l/m} - Hf^{(m)}_{t-l/m})' = cV^{-1}H\Gamma_{j-l}HV^{-1} + o_p(1)$. ∎

Proof of Lemma B.2 - (d). The proof is similar to the proof of Lemma A.2 - (b) in GP (2014). By using the identity we use in the proof of B.2-(b), we have

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Hf_{t-l/m}(\tilde{f}_{t-j/m} - Hf_{t-j/m})' &= H\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_{t-l/m}(A_{1,t-j/m} + A_{2,t-j/m} + A_{3,t-j/m} + A_{4,t-j/m})'\tilde{V}^{-1} \\ &\equiv \sqrt{T}H(d_{f1} + d_{f2} + d_{f3} + d_{f4})\tilde{V}^{-1}.\end{aligned}$$

We show the convergence limit for d_{fi}, for $i = 1,2,3,4$. We can show that all the terms except d_{f4} is negligible. For example, $d_{f1} = O_p(\delta^{-1}_{NT_H}T^{-1/2}) + O_p(T_H^{-1})$. To show this, we first rewrite d_{f1} as $\frac{1}{T}\sum_{t=1}^{T}f_{t-l/m}\big(\frac{1}{T_H}\sum_{s=1}^{T_H}(\tilde{f}_s - Hf_s)'\gamma_{s,t-j/m}\big) + \frac{1}{T}\sum_{t=1}^{T}f_{t-l/m}\big(\frac{1}{T_H}\sum_{s=1}^{T_H}f'_s\gamma_{s,t-j/m}\big)H'$. The first term of d_{f1} is $O_p(\delta^{-1}_{NT_H}T^{-1/2})$ by applying Assumptions A.1-A.2 and $\frac{1}{T_H}\sum_{s_h=1}^{T_H}\|\tilde{f}_s - Hf_s\|^2 = O_p(\delta^{-2}_{NT_H})$. The second term is $O_p(T_H^{-1})$ by Cauchy-Schwarz inequality and Assumptions A.1 and A.2. We can also show that $\|d_{f2}\| = O_p((TN)^{-1/2})$ by showing $\frac{1}{T_H}\sum_{s=1}^{T_H}\|\frac{1}{T}\sum_{t=1}^{T}f_{t-l/m}\zeta_{s,t-j/m}\|^2 = O_p((TN)^{-1})$ under Assumption A.4-(b). The third term is also bounded by Cauchy-Schwarz inequality such that $\|d_{f3}\| = O_p((NT)^{-1/2})$ and by applying Assumption A.4-(c). Finally, we decompose the last term into two parts as follows:

$$\begin{aligned}d_{f4} &= \frac{1}{T}\sum_{t=1}^{T}f_{t-l/m}\Big(\frac{1}{T_H}\sum_{s=1}^{T_H}(\tilde{f}_s - Hf_s)'\xi_{s,t-j/m}\Big) + \frac{1}{T}\sum_{t=1}^{T}f_{t-l/m}\Big(\frac{1}{T_H}\sum_{s=1}^{T_H}f'_s\xi_{s,t-j/m}\Big)H' \\ &\equiv d_{f4.1} + d_{f4.2}.\end{aligned}$$

By rearranging the second term, we have $d_{f4.2} = \frac{1}{\sqrt{T_H N}}\Big(\frac{1}{T}\sum_{s=1}^{T}f_{t-l/m}f'_{t-j/m}\Big)\Big(\frac{1}{\sqrt{T_H N}}\sum_{s=1}^{T_H}\Lambda' e_s f'_s\Big) = O_p(1/(\sqrt{T_H N}))$ by Assumptions A.4-(1) and A.3-(c). We can also rearrange the terms in the first term and write it as follows:

$$d_{f4.1} = \frac{1}{T}\sum_{t=1}^{T}f_{t-l/m}\Big[\frac{1}{T_H}\sum_{s=1}^{T_H}(\tilde{f}_s - Hf_s)'\Big(f'_{t-j/m}\frac{\Lambda' e_s}{N}\Big)\Big] = \Big(\frac{1}{T}\sum_{t=1}^{T}f_{t-l/m}f'_{t-j/m}\Big)\Big(\frac{1}{T_H}\sum_{s=1}^{T_H}\frac{\Lambda' e_s}{N}(\tilde{f}_s - Hf_s)'\Big).$$

This is $O_p(1/N)$ under our assumptions. By using $\frac{1}{T_H}\sum_{s=1}^{T_H}\frac{\Lambda' e_s}{N}(\tilde{f}_s - Hf_s)' = \frac{1}{N}(\Gamma + o_p(1))Q'V^{-1}$, from the proof in GP (2014), we have

$$\sqrt{T}Hd_{f4.1} = H\Big(\frac{1}{T}\sum_{t=1}^{T} f_{t-l/m}f'_{t-j/m}\Big)\Big(\frac{\sqrt{T}}{N}(\Gamma + o_p(1))Q'V^{-1}\Big) = cQ_{j-l}\Gamma Q'V^{-1} + o_p(1)$$

Thus, $\sqrt{T}d_{f4.1}\tilde{V}^{-1} = cQ_{j-l}\Gamma Q'V^{-2} + o_p(1)$, where $Q_{j-l} = \frac{1}{T}\sum_{t=1}^{T}\tilde{f}_{t-j/m}f_{t-l/m} = \frac{1}{T_H}\sum_{t=1}^{T_H}\tilde{f}_t f_{t-(j-l)}$.
■

proof of Lemma B.3 - (a). We write the equation as follows:

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{F}_t(\theta) - F_t(\theta))(\tilde{F}_t(\theta) - HF_t(\theta))' &= \frac{1}{\sqrt{T}}\Big[\sum_{j=0}^{q} w_j(\theta)(\tilde{f}_{t-j/m} - Hf_{t-j/m})\Big]\Big[\sum_{j=0}^{q} w_j(\theta)(\tilde{f}_{t-j/m} - Hf_{t-j/m})\Big]' \\
&= \sum_{j=0}^{q} w_j(\theta)\Big[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{f}_{t-j/m} - Hf_{t-j/m})(\tilde{f}_{t-j/m} - Hf_{t-j/m})'\Big]w_j(\theta) \\
&\quad + \sum_{j=0}^{q}\sum_{l\neq j}^{q} w_j(\theta)\Big[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{f}_{t-j/m} - Hf_{t-j/m})(\tilde{f}_{t-l/m} - Hf_{t-l/m})'\Big]w_l(\theta) \\
&= cV^{-1}Q\Big\{\sum_{j=0}^{q} w_j^2(\theta)\Gamma + \sum_{j=0}^{q}\sum_{l\neq j}^{q} w_j(\theta)\Gamma_{j-l}w_l(\theta)\Big\}Q'V^{-1} + o_p(1).
\end{aligned}$$

By applying the results of Lemmas B.2-(a) and (b), we have the last equality. ■

proof of Lemma B.3 - (b). Similar to previous proof, we write the equation as:

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T} HF_t(\theta)(\tilde{F}_t(\theta) - HF_t(\theta))' &= \frac{1}{\sqrt{T}}\Big[\sum_{j=0}^{q} w_j(\theta)Hf_{t-j/m})\Big]\Big[\sum_{j=0}^{q} w_j(\theta)(\tilde{f}_{t-j/m} - Hf_{t-j/m})\Big]' \\
&= \sum_{j=0}^{q} w_j(\theta)\Big[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Hf_{t-j/m}(\tilde{f}_{t-j/m} - Hf_{t-j/m})'\Big]w_j(\theta) \\
&\quad + \sum_{j=0}^{q}\sum_{l\neq j}^{q} w_j(\theta)\Big[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} Hf_{t-l/m}(\tilde{f}_{t-j/m} - Hf_{t-j/m})'\Big]w_l(\theta) \\
&= c\Big\{\sum_{j=0}^{q} w_j^2(\theta)H + \sum_{j=0}^{q}\sum_{l\neq j}^{q} w_j(\theta)Q_{j-l}w_l(\theta)\Big\}\Gamma Q'V^{-2} + o_p(1).
\end{aligned}$$

By applying Lemmas B.2-(c) and (d), we have the last equality. ■

Next, we prove Theorem 2.2 and 2.1. For proving Theorem 2.2, we first prove when there is no cross-sectional dependence (only serial correlation) in the idiosyncratic term in the factor model, and then we prove when the cross-sectional dependence is added. Note that when the error term is only serially correlated, the estimator for Γ_k is $\tilde{\Gamma}_k = \frac{1}{N}\sum_{i=1}^N \tilde{\lambda}_i\tilde{\lambda}_i'\tilde{e}_{i,t_h}\tilde{e}_{i,t_h-k}$.

proof of Theorem 2.2.

If the idiosyncratic term is only serially correlated, by applying the proof in Bai (2003), we can write the estiamtor as follows

$$\begin{aligned}\tilde{\Gamma}_k &= \frac{1}{N(T_H-k)}\sum_{t_h=k+1}^{T_H}\frac{1}{N}\sum_{i=1}^{N}(\tilde{e}_{i,t_h}\tilde{e}_{i,t_h-k} - e_{i,t_h}e_{i,t_h-k} + e_{i,t_h}e_{i,t_h-k})\tilde{\lambda}_i\tilde{\lambda}_i' \\ &= o_p(1) + (H^{-1})'\Big(\frac{1}{N(T_H-k)}\sum_{t_h=k+1}^{T_H}\frac{1}{N}\sum_{i=1}^{N}e_{i,t_h}e_{i,t_h-k}\lambda_i\lambda_i'\Big)H^{-1}.\end{aligned}$$

Under Assumption A.2 - (d), we have $E(e_{i,t_h}e_{i,t_h-k}) = \sigma_{ii,k}$ and $\frac{1}{N(T_H-k)}\sum_{t_h=k+1}^{T_H}\sum_{i=1}^{N}(\tilde{e}_{i,t_h}\tilde{e}_{i,t_h-k} - e_{i,t_h}e_{i,t_h-k}) = o_p(1)$. Therefore,

$$\frac{1}{T_H-k}\sum_{t_h=k+1}^{T_H}\frac{1}{N}\sum_{i=1}^{N}\tilde{e}_{i,t_h}\tilde{e}_{i,t_h-k}\tilde{\lambda}_i\lambda_i \xrightarrow{p} \Gamma_k,$$

where $\Gamma_k = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\tilde{\lambda}_i\tilde{\lambda}_i'\Big(\frac{1}{T_H-k}\sum_{t_h=k+1}^{T_H}\tilde{e}_{i,t_h}\tilde{e}_{i,t_h-k}\Big)$. When the idiosyncratic term is serially and cross-setionally correlated, the proof is similar to the proof for Theorem 4 in Bai and Ng (2006). Under Assumption A.2 - (d), we have $\sigma_{ij,k} = E(e_{i,t_h}e_{j,t_h-k})$. Let $\tilde{\sigma}_{ij,k} = \frac{1}{T_H-k}\sum_{t_h=k+1}^{T_H}\tilde{e}_{i,t_h}\tilde{e}_{i,t_h-k}$ and $\Gamma_{n,k} = \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\sigma_{ij,k}\lambda_i\lambda_j'$. By the definition, $\Gamma_k = \lim_{n\to\infty}\Gamma_{n,k}$. Let $\bar{\Gamma}_{n,k} = \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\tilde{\sigma}_{ij,k}\lambda_i\lambda_j'$. Then, we can write

$$\tilde{\Gamma}_k - H^{-1'}\Gamma_k H^{-1} = \tilde{\Gamma}_k - H^{-1'}\bar{\Gamma}_{n,k}H^{-1} + H^{-1'}(\bar{\Gamma}_{n,k} - \Gamma_{n,k})H^{-1} + H^{-1'}(\Gamma_{n,k} - \Gamma_k)H^{-1}.$$

Since Γ_k is the limit of $\Gamma_{n,k}$, we have $\Gamma_{n,k} - \Gamma_k \to 0$. The remaining part to show is that $\bar{\Gamma}_{n,k} - \Gamma_{n,k} \xrightarrow{p} 0$ if $n/N \to 0$ and $n/T_H \to 0$, and $\tilde{\Gamma}_k - H^{-1'}\bar{\Gamma}_{n,k}H^{-1} \xrightarrow{p} 0$. For the first part, we rewrite it as follows:

$$
\begin{aligned}
\bar{\Gamma}_{n,k} - \Gamma_{n,k} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(\tilde{\sigma}_{ij,k} - \sigma_{ij,k})\lambda_i\lambda_j' \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{T_H - k}\sum_{t_h=k+1}^{T_H}(e_{i,t_h}e_{j,t_h-k} - \sigma_{ij,k})\lambda_i\lambda_j' - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{T_H - k}\sum_{t_h=k+1}^{T_H} e_{i,t_h}(c_{j,t_h-k} - \tilde{c}_{j,t_h-k})\lambda_i\lambda_j' \\
&\quad - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{T_H - k}\sum_{t_h=k+1}^{T_H} e_{j,t_h-k}(c_{i,t_h} - \tilde{c}_{i,t_h})\lambda_i\lambda_j' \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{T_H - k}\sum_{t_h=k+1}^{T_H}(c_{i,t_h} - \tilde{c}_{i,t_h})(c_{j,t_h-k} - \tilde{c}_{j,t_h-k})\lambda_i\lambda_j' \\
&= I + II + III + IV,
\end{aligned}
$$

where we obtain the second equality by $\tilde{e}_{i,t_h}\tilde{e}_{j,t_h-k} = e_{i,t_h}e_{j,t_h-k} - e_{i,t_h}(c_{j,t_h-k} - \tilde{c}_{j,t_h-k}) - e_{j,t_h-k}(c_{i,t_h} - \tilde{c}_{i,t_h}) + (c_{i,t_h} - \tilde{c}_{i,t_h})(c_{j,t_h-k} - \tilde{c}_{j,t_h-k})$. We can show that I is $O_p((T_H - k)^{-1/2})$ since it is zero mean process. For, II, by using $c_{j,t_h} - \tilde{c}_{j,t_h} = (H^{-1'}\lambda_j - \tilde{\lambda}_j)'\tilde{f}_{t_h} + \lambda_j' H^{-1}(Hf_{t_h} - \tilde{f}_{t_h})$ and we can decompose it into two parts. Then, following Bai and Ng (2006), we have $II \to 0$ if $\sqrt{n}/T_H \to 0$ and $n/\delta^2_{NT_H} \to 0$. Similarly, we have $III \to 0$ as $n/\delta^2_{NT_H} \to 0$. Finally, for IV, by Cauchy-Schwarz inequality, we have

$$
|IV| \leq \Big(\frac{1}{T_H - k}\sum_{t_h=k+1}^{T_H}\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(c_{i,t_h} - \tilde{c}_{i,t_h})\lambda_i\|^2\Big)^{1/2}\Big(\frac{1}{T_H - k}\sum_{t_h=k+1}^{T_H}\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(c_{j,t_h-k} - \tilde{c}_{j,t_h-k})\lambda_j\|^2\Big)^{1/2}.
$$

By using $c_{i,t_h} - \tilde{c}_{i,t_h} = (H^{-1}\lambda_i - \tilde{\lambda}_i)'\tilde{f}_{t_h} + \lambda_i' H^{-1}(Hf_{t_h} - \tilde{f}_{t_h})$, we have

$$
\begin{aligned}
\frac{1}{T_H - k}\sum_{t_h=k+1}^{T_H}\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(c_{i,t_h} - \tilde{c}_{i,t_h})\lambda_i\|^2 &\leq 2\Big(\frac{1}{T_H - k}\sum_{t_h=k+1}^{T_H}\|f_{t_h}\|^2\Big)\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\lambda_i(H^{-1'}\lambda_i - \tilde{\lambda}_i)'\|^2 \\
&\quad + 2\|H^{-1}\|^2\Big(\frac{1}{n}\sum_{i=1}^{n}\|\lambda_i\|^2\Big)^2 n\frac{1}{T_H - k}\sum_{t_h=k+1}^{T_H}\|\tilde{f}_{t_h} - Hf_{t_h}\|^2.
\end{aligned}
$$

The first part $\to 0$ as $\sqrt{n}/T \to 0$ and the second part $\to 0$ as $n/T_H \to 0$. The last remaining term is $\tilde{\Gamma}_k - H^{-1'}\bar{\Gamma}_{n,k}H^{-1}$. We can rewrite this term as follows:

$$\begin{aligned}\tilde{\Gamma}_k - H^{-1'}\bar{\Gamma}_{n,k}H^{-1} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\tilde{\sigma}_{ij,k}(\tilde{\lambda}_i\tilde{\lambda}_j' - H^{-1'}\lambda_i\lambda_j'H^{-1}) \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(\tilde{\sigma}_{ij,k} - \sigma_{ij,k})(\tilde{\lambda}_i\tilde{\lambda}_j' - H^{-1'}\lambda_i\lambda_j'H^{-1}) + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij,k}(\tilde{\lambda}_i\tilde{\lambda}_j' - H^{-1'}\lambda_i\lambda_j'H^{-1}) \\ &= I + II.\end{aligned}$$

Then, $I \to 0$ using the fact that it is zero mean process. For the second part, we can write as:

$$II = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}(\tilde{\lambda}_i - H^{-1}\lambda_i)\tilde{\lambda}_j' + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\lambda_i H^{-1}(\tilde{\lambda}_j - H^{-1'}\lambda_j)' = a + b.$$

Then, $a \to 0$ since $a = O_p(T_H^{-1/2}) + O_p(\delta_{NT_H}^{-2})$ and $b \to 0$ as $b = O_p(T^{-1/2}) + O_p(\delta_{NT_H}^{-2})$. ■

The proof for Proposition 2.1 is straightforward by applying Theorem 2.2.

C Proof of results in Section 3

Notation: P^* denotes the bootstrap probability measure, conditional on the original sample. The bootstrap measure P^* depends on the original sample size N, T and T_H, and sample realization ω, but for a simpler notation, we omit these and write P^* for $P^*_{NT,\omega}$. We write $T^*_{NT} = o_{p^*}(1)$, in probability, or $T^*_{NT} \xrightarrow{p*} 0$, in probability, for any bootstrap test statistics T^*_{NT}, if, when for any $\delta > 0$, $P^*(|T^*_{NT}| > \delta) = o_p(1)$. If for all $\delta > 0$, there exists $M_\delta < \infty$ such that $\lim_{N,T\to\infty} P[P^*(|T^*_{NT}| > M_\delta) > \delta] = 0$, we write as $T^*_{NT} = O_{p^*}(1)$, in probability. We write $T^*_{NT} \xrightarrow{d*} D$, in probability, if T^*_{NT} weakly converges to the distribution D under P^*, conditional on a sample with probability that converges to one, i.e. $E^*(f(T^*_{NT})) \xrightarrow{p} E(f(D))$ for all bounded and uniformly continuous function f.

Condition C.1* (*Weak Time Series and Cross Section Dependence in* $e^*_{it_h}$)

(a) $E^*(e^*_{i,t_h}) = 0$ *for all* (i, t_h).

(b) $\frac{1}{T_H}\sum_{t_h=1}^{T_H}\sum_{s_h=1}^{T_H}|\gamma^*_{s_h t_h}|^2 = O_p(1)$, *where* $\gamma^*_{s_h t_h} = E^*\big(\frac{1}{N}\sum_{i=1}^{N} e^*_{i,t_h}e^*_{i,s_h}\big)$.

(c) $\frac{1}{T_H^2}\sum_{t_h=1}^{T_H}\sum_{s_h=1}^{T_H} E^*\big|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(e^*_{i,t_h}e^*_{i,s_h} - E^*(e^*_{i,t_h}e^*_{i,s_h}))\big|^2 = O_p(1)$.

Condition C.2* (*Weak Dependence Among* $\tilde{f}_{t_h}$, $\tilde{\lambda}_i$, *and* $\tilde{e}^*_{i,t_h}$)

(a) $\frac{1}{T_H}\sum_{t_h=1}^{T_H}\sum_{s_h=1}^{T_H}\tilde{f}_{s_h}\tilde{f}'_{t_h}\gamma^*_{s_h t_h} = O_p(1)$.

(b) $\frac{1}{T_H}\sum_{t_h=1}^{T_H} E^*\|\frac{1}{\sqrt{T_H N}}\sum_{s_h=1}^{T_H}\sum_{i=1}^{N}\tilde{f}_{s_h}(e^*_{i,t_h}e^*_{i,s_h} - E^*(e^*_{i,t_h}e^*_{i,s_h}))\|^2 = O_p(1)$.

(c) $E^*\|\frac{1}{\sqrt{T_H N}}\sum_{t_h=1}^{T_H}\sum_{i=1}^{N}\tilde{f}_{t_h}\tilde{\lambda}'_i e^*_{i,t_h}\|^2 = O_p(1)$.

(d) $\frac{1}{T_H}\sum_{t_h=1}^{T_H} E^*\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\lambda}_i e^*_{i,t_h}\|^2 = O_p(1)$.

(e) $\frac{1}{T_H}\sum_{t_h=1}^{T_H}\big(\frac{\tilde{\Lambda}e^*_{t_h}}{\sqrt{N}}\big)\big(\frac{e^{*\prime}_{t_h}\tilde{\Lambda}}{\sqrt{N}}\big) - \tilde{\Gamma} = o_{p^*}(1)$, *in probability, where* $\tilde{\Gamma} \equiv \frac{1}{T_H}\sum_{t_h=1}^{T_H} Var^*\big(\frac{1}{\sqrt{N}}\tilde{\Lambda}' e^*_{t_h}\big) > 0$.

Condition C.3* (***Serial Dependence among*** $\tilde{f}_{t_h}$, $\tilde{\lambda}_i$, ***and*** $\tilde{e}^*_{i,t_h}$)

(a) $\frac{1}{T_H}\sum_{t_h=1}^{T_H} E^*\|\frac{1}{\sqrt{T_H N}}\sum_{s_h=1}^{T_H}\sum_{i=1}^{N}\tilde{f}_{s_h}(e^*_{i,t_h}e^*_{i,s_h-k} - E^*(e^*_{i,t_h}e^*_{i,s_h-k}))\|^2 = O_p(1)$ *for all* k.

(b) $E^*\|\frac{1}{\sqrt{T_H N}}\sum_{t_h=1}^{T_H}\tilde{f}_{t_h}e^{*\prime}_{t_h-k}\tilde{\Lambda}\|^2 = O_p(1)$ *for all* k.

(c) $\frac{1}{T_H}\sum_{t_h=1}^{T_H}\big(\frac{\tilde{\Lambda}e^*_{t_h}}{\sqrt{N}}\big)\big(\frac{e^{*\prime}_{t_h-k}\tilde{\Lambda}}{\sqrt{N}}\big) - \tilde{\Gamma}_k = o_{p^*}(1)$, *in probability, where* $\tilde{\Gamma}_k \equiv \frac{1}{T_H}\sum_{t_h=1}^{T_H} Cov^*\big(\frac{\tilde{\Lambda}' e^*_{t_h}}{\sqrt{N}}, \frac{\tilde{\Lambda}' e^*_{t_h-k}}{\sqrt{N}}\big) > 0$.

Condition C.4* (***Weak Dependence Between*** e^*_{i,t_h} ***and*** ε^*_t)

(a) $\frac{1}{T}\sum_{t=1}^{T} E^*\big|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T}\sum_{i=1}^{N}\varepsilon^*_s(e^*_{i,t-j/m}e^*_{i,s-j/m} - E^*(e^*_{i,t-j/m}e^*_{i,s-j/m}))\big|^2 = O_p(1)$ *for* $j = 0, \ldots, m-1$.

(b) $E^*\|\frac{1}{\sqrt{TN}}\sum_{t=1}^{T}\sum_{i=1}^{N}\tilde{\lambda}_i e^*_{i,t-j/m}\varepsilon^*_t\|^2 = O_p(1)$, *where* $E(e^*_{i,t-j/m}) = 0$ *for all* (i,t) *and* $j = 0, \ldots, m-1$.

Condition C.5* (***Bootstrap CLT***)

(a) $E^*(\varepsilon^*_t) = 0$ *and* $\frac{1}{T}\sum_{t=1}^{T} E^*|\varepsilon_t|^2 = O_p(1)$.

(b) $\tilde{\Omega}^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{g}_{\alpha,t}\varepsilon^*_t \xrightarrow{d*} N(0, I_{r+p})$, *in probability, where* $E^*\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{g}_{\alpha,t}\varepsilon^*_t\|^2 = O_p(1)$ *and* $\tilde{g}_{\alpha,t} = \partial g(\tilde{F}_t;\alpha)/\partial\alpha$, *and* $\tilde{\Omega} \equiv Var^*\big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{g}_{\alpha,t}\varepsilon^*_t\big) > 0$.

Conditions C.1* to Condition C.4* are the bootstrap analogue of Assumptions Assumption A.1 to Assumption A.6 in Appendix A. Condition C.1*-C.2* are similar to the bootstrap high level conditions in GP (2014). The mean of bootstrap residuals are required to be zeros for all (i, t_h) and t, which implies that we need to recenter the residuals when we resample them. Unlike in GP

(2014), since our bias contains the serial dependence, we impose weak serial dependence among $\tilde{f}_{t_h}$, $\tilde{\lambda}_i$ and $\tilde{e}^*_{i,t_h}$ in Condition C.3*. Note that since $\tilde{f}_{t_h}$ and $\tilde{\lambda}_i$ are fixed, serial dependence in the factors can be implied by restricting the serial dependence of e_{i,t_h}. Condition C.4* is also similar to GP (2014), where we restrict the dependence between two bootstrap residuals. Finally, Condition C.5* implies that we can apply a central limit theorem on the score vector, $\tilde{g}_{\alpha,t}\varepsilon^*_t$.

Lemma C.1 $\frac{1}{T}\sum_{t=1}^T \varepsilon^*_t(\tilde{f}^*_{t-j/m} - H^*\tilde{f}_{t-j/m}) = o_{p^*}(1)$.

Lemma C.2 *If* $\sqrt{T}/N \to c$, *where* $0 \le c < \infty$,

(a) $\frac{1}{\sqrt{T}}\sum_{t=1}^T(\tilde{f}^*_{t-j/m} - H^*\tilde{f}^{(m)}_{t-j/m})(\tilde{F}^*_{t-j/m} - H^*\tilde{f}_{t-j/m})' = \frac{\sqrt{T}}{N}\tilde{V}^{*-1}H^*\tilde{\Gamma}H^*\tilde{V}^{*-1} + o_{p^*}(1)$,

(b) $\frac{1}{\sqrt{T}}\sum_{t=1}^T(\tilde{f}^*_{t-j/m} - H^*\tilde{f}_{t-j/m})(\tilde{f}^*_{t-l/m} - H^*\tilde{f}_{t-l/m})' = \frac{\sqrt{T}}{N}\tilde{V}^{*-1}H^*\tilde{\Gamma}_{j-l}H^*\tilde{V}^{*-1} + o_{p^*}(1)$,

(c) $\frac{1}{\sqrt{T}}\sum_{t=1}^T H^*\tilde{f}_{t-j/m}(\tilde{f}^*_{t-j/m} - H^*\tilde{f}_{t-j/m})' = \frac{\sqrt{T}}{N}H^*\tilde{\Gamma}\Big(\frac{1}{T_H}\sum_{s_h=1}^{T_H}\tilde{f}_{s_h}\tilde{f}^{*\prime}_{s_h}\Big)\tilde{V}^{*-2} + o_{p^*}(1)$,

(d) $\frac{1}{\sqrt{T}}\sum_{t=1}^T H^*\tilde{f}_{t-l/m}(\tilde{f}^*_{t-j/m} - H^*\tilde{f}_{t-j/m})' = \frac{\sqrt{T}}{N}H^*\Big(\frac{1}{T}\sum_{t=1}^T\tilde{f}_{t-l/m}\tilde{f}'_{t-j/m}\Big)\tilde{\Gamma}\Big(\frac{1}{T_H}\sum_{s_h=1}^{T_H}\tilde{f}_{s_h}\tilde{f}^{*\prime}_{s_h}\Big)\tilde{V}^{*-2} + o_{p^*}(1)$.

Lemma C.3 *If* $\sqrt{T}/N \to c$, *where* $0 \le c < \infty$,

(a) $\frac{1}{\sqrt{T}}\sum_{t=1}^T(\tilde{F}^*_t(\tilde{\theta}) - H^*\tilde{F}_t(\tilde{\theta}))(\tilde{F}^*_t(\tilde{\theta}) - H^*\tilde{F}_t(\tilde{\theta}))'$
$= cH^*_0\tilde{V}^{-1}\Big(\sum_{j=0}^q w_j(\tilde{\theta})\tilde{\Gamma}w_j(\tilde{\theta}) + \sum_{j=0}^q w_j(\tilde{\theta})\tilde{\Gamma}_{j-l}w_l(\tilde{\theta})\Big)\tilde{V}^{-1}H^*_0 + o_{p^*}(1)$,

(b) $\frac{1}{\sqrt{T}}\sum_{t=1}^T H^*\tilde{F}_t(\tilde{\theta})(\tilde{F}^*_t(\tilde{\theta}) - H^*\tilde{F}_t(\tilde{\theta}))'$
$= cH^*_0\Big[\sum_{j=0}^q w_j^2(\tilde{\theta}) + \sum_{j=0}^q\sum_{l\neq j}^q w_j(\tilde{\theta})\Big(\frac{1}{T}\sum_{t=1}^T\tilde{f}_{t-l/m}\tilde{f}'_{t-j/m}\Big)w_l(\tilde{\theta})\Big]\tilde{\Gamma}\tilde{V}^{-2}H^*_0 + o_{p^*}(1)$.

proof of Theorem D.1. Since in the bootstrap world, $\tilde{\alpha}^*$ maximizes the following objective function:

$$\tilde{Q}^*_T(\tilde{\alpha}) = -\frac{1}{T}\sum_{t=1}^T[y_t - g(\tilde{F}^*_t;\tilde{\alpha})]^2.$$

where $g(\tilde{F}^*_t;\tilde{\alpha}) = \tilde{\beta}'H^{*-1}\tilde{F}^*_t(\tilde{\theta})$. Then, we have

$$\sqrt{T}(\tilde{\alpha}^* - (\Phi^*_0)^{-1}\tilde{\alpha}) = -\Big[\frac{1}{T}\sum_{t=1}^T H(\tilde{F}^*_t;\tilde{\alpha}_T)\Big]^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^T s(\tilde{F}^*_t;\tilde{\alpha}),$$

where $s(\tilde{F}^*_t;\tilde{\alpha})$ is a score vector and $H(\tilde{F}^*_t;\tilde{\alpha})$ is a Hessian matrix in the bootstrap world. $\tilde{\alpha}_T$ is intermediate between $\tilde{\alpha}$ and $\tilde{\alpha}^*$. We analyse each term.

1.

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}s(\tilde{F}_t^*;\tilde{\alpha})=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\varepsilon_t^*+\tilde{\beta}'H^{*-1}(H^*\tilde{F}_t(\tilde{\theta})-\tilde{F}_t^*(\tilde{\theta}))]\frac{\partial g(\tilde{F}_t^*;\tilde{\alpha})}{\partial\tilde{\alpha}}$$

We can write the partial derivative as:

$$\frac{\partial g(\tilde{F}_t^*;\tilde{\alpha})}{\partial\tilde{\alpha}}=\Phi^*\frac{\partial g(\tilde{F}_t;\alpha)}{\partial\alpha}+P_t^* \text{ where } P_t^*=\begin{bmatrix}\tilde{F}_t^*(\tilde{\theta})-H^*\tilde{F}_t(\theta)\\ (\frac{\partial\tilde{F}_t^*(\tilde{\theta})'}{\partial\tilde{\theta}}H^{*-1'}\tilde{\beta}-\frac{\partial\tilde{F}_t(\theta)'}{\partial\theta}H^{-1'}\beta)\end{bmatrix}$$

where $\Phi^*=diag(H^*,I_p)$. Using this decomposition, we can analyse $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t^*\frac{\partial g(\tilde{F}_t^*;\tilde{\alpha})}{\partial\tilde{\alpha}}$ into two parts:

(a)

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t^*(\tilde{F}_t^*(\tilde{\theta})-H^*\tilde{F}_t(\theta))&=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t^*\Big[\sum_{j=0}^{q}w_j(\tilde{\theta})(\tilde{F}_{t-j/m}^*-H^*\tilde{F}_{t-j/m})+\sum_{j=0}^{q}(w_j(\tilde{\theta})-w_j(\theta))H^*\tilde{F}_{t-j/m}\Big]\\&=\sum_{j=0}^{q}w_j(\tilde{\theta})\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t^*(\tilde{F}_{t-j/m}^*-H^*\tilde{F}_{t-j/m})\\&\quad+\sum_{j=0}^{q}(w_j(\tilde{\theta})-w_j(\theta))H^*\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t^*\tilde{F}_{t-j/m}\\&=o_{p^*}(1).\end{aligned}$$

Since $\tilde{\theta}\xrightarrow{p}\theta$ and assuming that the weighting function is continuous function, we can use continuous mapping theorem and have the second part as $o_p(1)$. The first part is $o_{p^*}(1)$ because of Lemma C.1.

(b) The second part can be argued similarly. It is easier to check for each row. k-th row in the second part is $(\frac{\partial\tilde{F}_{k,t}^*(\tilde{\theta}_k)}{\partial\tilde{\theta}_k}H_k^{*-1'}\tilde{\beta}_k-\frac{\partial\tilde{F}_{k,t}(\theta_k)}{\partial\theta_k}H_k^{-1'}\beta_k)$. Then, for this k-th row, we can write as:

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t^*(\frac{\partial\tilde{F}_{k,t}^*(\tilde{\theta}_k)}{\partial\tilde{\theta}_k}\tilde{\beta}_k-\frac{\partial\tilde{F}_{k,t}(\theta_k)}{\partial\theta_k}\beta_k)&=H_k^{*-1'}\tilde{\beta}_k\Big[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t^*\sum_{j=0}^{q}\frac{\partial w_{j,k}(\tilde{\theta}_k)}{\partial\tilde{\theta}_k}(\tilde{F}_{k,t-j/m}^*-H_k^*\tilde{F}_{k,t-j/m})\\&\quad+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t^*\sum_{j=0}^{q}\Big\{\frac{\partial w_{j,k}(\tilde{\theta}_k)}{\partial\tilde{\theta}_k}-\frac{\partial w_{j,k}(\theta_k)}{\partial\theta_k}\Big\}\tilde{F}_{k,t-j/m}\Big]\\&\quad+(\tilde{\beta}_k-H_k^{-1'}\beta_k)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t^*\Big[\sum_{j=0}^{q}\frac{\partial w_{j,k}(\theta_k)}{\partial\theta_k}\tilde{F}_{k,t-j/m}\Big]\\&=o_{p^*}(1)\end{aligned}$$

where H_k is the k-th diagonal element in the rotation matrix, H and β_k is the k-th slope parameter in β. We have the second equality because $\tilde{\beta}\xrightarrow{p}H^{-1'}\beta$ and Lemma C.1.

Therefore, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t^*\frac{\partial g(\tilde{F}_t;\tilde{\alpha})}{\partial\tilde{\alpha}}\xrightarrow{d^*}N(0,\Phi_0^*\tilde{\Omega}\Phi_0^*) \quad (23)$$

where $\Phi_0^*=\operatorname{plim}\Phi^*$ and $\tilde{\Omega}\equiv\operatorname{Var}^*(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t^*\tilde{g}_{\alpha,t})$ and $\tilde{g}_{\alpha,t}=\partial g(\tilde{F}_t;\alpha)/\partial\alpha$.

Now, we analyse the second term in the score vector $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{\beta}'H^{*-1}(H^*\tilde{F}_t(\tilde{\theta})-\tilde{F}_t^*(\tilde{\theta}))\frac{\partial g(\tilde{F}_t^*;\alpha)}{\partial\tilde{\alpha}}$ into two parts: with respect to $\tilde{\beta}$ and $\tilde{\theta}$, respectively.

(a) with respect to $\tilde{\beta}$:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(H^*\tilde{F}_t(\tilde{\theta})-\tilde{F}_t^*(\tilde{\theta}))\tilde{F}_t^*(\tilde{\theta})'H^{*-1'}\tilde{\beta}$$

$$=-\Big[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{F}_t^*(\tilde{\theta})-H^*\tilde{F}_t(\tilde{\theta}))(\tilde{F}_t^*(\tilde{\theta})-H^*\tilde{F}_t(\tilde{\theta}))'+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}H^*\tilde{F}_t(\tilde{\theta})(\tilde{F}_t^*(\tilde{\theta})-H^*\tilde{F}_t(\tilde{\theta}))'\Big]H^{*-1'}\tilde{\beta}$$

$$=-cH_0^*\Big[\tilde{V}^{-1}\Big\{\sum_{j=0}^{q}w_j(\tilde{\theta})\tilde{\Gamma}w_j(\tilde{\theta})+\sum_{j=0}^{q}\sum_{l\neq j}^{q}w_j(\tilde{\theta})\tilde{\Gamma}_{j-l}w_l(\tilde{\theta})\Big\}\tilde{V}^{-1}$$

$$+\Big\{\sum_{j=0}^{q}w_j^2(\tilde{\theta})+\sum_{j=0}^{q}\sum_{l\neq j}^{q}w_j(\tilde{\theta})\Big(\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_{t-j/m}\tilde{F}_{t-l/m}'\Big)w_l(\tilde{\theta})\Big\}\tilde{\Gamma}\tilde{V}^{-2}\Big]\tilde{\beta}$$

$$=-cH_0^*\tilde{B}_\beta+o_{p^*}(1)$$

in probability, where $\tilde{B}_\beta\equiv\Big[\tilde{V}^{-1}\Big\{\sum_{j=0}^{q}w_j(\tilde{\theta})\tilde{\Gamma}w_j(\tilde{\theta})+\sum_{j=0}^{q}\sum_{l\neq j}^{q}w_j(\tilde{\theta})\tilde{\Gamma}_{j-l}w_l(\tilde{\theta})\Big\}\tilde{V}^{-1}+\Big\{\sum_{j=0}^{q}w_j^2(\tilde{\theta})+\sum_{j=0}^{q}\sum_{l\neq j}^{q}w_j(\tilde{\theta})\Big(\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_{t-j/m}\tilde{F}_{t-l/m}'\Big)w_l(\tilde{\theta})\Big\}\tilde{\Gamma}\tilde{V}^{-2}\Big]\tilde{\beta}$.

(b) with respect to $\tilde{\theta}$:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial\tilde{F}_t^*(\tilde{\theta})}{\partial\tilde{\theta}}H^{*-1'}\tilde{\beta}\tilde{\beta}'H^{*-1}[H^*\tilde{F}_t(\tilde{\theta})-\tilde{F}_t^*(\tilde{\theta})]$$

$$=-c\tilde{\beta}\circ\Big[\tilde{V}^{-1}\Big\{\sum_{j=0}^{q}\frac{\partial w_j(\tilde{\theta})}{\partial\tilde{\theta}}\tilde{\Gamma}w_j(\tilde{\theta})+\sum_{j=0}^{q}\sum_{l\neq j}^{q}\frac{\partial w_j(\tilde{\theta})}{\partial\tilde{\theta}}\tilde{\Gamma}_{j-l}w_l(\tilde{\theta})\Big\}\tilde{V}^{-1}$$

$$+\Big\{\sum_{j=0}^{q}\frac{\partial w_j(\tilde{\theta})}{\partial\tilde{\theta}}w_j(\tilde{\theta})+\sum_{j=0}^{q}\sum_{l\neq j}^{q}\frac{\partial w_j(\tilde{\theta})}{\partial\tilde{\theta}}\Big(\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_{t-j/m}\tilde{F}_{t-l/m}'\Big)w_l(\tilde{\theta})\Big\}\tilde{\Gamma}\tilde{V}^{-2}\Big]\tilde{\beta}$$

$$=-c\tilde{B}_\theta+o_{p^*}(1),$$

in probability, where $\tilde{B}_\theta\equiv\tilde{\beta}\circ\Big[\tilde{V}^{-1}\Big\{\sum_{j=0}^{q}\frac{\partial w_j(\tilde{\theta})}{\partial\tilde{\theta}}\tilde{\Gamma}w_j(\tilde{\theta})+\sum_{j=0}^{q}\sum_{l\neq j}^{q}\frac{\partial w_j(\tilde{\theta})}{\partial\tilde{\theta}}\tilde{\Gamma}_{j-l}w_l(\tilde{\theta})\Big\}\tilde{V}^{-1}+\Big\{\sum_{j=0}^{q}\frac{\partial w_j(\tilde{\theta})}{\partial\tilde{\theta}}w_j(\tilde{\theta})+\sum_{j=0}^{q}\sum_{l\neq j}^{q}\frac{\partial w_j(\tilde{\theta})}{\partial\tilde{\theta}}\Big(\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_{t-j/m}\tilde{F}_{t-l/m}'\Big)w_l(\tilde{\theta})\Big\}\tilde{\Gamma}\tilde{V}^{-2}\Big]\tilde{\beta}$.

2.

$$\frac{1}{T}\sum_{t=1}^{T}H(\tilde{F}_t^*;\tilde{\alpha})=\frac{1}{T}\sum_{t=1}^{T}\xi_t\frac{\partial^2 g(\tilde{F}_t^*;\tilde{\alpha})}{\partial\tilde{\alpha}\partial\tilde{\alpha}'}-\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g(\tilde{F}_t^*;\tilde{\alpha})}{\partial\tilde{\alpha}}\frac{\partial g(\tilde{F}_t^*;\tilde{\alpha})}{\partial\tilde{\alpha}'}$$

Then, the first term is $o_{p^*}(1)$ by Condition C.5*-(b) and the results in the proof for Lemma C.2. The second term converges in probability as following:

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g(\tilde{F}_t^*;\tilde{\alpha})}{\partial\tilde{\alpha}}\frac{\partial g(\tilde{F}_t^*;\tilde{\alpha})}{\partial\tilde{\alpha}'}\xrightarrow{p^*}\Phi_0^*E\Big[\frac{\partial g(\tilde{F}_t;\alpha)}{\partial\alpha}\frac{\partial g(\tilde{F}_t;\alpha)}{\partial\alpha'}\Big]\Phi_0^*\equiv\Phi_0^*\tilde{\Sigma}\Phi_0^* \tag{24}$$

where $E\Big[\frac{\partial g(\tilde{F}_t;\alpha)}{\partial\alpha}\frac{\partial g(\tilde{F}_t;\alpha)}{\partial\alpha'}\Big]\equiv\tilde{\Sigma}$. We can obtain this by rewriting $\frac{\partial g(\tilde{F}_t^*;\tilde{\alpha})}{\partial\tilde{\alpha}}=\Phi^*\frac{\partial g(\tilde{F}_t;\alpha)}{\partial\alpha}+P_t^*$. Then, $\frac{1}{T}\sum_{t=1}^{T}\frac{\partial g(\tilde{F}_t;\tilde{\alpha})}{\partial\tilde{\alpha}}P_t^{*'}=o_{p^*}(1)$, in probability and $\frac{1}{T}\sum_{t=1}^{T}P_t^*P_t^{*'}=o_{p^*}(1)$, in probability.

Thus, we have

$$\sqrt{T}(\tilde{\alpha}^*-(\Phi_0^*)^{-1}\tilde{\alpha})\xrightarrow{d^*}N(-c(\Phi_0^*\tilde{\Sigma}\Phi_0^*)^{-1}\Phi_0^*\tilde{B}_\alpha,\Phi_0^{*-1}\tilde{\Sigma}^{-1}\tilde{\Omega}\tilde{\Sigma}^{-1}\Phi_0^{*-1}), \tag{25}$$

in probability, where $\tilde{B}_\alpha = (\tilde{B}_\beta, \tilde{B}_\theta)'$. Under Assumption A.1-Assumption A.6, $\text{plim}\, \tilde{V} = V$, $\text{plim}\, \tilde{\alpha} = \Phi^{-1}\alpha$, $\text{plim}\, \tilde{\Gamma} = H\Gamma H$, $\text{plim}\, \tilde{\Gamma}_{j-l} = H\Gamma_{j-l}H$, $\text{plim}\, \Phi^* = \Phi_0^*$ and $\text{plim}\, \tilde{\Omega} = \Phi_0\Omega\Phi_0$ implies that $\sqrt{T}(\tilde{\alpha}^* - (\Phi_0^*)^{-1}\tilde{\alpha}) \xrightarrow{d^*} N(-c\Phi^{*-1}\Delta_\alpha, \Phi_0^{*-1}\Sigma_\alpha\Phi_0^{*-1})$, in probability.

■

The proof for Lemma D.1 is similar to the proof of Lemma B.2 in GP (2014) and Lemma D.2 - (a) and (c) are similar to the proof of Lemma B.3 - (a) and (b) in GP (2014), respectively. Thus, we omit the proof here and only show the proof for (b) and (d) below.

proof of Lemma C.2.

(proof of part(b))

$$\frac{1}{T}\sum_{t=1}^{T}(\tilde{F}^*_{t-j/m} - H^*\tilde{F}_{t-j/m})(\tilde{F}^*_{t-l/m} - H^*\tilde{F}_{t-l/m})'$$
$$= \tilde{V}^{*-1}\frac{1}{T}\sum_{t=1}^{T}(A^*_{1,t-j/m} + A^*_{2,t-j/m} + A^*_{3,t-j/m} + A^*_{4,t-j/m})(A^*_{1,t-l/m} + A^*_{2,t-l/m} + A^*_{3,t-l/m} + A^*_{4,t-l/m})'\tilde{V}^{*-1}.$$

Then, ignoring $\tilde{V}^{*-1} = O_{p^*}(1)$, we have $\frac{1}{T}\sum_{t=1}^{T} A^*_{1,t-j/m}A^{*'}_{1,t-l/m} = O_{p^*}(T^{-1})$, $\frac{1}{T}\sum_{t=1}^{T} A^*_{2,t-j/m}A^{*'}_{2,t-l/m} = O_{p^*}(N^{-1}\delta^{-2}_{NT_H})$, $\frac{1}{T}\sum_{t=1}^{T} A^*_{3,t-j/m}A^{*'}_{3,t-l/m} = \frac{1}{N}H^*\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\tilde{\lambda}'e^*_{t-j/m}}{\sqrt{N}}\right)\left(\frac{e^{*'}_{t-l/m}\tilde{\lambda}}{\sqrt{N}}\right)H^* + o_{p^*}(1)$, $\frac{1}{T}\sum_{t=1}^{T} A^*_{4,t-j/m}A^{*'}_{4,t-l/m} = O_{p^*}(N^{-1}\delta^{-2}_{NT_H})$, $\frac{1}{T}\sum_{t=1}^{T} A^*_{1,t-j/m}A^{*'}_{2,t-l/m} = O_{p^*}(T^{-1/2}N^{-1/2}\delta^{-1}_{NT_H})$, $\frac{1}{T}\sum_{t=1}^{T} A^*_{1,t-j/m}A^{*'}_{3,t-l/m} = O_{p^*}(T^{-1/2}N^{-1/2})$, $\frac{1}{T}\sum_{t=1}^{T} A^*_{1,t-j/m}A^{*'}_{4,t-l/m} = O_{p^*}(T^{-1/2}N^{-1/2})$, $\frac{1}{T}\sum_{t=1}^{T} A^*_{2,t-j/m}A^{*'}_{3,t-l/m} = O_{p^*}(N^{-1}\delta^{-2}_{NT_H})$, $A^*_{2,t-j/m}A^{*'}_{4,t-l/m} = O_{p^*}(N^{-1}\delta^{-1}_{NT_H})$ and $\frac{1}{T}\sum_{t=1}^{T} A^*_{3,t-j/m}A^{*'}_{4,t-l/m} = O_{p^*}(N^{-1}\delta^{-1}_{NT_H})$. Thus, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{F}^*_{t-j/m} - H^*\tilde{F}_{t-j/m})(\tilde{F}^*_{t-l/m} - H^*\tilde{F}_{t-l/m})' = \frac{\sqrt{T}}{N}\tilde{V}^{*-1}H^*\tilde{\Gamma}_{j-l}H^*\tilde{V}^{*-1} + o_{p^*}(1).$$

where we define $\tilde{\Gamma}_{j-l} \equiv \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\tilde{\lambda}'e^*_{t-j/m}}{\sqrt{N}}\right)\left(\frac{e^{*'}_{t-l/m}\tilde{\lambda}}{\sqrt{N}}\right)$.

(proof of part(d))

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}H^*\tilde{F}_{t-j/m}(\tilde{F}^*_{t-l/m} - H^*\tilde{F}_{t-l/m})' = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}H^*\tilde{F}_{t-j/m}(A^*_{1,t-l/m} + A^*_{2,t-l/m} + A^*_{3,t-l/m} + A^*_{4,t-l/m})'\tilde{V}^{*-1}$$
$$\equiv \sqrt{T}H^*(d^*_{f1} + d^*_{f2} + d^*_{f3} + d^*_{f4})'\tilde{V}^{*-1}$$

where $d^*_{fi} \equiv \frac{1}{T}\sum_{t=1}^{T}\tilde{F}_{t-j/m}A^{*'}_{i,t-l/m}$ for $i = 1,2,3,4$. Then, we can obtain $d^*_{f1} = O_{p^*}(\delta^{-1}_{NT_H}T^{-1/2}) + O_{p^*}(T_H^{-1})$, $d^*_{f2} = O_{p^*}((TN)^{-1/2})$ by Condition C.3*-(a) and $d^*_{f3} = O_{p^*}((TN)^{-1/2})$ by Condition C.3*-(b). Finally, $d^*_{f4} = \frac{1}{N}\big(\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_{t-l/m}\tilde{F}'_{t-j/m}\big)\tilde{\Gamma}\big(\frac{1}{T}\sum_{t=1}^{T_H}\tilde{F}_t\tilde{F}^{*'}_t\big)\tilde{V}^{*-1} + o_{p^*}(1)$. Thus,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}H^*\tilde{F}_{t-j/m}(\tilde{F}^*_{t-l/m} - H^*\tilde{F}_{t-l/m})' = \frac{\sqrt{T}}{N}H^*\Big(\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_{t-l/m}\tilde{F}'_{t-j/m}\Big)\tilde{\Gamma}\Big(\frac{1}{T_H}\sum_{s=1}^{T_H}\tilde{F}_s\tilde{F}^{*'}_s\Big)\tilde{V}^{*-2} + o_{p^*}(1).$$

■

proof of Lemma C.3.

(proof of part(a))

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big[\sum_{j=0}^{q}w_j(\tilde{\theta})(\tilde{F}^*_{t-j/m}-H^*\tilde{F}_{t-j/m})\Big]\Big[\sum_{j=0}^{q}w_j(\tilde{\theta})(\tilde{F}^*_{t-j/m}-H^*\tilde{F}_{t-j/m})\Big]'\\
&=\sum_{j=0}^{q}w_j(\tilde{\theta})\Big[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{F}^*_{t-j/m}-H^*\tilde{F}_{t-j/m})(\tilde{F}^*_{t-j/m}-H^*\tilde{F}_{t-j/m})'\Big]w_j(\tilde{\theta})\\
&\quad+\sum_{j=0}^{q}\sum_{l\neq j}^{q}w_j(\tilde{\theta})\Big[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(\tilde{F}^*_{t-j/m}-H^*\tilde{F}_{t-j/m})(\tilde{F}^*_{t-l/m}-H^*\tilde{F}_{t-l/m})'\Big]w_l(\tilde{\theta})\\
&=c\tilde{V}^{*-1}H^*\Big(\sum_{j=0}^{q}w_j(\tilde{\theta})\tilde{\Gamma}w_j(\tilde{\theta})+\sum_{j=0}^{q}w_j(\tilde{\theta})\tilde{\Gamma}_{j-l}w_l(\tilde{\theta})\Big)H^*\tilde{V}^{*-1}+o_{p^*}(1)\\
&=cH_0^*\tilde{V}^{-1}\Big(\sum_{j=0}^{q}w_j(\tilde{\theta})\tilde{\Gamma}w_j(\tilde{\theta})+\sum_{j=0}^{q}w_j(\tilde{\theta})\tilde{\Gamma}_{j-l}w_l(\tilde{\theta})\Big)\tilde{V}^{-1}H_0^*+o_{p^*}(1),
\end{aligned}$$

in probability. For the final equality, we use Lemma B.1 in GP (2014), $\tilde{V}^*=H^*\tilde{V}H^{*'}+O_{p^*}(\delta_{NT_H}^{-2})=\tilde{V}+O_{p^*}(\delta_{NT_H}^{-2})$ and $H^*=H_0^*+O_{p^*}(\delta_{NT_H}^{-2})$ in probability.

(proof of part(b))

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big[\sum_{j=0}^{q}w_j(\tilde{\theta})(\tilde{F}^*_{t-j/m}-H^*\tilde{F}_{t-j/m})\Big]\Big[\sum_{j=0}^{q}w_j(\tilde{\theta})H^*\tilde{F}_{t-j/m}\Big]'\\
&=\sum_{j=0}^{q}w_j(\tilde{\theta})\frac{1}{\sqrt{T}}\sum_{t=1}^{T}H^*\tilde{F}_{t-j/m}(\tilde{F}^*_{t-j/m}-\tilde{F}_{t-j/m})'w_j(\tilde{\theta})+\sum_{j=0}^{q}\sum_{l\neq j}^{q}w_j(\tilde{\theta})\frac{1}{\sqrt{T}}\sum_{t=1}^{T}H^*\tilde{F}_{t-l/m}(\tilde{F}^*_{t-j/m}-\tilde{F}_{t-j/m})'w_l(\tilde{\theta})\\
&=cH^*\Big[\sum_{j=0}^{q}w_j^2(\tilde{\theta})+\sum_{j=0}^{q}\sum_{l\neq j}^{q}w_j(\tilde{\theta})\Big(\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_{t-l/m}\tilde{F}'_{t-j/m}\Big)w_l(\tilde{\theta})\Big]\tilde{\Gamma}\Big(\frac{1}{T_H}\sum_{s=1}^{T_H}\tilde{F}_s\tilde{F}_s^{*'}\Big)\tilde{V}^{*-2}+o_{p^*}(1)\\
&=cH_0^*\Big[\sum_{j=0}^{q}w_j^2(\tilde{\theta})+\sum_{j=0}^{q}\sum_{l\neq j}^{q}w_j(\tilde{\theta})\Big(\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_{t-l/m}\tilde{F}'_{t-j/m}\Big)w_l(\tilde{\theta})\Big]\tilde{\Gamma}\tilde{V}^{-2}H_0^*+o_{p^*}(1),
\end{aligned}$$

in probability. The final equality is by applying Lemma B.1. in GP (2014) and by $\frac{\tilde{F}^{*'}\tilde{F}}{T_H}\tilde{V}^{*-1}=\tilde{V}^{-1}H^*$ and $H^*\tilde{V}^{*-1}=\tilde{V}^{-1}H^*$.

■

Lemma C.4 *Suppose Assumption A.1-A.6 hold. If in addition either:*

1. $\{F_s\}$, $\{\lambda_i\}$ *and* $\{e_{i,t_h}\}$ *are mutually independent and for some* $p\geq 2$, $E|e_{it}|^{2p}\leq M<\infty$;
2. *for some* $p\geq 2$, $E|e_{it}|^{3p}\leq M<\infty$, $E\|\lambda_i\|^{3p}\leq M<\infty$ *and* $E\|F_t\|^{3p}\leq M<\infty$,

   it follows that,

   (a) $\frac{1}{T_H}\sum_{t=1}^{T_H}\|\tilde{F}_t-HF_t\|^p=O_p(1)$,

   (b) $\frac{1}{N}\sum_{i=1}^{N}\|\tilde{\lambda}_i-H^{-1'}\lambda_i\|^p=O_p(1)$,

(c) $\frac{1}{T_H N}\sum_{t=1}^{T_H}\sum_{i=1}^{N}\tilde{e}_{it}^p = O_p(1)$.

(d) $\frac{1}{T_H}\sum_{t=1}^{T_H}\|\tilde{F}_t\|^p = O_p(1)$,

(e) $\frac{1}{T_H N}\sum_{t=1}^{T_H}\sum_{i=1}^{N}\tilde{u}_{it}^p = O_p(1)$.

proof of Lemma C.4. (a)-(d) of Lemma C.4 is from GP (2014). The proofs can be found in GP (2014) page 16. Therefore, we present the proof only for (e). Note that $\tilde{u}_{it} = \tilde{e}_{it} - \sum_{j=1}^{p_i}\tilde{a}_{i,j}\tilde{e}_{i,t-j}$. By c_r inequality,

$$\frac{1}{NT_H}\sum_{i=1}^{N}\sum_{t=1}^{T_H}|u_{it}|^p \leq 2^{p-1}\frac{1}{NT_H}\sum_{i=1}^{N}\sum_{t=1}^{T_H}|\tilde{e}_{it}|^p + 2^{p-1}\frac{1}{NT_H}\sum_{i=1}^{N}\sum_{t=1}^{T_H}\sum_{j=1}^{p_i}|\tilde{a}_{ij}\tilde{e}_{i,t-j}|^p = O_p(1).$$

The first term is $O_p(1)$ by (c) in Lemma C.4. The second term is also $O_p(1)$ by Cauchy-Schwarz inequality as,

$$\frac{1}{NT_H}\sum_{i=1}^{N}\sum_{t=1}^{T_H}\sum_{j=1}^{p_i}|\tilde{a}_{ij}\tilde{e}_{i,t-j}|^p \leq \Big(\sum_{i=1}^{N}\sum_{j=1}^{p_i}|\tilde{a}_{ij}|^p\Big)\Big(\frac{1}{NT_H}\sum_{i=1}^{N}\sum_{t=1}^{T_H}|\tilde{e}_{i,t-j}|^p\Big).$$

Again, the second term is also $O_p(1)$ by (c). The first term is $O_p(1)$ since $\sum_{j=1}^{\infty}|a_{i,j}| < \infty$. ■

proof of Theorem 3.1. We prove each condition, respectively. For Condition C.1* - (a), $E^*(e_{it}^*) = E^*(\sum_{j=0}^{\infty}\tilde{b}_{i,j}u_{i,t-j}^*) = 0$ since $E^*(u_{i,t-j}^*) = E^*(\tilde{u}_{i,t-j}\eta_{i,t-j}) = 0$ by $\eta_{i,t-j} \sim$ i.i.d$N(0,1)$. For part (b), we use the MA(∞) representation and write $\gamma_{s,t}^*$ as

$$\begin{aligned}\gamma_{st}^* &= E^*\Big(\frac{1}{N}\sum_{i=1}^{N}e_{i,t}^*e_{i,s}^*\Big)\\ &= E^*\Big[\frac{1}{N}\sum_{i=1}^{N}\Big(\sum_{j=0}^{\infty}\tilde{b}_{i,j}u_{i,t-j}^*\Big)\Big(\sum_{j=0}^{\infty}\tilde{b}_{i,j}u_{i,s-j}^*\Big)\Big]\\ &= E^*\Big[\frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{\infty}\tilde{b}_{i,j}\tilde{b}_{i,s-t+j}\tilde{u}_{i,t}^2\Big]\end{aligned}$$

Then,

$$\frac{1}{T_H}\sum_{t=1}^{T_H}\sum_{s=1}^{T_H}|\gamma_{s,t}^*|^2 \leq \Big(\sum_{j=0}^{\infty}\tilde{b}_{i,j}\sum_{t=1}^{T_H}\sum_{s=1}^{T_H}\tilde{b}_{i,s-t+j}\Big)^2\frac{1}{NT_H}\sum_{t=1}^{T_H}\sum_{i=1}^{N}\tilde{u}_{it}^4$$

which we have $O_p(1)$ by applying $p = 4$ in Lemma C.4-(e). For part (c), we can write as:

$$\begin{aligned}E^*\Big|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(e_{it}^*e_{is}^* - E^*(e_{it}^*e_{is}^*))\Big|^2 &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}Cov^*(e_{it}^*e_{is}^*, e_{jt}^*e_{js}^*)\\ &= \sum_{l=0}^{\infty}\tilde{b}_{i,l}^2\tilde{b}_{i,s-t+l}^2\tilde{u}_{i,t-l}^2\tilde{u}_{i,s-l}^2 Var(\eta_{i,t-l}\eta_{i,s-l})\end{aligned}$$

By assuming $Var(\eta_{i,t-l}\eta_{i,s-l}) \leq \bar{\eta}$, part (c) is smaller than

$$\bar{\eta}\frac{1}{T_H^2}\sum_{t=1}^{T_H}\sum_{s=1}^{T_H}\frac{1}{N}\sum_{i=1}^{N}\sum_{l=0}^{\infty}\tilde{b}_{i,l}^2\tilde{b}_{i,s-t+l}^2\tilde{u}_{i,t-l}^2\tilde{u}_{i,s-l}^2 \leq \bar{\eta}\Big(\sum_{i=1}^{N}\sum_{l=0}^{\infty}\tilde{b}_{i,l}^2\tilde{b}_{i,s-t+l}^2\Big)\Big(\sum_{l=0}^{\infty}\frac{1}{NT_H}\sum_{i=1}^{N}\sum_{t=1}^{T_H}\tilde{u}_{i,t-l}^4\Big)$$

Since $\frac{1}{NT_H}\sum_{i=1}^{N}\sum_{t=1}^{T_H}\tilde{u}_{i,t-l}^4 = O_p(1)$ by applying $p = 4$ on Lemma C.4-(e). In Condition C.2*, for part (a), we write

γ^*_{st} using

$$\frac{1}{N}\sum_{i=1}^{N} e^*_{is}e^*_{it} = \frac{1}{N}\sum_{i=1}^{N}\Big(\sum_{j=0}^{\infty}\tilde{b}_{i,j}u^*_{i,t-j}\Big)\Big(\sum_{l=0}^{\infty}\tilde{b}_{i,l}u^*_{i,s-l}\Big)$$

Since $u^*_{i,t-j}u^*_{i,s-l} = 0$ if $s-l \neq t-j$, we can consider it as:

$$\frac{1}{T_H}\sum_{t=1}^{T_H}\sum_{s=1}^{T_H}\tilde{F}_s\tilde{F}'_t\Big(\frac{1}{N}\Big(\sum_{j=0}^{\infty}\tilde{b}_{i,j}\tilde{b}_{i,s-t+j}\tilde{u}^2_{i,t}\Big)$$
$$\leq \Big(\frac{1}{T_H}\sum_{t=1}^{T_H}\sum_{s=1}^{T_H}\|\tilde{F}_s\tilde{F}'_t\|^2\Big)^{1/2}\Big(\frac{1}{NT_H}\sum_{t=1}^{T_H}\sum_{s=1}^{T_H}\sum_{i=1}^{N}\Big(\sum_{j=0}^{\infty}|\tilde{b}_{i,j}||\tilde{b}_{i,s-t+j}||\tilde{u}_{i,t}|^4\Big)^{1/2} = O_p(1)$$

where the first term is bounded by $\frac{1}{T_H}\sum_{t=1}^{T_H}\|\tilde{F}_t\|^4$ and we have $\frac{1}{NT_H}\sum_{i=1}^{N}|\tilde{u}_{it}|^4 = O_p(1)$. For part (b), we have

$$\frac{1}{T_H}\sum_{t=1}^{T_H}\frac{1}{T_H}\sum_{s=1}^{T_H}\sum_{l=1}^{T_H}\tilde{F}'_s\tilde{F}_l\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}Cov^*(e^*_{it}e^*_{is}, e^*_{jt}e^*_{jl})$$

Since we assume the idiosyncratic error terms are cross-sectionally independent, we have $Cov^*(e^*_{it}e^*_{is}, e^*_{jt}e^*_{jl}) = 0$ for $i \neq j$, for any t, s, l. We consider the case where $i = j$, $Cov^*(e^*_{it}e^*_{is}, e^*_{it}e^*_{il})$. Thus, part (b) becomes,

$$\frac{1}{T_H}\sum_{t=1}^{T_H}E^*\|\frac{1}{\sqrt{T_H N}}\sum_{t=1}^{T_H}\sum_{i=1}^{N}\tilde{F}_s(e^*_{it}e^*_{is} - E^*(e^*_{it}e^*_{is}))\|^2$$
$$\leq \bar{\eta}\Big(\sum_{j=0}^{\infty}|\tilde{b}_{i,j}|^4\Big)\frac{1}{N}\sum_{i=1}^{N}\Big(\frac{1}{T_H}\sum_{t=1}^{T_H}\tilde{u}^2_{i,t}\Big)\Big(\frac{1}{T_H}\sum_{s=1}^{T_H}\tilde{F}'_s\tilde{F}_s\tilde{u}^2_{i,s}\Big)$$
$$\leq \bar{\eta}\Big(\sum_{j=0}^{\infty}|\tilde{b}_{i,j}|^4\Big)\Big(\frac{1}{NT_H}\sum_{i=1}^{N}\sum_{t=1}^{T_H}\tilde{u}^4_{i,t}\Big)^{1/2}\Big(\frac{1}{T_H}\sum_{s=1}^{T_H}\|\tilde{F}_s\|^4\frac{1}{NT_H}\sum_{i=1}^{N}\sum_{t=1}^{T_H}\tilde{u}^4_{i,s}\Big)^{1/2} = O_p(1),$$

where the inequality works with Cauchy-Schwarz inequality. The part (c) is $O_p(1)$ since

$$E^*\|\frac{1}{\sqrt{T_H}}\sum_{t=1}^{T_H}\frac{\tilde{\Lambda}'e^*_t}{\sqrt{N}}\tilde{F}'_t\|^2 = \frac{1}{T_H}\sum_{t=1}^{T_H}\|\tilde{F}_t\|^2\frac{1}{N}\sum_{i=1}^{N}\|\tilde{\lambda}_i\|^2 e^{*2}_{i,t}$$
$$= \frac{1}{T_H}\sum_{t=1}^{T_H}\|\tilde{F}_t\|^2\frac{1}{N}\sum_{i=1}^{N}\|\tilde{\lambda}_i\|^2\Big(\sum_{j=0}^{\infty}\tilde{b}_{i,j}u^*_{i,t}\Big)^2$$
$$\leq \frac{1}{T_H}\sum_{t=1}^{T_H}\|\tilde{F}_t\|^2\frac{1}{N}\sum_{i=1}^{N}\|\tilde{\lambda}_i\|^2\sum_{j=0}^{\infty}|\tilde{b}_{i,j}|^2\tilde{u}^2_{i,t-j}$$
$$\leq \Big(\frac{1}{T_H}\sum_{t=1}^{T_H}\|\tilde{F}_t\|^4\Big)^{1/2}\Big[\frac{1}{T_H}\sum_{t=1}^{T_H}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\tilde{\lambda}_i\|^2\sum_{j=0}^{\infty}|\tilde{b}_{i,j}|^2\tilde{u}^2_{i,t-j}\Big)^2\Big]^{1/2}$$

By Cauchy-Schwarz inequality, the term in square bracket is bounded by:

$$\frac{1}{T_H}\sum_{t=1}^{T_H}\sum_{j=0}^{\infty}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\tilde{\lambda}_i\|^2\tilde{u}^2_{i,t-j}|\tilde{b}_{i,j}|^2\Big)^2 \leq \frac{1}{N}\sum_{i=1}^{N}\|\tilde{\lambda}_i\|^4\sum_{j=0}^{\infty}|\tilde{b}_{i,j}|^2\frac{1}{T_H}\sum_{t=1}^{T_H}\frac{1}{N}\sum_{i=1}^{N}\tilde{u}^4_{i,t-j} = O_p(1),$$

where we have $\frac{1}{N}\sum_{i=1}^{N}\sum_{i=1}^{N}\|\tilde{\lambda}_i\|^4 = O_p(1)$ by Lemma C.4.

For part (d), we can write as:

$$\frac{1}{T_H}\sum_{t=1}^{T_H} E^*\|\frac{\tilde{\Lambda'} e_t^*}{\sqrt{N}}\|^2 = \frac{1}{N}\sum_{i=1}^{N}\|\tilde{\lambda}_i\|^2\Big(\frac{1}{T_H}\sum_{t=1}^{T_H} e_{it}^*\Big)$$
$$= \frac{1}{N}\sum_{i=1}^{N}\|\tilde{\lambda}_i\|^2\Big(\frac{1}{T_H}\sum_{t=1}^{T_H}\Big(\sum_{j=0}^{\infty}\tilde{b}_{ij}u_{it}^*\Big)^2\Big)$$
$$\leq \Big(\frac{1}{N}\sum_{i=1}^{N}\|\tilde{\lambda}_i\|^4\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{T_H}\sum_{t=1}^{T_H}\Big(\sum_{j=0}^{\infty}|\tilde{b}_{ij}|^4\Big)\tilde{u}_{it}^4\Big)^{1/2} = O_p(1),$$

where the last inequality is by Cauchy-Schwarz inequality. Then, for part (e), we have to show that

$$A^* = \frac{1}{T_H}\sum_{t=1}^{T_H}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\tilde{\lambda}_i\tilde{\lambda}_j'(e_{it}^*e_{jt}^* - E^*(e_{it}^*e_{jt}^*)) = o_p(1).$$

The term A^* is mean zero by construction, therefore, we need to show that the variance of A^* tends to zero in probability. For simplicity, we take $r = 1$. Then, the variance of A^* is:

$$Var(A^*) = \frac{1}{T_H^2}\sum_{t=1}^{T_H}\sum_{s=1}^{T_H}\frac{1}{N^2}\sum_{i,j,k,l}^{N}\tilde{\lambda}_i\tilde{\lambda}_j\tilde{\lambda}_l\tilde{\lambda}_k Cov^*(e_{it}^*e_{jt}^*, e_{ls}^*e_{ks}^*),$$

where $Cov^*(e_{it}^*e_{jt}^*, e_{ls}^*e_{ks}^*) = 0$ if $i \neq j \neq k \neq l$. When $i = j = k = l$, $Cov^*(e_{it}^*e_{jt}^*, e_{ls}^*e_{ks}^*) = \sum_{j=0}^{\infty}\tilde{b}_{ij}^2\tilde{b}_{i,s-t+j}\tilde{u}_{it-j}^4$ and when $i = k \neq j = l$, $Cov^*(e_{it}^*e_{jt}^*, e_{ls}^*e_{ks}^*) = \Big(\sum_{l=0}^{\infty}\tilde{b}_{i,l}^2\tilde{u}_{i,t-l}^2\Big)\Big(\sum_{l=0}^{\infty}\tilde{b}_{j,l}^2\tilde{u}_{j,t-l}^2\Big)$. By applying $p = 4$ in Lemma C.4, we have $Var^*(A^*) = o_p(1)$.

The proofs for Condition C.4* and Condition C.5* are same as in the proof for Condition C* and Condition D* in GP (2014). This is because we use wild bootstrap for resampling $\tilde{\varepsilon}_t = y_t - \tilde{\beta}'\tilde{F}_t(\tilde{\theta})$, which is same procedure as in GP (2014). ■

D Additional results

Table 7 shows the bias and 95% coverage rate of β when the idiosyncratic error term follows simple AR (1) process as:

$$e_{i,t_h} = \rho_i e_{i,t_h-1} + v_{i,t_h} \text{ for } t_h = 1,\ldots,T_H$$

where v_{i,t_h} is i.i.d. randomly generated from $N(0,1)$. ρ indicates the auto-regressive coefficient, which implies the persistence of auto-regressive process. For simplicity, we impose that each variable shares same autoregressive coefficient, $\rho_i = \rho$. In Table 7, we compare the results by varying persistence. We increase the coefficient from 0 to 0.7. When the persistence in the idiosyncratic error term is $\rho = 0.5$, the bias is around twice bigger than the bias where there is no serial-dependence. Moreover, the size of bias increase as the persistence increases.

Table 7: Bias and 95% coverage rate of β

N	T_H	$\rho = 0$		$\rho = 0.5$		$\rho = 0.6$		$\rho = 0.7$	
		bias	95%	bias	95%	bias	95%	bias	95%
50	150	-0.3380	84.7	-0.5887	68.02	-0.6808	60.42	-0.7993	49.18
	300	-0.3100	81.76	-0.5362	57.94	-0.6197	48.16	-0.7278	35.18
	600	-0.2890	74	-0.4970	40.96	-0.5746	29.32	-0.6761	17.2
100	150	-0.2022	89.82	-0.3763	83.18	-0.4450	79.34	-0.5372	72.62
	300	-0.1709	90.72	-0.3157	81.1	-0.3729	75.68	-0.4502	67.1
	600	-0.1565	88.7	-0.2849	75.36	-0.3358	67.44	-0.4047	56.16
200	150	-0.1343	91.48	-0.2639	87.6	-0.3163	85.38	-0.3890	81.8
	300	-0.1027	92.5	-0.1996	89.18	-0.2393	87.28	-0.2943	83.54
	600	-0.0865	92.44	-0.1647	88.02	-0.1968	85.48	-0.2411	80.7

(To be added in bootstrap part)

Let $\{e^*_{t_h} = (e^*_{1,t_h}, \ldots, e^*_{N,t_h})'\}$ be a bootstrap sample from $\{\tilde{e}_{t_h} = (\tilde{e}_{1,t_h}, \ldots, \tilde{e}_{N,t_h})'\}$, where $\tilde{e}_{t_h} = X_{t_h} - \tilde{\Lambda}\tilde{f}_{t_h}$ are the residuals from the original panel dataset. $\{\varepsilon^*_t\}$ are the resampled bootstrap residuals from $\{\tilde{\varepsilon}_t = y_t - g(\tilde{F}_t; \tilde{\alpha})\}$. Using these two bootstrap samples, $\{e^*_{t_h}\}$ and $\{\varepsilon^*_t\}$, the bootstrap data generating process (DGP) is as follows:

$$X^*_{t_h} = \tilde{\Lambda}\tilde{f}_{t_h} + e^*_{t_h}, \text{ for } t_h = 1, \ldots, T_H, \tag{26}$$

$$y^*_t = \tilde{\beta}_0 + \tilde{\beta}'_1 \tilde{F}_t(\tilde{\theta}) + \varepsilon^*_t, \text{ for } t = 1, \ldots, T. \tag{27}$$

We follow a two-step process that is similar to the procedure used in the original sample. For the first step, we estimate the factors from a new bootstrap panel dataset, $X^*_{t_h}$. The estimated factors are denoted by $\tilde{f}^*_{t_h}$. In the second step, we estimate $\tilde{\alpha}$ by regressing y^*_t on 1 and $\tilde{F}^*_t(\tilde{\theta})$, which are temporally aggregated bootstrap factors. Note that the weighting parameter we use in this step is $\tilde{\theta}$, which are the true weighting parameters in the bootstrap world. The resulting estimators from this step are denoted by $\tilde{\alpha}^* = (\tilde{\beta}^{*\prime}, \tilde{\theta}^{*\prime})'$, where $\tilde{\beta}^* = (\tilde{\beta}^*_0, \tilde{\beta}^{*\prime}_1)'$. These estimators are the bootstrap analogue of $\tilde{\alpha}$, which are the NLS estimators from the original sample.

Our objective is to demonstrate the consistency of the bootstrap estimators under the bootstrap DGP as specified in (26) and (27). In order to do that, we provide a set of high-level conditions, similar to those in GP (2014) for factor-augmented regression models. We extend the conditions in GP (2014) to incorporate serial dependence among the "true" factors, factor loadings, and the idiosyncratic world in the bootstrap world. As most of the conditions are similar to GP (2014), we leave the conditions in Appendix C.

D.1 Bootstrap distribution

In this section, we derive the bootstrap distribution of the estimators of bootstrap DGP, (26) and (27). By Conditions C.1* - C.5* in Appendix C, the estimated factors $\tilde{f}_t^*$ consistently estimate the rotated version of true "latent" bootstrap factors, $H^*\tilde{f}_t$. The rotation matrix H^* is given by $\tilde{V}^{*-1}\frac{\tilde{f}^{*\prime}\tilde{f}}{T_H}\frac{\tilde{\Lambda}'\tilde{\Lambda}}{N}$, where $\tilde{V}^*$ is the $r \times r$ diagonal matrix containing on the main diagonal the r largest eigenvalues of $X^*X^{*\prime}/NT_H$, in decreasing order. This matrix is the bootstrap analogue of the rotation matrix in the original sample, $H = \tilde{V}^{-1}\frac{\tilde{f}'f}{T_H}\frac{\Lambda'\Lambda}{N}$. In the original sample, because the rotation matrix, H depends on true latent factors, f_t, H cannot be determined. However, the indeterminacy of the rotation matrix is not a problem in the bootstrap world, as H^* does not depend on the population values. Moreover, H^* is asymptotically equal to $H_0^* = \text{diag}(\pm 1)$, where the sign is determined by the sign of $\tilde{f}^{*\prime}\tilde{f}/T_H$. This implies that the bootstrap factors are identified up to a change of sign.

Another thing to note is that the NLS estimators in (27) also rotate due to the rotation in the factors in the bootstrap world. We can see this by rewriting (27) as follows:

$$y_t^* = \tilde{\beta}_0 + \tilde{\beta}_1' H^{*-1}\tilde{F}_t^*(\tilde{\theta}) + \tilde{\beta}_1' H^{*-1}(H^*\tilde{F}_t(\tilde{\theta}) - \tilde{F}_t^*(\tilde{\theta})) + \varepsilon_t^* = g(\tilde{F}_t^*; \tilde{\alpha}) + \xi_t^*,$$

where $g(\tilde{F}_t^*; \tilde{\alpha}) \equiv \tilde{\beta}_0 + \tilde{\beta}_1' H^{*-1}\tilde{F}_t^*(\tilde{\theta})$ and $\xi_t \equiv \tilde{\beta}_1' H^{*-1}(H^*\tilde{F}_t(\tilde{\theta}) - \tilde{F}_t^*(\tilde{\theta})) + \varepsilon_t^*$. Thus, $\tilde{\alpha}^*$ estimates $(\Phi^*)^{-1}\tilde{\alpha}$, where $\Phi^* = diag(H^*, I_p)$ is a block diagonal matrix. $(\Phi^*)^{-1}\tilde{\alpha}$ are the rotated version of NLS estimators in the original sample. As H^* is asymptotically equal to H_0^*, $(\Phi^*)^{-1}\tilde{\alpha}$ is equal to $(\Phi_0^*)^{-1}\tilde{\alpha}$, where $\Phi_0^* = \text{diag}(H_0^*, I_p)$, and $(\Phi_0^*)^{-1}\tilde{\alpha}$ is the sign-adjusted version of $\tilde{\alpha}$.

Prior to characterizing the asymptotic bootstrap distribution of $\sqrt{T}(\tilde{\alpha}^* - (\Phi_0^*)^{-1}\tilde{\alpha})$ under the assumption of $\sqrt{T}/N \to c$, where $0 \leq c < \infty$, we include following additional conditions.

Condition 1* $\text{plim}\,\tilde{\Omega} = \Phi_0\Omega\Phi_0'$.

Condition 2* $\text{plim}\,\tilde{\Gamma} = H_0\Gamma H_0'$ *and* $\text{plim}\,\tilde{\Gamma}_{j-l} = H_0\Gamma_{j-l}H_0'$.

$\tilde{\Omega} = Var^* \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{g}_{\alpha,t} \varepsilon_t^* \right)$ is the bootstrap variance of the score vector in the bootstrap world, where $\tilde{g}_{\alpha,t} \equiv \partial g(\tilde{F}_t; \alpha)/\partial \alpha$. It is a bootstrap analogue of Ω. The Condition 1* implies that the bootstrap version is rotated with a block diagonal matrix, Φ_0. This is because the score vector is $\tilde{g}_{\alpha,t} = \left(F_t'(\theta)H', \beta' \frac{\partial F_t(\theta)}{\partial \theta'} \right)'$ is a rotated version of $g_{\alpha,t}$, where the rotation is given by Φ_0. Similarly, $\tilde{\Gamma}$ and $\tilde{\Gamma}_{j-l}$, defined in Condition C.2* and Condition C.3* in Appendix C, are the bootstrap analogues of Γ and Γ_{j-l}, respectively. They are also rotated with the rotation matrix, H. Conditions 1* and 2* imply that it is crucial how we mimic the error terms of the MIDAS regression and the idiosyncratic factor error terms in the bootstrap world. Moreover, in our context, since the bias depends on both serial and cross-sectional dependence of e_{t_h}, the idiosyncratic error term in the bootstrap world should mimic the dependence in the time series and cross-sectional dimension.

Theorem D.1 *Let the Assumptions A.1-A.5 in Appendix A hold and consider any residual-based bootstrap scheme for which Conditions C.1*-C.4* are verified. Suppose $\sqrt{T}/N \to c,\ 0 \leq c < \infty$. In addition, let the two following conditions hold: (1) Condition 1* is verified and (2) $c = 0$ or Condition 2* is verified; then as $N, T \to \infty$,*

$$\sqrt{T}(\tilde{\alpha}^* - (\Phi_0^*)^{-1}\tilde{\alpha}) \xrightarrow{d^*} N(-c(\Phi_0^*)^{-1}\Delta_\alpha, (\Phi_0^*)^{-1}\Sigma_\alpha(\Phi_0^*)^{-1}),$$

in probability and Δ_α and Σ_α are defined in Theorem 2.1.

According to Theorem D.1, the distribution of $\sqrt{T}(\tilde{\alpha}^* - (\Phi^*)^{-1}\tilde{\alpha})$ follows a normal distribution with a non-zero mean vector, $-c(\Phi_0^*)^{-1}\Delta_\alpha$. The asymptotic bias is proportional to $(H_0^*)^{-1}\tilde{\beta}$. However, the weighting parameters $\tilde{\theta}^*$ are not affected by the rotation problem.

As in GP (2014), we need to match the bootstrap distribution with the limiting distribution of the estimators in the original sample to achieve bootstrap consistency since our rotation matrix H^* may not be an identity matrix. To do that, we consider the rotated version of our bootstrap results, given by $\sqrt{T}(\Phi^*\tilde{\alpha}^* - \tilde{\alpha})$. Φ^* is asymptotically equal to $\Phi_0^* = \text{diag}(H_0^*, I_p)$, where $H_0 = \text{diag}(\pm 1)$ and the sign is determined by the sign of $\tilde{f}^{*\prime}\tilde{f}$. Because we know $\tilde{f}^*$ and $\tilde{f}$ in the bootstrap world, we can compute Φ_0^*. For the consistency of the rotated bootstrap results, we rely on the Corollary 3.1. in GP (2014) such that $\sup_{x \in \mathbb{R}^{r+p}} |P^*(\sqrt{T}(\Phi^*\tilde{\alpha}^* - \tilde{\alpha}) \leq x) - P(\sqrt{T}(\tilde{\alpha} - \alpha) \leq x)| \xrightarrow{p} 0$. This corollary justifies the use of a residual-based bootstrap method in the context of the factor-MIDAS regression models. If $c = 0$, then there will be no bias and it is important to satisfy Condition 1*. However, if $c > 0$, Condition 2* must also be satisfied to capture the bias. This means that the idiosyncratic error term in the bootstrap world should mimic the dependence of the error term in the factor model of the original sample.

References

Aastveit, K. A., Foroni, C., & Ravazzolo, F. (2017). Density forecasts with midas models. *Journal of Applied Econometrics*, *32*(4), 783-801. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2545` doi: https://doi.org/10.1002/jae.2545

Andreou, E., Gagliardini, P., Ghysels, E., & Rubin, M. (2019). Inference in group factor models with an application to mixed-frequency data. *Econometrica*, *87*(4), 1267–1305.

Andreou, E., Ghysels, E., & Kourtellos, A. (2013). Should macroeconomic forecasters use daily financial data and how? *Journal of Business & Economic Statistics*, *31*(2), 240–251.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, *71*(1), 135–171.

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*(1), 191–221.

Bai, J., & Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, *74*(4), 1133–1150.

Bai, J., & Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, *176*(1), 18–29.

Beyhum, J., & Striaukas, J. (2023). Factor-augmented sparse midas regression for nowcasting. *arXiv preprint arXiv:2306.13362*.

Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli*, 123–148.

Clements, M. P., & Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the united states. *Journal of Business & Economic Statistics*, *26*(4), 546–554.

Clements, M. P., & Galvão, A. B. (2009). Forecasting us output growth using leading indicators: An appraisal using midas models. *Journal of Applied Econometrics*, *24*(7), 1187–1206.

Ferrara, L., & Marsilli, C. (2019). Nowcasting global economic growth: A factor-augmented mixed-frequency approach. *The World Economy*, *42*(3), 846-875. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1111/twec.12708` doi: https://doi.org/10.1111/

twec.12708

Foroni, C., Marcellino, M., & Schumacher, C. (2015). Unrestricted mixed data sampling (midas): Midas regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *178*(1), 57–82.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004, 06). The midas touch: mixed data sampling regression.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2005). There is a risk-return trade-off after all. *Journal of Financial Economics*, *76*(3), 509–548.

Ghysels, E., Santa-Clara, P., & Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, *131*(1-2), 59–95.

Ghysels, E., Sinko, A., & Valkanov, R. (2007, 02). Midas regressions: Further results and new directions. *Econometric Reviews*, *26*, 53-90. doi: 10.2139/ssrn.885683

Ghysels, E., Valkanov, R. I., & Serrano, A. R. (2009). Multi-period forecasts of volatility: Direct, iterated, and mixed-data approaches. In *Efa 2009 bergen meetings paper.*

Gonçalves, S., Koh, J., & Perron, B. (2023). *Bootstrap inference for group factor models.*

Gonçalves, S., & Perron, B. (2014). Bootstrapping factor-augmented regression models. *Journal of Econometrics*, *182*(1), 156–173.

Gonçalves, S., & Perron, B. (2020). Bootstrapping factor models with cross sectional dependence. *Journal of Econometrics*, *218*(2), 476–495.

Gonçalves, S., Perron, B., & Djogbenou, A. (2017). Bootstrap prediction intervals for factor models. *Journal of Business & Economic Statistics*, *35*(1), 53–69.

Kim, H. H., & Swanson, N. R. (2018). Methods for backcasting, nowcasting and forecasting using factor-midas: With an application to korean gdp. *Journal of Forecasting*, *37*(3), 281–302.

Kreiss, J.-P., Paparoditis, E., & Politis, D. N. (2011). On the range of validity of the autoregressive sieve bootstrap. *The Annals of Statistics*, *39*(4), 2103–2130.

Ludvigson, S. C., & Ng, S. (2009). *A factor analysis of bond risk premia* (Tech. Rep.). National Bureau of Economic Research.

Marcellino, M., & Schumacher, C. (2010). Factor midas for nowcasting and forecasting with ragged-edge data: A model comparison for german gdp. *Oxford Bulletin of Economics and Statistics*, *72*(4), 518–550.

McCracken, M., & Ng, S. (2020). *Fred-qd: A quarterly database for macroeconomic research* (Tech. Rep.). National Bureau of Economic Research.

McCracken, M. W., & Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, *34*(4), 574–589.

Monteforte, L., & Moretti, G. (2013). Real-time forecasts of inflation: The role of financial variables. *Journal of Forecasting*, *32*(1), 51–61.