Selecting the Relevant Variables for Factor Estimation in a FAVAR Models*

John C. Chao[1] and Norman R. Swanson[2]

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

July 19, 2022

Abstract

When specifying and estimating latent factor models, a common assumption made is one of factor pervasiveness, which implies that all available predictor variables in a dataset load significantly on the underlying factors, with the possible exception of a negligible number of them. In this paper, build on the recent nacent literature that examines how to relax this assumption (see e.g., Giglio, Xiu, and Zhang (2021), Freyaldenhoven (2021a,b), and Bai and Ng (2021)), and analyze the scenario where there is significant underlying heterogeneity in the sense that some of the variables load significantly on the underlying factors, while others are irrelevant. Consistent factor estimation is shown to be feasible, even under factor nonpervasiveness, if one first pre-screens all available variables and prunes out the irrelevant ones. For this purpose, we introduce, within a factor-augmented VAR framework, a novel variable selection procedure that, with probability approaching one, correctly distinguishes between relevant and irrelevant variables. Our methodology enables the consistent estimation of conditional mean functions of factor-augmented forecast equations, even when the conventional assumption of factor pervasiveness is violated.

Keywords: Factor analysis, factor augmented vector autoregression, forecasting, moderate deviation, principal components, self-normalization, variable selection.

JEL Classification: C32, C33, C38, C52, C53, C55.

* *Corresponding Author:* John C. Chao, Department of Economics, 7343 Preinkert Drive, University of Maryland, jcchao@umd.edu.

Norman R. Swanson, Department of Economics, 9500 Hamilton Street, Rutgers University, nswanson@econ.rutgers.edu. The authors are grateful to Simon Freyaldenhoven, Yuan Liao, Minchul Shin, Xiye Yang, and seminar participants at the University of Glasgow, the University of Riverside, the Federal Reserve Bank of Pihadelphia, the 2022 Summer Econometrics Society Meetings, and the 2022 International Association of Applied Econometrics Association meetings for useful comments received on earlier versions of this paper. Chao thanks the University of Maryland for research support.

1 Introduction

As a result of the astounding rate at which raw information is currently being accumulated, there is a clear need for variable selection, dimension reduction and shrinkage techniques when analyzing big data using machine learning techniques. This has led to a profusion of novel research in areas ranging from the analysis of high dimensional and/or high frequency datasets to the development of new statistical learning methods. Needless to say, there are many critical unanswered questions in this burgeoning literature. One such question, which we address in this paper stems from the pathbreaking work due to Bai and Ng (2002), Stock and Watson (2002a,b), Bai (2003), Forni, Hallin, Lippi, and Reichlin (2005), and Bai and Ng (2008). In these papers, the authors develop methods for constructing forecasts based on factor-augmented regression models. An obvious appeal of using factor analytical methods for this problem is the capacity for dimension reduction, so that in terms of the specification of the forecasting equation, employment of a factor structure allows the parsimonious representation of information embedded in a possibly high-dimensional vector of predictor variables.

Within this context, we note that a key assumption commonly used in the literature to obtain consistent factor estimation is the so-called factor pervasiveness assumption which presupposes that all available variables in a dataset, load significantly on the underlying latent factors, with the possible exception of a negligible number of them[1]. Such an assumption places stringent requirements on the relationship between variables in a given dataset and, thus, may not be satisfied by many datasets that are available for empirical research. For this reason, the pervasiveness assumption has recently been relaxed in a number of interesting contexts in recent papers, including Giglio, Xiu, and Zhang (2021), Freyaldenhoven (2021a,b), and Bai and Ng (2021), for example. Our paper adds to this nascent literature by analyzing the likely scenario that there is significant underlying heterogeneity, so that some variables are relevant in the sense that they load significantly on the underlying fac-

[1] A more formal discussion of this factor pervasiveness assumption is given in part (a) of Remark OA1.1 of the Online Appendix to this paper, Chao and Swanson (2022c).

tors, whereas others are irrelevant, in the sense that they do not share any common dynamic structure with the other variables in the dataset. In scenarios such as this where the assumption of factor pervasiveness does not hold, inconsistency in factor estimation may result if one were to naively use all available variables to estimate the underlying factors, without regard to whether they are relevant or not. See Chao and Swanson (2022a), for a particularly pathological example where an estimated factor, $\widehat{f}_t$, approaches 0 in probability, regardless of what the true value of f_t happens to be - a situation which can arise when the underlying factors are nonpervasive. Not being able to obtain consistent estimates of the underlying factors, in turn, would clearly cause problems for empirical researchers, such as when the objective is to estimate forecast functions that incorporate estimated factors. On the other hand, if one pre-screens the variables and successfully prunes out the irrelevant ones, then consistent estimation can be achieved, under appropriate conditions. For this reason, a main contribution of this paper is to introduce a novel variable selection procedure which allows empirical researchers to correctly distinguish the relevant from the irrelevant variables prior to factor estimation, with probability approaching one. We study this problem within a factor-augmented VAR (FAVAR) framework - a setup which has the advantage that it allows time series forecasts to be made using information sets much richer than that used in traditional VAR models. While the present paper focuses on the development of the variable selection procedure itself as well as the analysis of its asymptotic properties; we have shown in an earlier, extended version of this paper, Chao and Swanson (2022a), that the use of our methodology will allow the conditional mean function of a factor-augmented forecast equation to be consistently estimated in a wide range of situations, including cases where violation of factor pervasiveness is such that consistent estimation is precluded in the absence of variable pre-screening.[2] Moreover, there are also clear benefits to using our procedure even in cases where weaker pervasiveness assumptions, such as that discussed in Bai and Ng (2021), characterize the data.

The research reported here is related to the well-known supervised principal com-

[2]See Theorem 5 of Chao and Swanson (2022a) which is available at http://econweb.umd.edu/~chao/Research/research_files/ConEstVarSelForecast-03-18-2022-main.pdf, and the the proof of Theorem 5 in that paper, which can be found in the Technical Appendix to that paper, Chao and Swanson (2022b), which is available at http://econweb.umd.edu/~chao/Research/research_files/AppConFacVarSel-03-18-2022.pdf.

ponents method proposed by Bair, Hastie, Paul, and Tibshirani (2006). Additionally, our research is related to some interesting recent work by Giglio, Xiu, and Zhang (2021), who propose a method for selecting test assets, with the objective of estimating risk premia in a Fama-MacBeth type framework. A crucial difference between the variable selection procedure proposed in our paper and those proposed in these papers is that we use a score statistic that is self-nomalized, whereas the aforementioned papers do not make use of statistics that involve self-normalization. An important advantage of self-normalized statistics is their ability to accommodate a much wider range of possible tail behavior in the underlying distributions, relative to their non-self-normalized counterparts. This makes self-normalized statistics better suited for various types of economic and financial applications, where the data are known not to exhibit the type of exponentially decaying tail behavior assumed in much of the statistics literature on high-dimensional models. In addition, the type of models studied in Bair, Hastie, Paul, and Tibshirani (2006) and Giglio, Xiu, and Zhang (2021) differ significantly from the FAVAR model studied here. In particular, Bair, Hastie, Paul, and Tibshirani (2006) study a one-factor model in an *i.i.d.* Gaussian framework, precluding complications associated with the introduction of dependence and non-normality. Giglio, Xiu, and Zhang (2021), on the other hand, make certain high-level assumptions which can accommodate some dependence both cross-sectionally and intertemporally, but the model that they consider is very different from the dynamic vector time series model studied in the sequel.

In another important related paper, Bai and Ng (2021) provide results which show that factors can still be estimated consistently in situations where factor loadings are weaker than implied by the conventional pervasiveness assumption. As might be expected, in their framework the rate of convergence of the factor estimator is slower and additional assumptions are needed. For example, as discussed in the next section of this paper, their factor consistency result depends on how severely the conventional factor pervasiveness assumption is violated. Our methodology makes related assumptions, and additionally outlines an empirical approach to assessing how severely the pervasiveness assumption is violated. One important contribution in this dimension is that we show that variable pre-selection can result consistent estimation even when the Bai and Ng condition is not acheived. The notion of how severely pervasiveness is violated is of importance, given that various authors have documented cases in economics-related research where empirical results suggest that the underlying factors may be quite weak, in which case the pervasiveness type assumption made in Bai and Ng (2021) plays an important role in the theoretical properties of factor estimators[3]. In summary, our paper aims to build on the results developed by Bai and Ng (2021) and others in contexts where factor estimator properties are impacted by failure of the conventional pervasiveness assumption.

It is also worth pointing out that our variable selection procedure differs substantially from the approach to variable/model selection taken in much of the traditional econometrics literature. In particular, we show that important moderate deviation results obtained recently by Chen, Shao, Wu, and Xu (2016) can be used to help control the probability of a Type I error, i.e., the error that an irrelevant variable which is not informative about the underlying factors is falsely selected as a relevant variable. This is so even in situations where the number of irrelevant variables is very large and even if the tails of the underlying distributions do not satisfy the kind of sub-exponential behavior typically assumed by large deviation inequalities used in high-dimensional analysis. Hence, we are able to design a variable selection procedure where the probability of a Type I error goes to zero, as the sample sizes grow to infinity. This fact, taken together with the fact that the probability of a Type II error for our procedure also goes to zero asymptotically, allows us to establish that our variable selection procedure is completely consistent, in the sense that the probabilities of both Type I and Type II errors go to zero in the limit. This property of complete consistency is important because if we try simply to control the probability of a Type I error at some predetermined non-zero level, which is the typical approach in multiple hypothesis testing, then we will not in general be able to estimate the factors consistently, even up to an invertible matrix transformation, and in consequence, we will have fallen short of our ultimate goal of obtaining a consistent estimate of the conditional mean function of the factor-augmented forecasting equation.

The rest of the paper is organized as follows. In Section 2, we discuss the FAVAR model and the assumptions that we impose on this model. We also describe our

[3]See, for example, the discussions in Jagannathan and Wang (1998), Kan and Zhang (1999), Harding (2008), Kleibergen (2009), Ontaski (2012), Bryzgalova (2016), Burnside (2016), Gospodinov, Kan, and Robotti (2017), Anatolyev and Mikusheva (2021), and Freyaldenhoven (2021a,b).

variable selection procedure and provide theoretical results establishing the complete consistency of this procedure. Section 3 presents the results of a promising Monte Carlo study on the finite sample performance of our variable selection method, and makes recommendations regarding the calibration of the tuning parameter used in the said method. Section 4 offers some concluding remarks. Proofs of the main theorems and of the supporting lemmas are given in the Appendix to this paper. In addition, we have prepared a separate Online Appendix, Chao and Swanson (2022c), which provides additional results showing that proper selection of variables using our proposed methodology can lead to the consistent estimation of latent factors even in situations where the assumption of factor pervasiveness is violated.

Before proceeding, we first say a few words about some of the frequently used notation in this paper. Throughout, let $\lambda_{(j)}(A)$, $\lambda_{\max}(A)$, and $\lambda_{\min}(A)$ denote, respectively, the j^{th} largest eigenvalue, the maximal eigenvalue, and the minimal eigenvalue of a square matrix A. Similarly, let $\sigma_{(j)}(B)$, $\sigma_{\max}(B)$, and $\sigma_{\min}(B)$ denote, respectively, the j^{th} largest singular value, the maximal singular value, and the minimal singular value of a matrix B, which is not restricted to be a square matrix. In addition, let $\|a\|_2$ denote the usual Euclidean norm when applied to a (finite-dimensional) vector a. Also, for a matrix A, $\|A\|_2 \equiv \max\left\{\sqrt{\lambda(A'A)} : \lambda(A'A) \text{ is an eigenvalue of } A'A\right\}$ denotes the matrix spectral norm. For two sequences, $\{x_T\}$ and $\{y_T\}$, write $x_T \sim y_T$ if $x_T/y_T = O(1)$ and $y_T/x_T = O(1)$, as $T \to \infty$. Furthermore, let $|z|$ denote the absolute value or the modulus of the number z; let $\lfloor\cdot\rfloor$ denote the floor function, so that $\lfloor x \rfloor$ gives the integer part of the real number x, and let $\iota_p = (1, 1, ..., 1)'$ denote a $p \times 1$ vector of ones. Finally, for a sequence of random variables $u_{i,t+m}, u_{i,t+m+1}, u_{i,t+m+2}, ...$; we let $\sigma(u_{i,t+m}, u_{i,t+m+1}, u_{i,t+m+2}, ...)$ denote the σ-field generated by this sequence of random variables.

2 Model, Assumptions, and Variable Selection in High Dimensions

Consider the following p^{th}-order factor-augmented vector autoregression (FAVAR):

$$W_{t+1} = \mu + A_1 W_t + \cdots + A_p W_{t-p+1} + \varepsilon_{t+1}, \tag{1}$$

where

$$\underset{(d+K)\times 1}{W_{t+1}} = \begin{pmatrix} \underset{d\times 1}{Y_{t+1}} \\ \underset{K\times 1}{F_{t+1}} \end{pmatrix}, \quad \underset{(d+K)\times 1}{\varepsilon_{t+1}} = \begin{pmatrix} \underset{d\times 1}{\varepsilon^Y_{t+1}} \\ \underset{K\times 1}{\varepsilon^F_{t+1}} \end{pmatrix}, \quad \underset{(d+K)\times 1}{\mu} = \begin{pmatrix} \underset{d\times 1}{\mu_Y} \\ \underset{K\times 1}{\mu_F} \end{pmatrix}, \text{ and}$$

$$\underset{(d+K)\times(d+K)}{A_g} = \begin{pmatrix} \underset{d\times d}{A_{YY,g}} & \underset{d\times K}{A_{YF,g}} \\ \underset{K\times d}{A_{FY,g}} & \underset{K\times K}{A_{FF,g}} \end{pmatrix}, \text{ for } g = 1, ..., p.$$

Here, Y_t denotes the vector of observable economic variables, and F_t is a vector of unobserved (latent) factors. In our analysis of this model, it will often be convenient to rewrite the FAVAR in several alternative forms, which will facilitate writing down assumptions and conditions used in the sequel. We thus briefly outline two alternative representations of the above model. First, it is easy to see that the system of equations given in (1) can be written in the form:

$$Y_{t+1} = \mu_Y + A_{YY}\underline{Y}_t + A_{YF}\underline{F}_t + \varepsilon^Y_{t+1}, \tag{2}$$

$$F_{t+1} = \mu_F + A_{FY}\underline{Y}_t + A_{FF}\underline{F}_t + \varepsilon^F_{t+1}, \tag{3}$$

where

$$\begin{aligned} \underset{d\times dp}{A_{YY}} &= \begin{pmatrix} A_{YY,1} & A_{YY,2} & \cdots & A_{YY,p} \end{pmatrix}, \ \underset{d\times Kp}{A_{YF}} = \begin{pmatrix} A_{YF,1} & A_{YF,2} & \cdots & A_{YF,p} \end{pmatrix}, \\ \underset{K\times dp}{A_{FY}} &= \begin{pmatrix} A_{FY,1} & A_{FY,2} & \cdots & A_{FY,p} \end{pmatrix}, \ \underset{K\times Kp}{A_{FF}} = \begin{pmatrix} A_{FF,1} & A_{FF,2} & \cdots & A_{FF,p} \end{pmatrix}, \\ \underset{dp\times 1}{\underline{Y}_t} &= \begin{pmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{pmatrix}, \text{ and } \underset{Kp\times 1}{\underline{F}_t} = \begin{pmatrix} F_t \\ F_{t-1} \\ \vdots \\ F_{t-p+1} \end{pmatrix}. \end{aligned} \tag{4}$$

Another useful representation of the FAVAR model is the so-called companion form, wherein the p^{th}-order model given in expression (1) is written in terms of a first-order model:

$$\underset{(d+K)p\times 1}{\underline{W}_t} = \alpha + A\underline{W}_{t-1} + E_t,$$

where $\underline{W}_t = \begin{pmatrix} W_t' & W_{t-1}' & \cdots & W_{t-p+2}' & W_{t-p+1}' \end{pmatrix}'$ and where

$$\alpha = \begin{pmatrix} \mu \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, A = \begin{pmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_{d+K} & 0 & \cdots & 0 & 0 \\ 0 & I_{d+K} & \ddots & \vdots & 0 \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & I_{d+K} & 0 \end{pmatrix}, \text{ and } E_t = \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}. \tag{5}$$

In addition to observations on Y_t, suppose that the data set available to researchers includes a vector of time series variables which are related to the unobserved factors in the following manner:

$$Z_t = \Gamma \underline{F}_t + u_t, \tag{6}$$

where $\underset{N\times 1}{Z_t} = (Z_{1t}, Z_{2t}, ..., Z_{Nt})'$. Assume, however, that not all components of Z_t provide useful information for estimating the unobserved vector $\underline{F}_t$, so that the $N \times Kp$ parameter matrix Γ may have some rows whose elements are all zero. More precisely, let the $1 \times Kp$ vector γ_i' denote the i^{th} row of Γ, and assume that the rows of the matrix Γ can be divided into two classes:

$$H = \{k \in \{1, ..., N\} : \gamma_k = 0\} \text{ and} \tag{7}$$

$$H^c = \{k \in \{1, ..., N\} : \gamma_k \neq 0\}. \tag{8}$$

Now, let $\mathcal{P}$ be a permutation matrix which reorders the components of Z_t such that $\mathcal{P}Z_t = \begin{pmatrix} Z_t^{(1)\prime} & Z_t^{(2)\prime} \end{pmatrix}'$, where

$$\underset{N_1\times 1}{Z_t^{(1)}} = \Gamma_1 \underline{F}_t + u_t^{(1)} \tag{9}$$

$$\underset{N_2\times 1}{Z_t^{(2)}} = u_t^{(2)}. \tag{10}$$

The above representation suggests that the components of $Z_t^{(1)}$ can be interpreted as the relevant variables for the purpose of factor estimation, as the information that they supply will be helpful in estimating $\underline{F}_t$. On the other hand, the components of the subvector $Z_t^{(2)}$ are irrelevant variables (or pure "noise" variables), as they do not load on the underlying factors and only add noise if they are included in the factor estimation process. Given that an empirical researcher will typically not have prior knowledge as to which variables are elements of $Z_t^{(1)}$ and which are elements of $Z_t^{(2)}$, it will be nice to have a variable selection procedure which will allow us to properly identify the components of $Z_t^{(1)}$ and to use only these variables when we try to estimate $\underline{F}_t$. On the other hand, if we unknowingly include too many components of $Z_t^{(2)}$ in the estimation process, then inconsistent factor estimation can arise. This is demonstrated in an example analyzed recently in Chao and Swanson (2022a) which considers a setting similar to the specification given in expressions (6)-(10) above, but for the case of a simple one-factor model. More precisely, Chao and Swanson (2022a) give an example which shows that, in this situation without variable pre-screening, the usual principal-component-based factor estimator $\widehat{f}_t \xrightarrow{p} 0$ regardless of the true value f_t under the additional rate condition that $N/\left(TN_1^{(1+\kappa)}\right) = c + o\left(N_1^{-1}\right)$, where c and κ are constants such that $0 < c < \infty$ and $0 < \kappa < 1$ and where N_1 is the number of relevant variables, N_2 is the number of irrelevant variables, and $N = N_1 + N_2$. This example shows the kind of severe inconsistency in factor estimation that could result if the commonly assumed condition of factor pervasiveness (which essentially requires that $N_1 \sim N$) does not hold[4]. This example is also related to a growing number of results which have appeared in the statistics literature showing the possible inconsistency of sample eigenvectors as estimators of population eigenvectors in high dimensional situations. See, for example, Paul (2007), Johnstone and Lu (2009), Shen, Shen, Zhu, and Marron (2016), and Johnstone and Paul (2018).

Turning again to the interesting and thought-provoking recent paper by Bai and Ng (2021), recall that they show that factors can still be estimated consistently in certain situations, even when factor loadings are weaker than implied by the conventional pervasiveness assumption, but that in such cases the rate of convergence is slower and additional assumptions are needed. To understand the relationship between their results and our setup, note that a key condition for the consistency result

[4]The reason why we refer to the result given in Chao and Swanson (2022a) as a severe form of inconsistency in factor estimation is because inconsistency of this type will preclude the consistent estimation of the conditional mean function of a factor-augmented forecast equation. This is different from the case where the factors may be estimated consistently up to a non-zero scalar multiplication or, more generally, up to an invertible matrix transformation. In the latter case, consistent estimation of the conditional mean function of a factor-augmented forecast equation can still be attained.

given in their paper, when expressed in terms of our setup, is the assumption that $N/(TN_1) \to 0$. When violation of the factor pervasiveness condition is more severe than that characterized by this rate condition (i.e., if $N/(TN_1) \to c_1$, for some positive constant c_1 or if $N/(TN_1) \to \infty$), then factors will be estimated inconsistently unless there is some method which can correctly identify relevant variables, and only these variables are used to estimate the factors. Indeed, the Online Appendix for this paper, we add to the results given in Bai and Ng (2021) by giving a result which shows that if one pre-screens variables using the variable selection method proposed below, then consistent factor estimation can be achieved, even if the rate condition that $N/(TN_1) \to 0$ is not satisfied. In general, knowledge about the severity with which the conventional factor pervasiveness assumption may be violated must ultimately be gathered on a case-by-case basis, and depends on the dataset used for a particular study. As mentioned earlier, this is important, as various authors have already documented cases where the empirical evidence suggests that the underlying factors may be quite weak, suggesting that the rate condition given in Bai and Ng (2021) may be important to assess. For example, see Jagannathan and Wang (1998), Kan and Zhang (1999), Harding (2008), Kleibergen (2009), Onatski (2012), Bryzgalova (2016), Burnside (2016), Gospodinov, Kan, and Robotti (2017), Anatolyev and Mikusheva (2021), and Freyaldenhoven (2021a,b). In such cases, it is of interest to explore the possibility that weakness in loadings is not uniform across all variables, but rather is due to the fact that only a fraction of the Z_{it} variables loads significantly on the underlying factors. Furthermore, even if the empirical situation of interest is one where, strictly speaking, the condition $N/(TN_1) \to 0$ does hold, it may still be beneficial in some such instances to do variable pre-screening. This is particularly true in situations where the condition $N/(TN_1) \to 0$ is "barely" satisfied, in which case one would expect to pay a rather hefty finite sample price for not pruning out variables that do not load significantly on the underlying factors, since the variables may add unwanted noise to the estimation process. For these reasons, we believe that there is a need to develop methods enabling empirical researchers to pre-screen the components of Z_t, so that variables which are informative and helpful to the estimation process can be properly identified.

To provide a variable selection procedure with provable guarantees, we must first specify a number of conditions on the FAVAR model defined above.

Assumption 2-1: Suppose that:

$$\det\left\{I_{(d+K)} - A_1 z - \cdots - A_p z^p\right\} = 0, \text{ implies that } |z| > 1. \tag{11}$$

Assumption 2-2: Let ε_t satisfy the following set of conditions: (a) $\{\varepsilon_t\}$ is an independent sequence of random vectors with $E[\varepsilon_t] = 0$ $\forall t$; (b) there exists a positive constant C such that $\sup_t E\|\varepsilon_t\|_2^6 \leq C < \infty$; and (c) ε_t admits a density g_{ε_t} such that, for some positive constant $M < \infty$, $\sup_t \int |g_{\varepsilon_t}(\upsilon - u) - g_{\varepsilon_t}(\upsilon)| d\upsilon \leq M\|u\|$, whenever $\|u\| \leq \overline{\kappa}$ for some constant $\overline{\kappa} > 0$.

Assumption 2-3: Let $u_{i,t}$ be the i^{th} element of the error vector u_t in expression (6), and we assume that it satisfies the following conditions: (a) $E[u_{i,t}] = 0$ for all i and t; (b) there exists a positive constant $\overline{C}$ such that $\sup_{i,t} E|u_{i,t}|^7 \leq \overline{C} < \infty$, and there exists a constant $\underline{C} > 0$ such that $\inf_{i,t} E\left[u_{i,t}^2\right] \geq \underline{C}$; and (c) define $\mathcal{F}_{i,-\infty}^t = \sigma(\ldots., u_{i,t-2}, u_{i,t-1}, u_t)$, $\mathcal{F}_{i,t+m}^{\infty} = \sigma(u_{i,t+m}, u_{i,t+m+1}, u_{i,t+m+2}, \ldots.)$, and $\beta_i(m) = \sup_t E\left[\sup\left\{\left|P\left(B|\mathcal{F}_{i,-\infty}^t\right) - P(B)\right| : B \in \mathcal{F}_{i,t+m}^{\infty}\right\}\right]$. Assume that there exist constants $a_1 > 0$ and $a_2 > 0$ such that

$$\beta_i(m) \leq a_1 \exp\{-a_2 m\}, \text{ for all } i.$$

Assumption 2-4: ε_t and $u_{i,s}$ are independent, for all i, t, and s.

Assumption 2-5: There exists a positive constant $\overline{C}$, such that $\sup_{i\in H^c}\|\gamma_i\|_2 \leq \overline{C} < \infty$ and $\|\mu\|_2 \leq \overline{C} < \infty$, where $\mu = (\mu_Y', \mu_F')'$.

Assumption 2-6: Let A be as defined in expression (5) above, and let the modulus of the eigenvalues of the matrix $I_{(d+K)p} - A$ be sorted so that:

$$\left|\lambda^{(1)}\left(I_{(d+K)p} - A\right)\right| \geq \left|\lambda^{(2)}\left(I_{(d+K)p} - A\right)\right| \geq \cdots \geq \left|\lambda^{((d+K)p)}\left(I_{(d+K)p} - A\right)\right| = \overline{\phi}_{\min}.$$

Suppose that there is a constant $\underline{C} > 0$ such that

$$\sigma_{\min}\left(I_{(d+K)p} - A\right) \geq \underline{C}\overline{\phi}_{\min} \tag{12}$$

In addition, there exists a positive constant $\overline{C} < \infty$ such that, for all positive integer

j,

$$\sigma_{\max}\left(A^j\right) \leq \overline{C} \max\left\{\left|\lambda_{\max}\left(A^j\right)\right|, \left|\lambda_{\min}\left(A^j\right)\right|\right\}. \tag{13}$$

Remark 2.1:

(a) Note that Assumption 2-1 is the stability condition that one typically assumes for a stationary VAR process. One difference is that we allow for possible heterogeneity in the distribution of ε_t across time, so that our FAVAR process is not necessarily a strictly stationary process. Under Assumption 2-1, there exists a vector moving average representation for the FAVAR process.

(b) It is well known that $\det\left\{I_{(d+K)} - Az\right\} = \det\left\{I_{(d+K)} - A_1 z - \cdots - A_p z^p\right\}$, where A is the coefficient matrix of the companion form given in expression (5). See, for example, page 16 of Lütkepohl (2005). It follows that Assumption 2-1 is equivalent to the condition that

$$\det\left\{I_{(d+K)} - Az\right\} = 0 \text{ implies that } |z| > 1. \tag{14}$$

In addition, Assumption 2-1 is also, of course, equivalent to the assumption that all eigenvalues of A have modulus less than 1.

(c) Assumption 2-6 imposes a condition whereby the extreme singular values of the matrices A^j and $I_{(d+K)p} - A$ have bounds that depend on the extreme eigenvalues of these matrices. More primitive conditions for such a relationship between the singular values and the eigenvalues of a (not necessarily symmetric) matrix have been studied in the linear algebra literature. In fact, an easy extension of a well-known result by Ruhe (1975) yields the following lemma[5]:

Lemma 1: *Let A be an $n \times n$ square matrix with (ordered) singular values given by:*

$$\sigma_{(1)}(A) \geq \sigma_{(2)}(A) \geq \cdots \geq \sigma_{(n)}(A) \geq 0.$$

Suppose that A is diagonalizable, i.e., $A = S\Lambda S^{-1}$, where Λ is diagonal matrix whose

[5] We do not give a proof of Lemma 1 in this paper because its proof follows from a straightforward extension of the proof of Theorem 1 of Ruhe (1975) and because this lemma is not central to our paper and we only state it here in order to provide a motivation for Assumption 2-6. An explicit proof of this lemma is given in the Technical Appendix of an earlier version of this paper, Chao and Swanson (2022b). See, in particular, Lemma C-9 and its proof in Appendix C of Chao and Swanson (2022b).

diagonal elements are the eigenvalues of A. Let the modulus of these eigenvalues be ordered as follows:

$$\left|\lambda^{(1)}(A)\right| \geq \left|\lambda^{(2)}(A)\right| \geq \cdots \geq \left|\lambda^{(n)}(A)\right|.$$

Then, for any $k \in \{1,...,n\}$ and for any positive integer j, we have that:

$$\chi(S)^{-1}\left|\lambda^{(k)}\left(A^j\right)\right| \leq \sigma_{(k)}\left(A^j\right) \leq \chi(S)\left|\lambda^{(k)}\left(A^j\right)\right|$$

where $\chi(S) = \sigma_{(1)}(S)\,\sigma_{(1)}(S^{-1})$.

Note that in the special case where the matrices A and $I_{(d+K)p} - A$ are diagonalizable, the inequalities given in expressions (12) and (13) are a direct consequence of this lemma. On the other hand, Assumption 2-6 takes into account other situations where expressions (12) and (13) are valid even though the matrices A and $I_{(d+K)p} - A$ are not diagonalizable.

(d) Note that Assumptions 2-1, 2-2, and 2-6 together imply that the process $\{W_t\}$ generated by the FAVAR model given in expression (1) is a β-mixing process with β-mixing coefficient satisfying:

$$\beta_W(m) \leq a_1 \exp\{-a_2 m\},$$

for some positive constants a_1 and a_2, with $\beta_W(m) = \sup_t E\left[\sup\left\{\left|P\left(B|\mathcal{A}_{-\infty}^t\right) - P(B)\right| : B \in \mathcal{A}_{t+m}^{\infty}\right\}\right]$, and with $\mathcal{A}_{-\infty}^t = \sigma(..., W_{t-2}, W_{t-1}, W_t)$ and $\mathcal{A}_{t+m}^{\infty} = \sigma(W_{t+m}, W_{t+m+1}, W_{t+m+2}, ...)$[6]. Note, in addition, that Assumption 2-2 (c) rules out situations such as that given in the famous counterexample presented by Andrews (1984) which shows that a first-order autoregression with errors having a discrete Bernoulli distribution is not α-mixing, even if it satisfies the stability condition. Conditions similar to Assumption 2-2(c) have also appeared in previous papers, such as Gorodetskii (1977) and Pham and Tran (1985), which seek to provide sufficient conditions for establishing the α or β mixing properties of linear time series processes.

[6] This can be shown by applying Theorem 2.1 of Pham and Tran (1985). A proof of this result is also given in the Technical Appendix of an earlier version of this paper, Chao and Swanson (2022b). See, in particular, Lemma C-11 and its proof in Appendix C of Chao and Swanson (2022b).

Our variable selection procedure is based on a self-normalized statistic and makes use of some pathbreaking moderate deviation results for weakly dependent processes recently obtained by Chen, Shao, Wu, and Xu (2016). An advantage of using a self-normalized statistic is that doing so allows us to impose much weaker moment conditions, even when N is much larger than T. In particular, as can be seen from Assumptions 2-2 and 2-3 above, we only make moment conditions that are of a polynomial order on the errors processes $\{\varepsilon_t\}$ and $\{u_{it}\}$. Such conditions are substantially weaker than assumption of Gaussianity or of sub-exponential tail behavior which has been made in papers studying high-dimensional factor models and/or high-dimensional covariance matrices, without employing statistics that are self-normalized[7].

To accommodate data dependence, we consider self-nomalized statistics that are constructed from observations which are first split into blocks in a manner similar to the kind of construction one would employ in implementing a block bootstrap or in proving a central limit theorem using the blocking technique. Two such statistics are proposed in this paper. The first of these statistics has the form of an ℓ_∞ norm and is given by:

$$\max_{1\leq \ell\leq d}|S_{i,\ell,T}| = \max_{1\leq \ell\leq d}\left|\frac{\overline{S}_{i,\ell,T}}{\sqrt{\overline{V}_{i,\ell,T}}}\right|, \tag{15}$$

where

$$\overline{S}_{i,\ell,T} = \sum_{r=1}^{q}\sum_{t=(r-1)\tau+p}^{(r-1)\tau+\tau_1+p-1} Z_{it}y_{\ell,t+1} \text{ and} \tag{16}$$

$$\overline{V}_{i,\ell,T} = \sum_{r=1}^{q}\left[\sum_{t=(r-1)\tau+p}^{(r-1)\tau+\tau_1+p-1} Z_{it}y_{\ell,t+1}\right]^2. \tag{17}$$

Here, Z_{it} denotes the i^{th} component of Z_t , $y_{\ell,t+1}$ denotes the ℓ^{th} component of Y_{t+1}, $\tau_1 = \lfloor T_0^{\alpha_1}\rfloor$, and $\tau_2 = \lfloor T_0^{\alpha_2}\rfloor$, where $1 > \alpha_1 \geq \alpha_2 > 0$, $\tau = \tau_1 + \tau_2$, $q = \lfloor T_0/\tau\rfloor$, and $T_0 = T - p + 1$. Note that the statistic given in expression (15) can be interpreted as the maximum of the (self-normalized) sample covariances between the i^{th} component of Z_t and the components of Y_{t+1}. Our second statistic has the form of a pseudo-L_1

[7] See, for example, Bickel and Levina (2008) and Fan, Liao, and Mincheva (2011, 2013).

norm and is given by:

$$\sum_{\ell=1}^{d} \varpi_\ell \left| S_{i,\ell,T} \right| = \sum_{\ell=1}^{d} \varpi_\ell \left| \frac{\overline{S}_{i,\ell,T}}{\sqrt{\overline{V}_{i,\ell,T}}} \right|,$$

where $\overline{S}_{i,\ell,T}$ and $\overline{V}_{i,\ell,T}$ are as defined in expressions (16) and (17) above and where $\{\varpi_\ell : \ell = 1, .., d\}$ denotes pre-specified weights, such that $\varpi_\ell \geq 0$, for every $\ell \in \{1, ..., d\}$ and $\sum_{\ell=1}^{d} \varpi_\ell = 1$. Both of these statistics employ a blocking scheme similar to that proposed in Chen, Shao, Wu, and Xu (2016), where, in order to keep the effects of dependence under control, the construction of these statistics is based only on observations in every other block. To see this, note that if we write out the "numerator" term $\overline{S}_{i,\ell,T}$ in greater detail, we have that:

$$\begin{aligned} \overline{S}_{i,\ell,T} &= \sum_{t=p}^{\tau_1+p-1} Z_{it} y_{\ell,t+1} + \sum_{t=\tau+p}^{\tau+\tau_1+p-1} Z_{it} y_{\ell,t+1} \\ &\quad + \sum_{t=2\tau+p}^{2\tau+\tau_1+p-1} Z_{it} y_{\ell,t+1} + \cdots + \sum_{t=(q-1)\tau+p}^{(q-1)\tau+\tau_1+p-1} Z_{it} y_{\ell,t+1} \end{aligned} \tag{18}$$

Comparing the first term and the second terms on the right-hand side of expression (18), we see that the observations $Z_{it} y_{\ell,t+1}$, for $t = \tau_1 + p, ..., \tau + p - 1$, have not been included in the construction of the sum. Similar observations hold when comparing the second and the third terms, and so on.

It should also be pointed out that although we make use of some of their fundamental results on moderate deviation, both the model studied in our paper and the objective of our paper are very different from that of Chen, Shao, Wu, and Xu (2016). Whereas Chen, Shao, Wu, and Xu (2016) focus their analysis on problems of testing and inference for the mean of a scalar weakly dependent time series using self-normalized Student-type test statistics, our paper applies the self-normalization approach to a variable selection problem in a FAVAR setting. Indeed, the problem which we study is in some sense more akin to a model selection problem rather than a multiple hypothesis testing problem. In order to consistently estimate the factors (at least up to an invertible matrix transformation), we need to develop a variable selection procedure whereby both the probability of a false positive and the probability of a false negative converge to zero as $N_1,\ N_2,\ T\rightarrow\infty$[8]. This is different from the typical multiple hypothesis testing approach whereby one tries to control the familywise error rate (or, alternatively, the false discovery rate), so that it is no greater than 0.05, say, but does not try to ensure that this probability goes to zero as the sample size grows.

To determine whether the i^{th} component of Z_t is a relevant variable for the purpose of factor estimation, we propose the following procedure. Define $i\in\widehat{H}^c$ to indicate that the procedure has classified Z_{it} to be a relevant variable for the purpose of factor estimation. Similarly, define $i\in\widehat{H}$ to indicate that the procedure has classified Z_{it} to be an irrelevant variable. Now, let $\mathbb{S}^+_{i,T}$ denote either the statistic $\max_{1\le\ell\le d}|S_{i,\ell,T}|$ or the statistic $\sum_{\ell=1}^{d}\varpi_\ell|S_{i,\ell,T}|$. Our variable selection procedure is based on the decision rule:

$$i\in\begin{cases}\widehat{H}^c & \text{if } \mathbb{S}^+_{i,T}\ge\Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\\ \widehat{H} & \text{if } \mathbb{S}^+_{i,T}<\Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\end{cases}, \tag{19}$$

where $\Phi^{-1}(\cdot)$ denotes the quantile function or the inverse of the cumulative distribution function of the standard normal random variable, and where φ is a tuning parameter which may depend on N. Some conditions on φ will be given in Assumption 2-10 below.

Remark 2.2:

(a) To understand why using the quantile function of the standard normal as the threshold function for our procedure is a natural choice, note first that, by a slight modification of the arguments given in the proof of Lemma A3[9], we can show that, as $T\rightarrow\infty$

$$P\left(|S_{i,\ell,T}|\ge z\right)=2\left[1-\Phi(z)\right](1+o(1)), \tag{20}$$

which holds for all i and ℓ and for all z such that $0\le z\le c_0\min\left\{T^{(1-\alpha_1)/6}/L(T),T^{\alpha_2/2}\right\}$, where $L(T)$ denotes a slowly varying function such that $L(T)\rightarrow\infty$ as $T\rightarrow\infty$. In view of expression (20), we can interpret moderate deviation as providing an asymptotic approximation of the (two-sided) tail behavior of the statistic, $S_{i,\ell,T}$, based on the tails of the standard normal distribution.

[8] Here, a false positive refers to mis-classifying a variable, Z_{it}, as a relevant variable for the purpose of factor estimation when its factor loading $\gamma_i'=0$, whereas a false negative refers to the opposite case, where $\gamma_i'\neq0$, but the variable Z_{it} is mistakenly classified as irrelevant.

[9] The statement and proof of Lemma A3 are provided below in the Appendix to this paper.

Now, suppose initially that we wish simply to control the probability of a Type I error for testing the null hypothesis $H_0 : \gamma_i = 0$ (i.e., the i^{th} variable does not load on the underlying factors) at some fixed significance level α. Then, expression (20) suggests that a natural way to do this is to set $z = \Phi^{-1}(1 - \alpha/2)$. This is because, given that the quantile function $\Phi^{-1}(\cdot)$ is, by definition, the inverse function of the cdf $\Phi(\cdot)$, we have that:

$$P\left(|S_{i,\ell,T}| \geq \Phi^{-1}(1 - \alpha/2)\right) \leq 2\left[1 - \Phi\left(\Phi^{-1}(1 - \alpha/2)\right)\right](1 + o(1)) = \alpha(1 + o(1)),$$

so that the probability of a Type I error is controlled at the desired level α asymptotically. Note also that an advantage of moderate deviation theory is that it gives a characterization of the relative approximation error, as opposed to the absolute approximation error. As a result, the approximation given is useful and meaningful even when α is very small, which is of importance to us since we are interested in situations where we might want to let α go to zero, as sample size approaches infinity.

We give the above example to provide some intuition concerning the form of the threshold function that we have specified. The variable selection problem that we actually consider is more complicated than what is illustrated by this example, since we need to control the probability of a Type I error (or of a false positive) not just for a single test involving the i^{th} variable but for all variables simultaneously. Moreover, as noted previously, we also need the probability of a false positive to go to zero asymptotically, if we want to be able to estimate the factors consistently, even up to an invertible matrix transformation. We show in Theorem 1 below that these objectives can all be accomplished using the threshold function specified in expression (19), since a threshold function of this form makes it easy for us to properly control the probability of a false positive in large samples.

(b) The threshold function used here is reminiscent of the one employed in a celebrated paper by Belloni, Chen, Chernozhukov, and Hansen (2012). More specifically, Belloni, Chen, Chernozhukov, and Hansen (2012) use a similar threshold function to help set the penalty level for Lasso estimation of the first-stage equation of an IV regression model assuming *i.n.i.d.* data. In spite of the similarity in the form of the threshold function, the problem studied in that paper is very different from the one which we analyze. In consequence, the conditions we specify for setting the tuning parameter φ will also be quite different from what they recommend in their paper.

Under appropriate conditions, the variable selection procedure described above can be shown to be consistent, in the sense that both the probability of a false positive, i.e. $P\left(i \in \widehat{H}^c | i \in H\right)$, and the probability of a false negative, i.e., $P\left(i \in \widehat{H} | i \in H^c\right)$, approach zero as $N_1, N_2, T \to \infty$. To show this result, we must first state a number of additional assumptions.

Assumption 2-7: There exists a positive constant $\underline{c}$ such that for all $r \geq 1$ and $\tau_1 \geq 1$:

$$\min_{1\leq \ell \leq d} \min_{i\in H} \min_{r\in\{1,\ldots,q\}} E\left\{\left[\frac{1}{\sqrt{\tau_1}} \sum_{t=(r-1)\tau+p}^{(r-1)\tau+\tau_1+p-1} y_{\ell,t+1} u_{it}\right]^2\right\} \geq \underline{c},$$

where, as defined earlier, $\tau_1 = \lfloor T_0^{\alpha_1} \rfloor$, $\tau_2 = \lfloor T_0^{\alpha_2} \rfloor$ for $1 > \alpha_1 \geq \alpha_2 > 0$ and $q = \left\lfloor \frac{T_0}{\tau_1+\tau_2} \right\rfloor$, and $T_0 = T - p + 1$.

Assumption 2-8: Let $i \in H^c = \{k \in \{1, ..., N\} : \gamma_k \neq 0\}$. Suppose that there exists a positive constant, $\underline{c}$, such that, for all N_1, N_2,and T sufficiently large:

$$\begin{aligned}
& \min_{1\leq \ell \leq d} \min_{i\in H^c} \left| \frac{\mu_{i,\ell,T}}{q\tau_1} \right| \\
= & \min_{1\leq \ell \leq d} \min_{i\in H^c} \left| \frac{1}{q} \sum_{r=1}^{q} \frac{1}{\tau_1} \sum_{t=(r-1)\tau+p}^{(r-1)\tau+\tau_1+p-1} \gamma_i' \left\{ E\left[\underline{F}_t\right] \mu_{Y,\ell} + E\left[\underline{F}_t \underline{Y}_t'\right] \alpha_{YY,\ell} + E\left[\underline{F}_t \underline{F}_t'\right] \alpha_{YF,\ell} \right\} \right| \\
\geq & \ \underline{c} > 0,
\end{aligned}$$

where $\mu_{Y,\ell} = e_{\ell,d}' \mu_Y$, $\alpha_{YY,\ell} = A_{YY}' e_{\ell,d}$, and $\alpha_{YF,\ell} = A_{YF}' e_{\ell,d}$. Here, $e_{\ell,d}$ is a $d \times 1$ elementary vector whose ℓ^{th} component is 1 and all other components are 0.

Assumption 2-9: Suppose that, as N_1, N_2, and $T \to \infty$, the following rate conditions hold:

(a) $\sqrt{\ln N} / \min\left\{T^{(1-\alpha_1)/6}, T^{\alpha_2/2}\right\} \to 0$, where $1 > \alpha_1 \geq \alpha_2 > 0$ and $N = N_1 + N_2$.

(b) $N_1 / T^{3\alpha_1} \to 0$ where α_1 is as defined in part (a) above.

Assumption 2-10: Let φ satisfy the following two conditions: (a) $\varphi \to 0$ as $N_1, N_2 \to \infty$, and (b) there exists some constant $a > 0$, such that $\varphi \geq 1/N^a$, for all N_1, N_2 sufficiently large.

Remark 2.3:

(a) Assumption 2-9 imposes the condition that there exists a positive constant, $\underline{c}$, such that, for all N_1, N_2, and T sufficiently large:

$$\min_{1\leq \ell \leq d} \min_{i\in H^c} \left| \frac{1}{q} \sum_{r=1}^{q} \frac{1}{\tau_1} \sum_{t=(r-1)\tau+p}^{(r-1)\tau+\tau_1+p-1} \gamma_i' \left\{ E\left[\underline{F}_t\right] \mu_{Y,\ell} + E\left[\underline{F}_t \underline{Y}_t'\right] \alpha_{YY,\ell} + E\left[\underline{F}_t \underline{F}_t'\right] \alpha_{YF,\ell} \right\} \right|$$
$$\geq \quad \underline{c} > 0.$$

This is a fairly mild condition which allows us to differentiate the alternative hypothesis, $i \in H^c$, from the null hypothesis, $i \in H$, since if $i \in H$, then it is clear that:

$$\frac{\mu_{i,\ell,T}}{q\tau_1} = \frac{1}{q} \sum_{r=1}^{q} \frac{1}{\tau_1} \sum_{t=(r-1)\tau+p}^{(r-1)\tau+\tau_1+p-1} \gamma_i' \left\{ E\left[\underline{F}_t\right] \mu_{Y,\ell} + E\left[\underline{F}_t \underline{Y}_t'\right] \alpha_{YY,\ell} + E\left[\underline{F}_t \underline{F}_t'\right] \alpha_{YF,\ell} \right\} = 0,$$

given that $\gamma_i = 0$. Note that this assumption does rule out certain specialized situations, such as the case when $\mu_{Y,\ell} = 0$, $\alpha_{YY,\ell} = 0$, and $\alpha_{YF,\ell} = 0$, for some $\ell \in \{1, ..., d\}$. However, we do not consider such cases to be of much practical interest since, for example, if $\mu_{Y,\ell} = 0$, $\alpha_{YY,\ell} = 0$, and $\alpha_{YF,\ell} = 0$ for some ℓ then expression (2) above implies that the ℓ^{th} component of Y_{t+1} will have the representation

$$y_{\ell,t+1} = \mu_{Y,\ell} + \underline{Y}_t' \alpha_{YY,\ell} + \underline{F}_t' \alpha_{YF,\ell} + \varepsilon_{\ell,t+1}^{Y} = \varepsilon_{\ell,t+1}^{Y},$$

so that, in this case, $y_{\ell,t+1}$ depends neither on $\underline{Y}_t = \left(Y_t', Y_{t-1}', ..., Y_{t-p+1}'\right)'$ nor on $\underline{F}_t = \left(F_t', F_{t-1}', ..., F_{t-p+1}'\right)$. This is, of course, an unrealistic model for $y_{\ell,t+1}$ since it would not even be a dependent process in this case.

(b) Bai and Ng (2008) address the important issue of pre-selecting variables Z_{it} based on their predictability for Y_{t+1}. Our selection approach is related to theirs. However, it is worth stressing that for the FAVAR model considered here, whether Z_{it} helps predict future values of Y_t (say, Y_{t+h}) depends on two things: (i) whether Z_{it} loads significantly on the underlying factors $\underline{F}_t$ (i.e., whether $\gamma_i \neq 0$ or not) and (ii) whether at least some components of $\underline{F}_t$ are helpful for predicting certain components of Y_{t+h}. The variable selection procedure which we propose focuses on the first issue but not the second. Thus, we focus on obtaining factor estimates with desirable asymptotic properties before trying to assess which factor(s) may or may not be useful for predicting Y_{t+h}. Note that, for a given t, the precision with which $\underline{F}_t$ is estimated depends primarily on the size of the cross-sectional dimension, and the exclusion of any relevant Z_{it} (with $\gamma_i \neq 0$) will have the negative effect of reducing the sample size used for this estimation. More importantly, as is discussed in greater detail above and in the Online Appendix[10], if we exclude a significant number of variables (at the variable selection stage) that load strongly on at least some of the factors, this can result in $\underline{F}_t$ being inconsistently estimated. While the question of predictability is certainly an important one, the answer we get for this question can, in some situations, be at odds with the objective of achieving consistent factor estimation. This is because while $\gamma_i' = 0$ does imply that $Z_{i\cdot}$ will not be helpful for predicting future values of Y, the reverse is not necessarily true. On the other hand, to ensure consistent estimation of the factors, we would like to use every data point Z_{it}, for which $\gamma_i' \neq 0$. Furthermore, if it is true that some of the factors load primarily on variables which are uninformative predictors for certain components of Y_{t+h}, then that will show up in the form of certain parameter restrictions on the forecasting equation, in which case the best way to address this problem is to perform hypothesis testing or model selection on the forecasting equation itself, after the unobserved factors have first been properly estimated.

The following two theorems give our main theoretical results on the variable selection procedure described above.

Theorem 1: *Let $H = \{k \in \{1, ..., N\} : \gamma_k = 0\}$. Suppose that Assumptions 2-1, 2-2, 2-3, 2-4, 2-5, 2-6, 2-7, 2-9 (a) and 2-10 hold. Let $\Phi^{-1}(\cdot)$ denote the inverse of the cumulative distribution function of the standard normal random variable, or, alternatively, the quantile function of the standard normal distribution. Then the following statements are true:*

(a) *Let $\{\varpi_\ell : \ell = 1, .., d\}$ be pre-specified weights such that $\varpi_\ell \geq 0$ for every $\ell \in \{1, ..., d\}$ and $\sum_{\ell=1}^{d} \varpi_\ell = 1$, then:*

$$P\left(\max_{i \in H} \sum_{\ell=1}^{d} \varpi_\ell \left|S_{i,\ell,T}\right| \geq \Phi^{-1}\left(1 - \frac{\varphi}{2N}\right)\right) = O\left(\frac{N_2 \varphi}{N}\right) = o(1),$$

where $N = N_1 + N_2$.

[10] See part (a) of Remark OA1.2 of the Online Appendix.

(b)
$$P\left(\max_{i\in H}\max_{1\leq\ell\leq d}|S_{i,\ell,T}|\geq\Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right)=O\left(\frac{N_2\varphi}{N}\right)=o\left(1\right).$$

Theorem 2: *Let $H^c=\{k\in\{1,....,N\}:\gamma_k\neq 0\}$. Suppose that Assumptions 2-1, 2-2, 2-3, 2-5, 2-6, 2-8, 2-9, and 2-10 hold. Then the following statements are true.*

(a) *Let $\{\varpi_\ell:\ell=1,..,d\}$ be pre-specified weights such that $\varpi_\ell\geq 0$ for every $\ell\in\{1,...,d\}$ and $\sum_{\ell=1}^{d}\varpi_\ell=1$, then:*
$$P\left(\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell\left|S_{i,\ell,T}\right|\geq\Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right)\rightarrow 1.$$

(b)
$$P\left(\min_{i\in H^c}\max_{1\leq\ell\leq d}|S_{i,\ell,T}|\geq\Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right)\rightarrow 1.$$

Remark 2.4:
(a) Theorem 1 shows that, for both of our statistics, the probability of a false positive approaches zero uniformly over all $i\in H$ as $N_1,N_2,T\rightarrow\infty$. The results of Theorem 2 further imply that, for both of our statistics, the probability of a false negative also approaches zero, uniformly over all $i\in H^c$ as $N_1,N_2,T\rightarrow\infty$. Together, these two theorems show that our variable selection procedure is (completely) consistent in the sense that the probability of committing a misclassification error vanishes as $N_1,N_2,T\rightarrow\infty$.
(b) Note that our variable selection procedure delivers a consistent estimate of N_1 (i.e., $\widehat{N}_1$), and shown in Lemma OA-2 part (a) of the Online Appendix, where we establish that $\widehat{N}_1/N_1\xrightarrow{p}1$. The estimator $\widehat{N}_1$ is useful to empirical researchers implementing the methodology developed in this paper, and also to empiricists interested in assessing the rate condition given in Bai and Ng (2021) for consistent factor estimation (i.e., Assumption A4 in their paper) that is discussed above. This is another way in which the methods developed in this paper build on the work of Bai and Ng (2021).
(c) In addition, note that knowledge of the number of factors is not needed to implement our variable selection procedure. In the case where the number of factors

needs to be determined empirically, an applied researcher can first use our procedure to select the relevant variables and then apply an information criterion such as that proposed in Bai and Ng (2002) to estimate the number of factors.

3 Monte Carlo Study

In this section, we report some simulation results on the finite sample performance of our variable selection procedure. The model used in the Monte Carlo study is the following tri-variate FAVAR(1) process:

$$
\begin{aligned}
W_t &= \mu + AW_{t-1} + \varepsilon_t, && (21) \\
Z_t &= \gamma F_t + u_t, && (22)
\end{aligned}
$$

where

$$
W_t = \begin{pmatrix} Y_{1t} \\ Y_{2t} \\ F_t \end{pmatrix}, \mu = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}, A = \begin{pmatrix} 0.9 & 0.3 & 0.5 \\ 0 & 0.7 & 0.1 \\ 0 & 0.6 & 0.7 \end{pmatrix}, \text{ and } \gamma = \begin{pmatrix} \iota_{N_1} \\ \underset{N_2 \times 1}{0} \end{pmatrix},
$$

with ι_{N_1} denoting an $N_1 \times 1$ vector of ones. We consider different configurations of N, N_1, and T, as given in the tables below. For the error process in equation (21), we take $\{\varepsilon_t\} \equiv i.i.d.N(0, \Sigma_\varepsilon)$, where:

$$
\Sigma_\varepsilon = \begin{pmatrix} 1.3 & 0.99 & 0.641 \\ 0.99 & 0.81 & 0.009 \\ 0.641 & 0.009 & 5.85 \end{pmatrix}.
$$

The error process, $\{u_{it}\}$, in equation (22) is allowed to exhibit both temporal and cross-sectional dependence and also conditional heteroskedasticity. More specifically, we let $u_{it} = 0.8u_{it-1} + \zeta_{it}$, and following the approach for modeling cross-sectional dependence given in the Monte Carlo design of Stock and Watson (2002a), we specify: $\zeta_{it} = (1 + b^2)\eta_{it} + b\eta_{i+1,t} + b\eta_{i-1,t}$, and set $b = 1$. In addition, $\eta_{it} = \omega_{it}\xi_{it}$, with $\{\xi_{it}\} \equiv i.i.d.N(0, 1)$ independent of $\{\varepsilon_t\}$, and ω_{it} follows a GARCH(1,1) process given by: $\omega_{it}^2 = 1 + 0.9\omega_{it-1}^2 + 0.05\eta_{it-1}^2$. To study the effects of varying the tuning

Table 1: $\mathbb{S}_{i,T}^{+} = \max_{1 \le \ell \le d} |S_{i,\ell,T}|$

		$N = 100$	$N_1 = 50$	$T = 100$	$\tau = 5$		
		$\varphi = N^{-0.2}$	$\varphi = N^{-0.3}$	$\varphi = N^{-0.4}$	$\varphi = N^{-0.5}$	$\varphi = N^{-0.6}$	$\varphi = N^{-0.7}$
$\tau_1 = 2$	FPR	0.01690	0.00960	0.00464	0.00218	0.00096	0.00034
	FNR	0.00218	0.00548	0.01328	0.03204	0.07274	0.15890
$\tau_1 = 3$	FPR	0.02078	0.01156	0.00632	0.00288	0.00128	0.00048
	FNR	0.00126	0.00350	0.00866	0.02234	0.05374	0.12050
$\tau_1 = 4$	FPR	0.02544	0.01468	0.00826	0.00408	0.00194	0.00070
	FNR	0.00090	0.00228	0.00582	0.01582	0.04010	0.09362
$\tau_1 = 5$	FPR	0.03208	0.01980	0.01100	0.00584	0.00288	0.00122
	FNR	0.00052	0.00164	0.00430	0.01140	0.02988	0.07190

Results based on 1000 simulations.

parameter, we let $\varphi = N^{-\vartheta}$, and consider six different values of ϑ, i.e., $\vartheta = 0.2$, 0.3, 0.4, 0.5, 0.6, and 0.7. We also attempt to shed light on the effects of forming blocks of different sizes on the performance of our procedure. To do this, for $T = 100$, we set $\tau_1 = 2$, 3, 4, and 5; for $T = 200$, we set $\tau_1 = 5$, 6, 8, and 10; and for $T = 600$, we set $\tau_1 = 6$, 8, 10, and 12. In addition, we present results for both statistics, i.e. $\max_{1 \le \ell \le d} |S_{i,\ell,T}|$ and $\sum_{\ell=1}^{d} \varpi_\ell |S_{i,\ell,T}|$. Note that $d = 2$ in our setup; and, for the statistic $\sum_{\ell=1}^{2} \varpi_\ell |S_{i,\ell,T}|$, we set $\varpi_1 = \varpi_2 = 1/2$.

The results of our Monte Carlo study are reported in Tables 1-8. In these tables, we let FPR denote the "False Positive Rate" or the "Type I" error rate, i.e., the proportion of cases where an irrelevant variable Z_{it}, with associated coefficient $\gamma_i = 0$, is erroneously selected as a relevant variable. We let FNR denote the "False Negative Rate" or the "Type II" error rate, i.e., the proportion of cases where a relevant variable is erroneously identified as being irrelevant.

Looking across each row of the tables, note that FPRs decrease when moving from left to right, whereas FNRs increase. This is not surprising, because moving from $\varphi = N^{-0.2}$ to $\varphi = N^{-0.7}$ for a given N results in smaller values of the tuning parameter φ, and the specified threshold $\Phi^{-1}\left(1 - \frac{\varphi}{2N}\right)$ thus becomes larger. Overall, our results indicate that choosing $\varphi = N^{-\vartheta}$ with $\vartheta = 0.2$, 0.3, or 0.4 leads to very good performance, since with these choices, neither FPR nor FNR exceeds 0.1 in any of the cases studied here. In fact, both are smaller than 0.05 in a vast majority of the cases. In contrast, choosing $\vartheta = 0.6$ or 0.7 can lead to high FNRs, as these values can set our threshold at such a high level that our procedure ends up having very little

Table 2: $\mathbb{S}^{+}_{i,T} = \sum_{\ell=1}^{d} \varpi_\ell \, |S_{i,\ell,T}|$

		$N = 100$	$N_1 = 50$	$T = 100$	$\tau = 5$		
		$\varphi = N^{-0.2}$	$\varphi = N^{-0.3}$	$\varphi = N^{-0.4}$	$\varphi = N^{-0.5}$	$\varphi = N^{-0.6}$	$\varphi = N^{-0.7}$
$\tau_1 = 2$	FPR	0.01460	0.00810	0.00382	0.00174	0.00076	0.00028
	FNR	0.00284	0.00700	0.01674	0.04058	0.09412	0.19952
$\tau_1 = 3$	FPR	0.01810	0.00996	0.00526	0.00226	0.00092	0.00032
	FNR	0.00172	0.00450	0.01100	0.02860	0.06942	0.15378
$\tau_1 = 4$	FPR	0.02224	0.01276	0.00702	0.00338	0.00162	0.00044
	FNR	0.00118	0.00310	0.00828	0.02082	0.05194	0.12132
$\tau_1 = 5$	FPR	0.02796	0.01714	0.00924	0.00502	0.00232	0.00080
	FNR	0.00084	0.00222	0.00574	0.01508	0.03948	0.09456

Results based on 1000 simulations.

Table 3: $\mathbb{S}^{+}_{i,T} = \max_{1 \le \ell \le d} |S_{i,\ell,T}|$

		$N = 200$	$N_1 = 100$	$T = 100$	$\tau = 5$		
		$\varphi = N^{-0.2}$	$\varphi = N^{-0.3}$	$\varphi = N^{-0.4}$	$\varphi = N^{-0.5}$	$\varphi = N^{-0.6}$	$\varphi = N^{-0.7}$
$\tau_1 = 2$	FPR	0.00578	0.00239	0.00085	0.00020	0.00005	0.00000
	FNR	0.01074	0.02997	0.07812	0.18957	0.39889	0.68275
$\tau_1 = 3$	FPR	0.00775	0.00324	0.00126	0.00038	0.00006	0.00001
	FNR	0.00724	0.02088	0.05676	0.14547	0.32908	0.60780
$\tau_1 = 4$	FPR	0.00981	0.00457	0.00170	0.00057	0.00014	0.00002
	FNR	0.00517	0.01494	0.04224	0.11350	0.27048	0.53471
$\tau_1 = 5$	FPR	0.01334	0.00609	0.00266	0.00094	0.00023	0.00004
	FNR	0.00362	0.01133	0.03244	0.08901	0.22162	0.46424

Results based on 1000 simulations.

Table 4: $\mathbb{S}^{+}_{i,T} = \sum_{\ell=1}^{d} \varpi_\ell \, |S_{i,\ell,T}|$

		$N = 200$	$N_1 = 100$	$T = 100$	$\tau = 5$		
		$\varphi = N^{-0.2}$	$\varphi = N^{-0.3}$	$\varphi = N^{-0.4}$	$\varphi = N^{-0.5}$	$\varphi = N^{-0.6}$	$\varphi = N^{-0.7}$
$\tau_1 = 2$	FPR	0.00486	0.00196	0.00064	0.00014	0.00002	0.00000
	FNR	0.01415	0.03813	0.09966	0.23933	0.48356	0.77511
$\tau_1 = 3$	FPR	0.00657	0.00268	0.00098	0.00024	0.00005	0.00001
	FNR	0.00921	0.02714	0.07372	0.18714	0.40894	0.70884
$\tau_1 = 4$	FPR	0.00841	0.00378	0.00133	0.00043	0.00004	0.00002
	FNR	0.00661	0.01975	0.05564	0.14734	0.34279	0.63906
$\tau_1 = 5$	FPR	0.01124	0.00509	0.00213	0.00069	0.00017	0.00002
	FNR	0.00477	0.01475	0.04258	0.11741	0.28620	0.56845

Results based on 1000 simulations.

Table 5: $\mathbb{S}^{+}_{i,T} = \max_{1 \le \ell \le d} |S_{i,\ell,T}|$

		$N = 400$	$N_1 = 200$	$T = 200$	$\tau = 10$		
		$\varphi = N^{-0.2}$	$\varphi = N^{-0.3}$	$\varphi = N^{-0.4}$	$\varphi = N^{-0.5}$	$\varphi = N^{-0.6}$	$\varphi = N^{-0.7}$
$\tau_1 = 5$	FPR	0.00035	0.00009	0.00003	0.00001	0.00000	0.00000
	FNR	0.00200	0.01116	0.05764	0.23070	0.61173	0.94453
$\tau_1 = 6$	FPR	0.00040	0.00010	$2.5{\times}10^{-5}$	$5.0{\times}10^{-6}$	0.00000	0.00000
	FNR	0.00128	0.00740	0.04154	0.18482	0.54582	0.92176
$\tau_1 = 8$	FPR	0.00054	0.00015	0.00005	0.00001	0.00000	0.00000
	FNR	0.00054	0.00369	0.02191	0.11627	0.41851	0.85806
$\tau_1 = 10$	FPR	0.00093	0.00031	0.00008	$1.5{\times}10^{-5}$	$5.0{\times}10^{-6}$	0.00000
	FNR	0.00026	0.00194	0.01218	0.07226	0.30765	0.76833

Results based on 1000 simulations.

Table 6: $\mathbb{S}^{+}_{i,T} = \sum_{\ell=1}^{d} \varpi_\ell |S_{i,\ell,T}|$

		$N = 400$	$N_1 = 200$	$T = 200$	$\tau = 10$		
		$\varphi = N^{-0.2}$	$\varphi = N^{-0.3}$	$\varphi = N^{-0.4}$	$\varphi = N^{-0.5}$	$\varphi = N^{-0.6}$	$\varphi = N^{-0.7}$
$\tau_1 = 5$	FPR	0.00030	$8.5{\times}10^{-5}$	$2.5{\times}10^{-5}$	$5.0{\times}10^{-6}$	0.00000	0.00000
	FNR	0.00231	0.01355	0.06894	0.26683	0.67266	0.96749
$\tau_1 = 6$	FPR	0.00034	$9.5{\times}10^{-5}$	0.00002	$5.0{\times}10^{-6}$	0.00000	0.00000
	FNR	0.00148	0.00901	0.05058	0.21713	0.60968	0.95287
$\tau_1 = 8$	FPR	0.00046	0.00013	0.00004	0.00001	0.00000	0.00000
	FNR	0.00068	0.00448	0.02712	0.14045	0.48133	0.90649
$\tau_1 = 10$	FPR	0.00079	0.00026	$7.5{\times}10^{-5}$	0.00001	$5.0{\times}10^{-6}$	0.00000
	FNR	0.00034	0.00246	0.01535	0.08934	0.36382	0.83510

Results based on 1000 simulations.

Table 7: $\mathbb{S}^{+}_{i,T} = \max_{1 \le \ell \le d} |S_{i,\ell,T}|$

		$N = 1000$	$N_1 = 500$	$T = 600$	$\tau = 12$		
		$\varphi = N^{-0.2}$	$\varphi = N^{-0.3}$	$\varphi = N^{-0.4}$	$\varphi = N^{-0.5}$	$\varphi = N^{-0.6}$	$\varphi = N^{-0.7}$
$\tau_1 = 6$	FPR	0.00044	0.00017	$7.4{\times}10^{-5}$	$2.8{\times}10^{-5}$	0.00001	$2.0{\times}10^{-6}$
	FNR	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
$\tau_1 = 8$	FPR	0.00054	0.00023	$9.6{\times}10^{-5}$	$4.2{\times}10^{-5}$	$1.6{\times}10^{-5}$	$8.0{\times}10^{-6}$
	FNR	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
$\tau_1 = 10$	FPR	0.00080	0.00038	0.00018	0.00007	$3.6{\times}10^{-5}$	$2.0{\times}10^{-5}$
	FNR	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
$\tau_1 = 12$	FPR	0.00127	0.00068	0.00031	0.00015	$6.8{\times}10^{-5}$	$3.0{\times}10^{-5}$
	FNR	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Results based on 1000 simulations.

Table 8: $\mathbb{S}^+_{i,T} = \sum_{\ell=1}^{d} \varpi_\ell \, |S_{i,\ell,T}|$

		$N = 1000$	$N_1 = 500$	$T = 600$	$\tau = 12$		
		$\varphi = N^{-0.2}$	$\varphi = N^{-0.3}$	$\varphi = N^{-0.4}$	$\varphi = N^{-0.5}$	$\varphi = N^{-0.6}$	$\varphi = N^{-0.7}$
$\tau_1 = 6$	FPR	0.00038	0.00015	0.00006	$2.6{\times}10^{-5}$	0.00001	$2.0{\times}10^{-6}$
	FNR	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
$\tau_1 = 8$	FPR	0.00049	0.00020	$8.2{\times}10^{-5}$	$3.4{\times}10^{-5}$	$1.4{\times}10^{-5}$	$6.0{\times}10^{-6}$
	FNR	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
$\tau_1 = 10$	FPR	0.00072	0.00033	0.00016	0.00006	$3.2{\times}10^{-5}$	$1.8{\times}10^{-5}$
	FNR	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
$\tau_1 = 12$	FPR	0.00115	0.00062	0.00028	0.00014	$6.0{\times}10^{-5}$	$2.8{\times}10^{-5}$
	FNR	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Results based on 1000 simulations.

power. A particularly attractive choice of the tuning parameter is to take $\varphi = N^{-0.4}$. As discussed in part (b) of Remark OA1.1 of the Online Appendix, this choice of the tuning parameter allows the rate condition given in Assumption OA-4 part (c) of the Online Appendix to be satisfied as long as $N_1 \to \infty$, so that we do not need to make a further assumption on the rate at which N_1 grows. Assumption OA-4, in turn, is a condition that is needed to ensure consistent factor estimation using the selected variables. See Theorem 3 and Remark OA1.1(b) of the Online Appendix for further discussion.

Looking down the columns of each table, note that FPR tends to increase as τ_1 increases, whereas FNR tends to decrease as τ_1 increases. As an explanation for this result, note first that the smaller is τ_1 relative to τ, the larger is τ_2 (since $\tau = \tau_1 + \tau_2$), and thus the larger is the number of observations removed when constructing the self-normalized block sums. Intuitively, this can lead to better accommodation of the effects of dependence and better moderate deviation approximations under the null hypothesis, resulting in a lower FPR. However, removal of a larger number of observations can also lead to a reduction in power, when the alternative hypothesis is correct, so that a negative consequence of having a smaller τ_1 relative to τ is that FNR will tend to be higher in this case. The opposite, of course, occurs when we try to specify a larger τ_1 relative to τ.

Our results also show that when the sample sizes are large enough such as the cases presented in Tables 7 and 8, where $T = 600$ and $N = 1000$, then both FPR and FNR are small for all of the cases that we consider. This is in accord with the results of our theoretical analysis, which shows that our variable selection procedure is completely consistent in the sense that both the probability of a false positive and the probability of a false negative approach zero, as the sample sizes go to infinity.

A final observation based on these Monte Carlo results is that there does not seem to be a great deal of difference in the performance of the statistic $\max_{1\leq \ell \leq d} |S_{i,\ell,T}|$ vis-à-vis the statistic $\sum_{\ell=1}^{d} \varpi_\ell |S_{i,\ell,T}|$. Overall, the statistic $\max_{1\leq \ell \leq d} |S_{i,\ell,T}|$ seems to be a bit better at controlling FNR, whereas the statistic $\sum_{\ell=1}^{d} \varpi_\ell |S_{i,\ell,T}|$ seems a bit better at controlling FPR.

4 Conclusion

In this paper, we propose a new variable selection procedure based on two alternative self-normalized score statistics and provide asymptotic analyses showing that our procedure, based on either of these statistics, correctly identify the set of variables which load significantly on the underlying factors, with probability approaching one as the sample sizes go to infinity. Our research is motivated by the observation that inconsistency in factor estimation could result in high dimensional settings when the conventional assumption of factor pervasiveness does not hold. Hence, in such settings, it is particularly important to pre-screen the variables in terms of their association with the underlying factors prior to estimation. We conduct a small Monte Carlo study which yields encouraging evidence about the finite sample properties of our variable selection procedure. In addition, in an Online Appendix, Chao and Swanson (2022c), which accompanies this paper, we prove that consistent estimation of factors (up to an invertible matrix transformation) can be achieved by estimating factors using only those variables selected by our method, and this is so even in situations where the standard pervasiveness assumption does not hold. Finally, it is worth noting that in an earlier version of this paper (Chao and Swanson (2022a)) we show that by plugging factors estimated in such a manner into the factor-augmented forecasting equation implied by the FAVAR model, the conditional mean function of the forecasting equation can be consistently estimated, even for the case of multi-step ahead forecasts. In sum, the collective body of results discussed in this paper indicates that the variable selection methodology introduced in this paper can be useful to empirical researchers as they engage in the important tasks of factor estimation and the construction of point forecasts based on factor-augmented forecasting equations.

References

[1] Anatolyev, S. and A. Mikusheva (2021): "Factor Models with Many Assets: Strong Factors, Weak Factors, and the Two-Pass Procedure," *Journal of Econometrics*, forthcoming.

[2] Andrews, D.W.K. (1984): "Non-strong Mixing Autoregressive Processes," *Journal of Applied Probability*, 21, 930-934.

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

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

[5] Bai, J. and S. Ng (2008): "Forecasting Economic Time Series Using Targeted Predictors," *Journal of Econometrics*, 146, 304-317.

[6] Bai, J. and S. Ng (2021): "Approximate Factor Models with Weaker Loading," Working Paper, Columbia University.

[7] Bair, E., T. Hastie, D. Paul, and R. Tibshirani (2006): "Prediction by Supervised Principal Components," *Journal of the American Statistical Association*, 101, 119-137.

[8] Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012): "Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain," *Econometrica*, 80, 2369-2429.

[9] Bickel, P. J. and E. Levina (2008): "Covariance Regularization by Thresholding," *Annals of Statistics*, 36, 2577-2604.

[10] Bryzgalova, S. (2016): "Spurious Factors in Linear Asset Pricing Models," Working Paper, Stanford Graduate School of Business.

[11] Burnside, C. (2016): "Identification and Inference in Linear Stochastic Discount Factor Models with Excess Returns," *Journal of Financial Econometrics*, 14, 295-330.

[12] Chao, J. C. and N. R. Swanson (2022a): "Consistent Estimation, Variable Selection, and Forecasting in Factor-Augmented VAR Models," Working Paper, Rutgers University and University of Maryland.

[13] Chao, J. C. and N. R. Swanson (2022b): Technical Appendix to "Consistent Estimation, Variable Selection, and Forecasting in Factor-Augmented VAR Models," Working Paper, Rutgers University and University of Maryland.

[14] Chao, J. C. and N. R. Swanson (2022c): Online Appendix to "Selecting the Relevant Variables for Factor Estimation in a Factor-Augmented VAR Model," Working Paper, Rutgers University and University of Maryland.

[15] Chen, X., Q. Shao, W. B. Wu, and L. Xu (2016): "Self-normalized Cramér-type Moderate Deviations under Dependence," *Annals of Statistics*, 44, 1593-1617.

[16] Davidson. J. (1994): *Stochastic Limit Theory: An Introduction for Econometricians*. New York: Oxford University Press.

[17] Fan, J., Y. Liao, and M. Mincheva (2011): "High-dimensional Covariance Matrix Estimation in Approximate Factor Models," *Annals of Statistics*, 39, 3320-3356.

[18] Fan, J., Y. Liao, and M. Mincheva (2013): "Large Covariance Estimation by Thresholding Principal Orthogonal Complements," *Journal of the Royal Statistical Society, Series B*, 75, 603-680.

[19] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005): "The Generalized Dynamic Factor Model, One-Sided Estimation and Forecasting," *Journal of the American Statistical Association*, 100, 830-840.

[20] Freyaldenhoven, S. (2021a): "Factor Models with Local Factors - Determining the Number of Relevant Factors," *Journal of Econometrics*, forthcoming.

[21] Freyaldenhoven, S. (2021b): "Identification through Sparsity in Factor Models: The ℓ_1-Rotation Criterion," Working Paper, Federal Reserve Bank of Philadelphia.

[22] Giglio, S., D. Xiu, and D. Zhang (2021): "Test Assets and Weak Factors," Working Paper, Yale School of Management and the Booth School of Business, University of Chicago.

[23] Goroketskii, V. V. (1977): "On the Strong Mixing Property for Linear Sequences," *Theory of Probability and Applications*, 22, 411-413.

[24] Gospodinov, N., R. Kan, and C. Robotti (2017): "Spurious Inference in Reduced-Rank Asset Pricing Models," *Econometrica*, 85, 1613-1628.

[25] Harding, M. C. (2008): "Explaining the Single Factor Bias of Arbitrage Pricing Models in Finite Samples," *Economics Letters*, 99, 85-88.

[26] Jagannathan, R. and Z. Wang (1998): "An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-Sectional Regression," *Journal of Finance*, 53, 1285-1309.

[27] Johnstone, I. M. and A. Lu (2009): "On Consistency and Sparsity for Principal Components Analysis in High Dimensions," *Journal of the American Statistical Association*, 104, 682-697.

[28] Johnstone, I. M. and D. Paul (2018): "PCA in High Dimensions: An Orientation," *Proceedings of the IEEE*, 106, 1277-1292.

[29] Kan, R. and C. Zhang (1999): "Two-Pass Tests of Asset Pricing Models with Useless Factors," *Journal of Finance*, 54, 203-235.

[30] Kleibergen, F. (2009): "Tests of Risk Premia in Linear Factor Models," *Journal of Econometrics*, 149, 149-173.

[31] Lütkepohl, H. (2005): *New Introduction to Multiple Time Series Analysis*. New York: Springer.

[32] Onatski, A. (2012): "Asymptotics of the Principal Components Estimator of Large Factor Models with Weakly Influential Factors," *Journal of Econometrics*, 168, 244-258.

[33] Paul, D. (2007): "Asymptotics of Sample Eigenstructure for a Large Dimensional Spiked Covariance Model," *Statistica Sinica*, 17, 1617-1642.

[34] Pham, T. D. and L. T. Tran (1985): "Some Mixing Properties of Time Series Models," *Stochastic Processes and Their Applications*, 19, 297-303.

[35] Ruhe, A. (1975): "On the Closeness of Eigenvalues and Singular Values for Almost Normal Matrices," *Linear Algebra and Its Applications*, 11, 87-94.

[36] Shen, D., H. Shen, H. Zhu, J.S. Marron (2016): "The Statistics and Mathematics of High Dimension Low Sample Size Asymptotics," *Statistica Sinica*, 26, 1747-1770.

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

[38] Stock, J. H. and M. W. Watson (2002b): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20, 147-162.

5 Appendix: Proofs of Theorems

This appendix contains the proofs of Theorems 1 and 2, as well as that of three supporting lemmas: Lemma A1, Lemma A2, and Lemma A3. The proofs of Theorems 1 and 2 are given first, followed by the statements and proofs of Lemmas A1-A3.

Proof of Theorem 1: To show part (a), first set $z = \Phi^{-1}\left(1 - \frac{\varphi}{2N}\right)$, where $N = N_1 + N_2$. Note that, under Assumption 2-10, we can easily show that
$\Phi^{-1}\left(1 - \frac{\varphi}{2N}\right) \leq \sqrt{2\left(1+a\right)}\sqrt{\ln N}$, for all N_1, N_2 sufficiently large.[11] By part (a) of Assumption 2-9, $\sqrt{\ln N}/\min\left\{T^{(1-\alpha_1)/6}, T^{\alpha_2/2}\right\} \to 0$ as $N_1, N_2, T \to \infty$; this, in

[11] An explicit proof of this result is given in the Technical Appendix of an earlier version of this paper, Chao and Swanson (2022b). In particular, this inequality is shown in part (b) of Lemma C-16 in Appendix C of Chao and Swanson (2022b).

turn, implies that, for some positive constant c_0, $\Phi^{-1}\left(1-\frac{\varphi}{2N}\right)$ satisfies the inequality constraint $0 \leq \Phi^{-1}\left(1-\frac{\varphi}{2N}\right) \leq c_0 \min\left\{T^{(1-\alpha_1)/6}, T^{\alpha_2/2}\right\}$ for all N_1, N_2, T sufficiently large, so that $\Phi^{-1}\left(1-\frac{\varphi}{2N}\right)$ lies within the range of values of z for which the moderate deviation inequality given in Lemma A3 holds. Thus, plugging $\Phi^{-1}\left(1-\frac{\varphi}{2N}\right)$ into the moderate deviation inequality (23) given in Lemma A3 below, we see that there exists a positive constant A such that:

$$
\begin{aligned}
&P\left(|S_{i,\ell,T}| \geq \Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right) \\
&\leq 2\left[1-\Phi\left(\Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right)\right]\left\{1+A\left[1+\Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right]^3 T^{-\frac{1-\alpha_1}{2}}\right\} \\
&= 2\left[1-\left(1-\frac{\varphi}{2N}\right)\right]\left\{1+A\left[1+\Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right]^3 T^{-\frac{1-\alpha_1}{2}}\right\} \\
&= \frac{\varphi}{N}\left\{1+A\left[1+\Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right]^3 T^{-\frac{1-\alpha_1}{2}}\right\},
\end{aligned}
$$

for $\ell \in \{1, ..., d\}$, for $i \in H = \{k \in \{1, ..., N\} : \gamma_k = 0\}$, and for all N_1, N_2, T sufficiently large. Next, note that:

$$
\begin{aligned}
&P\left(\max_{i\in H}\sum_{\ell=1}^{d} \varpi_\ell |S_{i,\ell,T}| \geq \Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right) \\
&\leq P\left(\bigcup_{i\in H}\bigcup_{1\leq\ell\leq d}\left\{|S_{i,\ell,T}| \geq \Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right\}\right) \quad \left(\text{since } 0 \leq \varpi_\ell \leq 1 \text{ and } \sum_{\ell=1}^{d}\varpi_\ell = 1\right) \\
&\leq \sum_{i\in H}\sum_{\ell=1}^{d} P\left(|S_{i,\ell,T}| \geq \Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right) \quad \text{(by union bound)} \\
&\leq \sum_{i\in H}\sum_{\ell=1}^{d} \frac{\varphi}{N}\left\{1+A\left[1+\Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right]^3 T^{-(1-\alpha_1)\frac{1}{2}}\right\} \\
&= d\frac{N_2\varphi}{N}\left\{1+A\left[1+\Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right]^3 T^{-(1-\alpha_1)\frac{1}{2}}\right\}
\end{aligned}
$$

Using the inequality $\Phi^{-1}\left(1-\frac{\varphi}{2N}\right) \leq \sqrt{2(1+a)}\sqrt{\ln N}$ discussed above, we further obtain, for all N_1, N_2, T sufficiently large:

$$
\begin{aligned}
&P\left(\max_{i\in H}\sum_{\ell=1}^{d} \varpi_\ell |S_{i,\ell,T}| \geq \Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right) \leq \frac{dN_2\varphi}{N}\left\{1+\frac{A}{T^{(1-\alpha_1)/2}}\left[1+\Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right]^3\right\} \\
&\leq \frac{dN_2\varphi}{N}\left\{1+2^2 A T^{-\frac{(1-\alpha_1)}{2}} + 2^2 A\left[\Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right]^3 T^{-\frac{(1-\alpha_1)}{2}}\right\} \\
&\left(\text{by the inequality } \left|\sum_{i=1}^{m} a_i\right|^r \leq c_r \sum_{i=1}^{m}|a_i|^r \text{ where } c_r = m^{r-1} \text{ for } r \geq 1\right) \\
&\leq \frac{dN_2\varphi}{N}\left\{1+4AT^{-\frac{(1-\alpha_1)}{2}} + 4A\left[\sqrt{2(1+a)}\sqrt{\ln N}\right]^3 T^{-\frac{(1-\alpha_1)}{2}}\right\} \\
&= \frac{dN_2\varphi}{N}\left\{1+4AT^{-\frac{(1-\alpha_1)}{2}} + 2^{\frac{7}{2}}A(1+a)^{\frac{3}{2}}\frac{(\ln N)^{\frac{3}{2}}}{T^{\frac{1-\alpha_1}{2}}}\right\}.
\end{aligned}
$$

Finally, note that rate condition given in part (a) of Assumption 2-9 (i.e., $\sqrt{\ln N}/\min\left\{T^{(1-\alpha_1)/6}, T^{\alpha_2/2}\right\} \to 0$ as $N_1, N_2, T \to \infty$) implies that $(\ln N)^{\frac{3}{2}}/T^{\frac{1-\alpha_1}{2}} \to 0$ as $N_1, N_2, T \to \infty$, from which it follows that:

$$P\left(\max_{i\in H}\sum_{\ell=1}^{d}\varpi_\ell\left|S_{i,\ell,T}\right| \geq \Phi^{-1}\left(1-\tfrac{\varphi}{2N}\right)\right)$$

$$\leq \tfrac{dN_2\varphi}{N}\left\{1+4AT^{-\frac{(1-\alpha_1)}{2}}+2^{\frac{7}{2}}A\left(1+a\right)^{\frac{3}{2}}\frac{(\ln N)^{\frac{3}{2}}}{T^{\frac{1-\alpha_1}{2}}}\right\} = \tfrac{dN_2\varphi}{N}\left[1+o\left(1\right)\right] = O\left(\tfrac{N_2\varphi}{N}\right) = o\left(1\right).$$

Next, to show part (b), note that, by a similar argument as that given for part (a) above, we have:

$$P\left(\max_{i\in H}\max_{1\leq\ell\leq d}\left|S_{i,\ell,T}\right| \geq \Phi^{-1}\left(1-\tfrac{\varphi}{2N}\right)\right)$$

$$= P\left(\bigcup_{i\in H}\bigcup_{1\leq\ell\leq d}\left\{\left|S_{i,\ell,T}\right| \geq \Phi^{-1}\left(1-\tfrac{\varphi}{2N}\right)\right\}\right)$$

$$\leq \tfrac{dN_2\varphi}{N}\left\{1+\tfrac{4A}{T^{(1-\alpha_1)/2}}+\tfrac{2^{\frac{7}{2}}A(1+a)^{\frac{3}{2}}(\ln N)^{\frac{3}{2}}}{T^{(1-\alpha_1)/2}}\right\} = \tfrac{dN_2\varphi}{N}\left[1+o\left(1\right)\right] = O\left(\tfrac{N_2\varphi}{N}\right) = o\left(1\right). \;\square$$

Proof of Theorem 2: To show part (a), note that:

$$P\left(\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell\left|S_{i,\ell,T}\right| \geq \Phi^{-1}\left(1-\tfrac{\varphi}{2N}\right)\right)$$

$$= P\left(\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell\left|\tfrac{\overline{S}_{i,,\ell,T}-\mu_{i,\ell,T}}{\sqrt{\overline{V}_{i,\ell,T}}}+\tfrac{\mu_{i,,\ell,T}}{\sqrt{\overline{V}_{i,\ell,T}}}\right| \geq \Phi^{-1}\left(1-\tfrac{\varphi}{2N}\right)\right)$$

$$\geq P\left(\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell\left\{\left|\tfrac{\mu_{i,\ell,T}}{\sqrt{\overline{V}_{i,\ell,T}}}\right|-\left|\tfrac{\overline{S}_{i,\ell,T}-\mu_{i,\ell,T}}{\sqrt{\overline{V}_{i,\ell,T}}}\right|\right\} \geq \Phi^{-1}\left(1-\tfrac{\varphi}{2N}\right)\right)$$

$$= P\left(\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell\left\{\left|\tfrac{\mu_{i,\ell,T}}{\sqrt{\overline{V}_{i,\ell,T}}}\right|\left[1-\left|\tfrac{\sqrt{\overline{V}_{i,\ell,T}}}{\mu_{i,\ell,T}}\right|\left|\tfrac{\overline{S}_{i,\ell,T}-\mu_{i,\ell,T}}{\sqrt{\overline{V}_{i,\ell,T}}}\right|\right]\right\} \geq \Phi^{-1}\left(1-\tfrac{\varphi}{2N}\right)\right)$$

$$= P\left(\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell\left\{\left|\tfrac{\mu_{i,\ell,T}}{\sqrt{\overline{V}_{i,\ell,T}}}\right|\left[1-\left|\tfrac{\overline{S}_{i,\ell,T}-\mu_{i,\ell,T}}{\mu_{i,\ell,T}}\right|\right]\right\} \geq \Phi^{-1}\left(1-\tfrac{\varphi}{2N}\right)\right),$$

where $\mu_{i,\ell,T}=\sum_{r=1}^{q}\sum_{t=b_1(r)}^{b_2(r)}\gamma_i'\left\{E\left[\underline{F}_t\right]\mu_{Y,\ell}+E\left[\underline{F}_t\underline{Y}_t'\right]\alpha_{YY,\ell}+E\left[\underline{F}_t\underline{F}_t'\right]\alpha_{YF,\ell}\right\}$,for $b_1\left(r\right)=\left(r-1\right)\tau+p$ and $b_2\left(r\right)=b_1\left(r\right)+\tau_1-1$. Next, let $\pi_{i,\ell,T}=\sum_{r=1}^{q}\left(\sum_{t=b_1(r)}^{b_2(r)}\left\{\gamma_i'E\left[\underline{F}_t\right]\mu_{Y,\ell}+\gamma_i'E\left[\underline{F}_t\underline{Y}_t'\right]\alpha_{YY,\ell}+\gamma_i'E\left[\underline{F}_t\underline{F}_t'\right]\alpha_{YF,\ell}\right\}\right)^2$, and we

see that, under Assumption 2-8, there exists a positive constant $\underline{c}$ such that for every $\ell\in\{1,...,d\}$ and for all N_1, N_2, and T sufficiently large:

$$\min_{i\in H^c}\left\{\pi_{i,\ell,T}/\left(q\tau_1^2\right)\right\}$$

$$= \min_{i\in H^c}\tfrac{1}{q}\sum_{r=1}^{q}\left(\tfrac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\left\{\gamma_i'E\left[\underline{F}_t\right]\mu_{Y,\ell}+\gamma_i'E\left[\underline{F}_t\underline{Y}_t'\right]\alpha_{YY,\ell}+\gamma_i'E\left[\underline{F}_t\underline{F}_t'\right]\alpha_{YF,\ell}\right\}\right)^2$$

$$= \min_{i\in H^c} \frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} E\left[\gamma_i' \underline{F}_t y_{\ell,t+1}\right]\right)^2$$
$$\geq \min_{i\in H^c} \left(\frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} E\left[\gamma_i' \underline{F}_t y_{\ell,t+1}\right]\right)^2 \quad \text{(by Jensen's inequality)}$$
$$= \min_{i\in H^c} \left|\frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} \left\{\gamma_i' E\left[\underline{F}_t\right]\mu_{Y,\ell} + \gamma_i' E\left[\underline{F}_t \underline{Y}_t'\right]\alpha_{YY,\ell} + \gamma_i' E\left[\underline{F}_t \underline{F}_t'\right]\alpha_{YF,\ell}\right\}\right|^2$$
$$\geq \underline{c}^2 > 0 \quad \text{(in light of Assumption 2-8)}.$$

It follows that for all N_1, N_2, and T sufficiently large:

$$P\left(\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell \left|S_{i,\ell,T}\right| \geq \Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right)$$
$$\geq P\left(\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell \left\{\left|\frac{\mu_{i,\ell,T}}{\sqrt{\overline{V}_{i,\ell,T}}}\right|\left[1-\left|\frac{\overline{S}_{i,\ell,T}-\mu_{i,\ell,T}}{\mu_{i,\ell,T}}\right|\right]\right\} \geq \Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right)$$
$$= P\left(\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell \left\{\left|\frac{\sqrt{q}\left[\mu_{i,\ell,T}/(q\tau_1)\right]}{\sqrt{\pi_{i,\ell,T}/\left(q\tau_1^2\right)}}\right|\left|\frac{\sqrt{\pi_{i,\ell,T}/\left(q\tau_1^2\right)}}{\sqrt{\pi_{i,\ell,T}/\left(q\tau_1^2\right)}+\sqrt{\overline{V}_{i,\ell,T}/\left(q\tau_1^2\right)}-\sqrt{\pi_{i,\ell,T}/\left(q\tau_1^2\right)}}\right|\right.\right.$$
$$\left.\left.\times\left[1-\left|\frac{\overline{S}_{i,\ell,T}-\mu_{i,\ell,T}}{\mu_{i,\ell,T}}\right|\right]\right\} \geq \Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right)$$
$$= P\left(\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell \left\{\left|\frac{\sqrt{q}\left[\mu_{i,\ell,T}/(q\tau_1)\right]}{\sqrt{\pi_{i,\ell,T}/\left(q\tau_1^2\right)}}\right|\left|\frac{1}{1+\left(\sqrt{\overline{V}_{i,\ell,T}}-\sqrt{\pi_{i,\ell,T}}\right)/\sqrt{\pi_{i,\ell,T}}}\right|\right.\right.$$
$$\left.\left.\times\left[1-\left|\frac{\overline{S}_{i,\ell,T}-\mu_{i,\ell,T}}{\mu_{i,\ell,T}}\right|\right]\right\} \geq \Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right)$$
$$\geq P\left(\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell \left\{\left|\frac{\sqrt{q}\left(\mu_{i,\ell,T}/(q\tau_1)\right)}{\sqrt{\pi_{i,\ell,T}/\left(q\tau_1^2\right)}}\right|\frac{1}{1+\max_{k\in H^c}\left|\sqrt{\overline{V}_{k,\ell,T}}-\sqrt{\pi_{k,\ell,T}}\right|/\sqrt{\pi_{k,\ell,T}}}\right.\right.$$
$$\left.\left.\times\left[1-\max_{k\in H^c}\left|\frac{\overline{S}_{k,\ell,T}-\mu_{k,\ell,T}}{\mu_{k,\ell,T}}\right|\right]\right\} \geq \Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right)$$
$$\geq P\left(\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell \left\{\left|\frac{\sqrt{q}\left[\mu_{i,\ell,T}/(q\tau_1)\right]}{\sqrt{\pi_{i,\ell,T}/\left(q\tau_1^2\right)}}\right|\frac{1}{1+\max_{k\in H^c}\sqrt{\left|\overline{V}_{k,\ell,T}-\pi_{k,\ell,T}\right|/\pi_{k,\ell,T}}}\right.\right.$$
$$\left.\left.\times\left[1-\max_{k\in H^c}\left|\frac{\overline{S}_{k,\ell,T}-\mu_{k,\ell,T}}{\mu_{k,\ell,T}}\right|\right]\right\} \geq \Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right)$$
$$\left(\text{making use of the inequality } \left|\sqrt{x}-\sqrt{y}\right| \leq \sqrt{|x-y|} \text{ for } x\geq 0 \text{ and } y \geq 0\right)$$
$$= P\left(\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell \left\{\left|\frac{\sqrt{q}\left[\mu_{i,\ell,T}/(q\tau_1)\right]}{\sqrt{\pi_{i,\ell,T}/\left(q\tau_1^2\right)}}\right|\frac{1-\max_{k\in H^c}\left|\mathcal{E}_{k,\ell,T}\right|}{1+\max_{k\in H^c}\sqrt{\left|\mathcal{V}_{k,\ell,T}\right|}}\right\} \geq \Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right),$$

where $\mathcal{E}_{k,\ell,T} = \left(\overline{S}_{k,\ell,T}-\mu_{k,\ell,T}\right)/\mu_{k,\ell,T}$ and $\mathcal{V}_{k,\ell,T} = \left(\overline{V}_{k,\ell,T}-\pi_{k,\ell,T}\right)/\pi_{k,\ell,T}$. By the result of part (a) of Lemma A2 (given below), there exists a sequence of positive numbers $\{\epsilon_T\}$ such that, as $T\to\infty$, $\epsilon_T \to 0$ and $P\left(\max_{1\leq\ell\leq d}\max_{k\in H^c}\left|\mathcal{E}_{k,\ell,T}\right| \geq \epsilon_T\right) \to 0$. In addition, by the result of part (b) of Lemma A2, there exists a sequence of positive numbers $\{\epsilon_T^*\}$ such that, as $T \to \infty$, $\epsilon_T^* \to 0$ and $P\left(\max_{1\leq\ell\leq d}\max_{k\in H^c}|\mathcal{V}_{k,\ell,T}| \geq \epsilon_T^*\right) \to 0$. Further define $\overline{\mathbb{E}}_T = \max_{1\leq\ell\leq d}\max_{k\in H^c}|\mathcal{E}_{k,\ell,T}|$ and $\overline{\mathbb{V}}_T = \max_{1\leq\ell\leq d}\max_{k\in H^c}|\mathcal{V}_{k,\ell,T}|$; and note that, for all N_1, N_2, and T sufficiently large,

$$
\begin{aligned}
&P\left(\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell\left|S_{i,\ell,T}\right| \geq \Phi^{-1}\left(1-\tfrac{\varphi}{2N}\right)\right)\\
&\geq P\left(\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell\left\{\left|\frac{\sqrt{q}\left[\mu_{i,\ell,T}/(q\tau_1)\right]}{\sqrt{\pi_{i,\ell,T}/\left(q\tau_1^2\right)}}\right|\frac{1-\max_{k\in H^c}\left|\mathcal{E}_{k,\ell,T}\right|}{1+\max_{k\in H^c}\sqrt{\left|\mathcal{V}_{k,\ell,T}\right|}}\right\} \geq \Phi^{-1}\left(1-\tfrac{\varphi}{2N}\right)\right)\\
&\geq P\left(\frac{1-\max_{1\leq\ell\leq d}\max_{k\in H^c}\left|\mathcal{E}_{k,\ell,T}\right|}{1+\max_{1\leq\ell\leq d}\max_{k\in H^c}\sqrt{\left|\mathcal{V}_{k,\ell,T}\right|}}\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell\left|\frac{\sqrt{q}\left[\mu_{i,\ell,T}/(q\tau_1)\right]}{\sqrt{\pi_{i,\ell,T}/\left(q\tau_1^2\right)}}\right| \geq \Phi^{-1}\left(1-\tfrac{\varphi}{2N}\right)\right)\\
&= P\left(\frac{1-\overline{\mathbb{E}}_T}{1+\sqrt{\overline{\mathbb{V}}_T}}\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell\left|\frac{\mu_{i,\ell,T}}{\sqrt{\pi_{i,\ell,T}}}\right| \geq \Phi^{-1}\left(1-\tfrac{\varphi}{2N}\right)\right)\\
&\geq P\left(\left\{\left|\frac{1-\epsilon_T}{1+\sqrt{\epsilon_T^*}}\right|\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell\left|\frac{\mu_{i,\ell,T}}{\sqrt{\pi_{i,\ell,T}}}\right| \geq \Phi^{-1}\left(1-\tfrac{\varphi}{2N}\right)\right\}\cap\left\{\overline{\mathbb{E}}_T<\epsilon_T\right\}\cap\left\{\overline{\mathbb{V}}_T<\epsilon_T^*\right\}\right)\\
&+P\left(\left\{\frac{1-\overline{\mathbb{E}}_T}{1+\sqrt{\overline{\mathbb{V}}_T}}\min_{i\in H}\sum_{\ell=1}^{d}\varpi_\ell\left|\frac{\mu_{i,\ell,T}}{\sqrt{\pi_{i,\ell,T}}}\right| \geq \Phi^{-1}\left(1-\tfrac{\varphi}{2N}\right)\right\}\cap\left\{\overline{\mathbb{E}}_T\geq\epsilon_T\cup\overline{\mathbb{V}}_T\geq\epsilon_T^*\right\}\right)\\
&\geq P\left(\left\{\left|\frac{1-\epsilon_T}{1+\sqrt{\epsilon_T^*}}\right|\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell\left|\frac{\mu_{i,\ell,T}}{\sqrt{\pi_{i,\ell,T}}}\right| \geq \Phi^{-1}\left(1-\tfrac{\varphi}{2N}\right)\right\}\cap\left\{\overline{\mathbb{E}}_T<\epsilon_T\right\}\cap\left\{\overline{\mathbb{V}}_T<\epsilon_T^*\right\}\right)\\
&+P\left(\left\{\frac{1-\overline{\mathbb{E}}_T}{1+\sqrt{\overline{\mathbb{V}}_T}}\min_{i\in H}\sum_{\ell=1}^{d}\varpi_\ell\left|\frac{\mu_{i,\ell,T}}{\sqrt{\pi_{i,\ell,T}}}\right| \geq \Phi^{-1}\left(1-\tfrac{\varphi}{2N}\right)\right\}\cap\left\{\overline{\mathbb{E}}_T\geq\epsilon_T\right\}\right)\\
&= P\left(\left\{\left|\frac{1-\epsilon_T}{1+\sqrt{\epsilon_T^*}}\right|\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell\left|\frac{\mu_{i,\ell,T}}{\sqrt{\pi_{i,\ell,T}}}\right| \geq \Phi^{-1}\left(1-\tfrac{\varphi}{2N}\right)\right\}\cap\left\{\overline{\mathbb{E}}_T<\epsilon_T\right\}\cap\left\{\overline{\mathbb{V}}_T<\epsilon_T^*\right\}\right)\\
&\quad+o(1).
\end{aligned}
$$

where the last equality above follows from the fact that

$$
\begin{aligned}
&P\left(\left\{\frac{1-\overline{\mathbb{E}}_T}{1+\sqrt{\overline{\mathbb{V}}_T}}\min_{i\in H}\sum_{\ell=1}^{d}\varpi_\ell\left|\frac{\mu_{i,\ell,T}}{\sqrt{\pi_{i,\ell,T}}}\right| \geq \Phi^{-1}\left(1-\tfrac{\varphi}{2N}\right)\right\}\cap\left\{\overline{\mathbb{E}}_T\geq\epsilon_T\right\}\right)\\
&\leq P\left(\overline{\mathbb{E}}_T\geq\epsilon_T\right)=o(1)
\end{aligned}
$$

Moreover, making use of Assumption 2-8, the result given in Lemma A1, and the fact that $q = \lfloor T_0/\tau\rfloor \sim T^{1-\alpha_1}$, we see that, there exists positive constants $\underline{c}$ and $\overline{C}$ such that:

$$
\min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell\left|\frac{\mu_{i,\ell,T}}{\sqrt{\pi_{i,\ell,T}}}\right| = \min_{i\in H^c}\sum_{\ell=1}^{d}\varpi_\ell\frac{\sqrt{q}\left|\mu_{i,\ell,T}/(q\tau_1)\right|}{\sqrt{\pi_{i,\ell,T}/\left(q\tau_1^2\right)}}
$$

$$\geq \sqrt{q} \sum_{\ell=1}^{d} \varpi_\ell \frac{\min_{i \in H^c} \left| \mu_{i,\ell,T} / (q\tau_1) \right|}{\sqrt{\max_{i \in H^c} \pi_{i,\ell,T} / \left(q\tau_1^2\right)}} \geq \sqrt{q} \sum_{\ell=1}^{d} \varpi_\ell \frac{c}{\sqrt{C}} = \sqrt{q} \frac{c}{\sqrt{C}} \sim \sqrt{q} \sim \sqrt{\frac{T_0}{\tau}} \sim T^{(1-\alpha_1)/2} \ .$$

On the other hand, applying the inequality $\Phi^{-1}\left(1 - \frac{\varphi}{2N}\right) \leq \sqrt{2\left(1+a\right)}\sqrt{\ln N} \sim \sqrt{\ln N}$[12], we further deduce that, as $N_1, N_2, T \to \infty$,

$$\frac{1}{\Phi^{-1}\left(1-\frac{\varphi}{2N}\right)} \min_{i \in H^c} \sum_{\ell=1}^{d} \varpi_\ell \left| \frac{\mu_{i,\ell,T}}{\sqrt{\pi_{i,\ell,T}}} \right| \geq \frac{c}{\sqrt{C}} \sqrt{\frac{q}{2\left(1+a\right)\ln N}} \sim \sqrt{\frac{T^{(1-\alpha_1)}}{\ln N}} \to \infty.$$

This is true because the condition $\sqrt{\ln N} / \min\left\{T^{(1-\alpha_1)/6}, T^{\alpha_2/2}\right\} \to 0$ as $N_1, N_2, T \to \infty$ (given in Assumption 2-9 part (a)) implies that $\ln N / T^{(1-\alpha_1)} \to 0$ as $N_1, N_2, T \to \infty$. Hence, there exists a natural number M such that, for all $N_1 \geq M, N_2 \geq M$, and $T \geq M$, we have $\left|\frac{1-\epsilon_T}{1+\sqrt{\epsilon_T^*}}\right| \min_{i \in H^c} \sum_{\ell=1}^{d} \varpi_\ell \left|\frac{\mu_{i,\ell,T}}{\sqrt{\pi_{i,\ell,T}}}\right| \geq \Phi^{-1}\left(1 - \frac{\varphi}{2N}\right)$ so that:

$$\begin{aligned}
&P\left(\min_{i \in H^c} \sum_{\ell=1}^{d} \varpi_\ell \left|S_{i,\ell,T}\right| \geq \Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right) \\
&\geq P\left(\left\{\left|\frac{1-\epsilon_T}{1+\sqrt{\epsilon_T^*}}\right| \min_{i \in H^c} \sum_{\ell=1}^{d} \varpi_\ell \left|\frac{\mu_{i,\ell,T}}{\sqrt{\pi_{i,\ell,T}}}\right| \geq \Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right\} \cap \left\{\overline{\mathbb{E}}_T < \epsilon_T\right\} \cap \left\{\overline{\mathbb{V}}_T < \epsilon_T^*\right\}\right) \\
&\quad + o\left(1\right) \\
&= P\left(\left\{\overline{\mathbb{E}}_T < \epsilon_T\right\} \cap \left\{\overline{\mathbb{V}}_T < \epsilon_T^*\right\}\right) + o\left(1\right) \\
&\quad \left(\text{for all } N_1 \geq M, N_2 \geq M, \text{and } T \geq M\right) \\
&\geq P\left(\overline{\mathbb{E}}_T < \epsilon_T\right) + P\left(\overline{\mathbb{V}}_T < \epsilon_T^*\right) - 1 + o\left(1\right) \\
&\quad \left(\text{using the inequality } P\left\{\bigcap_{i=1}^{m} A_i\right\} \geq \sum_{i=1}^{m} P\left(A_i\right) - \left(m-1\right)\right) \\
&= 1 - P\left(\overline{\mathbb{E}}_T \geq \epsilon_T\right) + 1 - P\left(\overline{\mathbb{V}}_T \geq \epsilon_T^*\right) - 1 + o\left(1\right) \\
&= 1 - P\left(\overline{\mathbb{E}}_T \geq \epsilon_T\right) - P\left(\overline{\mathbb{V}}_T \geq \epsilon_T^*\right) + o\left(1\right) \\
&= 1 + o\left(1\right).
\end{aligned}$$

Next, to show part (b), note that, by applying the result in part (a), we have that:

$$\begin{aligned}
&P\left(\min_{i \in H^c} \max_{1 \leq \ell \leq d} \left|S_{i,\ell,T}\right| \geq \Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right) \\
&\geq P\left(\min_{i \in H^c} \sum_{\ell=1}^{d} \varpi_\ell \left|S_{i,\ell,T}\right| \geq \Phi^{-1}\left(1-\frac{\varphi}{2N}\right)\right) = 1 + o\left(1\right). \ \square
\end{aligned}$$

[12] As noted previously, an explicit proof of this result is given in the Technical Appendix of an earlier version of this paper, Chao and Swanson (2022b). In particular, this inequality is shown in part (b) of Lemma C-16 in Appendix C of Chao and Swanson (2022b).

Lemma A1: Let $\underline{Y}_t$ and $\underline{F}_t$ be defined in expression (4), and define $b_1(r) = (r-1)\tau + p$ and $b_2(r) = b_1(r) + \tau_1 - 1$. Under Assumptions 2-1, 2-2, 2-5, 2-6, and 2-9(b); there exists a positive constant C such that:

$$\max_{1\leq \ell \leq d,\ i\in H^c}\left(\frac{\pi_{i,\ell,T}}{q\tau_1^2}\right)$$

$$= \max_{1\leq \ell \leq d,\ i\in H^c}\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\gamma_i'\left\{E\left[\underline{F}_t\right]\mu_{Y,\ell} + E\left[\underline{F}_t\underline{Y}_t'\right]\alpha_{YY,\ell} + E\left[\underline{F}_t\underline{F}_t'\right]\alpha_{YF,\ell}\right\}\right)^2$$

$\leq C < \infty$, for all N_1, N_2, T sufficiently large.

Proof of Lemma A1: To proceed, let $\phi_{\max} = \max\left\{\left|\lambda_{\max}(A)\right|, \left|\lambda_{\min}(A)\right|\right\}$ and, for $\ell \in \{1,...,d\}$, let $e_{\ell,d}$ denote a $d\times 1$ elementary vector whose ℓ^{th} component is 1 and all other components are 0. Now, note that:

$$\max_{1\leq \ell \leq d,\ i\in H^c}\left\{\pi_{i,\ell,T}/\left(q\tau_1^2\right)\right\}$$

$$= \max_{1\leq \ell \leq d,\ i\in H^c}\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\gamma_i'\left\{E\left[\underline{F}_t\right]\mu_{Y,\ell} + E\left[\underline{F}_t\underline{Y}_t'\right]\alpha_{YY,\ell} + E\left[\underline{F}_t\underline{F}_t'\right]\alpha_{YF,\ell}\right\}\right)^2$$

$$\leq \max_{1\leq \ell \leq d,\ i\in H^c}\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\left\{E\left[\left|\gamma_i'\underline{F}_t\right|\right]\left|\mu_{Y,\ell}\right| + E\left[\left|\gamma_i'\underline{F}_t\underline{Y}_t'A_{YY}'e_{\ell,d}\right|\right]\right.\right.$$
$$\left.\left. + E\left[\left|\gamma_i'\underline{F}_t\underline{F}_t'A_{YF}'e_{\ell,d}\right|\right]\right\}\right)^2$$

(by triangle and Jensen's inequalities)

$$\leq \max_{i\in H^c}\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\left\{\sqrt{\left\|\gamma_i\right\|_2^2}\sqrt{E\left\|\underline{F}_t\right\|_2^2}\max_{1\leq \ell\leq d}\left|\mu_{Y,\ell}\right|\right.\right.$$
$$+\sqrt{\gamma_i'E\left[\underline{F}_t\underline{F}_t'\right]\gamma_i}\sqrt{\max_{1\leq \ell\leq d}e_{\ell,d}'A_{YY}E\left[\underline{Y}_t\underline{Y}_t'\right]A_{YY}'e_{\ell,d}}$$
$$\left.\left.+\sqrt{\gamma_i'E\left[\underline{F}_t\underline{F}_t'\right]\gamma_i}\sqrt{\max_{1\leq \ell\leq d}e_{\ell,d}'A_{YF}E\left[\underline{F}_t\underline{F}_t'\right]A_{YF}'e_{\ell,d}}\right\}\right)^2$$

$$\leq \left(\max_{i\in H^c}\left\|\gamma_i\right\|_2^2\right)\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\left\{\sqrt{E\left\|\underline{F}_t\right\|_2^2}\max_{1\leq \ell\leq d}\left|\mu_{Y,\ell}\right|\right.\right.$$
$$+\sqrt{E\left\|\underline{F}_t\right\|_2^2}\sqrt{E\left\|\underline{Y}_t\right\|_2^2}\sqrt{\max_{1\leq \ell\leq d}e_{\ell,d}'A_{YY}A_{YY}'e_{\ell,d}}$$
$$\left.\left.+E\left\|\underline{F}_t\right\|_2^2\sqrt{\max_{1\leq \ell\leq d}e_{\ell,d}'A_{YF}A_{YF}'e_{\ell,d}}\right\}\right)^2$$

$$\leq \left(\max_{i\in H^c}\left\|\gamma_i\right\|_2^2\right)\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\left\{\sqrt{E\left\|\underline{F}_t\right\|_2^2}\max_{1\leq \ell\leq d}\left|\mu_{Y,\ell}\right|\right.\right.$$
$$\left.\left.+\sqrt{E\left\|\underline{F}_t\right\|_2^2}\sqrt{E\left\|\underline{Y}_t\right\|_2^2}C^{\dagger}\phi_{\max} + E\left\|\underline{F}_t\right\|_2^2C^{\dagger}\phi_{\max}\right\}\right)^2,$$

where the last inequality follows from the fact that, by making use of Assumption 2-6, it is easy to show that there exists a constant $C^{\dagger} > 0$ such that $\sqrt{\max_{1\leq\ell\leq d} e_{\ell,d}' A_{YY} A_{YY}' e_{\ell,d}} \leq \|A_{YY}\|_2 \sqrt{\max_{1\leq\ell\leq d} e_{\ell,d}' e_{\ell,d}} = \|A_{YY}\|_2 \leq C^{\dagger}\phi_{\max}$ and, similarly, $\sqrt{\max_{1\leq\ell\leq d} e_{\ell,d}' A_{YF} A_{YF}' e_{\ell,d}} \leq \|A_{YF}\|_2 \leq C^{\dagger}\phi_{\max}$.[13] Hence,

$$
\begin{aligned}
&\max_{1\leq\ell\leq d} \max_{k\in H^c} \left\{\pi_{i,\ell,T}/\left(q\tau_1^2\right)\right\} \\
&\leq \left(\max_{i\in H^c} \|\gamma_i\|_2^2\right) \frac{1}{q} \sum_{r=1}^{q} \left(\frac{1}{\tau_1} \sum_{t=b_1(r)}^{b_2(r)} \left\{\sqrt{E\left\|\underline{F}_t\right\|_2^2} \max_{1\leq\ell\leq d} \left|\mu_{Y,\ell}\right|\right.\right. \\
&\qquad \left.\left. + \sqrt{E\left\|\underline{F}_t\right\|_2^2} \sqrt{E\left\|\underline{Y}_t\right\|_2^2} C^{\dagger}\phi_{\max} + E\left\|\underline{F}_t\right\|_2^2 C^{\dagger}\phi_{\max}\right\}\right)^2 \\
&\leq \left(\max_{i\in H^c} \|\gamma_i\|_2^2\right) \frac{1}{q} \sum_{r=1}^{q} \frac{1}{\tau_1} \sum_{t=b_1(r)}^{b_2(r)} E\left\|\underline{F}_t\right\|_2^2 \left(\|\mu_Y\|_2^2 + \left[\sqrt{E\left\|\underline{Y}_t\right\|_2^2} + \sqrt{E\left\|\underline{F}_t\right\|_2^2}\right] C^{\dagger}\phi_{\max}\right)^2 \\
&\leq C < \infty,
\end{aligned}
$$

for some positive constant C such that $C \geq \left(\max_{i\in H^c} \|\gamma_i\|_2^2\right) E\left\|\underline{F}_t\right\|_2^2 \left(\|\mu_Y\|_2^2 + \left[\sqrt{E\left\|\underline{Y}_t\right\|_2^2} + \sqrt{E\left\|\underline{F}_t\right\|_2^2}\right] C^{\dagger}\phi_{\max}\right)^2$, where such a constant exists because $\max_{i\in H^c} \|\gamma_i\|_2^2$ and $\|\mu_Y\|_2^2$ are both bounded given Assumption 2-5; because $0 < \phi_{\max} < 1$ given Assumption 2-1; and because, under Assumptions 2-1, 2-2(a)-(b), 2-5, and 2-6; one can easily show that there exists a positive constant C^* such that $E\left\|\underline{F}_t\right\|_2^2 \leq C^*$ and $E\left\|\underline{Y}_t\right\|_2^2 \leq C^*$.[14] $\square$

Lemma A2: Suppose that Assumptions 2-1, 2-2, 2-3, 2-5, 2-6, and 2-8 hold and suppose that $N_1, N_2, T \to \infty$ such that $N_1/\tau_1^3 = N_1/\left\lfloor T_0^{\alpha_1}\right\rfloor^3 \to 0$. Then, the following statements are true: (a) $\max_{1\leq\ell\leq d,\ i\in H^c} \left|\left(\overline{S}_{i,\ell,T} - \mu_{i,\ell,T}\right)/\mu_{i,\ell,T}\right| \xrightarrow{p} 0$ and (b) $\max_{1\leq\ell\leq d,\ i\in H^c} \left|\left(\overline{V}_{i,\ell,T} - \pi_{i,\ell,T}\right)/\pi_{i,\ell,T}\right| \xrightarrow{p} 0$.

Proof of Lemma A2: To show part (a), let $b_1(r) = (r-1)\tau + p$ and $b_2(r) = b_1(r) + \tau_1 - 1$. Note first that:

$\max_{1\leq\ell\leq d,\ i\in H^c} \left|\left(\overline{S}_{i,\ell,T} - \mu_{i,\ell,T}\right)/\left(q\tau_1\right)\right|$

[13] Explicit proofs of these two inequalities are given in the Technical Appendix of an earlier version of this paper, Chao and Swanson (2022b). In particular, these inequalities are shown in parts (a) and (b) of Lemma C-7 in Appendix C of Chao and Swanson (2022b).

[14] An explicit proof that, under Assumptions 2-1, 2-2(a)-(b), 2-5, and 2-6; there exists some positive constant $C^{\#}$ such that $E\left\|\underline{F}_t\right\|_2^6 \leq C^{\#}$ and $E\left\|\underline{Y}_t\right\|_2^6 \leq C^{\#}$ is given in the Technical Appendix of an earlier version of this paper, Chao and Swanson (2022b). In particular, this result is shown in Lemma C-5 in Appendix C of Chao and Swanson (2022b). By Liapunov's inequality, the existence of the sixth moment then implies the existence of the second moment.

$$= \max_{1\leq \ell \leq d,\ i\in H^c} \left| \frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} \gamma_i'\underline{F}_t\left[\mu_{Y,\ell}+\underline{Y}_t'\alpha_{YY,\ell}+\underline{F}_t'\alpha_{YF,\ell}\right] \right.$$

$$+\frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} \gamma_i'\underline{F}_t\varepsilon_{\ell,t+1}+\frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} y_{\ell,t+1}u_{it}$$

$$\left. -\frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} \left\{\gamma_i'E\left[\underline{F}_t\right]\mu_{Y,\ell}+\gamma_i'E\left[\underline{F}_t\underline{Y}_t'\right]\alpha_{YY,\ell}+\gamma_i'E\left[\underline{F}_t\underline{F}_t'\right]\alpha_{YF,\ell}\right\} \right|$$

$$\leq \max_{1\leq \ell \leq d,\ i\in H^c} \left| \frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} \gamma_i'\left(\underline{F}_t-E\left[\underline{F}_t\right]\right)\mu_{Y,\ell} \right|$$

$$+\max_{1\leq \ell \leq d,\ i\in H^c} \left| \frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} \gamma_i'\left(\underline{F}_t\underline{Y}_t'-E\left[\underline{F}_t\underline{Y}_t'\right]\right)\alpha_{YY,\ell} \right|$$

$$+\max_{1\leq \ell \leq d,\ i\in H^c} \left| \frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} \gamma_i'\left(\underline{F}_t\underline{F}_t'-E\left[\underline{F}_t\underline{F}_t'\right]\right)\alpha_{YF,\ell} \right|$$

$$+\max_{1\leq \ell \leq d, i\in H^c} \left| \frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} \gamma_i'\underline{F}_t\varepsilon_{\ell,t+1} \right| +\max_{1\leq \ell \leq d, i\in H^c} \left| \frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} y_{\ell,t+1}u_{it} \right|$$

$$= o_p(1).$$

The last equality above follows from applying the Slutsky's theorem given that, by some tedious calculations, one can show the following five results:[15]

Result (i): $\max_{1\leq \ell \leq d,\ i\in H^c} \left| \frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} \gamma_i'\left(\underline{F}_t-E\left[\underline{F}_t\right]\right)\mu_{Y,\ell} \right| = o_p(1)$,

Result (ii): $\max_{1\leq \ell \leq d,\ i\in H^c} \left| \frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} \gamma_i'\left(\underline{F}_t\underline{Y}_t'-E\left[\underline{F}_t\underline{Y}_t'\right]\right)\alpha_{YY,\ell} \right| = o_p(1)$,

Result (iii): $\max_{1\leq \ell \leq d,\ i\in H^c} \left| \frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} \gamma_i'\left(\underline{F}_t\underline{F}_t'-E\left[\underline{F}_t\underline{F}_t'\right]\right)\alpha_{YF,\ell} \right| = o_p(1)$,

Result (iv): $\max_{1\leq \ell \leq d,\ i\in H^c} \left| \frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} \gamma_i'\underline{F}_t\varepsilon_{\ell,t+1} \right| = o_p(1)$,

[15] Detailed proofs of these results are given in the Technical Appendix (more specifically, in Appendix C) of an earlier version of our paper, Chao and Swanson (2022b). In particular, result (i) is shown in part (c) of Lemma C-12; result (ii) is shown in part (a) of Lemma C-12; result (iii) is shown in part (b) of Lemma C-12; result (iv) is shown in part (a) of Lemma C-6; and result (v) is shown in part (c) of Lemma C-6.

Result (v): $\max_{1\leq \ell\leq d,\ i\in H^c}\left|\frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} y_{\ell,t+1}u_{it}\right| = o_p(1)$.

Moreover, by Assumption 2-8, there exists a positive constant $\underline{c}$ such that for all N_1, N_2, and T sufficiently large; we have that:

$$\begin{aligned}
&\min_{1\leq \ell\leq d,\ i\in H^c}\left|\frac{\mu_{i,\ell,T}}{q\tau_1}\right| \\
=\ &\min_{1\leq \ell\leq d,\ i\in H^c}\left|\frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\gamma_i'\left\{E\left[\underline{F}_t\right]\mu_{Y,\ell}+E\left[\underline{F}_t\underline{Y}_t'\right]\alpha_{YY,\ell}+E\left[\underline{F}_t\underline{F}_t'\right]\alpha_{YF,\ell}\right\}\right| \\
\geq\ &\underline{c}>0.
\end{aligned}$$

It thus follows that: $\max_{1\leq \ell\leq d,\ i\in H^c}\left|\left(\overline{S}_{i,\ell,T}-\mu_{i,\ell,T}\right)/\mu_{i,\ell,T}\right|$
$\leq \max_{1\leq \ell\leq d,\ i\in H^c}\left|\left(\overline{S}_{i,\ell,T}-\mu_{i,\ell,T}\right)/\left(q\tau_1\right)\right|/\min_{1\leq \ell\leq d,\ i\in H^c}\left|\mu_{i,\ell,T}/\left(q\tau_1\right)\right| = o_p(1)$.

To show part (b), first let $\psi_\ell\left(\underline{F}_t,\underline{Y}_t'\right)=\left(\underline{F}_t-E\left[\underline{F}_t\right]\right)\mu_{Y,\ell}$
$+\left(\underline{F}_t\underline{Y}_t'-E\left[\underline{F}_t\underline{Y}_t'\right]\right)\alpha_{YY,\ell}+\left(\underline{F}_t\underline{F}_t'-E\left[\underline{F}_t\underline{F}_t'\right]\right)\alpha_{YF,\ell}$; and note that:

$$\pi_{i,\ell,T}/\left(q\tau_1^2\right)=\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\gamma_i'\left\{E\left[\underline{F}_t\right]\mu_{Y,\ell}+E\left[\underline{F}_t\underline{Y}_t'\right]\alpha_{YY,\ell}+E\left[\underline{F}_t\underline{F}_t'\right]\alpha_{YF,\ell}\right\}\right)^2.$$

Now, applying the triangle and Cauchy-Schwarz inequalities, we see that:

$$\begin{aligned}
&\max_{1\leq \ell\leq d,\ i\in H^c}\left|\left(\overline{V}_{i,\ell,T}-\pi_{i,\ell,T}\right)/\left(q\tau_1^2\right)\right| \\
&\leq \max_{1\leq \ell\leq d,\ i\in H^c}\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\gamma_i'\psi_\ell\left(\underline{F}_t,\underline{Y}_t'\right)\right)^2 \\
&+2\left\{\sqrt{\max_{1\leq \ell\leq d,\ i\in H^c}\frac{\pi_{i,\ell,T}}{q\tau_1^2}}\sqrt{\max_{1\leq \ell\leq d,\ i\in H^c}\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\gamma_i'\psi_\ell\left(\underline{F}_t,\underline{Y}_t'\right)\right)^2}\right\} \\
&+\max_{1\leq \ell\leq d, i\in H^c}\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\gamma_i'\underline{F}_t\varepsilon_{\ell,t+1}\right)^2+\max_{1\leq \ell\leq d, i\in H^c}\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}y_{\ell,t+1}u_{it}\right)^2 \\
&+2\left\{\sqrt{\max_{1\leq \ell\leq d, i\in H^c}\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\gamma_i'\underline{F}_t\varepsilon_{\ell,t+1}\right)^2}\right. \\
&\quad\left.\times\sqrt{\max_{1\leq \ell\leq d, i\in H^c}\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}y_{\ell,t+1}u_{it}\right)^2}\right\}
\end{aligned}$$

$$+2\sqrt{\max_{1\leq\ell\leq d,\ i\in H^c}\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\gamma_i'\underline{F}_t\left[\mu_{Y,\ell}+\underline{Y}_t'\alpha_{YY,\ell}+\underline{F}_t'\alpha_{YF,\ell}\right]\right)^2}$$
$$\times\sqrt{\max_{1\leq\ell\leq d,\ i\in H^c}\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}y_{\ell,t+1}u_{it}\right)^2}$$
$$+2\sqrt{\max_{1\leq\ell\leq d,\ i\in H^c}\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\gamma_i'\underline{F}_t\left[\mu_{Y,\ell}+\underline{Y}_t'\alpha_{YY,\ell}+\underline{F}_t'\alpha_{YF,\ell}\right]\right)^2}$$
$$\times\sqrt{\max_{1\leq\ell\leq d,\ i\in H^c}\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\gamma_i'\underline{F}_t\varepsilon_{\ell,t+1}\right)^2}$$
$$=o_p(1).$$

The last equality above follows from applying the Slutsky's theorem given that there exists a positive constant C such that $\max_{1\leq\ell\leq d,i\in H^c}\pi_{i,\ell,T}/\left(q\tau_1^2\right)\leq C<\infty$ for all N_1, N_2, T sufficiently large, as shown previously in Lemma A1, and given that, by some tedious calculations, one can show the following four results:[16]

Result (vi): $\max_{1\leq\ell\leq d,i\in H^c}\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\gamma_i'\psi_\ell\left(\underline{F}_t,\underline{Y}_t'\right)\right)^2=o_p(1),$

Result (vii): $\max_{1\leq\ell\leq d,i\in H^c}\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\gamma_i'\underline{F}_t\varepsilon_{\ell,t+1}\right)^2=o_p(1),$

Result (viii): $\max_{1\leq\ell\leq d,i\in H^c}\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}y_{\ell,t+1}u_{it}\right)^2=o_p(1),$

Result (ix): $\max_{1\leq\ell\leq d,i\in H^c}\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\gamma_i'\underline{F}_t\left[\mu_{Y,\ell}+\underline{Y}_t'\alpha_{YY,\ell}+\underline{F}_t'\alpha_{YF,\ell}\right]\right)^2=O_p(1).$

Moreover, note that, for all N_1, N_2, and T sufficiently large,

$\min_{1\leq\ell\leq d,\ i\in H^c}\left\{\pi_{i,\ell,T}/\left(q\tau_1^2\right)\right\}$
$$=\min_{1\leq\ell\leq d,i\in H^c}\frac{1}{q}\sum_{r=1}^{q}\left(\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)}\gamma_i'\left\{E\left[\underline{F}_t\right]\mu_{Y,\ell}+E\left[\underline{F}_t\underline{Y}_t'\right]\alpha_{YY,\ell}+E\left[\underline{F}_t\underline{F}_t'\right]\alpha_{YF,\ell}\right\}\right)^2$$

[16]Detailed proofs of these results are given in the Technical Appendix (more specifically, in Appendix C) of an earlier version of this paper, Chao and Swanson (2022b). In particular, result (vi) is shown in part (d) of Lemma C-12; result (vii) is shown in part (b) of Lemma C-6; result (viii) is shown in part (d) of Lemma C-6; and result (ix) is shown in part (f) of Lemma C-12.

$\geq \min_{1\leq \ell \leq d, i\in H^c} \left(\frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} \gamma_i' \left\{ E\left[\underline{F}_t\right]\mu_{Y,\ell} + E\left[\underline{F}_t \underline{Y}_t'\right]\alpha_{YY,\ell} + E\left[\underline{F}_t \underline{F}_t'\right]\alpha_{YF,\ell} \right\} \right)^2$

(by Jensen's inequality)

$= \min_{1\leq \ell \leq d, i\in H^c} \left| \frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} \gamma_i' \left\{ E\left[\underline{F}_t\right]\mu_{Y,\ell} + E\left[\underline{F}_t \underline{Y}_t'\right]\alpha_{YY,\ell} + E\left[\underline{F}_t \underline{F}_t'\right]\alpha_{YF,\ell} \right\} \right|^2$

$= \left(\min_{1\leq \ell \leq d, i\in H^c} \left| \frac{1}{q}\sum_{r=1}^{q}\frac{1}{\tau_1}\sum_{t=b_1(r)}^{b_2(r)} \gamma_i' \left\{ E\left[\underline{F}_t\right]\mu_{Y,\ell} + E\left[\underline{F}_t \underline{Y}_t'\right]\alpha_{YY,\ell} + E\left[\underline{F}_t \underline{F}_t'\right]\alpha_{YF,\ell} \right\} \right| \right)^2$

$\geq \underline{c}^2 > 0$ (by Assumption 2-8).

It, thus, follows that $\max_{1\leq \ell \leq d,\ i\in H^c} \left| \left(\overline{V}_{i,\ell,T} - \pi_{i,\ell,T}\right)/\pi_{i,\ell,T} \right|$

$\leq \max_{1\leq \ell \leq d,\ i\in H^c} \left| \left(\overline{V}_{i,\ell,T} - \pi_{i,\ell,T}\right)/\left(q\tau_1^2\right) \right| / \min_{1\leq \ell \leq d,\ i\in H^c} \left(\pi_{i,\ell,T}/\left(q\tau_1^2\right)\right) = o_p(1)\ \square.$

Lemma A3: Suppose that Assumptions 2-1, 2-2, 2-3, 2-4, 2-5, 2-6, and 2-7 hold. Let $\Phi(\cdot)$ denote the cumulative distribution function of the standard normal random variable. Then, there exists a positive constant A such that

$$P\left(\left|S_{i,\ell,T}\right| \geq z\right) \leq 2\left[1-\Phi(z)\right]\left\{1 + A(1+z)^3 T^{-(1-\alpha_1)\frac{1}{2}}\right\} \tag{23}$$

for $i \in H = \{k \in \{1, ..., N\} : \gamma_k = 0\}$, for $\ell \in \{1, ..., d\}$, for T sufficiently large, and for all z such that $0 \leq z \leq c_0 \min\left\{T^{(1-\alpha_1)/6}, T^{\alpha_2/2}\right\}$ with c_0 being a positive constant.

Proof of Lemma A3: Note first that, for any i such that $i \in H = \{k \in \{1, ..., N\} : \gamma_k = 0\}$, the formula for $S_{i,\ell,T}$ reduces to:

$S_{i,\ell,T} = \left(\sum_{r=1}^{q}\left[\sum_{t=(r-1)\tau+p}^{(r-1)\tau+\tau_1+p-1} y_{\ell,t+1}u_{it}\right]^2\right)^{-\frac{1}{2}} \sum_{r=1}^{q}\sum_{t=(r-1)\tau+p}^{(r-1)\tau+\tau_1+p-1} y_{\ell,t+1}u_{it}.$

Hence, to verify the conditions of Theorem 4.1 of Chen, Shao, Wu, and Xu (2016), we set $X_{it} = u_{it}y_{\ell,t+1}$, and note that $E\left[X_{it}\right] = E\left[u_{it}y_{\ell,t+1}\right] = E_Y\left[E\left[u_{it}\right]y_{\ell,t+1}\right] = 0$, where the second equality follows by the law of iterated expectations given that Assumption 2-4 implies the independence of u_{it} and $y_{\ell,t+1}$ and where the third equality follows by Assumption 2-3(a). Hence, the first part of condition (4.1) of Chen, Shao, Wu, and Xu (2016) is fulfilled. Moreover, in light of Assumption 2-3(b) and in light of the fact that, under Assumptions 2-1, 2-2(a)-(b), 2-5, and 2-6; one can show by straightforward calculations that there exists a positive constant $\overline{C}$ such that $E\left\|\underline{Y}_t\right\|_2^6 \leq \overline{C}$[17]; we see

[17] As noted previously, under Assumptions 2-1, 2-2(a)-(b), 2-5, and 2-6; an explicit proof that there exists some positive constant $C^{\#}$ such that $E\left\|\underline{Y}_t\right\|_2^6 \leq C^{\#}$ is given in the Technical Appendix

that there exists some positive constant c_1 such that, for every $\ell \in \{1..., d\}$,

$$\begin{aligned} E\left[|X_{it}|^{\frac{31}{10}}\right] &= E\left[|u_{it}y_{\ell,t+1}|^{\frac{31}{10}}\right] \leq \left(E|u_{it}|^{\frac{186}{29}}\right)^{\frac{29}{60}} \left(E|y_{\ell,t+1}|^{6}\right)^{\frac{31}{60}} \\ &\leq \left[\left(E|u_{it}|^{\frac{186}{29}}\right)^{\frac{29}{186}}\right]^{\frac{31}{10}} \left[E\left(\sum_{k=1}^{d}\sum_{j=0}^{p-1} y_{k,t+1-j}^{2}\right)^{3}\right]^{\frac{31}{60}} \\ &\leq \left[\left(E|u_{it}|^{7}\right)^{\frac{1}{7}}\right]^{\frac{31}{10}} \left[\left(E\left\|\underline{Y}_{t+1}\right\|_2^6\right)^{\frac{1}{6}}\right]^{\frac{31}{10}} \leq c_1^{\frac{31}{10}}, \end{aligned}$$

where the first and third inequalities above follow, respectively, by Hölder's and Liapunov's inequalities. Hence, the second part of condition (4.1) of Chen, Shao, Wu, and Xu (2016) is also fulfilled with $r = \frac{31}{10} > 2$. Moreover, note that, by Assumption 2-7, for all $r \geq 1$ and $\tau_1 \geq 1$:

$$E\left\{\left[\sum_{t=(r-1)\tau+p}^{(r-1)\tau+\tau_1+p-1} X_{it}\right]^2\right\} = \tau_1 E\left\{\left[\frac{1}{\sqrt{\tau_1}}\sum_{t=(r-1)\tau+p}^{(r-1)\tau+\tau_1+p-1} y_{\ell,t+1}u_{it}\right]^2\right\} \geq \tau_1 \underline{c},$$

so that condition (4.2) of Chen, Shao, Wu, and Xu (2016) is satisfied here. Now, making use of Assumption 2-3(c) and Assumption 2-4 and applying Theorem 2.1 of Pham and Tran (1985), it can be shown that $\left\{(y_{\ell,t+1}, u_{it})'\right\}$ is β mixing with β mixing coefficient satisfying $\beta(m) \leq \overline{a}_1 \exp\{-a_2 m\}$ for some constants $\overline{a}_1 > 0$ and $a_2 > 0$. Next, define $X_{it} = y_{\ell,t+1}u_{it}$ as before, and note that $\{X_{it}\}$ is a β-mixing process with β-mixing coefficient $\beta_{X,m}$ satisfying the condition $\beta_{X,m} \leq a_1 \exp\{-a_2 m\}$ for some constant $a_1 > 0$ and for all m sufficiently large, given that measurable functions of a finite number of β-mixing random variables are also β-mixing, with β-mixing coefficients having the same order of magnitude[18]. It follows that $\{X_{it}\}$ satisfies the β mixing condition (2.1) stipulated in Chen, Shao, Wu, and Xu (2016) for all $i \in H$. Hence, by applying Theorem 4.1 of Chen, Shao, Wu, and Xu (2016) for the case where

of an earlier version of this paper, Chao and Swanson (2022b). In particular, this result is shown in Lemma C-5 in Appendix C of Chao and Swanson (2022b).

[18]For α-mixing and ϕ-mixing, this result is given in Theorem 14.1 of Davidson (1994). However, using essentially the same argument as that given in the proof of Theorem 14.1, one can also prove a similar result for β-mixing. An explicit proof of this result is given in the Technical Appendix of an earlier version of this paper, Chao and Swanson (2022b). In particular, this result is shown in Lemma C-2 part (a) in Appendix C of Chao and Swanson (2022b).

$\delta = 1$[19], we obtain the Cramér-type moderate deviation result

$$\frac{P\left\{\overline{S}_{i,\ell,T}/\sqrt{\overline{V}_{i,\ell,T}} \geq z\right\}}{1-\Phi\left(z\right)} = 1 + O\left(1\right)\left(1+z\right)^{3} T^{-(1-\alpha_1)\frac{1}{2}}, \tag{24}$$

which holds for all $0 \leq z \leq c_0 \min\left\{T^{(1-\alpha_1)/6}, T^{\alpha_2/2}\right\}$ and for $|O\left(1\right)| \leq A$ with A being an absolute constant.

Next, consider obtaining a moderate deviation result for $P\left\{-\overline{S}_{i,\ell,T}/\sqrt{\overline{V}_{i,\ell,T}} \geq z\right\}/\left[1-\Phi\left(z\right)\right]$. As $\overline{S}_{i,\ell,T} = \sum_{r=1}^{q}\sum_{t=(r-1)\tau+p}^{(r-1)\tau+\tau_1+p-1}\left(-u_{it}y_{\ell,t+1}\right)$, we can take $X_{it} = -u_{it}y_{\ell,t+1}$, and note that, by calculations similar to those given above, we have $E\left[X_{it}\right] = E\left[-u_{it}y_{\ell,t+1}\right] = 0$, $E\left[\left|X_{it}\right|^{\frac{31}{10}}\right] = E\left[\left|-u_{it}y_{\ell,t+1}\right|^{\frac{31}{10}}\right] = E\left[\left|u_{it}y_{\ell,t+1}\right|^{\frac{31}{10}}\right] \leq c_1^{\frac{31}{10}}$, and

$$E\left\{\left[\sum_{t=(r-1)\tau+p}^{(r-1)\tau+\tau_1+p-1} X_{it}\right]^2\right\} = E\left\{\left[\sum_{t=(r-1)\tau+p}^{(r-1)\tau+\tau_1+p-1}\left(-u_{it}y_{\ell,t+1}\right)\right]^2\right\} \geq \underline{c}\tau_1.$$

Moreover, it is easily seen that $\{X_{it}\}$ (with $X_{it} = -u_{it}y_{\ell,t+1}$) also satisfies the β mixing condition (2.1) stipulated in Chen, Shao, Wu, and Xu (2016) for every i. Thus, by applying Theorem 4.1 of Chen, Shao, Wu, and Xu (2016), we also obtain the Cramér-type moderate deviation result

$$\frac{P\left\{-\overline{S}_{i,\ell,T}/\sqrt{\overline{V}_{i,\ell,T}} \geq z\right\}}{1-\Phi\left(z\right)} = 1 + O\left(1\right)\left(1+z\right)^{3} T^{-(1-\alpha_1)\frac{1}{2}}, \tag{25}$$

which holds for all $0 \leq z \leq c_0 \min\left\{T^{(1-\alpha_1)/6}, T^{\alpha_2/2}\right\}$ and for $|O\left(1\right)| \leq A$ with A being an absolute constant. Next, note that:

$$\begin{aligned}
\left|\frac{P\left(\left|S_{i,\ell,T}\right| \geq z\right)}{2\left[1-\Phi(z)\right]} - 1\right| &= \left|\frac{P\left(\left|\overline{S}_{i,\ell,T}/\sqrt{\overline{V}_{i,\ell,T}}\right| \geq z\right)}{2\left[1-\Phi(z)\right]} - 1\right| \\
&= \left|\frac{P\left(\left\{\overline{S}_{i,\ell,T}/\sqrt{\overline{V}_{i,\ell,T}} \geq z\right\} \cup \left\{-\overline{S}_{i,\ell,T}/\sqrt{\overline{V}_{i,\ell,T}} \geq z\right\}\right)}{2\left[1-\Phi(z)\right]} - 1\right| \\
&= \left|\frac{P\left(\overline{S}_{i,\ell,T}/\sqrt{\overline{V}_{i,\ell,T}} \geq z\right) + P\left(-\overline{S}_{i,\ell,T}/\sqrt{\overline{V}_{i,\ell,T}} \geq z\right)}{2\left[1-\Phi(z)\right]} - 1\right|
\end{aligned}$$

[19] Note that Theorem 4.1 of Chen, Shao, Wu and Xu (2016) requires that $0 < \delta \leq 1$ and $\delta < r - 2$. These conditions are satisfied here given that we choose $\delta = 1$ and $r = 31/10$.

$$\left(\text{since } \left\{\overline{S}_{i,\ell,T}/\sqrt{\overline{V}_{i,\ell,T}} \geq z\right\} \cap \left\{-\overline{S}_{i,\ell,T}/\sqrt{\overline{V}_{i,\ell,T}} \geq z\right\} = \varnothing \text{ w.p.1}\right)$$
$$\leq \tfrac{1}{2}\left|\tfrac{P\left(\overline{S}_{i,\ell,T}/\sqrt{\overline{V}_{i,\ell,T}} \geq z\right)}{[1-\Phi(z)]} - 1\right| + \tfrac{1}{2}\left|\tfrac{P\left\{-\overline{S}_{i,\ell,T}/\sqrt{\overline{V}_{i,\ell,T}} \geq z\right\}}{1-\Phi(z)} - 1\right|.$$

Thus, in light of expressions (24) and (25), we have that:

$$\begin{aligned}
&\left|\frac{P\left(\left|S_{i,\ell,T}\right| \geq z\right)}{2\left[1-\Phi\left(z\right)\right]} - 1\right| \\
\leq\;& \frac{1}{2}\left|\frac{P\left(\overline{S}_{i,\ell,T}/\sqrt{\overline{V}_{i,\ell,T}} \geq z\right)}{\left[1-\Phi\left(z\right)\right]} - 1\right| + \frac{1}{2}\left|\frac{P\left\{-\overline{S}_{i,\ell,T}/\sqrt{\overline{V}_{i,\ell,T}} \geq z\right\}}{1-\Phi\left(z\right)} - 1\right| \\
\leq\;& \frac{A}{2}\left(1+z\right)^{3} T^{-(1-\alpha_1)\frac{1}{2}} + \frac{A}{2}\left(1+z\right)^{3} T^{-(1-\alpha_1)\frac{1}{2}} = A\left(1+z\right)^{3} T^{-(1-\alpha_1)\frac{1}{2}}
\end{aligned}$$

It then follows that:

$$-A\left(1+z\right)^{3} T^{-(1-\alpha_1)\frac{1}{2}} \leq \frac{P\left(\left|S_{i,\ell,T}\right| \geq z\right)}{2\left[1-\Phi\left(z\right)\right]} - 1 \leq A\left(1+z\right)^{3} T^{-(1-\alpha_1)\frac{1}{2}} \tag{26}$$

Focusing on the right-hand part of the inequality in (26), we have that: $P\left(\left|S_{i,\ell,T}\right| \geq z\right)/\left(2\left[1-\Phi\left(z\right)\right]\right) - 1 \leq A\left(1+z\right)^{3} T^{-(1-\alpha_1)\frac{1}{2}}$. Simple rearrangement of this inequality then leads to the desired result:

$$P\left(\left|S_{i,\ell,T}\right| \geq z\right) \leq 2\left[1-\Phi\left(z\right)\right]\left\{1 + A\left(1+z\right)^{3} T^{-(1-\alpha_1)\frac{1}{2}}\right\},$$

which holds for all $i \in H = \{k \in \{1, ..., N\} : \gamma_k = 0\}$, for every $\ell \in \{1, ..., d\}$, for all T sufficiently large, and for all z such that $0 \leq z \leq c_0 \min\left\{T^{(1-\alpha_1)/6}, T^{\alpha_2/2}\right\}$. □