

The Role of Pricing errors in Linear Asset Pricing Models with Strong, Semi-strong, and Latent Factors

M. Hashem Pesaran
University of Southern California,
and Trinity College,
Cambridge

Ron P. Smith
Birkbeck, University
of London

Abstract

This paper examines the role of pricing errors in linear factor pricing models, allowing for observed strong and semi-strong factors, and latent weak factors. It focusses on the estimation of $\Phi_k = \lambda_k - \mu_k$ which plays a pivotal role, not only in the estimation of risk premia but also in tests of market efficiency, where λ_k and μ_k are respectively the risk premium and the mean of the k^{th} risk factor. It proposes a two-step estimator of Φ_k with Shanken type bias-correction, and derives its asymptotic distribution under a general setting that allows for idiosyncratic pricing errors, weak missing factors, as well as weak error cross-sectional dependence. The implications of semi-strong factors for the asymptotic distribution of the proposed estimator are also investigated. Small sample results from extensive Monte Carlo experiments show that the proposed estimator has the correct size with good power properties. The paper also provides an empirical application to a large number of U.S. securities with risk factors selected from a large number of potential risk factors according to their strength.

Reference Details

CWPE 2317
Published 10 February 2023

Key Words Factor strength, pricing errors, risk premia, missing factors, Fama-French factors, panel R2

JEL Codes C38, G10

Website www.econ.cam.ac.uk/cwpe

The role of pricing errors in linear asset pricing models with strong, semi-strong, and latent factors*

M. Hashem Pesaran

University of Southern California, and Trinity College, Cambridge

Ron P. Smith

Birkbeck, University of London

February 17, 2023

Abstract

This paper examines the role of pricing errors in linear factor pricing models, allowing for observed strong and semi-strong factors, and latent weak factors. It focusses on the estimation of $\phi_k = \lambda_k - \mu_k$ which plays a pivotal role, not only in the estimation of risk premia but also in tests of market efficiency, where λ_k and μ_k are respectively the risk premium and the mean of the k^{th} risk factor. It proposes a two-step estimator of ϕ_k with Shanken type bias-correction, and derives its asymptotic distribution under a general setting that allows for idiosyncratic pricing errors, weak missing factors, as well as weak error cross-sectional dependence. The implications of semi-strong factors for the asymptotic distribution of the proposed estimator are also investigated. Small sample results from extensive Monte Carlo experiments show that the proposed estimator has the correct size with good power properties. The paper also provides an empirical application to a large number of U.S. securities with risk factors selected from a large number of potential risk factors according to their strength.

JEL Classifications: C38, G10

Key Words: Factor strength, pricing errors, risk premia, missing factors, Fama-French factors, panel R^2 .

*We are grateful to Juan Espinosa Torres and Jing Kong for carrying out the computations and to Hayun Song for excellent research assistance, all are PhD students at USC. We are also grateful to Takashi Yamagata and Natalia Bailey for helpful comments.

1 Introduction

The linear factor pricing model (LFPM), widely used in finance, explains the time series of excess returns on each of $i = 1, 2, \dots, n$ securities in terms of a set of observed risk factors, f_{kt} , $k = 1, 2, \dots, K$, $t = 1, 2, \dots, T$. The popular Fama and MacBeth (1973, FM) two-pass procedure first runs regressions of individual excess returns on the risk factors to estimate their loadings, β_{ik} , which are then used in the second pass to estimate the risk premia of f_{kt} , namely λ_k , using cross section regressions of the time averages of security excess returns on the estimated factor loadings. In this paper we focus on estimation of $\phi_k = \lambda_k - \mu_k$, where $\mu_k = E(f_{kt})$ can be readily estimated from time series data on risk factors and does not require knowing the factor loadings (and hence is not subject to Shanken bias-correction). ϕ_k plays a pivotal role both in estimation of the risk premia as well as in tests of alpha. This follows since $a_i = c + \sum_{k=1}^K \beta_{ik}\phi_k + \eta_i$, where a_i is the intercept in the i^{th} return regression, c is the zero beta return, $\beta_{ik}\phi_k$ is what we refer to as the spanning error of the k^{th} risk factor, and η_i is the idiosyncratic pricing error specific to security i , originally introduced by Ross (1976). Under market efficiency $a_i = 0$ (or $a_i = c$ if the risk free rate is fixed but unknown), and the spanning error, $\beta_{ik}\phi_k$, must be zero for $k = 1, 2, \dots, K$, which in turn implies that $\phi_k = 0$, when the associated risk factor, f_{kt} , is sufficiently strong. A two-step estimator of $\phi = (\phi_1, \phi_2, \dots, \phi_K)'$ can be obtained from a regression of the estimated intercepts, \hat{a}_i , on the estimated loadings, $\hat{\beta}_{ik}$, from the LFPM. As with the two-step estimator of λ_k , the two-step estimator of ϕ_k will also be biased, and requires bias-correction. Accordingly, we follow Shanken (1992) and propose a bias-corrected estimator of ϕ , which we denote by $\tilde{\phi}_{nT}$.

The main theoretical contribution of the paper is to develop the asymptotic distribution of $\tilde{\phi}_{nT}$ under a wide variety of assumptions regarding the idiosyncratic pricing errors, error cross-sectional dependence, and the presence of missing (latent) factors. The paper also investigates the implications of factor strengths for the precision with which ϕ_0 (the true value of ϕ) can be estimated. The LFPM, following Chamberlain and Rothschild (1983), assumes that all the observed factors are strong and the eigenvalues of the covariance matrix of the errors are bounded, and hence do not admit a factor structure. There has been concern in the literature that the factors may be weak

or not pervasive in that many of the β_{ik} may be zero. In our analysis we use a measure of the strength of factor k , α_k , developed in Bailey, Kapetanios, and Pesaran (2016, 2021), which can be estimated from the proportion of non-zero coefficients, β_{ik} , of the factor in the LFPM. Similar strength measures can be used to measure the degree of pervasiveness of missing factors and pricing errors. Strong factors have strength equal one, semi-strong between one and a half, and weak factors less than a half. Use of this measure allows us to be precise about the degree of pervasiveness and show how the strengths of the observed factors, the missing factors and the pricing errors influence estimation and inference about risk premia.

As noted above, the intercepts in the LFPM return regressions, a_i , can be written in terms of two types of pricing error as: $a_i = c + \sum_{k=1}^K \beta_{ik}\phi_k + \eta_i$, where c is the zero beta return, $\beta_{ik}\phi_k$ is a spanning error related to factor k , and η_i is an idiosyncratic pricing error specific to security i . The risk premium of factor f_{kt} , $\lambda_k = \phi_k + \mu_k$, is the sum of the spanning error parameter, ϕ_k , and the factor mean, $\mu_k = E(f_{kt})$. In his theoretical derivations, Ross (1976) assumed the factors had mean zero: $\mu_k = 0$, so $\lambda_k = \phi_k$. Shanken (1992) considered risk premia of traded factors under market efficiency (setting $\phi_k = 0$), and hence $\lambda_k = \mu_k$. The factor mean, μ_k , can be estimated consistently at the regular \sqrt{T} rate directly using time series data $\{f_{kt}, \text{ for } t = 1, 2, \dots, T\}$ alone.

The main theoretical results of the paper are set out around five theorems under a number of key assumptions, with proofs provided in the Appendix. Theorem 1 shows that the standard Fama-MacBeth estimator is valid only when there are no pricing errors and when $n/T \rightarrow 0$. Theorem 2 shows that the Shanken bias-corrected estimator of λ_k continues to be consistent for a fixed T as $n \rightarrow \infty$, even in presence of weak pricing errors and weak missing common factors. Theorem 3 provides conditions under which the bias-corrected estimator, $\tilde{\phi}_{nT}$, is consistent for ϕ_0 , and derives the asymptotic distribution of $\tilde{\phi}_{nT}$, assuming the observed factors are strong. Theorem 4 extends the results to the case with semi-strong factors and considers the relevance of factor strength for the precision with which ϕ can be estimated. Finally, Theorem 5 gives conditions for consistent estimation of the asymptotic variance of $\tilde{\phi}_{nT}$, using a suitable threshold estimator of the covariance matrix. In summary, the asymptotic distribution of the bias corrected estimator, $\tilde{\phi}_{nT}$, is established in presence of missing factors and pricing errors, subject to certain regularity conditions on the strength of

the missing factors and the pricing errors.

The small sample properties of $\tilde{\phi}_{nT}$ are investigated using extensive Monte Carlo experiments, allowing for a mixture of strong and semi-strong observed factors, latent factors, pricing errors, GARCH effects and non-Gaussian errors. Small sample results are in line with our theoretical findings, and confirm that the bias-corrected estimator, $\tilde{\phi}_{nT}$, has the correct size and good power properties for samples with time series dimensions of $T = 120$ and $T = 240$. The results are quite robust to non-Gaussian errors, GARCH effects, pricing errors, missing weak factors and various forms of error cross section dependence. In line with our theoretical findings, the Monte Carlo evidence also showed that the precision with which ϕ_k is estimated falls with α_k , the strength of the k^{th} factor.

In our theoretical derivations and Monte Carlo investigations we assume that the observed risk factors are known. In practice they have to be selected from a large set of potential factors. Given that both the theory and Monte Carlo indicate the importance of factor strength, in the empirical application factor selection is based on a threshold for factor strength. This procedure is applied in a high dimensional setting with both a large number of securities (n from 1,090 to 1,175) and potential risk factors (m from 177 to 189). Various procedures could be used to determine whether a factor had a non-zero coefficient in each return regression. We used Lasso. A factor strength threshold of 0.75 produced a small number of plausible factors always including the three Fama-French factors. Estimates of ϕ_0 were obtained with samples of length $T = 240$ and $T = 120$ months, ending in four years, 2015, 2017, 2019, 2021. The spanning error parameter of the market factor was only significant in the $T = 240$ sample ending in 2021, which included the COVID-19 pandemic. The joint null hypothesis that $\phi_0 = \mathbf{0}$ was rejected in all four periods, suggesting market inefficiency. This was mainly due to two risk factors, a measure of short selling suggested by Dechow, Hutton, Meulbroek, and Sloan (2001), and the Fama-French value factor, HML.

Related Literature. There is a large literature that relates to our research agenda. On estimation of risk premia, following Fama and MacBeth and Shanken (1992), estimation of risk premia is further examined by Shanken and Zhou (2007), Kan, Robotti, and Shanken (2013), and Bai and Zhou (2015). The survey paper by Jagannathan, Skoulakis, and Wang (2010) provide further references.

Testing for market efficiency dates back to Jensen (1968) who proposes testing $a_i = 0$ for each i separately. Gibbons, Ross, and Shanken (1989) provide a joint test for the case where the errors are Gaussian and $n < T$. Gagliardini, Ossola, and Scaillet (2016) develop two-pass regressions of individual stock returns, allowing time-varying risk premia, and propose a standardised Wald test. Raponi, Robotti, and Zaffaroni (2019) propose a test of pricing error in cross section regression for fixed number of time series observations. They use a bias-corrected estimator of Shanken (1992) to standardise their test statistic. Ma, Lan, Su, and Tsai (2020) employ polynomial spline techniques to allow for time variations in factor loadings when testing for alphas. Feng, Lan, Liu, and Ma (2022) propose a max-of-square type test of the intercepts instead of the average used in the literature, and recommend using a combination of the two testing procedures. He, He, Huang, and Zhou (2022) propose two statistics, a Wald type statistic which require n and T to be of the same order of magnitude and a standardised t ratio. Kleibergen (2009) considers testing in the case where the loadings are small. Pesaran and Yamagata (2012, 2023) consider testing that the intercepts in the LFPM are zero when n is large relative to T and there may be non Gaussian errors and weakly cross-correlated errors.

Strong and weak factors in asset returns are considered by Anatolyev and Mikusheva (2022) and Connor and Korajczyk (2022). Beaulieu, Dufour, and Khalaf (2020) discuss the lack of identification of risk premia when many of the loadings are zero. There has also been concern about the consequences of omitted factors. Giglio and Xiu (2021) discuss the problem and try to deal with it using a three-pass method which is valid even when not all factors in the model are specified or observed using principal components of the test assets. Onatski (2012) and Lettau and Pelger (2020a,b) provide extensive discussions of weak factor and latent factors, respectively.

A large number of risk factors have been considered in the empirical literature. We use the Fama and French (1993) three factors in our Monte Carlo design. In our empirical application we use the five factors proposed by Fama and French (2015) and the large set of factors provided by Chen and Zimmermann (2022) . Harvey and Liu (2019) document over 400 factors published in top finance journals. Dello Preite, Uppal, Zaffaroni, and Zviadadze (2022) argue that despite the hundreds of systematic risk factors considered in the literature, there is still a sizable pricing error and that

this can be explained by asset specific risk that reflects market frictions and behavioral biases. There is a large Bayesian literature, including Chib, Zeng, and Zhao (2020), and Hwang and Rubesam (2022) on selecting factor models. The issue of factor selection is also addressed by Fama and French (2018)

The rest of the paper is organized as follows: Section 2 provides the framework for estimation of ϕ_0 . Section 3 sets out the assumptions and states the main theorems. Section 4 presents the Monte Carlo (MC) design, its calibration and a summary of the main findings. Section 5 discusses the problem of factor selection from a large number of potential factors. Section 6 gives the empirical application using monthly data on a large number of individual US stocks and risk factors over the period 1996-2021. Section 7 provides some concluding remarks.

Detailed mathematical proofs are given in an online mathematical Appendix. Further information on data sources, MC calibration, and MC results, plus some supplementary material for the empirical application are provided in two online supplements, A and B.

2 Estimation of risk premia using linear factor pricing model

We follow the literature and consider the LFPM

$$r_{it} = R_{it} - r_t^f = a_i + \beta_i' \mathbf{f}_t + u_{it}, \quad (1)$$

for $i = 1, 2, \dots, n$ and $t = 1, 2, \dots, T$, where R_{it} is the return on security i , r_t^f the risk free rate, and r_{it} is the excess return explained in terms of the $K \times 1$ vector of observed factors $\mathbf{f}_t = (f_{1t}, f_{2t}, \dots, f_{Kt})'$. The intercept a_i and the $K \times 1$ vector of factor loadings, $\beta_i = (\beta_{i1}, \beta_{i2}, \dots, \beta_{iK})'$, are unknown. The idiosyncratic errors, u_{it} have zero means and are assumed to be serially uncorrelated. The observed factors, \mathbf{f}_t , are assumed to be covariance stationary with the constant mean $\boldsymbol{\mu}_0 = E(\mathbf{f}_t)$.

It will prove convenient to write (1) in matrix notation by stacking the excess returns by $t = 1, 2, \dots, T$, for each security i

$$\mathbf{r}_{i0} = a_i \boldsymbol{\tau}_T + \mathbf{F} \beta_i + \mathbf{u}_{i0}, \quad \text{for } i = 1, 2, \dots, n, \quad (2)$$

where $\mathbf{r}_{i0} = (r_{i1}, r_{i2}, \dots, r_{iT})'$, $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_T)'$, $\mathbf{u}_{i0} = (u_{i1}, u_{i2}, \dots, u_{iT})'$, and $\boldsymbol{\tau}_T$ is a

$T \times 1$ vector of ones. Similarly, stacking the excess returns by i for each t we have

$$\mathbf{r}_{ot} = \mathbf{a}_n + \mathbf{B}_n \mathbf{f}_t + \mathbf{u}_{ot}, \text{ for } t = 1, 2, \dots, T, \quad (3)$$

where $\mathbf{r}_{ot} = (r_{1t}, r_{2t}, \dots, r_{nt})'$, $\mathbf{a}_n = (a_1, a_2, \dots, a_n)'$, $\mathbf{B}_n = (\boldsymbol{\beta}_{o1}, \boldsymbol{\beta}_{o2}, \dots, \boldsymbol{\beta}_{oK})$, $\boldsymbol{\beta}_{ok} = (\beta_{1k}, \beta_{2k}, \dots, \beta_{nk})'$, and $\mathbf{u}_{ot} = (u_{1t}, u_{2t}, \dots, u_{nt})'$, with $\mathbf{V}_u = E(\mathbf{u}_{ot} \mathbf{u}'_{ot})$. Under the Arbitrage Pricing Theory (APT) due to Ross (1976) the pricing errors, η_i are defined by

$$\eta_i = E(r_{it}) - c - \boldsymbol{\beta}'_i \boldsymbol{\lambda}_0, \text{ for } i = 1, 2, \dots, n \quad (4)$$

where c is zero-beta expected excess return, $\boldsymbol{\lambda}_0$ is the $K \times 1$ vector of risk premia. Ross assumes that the pricing errors are bounded such that

$$\sum_{i=1}^n \eta_i^2 < \infty. \quad (5)$$

The focus of the literature has been on testing for alpha, $a_i = 0$, and the estimation of the risk premia, $\boldsymbol{\lambda}_0$, using panel data on excess returns, $\{r_{it}, i = 1, 2, \dots, n; t = 1, 2, \dots, T\}$, and \mathbf{F} , the $T \times K$ matrix of observations on the factors.

2.1 Fama-MacBeth and Shanken estimators of risk premia

The risk premia are usually estimated using a two-pass procedure suggested by Fama and MacBeth (1973). The first-pass runs time series regressions of excess returns, r_{it} , on the K observed factors to give estimates of the factor loadings, $\boldsymbol{\beta}_i$:

$$\hat{\boldsymbol{\beta}}_{iT} = (\mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{r}_{io}. \quad (6)$$

The second-pass runs a cross section regression of average returns, $\bar{r}_{io} = T^{-1} \sum_{t=1}^T r_{it}$ on the estimated factor loadings, to obtain the FM estimator of $\boldsymbol{\lambda}_0$:

$$\hat{\boldsymbol{\lambda}}_{nT} = \left(\hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{B}}_{nT} \right)^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \bar{\mathbf{r}}_{no}, \quad (7)$$

where $\hat{\mathbf{B}}_{nT} = (\hat{\boldsymbol{\beta}}_{1T}, \hat{\boldsymbol{\beta}}_{2T}, \dots, \hat{\boldsymbol{\beta}}_{nT})'$, $\bar{\mathbf{r}}_{no} = (\bar{r}_{1T}, \bar{r}_{2T}, \dots, \bar{r}_{nT})'$, $\mathbf{M}_T = \mathbf{I}_T - T^{-1} \boldsymbol{\tau}_T \boldsymbol{\tau}'_T$, $\boldsymbol{\tau}_T$ is a T -dimensional vector of ones, $\mathbf{M}_n = \mathbf{I}_n - n^{-1} \boldsymbol{\tau}_n \boldsymbol{\tau}'_n$, and $\boldsymbol{\tau}_n$ is an n -dimensional vector of ones.

As is well known, when T is finite FM's two-pass estimator is biased due the errors in estimation of factor loadings that do not vanish. The small T bias of the two-pass estimator of $\boldsymbol{\lambda}_0$ has been a source of concern in the empirical literature. Under standard regularity conditions and as $n \rightarrow \infty$, we have

$$\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_0 \rightarrow_p \left[\boldsymbol{\Sigma}_{\beta\beta} + \frac{\bar{\sigma}^2}{T} \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1} \right]^{-1} \left(\boldsymbol{\Sigma}_{\beta\beta} \mathbf{d}_{fT} - \frac{\bar{\sigma}^2}{T} \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1} \boldsymbol{\lambda}_0 \right), \quad (8)$$

where

$$\Sigma_{\beta\beta} = \lim_{n \rightarrow \infty} (n^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{B}_n), \mathbf{d}_{fT} = \hat{\mu}_T - \mu_0, \text{ and } \bar{\sigma}^2 = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \sigma_i^2 > 0. \quad (9)$$

The bias of $\hat{\lambda}_{nT}$ is due to terms that involve \mathbf{d}_{fT} and $\bar{\sigma}^2$. Following Shanken (1992), $\bar{\sigma}_n^2$ can be consistently estimated (for a fixed T) by

$$\hat{\sigma}_{nT}^2 = \frac{\sum_{t=1}^T \sum_{i=1}^n \hat{u}_{it}^2}{n(T - K - 1)}, \quad (10)$$

where

$$\hat{u}_{it} = r_{it} - \hat{a}_{iT} - \hat{\beta}'_{iT} \mathbf{f}_t, \quad (11)$$

and as before \hat{a}_{iT} and $\hat{\beta}_{iT}$ are the OLS estimators of a_i and β_i . Using these results the bias-corrected version of the two-pass estimator is given by¹

$$\tilde{\lambda}_{nT} = \mathbf{H}_{nT}^{-1} \left(\frac{\hat{\mathbf{B}}'_{nT} \mathbf{M}_n \bar{\mathbf{r}}_{no}}{n} \right), \quad (12)$$

where

$$\mathbf{H}_{nT} = \frac{\hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{B}}_{nT}}{n} - T^{-1} \hat{\sigma}_{nT}^2 \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1}. \quad (13)$$

When all the risk factors are strong, under certain regularity conditions, there exists a fixed T_0 such that for all $T > T_0$, then

$$p \lim_{n \rightarrow \infty} \left(\tilde{\lambda}_{nT} \right) = \lambda_T^* = \lambda_0 + (\hat{\mu}_T - \mu_0). \quad (14)$$

Shanken refers to λ_T^* as "ex-post" risk premia to be distinguished from λ_0 , referred to as "ex ante" risk premia. See also Section 3.7 of Jagannathan, Skoulakis, and Wang (2010).

In this paper we exploit Shanken's bias correction procedure by applying it to $\phi_0 = \lambda_0 - \mu_0$ which we identify directly in terms of the observables. To this end, taking expectations of both sides of equation (1) and using the resultant expression for $E(r_{it})$ in (4), we have $E(r_{it}) = c + \beta'_i \lambda_0 + \eta_i = a_i + \beta'_i \mu_0$, which in turn yields

$$a_i = c + \beta'_i \phi_0 + \eta_i, \text{ for } i = 1, 2, \dots, n. \quad (15)$$

Therefore, $\phi_0 = \lambda_0 - \mu_0$ can be identified from the regression of a_i on β_i for $i = 1, 2, \dots, n$, so long as the idiosyncratic pricing errors, η_i , are sufficiently weak relative to the strengths of the risk factors in a sense which will be made precise below. Under market efficiency we must have $\beta'_i \phi_0 = 0$, for all i with η_i , $i = 1, 2, \dots, n$ sufficiently bounded such that the APT condition given by (5) is met. Accordingly, we refer to

¹See also Shanken and Zhou (2007), Kan, Robotti, and Shanken (2013), and Bai and Zhou (2015), and the survey paper by Jagannathan, Skoulakis, and Wang (2010) for further references.

$\{\beta_i' \phi_0, i = 1, 2, \dots, n\}$ as spanning errors, since they are spanned by the factor loadings, and refer to $\{\eta_i, i = 1, 2, \dots, n\}$ as the idiosyncratic pricing errors. Most tests of alpha developed in the literature do not distinguish between these two types of pricing errors and focus on testing $\{a_i = 0, i = 1, 2, \dots, n\}$, see the discussion in the introduction. Estimation of c and ϕ_0 and inference about their statistical significance complement the tests of alpha ($a_i = 0$) that do not distinguish between spanning and idiosyncratic pricing errors. Further, tests on c and the individual elements of ϕ_0 also help with the interpretation of the test outcomes by identifying particular risk factors that might be behind the non-zero spanning errors.

2.2 Estimation of ϕ_0

In view of (15), the estimation of ϕ_0 can be carried out following a two-step procedure whereby in the first step a_i and β_i are estimated from the least squares regressions of r_{it} on an intercept and \mathbf{f}_t , and these are then used in a second step regression to estimate ϕ_0 . Specifically,

$$\hat{\phi}_{nT} = \left(\hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{B}}_{nT} \right)^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{a}}_{nT}, \quad (16)$$

where $\hat{\mathbf{a}}_{nT} = (\hat{a}_{1T}, \hat{a}_{1T}, \dots, \hat{a}_{nT})' = \bar{\mathbf{r}}_{nT} - \hat{\mathbf{B}}_{nT} \hat{\boldsymbol{\mu}}_T$, and as before $\hat{\mathbf{B}}_{nT} = (\hat{\beta}_{1T}, \hat{\beta}_{2T}, \dots, \hat{\beta}_{nT})'$. This estimator is consistent for ϕ_0 so long as n and $T \rightarrow \infty$, and bias-corrections are necessary to ensure the large n consistency of the estimator when T is fixed. A Shanken type bias-corrected estimator of ϕ_0 is given by

$$\tilde{\phi}_{nT} = \mathbf{H}_{nT}^{-1} \left[\frac{\hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{a}}_{nT}}{n} + T^{-1} \hat{\sigma}_{nT}^2 \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1} \hat{\boldsymbol{\mu}}_T \right], \quad (17)$$

where \mathbf{H}_{nT} and $\hat{\sigma}_{nT}^2$ are given by (13) and (10), respectively. It is also easily established that

$$\tilde{\phi}_{nT} = \tilde{\boldsymbol{\lambda}}_{nT} - \hat{\boldsymbol{\mu}}_T, \quad (18)$$

and for a fixed T and as $n \rightarrow \infty$, we have

$$p \lim_{n \rightarrow \infty} \tilde{\phi}_{nT} = p \lim_{n \rightarrow \infty} \tilde{\boldsymbol{\lambda}}_{nT} - \hat{\boldsymbol{\mu}}_T.$$

Hence, upon using (14)

$$p \lim_{n \rightarrow \infty} \tilde{\phi}_{nT} = \boldsymbol{\lambda}_0 + (\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0) - \hat{\boldsymbol{\mu}}_T = \boldsymbol{\lambda}_0 - \boldsymbol{\mu}_0 = \phi_0, \quad (19)$$

and there exists a fixed T_0 such that for all $T > T_0$, $\tilde{\phi}_{nT}$ converges to ϕ_0 as $n \rightarrow \infty$. Also using (14) and (18), and noting that $\lambda_0 - \mu_0 = \phi_0$, interestingly we have

$$\tilde{\lambda}_{nT} - \lambda_T^* = \tilde{\phi}_{nT} + \hat{\mu}_T - \lambda_T^* = \tilde{\phi}_{nT} - \phi_0.$$

So inference using the Shanken bias-corrected estimator of λ around λ_T^* , is the same as making inference using $\tilde{\phi}_{nT}$ around ϕ_0 .

The asymptotic distribution of $\tilde{\phi}_{nT}$ depends on both n and T . Assuming the observed factors are strong and under certain regularity conditions, to be introduced below, we have

$$\sqrt{nT} \left(\tilde{\phi}_{nT} - \phi_0 \right) \rightarrow_d N \left(\mathbf{0}, \Sigma_{\beta\beta}^{-1} \mathbf{V}_\xi \Sigma_{\beta\beta}^{-1} \right), \quad (20)$$

where

$$\mathbf{V}_\xi = \left(1 + \lambda_0' \Sigma_f^{-1} \lambda_0 \right) p \lim_{n \rightarrow \infty} \left[n^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{V}_u \mathbf{M}_n \mathbf{B}_n \right].$$

The variance of $\tilde{\phi}_{nT}$ is consistently estimated by

$$\widehat{Var} \left(\tilde{\phi}_{nT} \right) = T^{-1} n^{-1} \mathbf{H}_{nT}^{-1} \hat{\mathbf{V}}_{\xi, nT} \mathbf{H}_{nT}^{-1}, \quad (21)$$

where \mathbf{H}_{nT} is given by (13),

$$\hat{\mathbf{V}}_{\xi, nT} = \left(1 + \hat{s}_{nT} \right) \left(n^{-1} \hat{\mathbf{B}}'_n \mathbf{M}_n \tilde{\mathbf{V}}_u \mathbf{M}_n \hat{\mathbf{B}}_n \right), \quad (22)$$

and

$$\hat{s}_{nT} = \tilde{\lambda}'_{nT} \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1} \tilde{\lambda}_{nT}, \quad (23)$$

$\tilde{\lambda}_{nT}$ is defined by (12), and $\tilde{\mathbf{V}}_u$ is a suitable estimator of $\mathbf{V}_u = E(\mathbf{u}_{ot} \mathbf{u}'_{ot})$. How to estimate \mathbf{V}_u and the conditions under which $\widehat{Var} \left(\tilde{\phi}_{nT} \right)$ is consistently estimated by $\widehat{Var} \left(\tilde{\phi}_{nT} \right)$ is discussed in sub-section 3.2.

Note that even when $\mathbf{V}_u = \sigma^2 \mathbf{I}_T$ the variance of $\tilde{\phi}_{nT}$ does not reduce to $\sigma^2 \Sigma_{\beta\beta}^{-1}$, the standard least squares formula used for the case of known factor loadings. When the loadings are estimated the scaling term $\left(1 + \lambda_0' \Sigma_f^{-1} \lambda_0 \right)$ is required and its neglect can lead to serious over-rejection even if $n/T \rightarrow 0$ as n and $T \rightarrow \infty$.

2.3 Factor strength

In this paper we deviate from the standard literature and allow the observed and latent factors to have different degrees of strength, depending on how pervasively they impact the security returns. Bailey, Kapetanios, and Pesaran (2021) define the strength of factor, f_{kt} , in terms of the number of its non-zero factor loadings. For a factor to be

strong almost all of its n loadings must differ from zero. Given our focus on estimation of risk premia, we adopt the following definition which directly relates to the covariance of β_i . See also Chudik, Pesaran, and Tosetti (2011).

Definition 1 (*Factor strengths*) *The strength of factor f_{kt} is measured by its degree of pervasiveness as defined by the exponent α_k in*

$$\sum_{i=1}^n (\beta_{ik} - \bar{\beta}_k)^2 = \Theta(n^{\alpha_k}), \quad (24)$$

and $0 < \alpha_k \leq 1$. We refer to $\{\alpha_k, k = 1, 2, \dots, K\}$ as factor strengths. Factor f_{kt} is said to be strong if $\alpha_k = 1$, semi-strong if $1 > \alpha_k > 1/2$, and weak if $0 \leq \alpha_k \leq 1/2$. Condition (24) applies irrespective of whether the loadings, β_{ik} , are viewed as deterministic or stochastic.

In the above definition $\Theta_p(n^{\alpha_k})$ denotes the rate at which additional securities add to the factor's strength and α_k can be viewed as a logarithmic expansion rate in terms of n and relates to the proportion of non-zero factor loadings. In the literature it is commonly assumed that the covariance matrix of factor loadings defined by

$$\Sigma_{\beta\beta} = p \lim_{n \rightarrow \infty} \left[n^{-1} \sum_{i=1}^n (\beta_i - \bar{\beta}_n) (\beta_i - \bar{\beta}_n)' \right], \quad (25)$$

is positive definite, where $\bar{\beta}_n = n^{-1} \sum_{i=1}^n \beta_i = (\bar{\beta}_1, \bar{\beta}_2, \dots, \bar{\beta}_K)'$. For $\Sigma_{\beta\beta}$ to be positive definite matrix it is *necessary* that all the K risk factors under consideration are strong in the sense that

$$p \lim_{n \rightarrow \infty} \left[n^{-1} \sum_{i=1}^n (\beta_{ik} - \bar{\beta}_k)^2 \right] > 0, \text{ for } k = 1, 2, \dots, K. \quad (26)$$

In terms of our definition of factor strength, $\Sigma_{\beta\beta}$ will be positive definite if all the observed factors are strong, namely if $\alpha_k = 1$ for $k = 1, 2, \dots, K$. However, such an assumption is quite restrictive and is unlikely to be satisfied for many risk factors being considered in the literature. Bailey, Kapetanios, and Pesaran (2021) show that, apart from the market factor, only a handful of 144 factors in the literature considered by Feng, Giglio, and Xiu (2020) come close to being strong.

2.4 Missing factor

We now turn to the structure of the errors, u_{it} , in the returns equations, and consider two possible sources of error cross-sectional dependence: a missing or latent factor and

production networks. The issue of missing factors has been investigated in the recent literature by Giglio and Xiu (2021) and Anatolyev and Mikusheva (2022). The issue of production networks has been investigated in the recent literature by Herskovic (2018), who derives two risk factors based on the changes in network concentration and network sparsity, and Gofman, Segal, and Wu (2020), who focus on the vertical dimension of production by modeling a supply chain, in terms of supplier-customer links. They find that the further away a firm is from final consumers the higher its return. They use this to create a factor TMB (top minus bottom). Both sources of cross-sectional error dependence could be important, since network dependence cannot be represented using latent factor models. See Section 3 of Chudik, Pesaran, and Tosetti (2011).

To allow for both forms of error cross-sectional dependence we consider the following decomposition of u_{it}

$$u_{it} = \gamma_i g_t + v_{it}, \quad (27)$$

where g_t is the missing (latent) factor and v_{it} is weakly cross-correlated in the sense of approximate factor models due to Chamberlain (1983) and Chamberlain and Rothschild (1983). Here we allow for a single missing factor to simplify the exposition, but note that increasing the number of missing factors has little impact on our analysis, so long as the number of missing factors is fixed. Using the normalization $E(g_t^2) = 1$, and assuming that $\gamma_i g_t$ and v_{it} are independently distributed then $E(u_{it}u_{jt}) = \sigma_{ij} = \gamma_i \gamma_j + \sigma_{v,ij}$, and as shown in Lemma A.1, $n^{-1} \sum_{i=1}^n \sum_{j=1}^n |\sigma_{ij}| = O(1)$ so long as the strength of g_t , $\alpha_\gamma < 1/2$ and $\lambda_{\max}(\mathbf{V}_v) < \infty$. This is despite the fact that $\lambda_{\max}(\mathbf{V}_u) = O(n^{\alpha_\gamma})$, where $\mathbf{V}_v = E(\mathbf{v}_i \mathbf{v}_i')$ and $\mathbf{V}_u = E(\mathbf{u}_i \mathbf{u}_i')$.² In the Monte Carlo experiments, we consider the possibility of missing factors, as well as weak spatial and network cross-dependence that satisfy conditions of approximate factor models.

2.5 Pricing errors

The APT condition (5), given by (18) in Theorem II of Ross (1976), ensures that under APT the (idiosyncratic) pricing errors are sparse. In this paper we relax the Ross's

²Note that Chamberlain's approximate factor model specification requires $\lambda_{\max}(\mathbf{V}_u) = O(1)$ and is violated if $\alpha_\gamma > 0$.

condition to

$$\sum_{i=1}^n \eta_i^2 = O(n^{\alpha_\eta}), \quad (28)$$

where the exponent α_η measures the degrees of pervasiveness of pricing errors. Deviations from APT are measured in terms of α_η ($0 \leq \alpha_\eta < 1$). We investigate the robustness of our proposed estimator of ϕ_0 to α_η . This extension is important for tests of market efficiency where the null of interest is $H_0 : a_i = c$ for all i in (15). We note that under the alternative hypothesis $H_1 : a_i = c + \beta_i' \phi + \eta_i$, therefore it is desirable to develop a test of $\phi_0 = \mathbf{0}$ which is robust to a wider class of pricing errors than those entertained originally by Ross, where $\alpha_\eta = 0$. As noted above, market efficiency requires $\phi_0 = \mathbf{0}$, but is compatible with a moderate degree of pricing errors characterized by α_η .

3 Assumptions and theorems

We make the following standard assumptions about \mathbf{f}_t , g_t , v_{it} , β_i , η_i , and γ_i (the drivers of asset returns):

Assumption 1 (*Observed common factors*) (a) The $K \times 1$ vector of observed risk factors, \mathbf{f}_t , follows the general linear process

$$\mathbf{f}_t = \boldsymbol{\mu} + \sum_{\ell=0}^{\infty} \boldsymbol{\Psi}_\ell \zeta_{t-\ell}, \quad (29)$$

where $\|\boldsymbol{\mu}\| < C$, $\zeta_t \sim IID(\mathbf{0}, \mathbf{I}_K)$, and $\boldsymbol{\Psi}_\ell$ are $K \times K$ exponentially decaying matrices such that $\|\boldsymbol{\Psi}_\ell\| < C\rho^\ell$ for some $C > 0$ and $0 < \rho < 1$. (b) The $T \times K$ data matrix $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_T)'$ is full column rank and there exists T_0 such that for all $T > T_0$, $\hat{\boldsymbol{\Sigma}}_f = T^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{F}$ is a positive definite matrix, $\lambda_{\max} [(T^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1}] < C$, $\hat{\boldsymbol{\Sigma}}_f \rightarrow_p \boldsymbol{\Sigma}_f = E(\mathbf{f}_t - \boldsymbol{\mu}_0)(\mathbf{f}_t - \boldsymbol{\mu}_0)' > \mathbf{0}$, where $\boldsymbol{\mu}_0$ is the true value of $\boldsymbol{\mu}$.

Assumption 2 (*Observed factor loadings*) (a) The factor loadings β_{ik} for $i = 1, 2, \dots, n$ and $k = 1, 2, \dots, K$ are stochastically bounded such that $\sup_{ik} E(\beta_{ik}^2) < C$,

$$\sum_{i=1}^n (\beta_{ik} - \bar{\beta}_k)^2 = \Theta_p(n^{\alpha_k}), \text{ for } k = 1, 2, \dots, K. \quad (30)$$

(b) The $n \times K$ matrix of factor loadings, $\mathbf{B}_n = (\boldsymbol{\beta}_{o1}, \boldsymbol{\beta}_{o2}, \dots, \boldsymbol{\beta}_{oK})$, where $\boldsymbol{\beta}_{ok} = (\beta_{1k}, \beta_{2k}, \dots, \beta_{nk})'$

satisfy

$$0 < c < \lambda_{\min}(\mathbf{D}_\alpha^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{B}_n \mathbf{D}_\alpha^{-1}) < \lambda_{\max}(\mathbf{D}_\alpha^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{B}_n \mathbf{D}_\alpha^{-1}) < C < \infty, \quad (31)$$

for some small and large positive constants, c and C , where $\mathbf{M}_n = \mathbf{I}_n - n^{-1} \boldsymbol{\tau}_n \boldsymbol{\tau}'_n$, $\boldsymbol{\tau}_n = (1, 1, \dots, 1)'$, and \mathbf{D}_α is the $n \times n$ diagonal matrix

$$\mathbf{D}_\alpha = \text{Diag}(n^{\alpha_1/2}, n^{\alpha_2/2}, \dots, n^{\alpha_K/2}). \quad (32)$$

Assumption 3 (latent factor) (a) The latent factor, g_t , in (27) is distributed independently of $\mathbf{f}_{t'}$, for all t and t' , g_t is serially independent with mean zero, $E(g_t) = 0$, $E(g_t^2) = 1$, and a finite fourth order moment, $\sup_t E(g_t^4) < C$. (b) The loadings γ_i are such that $\sup_i |\gamma_i| < C$ and

$$\sum_{i=1}^n |\gamma_i| = O(n^{\alpha_\gamma}). \quad (33)$$

Assumption 4 (idiosyncratic errors) (a) The errors $\{v_{it}, i = 1, 2, \dots, n; t = 1, 2, \dots, T\}$ are distributed independently of the factors $f_{k,t}$, and g_t , for all i, t, t' and $k = 1, 2, \dots, K$, and their associated loadings β_{ik} , and γ_i . They are serially independent with $E(v_{it}) = 0$ and finite fourth order moments $E(v_{it}^4) < \infty$, and covariances $E(v_{it} v_{jt}) = \sigma_{v,ij}$, such that

$$\sup_i \sum_{j=1}^n |\sigma_{v,ij}| < \infty, \text{ and } \sup_i \sum_{j=1}^n \text{Cov}(v_{it}^2, v_{jt}^2) < \infty, \quad (34)$$

with $\lambda_{\min}(\mathbf{V}_v) > 0$, where $\mathbf{V}_v = (\sigma_{v,ij})$. (b) The degree of cross-sectional dependence of v_{it} is sufficiently weak so that

$$T^{-1/2} n^{-1/2} \sum_{t=1}^T \sum_{i=1}^n (\beta_{ik} - \bar{\beta}_k) v_{it} \rightarrow_d N(0, \omega_k^2), \text{ for } k = 1, 2, \dots, K, \quad (35)$$

where

$$\omega_k^2 = p \lim_{n \rightarrow \infty} n^{-\alpha_k} \sum_{i=1}^n \sum_{j=1}^n (\beta_{ik} - \bar{\beta}_k)(\beta_{jk} - \bar{\beta}_k) \sigma_{v,ij}. \quad (36)$$

Assumption 5 (Pricing errors) The pricing errors, η_i , for $i = 1, 2, \dots, n$ are individually bounded, $\sup_j |\eta_j| < C$, and are distributed independently of the factor loadings, β_{jk} , and γ_j for all i, j and $k = 1, 2, \dots, K$, as well as satisfying the condition

$$\sum_{i=1}^n |\eta_i| = O(n^{\alpha_\eta}), \quad (37)$$

with $\alpha_\eta < 1/2$.

Remark 1 Under Assumption 1 $E(\mathbf{f}_t) = \mu$, and $\text{Var}(\mathbf{f}_t) = \Sigma_f = \sum_{\ell=0}^{\infty} \Psi_{\ell} \Psi_{\ell}'$. Also since $\|\Sigma_f\| \leq \sum_{\ell=0}^{\infty} \|\Psi_{\ell}\|^2$ it then follows from part (a) of Assumption 1 that $\|\Sigma_f\| < C$.

Remark 2 Under Assumption 2

$$\mathbf{D}_{\alpha}^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{B}_n \mathbf{D}_{\alpha}^{-1} \rightarrow_p \Sigma_{\beta\beta}(\alpha) > 0, \quad (38)$$

where $\Sigma_{\beta\beta}(\alpha)$ is a $k \times k$ symmetric positive definite matrix which is a function of $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_K)'$. This follows from (31) since for any non-zero $n \times 1$ vector \mathbf{c} ,

$$\mathbf{c}' \mathbf{D}_{\alpha}^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{B}_n \mathbf{D}_{\alpha}^{-1} \mathbf{c} \geq (\mathbf{c}' \mathbf{c}) \lambda_{\min}(\mathbf{D}_{\alpha}^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{B}_n \mathbf{D}_{\alpha}^{-1}) > 0.$$

In the standard case where the factors are all strong ($\alpha_k = 1$ for all k), the above limit reduces to $n^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{B}_n \rightarrow_p \Sigma_{\beta\beta}(\tau_K) = \Sigma_{\beta\beta} > 0$.

Remark 3 The high level condition (35) in Assumption 4 is required for establishing the asymptotic normality of the estimator of ϕ_0 , and is clearly met when v_{it} and/or β_{ik} are independently distributed. It is also possible to establish (35) under weaker conditions assuming that v_{it} and/or β_{ik} satisfy some time-series type mixing conditions applied to cross section.

Remark 4 The exponent parameter, α_{η} , of the pricing condition in (37), can be viewed as the degree to which pricing errors are pervasive in large economies (as $n \rightarrow \infty$).

Letting $\boldsymbol{\eta}_n = (\eta_1, \eta_2, \dots, \eta_n)'$ we have

$$\sum_{i=1}^n \eta_i^2 = \|\boldsymbol{\eta}_n\|^2 \leq \|\boldsymbol{\eta}_n\|_{\infty} \|\boldsymbol{\eta}_n\|_1 = \sup_j |\eta_j| \left(\sum_{i=1}^n |\eta_i| \right), \quad (39)$$

and under Assumption (5) it also follows that

$$\sum_{i=1}^n \eta_i^2 = O(n^{\alpha_{\eta}}). \quad (40)$$

Similarly

$$\sum_{i=1}^n \gamma_i^2 = O(n^{\alpha_{\gamma}}). \quad (41)$$

Remark 5 Whilst (37) implies (40), the reverse does not follow. By allowing for $\alpha_{\eta} > 0$ we are relaxing the Ross's boundedness condition that requires setting $\alpha_{\eta} = 0$.

Remark 6 The assumption that the observed and missing factors, \mathbf{f}_t and g_t , are distributed independently is not restrictive and can be relaxed. For example, suppose that

$$g_t = \mu_g + \boldsymbol{\theta}' \mathbf{f}_t + v_{gt},$$

where \mathbf{f}_t and v_{gt} are independently distributed. Then using (27) we have

$$u_{it} = \gamma_i \mu_g + \gamma_i (\boldsymbol{\theta}' \mathbf{f}_t) + \gamma_i v_{gt} + v_{it},$$

and the return equation (1) can be written as

$$r_{it} = (a_i + \gamma_i \mu_g) + (\boldsymbol{\beta}_i + \gamma_i \boldsymbol{\theta})' \mathbf{f}_t + \gamma_i v_{gt} + v_{it},$$

with v_{gt} now acting as the missing common factor, which, by construction, is distributed independently of \mathbf{f}_t .

Remark 7 Assumptions 3 and 4 ensure that u_{it} have zero means and are serially independent, but allow u_{it} to be cross-sectionally weakly correlated. This assumption is not strong for asset pricing models, since realized returns are only mildly serially correlated and that can be captured by the serial correlation in the observed factors.

As we shall see, to estimate and conduct inference on the risk premia associated with the observed factors, f_{kt} , we require $\alpha_k > \alpha_\gamma < 1/2$, where α_γ denotes the strength of the latent factor, g_t , and similarly defined by $\sum_{i=1}^n \gamma_i^2 = \Theta(n^{\alpha_\gamma})$. Namely, the latent factor must be sufficiently weak so that ignoring it will be inconsequential, and observed factors sufficiently strong so that they can be distinguished from the weak latent factor.

The main theoretical results of the paper are set out around five theorems. Theorem 1 considers the Fama-MacBeth two-step estimator and derives its limiting property as n and $T \rightarrow \infty$. To eliminate the bias of Fama-MacBeth estimator we require $n/T \rightarrow 0$, and to eliminate the effects of pricing errors we need $Tn^{\alpha_\gamma}/n \rightarrow 0$, which results in a contradiction. Thus the Fama-MacBeth estimator is valid only when there are no pricing errors ($\eta_i = 0$ for all i) and when $n/T \rightarrow 0$. Theorem 2 provides a proof that the estimator of $\bar{\sigma}_n^2$ (denoted by $\hat{\sigma}_{nT}^2$) proposed by Shanken (1992) continues to be unbiased for a fixed T as $n \rightarrow \infty$, even under the general setting of the current paper that allows for missing factors as well as pricing errors. Theorem 2 also establishes that $\hat{\sigma}_{nT}^2 - \bar{\sigma}_n^2 \rightarrow O_p(n^{-1/2}T^{-1/2})$, which is essential for establishing the results for the bias-corrected estimator of ϕ_0 , namely $\tilde{\phi}_{nT}$ given by (17), summarized in Theorem 3. This theorem provides conditions under which $\tilde{\phi}_{nT}$ is a consistent estimator of ϕ_0 , and derives its asymptotic distribution assuming the observed factors are strong, again allowing for pricing errors, a missing factor, and other forms of weak error cross-sectional dependence. Theorem 4 extends the results of Theorem 3 to the case where one or more of the observed risk factors are semi-strong and shows how factor strength

impacts the precision with which the elements of ϕ_0 are estimated. Finally, Theorem 5 presents the conditions under which the asymptotic variance of $\tilde{\phi}_{nT}$ can be consistently estimated.

Theorem 1 (*Small T bias of Fama-MacBeth estimator of λ*) Consider the multi-factor linear return model (3) with the missing factor g_t in u_{it} as defined by (27) and the associated risk premia, λ , defined by (4). Suppose that Assumptions 1, 2, 4, 3 and 5 hold and all observed factors are strong. Suppose further that the true value of the risk premia, λ_0 , is estimated by Fama-MacBeth two-pass estimator, $\hat{\lambda}_{nT}$, defined by (7). Then for any fixed $T > T_0$ such that $\lambda_{\min}(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}) > 0$, we have (as $n \rightarrow \infty$)

$$\hat{\lambda}_{nT} - \lambda_0 = (\hat{\mu}_T - \mu_0) - \frac{\bar{\sigma}^2}{T} \left[\Sigma_{\beta\beta} + \bar{\sigma}^2 \frac{1}{T} \left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T} \right)^{-1} \right]^{-1} \left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T} \right)^{-1} \lambda_T^* + o_p(1), \quad (42)$$

where $\hat{\mu}_T = T^{-1} \sum_{t=1}^T \mathbf{f}_t$,

$$\Sigma_{\beta\beta} = \lim_{n \rightarrow \infty} \left(\frac{\mathbf{B}'_n \mathbf{M}_n \mathbf{B}_n}{n} \right), \text{ and } \bar{\sigma}^2 = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \sigma_i^2 > 0.$$

The proof is provided in Section B.1 of the Appendix.

To derive the asymptotic distribution of $\hat{\lambda}_{nT} - \lambda_0$ it is required that both n and $T \rightarrow \infty$, jointly. Also, noting that

$$\hat{\lambda}_{nT} - \lambda_0 = (\hat{\mu}_T - \mu_0) + \left(\hat{\phi}_{nT} - \phi_0 \right),$$

it is clear that increasing n is not relevant for the distribution of $\hat{\mu}_T - \mu_0$, but joint n and T asymptotics are required when investigating the distribution of $\hat{\phi}_{nT} - \phi_0$. Focussing on the latter, and using result (B.12) in the Appendix, we have

$$\left(n^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{B}}_{nT} \right) \sqrt{nT} \left(\hat{\phi}_{nT} - \phi_0 \right) = n^{-1/2} T^{1/2} \mathbf{B}'_n \mathbf{M}_n \boldsymbol{\eta}_n + n^{-1/2} T^{1/2} \mathbf{G}'_T \mathbf{U}'_{nT} \mathbf{M}_n \boldsymbol{\eta}_n + n^{-1/2} T^{1/2} \mathbf{G}'_T \mathbf{U}'_{nT} \mathbf{M}_n \bar{\mathbf{u}}_{n\circ} - n^{-1/2} T^{1/2} \mathbf{G}'_T \mathbf{U}'_{nT} \mathbf{U}_{nT} \mathbf{G}_T \lambda_T^*.$$

Where $\mathbf{U}_{nT} = (\mathbf{u}_{1\circ}, \mathbf{u}_{2\circ}, \dots, \mathbf{u}_{n\circ})'$, $\mathbf{u}_{i\circ} = (u_{i1}, u_{i2}, \dots, u_{iT})'$, $\mathbf{G}_T = \mathbf{M}_T \mathbf{F} (\mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1}$, $\bar{\mathbf{u}}_{n\circ} = (\bar{u}_{1\circ}, \bar{u}_{2\circ}, \dots, \bar{u}_{n\circ})'$, and $\bar{u}_{i\circ} = T^{-1} \sum_{t=1}^T u_{it}$. Consider first the terms that include the pricing errors, $\boldsymbol{\eta}_n$, and using the results in Lemma A.2 note that $n^{-1/2} T^{1/2} \mathbf{B}'_n \mathbf{M}_n \boldsymbol{\eta}_n = O_p(T^{1/2} n^{-1/2 + \alpha_\eta})$, $n^{-1/2} T^{1/2} \mathbf{G}'_T \mathbf{U}'_{nT} \mathbf{M}_n \boldsymbol{\eta}_n = O_p\left(n^{-1/2 + \frac{\alpha_\eta + \alpha_\gamma}{2}}\right)$.

It is clear that the effects of pricing errors on the distribution of $\hat{\phi}_{nT}$ vanish only if $T^{1/2}n^{-1/2+\alpha_\eta} \rightarrow 0$, and $\alpha_\eta + \alpha_\gamma < 1$. Also

$$\begin{aligned} n^{-1/2}T^{1/2}\mathbf{G}'_T\mathbf{U}'_{nT}\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} &= O_p(T^{-1/2}), \\ n^{-1/2}T^{1/2}\mathbf{G}'_T\mathbf{U}'_{nT}\mathbf{U}_{nT}\mathbf{G}_T\boldsymbol{\lambda}_T^* &= \sqrt{\frac{n}{T}}\bar{\sigma}_n^2\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}_T^* + O_p(T^{-1/2}). \end{aligned}$$

Finally, for the first two terms involving \mathbf{B}_n and \mathbf{U}_{nT} we have

$$n^{-1/2}T^{1/2}\mathbf{B}'_n\mathbf{M}_n(\bar{\mathbf{u}}_{n\circ} - \mathbf{U}_{nT}\mathbf{G}_T\boldsymbol{\lambda}_T^*) = O_p(1). \quad (43)$$

It is clear that the small T bias of the asymptotic distribution of the two-step estimator, given by $\sqrt{\frac{n}{T}}\bar{\sigma}_n^2\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}_T^*$, does not vanish unless, $n/T \rightarrow 0$. At the same time for the pricing errors to have no impact on the distribution of the two-step estimator we must have $Tn^{\alpha_\eta}/n \rightarrow 0$. Both conditions cannot be met simultaneously. It is possible to derive the asymptotic distribution of $\hat{\phi}_{nT}$, and hence that of $\hat{\boldsymbol{\lambda}}_{nT}$, when $n/T \rightarrow 0$ and $\boldsymbol{\eta} = \mathbf{0}$, but these are quite restrictive conditions, and to avoid them we follow Shanken (1992) and instead consider a bias-corrected version of $\hat{\phi}_{nT}$, namely $\tilde{\phi}_{nT}$ given by (17). As noted earlier $\tilde{\phi}_{nT} = \tilde{\boldsymbol{\lambda}}_{nT} - \hat{\boldsymbol{\mu}}_T$, where $\tilde{\boldsymbol{\lambda}}_{nT}$ is the bias-corrected version of $\hat{\boldsymbol{\lambda}}_{nT}$ originally proposed by Shanken.

To investigate the asymptotic properties of $\tilde{\phi}_{nT}$ we first need to establish conditions under which $\hat{\sigma}_{nT}^2$, defined by (10), is a consistent estimator of $\bar{\sigma}_n^2 = n^{-1}\sum_{i=1}^n\sigma_i^2$, which enters the bias-corrected estimator. The proof of consistency in the literature does not allow for missing factors or pricing errors and only considers the case where T is fixed as $n \rightarrow \infty$. For derivation of asymptotic distribution of $\tilde{\phi}_{nT}$ we also need to consider the limiting properties of $\hat{\sigma}_{nT}^2$ under joint n and T asymptotics. The following theorem provides the required results for $\hat{\sigma}_{nT}^2$ as an estimator of $\bar{\sigma}_n^2$.

Theorem 2 Consider $\hat{\sigma}_{nT}^2$, the estimator of $\bar{\sigma}_n^2$ given by (see (10)),

$$\hat{\sigma}_{nT}^2 = \frac{\sum_{t=1}^T \sum_{i=1}^n \hat{u}_{it}^2}{n(T-K-1)}, \quad (44)$$

and suppose that Assumptions 1, 3, and 4, are satisfied. Then for a fixed T

$$\lim_{n \rightarrow \infty} E\left(\hat{\sigma}_{nT}^2\right) = \bar{\sigma}^2, \quad (45)$$

where $\bar{\sigma}^2 = \lim_{n \rightarrow \infty} \bar{\sigma}_n^2$, and $\bar{\sigma}_n^2 = n^{-1}\sum_{i=1}^n\sigma_i^2$. Furthermore

$$\hat{\sigma}_{nT}^2 - \bar{\sigma}_n^2 = O_p(T^{-1/2}n^{-1/2}). \quad (46)$$

For a proof see sub-section B.2 in the Appendix.

Result (46) shows that $\hat{\sigma}_{nT}^2$ continues to be a consistent estimator of $\bar{\sigma}^2 = \lim_{n \rightarrow \infty} \bar{\sigma}_n^2$ for a fixed T as $n \rightarrow \infty$, even in the presence of pricing errors and a missing common

factor. This result also holds when one or more of the factors are semi-strong.

Equipped with the above result we are now in a position to present the theorem that sets out the asymptotic distribution of $\tilde{\phi}_{nT}$.

Theorem 3 Consider, $\tilde{\phi}_{nT}$, the bias-corrected estimators of ϕ_0 given by (17). Suppose Assumptions 1, 2, 4, 3 and 5 hold, all the observed factors are strong, ($\alpha_k = 1$, for $k = 1, 2, \dots, K$), and the strength of the missing factor, α_γ defined by (33), satisfies $\alpha_\gamma < 1/2$.

$$\begin{aligned} \tilde{\phi}_{nT} - \phi_0 = & O_p \left(T^{-1/2} n^{-1/2} \right) + O_p \left(T^{-1/2} n^{-1 + \frac{\alpha_\gamma + \alpha_\eta}{2}} \right) \\ & + O_p \left(n^{-1 + \alpha_\eta} \right) + O_p \left(T^{-1} n^{-1/2} \right), \end{aligned} \quad (47)$$

where α_η denotes the degree of pervasiveness of the pricing errors defined by (28). (a)

When T is fixed, $\alpha_\gamma < 1/2$ and $\alpha_\eta < 1$, then there exists T_0 such that for all $T > T_0$

$$p \lim_{n \rightarrow \infty} \left(\tilde{\phi}_{nT} \right) = \phi_0. \quad (48)$$

Also

$$\sqrt{nT} \left(\tilde{\phi}_{nT} - \phi_0 \right) = \Sigma_{\beta\beta}^{-1} \boldsymbol{\xi}_{nT} + O_p \left(n^{-\frac{1}{2} + \frac{\alpha_\eta + \alpha_\gamma}{2}} \right) + O_p \left(T^{1/2} n^{-1/2 + \alpha_\eta} \right) + O_p \left(T^{-1/2} \right), \quad (49)$$

where $\Sigma_{\beta\beta} = p \lim_{n \rightarrow \infty} \left(n^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{B}_n \right)$,

$$\boldsymbol{\xi}_{nT} = n^{-1/2} T^{-1/2} \mathbf{B}'_n \mathbf{M}_n \mathbf{U}_{nT} \mathbf{a}_T, \quad (50)$$

and $\mathbf{a}_T = \tau_T - \mathbf{M}_T \mathbf{F} (T^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1} \boldsymbol{\lambda}_T^*$. (b) If $\alpha_\gamma < 1/2$, $\alpha_\eta < 1/2$, and $\sqrt{\frac{T}{n}} n^{\alpha_\eta} \rightarrow 0$, as n and $T \rightarrow \infty$ jointly, then

$$\sqrt{nT} \left(\tilde{\phi}_{nT} - \phi_0 \right) \rightarrow_d N \left(\mathbf{0}, \Sigma_{\beta\beta}^{-1} \mathbf{V}_\xi \Sigma_{\beta\beta}^{-1} \right), \quad (51)$$

where

$$\mathbf{V}_\xi = \left(1 + \boldsymbol{\lambda}'_0 \Sigma_f^{-1} \boldsymbol{\lambda}_0 \right) p \lim_{n \rightarrow \infty} \left(n^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{V}_u \mathbf{M}_n \mathbf{B}_n \right). \quad (52)$$

For a proof see sub-section B.3 in the Appendix.

Result (47) establishes the finite T consistency of $\tilde{\phi}_{nT}$ for ϕ_0 so long as $\alpha_\gamma < 1/2$ and $\alpha_\eta < 1$, thus extending the Shanken result to a much more general setting. To the best of our knowledge the asymptotic distribution in (51) is new and shows that the asymptotic covariance matrix of $\tilde{\phi}_{nT}$ includes the term $\boldsymbol{\lambda}_T^{*'} (T^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1} \boldsymbol{\lambda}_T^*$, that arises from the first stage estimation of the factor loadings, and must be included in the analysis for valid inference. It is also clear that this additional term does not vanish with $T \rightarrow \infty$, and tends to $\boldsymbol{\lambda}'_0 \Sigma_f^{-1} \boldsymbol{\lambda}_0 \geq (\boldsymbol{\lambda}'_0 \boldsymbol{\lambda}_0) \lambda_{\max}(\Sigma_f^{-1}) = (\boldsymbol{\lambda}'_0 \boldsymbol{\lambda}_0) \lambda_{\min}(\Sigma_f) > 0$, which is strictly non-zero unless $\boldsymbol{\lambda}_0 = \mathbf{0}$. Shanken type bias correction addresses the mean of the asymptotic distribution of $\tilde{\phi}_{nT}$, but not its covariance.

The $O_p(T^{-1/2})$ term in (49) arises from the sampling errors involved in the estimation of the factor loadings and $\bar{\sigma}_n^2$, and tends to zero at the regular \sqrt{T} rate. But n has to be sufficiently large to eliminate the effects of pricing errors on identification of ϕ_0 , as dictated by condition $\sqrt{\frac{T}{n}}n^{\alpha_n} \rightarrow 0$, as n and $T \rightarrow \infty$.³ The requirement that T need not be too large relative to n for estimation of ϕ_0 is consistent with separating the estimation of ϕ_0 from that of μ_0 , allowing the use a relatively small T and a large n to estimate ϕ_0 and a relatively large T when estimating μ_0 .

3.1 What if one or more of the risk factors are semi-strong?

We now turn to an intermediate case where one or more of the observed factors are semi-strong, in the sense that their factor strength, α_k lies between 1/2 and 1. The case of weak risk factors is already covered in the proceeding analysis, and such factors can be included in the error term, u_{it} , with little consequence for the estimation of risk premia of the remaining factors that are strong or semi-strong. Weak factors do not have any explanatory power and can be dropped from the analysis.

When one or more of the observed factors is semi-strong $\Sigma_{\beta\beta}$ is no longer positive definite and Theorem 3 does not apply, but it is possible to adapt the proofs to establish the limiting properties of $\tilde{\phi}_{k,nT}$ (the k^{th} element of $\tilde{\phi}_{nT}$) for different values of α_k .

To this end, analogously to $\tilde{\phi}_{nT}$, we introduce the following estimator of ϕ_0

$$\tilde{\phi}_{nT}(\alpha) = \mathbf{H}_{nT}^{-1}(\alpha) \left[\mathbf{D}_\alpha^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{a}}_{nT} + \frac{n \hat{\sigma}_{nT}^2}{T} \mathbf{D}_\alpha^{-1} \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1} \hat{\boldsymbol{\mu}} \right], \quad (53)$$

where

$$\mathbf{H}_{nT}(\alpha) = \mathbf{D}_\alpha^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{B}}_{nT} \mathbf{D}_\alpha^{-1} - \frac{n \hat{\sigma}_{nT}^2}{T} \mathbf{D}_\alpha^{-1} \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1} \mathbf{D}_\alpha^{-1}. \quad (54)$$

It is now easily seen that

$$\mathbf{D}_\alpha \left(\tilde{\phi}_{nT}(\alpha) - \phi_0 \right) = \mathbf{H}_{nT}^{-1}(\alpha) \mathbf{q}_{nT}(\alpha), \quad (55)$$

where

$$\mathbf{q}_{nT}(\alpha) = \mathbf{D}_\alpha^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{a}}_{nT} + \frac{n \hat{\sigma}_{nT}^2}{T} \mathbf{D}_\alpha^{-1} \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1} \hat{\boldsymbol{\mu}}_T - \mathbf{H}_{nT}(\alpha) \mathbf{D}_\alpha \phi_0. \quad (56)$$

\mathbf{D}_α is defined by (32), and $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_K)'$. It is easily established that numerically $\tilde{\phi}_{nT}(\alpha)$ is identical to $\tilde{\phi}_{nT}$, and its introduction is primarily for the purpose of

³The condition $\sqrt{\frac{T}{n}}n^{\alpha_n} \rightarrow 0$ can be weakened somewhat to $\sqrt{\frac{T}{n}}n^{\alpha_n/2} \rightarrow 0$ if we also assume that $\beta_{ik} - \bar{\beta}_k$ are independently distributed over i , but will still require n to be larger than T .

establishing the limiting properties of $\tilde{\phi}_{k,nT} - \phi_{0,k}$ that do depend on α_k . Note that

$$\mathbf{H}_{nT}(\boldsymbol{\alpha}) = n\mathbf{D}_\alpha^{-1}\mathbf{H}_{nT}\mathbf{D}_\alpha^{-1}, \text{ and } \mathbf{q}_{nT}(\boldsymbol{\alpha}) = n\mathbf{D}_\alpha^{-1}\mathbf{s}_{nT}$$

where \mathbf{s}_{nT} and \mathbf{H}_{nT} are already defined by (B.23) and (B.24). Using these in (55) we have

$$\mathbf{D}_\alpha \left(\tilde{\phi}_{nT}(\boldsymbol{\alpha}) - \phi_0 \right) = \left(n\mathbf{D}_\alpha^{-1}\mathbf{H}_{nT}\mathbf{D}_\alpha^{-1} \right)^{-1} n\mathbf{D}_\alpha^{-1}\mathbf{s}_{nT} = \mathbf{D}_\alpha\mathbf{H}_{nT}^{-1}\mathbf{s}_{nT},$$

and it follows that $\tilde{\phi}_{nT}(\boldsymbol{\alpha}) - \phi_0 = \mathbf{H}_{nT}^{-1}\mathbf{s}_{nT} = \tilde{\phi}_{nT}(\boldsymbol{\tau}_K) = \tilde{\phi}_{nT}$. See (B.22).

The convergence results for $\tilde{\phi}_{nT}(\boldsymbol{\alpha})$ are set out in the following theorem.

Theorem 4 Consider, $\tilde{\phi}_{nT}(\boldsymbol{\alpha})$, the bias-corrected estimators of ϕ_0 given by (53), and suppose Assumptions 1, 2, 4, 3 and 5 hold, the strength of observed factors, $\mathbf{f}_t = (f_{1t}, f_{2t}, \dots, f_{Kt})'$, is given by $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_K)'$, and the strength of the missing factor, g_t , defined by (33) is α_γ . Let $\alpha_{\min} = \min_k(\alpha_k)$ and suppose that $\alpha_\gamma < 1/2$. Then

$$\mathbf{H}_{nT}(\boldsymbol{\alpha}) = \mathbf{D}_\alpha^{-1}\mathbf{B}'_n\mathbf{M}_n\mathbf{B}_n\mathbf{D}_\alpha^{-1} + O_p(T^{-1}n^{-\alpha_{\min}+1/2}),$$

where $\mathbf{H}_{nT}(\boldsymbol{\alpha})$ is given by (54), and by part (b) of Assumption 2, $\mathbf{H}_{nT}(\boldsymbol{\alpha}) \rightarrow_p \boldsymbol{\Sigma}_{\beta\beta}(\boldsymbol{\alpha}) > 0$, for any fixed $T > T_0$ such that $\lambda_{\max} \left[\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T} \right)^{-1} \right] < C$ and $\alpha_{\min} > 1/2 > \alpha_\gamma$. Also

$$\begin{aligned} \tilde{\phi}_{k,nT}(\boldsymbol{\alpha}) - \phi_{0,k} &= O_p \left(n^{-(\alpha_k + \alpha_{\min})/2 + 1/2} T^{-1/2} \right) + O_p \left(n^{\frac{-(\alpha_k + \alpha_{\min}) + (\alpha_\eta + \alpha_\gamma)}{2}} T^{-1/2} \right) \\ &+ O_p \left(n^{-(\alpha_k + \alpha_{\min})/2 + \alpha_\eta} \right) + O_p \left(n^{-(\alpha_k + \alpha_{\min})/2 + 1/2} T^{-1} \right). \end{aligned} \quad (57)$$

See sub-section B.4 of the Appendix for a proof.

The result in (57) establishes the consistency of $\tilde{\phi}_{k,nT}(\boldsymbol{\alpha}) = \tilde{\phi}_{k,nT}$ even if f_{kt} is semi-strong so long as $n \rightarrow \infty$, and $\alpha_{\min} > 1/2$, $\alpha_k + \alpha_{\min} > \alpha_\eta + \alpha_\gamma$, and $\alpha_k + \alpha_{\min} > 2\alpha_\eta$. Clearly, these results reduce to the case of strong factors where $\alpha_{\min} = \alpha_k = 1$. Turning to the asymptotic distribution of $\tilde{\phi}_{k,nT}(\boldsymbol{\alpha})$, again only convergence rates are affected, and instead of the regular rate of \sqrt{nT} , we have $\sqrt{T}n^{(\alpha_k + \alpha_{\min} - 1)/2}$, and using (57) we have

$$\begin{aligned} &\sqrt{T}n^{(\alpha_k + \alpha_{\min} - 1)/2} \left(\tilde{\phi}_{k,nT}(\boldsymbol{\alpha}) - \phi_{0,k} \right) \\ &= O_p(1) + O_p \left(n^{-1/2 + \frac{(\alpha_\eta + \alpha_\gamma)}{2}} \right) + O_p \left(\sqrt{T}n^{-1/2 + \alpha_\eta} \right) + O_p(T^{-1/2}). \end{aligned} \quad (58)$$

The conditions needed for eliminating the effects of the pricing errors are the same as before and are given by $\alpha_\eta + \alpha_\gamma < 1$ and $\sqrt{T}n^{-1/2 + \alpha_\eta} \rightarrow 0$. The asymptotic distribution is unaffected except for the slower rate of convergence alluded to above. It is also of interest to note that adding semi-strong factors can adversely affect the convergence

rate of the strong factor with $\alpha_k = 1$. As an example suppose the asset pricing model contains two factors, one strong, $\alpha_1 = 1$ and one semi strong with $\alpha_2 < 1$. Then the convergence rate of $\tilde{\phi}_{1,nT}(\boldsymbol{\alpha}) - \phi_{0,1}$ is given by $\sqrt{T} n^{(1+\alpha_2-1)/2}$ which is slower than the rate we would have obtained for $\tilde{\phi}_{1,nT}(\boldsymbol{\alpha}) - \phi_{0,1}$ if both factors were strong ($\alpha_{min} = \alpha_k = 1$), namely the regular rate of \sqrt{nT} .

3.2 Consistent estimation of the variance of $\tilde{\phi}_{nT}$

To carry out inference on ϕ_0 , or any of its elements individually, we require a consistent estimator of $Var(\tilde{\phi}_{nT})$. Using (51) and (52) we first note that $\boldsymbol{\Sigma}_{\beta\beta}$ is consistently estimated by \mathbf{H}_{nT} given by (13). Therefore, it is sufficient to find a suitable estimator of $\mathbf{V}_u = (\sigma_{ij})$ such that \mathbf{V}_ξ given by (52) is consistently estimated. Under suitable sparsity restrictions \mathbf{V}_u can be consistently estimated using the various thresholding procedures advanced in the statistical literature by Bickel and Levina (2008a,b), Cai and Liu (2011), and Bailey, Pesaran, and Smith (2019, BPS). Fan, Liao, and Mincheva (2011, 2013) also show that the adaptive threshold technique of Cai and Liu applies equally to the residuals from an approximate factor model. Here we consider the threshold estimator proposed by BPS which does not require cross-validation and is shown to have desirable small sample properties. It is given by $\tilde{\mathbf{V}}_u = (\tilde{\sigma}_{ij})$

$$\begin{aligned} \tilde{\sigma}_{ii} &= \hat{\sigma}_{ii} \\ \tilde{\sigma}_{ij} &= \hat{\sigma}_{ij} \mathbf{1} [|\hat{\rho}_{ij}| > T^{-1/2} c_\alpha(n, \delta)], \quad i = 1, 2, \dots, n-1, \quad j = i+1, \dots, n, \end{aligned} \quad (59)$$

where

$$\hat{\sigma}_{ij} = \frac{1}{T} \sum_{t=1}^T \hat{u}_{it} \hat{u}_{jt}, \quad \hat{\rho}_{ij} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii} \hat{\sigma}_{jj}}}, \quad \hat{u}_{it} = r_{it} - \hat{\mathbf{a}}_{i,T} - \hat{\boldsymbol{\beta}}'_{i,T} \mathbf{f}_t, \quad (60)$$

and $c_p(n, d) = \Phi^{-1}(1 - \frac{p}{2n^d})$, is a normal critical value function, p is the nominal size of testing of $\sigma_{ij} = 0$, ($i \neq j$) and d is chosen to take account of the $n(n-1)/2$ multiple tests being carried out. Monte Carlo experiments carried out by BPS suggest setting $d = 2$. The variance estimator given by (59) does not require a knowledge of the factor strength and applies to risk factors of differing degrees.

Under Assumptions 1, 3, and 4, $\|\mathbf{V}_u\| = O(n^{\alpha_\gamma})$, and using results in Fan, Liao, and Mincheva (2011, 2013) we have

$$\|\tilde{\mathbf{V}}_u - \mathbf{V}_u\| = O_p\left(n^{\alpha_\gamma} \sqrt{\frac{\ln(n)}{T}}\right). \quad (61)$$

Consider the following estimator of \mathbf{V}_ξ

$$\hat{\mathbf{V}}_{\xi,nT} = (1 + \hat{s}_{nT}) \left(n^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \tilde{\mathbf{V}}_u \mathbf{M}_n \hat{\mathbf{B}}_{nT} \right).$$

where $\hat{s}_{nT} = \tilde{\boldsymbol{\lambda}}'_{nT} (T^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1} \tilde{\boldsymbol{\lambda}}_{nT}$. Under Assumption 1, $T^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{F} \rightarrow_p \boldsymbol{\Sigma}_f$ and using the results above we have $\tilde{\boldsymbol{\lambda}}_{nT} = \tilde{\boldsymbol{\phi}}_{nT} + \hat{\boldsymbol{\mu}}_T \rightarrow_p \boldsymbol{\phi}_0 + \boldsymbol{\mu}_0 = \boldsymbol{\lambda}_0$. Hence, $\hat{s}_{nT} \rightarrow_p \boldsymbol{\lambda}'_0 \boldsymbol{\Sigma}_f^{-1} \boldsymbol{\lambda}_0$ as $n, T \rightarrow \infty$, jointly, and it is sufficient to show that

$$n^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \tilde{\mathbf{V}}_u \mathbf{M}_n \hat{\mathbf{B}}_{nT} - n^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{V}_u \mathbf{M}_n \mathbf{B}_n \rightarrow_p \mathbf{0}. \quad (62)$$

The following theorem provides a formal statement of the conditions under which $\hat{\mathbf{V}}_{\xi,nT}$ is a consistent estimator of \mathbf{V}_ξ .

Theorem 5 *Suppose Assumptions 1, 2, 4, 3 and 5 hold, and all the observed factors are strong, ($\alpha_k = 1$, for $k = 1, 2, \dots, K$), and the strength of the missing factor, g_t , defined by (33), $\alpha_\gamma < 1/2$. Then*

$$\left\| \hat{\mathbf{V}}_{\xi,nT} - \mathbf{V}_\xi \right\| = O_p \left(n^{\alpha_\gamma} \sqrt{\frac{\ln(n)}{T}} \right), \quad (63)$$

where

$$\hat{\mathbf{V}}_{\xi,nT} = (1 + \hat{s}_{nT}) \left(n^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \tilde{\mathbf{V}}_u \mathbf{M}_n \hat{\mathbf{B}}_{nT} \right), \quad (64)$$

$\tilde{\mathbf{V}}_u = (\tilde{\sigma}_{ij})$, $\tilde{\sigma}_{ij}$ is the threshold estimator of σ_{ij} given by (59), and

$$\mathbf{V}_\xi = (1 + \boldsymbol{\lambda}'_0 \boldsymbol{\Sigma}_f^{-1} \boldsymbol{\lambda}_0) \underset{n \rightarrow \infty}{p \lim} \left(n^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{V}_u \mathbf{M}_n \mathbf{B}_n \right), \quad (65)$$

For a proof see sub-section B.5 in the Appendix.

This theorem shows that consistent estimation of $Var \left(\tilde{\boldsymbol{\phi}}_{nT} \right)$ can be achieved by using a suitable threshold estimator of \mathbf{V}_u , so long as the strength of the missing factor, α_γ , is sufficiently weak in the sense that $n^{\alpha_\gamma} \sqrt{\ln(n)/T} \rightarrow 0$ as $n, T \rightarrow \infty$.

4 Small sample properties of the estimators and tests for $\boldsymbol{\phi}_0$

4.1 Monte Carlo Design

This section presents Monte Carlo simulations to investigate the small sample properties of estimators and tests for $\boldsymbol{\phi}_0$. In the empirical application of the next section the factors are selected from a large list. But here we assume $K = 3$ and mimic

the 3 Fama-French factors, namely the market return minus the risk free rate, MKT, the value factor (high minus low book to market portfolios, HML) and the size factor (small minus big portfolios, SMB). These are denoted by f_{kt} , $k = M, H, S$.⁴ For further details see Section S3 of the online supplement A.

4.1.1 Loadings and factor strengths

To calibrate the loadings, β_{ik} , we used excess returns on a large number securities observed over the shorter sample covering the 20 years 2002m1 –2021m12 ($T = 240$). Monthly returns for NYSE and NASDAQ stocks code 10 and 11 from CRSP were downloaded from Wharton Research Data Services and converted to excess returns over the risk free rate, taken from Kenneth French’s webpages, in percent per month. Only stocks with available data for the full sample were included, yielding a balanced panel, and to avoid outliers influencing the results, stocks with a kurtosis greater than 16 were excluded. There were 1289 stocks before exclusion on the basis of kurtosis and 1175 after. The summary statistics giving mean, median, standard deviation of the estimates of β_{ik} and their histograms are provided in the online supplement A.

For factor strength, we considered a range of DGPs. Given the evidence that most factors, other than the market factor, are not strong, we focus on the case where there is one strong factor, namely the market factor with $\alpha_M = 1$, plus two semi-strong factors, with the value factor, HML , being quite strong with $\alpha_H = 0.85$, and the size factor, SML , being only moderately strong with $\alpha_S = 0.65$. These estimates are also informed by the results provided in Bailey, Kapetanios, and Pesaran (2021) who propose methods for estimation of factor strength. For a given factor strength, α_k , the associated loadings, β_{ik} , are generated as $\boldsymbol{\beta}_k = (\beta_{1k}, \beta_{2k}, \dots, \beta_{\alpha_k}, 0, 0, \dots, 0)$ where $n_{\alpha_k} = \lfloor n^{\alpha_k} \rfloor$ the integer part of n^{α_k} , with non-zero and zero values of $\boldsymbol{\beta}_k$ given by

$$\begin{aligned} \beta_{ik} &\sim IIDN(\mu_{\beta_k}, \sigma_{\beta_k}^2), \text{ for } i = 1, 2, \dots, \lfloor n^{\alpha_k} \rfloor, \\ \beta_{ik} &= 0 \text{ for } i = \lfloor n^{\alpha_k} \rfloor + 1, \lfloor n^{\alpha_k} \rfloor + 2, \dots, n, \end{aligned}$$

where $\lfloor n^{\alpha_k} \rfloor$ denotes the integer part of n^{α_k} . Since the security returns are randomly generated, it does not matter how zero and non-zero values of β_{ik} are distributed across i . Also, the zero loadings can also be replaced by an exponentially decaying sequence

⁴Data on factors and the risk free rate are downloaded from Kenneth French’s data library: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

without any implications for the simulation results.⁵ We also set

$$\begin{aligned}\mu_{\beta_M} &= 1, \sigma_{\beta_M} = 0.4; \mu_{\beta_H} = 0.2, \sigma_{\beta_H} = 0.5 \\ \mu_{\beta_S} &= 0.6, \sigma_{\beta_S} = 0.5,\end{aligned}$$

which match the mean and standard deviation of the estimates of β_{ik} . See above.

4.1.2 Generation of pricing errors

The pricing errors in (S3.2) can be considered as firm-specific characteristics and are set as $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_{n_\eta}, 0, 0, \dots, 0)'$. The non-zero loadings of $\boldsymbol{\eta}$ for $i \leq n_\eta = \lfloor n^{\alpha_\eta} \rfloor$ are drawn from $IIDU(0.7, 0.9)$, and $\eta_i = 0$ for $i = n_\eta + 1, n_\eta + 2, \dots, n$. We consider $\alpha_\eta = (0, 0.3)$. When $\alpha_\eta = 0$ we have $\eta_i = 0$, for all i . As in the case of factor loadings the non-zero values of $\boldsymbol{\eta}$ must be randomly allocated to different groups.

4.1.3 Generation of return equation errors

The return equation errors, u_{it} , are generated following (27) as a combination of a missing factor, $g_t \sim IIDN(0, 1)$ plus an idiosyncratic error, v_{jt} . The loadings $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_{n_\gamma}, 0, 0, \dots, 0)'$ of the missing factor are set as

$$\begin{aligned}\gamma_i &\sim IIDU(0.7, 0.9), \text{ for } i = 1, 2, \dots, \lfloor n^{\alpha_\gamma} \rfloor, \\ \gamma_i &= 0, \text{ for } i = \lfloor n^{\alpha_\gamma} \rfloor + 1, \lfloor n^{\alpha_\gamma} \rfloor + 2, \dots, n,\end{aligned}$$

where α_γ is the strength of the missing factor g_t . We consider $\alpha_\gamma = 1/4$ and $1/2$.

For the idiosyncratic errors, v_{it} , we consider spatial as well as a block diagonal specification, with the spatial specification including a diagonal specification as the special case. Under the spatial specification the idiosyncratic errors are generated as the first order spatial autoregressive model $v_{it} = \rho_\varepsilon \sum_{j=1}^n w_{ij} v_{jt} + \kappa \varepsilon_{it}$, which can be written in matrix notation as $\mathbf{v}_t = \rho_\varepsilon \mathbf{W} \mathbf{v}_t + \kappa \boldsymbol{\varepsilon}_t$, and solved for as $\mathbf{v}_t = \kappa (\mathbf{I}_n - \rho_\varepsilon \mathbf{W})^{-1} \boldsymbol{\varepsilon}_t$. Adding the missing factor now yields

$$\mathbf{u}_t = \boldsymbol{\gamma} g_t + \kappa (\mathbf{I}_n - \rho_\varepsilon \mathbf{W})^{-1} \boldsymbol{\varepsilon}_t. \quad (66)$$

The spatial coefficient ρ_ε is such that $|\rho_\varepsilon| < 1$, $\mathbf{W} = (w_{ij})$ with $w_{ii} = 0$, and $\sum_{j=1}^n w_{ij} = 1$. The diagonal case is obtained by setting $\rho_\varepsilon = 0$, with $\rho_\varepsilon = 0.5$ characterizing the SAR specification. The weight matrix $\mathbf{W} = (w_{ij})$ is set to follow the familiar rook pattern where all its elements are set to zero except for $w_{i+1,i} = w_{j-1,j} = 0.5$ for $i = 1, 2, \dots, n-2$ and $j = 3, 4, \dots, n$, with $w_{1,2} = w_{n,n-1} = 1$.

⁵See also footnote 5 of Bailey, Kapetanios, and Pesaran (2016, p.942).

Under the block error covariance specification, \mathbf{v}_t is generated as $\mathbf{v}_t = \kappa \hat{\mathbf{S}} \boldsymbol{\varepsilon}_t$, where $\hat{\mathbf{S}}$ is a block diagonal matrix with its b^{th} block given by $\hat{\mathbf{S}}_b$ for $b = 1, 2, \dots, B$, and $\boldsymbol{\varepsilon}_t = (\varepsilon'_{1t}, \varepsilon'_{2t}, \dots, \varepsilon'_{Bt})'$, and $\boldsymbol{\varepsilon}_{bt} = (\varepsilon_{b,1t}, \varepsilon_{b,2t}, \dots, \varepsilon_{b,n_b,t})'$. $\hat{\mathbf{S}}$ is set as a Cholesky factor of the correlation matrix of \mathbf{u}_t . Denoting this correlation matrix by $\hat{\mathbf{R}}_u$,

$$\hat{\mathbf{R}}_u = \left[\text{Diag}(\hat{\mathbf{V}}_{Bu}) \right]^{-1/2} \hat{\mathbf{V}}_{Bu} \left[\text{Diag}(\hat{\mathbf{V}}_{Bu}) \right]^{-1/2} = \text{Diag}(\hat{\mathbf{R}}_{bu}, b = 1, 2, \dots, B),$$

where $\hat{\mathbf{V}}_{Bu}$ is the threshold estimator of \mathbf{V}_u subject to the additional restriction that \mathbf{V}_u is block diagonal. For each block $\hat{\mathbf{R}}_{bu}$ we set the number of distinct non-zero elements of this block equal to the integer part of $[n_b(n_b - 1)/2] \times q_b$ where q_b is the proportion of non-zero distinct elements in block b of our calibrated sample and computed by the calibration over the sample 2001m10 – 2021m9. The non-zero elements are drawn randomly from $IIDU(0, 0.5)$. Similarly, adding the missing factor, we have

$$\mathbf{u}_t = \gamma g_t + \kappa \hat{\mathbf{S}} \boldsymbol{\varepsilon}_t. \quad (67)$$

The block diagonal structure is intended to capture possible within industry correlations not picked up by observed or weak missing factors, with each block representing an industry or sector. To calibrate the block structure estimates of the pair-wise correlations between the residuals of the return regressions using the Fama-French three factors of the $T = 240$ sample ending in 2021 were obtained. Then all the statistically insignificant correlations were set to zero, allowing for the multiple testing nature of the tests. For the majority of securities (668 out of the 1168), the pair-wise return correlations were not statistically significant. The securities with a relatively large number of non-zero correlations were either in the banking or energy related industries. Considering stocks by 2-digit SIC classifications, a division into $B = 14$ contiguous groups ranging in size from 33 to 145 stocks, seemed sensible. More detail on the process is given in Section S4 of the online supplement A.

The primitive errors, ε_{it} for $i = 1, 2, \dots, n$ in (66) and (67) are generated as $\varepsilon_{it} = \sqrt{\sigma_{ii}} \varpi_{it}$, where $\varpi_{it} \sim IIDN(0, 1)$, and $\varepsilon_{it} = \sqrt{\sigma_{ii}} \left[\sqrt{\frac{v-2}{v}} \varpi_{it} \right]$, where $\varpi_{it} \sim IID t(v)$, with $t(v)$ denotes a standard t distributed variate with $v = 5$ degrees of freedom. Also $\sigma_{ii} \sim IID 0.5(1 + \chi_1^2)$, for $i = 1, 2, \dots, n_b$ and $b = 1, 2, \dots, B$. In this way, it is ensured that $Var(\varepsilon_{b,it}) = \sigma_{b,ii}$, and on average $E[Var(\varepsilon_{b,it})] = E(\sigma_{ii}) = 1$, under both Gaussian and t -distributed errors. Note that $Var(\nu_{b,it}) = v/(v - 2)$. All the experiments are designed to give an R^2 of about 0.3, similar to that obtained in the empirical applications. For further details see sub-section S-3.5 of the online supplement A.

4.1.4 Experiments

In total, we consider 12 experimental designs: six designs with Gaussian errors and six with $t(5)$ distributed errors. We considered designs with GARCH effects, with and without pricing errors, η_i , and with and without the missing factor, g_t . We also considered designs with spatial patterns in the idiosyncratic errors, v_{it} . All experiments are implemented using $R = 2,000$ replications. Details of the 12 experiments are summarized in Table S-1 of the online supplement B (MC results).

4.1.5 Alternative estimators of \mathbf{V}_u

Subsection 3.2 considered consistent estimation of the variance of $\tilde{\phi}_{nT}$ using $\tilde{\mathbf{V}}_u$ a threshold estimator for \mathbf{V}_u , given by equation (59). For comparison purposes we also considered two other estimators of \mathbf{V}_u . These were the sample covariance matrix $\hat{\mathbf{V}}_u = \sum_{t=1}^T \hat{\mathbf{u}}_t \hat{\mathbf{u}}_t' / T$ and a diagonal covariance matrix, where the off-diagonal elements of $\hat{\mathbf{V}}_u$, $\hat{\sigma}_{ij}$, are set to zero. Thus we have three designs for the return error covariance matrix, \mathbf{V}_u , and three different estimators of it. A comparison of the results for the different covariance matrices is available on request. The diagonal estimator, as to be expected, performed poorly when the true covariance matrix was not diagonal, particularly for the spatial error covariance matrix and when the strength of the missing factor was close to 1/2. For these designs the sample and threshold estimators of the covariance matrix generally performed similarly and given that there is a theoretical justification for the threshold estimator and there are structures of the error covariance matrix for which the sample estimator is unlikely to perform well we report the results using the threshold estimator in the simulations below.

4.2 Monte Carlo results

We focus on a comparison of two-step (defined by (16)) and the bias-corrected (BC) estimator (defined by (17)), and report bias, root mean square error (RMSE) and size for testing $H_{0j} : \phi_{0k} = 0$, $k = M, H, S$ at the five per cent nominal level, for all $n = 100, 500, 1,000, 3,000$ and $T = 60, 120, 240$ combinations. The results for all 12 experiments are summarized in Tables S-A-E1-S-A-E12 in the online supplement B. In terms of bias and RMSE the two-step estimator does much better than the bias-

corrected (BC) estimator when $T = 60$ and $n = 100$, but this gap closes quickly as n is increased. In fact for $T = 60$ and $n = 3,000$, the bias and RMSE of the BC estimator (at 0.0010 and 0.0607) are much less than those of the two-step estimator (at -0.0080 and 0.1489). This pattern continues to hold when $T = 120$ and 240. Bias correction can cause the RMSE to "blow up" for small samples, such as $n = 100$ and $T = 60$, but for $n = 500$ and above the bias-corrected estimator always has a smaller RMSE than the two-step estimator. As discussed in the theoretical section, having a large n is important for the properties of the estimators.

But most importantly, the two-step estimator is subject to substantial size distortions, particularly when T is small relative to n . As predicted by the theory, the degree of over-rejection of the tests based on FM estimator falls with T , but increases with n . For example, the two-step test sizes rise from 11.1% when $T = 60$ and $n = 100$ to 60.9% when $T = 60$ and $n = 3,000$. Increasing T reduces the size distortion of the two-step estimator but test sizes are still substantially above the 5% nominal value when n is large. The strong tendency of the tests based on the two-step estimator to over-reject could be an important contributory factor leading to false discovery of a large number of apparently significant factors in the literature. In contrast, sizes of the tests based on the BC estimator, using the variance estimator given by (21), are all close to its nominal value, irrespective of the factor strength or sample size combinations. We only note some elevated test sizes in the case of the experimental design 12, and when we consider the semi-strong factors. The highest test size of 7.85 per cent is obtained for the least strong factor, f_{st} , when $n = 3000$ and $T = 60$. See Table S-A-E12 of the online supplement B.

We also experimented with raising α_η from 0.3 to 0.5, making the pricing errors much more pervasive and ρ_ε from 0.5 to 0.85, introducing more spatial correlation. This increased the rejection rate in experiments 9 and 10.

4.2.1 Empirical power functions

Plots of the empirical power functions for testing the null hypothesis that the spanning coefficients are equal to zero ($H_{0j} : \phi_{0k} = 0, k = M, H, S$) are also provided in the online supplement B (Figures S-A-E1 to S-A-E12). All power curves have the familiar bell curve shape and tend to unity as n and/or T are increased, showing the test has

satisfactory power, particularly for n sufficiently large even when $T = 60$. Again the power functions are quite similar across the 12 different experiments and show similar patterns for strong and semi-strong factors. However, this similarity hides the fact that the test of $\phi_k = 0$ for the strong factor is much more powerful than corresponding tests for the semi-strong factors, with the test power declining as factor strength is reduced.

4.2.2 Differences in performance of strong and semi-strong factors

These differences in the effects of factor strength on the power of the test of $\phi_k = 0$ are in line with our theoretical results, and are also reflected in the rate at which the RMSE of the estimators of ϕ_M , ϕ_H and ϕ_S fall with n . For example, using results in Table S-B-E12, for the bias-corrected estimator and design 12 with $T = 240$, the ratio of RMSE of $n = 3,000$ to $n = 100$ is 17% for the strong factor ($\alpha_M = 1$), 23% for the first semi-strong factor with $\alpha_H = 0.85$, and 36% for the second semi-strong factor with $\alpha_S = 0.65$. As strength falls one needs larger cross section samples of securities to attain the same level of precision. In the case of the two-step estimator there was the same pattern, but the fall in the RMSE with n was much slower. The ratio for $T = 240$ of RMSE of $n = 3,000$ to $n = 100$ is 25% for ϕ_M , rather than 17%, which it was for the bias-corrected estimator; 48% rather than 23% for ϕ_H , and the coefficient of the third least strong factor the RMSE fell then increased with n .

4.2.3 Semi-strong versus weak factors

So far, we have assumed that the DGP is correctly specified with two semi-strong and no observed weak factors. Here we consider the implications of incorrectly excluding semi-strong factors or correctly including weak factors on the small sample properties of the BC estimator of ϕ_M the coefficient of the strong factor. Using the same DGP (which includes one strong factor and two semi-strong factors), we carried out additional MC experiments (designs 1-12) where we also estimated ϕ_M without the semi-strong factors being included in the regression. Comparative results with and without the semi-strong factors. The results are summarized in Tables S-C-E1 to S-C-E12 of the online supplement B. We find that incorrectly excluding semi-strong factors can be quite costly, both in terms of bias and RMSE as well as size distortions. In terms of

RMSE it was almost always better to estimate the model with the semi strong factors included. The exception was for the case of $T = 60$, $n = 100$, where including the semi-strong factors with a very small sample size caused the RMSE to blow up. Size distortions resulting from the exclusion of the semi-strong factors tended to be more pronounced for large n and T samples. These conclusions were not sensitive to the choice of the experimental design. For these experiments the lesson seems to be that it is important to have n large and include relevant semi-strong factors providing that they are sufficiently strong.

When the DGP includes one strong factor ($\alpha_M = 1$) and two weak factors ($\alpha_H = \alpha_S = 0.5$), in terms of the bias and RMSE for ϕ_M it is unambiguously better to exclude the weak factors from the regression, even though they are in the DGP. Weak factors are best treated as missing and absorbed in the error term.

4.3 Main conclusions from MC experiments

The conclusions from the Monte Carlo simulations are that the bias corrected estimator of ϕ_0 , generally works well. Although, it can generate a large RMSE for small n and T , this can be solved by increasing n . This performance is robust to non-Gaussian errors, GARCH effects, missing weak factors, pricing errors and cross-sectional dependence of the types considered in these experiments. The size is generally correct and the power good. The rate at which RMSEs decline with n depends on the strength of the underlying factors. Semi-strong factors need much larger values of n for precise estimation. Tests of the joint significance of $\phi_M = \phi_H = \phi_S = 0$, not reported here, also performed well, as might be expected given the good power performance of the separate induced tests which are reported. Including weak factors could be harmful, but there are potential advantages of adding semi-strong factors, although issue of how best to select such factors is an open question to which we now turn.

5 Factor selection when the number of securities and the number of factors are both large

Our theoretical derivations and Monte Carlo simulations both assume that the number of risk factors included in the return regressions is fixed and the factors are known. For the empirical application we face the additional challenge of selecting a small number of relevant risk factors from a possibly large number of potential factors, m . This problem has been the subject of a number of recent studies. Harvey, Liu, and Zhu (2016) propose a multiple testing approach aimed at controlling the false discovery rate in the process of factor selection, and in a more recent paper Harvey and Liu (2020) suggest using a double-bootstrap method to calibrate the t-statistic used for controlling the desired level of the false discovery rate. Giglio and Xiu (2021) suggest applying the double-selection Lasso procedure by Belloni, Chernozhukov, and Hansen (2014) to second pass regressions. None of these methods distinguish between strong, semi-strong or weak factors in their selection process, whereas the theory and simulations presented above indicate the importance of factor strength for estimation and inference.

Here we propose an alternative selection procedure where we first estimate the strength of all the m factors under consideration, and then select factors with strength above a given threshold, the value of which is informed by the convergence results of Theorem 4. This theorem showed that if factor f_{kt} has strength α_k , then the BC estimator of ϕ_k converges to its true value, ϕ_{0k} , at the rate of $n^{\alpha_k/2}$. As is recognized in non-parametric estimation literature, if the rate of convergence is less than $1/3$, corresponding to $\alpha_k = 2/3$, the gain in precision with n is so slow that the estimator may not be that useful.⁶ Armed with the threshold of $2/3$, the main issue is how to estimate factor strength. This problem is already addressed in Bailey, Kapetanios, and Pesaran (2021, BKP) when m is fixed. In this setting they base their estimation on the statistical significance of f_{kt} in the first stage time-series regressions of excess returns on all the factors under consideration, whilst allowing for the n multiple testing problem which their approach entails. When m (the number of factors) is also large

⁶The Manski (1985) maximum score estimator for a binary response model has $n^{1/3}$ convergence and this is regarded as very slow and there are suggested modifications such as Horowitz (1992) to increase the rate of convergence to $n^{2/5}$.

the first stage regressions will also be subject to the multiple testing problem and penalized regression techniques such as Lasso or the one covariate at a time (OCMT) selection procedure technique proposed by Chudik, Kapetanios, and Pesaran (2018) could be used. Irrespective of selection technique used at the level of individual security returns, we end up with n different subsets of the m factors under consideration. Factor strengths can then be estimated similarly to BKP from their selection frequencies across the n securities.

To be more specific, denote the set of m factors under consideration by \mathcal{S} and denote the set of selected factors for security i by \hat{S}_i and their numbers by $\hat{m}_i = |\hat{S}_i|$. Clearly $\hat{S}_i \subseteq \mathcal{S}$, and $\hat{m}_i \leq m$, for $i = 1, 2, \dots, n$. Then compute the proportion of stocks in which the k^{th} factor is selected, $\hat{\pi}_k$, for $k \in \{1, 2, \dots, m\}$ based on \hat{S}_i , $i = 1, 2, \dots, n$ by $\hat{\pi}_k = \frac{1}{n} \sum_{i=1}^n \mathcal{I}\{k \in \hat{S}_i\}$. Then the strength of the k^{th} factor is measured by

$$\hat{\alpha}_k = \begin{cases} 1 + \frac{\ln \hat{\pi}_k}{\ln n}, & \text{if } \hat{\pi}_k > 0, \\ 0, & \text{if } \hat{\pi}_k = 0. \end{cases} \quad (68)$$

The transformation from $\hat{\pi}_k$ to $\hat{\alpha}_k$ is explained and justified in BKP, where it is shown that considering strength aids interpretation because it is not dependent on n . It is beyond the scope of the present paper to provide theoretical justification for the proposed factor selection procedure, but using extensive Monte Carlo experiments Yoo (2022) has shown that the proposed method has desirable small sample properties whether Lasso or OCMT is used for factor selection at the level of individual security returns.

6 An empirical application using a large number of U.S. securities and a large number of risk factors

This section uses the results above in the explanation of monthly returns for a large number, n , of U.S. securities, by a large active set of m potential risk factors. We first briefly describe the sources and characteristics of the data for the stock returns and factors, which cover different sub-samples over the period 1996m1 – 2021m12. We then consider the selection of a subset of K factors from the active set, and conditional on the selected factors present estimates of the spanning error parameters, ϕ_0 , of the

selected factors and discuss their statistical significance.

Monthly returns (inclusive of dividends) for NYSE and NASDAQ stocks from CRSP with codes 10 and 11 were downloaded on July 2 2022 from Wharton Research Data Services. They were converted to excess returns by subtracting the risk free rate, which was taken from Kenneth French’s data base. To obtain balanced panels of stock returns and factors, only variables for which there was data for the full sample under consideration were used. Excess returns are measured in percent per month. To avoid outliers influencing the results, stocks with a kurtosis greater than 16 were excluded. Four main samples were considered, each had 20 years of data, $T = 240$, ending in 2015m12, 2017m12, 2019m12, 2021m12. Filtering out the stocks with kurtosis larger than 16 removed about 100 of the roughly 1200 stocks. The number of stocks (n) considered for each of the four $T = 240$ samples are given in panel A of Table 1. Further detail is given in Section S5 of the online supplement A.⁷

6.1 Factor selection

For factor selection we used all the 240 observations. But for estimation of the spanning error parameters, ϕ_0 , we also consider the sub-sample of size $T = 120$ ending in the years 2015, 2017, 2019 and 2021. The Monte Carlo experiments indicated that when $T = 60$ the performance of the estimators was poor, so we do not report results for sub-samples of that size. The set of factors considered combine the 5 Fama-French factors with the 207 factors from the Chen and Zimmermann (2022), Open Source Asset Pricing webpages, both downloaded July 6 2022. Only factors with data for the full sample were considered so the return regressions constitute a balanced panel. The number of factors in each of the four 20 year samples ending in the years 2015, 2017, 2019, 2021 is also given in panel A of Table 1, and range between 187 to 199. Summary statistics for the factors in the active set are given in DS8 of the online supplement A.

To implement the factor selection procedure set out in Section 5, Lasso is used to carry out selection in the return regressions for individual securities. As is well known, Lasso does not work well with too many highly correlated regressors, therefore, factors with an absolute correlation with the market factor greater than 0.70 were dropped.

⁷Summary statistics for the excess returns across the different samples are given in Table DS7 of the online supplement A.

This still left between 177 and 190 risk factors in the active set \mathcal{S} , depending on the sample period (see panel A of Table 1).

Specifically, Lasso was applied n times to the regressions of excess returns, $r_{it} = R_{it} - r_t^f$, for $i = 1, 2, \dots, n$, on the 177 to 190 factors in the active set, \mathcal{S} , to select the sub-set \hat{S}_i for each i over the four 20-year samples, separately. Following the literature, the tuning parameters in the Lasso algorithm were set by ten-fold cross-validation.⁸ Interestingly, the market factor was selected by Lasso for almost all the securities, thus confirming the pervasive nature of the market factor. No other factor came close to being selected for all the securities. Lasso tended to choose a lot of non-market factors and every factor got chosen in at least one return regression. The mean number of non-market factors chosen by Lasso fell from 11.6 in the 2015 sample to 9.9 in the 2021 sample. The median was lower, falling from 10 to 8. There was a long right tail because Lasso tended to choose a very large number of non-market factors for some securities, ranging from a maximum of 48 in the 2021 sample to 54 in the 2017 sample.

Apart from the market factor, there are no systematic patterns for the rest of selected factors in the return equations for the individual securities. Following the theory set out in Section 5, the K factors used to estimate the spanning error parameters are chosen on the basis of their factor strength.⁹ A minimum threshold of 0.7 was used. This is just above the $2/3$ strength which results in irregular estimators of ϕ_0 that converge at rate $n^{1/3}$ and require very large values of n to achieve a reasonable degree of precision, as also shown by the Monte Carlo evidence. In addition, a large number of securities (n) are needed to obtain reasonably precise estimates of factor strength when it is below $2/3$. See the Monte Carlo evidence provided by Bailey, Kapetanios, and Pesaran (2021).

Given the lower bound of 0.7 threshold values of 0.70, 0.75, and 0.80 were considered. The number of selected factors for different choices of factor strength threshold is given in panel B of Table 1, for the four different samples.

Using the threshold of 0.70, 17 factors (inclusive of the market factor) were selected for the sample ending in 2015, with the number of selected factors declining to 15, 13

⁸The post-Lasso and one covariate multiple testing (OCMT) approach of Chudik, Kapetanios, and Pesaran (2018, CKP). were also investigated, but Lasso seemed to work reasonably well. The details of the Lasso procedure used are given in Section 2.2 of the online supplement of CKP paper.

⁹The idea of using factor strength could also be viewed as a kind of averaging of the factors selected in individual return regressions.

and 11, for the samples ending 2017, 2019 and 2021, respectively. At the other extreme, setting the threshold at 0.80, the number of selected factors dropped to 4 for the samples ending in 2015, 2017 and 2021, and 2 for the sample ending in 2021. Since 17 factors seemed too large and 2 factors too small, the threshold value was set at the intermediate value of 0.75. We considered always conditioning on the market factor, but since Lasso almost always selected it, this was unnecessary.

Table 1: Summary statistics for the number of stocks and number of selected factors using the factor strength threshold of 0.75, for four twenty years ($T = 240$) samples ending in 2021, 2019, 2017 and 2015

$T = 240$ with end dates	2021	2019	2017	2015
Panel A: Number of stocks and factors under consideration				
Number of stocks	1289	1276	1243	1181
Number of stocks with kurtosis < 16	1175	1143	1132	1090
Number of non-market factors	187	198	199	197
Number of non-market factors with $r < 0.70$	177	189	190	189
Panel B: Number of selected factors by strength threshold				
Number with strength > 0.80	2	4	4	4
Number with strength > 0.75	4	6	7	7
Number with strength > 0.70	11	13	15	17

Notes. Panel A shows the number of stocks and risk factors used before and after filtering by the specified criterion. Panel B shows the number of risk factors selected with strength greater than the specified threshold level using Lasso to select factors at the level of individual securities.

The list of selected factors for the four samples are given in Table 2. The three Fama-French factors, Market, HML and SMB, are all selected in all four periods.¹⁰ Of the Fama-French three, only the market factor is strong, with estimated strength in excess of 0.98 across the four periods.¹¹ The other factor which is selected across all the four periods is "Short-Interest". This is proposed by Dechow, Hutton, Meulbroek,

¹⁰We also considered selecting the risk factors using the generalized one covariate at a time (OCMT) method proposed by Sharifvaghefi (2022). Using GOCMT the Fama-French three factors were again amongst the five strongest factors selected. The use of GOCMT for factor selection is also investigated by Yoo (2022), using Monte Carlo and empirical applications.

¹¹These results also support the choice of 3 Fama-French factors and their strength used in our Monte Carlo simulations.

and Sloan (2001) who argue that short-sellers target firms that are priced high relative to fundamentals. It measures the extent to which investors are shorting the market as reflected in Compustat data. Two additional factors are selected in periods ending in 2019 and earlier. One is "Beta Tail Risk" proposed by Kelly and Jiang (2014) which estimates a time-varying tail exponent from the cross section of returns. The other is "Cash Based Operating Profitability" (CBOP) suggested by Ball, Gerakos, Linnainmaa, and Nikolaev (2016). This is operating profit less accruals, with working capital and R&D adjustments. For periods ending in 2017 and 2015 the "Sin Stock" indicator proposed by Hong and Kacperczyk (2009) is also selected. It takes the value of unity if the stock in question is involved in producing alcohol, tobacco, and gaming. They find that such stocks are held less by norm-constrained institutions such as pension plans.

The factor strengths are relatively stable across the periods, with many of the estimates close to the threshold value of 0.75. Apart from the market factor only SMB, Short Interest and Beta Tail Risk factors have strengths in excess of 0.85 when averaged across the four periods. From the large number of factors in the active set we have ended up with only four factors that are reasonably strong and their spanning risk parameter and associated risk premia can be estimated reasonably accurately.

The strengths of the selected factors are also closely related to the average measures of fit often used in the literature. Here we consider both $Ave\bar{R}^2 = n^{-1} \sum_{i=1}^n \bar{R}_i^2$, a simple average of the fit of the individual return regressions adjusted for degrees of freedom, $\bar{R}_i^2 = 1 - (T - K - 1)^{-1} \sum_{t=1}^T \hat{u}_{it}^2 / T^{-1} \sum_{t=1}^T (r_{it} - \bar{r}_{io})^2$, and the adjusted pooled R^2 defined by $\overline{PR}^2 = 1 - \hat{\sigma}_{nT}^2 / s_{r,nT}^2$, where $\hat{\sigma}_{nT}^2$ is the bias-corrected estimator of $\bar{\sigma}_n^2$ defined by (44) and $= (nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T (r_{it} - \bar{r}_{io})^2$. Both of these measures behave very similarly, but the pooled version is less sensitive to outliers. As shown in Appendix S6, for sufficiently large n and T , \overline{PR}^2 is dominated by the contribution of the most strong factor(s). Since the only strong factor selected is the market factor in Table 3 we report the $Ave\bar{R}^2$ and \overline{PR}^2 in the case of return regressions which just include the market factor and those which include all other factors with strength in excess of 0.75. First, we note that the \overline{PR}^2 values are generally lower than the $Ave\bar{R}^2$. Second, the additional factors do add to the fit, but their relative contributions vary considerably across sample sizes and periods. In general, the marginal contribution of

Table 2: Selected factors with estimated strength in excess of the threshold 0.75 for the samples of size $T=240$ ending in 2021, 2019, 2017 and 2015

End date	2021	2019	2017	2015
Selected Factors	Estimated strength (α)			
Mkt.	0.99	0.98	0.98	0.98
SMB	0.90	0.84	0.86	0.86
Short Interest	0.77	0.85	0.85	0.83
HML	0.76	0.77	0.76	0.75
BetaTailRisk		0.87	0.86	0.86
CBOP		0.76	0.77	0.77
Sin Stock		.	0.76	0.76

Notes: The risk factors listed are the market factor (Mkt.), size (SMB), Short Interest that measures the extent of short sales in the market, the value factor (HML), the cash-based operating profitability factor (CBOP), Beta Tail Risk, and Sin Stock which is a binary indicator taking the value of unity if the stock in question is involved in so called "Sin" industries producing alcohol, tobacco, and gaming. Further details on these risk factors are provided by Chen and Zimmermann (2022).

non-market factors tend to be smaller when T is larger, which is consistent with the theory for adjusted pooled R^2 set out in Section S6 of the online supplement A.

6.2 Estimation of ϕ_0 for selected risk factors

Given the selected factors we then estimate their spanning error parameters, ϕ_0 , for $T = 240$, and $T = 120$ samples. The risk factors used in both samples are selected using $T = 240$ observations. As noted earlier, we do not report estimates for $T = 60$ as they are likely to be misleading given the Monte Carlo evidence discussed above.

The bias-corrected estimates of ϕ_0 for samples ending 2015, 2017, 2019, and 2021 are summarized in Table 4. Panel A of this table gives the estimates using $T = 240$ months from the end of each period, whilst panel B reports the estimates based on $T = 120$. As can be seen, the spanning error parameter associated with the market factor is statistically significant only once. This is in the 20 year sample ending in 2021, which includes the disruption of financial markets by the COVID-19 pandemic. In contrast, the spanning error parameters of the Short Interest and HML risk factors

Table 3: Average and pooled R squared for the return regressions when using market factor alone or factors chosen by Pooled Lasso plus 0.75 threshold

End Year		2021		2019		2017		2015	
No. of stocks, n		1175		1143		1132		1090	
No. of selected factors		4		6		7		7	
		Mkt.	Selected	Mkt.	Selected	Mkt.	selected	Mkt.	selected
$Ave\bar{R}^2$	T								
	240	0.22	0.28	0.18	0.26	0.16	0.25	0.16	0.24
	120	0.23	0.30	0.24	0.31	0.25	0.33	0.25	0.32
\overline{PR}^2	240	0.20	0.24	0.17	0.24	0.15	0.23	0.15	0.23
	120	0.18	0.24	0.19	0.24	0.22	0.29	0.22	0.29

Note. This table shows for each of the four end years and the two sample sizes the adjusted average and pooled R^2 ($Ave\bar{R}^2$ and \overline{PR}^2) for the return regressions when using market factor alone or factors selected with strength higher than 0.75. The list of selected factors are given in Table 2.

are statistically significant at the 5% level in all the four periods, when $T = 240$. For the shorter sample sub-sample ($T = 120$), the spanning error parameter of the Short Interest risk factor are statistically significant in two of the four periods at the 1% level and that of HML is statistically significant in two of the periods once at the 5% level and once at the 1% level. For both Short Interest and HML factors the statistically significant coefficients were all negative. The ϕ parameter of the size factor (SMB) is statistically significant only once at the 10% level in the $T = 120$ sample ending in 2021. The Cash Based Operating Profitability and the Sin Stocks risk factors were both significant only once at the 10% level and the Beta Tail Risk not at all.

Table 4: Bias-corrected estimates of spanning parameters of the selected factors for the T=240 and T= 120 ending in the specified years

End date	2021	2019	2017	2015
	ϕ_0	ϕ_0	ϕ_0	ϕ_0
PANEL A: T = 240				
Mkt.	-0.49*** (0.09)	-0.38 (0.21)	-0.20 (0.34)	-0.07 (0.37)
SMB	0.12 (0.08)	-0.09 (0.08)	-0.02 (0.08)	-0.03 (0.08)
Short Interest	-1.16*** (0.14)	-1.07*** (0.17)	-1.21*** (0.17)	-1.25*** (0.15)
HML	-0.31*** (0.08)	-0.30** (0.12)	-0.47** (0.16)	-0.56*** (0.16)
CBOP		-0.23 (0.55)	-0.52 (0.73)	-0.64 (0.83)
Beta Tail Risk		0.17 (0.13)	-0.07 (0.19)	-0.14 (0.16)
Sin Stocks			-0.55* (0.24)	-0.40 (0.25)
PANEL B: T = 120				
Mkt.	-0.24 (0.23)	0.30 (0.93)	0.12 (0.40)	0.84 (0.68)
SMB	0.36* (0.18)	0.12 (0.22)	0.23 (0.24)	0.21 (0.21)
Short Interest	-0.41 (0.37)	-0.57 (0.36)	-1.45*** (0.22)	-1.13*** (0.22)
HML	-1.01*** (0.22)	-0.83** (0.27)	-0.23 (0.18)	-0.23 (0.18)
CBOP		0.97 (0.77)	-1.35* (0.56)	-0.49 (0.81)
BetaTail Risk		0.28 (0.18)	-0.27 (0.26)	-0.58 (0.39)
Sin Stocks			-0.20 (0.21)	0.24 (0.30)

Notes: Figures in brackets are asymptotic standard errors. * * * indicates statistical significance at the 1% level, ** at the 5% level, and * at the 10% level. See Table 2 for the description of the selected risk factors.

7 Concluding remarks

In this paper we have highlighted the importance of decomposing the risk premia, $\boldsymbol{\lambda}$, into the the factor mean, $\boldsymbol{\mu}$, and $\boldsymbol{\phi}$, and writing the alpha of security i , α_i , in terms of $\boldsymbol{\phi}$ and the idiosyncratic pricing errors. Given the pivotal role played by $\boldsymbol{\phi}$, both for estimating the risk premia and for tests of market efficiency, we have exclusively focussed on estimation of $\boldsymbol{\phi}$, and its asymptotic distribution under quite a general setting that allows for missing factors and idiosyncratic pricing errors. Since factor means, $\boldsymbol{\mu}$, can be estimated at the regular rate of $T^{-1/2}$ from time series data, it is relatively straightforward to develop a mixed strategy for estimation of $\boldsymbol{\lambda}$ by adding a time series estimate of $\boldsymbol{\mu}$ to the bias-corrected estimator of $\boldsymbol{\phi}$. If we use the same time series sample, such an estimator reduces to the Shanken bias-corrected estimator of $\boldsymbol{\lambda}$. But in practice, given the concern over the instability of the factor loadings β_{ik} over time, one could use relatively long time series, say T_μ , when estimating $\boldsymbol{\mu}$, and a shorter time series, say $T_\phi < T_\mu$, when estimating $\boldsymbol{\phi}$. The distributional and small sample properties of such a mixed estimator of risk premia is a topic for further research.

Our theoretical and Monte Carlo results further highlight the important role played by factor strengths in estimation and inference on $\boldsymbol{\phi}$, and hence on $\boldsymbol{\lambda}$. Factors with strength below $2/3$ lead to estimates of $\boldsymbol{\phi}$ with convergence rate of $n^{-1/3}$ or worse, and their use in asset pricing models can be justified only when n is very large. We have also shown that weak factors, with strength below $1/2$ are best treated as missing and absorbed in the error term. We have shown that estimation of $\boldsymbol{\phi}$ for strong or semi-strong factors is robust to weak missing factors, and the explicit inclusion of weak factors in the empirical analysis is likely to have adverse spill over effects on the estimates of $\boldsymbol{\phi}$ for strong and semi-strong factors. In view of these results we have proposed a factor selection procedure where only factors with strength above $2/3$ are included in the asset pricing model. Developing a formal statistical theory for the proposed selection is another topic for future research.

Our Monte Carlo results also raise a point of caution. Whereas the proposed bias-corrected estimator of $\boldsymbol{\phi}$ performed well in the Monte Carlo experiments for samples of size $T = 120$ and $T = 240$, and for a wide variety of DGPs, its performance was rather

poor when $T = 60$ and $n < 3,000$. It is therefore advisable not to use our proposed estimation and inference method in case of short T samples, unless n is relatively large (in excess of 3,000) and the underlying factors are sufficiently strong (in excess of 0.85).

References

- ANATOLYEV, S., AND A. MIKUSHEVA (2022): “Factor models with many assets: strong factors, weak factors, and the two-pass procedure,” *Journal of Econometrics*, 229, 103–126.
- BAI, J., AND G. ZHOU (2015): “Fama-MacBeth two-pass regressions: improving risk premia estimates,” *Finance Research Letters*, 15, 31–40.
- BAILEY, N., G. KAPETANIOS, AND M. H. PESARAN (2016): “Exponent of cross-sectional dependence: Estimation and inference,” *Journal of Applied Econometrics*, 31, 929–960.
- (2021): “Measurement of factor strength: theory and practice,” *Journal of Applied Econometrics*, 36, 587–613.
- BAILEY, N., M. H. PESARAN, AND L. V. SMITH (2019): “A multiple testing approach to the regularisation of large sample correlation matrices,” *Journal of Econometrics*, 208, 507–534.
- BALL, R., J. GERAKOS, J. T. LINNAINMAA, AND V. NIKOLAEV (2016): “Accruals, cash flows, and operating profitability in the cross section of stock returns,” *Journal of Financial Economics*, 121, 28–45.
- BEAULIEU, M.-C., J.-M. DUFOUR, AND L. KHALAF (2020): “Arbitrage pricing, weak beta, strong beta: identification-robust and simultaneous inference,” *Cireq Working paper 15-2020*.
- BELLONI, A., V. CHERNOZHUKOV, AND C. HANSEN (2014): “Inference on treatment effects after selection among high-dimensional controls,” *The Review of Economic Studies*, 81, 608–650.
- BICKEL, P. J., AND E. LEVINA (2008a): “Covariance regularization by thresholding,” *The Annals of Statistics*, 36, 2577–2604.
- (2008b): “Regularized estimation of large covariance matrices,” *The Annals of Statistics*, 36, 199–227.
- CAI, T., AND W. LIU (2011): “Adaptive thresholding for sparse covariance matrix estimation,” *Journal of the American Statistical Association*, 106, 672–684.
- CHAMBERLAIN, G. (1983): “Funds, factors, and diversification in arbitrage pricing models,” *Econometrica*, 51, 1305–1323.
- CHAMBERLAIN, G., AND M. ROTHCHILD (1983): “Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets,” *Econometrica*, 51, 1281–1304.

- CHEN, A. Y., AND T. ZIMMERMANN (2022): “Open source cross-sectional asset pricing,” *Critical Finance Review*, 27, 207–264.
- CHIB, S., X. ZENG, AND L. ZHAO (2020): “On comparing asset pricing models,” *The Journal of Finance*, 75, 551–577.
- CHUDIK, A., G. KAPETANIOS, AND M. H. PESARAN (2018): “A one covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models,” *Econometrica*, 86, 1479–1512.
- CHUDIK, A., M. H. PESARAN, AND E. TOSETTI (2011): “Weak and strong cross-section dependence and estimation of large panels,” *The Econometrics Journal*, 14, C45–C90.
- CONNOR, G., AND R. A. KORAJCZYK (2022): “Semi-strong factors in asset returns,” *Journal of Financial Econometrics*, (forthcoming).
- DECHOW, P. M., A. P. HUTTON, L. MEULBROEK, AND R. G. SLOAN (2001): “Short-sellers, fundamental analysis, and stock returns,” *Journal of Financial Economics*, 61, 77–106.
- DELLO PREITE, M., R. UPPAL, P. ZAFFARONI, AND I. ZVIADADZE (2022): “What is Missing in Asset-Pricing Factor Models?,” *Available at SSRN 4135146*.
- FAMA, E. F., AND K. R. FRENCH (1993): “Common risk factors in the returns on stocks and bonds,” *Journal of Financial Economics*, 33, 3–56.
- (2015): “A five-factor asset pricing model,” *Journal of Financial Economics*, 116, 1–22.
- (2018): “Choosing factors,” *Journal of Financial Economics*, 128, 234–252.
- FAMA, E. F., AND J. D. MACBETH (1973): “Risk, return, and equilibrium: empirical tests,” *Journal of Political Economy*, 81, 607–636.
- FAN, J., Y. LIAO, AND M. MINCHEVA (2011): “High dimensional covariance matrix estimation in approximate factor models,” *Annals of Statistics*, 39, 3320–3356.
- (2013): “Large covariance estimation by thresholding principal orthogonal complements,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75, 603–680.
- FENG, G., S. GIGLIO, AND D. XIU (2020): “Taming the factor zoo: a test of new factors,” *The Journal of Finance*, 75, 1327–1370.
- FENG, L., W. LAN, B. LIU, AND Y. MA (2022): “High-dimensional test for alpha in linear factor pricing models with sparse alternatives,” *Journal of Econometrics*, 229, 152–175.
- GAGLIARDINI, P., E. OSSOLA, AND O. SCAILLET (2016): “Time-varying risk premium in large cross-sectional equity data sets,” *Econometrica*, 84, 985–1046.
- GIBBONS, M. R., S. A. ROSS, AND J. SHANKEN (1989): “A test of the efficiency of a given portfolio,” *Econometrica*, 57, 1121–1152.
- GIGLIO, S., AND D. XIU (2021): “Asset pricing with omitted factors,” *Journal of Political Economy*, 129, 1947–1990.

- GOFMAN, M., G. SEGAL, AND Y. WU (2020): “Production networks and stock returns: The role of vertical creative destruction,” *The Review of Financial Studies*, 33, 5856–5905.
- HARVEY, C. R., AND Y. LIU (2019): “A census of the factor zoo,” *Available at SSRN 3341728*.
- (2020): “False (and missed) discoveries in financial economics,” *The Journal of Finance*, 75, 2503–2553.
- HARVEY, C. R., Y. LIU, AND H. ZHU (2016): “... and the cross-section of expected returns,” *The Review of Financial Studies*, 29, 5–68.
- HE, A., S. HE, D. HUANG, AND G. ZHOU (2022): “Testing Asset Pricing Models Using Pricing Error Information,” *Available at SSRN 3143752*.
- HERSKOVIC, B. (2018): “Networks in production: asset pricing implications,” *The Journal of Finance*, 73, 1785–1818.
- HONG, H., AND M. KACPERCZYK (2009): “The price of sin: the effects of social norms on markets,” *Journal of Financial Economics*, 93, 15–36.
- HOROWITZ, J. L. (1992): “A smoothed maximum score estimator for the binary response model,” *Econometrica*, 60, 505–531.
- HWANG, S., AND A. RUBESAM (2022): “Bayesian selection of asset pricing factors using individual stocks,” *Journal of Financial Econometrics*, 20, 716–761.
- JAGANNATHAN, R., G. SKOULAKIS, AND Z. WANG (2010): “The analysis of the cross-section of security returns,” in *Handbook of Financial Econometrics: Applications*, ed. by Y. Ait-Sahalia, and L. P. Hansen, vol. 2, chap. 14, pp. 73–134. Elsevier, Amsterdam.
- JENSEN, M. C. (1968): “The performance of mutual funds in the period 1945-1964,” *The Journal of Finance*, 23, 389–416.
- KAN, R., C. ROBOTTI, AND J. SHANKEN (2013): “Pricing model performance and the two-pass cross-sectional regression methodology,” *The Journal of Finance*, 68, 2617–2649.
- KELLY, B., AND H. JIANG (2014): “Tail risk and asset prices,” *The Review of Financial Studies*, 27, 2841–2871.
- KLEIBERGEN, F. (2009): “Tests of risk premia in linear factor models,” *Journal of Econometrics*, 149, 149–173.
- LETTAU, M., AND M. PELGER (2020a): “Estimating latent asset-pricing factors,” *Journal of Econometrics*, 218, 1–31.
- (2020b): “Factors that fit the time series and cross-section of stock returns,” *The Review of Financial Studies*, 33, 2274–2325.
- MA, S., W. LAN, L. SU, AND C.-L. TSAI (2020): “Testing alphas in conditional time-varying factor models with high-dimensional assets,” *Journal of Business & Economic Statistics*, 38, 214–227.
- MANSKI, C. F. (1985): “Semiparametric analysis of discrete response: asymptotic

- properties of the maximum score estimator,” *Journal of Econometrics*, 27, 313–333.
- ONATSKI, A. (2012): “Asymptotics of the principal components estimator of large factor models with weakly influential factors,” *Journal of Econometrics*, 168, 244–258.
- PESARAN, M. H., AND T. YAMAGATA (2012): “Testing CAPM with a large number of assets,” *IZA Discussion Papers No. 6469*.
- (2023): “Testing for alpha in linear factor pricing models with a large number of securities,” *Journal of Financial Econometrics (forthcoming)*.
- RAPONI, V., C. ROBOTTI, AND P. ZAFFARONI (2019): “Testing beta-pricing models using large cross-sections,” *The Review of Financial Studies*, 33, 2796–2842.
- ROSS, S. A. (1976): “The arbitrage theory of capital asset pricing,” *Journal of Economic Theory*, 13, 341–360.
- SHANKEN, J. (1992): “On the estimation of beta-pricing models,” *The Review of Financial Studies*, 5, 1–33.
- SHANKEN, J., AND G. ZHOU (2007): “Estimating and testing beta pricing models: alternative methods and their performance in simulations,” *Journal of Financial Economics*, 84, 40–86.
- SHARIFVAGHEFI, M. (2022): “Variable selection in linear regressions with many highly correlated covariates,” *Available at SSRN 4159979*.
- YOO, J. (2022): “Factor strengths in high-dimensional setting with applications to finance and macroeconomics,” *PhD Job Market Paper*, Department of Economics, University of Southern California.

A Online Mathematical Appendix

A.1 Introduction

In this online appendix we first introduce the notations used in our mathematical treatment, and state and establish a number of lemmas. We then provide detailed proofs of Theorems 1, 2, 3, 4 and 5 presented in the paper.

Notations: Generic positive finite constants are denoted by C when large, and c when small. They can take different values at different instances. \rightarrow^p denotes convergence in probability as $n, T \rightarrow \infty$, jointly. $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ denote the maximum and minimum eigenvalues of \mathbf{A} . $\mathbf{A} > 0$ denotes that \mathbf{A} is a positive definite matrix. $\|\mathbf{A}\| = \lambda_{\max}^{1/2}(\mathbf{A}'\mathbf{A})$, $\|\mathbf{A}\|_F = [\text{Tr}(\mathbf{A}'\mathbf{A})]^{1/2}$, $\|\mathbf{A}\|_p = (E\|\mathbf{A}\|^p)^{1/p}$, for $p \geq 2$ denote spectral, Frobenius, and ℓ_p norms of matrix \mathbf{A} , respectively. If $\{f_n\}_{n=1}^{\infty}$ is any real sequence and $\{g_n\}_{n=1}^{\infty}$ is a sequences of positive real numbers, then $f_n = O(g_n)$, if there exists C such that $|f_n|/g_n \leq C$ for all n . $f_n = o(g_n)$ if $f_n/g_n \rightarrow 0$ as $n \rightarrow \infty$. Similarly, $f_n = O_p(g_n)$ if f_n/g_n is stochastically bounded, and $f_n = o_p(g_n)$, if $f_n/g_n \rightarrow_p 0$, where \rightarrow_p denotes convergence in probability. If $\{f_n\}_{n=1}^{\infty}$ and $\{g_n\}_{n=1}^{\infty}$ are both positive sequences of real numbers, then $f_n = \Theta(g_n)$ if there exists $n_0 \geq 1$ such that $\inf_{n \geq n_0} (f_n/g_n) \geq C$, and $\sup_{n \geq n_0} (f_n/g_n) \leq C$.

A.2 Statement of lemmas and their proofs

Lemma A.1 Consider the errors $\{u_{it}, i = 1, 2, \dots, n; t = 1, 2, \dots, T\}$ defined by (27) and suppose that Assumptions 3 and 4 hold with $\alpha_\gamma < 1/2$, and $\mathbf{V}_u = (\sigma_{ij})$. Set $\bar{u}_{i\circ} = T^{-1} \sum_{t=1}^T u_{it}$, and $E(u_{it}^2) = \sigma_i^2$. Then

$$\|\mathbf{V}_u\| = \lambda_{\max}(\mathbf{V}_u) \leq \sup_i \sum_{j=1}^n |\sigma_{ij}| = O(n^{\alpha_\gamma}), \quad (\text{A.1})$$

$$n^{-1} \sum_{i=1}^n \sum_{j=1}^n |\sigma_{ij}| = O(1), \text{ and } n^{-1} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij}^2 = O(1). \quad (\text{A.2})$$

Also, for any t and t'

$$a_{n,tt} = \frac{1}{n} \sum_{i=1}^n (u_{it}^2 - \sigma_i^2) = O_p(n^{-1/2}), \quad (\text{A.3})$$

$$\text{Var}(a_{n,tt}) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(u_{it}^2, u_{jt}^2) = O(n^{-1}), \quad (\text{A.4})$$

$$a_{n,tt'} = n^{-1} \sum_{i=1}^n u_{it} u_{it'} = O_p(n^{-1/2}), \text{ for } t \neq t', \quad (\text{A.5})$$

$$\text{Var}(a_{n,tt'}) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij}^2 = O(n^{-1}), \text{ for } t \neq t', \quad (\text{A.6})$$

$$b_{n,t} = \frac{1}{n} \sum_{i=1}^n (u_{it} \bar{u}_{i\circ} - T^{-1} \sigma_i^2) = O_p(T^{-1/2} n^{-1/2}), \quad (\text{A.7})$$

and

$$\text{Var}(b_{n,t}) = O(T^{-1} n^{-1}). \quad (\text{A.8})$$

Proof. Result (A.1) follows noting that under Assumptions 3 and 4, $\sigma_{ij} = \gamma_i \gamma_j + \sigma_{v,ij}$, $\sum_{j=1}^n |\sigma_{ij}| \leq \sup_i |\gamma_i| \sum_{j=1}^n |\gamma_j| + \sum_{j=1}^n |\sigma_{v,ij}|$, and by assumption $\sup_i |\gamma_i| < C$, $\sup_i \sum_{j=1}^n |\sigma_{v,ij}| < C$, and $\sum_{j=1}^n |\gamma_j| = O(n^{\alpha_\gamma})$. To prove (A.2)

$$\begin{aligned} n^{-1} \sum_{i=1}^n \sum_{j=1}^n |\sigma_{ij}| &\leq n^{-1} \sum_{i=1}^n \sum_{j=1}^n |\gamma_i| |\gamma_j| + n^{-1} \sum_{i=1}^n \sum_{j=1}^n |\sigma_{v,ij}| \\ &= n^{-1} \left(\sum_{i=1}^n |\gamma_i| \right)^2 + n^{-1} \sum_{i=1}^n \sum_{j=1}^n |\sigma_{v,ij}|. \end{aligned}$$

By assumption $n^{-1} \sum_{i=1}^n \sum_{j=1}^n |\sigma_{v,ij}| = O(1)$, and $\sum_{i=1}^n |\gamma_i| = O(n^{\alpha_\gamma})$, then, in view of (33) and (34) $n^{-1} \sum_{i=1}^n \sum_{j=1}^n |\sigma_{ij}| = O(n^{-1+2\alpha_\gamma}) + O(1) = O(1)$, since $\alpha_\gamma < 1/2$. Similarly $\sigma_{ij}^2 = \gamma_i^2 \gamma_j^2 + \sigma_{v,ij}^2 + 2\gamma_i \gamma_j \sigma_{v,ij}$, and

$$n^{-1} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij}^2 = n^{-1} \text{Tr}(\mathbf{V}_u^2) = n^{-1} \left(\sum_{i=1}^n \gamma_i^2 \right)^2 + n^{-1} \sum_{i=1}^n \sum_{j=1}^n \sigma_{v,ij}^2 + 2n^{-1} \gamma' \mathbf{V}_v \gamma.$$

But

$$\begin{aligned} n^{-1} \sum_{i=1}^n \sum_{j=1}^n \sigma_{v,ij}^2 &= n^{-1} \text{Tr}(\mathbf{V}_v^2) \leq \lambda_{\max}^2(\mathbf{V}_v) = O(1), \\ \gamma' \mathbf{V}_v \gamma &\leq (\gamma' \gamma) \lambda_{\max}(\mathbf{V}_v) = O(n^{\alpha_\gamma}). \end{aligned} \quad (\text{A.9})$$

Overall

$$n^{-1} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij}^2 = O(n^{-1+2\alpha_\gamma}) + O(1) + O(n^{-1+\alpha_\gamma}),$$

and since $\alpha_\gamma < 1/2$, then it follows that $n^{-1} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij}^2 = O(1)$. To prove (A.3) we first note that (using the normalization $E(g_t^2) = 1$)

$$u_{it}^2 - E(u_{it}^2) = (g_t^2 - 1)\gamma_i^2 + (v_{it}^2 - \sigma_{v,ii}) + 2g_t\gamma_i v_{it}, \quad (\text{A.10})$$

and

$$\begin{aligned} a_{n,tt} &= (g_t^2 - 1) \left(n^{-1} \sum_{i=1}^n \gamma_i^2 \right) + n^{-1} \sum_{i=1}^n (v_{it}^2 - \sigma_{v,ii}) + 2g_t \left(n^{-1} \sum_{i=1}^n \gamma_i v_{it} \right) \\ &= O(n^{-1+\alpha_\gamma}) + O_p(n^{-1/2}) + O(n^{-1+\alpha_\gamma/2}) = O_p(n^{-1/2}), \text{ since } \alpha_\gamma < 1/2. \end{aligned}$$

To establish (A.4), using (A.10) we first note that (for given loadings γ_i)

$$\text{Cov}(u_{it}^2, u_{jt}^2) = \gamma_i^2 \gamma_j^2 \left[E(g_t^2 - 1)^2 \right] + \text{Cov}(v_{it}^2, v_{jt}^2) + 4E(g_t^2) \gamma_i \gamma_j \sigma_{v,ij},$$

$$\begin{aligned} \text{Var}(a_{n,tt}) &= n^{-2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(u_{it}^2, u_{jt}^2) = \text{Var}(g_t^2) \left(n^{-1} \sum_{i=1}^n \gamma_i^2 \right)^2 \\ &\quad + n^{-2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(v_{it}^2, v_{jt}^2) + 4E(g_t^2) n^{-2} \gamma' \mathbf{V}_v \gamma = O(n^{-2+2\alpha_\gamma}) + O(n^{-1}) + O(n^{-2+\alpha_\gamma}), \end{aligned}$$

and since $\alpha_\gamma < 1/2$, then $\text{Var}(a_{n,tt}) = O(n^{-1})$, as required (which also corroborate (A.3)). Consider now (A.5) and since u_{it} is serially independent then $E(a_{n,tt'}) = 0$ for $t \neq t'$, and we have

$$E(u_{it} u_{it'} u_{jt} u_{jt'}) = E(u_{it} u_{jt} u_{it'} u_{jt'}) = E(u_{it} u_{jt}) E(u_{it'} u_{jt'}) = \sigma_{ij}^2 \text{ for } t \neq t',$$

and

$$\begin{aligned} Var(a_{n,tt'}) &= E(a_{n,tt'}^2) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n E(u_{it}u_{it'}u_{jt}u_{jt'}), \text{ for } t \neq t' \\ &= n^{-2} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij}^2 = O(n^{-1}), \text{ using (A.2),} \end{aligned}$$

which establishes (A.6) and (A.5). To prove (A.7) set $z_{it} = u_{it}\bar{u}_{i0} - T^{-1}\sigma_i^2$, and write $b_{n,t} = \frac{1}{n} \sum_{i=1}^n z_{it}$. Also note that $u_{it}\bar{u}_{i0} = T^{-1} \sum_{s=1}^T u_{it}u_{is}$, and given that $\{u_{it}\}$ is serially independent then $E(z_{it}) = 0$ and $E(b_{n,t}) = 0$. and

$$\begin{aligned} Var(b_{n,t}) &= n^{-2} \sum_{i=1}^n \sum_{j=1}^n E(z_{it}z_{jt}) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n E(u_{it}u_{jt}\bar{u}_{i0}\bar{u}_{j0}) - T^{-2}\bar{\sigma}_n^4 \\ &= n^{-2}T^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{s=1}^T \sum_{s'=1}^T E(u_{it}u_{jt}u_{is}u_{js'}) - T^{-2}\bar{\sigma}_n^4. \end{aligned} \quad (\text{A.11})$$

Also $E(u_{it}u_{jt}u_{is}u_{js'}) = 0$ for all t if $s \neq s'$. We are left with one case where $s = s' = t$, and one case where $s = s' \neq t$. Hence

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^n \sum_{s=1}^T \sum_{s'=1}^T E(u_{it}u_{jt}u_{is}u_{js'}) &= \sum_{i=1}^n \sum_{j=1}^n \sum_{s=1}^T E(u_{it}u_{jt}u_{is}u_{js}) \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{s=1, s=t}^T E(u_{it}u_{jt}u_{is}u_{js}) + \sum_{i=1}^n \sum_{j=1}^n \sum_{s=1, s \neq t}^T E(u_{it}u_{jt}u_{is}u_{js}) \\ &= \sum_{i=1}^n \sum_{j=1}^n E(u_{it}^2u_{jt}^2) + \sum_{i=1}^n \sum_{j=1}^n \sum_{s=1, s \neq t}^T E(u_{it}u_{jt}) E(u_{is}u_{js}) \\ &= \sum_{i=1}^n \sum_{j=1}^n [Cov(u_{it}^2, u_{jt}^2) + \sigma_i^2\sigma_j^2] + (T-1) \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij}^2, \end{aligned}$$

and hence

$$\sum_{i=1}^n \sum_{j=1}^n \sum_{s=1}^T \sum_{s'=1}^T E(u_{it}u_{jt}u_{is}u_{js'}) = \sum_{i=1}^n \sum_{j=1}^n Cov(u_{it}^2, u_{jt}^2) + \left(\sum_{i=1}^n \sigma_i^2 \right)^2 + (T-1) \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij}^2,$$

$$\begin{aligned}
\text{Var}(b_{n,t}) &= n^{-2} T^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{s=1}^T \sum_{s'=1}^T E(u_{it} u_{jt} u_{is} u_{js'}) - T^{-2} \bar{\sigma}_n^4 \\
&= T^{-2} n^{-2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(u_{it}^2, u_{jt}^2) + T^{-2} \bar{\sigma}_n^4 + (T-1) \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij}^2 - T^{-2} \bar{\sigma}_n^4.
\end{aligned}$$

Using this result in (A.11) we have

$$\text{Var}(b_{n,t}) = n^{-2} T^{-2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(u_{it}^2, u_{jt}^2) + \frac{T-1}{T^2} \left(n^{-2} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij}^2 \right).$$

Now using (A.4) and (A.2) we have $\text{Var}(b_{n,t}) = O(T^{-1}n^{-1})$, which establishes (A.8), and result (A.7) follows by Markov inequality. ■

Lemma A.2 Consider the $n \times T$ error matrix $\mathbf{U}_{nT} = (\mathbf{u}_{1o}, \mathbf{u}_{2o}, \dots, \mathbf{u}_{no})'$, where $\mathbf{u}_{io} = (u_{i1}, u_{i2}, \dots, u_{iT})'$, the $n \times k$ matrix of factor loadings, $\mathbf{B}_n = (\boldsymbol{\beta}_{o1}, \boldsymbol{\beta}_{o2}, \dots, \boldsymbol{\beta}_{oK})$, where $\boldsymbol{\beta}_{ok} = (\beta_{1k}, \beta_{2k}, \dots, \beta_{nk})'$, the $n \times 1$ vector of pricing errors $\boldsymbol{\eta}_n = (\eta_1, \eta_2, \dots, \eta_n)'$, with the pervasiveness coefficient, α_η , the observed factors, $\mathbf{f}_k = (f_{k1}, f_{k2}, \dots, f_{kT})'$ are strong (with $\alpha_k = 1$, for $k = 1, 2, \dots, K$), the missing factor g_t , has strength $\alpha_\gamma < 1/2$, $\mathbf{G}_T = \mathbf{M}_T \mathbf{F} (\mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1}$, $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_K)$, $\mathbf{M}_n = \mathbf{I}_n - \frac{1}{n} \boldsymbol{\tau}_n \boldsymbol{\tau}_n'$, $\bar{\mathbf{u}}_{no} = (\bar{u}_{1o}, \bar{u}_{2o}, \dots, \bar{u}_{no})'$, $\bar{u}_{io} = T^{-1} \sum_{t=1}^T u_{it}$, $\bar{\sigma}_n^2 = n^{-1} \sum_{i=1}^n \sigma_i^2$, and $\boldsymbol{\tau}_n$ and $\boldsymbol{\tau}_T$ are, respectively, $n \times 1$ and $T \times 1$ vectors of ones. Suppose that Assumptions 1, 3, 4, 2 and 5 hold. Then

$$n^{-1} \mathbf{B}_n' \mathbf{M}_n \boldsymbol{\eta}_n = O_p(n^{-1+\alpha_\eta}), \quad (\text{A.12})$$

$$n^{-1} \mathbf{B}_n' \mathbf{M}_n \bar{\mathbf{u}}_{no} = O_p(T^{-1/2} n^{-1/2}), \quad (\text{A.13})$$

$$n^{-1} \mathbf{B}_n' \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_T = O_p(T^{-1/2} n^{-1/2}), \quad (\text{A.14})$$

$$n^{-1} \mathbf{G}_T' \mathbf{U}_{nT}' \mathbf{M}_n \boldsymbol{\eta}_n = O_p\left(T^{-1/2} n^{-1+\frac{\alpha_\eta+\alpha_\gamma}{2}}\right), \quad (\text{A.15})$$

$$n^{-1} \mathbf{G}_T' \mathbf{U}_{nT}' \mathbf{M}_n \bar{\mathbf{u}}_{no} = O_p(T^{-1} n^{-1/2}), \quad (\text{A.16})$$

$$\mathbf{G}_T' (n^{-1} \mathbf{U}_{nT}' \mathbf{M}_n \mathbf{U}_{nT} - \bar{\sigma}_n^2 \mathbf{I}_T) \mathbf{G}_T = O_p(T^{-1} n^{-1/2}). \quad (\text{A.17})$$

Proof. To establish (A.12) we first note that the k^{th} element of $n^{-1} \mathbf{B}_n' \mathbf{M}_n \boldsymbol{\eta}_n$ can be written as

$$\pi_{k,n} = n^{-1} \sum_{i=1}^n (\beta_{ik} - \bar{\beta}_k) \eta_i, \quad (\text{A.18})$$

where $\bar{\beta}_k = n^{-1} \boldsymbol{\tau}'_n \boldsymbol{\beta}_{ok}$. Since η_j and β_{ik} are distributed independently for all i and j , then $E(\pi_{k,n}) = 0$,¹²

$$E(|\pi_{k,n}| | \boldsymbol{\eta}_n) = n^{-1} \sum_{i=1}^n E|\beta_{ik} - \bar{\beta}_k| |\eta_i| \leq \left[\sup_{i,k} E|\beta_{ik} - \bar{\beta}_k| \right] \left(n^{-1} \sum_{i=1}^n |\eta_i| \right) = O(n^{-1+\alpha_\eta}),$$

for $k = 1, 2, \dots, K$, and by Markov inequality we have $n^{-1} \mathbf{B}'_n \mathbf{M}_n \boldsymbol{\eta}_n = O_p(n^{-1+\alpha_\eta})$, as required. To establish (A.13), noting that

$$n^{-1} \bar{\mathbf{u}}'_{no} \mathbf{M}_n \mathbf{B}_n = \left[n^{-1} \bar{\mathbf{u}}'_{no} (\boldsymbol{\beta}_{o1} - \tau_n \bar{\beta}_1), n^{-1} \bar{\mathbf{u}}'_{no} (\boldsymbol{\beta}_{o2} - \tau_n \bar{\beta}_2), \dots, n^{-1} \bar{\mathbf{u}}'_{no} (\boldsymbol{\beta}_{oK} - \tau_n \bar{\beta}_K) \right],$$

and the k^{th} element of $n^{-1} \mathbf{B}'_n \mathbf{M}_n \bar{\mathbf{u}}_{no}$ is given by $c_{k,nT} = n^{-1} \sum_{i=1}^n (\beta_{ik} - \bar{\beta}_k) \bar{u}_{io}$. We have $E(c_{k,nT}) = 0$, and $Var(c_{k,nT} | \boldsymbol{\beta}_{ok}) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n (\beta_{ik} - \bar{\beta}_k) (\beta_{jk} - \bar{\beta}_k) E(\bar{u}_{io} \bar{u}_{jo})$, with $E(\bar{u}_{io} \bar{u}_{jo}) = T^{-1} \sigma_{ij}$. Hence (recalling that $\mathbf{V}_u = (\sigma_{ij})$)

$$Var(c_{nT,k}) = T^{-1} n^{-2} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} E[(\beta_{ik} - \bar{\beta}_k) (\beta_{jk} - \bar{\beta}_k)], \quad (\text{A.19})$$

and by Cauchy–Schwarz inequality and Assumption 2 we have $E[(\beta_{ik} - \bar{\beta}_k) (\beta_{jk} - \bar{\beta}_k)] \leq \left[E(\beta_{ik} - \bar{\beta}_k)^2 \right]^{1/2} \left[E(\beta_{jk} - \bar{\beta}_k)^2 \right]^{1/2} < C$. Hence, $Var(c_{nT,k}) \leq CT^{-1} n^{-2} \sum_{i=1}^n \sum_{j=1}^n |\sigma_{ij}|$. Since $\alpha_\gamma < 1/2$, then by (A.2) we have $n^{-1} \sum_{i=1}^n \sum_{j=1}^n |\sigma_{ij}| = O(1)$ and $Var(c_{nT,k}) = O(T^{-1} n^{-1})$. Thus by Markov inequality it follows that $c_{nT,k} = O_p(T^{-1/2} n^{-1/2})$, for $k = 1, 2, \dots, K$, and (A.13) follows. To establish (A.14) using $\mathbf{G}_T = \mathbf{M}_T \mathbf{F} (\mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1}$ we have

$$n^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_T = n^{-1} T^{-1} (\mathbf{B}'_n \mathbf{M}_n \mathbf{U}_{nT} \mathbf{M}_T \mathbf{F}) (T^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1},$$

and since $(T^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1}$ is a positive definite matrix then it is sufficient to consider the probability order of the $K \times K$ matrix $T^{-1} n^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{U}_{nT} \mathbf{M}_T \mathbf{F} = (q_{kk'})$, where

$$q_{kk'} = T^{-1} n^{-1} \sum_{i=1}^n \sum_{t=1}^T (\beta_{ik} - \bar{\beta}_k) (f_{k't} - \bar{f}_{k'}) u_{it} = n^{-1} \sum_{i=1}^n (\beta_{ik} - \bar{\beta}_k) \psi_{iT},$$

¹²Note that $\sup_{i,k} E|\beta_{ik} - \bar{\beta}_k| < C$ follows from $\sup_i E|\beta_{ik}|^2 < C$, required by Assumption 2.

where $\psi_{iT} = T^{-1} \sum_{t=1}^T (f_{k't} - \bar{f}_{k'}) u_{it}$. Thus

$$\text{Var}(q_{kk'} | \boldsymbol{\beta}_{ok}) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n (\beta_{ik} - \bar{\beta}_k) (\beta_{jk} - \bar{\beta}_k) \text{Cov}(\psi_{iT}, \psi_{jT}). \quad (\text{A.20})$$

Also

$$\text{Cov}(\psi_{iT}, \psi_{jT} | \mathbf{F}) = E(\psi_{iT} \psi_{jT} | \mathbf{F}) = T^{-2} \sum_{t=1}^T \sum_{t'=1}^T (f_{k't} - \bar{f}_{k'}) (f_{k't'} - \bar{f}_{k'}) E(u_{it} u_{jt'}),$$

and $E(u_{it} u_{jt'}) = \sigma_{ij}$ for $t = t'$ and 0 otherwise ($t \neq t'$). Then,

$$\text{Cov}(\psi_{iT}, \psi_{jT} | \mathbf{F}) = T^{-2} \sum_{t=1}^T (f_{k't} - \bar{f}_{k'})^2 E(u_{it} u_{jt}) = \sigma_{ij} T^{-2} \sum_{t=1}^T (f_{k't} - \bar{f}_{k'})^2.$$

Using this result in (A.20) now yields

$$\text{Var}(q_{kk'}) = T^{-1} n^{-1} \left[T^{-1} \sum_{t=1}^T E(f_{k't} - \bar{f}_{k'})^2 \right] n^{-1} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} E[(\beta_{ik} - \bar{\beta}_k) (\beta_{jk} - \bar{\beta}_k)].$$

The first term of the above is bounded since by assumption \mathbf{f}_t is stationary. The second term is bounded as established above (see the derivations below (A.19)). Hence $\text{Var}(q_{kk'}) = O(T^{-1} n^{-1})$ and result (A.14) follows. To establish (A.15) note that $n^{-1} \mathbf{G}'_T \mathbf{U}'_{nT} \mathbf{M}_n \boldsymbol{\eta}_n = (T^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1} (n^{-1} T^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{U}'_{nT} \mathbf{M}_n \boldsymbol{\eta}_n)$, where $(T^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1} = O_p(1)$. Also the k^{th} element of $n^{-1} T^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{U}'_{nT} \mathbf{M}_n \boldsymbol{\eta}_n$ is given by $p_k = T^{-1} \sum_{t=1}^T (f_{kt} - \bar{f}_k) c_{n,t}$, where $c_{n,t} = n^{-1} \sum_{i=1}^n u_{it} \eta_i = n^{-1} \boldsymbol{\eta}'_n \mathbf{u}_{ot}$. Under Assumption (4) f_{kt} and $c_{n,t}$ are distributed independently, $E(p_{n,t}) = 0$, and $c_{n,t}$ are also serially uncorrelated we have $\text{Var}(p_k) = T^{-2} \sum_{t=1}^T E(f_{kt} - \bar{f}_k)^2 \text{Var}(c_{n,t})$. Also noting that $\text{Var}(c_{n,t} | \boldsymbol{\eta}) = n^{-2} \boldsymbol{\eta}'_n \mathbf{V}_u \boldsymbol{\eta}_n$, then

$$\text{Var}(p_k) = \left[T^{-1} \sum_{t=1}^T E(f_{kt} - \bar{f}_k)^2 \right] (T^{-1} n^{-2} \boldsymbol{\eta}'_n \mathbf{V}_u \boldsymbol{\eta}_n).$$

The first term is bounded, and it follows that

$$\text{Var}(p_k | \boldsymbol{\eta}_n) < CT^{-1} n^{-2} \boldsymbol{\eta}'_n \mathbf{V}_u \boldsymbol{\eta}_n \leq CT^{-1} n^{-2} (\boldsymbol{\eta}'_n \boldsymbol{\eta}_n) \lambda_{\max}(\mathbf{V}_u) = O(T^{-1} n^{-2 + \alpha_\eta + \alpha_\gamma}),$$

and (A.15) follows. To establish (A.16) note that

$$n^{-1}\mathbf{G}'_T\mathbf{U}'_{nT}\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} = (T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1} (n^{-1}T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{U}'_{nT}\mathbf{M}_n\bar{\mathbf{u}}_{n\circ}),$$

and the k^{th} element of $n^{-1}T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{U}'_{nT}\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} = (d_k)$ is given by

$$\begin{aligned} d_k &= \frac{1}{nT} \sum_{t=1}^T (f_{kt} - \bar{f}_k) \sum_{i=1}^n (u_{it} - \bar{u}_{i\circ}) \bar{u}_{i\circ} = \frac{1}{nT} \sum_{t=1}^T \sum_{i=1}^n (f_{kt} - \bar{f}_k) (u_{it}\bar{u}_{i\circ} - \bar{u}_{i\circ}\bar{u}_{i\circ}) \\ &= \frac{1}{T} \sum_{t=1}^T (f_{kt} - \bar{f}_k) \left[n^{-1} \sum_{i=1}^n (u_{it}\bar{u}_{i\circ} - T^{-1}\bar{\sigma}_n^2) \right] = \frac{1}{T} \sum_{t=1}^T (f_{kt} - \bar{f}_k) b_{n,t}, \end{aligned}$$

where $b_{n,t} = n^{-1} \sum_{i=1}^n (u_{it}\bar{u}_{i\circ} - T^{-1}\bar{\sigma}_n^2)$. By assumption u_{it} and $f_{k't}$ are distributed independently and $b_{n,t}$ are also serially independent. Then $E(d_k) = 0$, and

$$\text{Var}(d_k) = \frac{1}{T^2} \sum_{t=1}^T (f_{kt} - \bar{f}_k)^2 \text{Var}(b_{n,t}).$$

Now using (A.8) in Lemma A.1, $\text{Var}(b_{n,t}) = O(T^{-1}n^{-1})$, and overall we have $\text{Var}(d_k) = O(T^{-2}n^{-1})$, and by Markov inequality $d_k = O_p(T^{-1}n^{-1/2})$, as required. Finally, consider (A.17) and note that

$$\begin{aligned} &\mathbf{G}'_T (n^{-1}\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{U}_{nT} - \bar{\sigma}_n^2\mathbf{I}_T) \mathbf{G}_T \\ &= T^{-2} (T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1} \mathbf{F}'\mathbf{M}_T [n^{-1}\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{U}_{nT} - \bar{\sigma}_n^2\mathbf{I}_T] \mathbf{M}_T\mathbf{F} (T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}. \end{aligned}$$

Since by assumption $(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1}$ is a positive definite matrix for all T , and K is fixed then it is sufficient to derive the probability order of the (k, k') element of of the $K \times K$ matrix $\Delta_{nT} = (\delta_{kk'})$

$$\begin{aligned} \Delta_{nT} &= T^{-2}\mathbf{F}'\mathbf{M}_T [n^{-1}\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{U}_{nT} - \bar{\sigma}_n^2\mathbf{I}_T] \mathbf{M}_T\mathbf{F} \\ &= T^{-1} [n^{-1}T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{U}_{nT}\mathbf{M}_T\mathbf{F} - \bar{\sigma}_n^2T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}] \\ &= T^{-1} (n^{-1}T^{-1}\mathbf{A}\mathbf{M}_n\mathbf{A}' - \bar{\sigma}_n^2T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}) = T^{-1}(\mathbf{S} - \mathbf{R}), \end{aligned} \quad (\text{A.21})$$

where $\mathbf{A} = \mathbf{F}'\mathbf{M}_T\mathbf{U}'_{nT} = (a_{ki})$. Denote the (k, k') elements of $\mathbf{S} = n^{-1}T^{-1}\mathbf{A}\mathbf{M}_n\mathbf{A}'$ and $\mathbf{R} = \bar{\sigma}_n^2T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}$ by $s_{kk'}$ and $r_{kk'}$, respectively, and note that

$$r_{kk'} = \left(n^{-1} \sum_{i=1}^n \sigma_i^2 \right) \left[T^{-1} \sum_{t=1}^T (f_{kt} - \bar{f}_k)(f_{k't} - \bar{f}_{k'}) \right] \quad (\text{A.22})$$

and $s_{kk'} = n^{-1}T^{-1} \sum_{i=1}^n (a_{ki} - \bar{a}_k) a_{k'i}$, where $a_{ki} = \sum_{t=1}^T \tilde{f}_{kt} u_{it}$, and $\bar{a}_k = \sum_{t=1}^T \tilde{f}_{kt} \bar{u}_{ot}$, where $\tilde{f}_{kt} = f_{kt} - \bar{f}_k$. Then

$$\begin{aligned}
s_{kk'} &= n^{-1}T^{-1} \sum_{i=1}^n \left[\sum_{t=1}^T \tilde{f}_{kt} (u_{it} - \bar{u}_{ot}) \right] \left[\sum_{t=1}^T \tilde{f}_{k't} u_{it} \right] \\
&= n^{-1}T^{-1} \sum_{t=1}^T \sum_{t'=1}^T \sum_{i=1}^n (u_{it} - \bar{u}_{ot}) u_{it'} \tilde{f}_{kt} \tilde{f}_{k't'} \\
&= n^{-1}T^{-1} \sum_{t=1}^T \sum_{t'=1}^T \sum_{i=1}^n u_{it} u_{it'} \tilde{f}_{kt} \tilde{f}_{k't'} - T^{-1} \sum_{t=1}^T \sum_{t'=1}^T \bar{u}_{ot} \bar{u}_{ot'} \tilde{f}_{kt} \tilde{f}_{k't'}. \tag{A.23}
\end{aligned}$$

Using this result and $r_{kk'}$ given by (A.22) in (A.21) now yields

$$\begin{aligned}
T\delta_{kk'} &= (s_{kk'} - r_{kk'}) \\
&= n^{-1}T^{-1} \sum_{t=1}^T \sum_{t'=1}^T \sum_{i=1}^n u_{it} u_{it'} \tilde{f}_{kt} \tilde{f}_{k't'} - T^{-1} \sum_{t=1}^T \sum_{t'=1}^T \bar{u}_{ot} \bar{u}_{ot'} \tilde{f}_{kt} \tilde{f}_{k't'} \\
&\quad - \bar{\sigma}_n^2 \left(T^{-1} \sum_{t=1}^T \tilde{f}_{kt} \tilde{f}_{k't} \right) - \bar{\sigma}_n^2 \left(T^{-1} \sum_{t=1}^T \tilde{f}_{kt} \tilde{f}_{k't} \right).
\end{aligned}$$

Also writing the first two terms as sums of the elements with $t = t'$ and those with $t \neq t'$, we have

$$\begin{aligned}
T\delta_{kk'} &= n^{-1}T^{-1} \sum_{t=1}^T \sum_{i=1}^n u_{it}^2 \tilde{f}_{kt} \tilde{f}_{k't} - \bar{\sigma}_n^2 \left(T^{-1} \sum_{t=1}^T \tilde{f}_{kt} \tilde{f}_{k't} \right) - T^{-1} \sum_{t=1}^T \bar{u}_{ot}^2 \tilde{f}_{kt} \tilde{f}_{k't} \\
&\quad + n^{-1}T^{-1} \sum_{t \neq t'}^T \sum_{i=1}^n u_{it} u_{it'} \tilde{f}_{kt} \tilde{f}_{k't'} - T^{-1} \sum_{t \neq t'}^T \bar{u}_{ot} \bar{u}_{ot'} \tilde{f}_{kt} \tilde{f}_{k't'} \\
&= T^{-1} \sum_{t=1}^T a_{n,tt} \tilde{f}_{kt} \tilde{f}_{k't} - T^{-1} \sum_{t=1}^T \sum_{t'=1}^T \bar{u}_{ot} \bar{u}_{ot'} \tilde{f}_{kt} \tilde{f}_{k't} + T^{-1} \sum_{t \neq t'}^T a_{n,tt'} \tilde{f}_{kt} \tilde{f}_{k't'}, \\
T\delta_{kk'} &= A_{kk'} - B_{kk'} + C_{kk'}, \tag{A.24}
\end{aligned}$$

where $a_{n,tt} = n^{-1} \sum_{i=1}^n (u_{it}^2 - \sigma_i^2)$, and $a_{n,tt'} = n^{-1} \sum_{i=1}^n u_{it} u_{it'}$. Since u_{it} and $f_{kt'}$ are distributed independently then $E(A_{kk'}) = 0$, and

$$\text{Var}(A_{kk'}) = T^{-2} \sum_{t=1}^T \sum_{t'=1}^T E(a_{n,tt} a_{n,tt'}) E\left(\tilde{f}_{kt} \tilde{f}_{k't} \tilde{f}_{k't} \tilde{f}_{k't'}\right).$$

Since u_{it} is serially independent, then $E(a_{n,tt}a_{n,t't'}) = 0$ for $t \neq t'$ and

$$\text{Var}(A_{kk'}) = T^{-2} \sum_{t=1}^T E(a_{n,tt}^2) E(\tilde{f}_{kt}^2 \tilde{f}_{k't}^2) \leq \left[\sup_{k,k',t} E(\tilde{f}_{kt}^2 \tilde{f}_{k't}^2) \right] \left[T^{-2} \sum_{t=1}^T E(a_{n,tt}^2) \right].$$

$\sup_{k,k',t} E(\tilde{f}_{kt}^2 \tilde{f}_{k't}^2) < C$ since by assumption $\sup_{t,k} E(\tilde{f}_{kt}^4) < C$. Also (using (A.3) of Lemma A.1)

$$E(a_{n,tt}^2) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(u_{it}^2, u_{jt}^2) = O(n^{-1}),$$

and we have $A_{kk'} = O_p(T^{-1}n^{-1/2})$. Now write $B_{kk'}$ as

$$B_{kk'} = T^{-1} \sum_{t=1}^T \sum_{t'=1}^T \bar{u}_{ot} \bar{u}_{ot'} \tilde{f}_{kt} \tilde{f}_{k't} = \left(T^{-1/2} \sum_{t=1}^T \bar{u}_{ot} \tilde{f}_{kt} \right) \left(T^{-1/2} \sum_{t=1}^T \bar{u}_{ot} \tilde{f}_{k't} \right) = q_{ok} q_{ok'},$$

where $q_{ok} = T^{-1/2} \sum_{t=1}^T \bar{u}_{ot} \tilde{f}_{kt}$. Also $E(q_{ok}) = 0$, and $\text{Var}(q_{ok}) = T^{-1} \sum_{t=1}^T E(\bar{u}_{ot}^2) E(\tilde{f}_{kt}^2) = E(\bar{u}_{ot}^2) \left[T^{-1} \sum_{t=1}^T E(\tilde{f}_{kt}^2) \right]$, where $E(\bar{u}_{ot}^2) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} = O(n^{-1})$. Hence, $T^{-1/2} \sum_{t=1}^T \bar{u}_{ot} \tilde{f}_{kt} = O_p(n^{-1/2})$, and $B_{kk'} = O_p(n^{-1})$. Finally, consider $C_{kk'} = T^{-1} \sum_{t \neq t'}^T a_{n,tt'} \tilde{f}_{kt} \tilde{f}_{k't'}$, and note that $E(a_{n,tt'}) = n^{-1} \sum_{i=1}^n E(u_{it} u_{it'}) = 0$ for all $t \neq t'$, which ensures that $E(C_{kk'}) = 0$. Further,

$$\text{Var}(C_{kk'}) = T^{-2} \sum_{t \neq t'}^T \sum_{s \neq s'}^T E(a_{n,tt'} a_{n,ss'}) E(\tilde{f}_{kt} \tilde{f}_{k't'} \tilde{f}_{ks} \tilde{f}_{k's'}) = T^{-2} \sum_{t \neq t'}^T E(a_{n,tt'}^2) E(\tilde{f}_{kt}^2 \tilde{f}_{k't'}^2),$$

$E(a_{n,tt'}^2) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij}^2$. See (A.6) in Lemma A.1. Since by assumption $E(\tilde{f}_{kt}^4) < C$, then we have

$$\text{Var}(C_{kk'}) = \left(n^{-2} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij}^2 \right) \left[T^{-2} \sum_{t \neq t'}^T E(\tilde{f}_{kt}^2 \tilde{f}_{k't'}^2) \right] < C \left(n^{-2} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij}^2 \right),$$

and using (A.2), it follows that $\text{Var}(C_{kk'}) = O(n^{-1})$, and hence $C_{kk'} = O_p(n^{-1/2})$. Using this result and the ones obtained for $A_{kk'}$ and $B_{kk'}$ in (A.24) now yields $\delta_{kk'} = T^{-1} [O_p(T^{-1}n^{-1/2}) + O_p(n^{-1}) + O_p(n^{-1/2})] = O_p(T^{-1}n^{-1/2})$, as required. ■

B Proof of theorems in the paper

B.1 Proof of theorem 1

Consider the two-pass estimator of $\boldsymbol{\lambda}$ defined by (7), which we reproduce here for convenience

$$\hat{\boldsymbol{\lambda}}_{nT} = \left(\hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{B}}_{nT} \right)^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \bar{\mathbf{r}}_{no},$$

where $\hat{\mathbf{B}}_{nT} = (\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2, \dots, \hat{\boldsymbol{\beta}}_n)'$, $\bar{\mathbf{r}}_{no} = (\bar{r}_{1o}, \bar{r}_{2o}, \dots, \bar{r}_{no})'$, $\bar{r}_{io} = T^{-1} \sum_{t=1}^T r_{it}$,

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{r}_{io}, \quad (\text{B.1})$$

and $\mathbf{r}_{io} = (r_{i1}, r_{i2}, \dots, r_{iT})'$. Under the factor model (2)

$$\mathbf{r}_{io} = \mathbf{a}_i \boldsymbol{\tau}_T + \mathbf{F} \boldsymbol{\beta}_i + \mathbf{u}_{io}, \quad (\text{B.2})$$

where $\mathbf{u}_{io} = (u_{i1}, u_{i2}, \dots, u_{iT})'$, and hence

$$\hat{\boldsymbol{\beta}}_i = \boldsymbol{\beta}_i + (\mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{u}_{io}. \quad (\text{B.3})$$

Stacking these results over i yields

$$\hat{\mathbf{B}}_{nT} = \mathbf{B}_n + \mathbf{U}_{nT} \mathbf{G}_T, \quad (\text{B.4})$$

where $\mathbf{U}_{nT} = (\mathbf{u}_{1o}, \mathbf{u}_{2o}, \dots, \mathbf{u}_{no})'$, and

$$\mathbf{G}_T = \mathbf{M}_T \mathbf{F} (\mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1}. \quad (\text{B.5})$$

Also

$$\mathbf{r}_{ot} = \mathbf{a}_n + \mathbf{B}_n \mathbf{f}_t + \mathbf{u}_{ot}, \quad (\text{B.6})$$

where $\mathbf{u}_{ot} = (u_{1t}, u_{2t}, \dots, u_{nt})'$, $\mathbf{a}_n = (a_1, a_2, \dots, a_n)'$. Using (15) in the paper,

$$\mathbf{a}_n = c \boldsymbol{\tau}_n + \mathbf{B}_n \boldsymbol{\phi}_0 + \boldsymbol{\eta}_n, \quad (\text{B.7})$$

and (B.6) can be written as $\mathbf{r}_{ot} = c \boldsymbol{\tau}_n + \mathbf{B}_n (\boldsymbol{\phi}_0 + \mathbf{f}_t) + \mathbf{u}_{ot} + \boldsymbol{\eta}_n$. Now averaging over t yields

$$\bar{\mathbf{r}}_{no} = c \boldsymbol{\tau}_n + \mathbf{B}_n \boldsymbol{\lambda}_T^* + \bar{\mathbf{u}}_{no} + \boldsymbol{\eta}_n, \quad (\text{B.8})$$

where $\bar{\mathbf{r}}_{no} = T^{-1} \sum_{t=1}^T \mathbf{r}_{ot} = (\bar{r}_{1o}, \bar{r}_{2o}, \dots, \bar{r}_{no})'$, $\bar{\mathbf{u}}_{no} = T^{-1} \sum_{t=1}^T \mathbf{u}_{ot} = (\bar{u}_{1o}, \bar{u}_{2o}, \dots, \bar{u}_{no})'$, and

$$\boldsymbol{\lambda}_T^* = \boldsymbol{\phi}_0 + \hat{\boldsymbol{\mu}}_T = \boldsymbol{\lambda}_0 + (\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0). \quad (\text{B.9})$$

Using (B.8) in (7) we have

$$\begin{aligned} \hat{\boldsymbol{\lambda}}_{nT} &= \left(\hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{B}}_{nT} \right)^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n (c\boldsymbol{\tau}_n + \mathbf{B}_n \boldsymbol{\lambda}_T^* + \bar{\mathbf{u}}_{no} + \boldsymbol{\eta}_n) \\ &= \left(\hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{B}}_{nT} \right)^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \left[\hat{\mathbf{B}}_{nT} \boldsymbol{\lambda}_T^* - \left(\hat{\mathbf{B}}_{nT} - \mathbf{B}_n \right) \boldsymbol{\lambda}_T^* + \bar{\mathbf{u}}_{no} + \boldsymbol{\eta}_n \right]. \end{aligned} \quad (\text{B.10})$$

Also using (B.9), and recalling that $\boldsymbol{\lambda}_0 = \boldsymbol{\phi}_0 + \boldsymbol{\mu}_0$, we have $\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_T^* = \hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_0 - (\boldsymbol{\lambda}_T^* - \boldsymbol{\lambda}_0) = \left(\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_0 \right) - (\boldsymbol{\phi}_0 + \hat{\boldsymbol{\mu}}_T - \boldsymbol{\lambda}_0)$, which yields $\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_0 = \hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_T^* + (\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0)$. Furthermore, $\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_T^* = \hat{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0$, where $\hat{\boldsymbol{\phi}}_{nT}$ is the two-step estimator of $\boldsymbol{\phi}_0$ given by (16). This results follows noting that $\hat{\boldsymbol{\lambda}}_{nT} = \hat{\boldsymbol{\phi}}_{nT} + \hat{\boldsymbol{\mu}}_T$, and $\boldsymbol{\lambda}_T^* = \boldsymbol{\phi}_0 + \hat{\boldsymbol{\mu}}_T$. Therefore,

$$\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_0 = \left(\hat{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0 \right) + (\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0). \quad (\text{B.11})$$

We focus on deriving the asymptotic distribution of $\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_T^* = \hat{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0$ since the panel (cross section) dimension does not apply to the second component, $(\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0)$. Now using (B.4) in (B.10) and after some simplifications we have

$$\left(\hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{B}}_{nT} \right) \hat{\boldsymbol{\lambda}}_{nT} = \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \left[\hat{\mathbf{B}}_{nT} \boldsymbol{\lambda}_T^* - \left(\hat{\mathbf{B}}_{nT} - \mathbf{B}_n \right) \boldsymbol{\lambda}_T^* + \bar{\mathbf{u}}_{no} + \boldsymbol{\eta}_n \right],$$

or

$$\left(n^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{B}}_{nT} \right) \left(\hat{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0 \right) = \mathbf{p}_{nT}, \quad (\text{B.12})$$

where

$$\begin{aligned} \mathbf{p}_{nT} &= n^{-1} \mathbf{B}'_n \mathbf{M}_n \boldsymbol{\eta}_n + n^{-1} \mathbf{G}'_T \mathbf{U}'_{nT} \mathbf{M}_n \boldsymbol{\eta}_n + n^{-1} \mathbf{B}'_n \mathbf{M}_n \bar{\mathbf{u}}_{no} + n^{-1} \mathbf{G}'_T \mathbf{U}'_{nT} \mathbf{M}_n \bar{\mathbf{u}}_{no} \\ &\quad - n^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_T \boldsymbol{\lambda}_T^* - n^{-1} \mathbf{G}'_T \mathbf{U}'_{nT} \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_T \boldsymbol{\lambda}_T^*, \end{aligned} \quad (\text{B.13})$$

also

$$\begin{aligned} n^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{B}}_{nT} &= n^{-1} (\mathbf{B}_n + \mathbf{U}_{nT} \mathbf{G}_T)' \mathbf{M}_n (\mathbf{B}_n + \mathbf{U}_{nT} \mathbf{G}_T) \\ &= n^{-1} (\mathbf{B}'_n \mathbf{M}_n \mathbf{B}_n) + n^{-1} (\mathbf{G}'_T \mathbf{U}'_{nT} \mathbf{M}_n \mathbf{B}_n) + \\ &\quad n^{-1} (\mathbf{B}'_n \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_T) + n^{-1} (\mathbf{G}'_T \mathbf{U}'_{nT} \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_T), \end{aligned} \quad (\text{B.14})$$

Now using the results in Lemma A.2 for case where all the observed factors are strong, for a fixed T and as $n \rightarrow \infty$ we have

$$n^{-1} \left(\hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{B}}_{nT} \right) = \boldsymbol{\Sigma}_{\beta\beta} + \bar{\sigma}^2 \mathbf{G}'_T \mathbf{G}_T + o_p(1), \quad (\text{B.15})$$

where, using (B.5),

$$\mathbf{G}'_T \mathbf{G}_T = \frac{1}{T} \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1}. \quad (\text{B.16})$$

Similarly, for the terms on the right hand side of (B.13) we have

$$\mathbf{p}_{nT} = -\frac{\bar{\sigma}^2}{T} \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1} \boldsymbol{\lambda}_T^* + O_p(n^{-1+\alpha_\eta}) + O_p(T^{-1/2} n^{-1/2}) + O_p\left(T^{-1/2} n^{-1+\frac{\alpha_\eta+\alpha_\gamma}{2}}\right).$$

It is now easily seen that for a fixed T , and if $\alpha_\eta < 1$ and $\alpha_\gamma < 1/2$, then as $n \rightarrow \infty$

$$\mathbf{p}_{nT} \rightarrow_p -\frac{\bar{\sigma}^2}{T} \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1} \boldsymbol{\lambda}_T^*.$$

Also, for a fixed T by Assumption 1, $\frac{\bar{\sigma}^2}{T} \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1}$ is a positive definite matrix, and by (B.15) $n^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{B}}_{nT} \rightarrow_p \boldsymbol{\Sigma}_{\beta\beta} + \frac{\bar{\sigma}^2}{T} \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1}$ which is also a positive definite matrix, noting that under Assumption 2 $\boldsymbol{\Sigma}_{\beta\beta}$ is a positive definite matrix. Using these results in (B.12) we now have

$$\hat{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0 = -\frac{\bar{\sigma}^2}{T} \left[\boldsymbol{\Sigma}_{\beta\beta} + \bar{\sigma}^2 \frac{1}{T} \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1} \right]^{-1} \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1} \boldsymbol{\lambda}_T^* + o_p(1), \text{ for a fixed } T \text{ as } n \rightarrow \infty.$$

The bias of estimating λ_0 by the two-step estimator also contains the bias of estimating $\boldsymbol{\mu}_0$. Using the above result in (B.11) we now have (for a fixed T and as $n \rightarrow \infty$)

$$\hat{\boldsymbol{\lambda}}_{nT} - \boldsymbol{\lambda}_0 = (\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0) - \frac{\bar{\sigma}^2}{T} \left[\boldsymbol{\Sigma}_{\beta\beta} + \bar{\sigma}^2 \frac{1}{T} \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1} \right]^{-1} \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1} \boldsymbol{\lambda}_T^* + o_p(1),$$

which establishes Theorem 1.

B.2 Proof of theorem 2

Using the expression for \hat{u}_{it} given by (11), we have $\hat{u}_{it} = a_i - \hat{a}_{iT} - \left(\hat{\beta}_{i,T} - \beta_i \right)' \mathbf{f}_t + u_{it}$. Since \hat{u}_{it} are OLS residuals then for each i , we also have $T^{-1} \sum_{t=1}^T \hat{u}_{it} = 0$, and the above can be written equivalently as $\hat{u}_{it} = u_{it} - \bar{u}_{i\circ} - \left(\hat{\beta}_{i,T} - \beta_i \right)' (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T)$, for $i = 1, 2, \dots, n$, and stacking over i now yields

$$\hat{\mathbf{u}}_t = \mathbf{u}_t - \bar{\mathbf{u}} - \left(\hat{\mathbf{B}}_{nT} - \mathbf{B}_n \right) (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T) = \mathbf{u}_t - \bar{\mathbf{u}} - \mathbf{U}_{nT} \mathbf{G}_T (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T), \quad (\text{B.17})$$

and stacking over t

$$\hat{\mathbf{U}}_{nT} = \mathbf{U}_{nT} \mathbf{M}_T - \mathbf{U}_{nT} \mathbf{G}_T \mathbf{F}' \mathbf{M}_T = \mathbf{U}_{nT} (\mathbf{M}_T - \mathbf{G}_T \mathbf{F}' \mathbf{M}_T).$$

But $\mathbf{G}_T = \mathbf{M}_T \mathbf{F} (\mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1}$, and we have

$$\hat{\mathbf{U}}_{nT} = \mathbf{U}_{nT} \mathbf{R}_T, \quad \mathbf{R}_T = \mathbf{M}_T - \mathbf{M}_T \mathbf{F} (\mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1} \mathbf{F}' \mathbf{M}_T,$$

where $\mathbf{R}_T^2 = \mathbf{R}_T = \mathbf{R}_T'$, $Tr(\mathbf{R}_T) = T - 1 - K$. Then

$$\hat{\sigma}_{nT}^2 = \frac{\sum_{t=1}^T \sum_{i=1}^n \hat{u}_{it}^2}{n(T - K - 1)} = \frac{Tr \left(n^{-1} \hat{\mathbf{U}}_{nT}' \hat{\mathbf{U}}_{nT} \right)}{T - K - 1}.$$

Also

$$\begin{aligned} n^{-1} T^{-1} E [Tr(\mathbf{U}_{nT}' \mathbf{U}_{nT})] &= n^{-1} T^{-1} E \left(\sum_{t=1}^T \sum_{i=1}^n u_{it}^2 \right) = n^{-1} \sum_{i=1}^n \sigma_i^2 = \bar{\sigma}_n^2, \\ E(n^{-1} \mathbf{U}_{nT}' \mathbf{U}_{nT}) &= n^{-1} \sum_{i=1}^n E(\mathbf{u}_{i\circ} \mathbf{u}_{i\circ}') = \bar{\sigma}_n^2 \mathbf{I}_T. \end{aligned}$$

Let $v = T - K - 1$ and note that

$$\begin{aligned} v \hat{\sigma}_{nT}^2 &= Tr \left(n^{-1} \hat{\mathbf{U}}_{nT}' \hat{\mathbf{U}}_{nT} \right) = Tr \left(n^{-1} \mathbf{U}_{nT}' \mathbf{U}_{nT} \mathbf{R}_T \right) = \\ &= Tr \left(n^{-1} \mathbf{U}_{nT}' \mathbf{U}_{nT} \mathbf{M}_T \right) - Tr \left(n^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{U}_{nT}' \mathbf{U}_{nT} \mathbf{M}_T \mathbf{F} (\mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1} \right) \\ &= Tr \left(n^{-1} \mathbf{U}_{nT}' \mathbf{U}_{nT} \right) - T^{-1} \tau_T' \left(n^{-1} \mathbf{U}_{nT}' \mathbf{U}_{nT} \right) \tau_T - Tr \left[\mathbf{Q}' \left(n^{-1} \mathbf{U}_{nT}' \mathbf{U}_{nT} \right) \mathbf{Q} \right] \end{aligned}$$

where $\mathbf{Q} = \mathbf{M}_T \mathbf{F} (T^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1/2}$. Consider the first term and note that

$$n^{-1} T r (\mathbf{U}'_{nT} \mathbf{U}_{nT}) = \sum_{t=1}^T \left[n^{-1} \sum_{i=1}^n (u_{it}^2 - \sigma_i^2) \right] + T \bar{\sigma}_n^2. \quad (\text{B.18})$$

Similarly

$$\begin{aligned} T^{-1} \tau'_T (n^{-1} \mathbf{U}'_{nT} \mathbf{U}_{nT}) \tau_T &= T^{-1} n^{-1} \tau'_T [\mathbf{U}'_{nT} \mathbf{U}_{nT} - E(\mathbf{U}'_{nT} \mathbf{U}_{nT})] \tau_T + \bar{\sigma}_n^2, \\ T r [\mathbf{Q}' (n^{-1} \mathbf{U}'_{nT} \mathbf{U}_{nT}) \mathbf{Q}] &= T r [\mathbf{Q}' [\mathbf{U}'_{nT} \mathbf{U}_{nT} - E(\mathbf{U}'_{nT} \mathbf{U}_{nT})] \mathbf{Q}] + K \bar{\sigma}_n^2. \end{aligned}$$

Hence

$$\widehat{\sigma}_{nT}^2 - \bar{\sigma}_n^2 = (T/v) (a_{nT} + b_{nT} + c_{nT}), \quad (\text{B.19})$$

where $a_{nT} = T^{-1} \sum_{t=1}^T [n^{-1} \sum_{i=1}^n (u_{it}^2 - \sigma_i^2)]$, $b_{nT} = T^{-2} n^{-1} \tau'_T [\mathbf{U}'_{nT} \mathbf{U}_{nT} - E(\mathbf{U}'_{nT} \mathbf{U}_{nT})] \tau_T$, and $c_{nT} = T^{-1} n^{-1} T r [\mathbf{Q}' [\mathbf{U}'_{nT} \mathbf{U}_{nT} - E(\mathbf{U}'_{nT} \mathbf{U}_{nT})] \mathbf{Q}]$. Due to the independence of \mathbf{F} and \mathbf{U}_{nT} , we have $E(a_{nT}) = 0$, $E(b_{nT}) = 0$ and $E(c_{nT}) = 0$, and for any fixed n and T $E(\widehat{\sigma}_{nT}^2 - \bar{\sigma}_n^2) = 0$, and for a fixed T , $\lim_{n \rightarrow \infty} E(\widehat{\sigma}_{nT}^2) = \bar{\sigma}^2$, and result (45) follows. To establish the probability order of $\widehat{\sigma}_{nT}^2 - \bar{\sigma}_n^2$, we consider the probability orders of a_{nT} , b_{nT} , and c_{nT} in turn, noting that that $T/v = T/(T - K - 1) = O(1)$. For a_{nT} , using result (A.3) in Lemma A.1, and noting that u_{it} are serially independent we have

$$a_{nT} = O_p(T^{-1/2} n^{-1/2}). \quad (\text{B.20})$$

Consider now b_{nT} and note that $b_{nT} = T^{-2} n^{-1} \sum_{i=1}^n [\tau'_T \mathbf{u}_{i\circ} \mathbf{u}'_{i\circ} \tau_T - E(\tau'_T \mathbf{u}_{i\circ} \mathbf{u}'_{i\circ} \tau_T)]$. Also $\tau'_T \mathbf{u}_{i\circ} \mathbf{u}'_{i\circ} \tau_T = \sum_{t=1}^T \sum_{t'=1}^T u_{it} u_{it'}$, and

$$\begin{aligned} b_{nT} &= T^{-2} n^{-1} \sum_{i=1}^n \sum_{t=1}^T \sum_{t'=1}^T [u_{it} u_{it'} - E(u_{it} u_{it'})] \\ &= T^{-2} \sum_{\substack{t=1 \\ t \neq t'}}^T n^{-1} \sum_{i=1}^n (u_{it}^2 - \sigma_i^2) + T^{-2} \sum_{t \neq t'}^T \left(n^{-1} \sum_{i=1}^n u_{it} u_{it'} \right) \\ &= T^{-2} \sum_{t=1}^T a_{n,tt} + T^{-2} \sum_{t \neq t'}^T a_{n,tt'}, \end{aligned}$$

where $a_{n,tt}$ and $a_{n,tt'}$ are both shown in Lemma A.1 to be $O_p(n^{-1/2})$. See equations (A.3) and (A.5). Therefore, given that $a_{n,tt}$ and $a_{n,tt'}$ with $t \neq t'$ are also distributed

independently over t we have

$$b_{nT} = O_p(n^{-1/2}T^{-1/2}). \quad (\text{B.21})$$

Denote the k^{th} column of \mathbf{Q} by $\mathbf{q}_k = (q_{k1}, q_{k2}, \dots, q_{kT})'$ (a $T \times 1$ vector) and write c_{nT} as

$$\begin{aligned} c_{nT} &= \sum_{k=1}^K T^{-1} \left[n^{-1} \sum_{i=1}^n \mathbf{q}'_k [\mathbf{u}_{i\circ} \mathbf{u}'_{i\circ} - E(\mathbf{u}_{i\circ} \mathbf{u}'_{i\circ})] \mathbf{q}_k \right] \\ &= \sum_{k=1}^K T^{-1} \left[n^{-1} \sum_{i=1}^n \sum_{t=1}^T \sum_{t'=1}^T q_{kt} q_{kt'} [u_{it} u_{it'} - E(u_{it} u_{it'})] \right]. \end{aligned}$$

Consider the k^{th} term of the above sum, and note that

$$c_{nT,k} = T^{-1} n^{-1} \sum_{i=1}^n \sum_{t=1}^T \sum_{t'=1}^T q_{kt} q_{kt'} [u_{it} u_{it'} - E(u_{it} u_{it'})] = T^{-1} \sum_{t=1}^T \sum_{t'=1}^T q_{kt} q_{kt'} a_{n,tt'},$$

where $a_{n,tt'} = n^{-1} \sum_{i=1}^n [u_{it} u_{it'} - E(u_{it} u_{it'})]$. is defined by (A.3) and (A.5) in Lemma A.1, with $\text{Var}(a_{n,tt'}) = O(n^{-1})$ for all t and t' . Also $a_{n,tt'}$ and $a_{n,ss'}$ are distributed independently if t or t' differ from s or s' . Therefore, for all t and t' for all t

$$\text{Var}(c_{nT,k}) = T^{-2} \sum_{t=1}^T \sum_{t'=1}^T q_{kt}^2 q_{kt'}^2 \text{Var}(a_{n,tt'}) \leq O(n^{-1}) \left(T^{-2} \sum_{t=1}^T q_{kt}^2 \right).$$

But it is easily verified that $\mathbf{Q}'\mathbf{Q} = \mathbf{I}_K$ which yields $\sum_{t=1}^T q_{kt}^2 = 1$, and $\text{Var}(c_{nT,k}) = O(T^{-2}n^{-1})$. Hence, $c_{nT,k} = O_p(T^{-1}n^{-1/2})$, for $k = 1, 2, \dots, K$, which establishes that $c_{nT} = O_p(T^{-1}n^{-1/2})$. The order result in (46) now follows using this result, (B.20) and (B.21) in (B.19).

B.3 Proof of theorem 3

The bias-corrected estimator of ϕ_0 is given by (17) which we reproduce here and re-write as

$$\mathbf{H}_{nT} \left(\tilde{\phi}_{nT} - \phi_0 \right) = \mathbf{s}_{nT}, \quad (\text{B.22})$$

where

$$\mathbf{s}_{nT} = \frac{\hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{a}}_{nT}}{n} + T^{-1} \hat{\sigma}_{nT}^2 \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1} \hat{\boldsymbol{\mu}}_T - \mathbf{H}_{nT} \boldsymbol{\phi}_0, \quad (\text{B.23})$$

$$\mathbf{H}_{nT} = \frac{\hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{B}}_{nT}}{n} - T^{-1} \hat{\sigma}_{nT}^2 \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1}. \quad (\text{B.24})$$

Also $\hat{\mathbf{a}}_{nT} = \bar{\mathbf{r}}_{no} - \hat{\mathbf{B}}_{nT} \hat{\boldsymbol{\mu}}_T$, and $\bar{\mathbf{r}}_{no} = c\boldsymbol{\tau}_n + \mathbf{B}_n \boldsymbol{\lambda}_T^* + \bar{\mathbf{u}}_{no} + \boldsymbol{\eta}_n$ (see (B.8)). Using these results and noting that $\boldsymbol{\lambda}_T^* = \boldsymbol{\lambda}_0 + (\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0) = \boldsymbol{\phi}_0 + \hat{\boldsymbol{\mu}}_T$, we have

$$\hat{\mathbf{a}}_{nT} = c\boldsymbol{\tau}_n + \mathbf{B}_n \boldsymbol{\phi}_0 + \bar{\mathbf{u}}_{no} + \boldsymbol{\eta}_n - \left(\hat{\mathbf{B}}_{nT} - \mathbf{B}_n \right) \hat{\boldsymbol{\mu}}_T$$

and $\bar{\mathbf{u}}_{no} = (\bar{u}_{1o}, \bar{u}_{2o}, \dots, \bar{u}_{no})'$. Then,

$$\begin{aligned} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{a}}_{nT} &= \left(\hat{\mathbf{B}}'_{nT} \mathbf{M}_n \mathbf{B}_n \right) \boldsymbol{\phi}_0 + \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \bar{\mathbf{u}}_{no} \\ &\quad + \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \boldsymbol{\eta}_n - \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \left(\hat{\mathbf{B}}_{nT} - \mathbf{B}_n \right) \hat{\boldsymbol{\mu}}_T. \end{aligned} \quad (\text{B.25})$$

Also

$$\hat{\mathbf{B}}_{nT} = \mathbf{B}_n + \mathbf{U}_{nT} \mathbf{G}_T, \quad (\text{B.26})$$

where $\mathbf{U}_{nT} = (\mathbf{u}_{1o}, \mathbf{u}_{2o}, \dots, \mathbf{u}_{no})'$ and \mathbf{G}_T is defined by (B.16). Using these results together with (B.24), the right hand side of (B.22) can be written as

$$\begin{aligned} \mathbf{s}_{nT} &= \frac{\mathbf{B}'_n \mathbf{M}_n \boldsymbol{\eta}_n}{n} + \frac{\mathbf{G}'_T \mathbf{U}'_{nT} \mathbf{M}_n \boldsymbol{\eta}_n}{n} + \frac{\mathbf{G}'_T \mathbf{U}'_{nT} \mathbf{M}_n \bar{\mathbf{u}}_{no}}{n} - \frac{\mathbf{B}'_n \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_T \boldsymbol{\lambda}_T^*}{n} \\ &\quad - \mathbf{G}'_T \left(\frac{\mathbf{U}'_{nT} \mathbf{M}_n \mathbf{U}_{nT}}{n} - \hat{\sigma}_{nT}^2 \right) \mathbf{G}_T \boldsymbol{\lambda}_T^*, \end{aligned}$$

where the last term can be decomposed as

$$\mathbf{G}'_T \left(\frac{\mathbf{U}'_{nT} \mathbf{M}_n \mathbf{U}_{nT}}{n} - \hat{\sigma}_{nT}^2 \right) \mathbf{G}_T \boldsymbol{\lambda}_T^* = \mathbf{G}'_T \left(\frac{\mathbf{U}'_{nT} \mathbf{M}_n \mathbf{U}_{nT}}{n} - \bar{\sigma}_n^2 \right) \mathbf{G}_T \boldsymbol{\lambda}_T^* - \left(\hat{\sigma}_{nT}^2 - \bar{\sigma}_n^2 \right) \mathbf{G}'_T \mathbf{G}_T \boldsymbol{\lambda}_T^*.$$

Similarly, using (B.14), we have

$$\begin{aligned} \mathbf{H}_{nT} &= n^{-1} (\mathbf{B}'_n \mathbf{M}_n \mathbf{B}_n) + n^{-1} (\mathbf{G}'_T \mathbf{U}'_{nT} \mathbf{M}_n \mathbf{B}_n) + \\ &\quad n^{-1} (\mathbf{B}'_n \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_T) + \mathbf{G}'_T \left(\frac{\mathbf{U}'_{nT} \mathbf{M}_n \mathbf{U}_{nT}}{n} - \bar{\sigma}_n^2 \right) \mathbf{G}_T - T^{-1} \left(\hat{\sigma}_{nT}^2 - \bar{\sigma}_n^2 \right) \left(\frac{\mathbf{F}' \mathbf{M}_T \mathbf{F}}{T} \right)^{-1}. \end{aligned}$$

Using Theorem 2 and since by assumption $T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F}$ is positive definite, then

$$\left(\widehat{\sigma}_{nT}^2 - \bar{\sigma}_n^2\right) \left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1} = O_p\left(T^{-1/2}n^{-1/2}\right).$$

Also using results in Lemma A.2, we have

$$\mathbf{H}_{nT} = n^{-1}(\mathbf{B}'_n\mathbf{M}_n\mathbf{B}_n) + O_p\left(T^{-1/2}n^{-1/2}\right). \quad (\text{B.27})$$

Hence, $\mathbf{H}_{nT} \rightarrow_p \boldsymbol{\Sigma}_{\beta\beta}$ for a fixed T as $n \rightarrow \infty$, so long as $\alpha_\gamma < 1/2$. Note also that by Assumption $\boldsymbol{\Sigma}_{\beta\beta}$ is positive definite. Further

$$\begin{aligned} \mathbf{s}_{nT} &= n^{-1}(\mathbf{B}'_n\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} - \mathbf{B}'_n\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T\boldsymbol{\lambda}_T^*) + n^{-1}(\mathbf{B}'_n + \mathbf{G}'_T\mathbf{U}'_{nT})\mathbf{M}_n\boldsymbol{\eta}_n + n^{-1}\mathbf{G}'_T\mathbf{U}'_{nT}\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} \\ &\quad - \mathbf{G}'_T\left(\frac{\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{U}_{nT}}{n} - \bar{\sigma}_n^2\right)\mathbf{G}_T\boldsymbol{\lambda}_T^* + T^{-1}\left(\widehat{\sigma}_{nT}^2 - \bar{\sigma}_n^2\right)\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}_T^*. \end{aligned} \quad (\text{B.28})$$

Using (A.12) and (A.15) (in Lemma A.2) we have

$$n^{-1}\mathbf{B}'_n\mathbf{M}_n\boldsymbol{\eta}_n + n^{-1}\mathbf{G}'_T\mathbf{U}'_{nT}\mathbf{M}_n\boldsymbol{\eta}_n = O_p\left(n^{-1+\alpha_\eta}\right) + O_p\left(T^{-1/2}n^{-1+\frac{\alpha_\eta+\alpha_\gamma}{2}}\right),$$

and (A.16) and (A.17)

$$n^{-1}\mathbf{G}'_T\mathbf{U}'_{nT}\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} - \mathbf{G}'_T\left(\frac{\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{U}_{nT}}{n} - \bar{\sigma}_n^2\right)\mathbf{G}_T\boldsymbol{\lambda}_T^* = O_p\left(T^{-1}n^{-1/2}\right).$$

Further by (46)

$$T^{-1}\left(\widehat{\sigma}_{nT}^2 - \bar{\sigma}_n^2\right)\left(\frac{\mathbf{F}'\mathbf{M}_T\mathbf{F}}{T}\right)^{-1}\boldsymbol{\lambda}_T^* = O_p\left(T^{-3/2}n^{-1/2}\right). \quad (\text{B.29})$$

Hence

$$\begin{aligned} \mathbf{s}_{nT} &= n^{-1}(\mathbf{B}'_n\mathbf{M}_n\bar{\mathbf{u}}_{n\circ} - \mathbf{B}'_n\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_T\boldsymbol{\lambda}_T^*) + O_p\left(n^{-1+\alpha_\eta}\right) \\ &\quad + O_p\left(T^{-1/2}n^{-1+\frac{\alpha_\eta+\alpha_\gamma}{2}}\right) + O_p\left(T^{-1}n^{-1/2}\right) + O_p\left(T^{-3/2}n^{-1/2}\right). \end{aligned} \quad (\text{B.30})$$

Using this result and (B.27) in (B.22) now yields (47), as required. To derive the asymptotic distribution of $\tilde{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0$ since by assumption $\alpha_\eta + \alpha_\gamma < 1$, then the dominant

term of \mathbf{s}_{nT} is given by

$$n^{-1} (\mathbf{B}'_n \mathbf{M}_n \bar{\mathbf{u}}_{n\circ} - \mathbf{B}'_n \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_T \boldsymbol{\lambda}_T^*) = O_p(T^{-1/2} n^{-1/2}), \quad (\text{B.31})$$

and to ensure that we end up with a non-degenerate, stable limiting distribution, $(\tilde{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0)$ needs to be scaled by \sqrt{nT} with n and $T \rightarrow \infty$, jointly. To this end we first note that when T is fixed $\mathbf{H}_{nT} \rightarrow_p \boldsymbol{\Sigma}_{\beta\beta}$, as $n \rightarrow \infty$ and we have

$$\sqrt{nT} (\tilde{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0) \stackrel{a}{\sim} \boldsymbol{\Sigma}_{\beta\beta}^{-1} (\sqrt{nT} \mathbf{s}_{nT}). \quad (\text{B.32})$$

Using (B.30)

$$\sqrt{nT} \mathbf{s}_{nT} = \boldsymbol{\xi}_{nT} + O_p\left(n^{-\frac{1}{2} + \frac{\alpha_n + \alpha_\gamma}{2}}\right) + O_p(T^{1/2} n^{-1/2 + \alpha_n}) + O_p(T^{-1/2}),$$

where

$$\boldsymbol{\xi}_{nT} = T^{1/2} n^{-1/2} (\mathbf{B}'_n \mathbf{M}_n \bar{\mathbf{u}}_{n\circ} - \mathbf{B}'_n \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_T \boldsymbol{\lambda}_T^*). \quad (\text{B.33})$$

Using the above results in (B.32) we now have

$$\sqrt{nT} (\tilde{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0) \stackrel{a}{\sim} \boldsymbol{\Sigma}_{\beta\beta}^{-1} \left[\boldsymbol{\xi}_{nT} + O_p\left(n^{-\frac{1}{2} + \frac{\alpha_n + \alpha_\gamma}{2}}\right) + O_p(T^{1/2} n^{-1/2 + \alpha_n}) + O_p(T^{-1/2}) \right]. \quad (\text{B.34})$$

Therefore, when condition $(T/n)^{1/2} n^{\alpha_n} \rightarrow 0$, as n and $T \rightarrow \infty$, is met we have

$$\sqrt{nT} (\tilde{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0) \stackrel{a}{\sim} \boldsymbol{\Sigma}_{\beta\beta}^{-1} \boldsymbol{\xi}_{nT} + o_p(1).$$

To derive the asymptotic distribution of $\boldsymbol{\xi}_{nT}$, we note that $\bar{\mathbf{u}}_{n\circ} = T^{-1} \mathbf{U}_{nT} \boldsymbol{\tau}_T$, and $\mathbf{G}_T \boldsymbol{\lambda}_T^* = T^{-1} \mathbf{M}_T \mathbf{F} (T^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1} \boldsymbol{\lambda}_T^*$. Then, using these results in (B.33)

$$\boldsymbol{\xi}_{nT} = (\xi_{k,nT}) = n^{-1/2} T^{-1/2} \mathbf{B}'_n \mathbf{M}_n \mathbf{U}_{nT} \mathbf{a}_T, \quad (\text{B.35})$$

where $\mathbf{a}_T = \boldsymbol{\tau}_T - \mathbf{M}_T \mathbf{F} (T^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1} \boldsymbol{\lambda}_T^* = (a_t)$. Also

$$s_{a,T}^2 = T^{-1} \sum_{t=1}^T a_t^2 = T^{-1} \mathbf{a}'_T \mathbf{a}_T = 1 + \boldsymbol{\lambda}_T^{*'} (T^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1} \boldsymbol{\lambda}_T^*, \quad (\text{B.36})$$

where $\boldsymbol{\lambda}_T^* = \boldsymbol{\phi}_0 + \hat{\boldsymbol{\mu}}_T = \boldsymbol{\lambda}_0 + (\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0)$, and $(\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}_0) = O_p(T^{-1/2})$. Further

$$s_{a,T}^2 \geq 1 \text{ and } s_{a,T}^2 \leq 1 + (\boldsymbol{\lambda}_T^* \boldsymbol{\lambda}_T^*) \lambda_{\max} [(T^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1}] < C, \quad (\text{B.37})$$

and $s_a^2 = \lim_{T \rightarrow \infty} s_{a,T}^2 = 1 + \boldsymbol{\lambda}'_0 \boldsymbol{\Sigma}_f^{-1} \boldsymbol{\lambda}'_0$. The k^{th} element of $\boldsymbol{\xi}_{nT}$ is given by $\xi_{k,nT} = n^{-1/2} T^{-1/2} \sum_{i=1}^n \sum_{t=1}^T a_t (\beta_{ik} - \bar{\beta}_k) u_{it}$, and using (27) we have

$$\begin{aligned} \xi_{k,nT} &= \left(T^{-1/2} \sum_{t=1}^T a_t g_t \right) \left[n^{-1/2} \sum_{i=1}^n (\beta_{ik} - \bar{\beta}_k) \gamma_i \right] \\ &\quad + n^{-1/2} T^{-1/2} \sum_{i=1}^n \sum_{t=1}^T a_t (\beta_{ik} - \bar{\beta}_k) v_{it}. \end{aligned}$$

Under Assumption 3 g_t is distributed independently of \mathbf{f}_t (and hence of a_t), as well as being serially independent. Also $\text{Var}(T^{-1/2} \sum_{t=1}^T a_t g_t) = s_{a,T}^2$ (recall that $E(g_t) = 0$ and

$E(g_t^2) = 1$), and we have $T^{-1/2} \sum_{t=1}^T a_t g_t = O_p(1)$. Further $E \left| n^{-1/2} \sum_{i=1}^n (\beta_{ik} - \bar{\beta}_k) \gamma_i \right| \leq \sup_{i,k} E |\beta_{ik} - \bar{\beta}_k| (n^{-1/2} \sum_{i=1}^n |\gamma_i|) = O(n^{-1/2+\alpha\gamma})$. Hence

$$\begin{aligned} \xi_{k,nT} &= n^{-1/2} T^{-1/2} \sum_{t=1}^T \sum_{i=1}^n a_t (\beta_{ik} - \bar{\beta}_k) v_{it} + O(n^{-1/2+\alpha\gamma}) \\ &= T^{-1/2} \sum_{t=1}^T a_t h_{nt} + O(n^{-1/2+\alpha\gamma}), \end{aligned} \quad (\text{B.38})$$

where $h_{nt} = n^{-1/2} \sum_{i=1}^n (\beta_{ik} - \bar{\beta}_k) v_{it}$. Under Assumption 2, $h_{nt} = n^{-1/2} \sum_{i=1}^n (\beta_{ik} - \bar{\beta}_k) v_{it} \rightarrow_d N(0, \omega_k^2)$, for $k = 1, 2, \dots, K$, where

$$\omega_k^2 = p \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \sum_{j=1}^n (\beta_{ik} - \bar{\beta}_k) (\beta_{jk} - \bar{\beta}_k) \sigma_{v,ij} > 0,$$

and $\omega_k^2 \leq \sup_{i,k} E(\beta_{ik} - \bar{\beta}_k)^2 \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \sum_{j=1}^n |\sigma_{v,ij}| < C$. Also, since v_{it} are serially independent then there exists T_0 such that for any fixed $T > T_0$ and as $n \rightarrow \infty$

$$T^{-1/2} \sum_{t=1}^T a_t h_{nt} \rightarrow_d N(0, \omega_k^2 (1 + s_{aT}^2)),$$

where s_{aT}^2 is defined by (B.36). Using this result in (B.38) and noting that $\alpha_\gamma < 1/2$, we also have for any fixed T and as $n \rightarrow \infty$,

$$\xi_{k,nT} \rightarrow_d N\left(0, \omega_k^2 (1 + s_{aT}^2)\right), \text{ for a fixed } T > T_0 \text{ and as } n \rightarrow \infty.$$

This result extends readily to the case where n and $T \rightarrow \infty$, jointly. In this case

$$\xi_{k,nT} \rightarrow_d N\left(0, \omega_k^2 (1 + s_a^2)\right), \text{ where } s_a^2 = 1 + \boldsymbol{\lambda}'_0 \boldsymbol{\Sigma}_f^{-1} \boldsymbol{\lambda}_0.$$

Similarly, We have (using $u_{it} = \gamma_i g_t + v_{it}$)

$$\begin{aligned} Cov(\xi_{k,nT}, \xi_{k',nT}) &= n^{-1} T^{-1} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^T a_t^2 (\beta_{ik} - \bar{\beta}_k) (\beta_{jk'} - \bar{\beta}_{k'}) E(u_{it} u_{jt}) = \\ &= (1 + s_{aT}^2) \left[n^{-1} \sum_{i=1}^n \sum_{j=1}^n \gamma_i \gamma_j (\beta_{ik} - \bar{\beta}_k) (\beta_{jk'} - \bar{\beta}_{k'}) \right] \\ &+ (1 + s_{aT}^2) \left[n^{-1} \sum_{i=1}^n \sum_{j=1}^n \sigma_{v,ij} (\beta_{ik} - \bar{\beta}_k) (\beta_{jk'} - \bar{\beta}_{k'}) \right]. \end{aligned}$$

But

$$\begin{aligned} E \left| n^{-1} \sum_{i=1}^n \sum_{j=1}^n \gamma_i \gamma_j (\beta_{ik} - \bar{\beta}_k) (\beta_{jk'} - \bar{\beta}_{k'}) \right| &\leq \sup_{i,k,k'} |(\beta_{ik} - \bar{\beta}_k) (\beta_{jk'} - \bar{\beta}_{k'})| E(\beta_{ik} - \bar{\beta}_k)^2 \left(n^{-1/2} \sum_{j=1}^n |\gamma_j| \right) \\ &\leq \sup_{i,k} E(\beta_{ik} - \bar{\beta}_k)^2 \left(n^{-1/2} \sum_{j=1}^n |\gamma_j| \right)^2 = O(n^{-1+2\alpha_\gamma}). \end{aligned}$$

and similarly

$$\left| n^{-1} \sum_{i=1}^n \sum_{j=1}^n \sigma_{v,ij} (\beta_{ik} - \bar{\beta}_k) (\beta_{jk'} - \bar{\beta}_{k'}) \right| \leq \sup_{i,k} E(\beta_{ik} - \bar{\beta}_k)^2 n^{-1} \sum_{i=1}^n \sum_{j=1}^n |\sigma_{v,ij}| < C.$$

Hence $Cov(\xi_{k,nT}, \xi_{k',nT}) < C$, for all k and k' . Using the above results in (B.35), and noting that K is fixed, we have $\boldsymbol{\xi}_{nT} \rightarrow_d N(\mathbf{0}, \mathbf{V}_\xi)$, as n and $T \rightarrow \infty$, where $\mathbf{V}_\xi = (1 + \boldsymbol{\lambda}'_0 \boldsymbol{\Sigma}_f^{-1} \boldsymbol{\lambda}_0) p \lim_{n \rightarrow \infty} (n^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{V}_u \mathbf{M}_n \mathbf{B}_n)$, noting that $s_{aT}^2 \rightarrow_p 1 + \boldsymbol{\lambda}'_0 \boldsymbol{\Sigma}_f^{-1} \boldsymbol{\lambda}_0$, where s_{aT}^2 is given by (B.36). Also recall from (B.34) that $\sqrt{nT} (\tilde{\boldsymbol{\phi}}_{nT} - \boldsymbol{\phi}_0) = \boldsymbol{\Sigma}_{\beta\beta}^{-1} \boldsymbol{\xi}_{nT} + O_p\left(n^{-\frac{1}{2} + \frac{\alpha_\eta + \alpha_\gamma}{2}}\right) + O_p(T^{1/2} n^{-1/2 + \alpha_\eta}) + O_p(T^{-1/2})$. Hence, result (51) follows since by assumption $\alpha_\eta < 1/2$, $\alpha_\gamma < 1/2$, and $(T/n)^{1/2} n^{\alpha_\gamma} \rightarrow 0$.

B.4 Proof of theorem 4

Using (55) and (54) and replacing $T^{-1} (T^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{F})^{-1}$ by $\mathbf{G}'_T \mathbf{G}_T$ we have (see (B.5))

$$\mathbf{H}_{nT}(\boldsymbol{\alpha}) \mathbf{D}_\alpha \left(\tilde{\boldsymbol{\phi}}_{nT}(\boldsymbol{\alpha}) - \boldsymbol{\phi}_0 \right) = \mathbf{q}_{nT}(\boldsymbol{\alpha}), \quad (\text{B.39})$$

where $\mathbf{q}_{nT}(\boldsymbol{\alpha}) = \mathbf{D}_\alpha^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{a}}_{nT} + n \hat{\sigma}_{nT}^2 \mathbf{D}_\alpha^{-1} \mathbf{G}'_T \mathbf{G}_T \hat{\boldsymbol{\mu}}_T - \mathbf{H}_{nT}(\boldsymbol{\alpha}) \mathbf{D}_\alpha \boldsymbol{\phi}_0$, and $\mathbf{H}_{nT}(\boldsymbol{\alpha}) = \mathbf{D}_\alpha^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \hat{\mathbf{B}}_{nT} \mathbf{D}_\alpha^{-1} - n \hat{\sigma}_{nT}^2 \mathbf{D}_\alpha^{-1} \mathbf{G}'_T \mathbf{G}_T \mathbf{D}_\alpha^{-1}$. But

$$\mathbf{H}_{nT}(\boldsymbol{\alpha}) = n \mathbf{D}_\alpha^{-1} \mathbf{H}_{nT} \mathbf{D}_\alpha^{-1}, \quad \text{and} \quad \mathbf{q}_{nT}(\boldsymbol{\alpha}) = n \mathbf{D}_\alpha^{-1} \mathbf{s}_{nT}, \quad (\text{B.40})$$

where \mathbf{s}_{nT} and \mathbf{H}_{nT} are already defined by (B.23) and (B.24). Consider first the limiting property of $\mathbf{H}_{nT}(\boldsymbol{\alpha})$, and using (B.26) note that

$$\mathbf{H}_{nT}(\boldsymbol{\alpha}) = \mathbf{D}_\alpha^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{B}_n \mathbf{D}_\alpha^{-1} + \mathbf{D}_\alpha^{-1} \mathbf{G}'_T \mathbf{U}'_{nT} \mathbf{M}_n \mathbf{B}_n \mathbf{D}_\alpha^{-1} + \mathbf{D}_\alpha^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_T \mathbf{D}_\alpha^{-1} + \mathbf{D}_\alpha^{-1} \mathbf{G}'_T \mathbf{U}'_{nT} \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_T \mathbf{D}_\alpha^{-1} - n \hat{\sigma}_{nT}^2 \mathbf{D}_\alpha^{-1} \mathbf{G}'_T \mathbf{G}_T \mathbf{D}_\alpha^{-1},$$

or

$$\mathbf{H}_{nT}(\boldsymbol{\alpha}) = \mathbf{D}_\alpha^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{B}_n \mathbf{D}_\alpha^{-1} + \mathbf{D}_\alpha^{-1} \mathbf{G}'_T \mathbf{U}'_{nT} \mathbf{M}_n \mathbf{B}_n \mathbf{D}_\alpha^{-1} + \mathbf{D}_\alpha^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_T \mathbf{D}_\alpha^{-1} + n \mathbf{D}_\alpha^{-1} \left[\mathbf{G}'_T (n^{-1} \mathbf{U}'_{nT} \mathbf{M}_n \mathbf{U}_{nT}) \mathbf{G}_T - \hat{\sigma}_{nT}^2 \mathbf{G}'_T \mathbf{G}_T \right] \mathbf{D}_\alpha^{-1}.$$

Further

$$\begin{aligned} & n \mathbf{D}_\alpha^{-1} \left[\mathbf{G}'_T (n^{-1} \mathbf{U}'_{nT} \mathbf{M}_n \mathbf{U}_{nT}) \mathbf{G}_T - \hat{\sigma}_{nT}^2 \mathbf{G}'_T \mathbf{G}_T \right] \mathbf{D}_\alpha^{-1} \\ &= n \mathbf{D}_\alpha^{-1} \left[\mathbf{G}'_T (n^{-1} \mathbf{U}'_{nT} \mathbf{M}_n \mathbf{U}_{nT}) \mathbf{G}_T - \bar{\sigma}_n^2 \mathbf{G}'_T \mathbf{G}_T \right] \mathbf{D}_\alpha^{-1} - \left(\hat{\sigma}_{nT}^2 - \bar{\sigma}_n^2 \right) n \mathbf{D}_\alpha^{-1} \mathbf{G}'_T \mathbf{G}_T \mathbf{D}_\alpha^{-1}. \end{aligned}$$

But by (46) $\hat{\sigma}_{nT}^2 - \bar{\sigma}_n^2 = O_p(T^{-1/2} n^{-1/2})$, and $\|\mathbf{D}_\alpha^{-1}\| = \lambda_{max}^{1/2}(\mathbf{D}_\alpha^{-2}) = n^{-\alpha_{\min}/2}$. Then, using results in Lemma A.2 we have

$$\left\| \mathbf{D}_\alpha^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_T \mathbf{D}_\alpha^{-1} \right\| \leq n \left\| \mathbf{D}_\alpha^{-1} \right\|^2 \left\| n^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_T \right\| = O_p \left(T^{-1/2} n^{-\alpha_{\min} + 1/2} \right),$$

and

$$\begin{aligned} & \left\| n \mathbf{D}_\alpha^{-1} \mathbf{G}'_T \left(n^{-1} \mathbf{U}'_{nT} \mathbf{M}_n \mathbf{U}_{nT} - \hat{\sigma}_{nT}^2 \mathbf{I}_T \right) \mathbf{G}_T \mathbf{D}_\alpha^{-1} \right\| \\ & \leq n \left\| \mathbf{D}_\alpha^{-1} \right\| \left\| \mathbf{G}'_T \left(n^{-1} \mathbf{U}'_{nT} \mathbf{M}_n \mathbf{U}_{nT} - \hat{\sigma}_{nT}^2 \mathbf{I}_T \right) \mathbf{G}_T \right\| = O_p \left(T^{-1} n^{-\alpha_{\min} + 1/2} \right). \end{aligned}$$

Hence

$$\mathbf{H}_{nT}(\boldsymbol{\alpha}) = \mathbf{D}_\alpha^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{B}_n \mathbf{D}_\alpha^{-1} + O_p(T^{-1} n^{-\alpha_{\min} + 1/2}),$$

and $\mathbf{H}_{nT}(\boldsymbol{\alpha}) \rightarrow_p \boldsymbol{\Sigma}_{\beta\beta}(\boldsymbol{\alpha})$, as $n \rightarrow \infty$, for a fixed T , so long as $\alpha_{\min} > 1/2 > \alpha_\gamma$. By Assumption 2, $\lim_{n \rightarrow \infty} (\mathbf{D}_\alpha^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{B}_n \mathbf{D}_\alpha^{-1}) = \boldsymbol{\Sigma}_{\beta\beta}(\boldsymbol{\alpha})$ is a positive definite matrix. Using this result in (B.39) we have

$$\mathbf{D}_\alpha \left(\tilde{\boldsymbol{\phi}}_{nT}(\boldsymbol{\alpha}) - \boldsymbol{\phi}_0 \right) \stackrel{a}{\sim} \boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\alpha}) \mathbf{q}_{nT}(\boldsymbol{\alpha}), \quad (\text{B.41})$$

and (since $\boldsymbol{\Sigma}_{\beta\beta}(\boldsymbol{\alpha})$ is positive definite, $\|\boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\alpha})\| < C$)

$$\left\| \tilde{\boldsymbol{\phi}}_{nT}(\boldsymbol{\alpha}) - \boldsymbol{\phi}_0 \right\| \leq \|\mathbf{D}_\alpha^{-1}\| \|\boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\alpha})\| \|\mathbf{q}_{nT}(\boldsymbol{\alpha})\| \leq C n^{-\alpha_{\min}/2} \|\mathbf{q}_{nT}(\boldsymbol{\alpha})\|.$$

Using (B.40)

$$\|\mathbf{q}_{nT}(\boldsymbol{\alpha})\| = \|n \mathbf{D}_\alpha^{-1} \mathbf{s}_{nT}\| \leq n^{1-\alpha_{\min}/2} \|\mathbf{s}_{nT}\|. \quad (\text{B.42})$$

Also from (B.30) we have

$$\begin{aligned} \mathbf{s}_{nT} &= n^{-1} (\mathbf{B}'_n \mathbf{M}_n \bar{\mathbf{u}}_{n\circ} - \mathbf{B}'_n \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_T \boldsymbol{\lambda}_T^*) + O_p(n^{-1+\alpha_\eta}) \\ &+ O_p\left(T^{-1/2} n^{-1+\frac{\alpha_\eta+\alpha_\gamma}{2}}\right) + O_p(T^{-1} n^{-1/2}). \end{aligned} \quad (\text{B.43})$$

Also using (B.31))

$$n^{-1} (\mathbf{B}'_n \mathbf{M}_n \bar{\mathbf{u}}_{n\circ} - \mathbf{B}'_n \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_T \boldsymbol{\lambda}_T^*) = O_p(T^{-1/2} n^{-1/2}). \quad (\text{B.44})$$

Substituting (B.44) in (B.43) and using the result in (B.42) we have

$$\begin{aligned} \|\mathbf{q}_{nT}(\boldsymbol{\alpha})\| &= O_p(n^{-\alpha_{\min}/2+1/2} T^{-1/2}) + O_p\left(T^{-1/2} n^{\frac{-\alpha_{\min}+(\alpha_\eta+\alpha_\gamma)}{2}}\right) \\ &+ O_p(n^{-\alpha_{\min}/2+\alpha_\eta}) + O_p(n^{-\alpha_{\min}/2+1/2} T^{-1}). \end{aligned} \quad (\text{B.45})$$

Denote the k^{th} element of $\mathbf{q}_{nT}(\boldsymbol{\alpha})$ by $q_{k,nT}(\boldsymbol{\alpha})$, we also have

$$\begin{aligned} q_{k,nT}(\boldsymbol{\alpha}) &= O_p(n^{-\alpha_{\min}/2+1/2} T^{-1/2}) + O_p\left(T^{-1/2} n^{\frac{-\alpha_{\min}+(\alpha_\eta+\alpha_\gamma)}{2}}\right) \\ &+ O_p(n^{-\alpha_{\min}/2+\alpha_\eta}) + O_p(n^{-\alpha_{\min}/2+1/2} T^{-1}). \end{aligned}$$

Also note that the k^{th} element of $\mathbf{D}_\alpha \left(\tilde{\phi}_{nT}(\boldsymbol{\alpha}) - \phi_0 \right)$ is given by $n^{\alpha_k/2} \left(\tilde{\phi}_{k,nT}(\boldsymbol{\alpha}) - \phi_{0,k} \right)$. Hence, in view of (B.41) and since $\boldsymbol{\Sigma}_{\beta\beta}^{-1}(\boldsymbol{\alpha})$ is a positive definite matrix then the probability order of $n^{\alpha_k/2} \left(\tilde{\phi}_{k,nT}(\boldsymbol{\alpha}) - \phi_{0,k} \right)$ must be the same as that of $q_{k,nT}$, and hence (as required)

$$\begin{aligned} \tilde{\phi}_{k,nT}(\boldsymbol{\alpha}) - \phi_{0,k} &= O_p \left(n^{-(\alpha_k + \alpha_{\min})/2 + 1/2} T^{-1/2} \right) + O_p \left(n^{\frac{-(\alpha_k + \alpha_{\min}) + (\alpha_\eta + \alpha_\gamma)}{2}} T^{-1/2} \right) \\ &\quad + O_p \left(n^{-(\alpha_k + \alpha_{\min})/2 + \alpha_\eta} \right) + O_p \left(n^{-(\alpha_k + \alpha_{\min})/2 + 1/2} T^{-1} \right). \end{aligned}$$

B.5 Proof of theorem 5

Using (64) and noting that $\hat{s}_{nT} \rightarrow_p \boldsymbol{\lambda}'_0 \boldsymbol{\Sigma}_f^{-1} \boldsymbol{\lambda}_0$, then

$$\begin{aligned} \hat{\mathbf{V}}_{\xi,nT} - \mathbf{V}_\xi &= (1 + \boldsymbol{\lambda}'_0 \boldsymbol{\Sigma}_f^{-1} \boldsymbol{\lambda}_0) \left[n^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \tilde{\mathbf{V}}_u \mathbf{M}_n \hat{\mathbf{B}}_{nT} - p \lim_{n \rightarrow \infty} \left(n^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{V}_u \mathbf{M}_n \mathbf{B}_n \right) \right] + o_p(1) \\ &= (1 + \boldsymbol{\lambda}'_0 \boldsymbol{\Sigma}_f^{-1} \boldsymbol{\lambda}_0) \left[n^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \tilde{\mathbf{V}}_u \mathbf{M}_n \hat{\mathbf{B}}_{nT} - n^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{V}_u \mathbf{M}_n \mathbf{B}_n \right] + o_p(1). \end{aligned} \tag{B.46}$$

Also using (B.4) we have

$$n^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \tilde{\mathbf{V}}_u \mathbf{M}_n \hat{\mathbf{B}}_{nT} = n^{-1} (\mathbf{B}_n + \mathbf{U}_{nT} \mathbf{G}_n)' \mathbf{M}_n \left(\tilde{\mathbf{V}}_u - \mathbf{V}_u + \mathbf{V}_u \right) \mathbf{M}_n (\mathbf{B}_n + \mathbf{U}_{nT} \mathbf{G}_n),$$

which, after some algebra, yields

$$n^{-1} \hat{\mathbf{B}}'_{nT} \mathbf{M}_n \tilde{\mathbf{V}}_u \mathbf{M}_n \hat{\mathbf{B}}_{nT} - n^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{V}_u \mathbf{M}_n \mathbf{B}_n = \sum_{j=1}^7 \mathbf{A}_{j,nT},$$

where

$$\begin{aligned} \mathbf{A}_{1,nT} &= n^{-1} \mathbf{B}'_n \mathbf{M}_n \left(\tilde{\mathbf{V}}_u - \mathbf{V}_u \right) \mathbf{M}_n \mathbf{B}_n, \quad \mathbf{A}_{2,nT} = n^{-1} \mathbf{G}'_n \mathbf{U}'_{nT} \mathbf{M}_n \left(\tilde{\mathbf{V}}_u - \mathbf{V}_u \right) \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_n, \\ \mathbf{A}_{3,nT} &= n^{-1} \mathbf{G}'_n \mathbf{U}'_{nT} \mathbf{M}_n \mathbf{V}_u \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_n, \quad \mathbf{A}_{4,nT} = n^{-1} \mathbf{G}'_n \mathbf{U}'_{nT} \mathbf{M}_n \left(\tilde{\mathbf{V}}_u - \mathbf{V}_u \right) \mathbf{M}_n \mathbf{B}_n, \\ \mathbf{A}_{5,nT} &= n^{-1} \mathbf{G}'_n \mathbf{U}'_{nT} \mathbf{M}_n \mathbf{V}_u \mathbf{M}_n \mathbf{B}_n, \quad \mathbf{A}_{6,nT} = n^{-1} \mathbf{B}'_n \mathbf{M}_n \left(\tilde{\mathbf{V}}_u - \mathbf{V}_u \right) \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_n, \\ \mathbf{A}_{7,nT} &= n^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{V}_u \mathbf{M}_n \mathbf{U}_{nT} \mathbf{G}_n. \end{aligned}$$

Considering the above terms in turn we note that

$$\|\mathbf{A}_{1,nT}\| \leq n^{-1} \|\mathbf{B}'_n \mathbf{M}_n\|^2 \left\| \tilde{\mathbf{V}}_u - \mathbf{V}_u \right\| = \lambda_{\max} \left(n^{-1} \mathbf{B}'_n \mathbf{M}_n \mathbf{B}_n \right) \left\| \tilde{\mathbf{V}}_u - \mathbf{V}_u \right\|.$$

Also, under Assumption 2 $\lambda_{\max}(n^{-1}\mathbf{B}'_n\mathbf{M}_n\mathbf{B}_n) < C$ and using (61) we have $\|\mathbf{A}_{1,nT}\| = O_p\left(n^{\alpha\gamma}\sqrt{\frac{\ln(n)}{T}}\right)$. Similarly

$$\|\mathbf{A}_{2,nT}\| \leq n^{-1}\|\mathbf{G}'_n\mathbf{U}'_{nT}\mathbf{M}_n\|^2\|\tilde{\mathbf{V}}_u - \mathbf{V}_u\| = \lambda_{\max}(n^{-1}\mathbf{G}'_n\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_n)\|\tilde{\mathbf{V}}_u - \mathbf{V}_u\|.$$

Then using (A.17) $n^{-1}\mathbf{G}'_n\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_n \rightarrow_p \bar{\sigma}_n^2\mathbf{G}'_T\mathbf{G}_T = T^{-1}\bar{\sigma}_n^2(T^{-1}\mathbf{F}'\mathbf{M}_T\mathbf{F})^{-1} = O(T^{-1})$, and it follows that $\|\mathbf{A}_{2,nT}\| = O_p\left(T^{-1}n^{\alpha\gamma}\sqrt{\frac{\ln(n)}{T}}\right)$. Turning to the third term

$$\|\mathbf{A}_{3,nT}\| \leq n^{-1}\|\mathbf{G}'_n\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_n\|\|\mathbf{V}_u\| = \lambda_{\max}(n^{-1}\mathbf{G}'_n\mathbf{U}'_{nT}\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_n)\|\mathbf{V}_u\|.$$

Also, by Lemma A.1 $\|\mathbf{V}_u\| = O(n^{\alpha\gamma})$, and hence $\|\mathbf{A}_{3,nT}\| = O_p(T^{-1}n^{\alpha\gamma})$, and $\|\mathbf{A}_{4,nT}\| \leq \|n^{-1/2}\mathbf{G}'_n\mathbf{U}'_{nT}\mathbf{M}_n\|\|n^{-1/2}\mathbf{M}_n\mathbf{B}_n\|\|\tilde{\mathbf{V}}_u - \mathbf{V}_u\|$. Also, as shown above $\|n^{-1/2}\mathbf{M}_n\mathbf{B}_n\| = O_p(1)$, $\|n^{-1/2}\mathbf{G}'_n\mathbf{U}'_{nT}\mathbf{M}_n\| = O_p(T^{-1/2})$, then

$$\|\mathbf{A}_{4,nT}\| = O_p\left(T^{-1/2}n^{\alpha\gamma}\sqrt{\frac{\ln(n)}{T}}\right),$$

$$\|\mathbf{A}_{5,nT}\| \leq \|n^{-1/2}\mathbf{G}'_n\mathbf{U}'_{nT}\mathbf{M}_n\|\|n^{-1/2}\mathbf{M}_n\mathbf{B}_n\|\|\mathbf{V}_u\| = O_p(T^{-1/2}n^{\alpha\gamma}),$$

$$\|\mathbf{A}_{6,nT}\| \leq \|n^{-1/2}\mathbf{G}'_n\mathbf{U}'_{nT}\mathbf{M}_n\|\|n^{-1/2}\mathbf{M}_n\mathbf{B}_n\|\|\tilde{\mathbf{V}}_u - \mathbf{V}_u\| = O_p\left(T^{-1/2}n^{\alpha\gamma}\sqrt{\frac{\ln(n)}{T}}\right),$$

and $\|\mathbf{A}_{7,nT}\| \leq \|n^{-1/2}\mathbf{B}'_n\mathbf{M}_n\|\|\mathbf{V}_u\|\|n^{-1/2}\mathbf{M}_n\mathbf{U}_{nT}\mathbf{G}_n\| = O_p(T^{-1/2}n^{\alpha\gamma})$. Overall,

$$\left\|n^{-1}\hat{\mathbf{B}}'_n\mathbf{M}_n\tilde{\mathbf{V}}_u\mathbf{M}_n\hat{\mathbf{B}}_n - n^{-1}\mathbf{B}'_n\mathbf{M}_n\mathbf{V}_u\mathbf{M}_n\mathbf{B}_n\right\| = O_p\left(n^{\alpha\gamma}\sqrt{\frac{\ln(n)}{T}}\right),$$

which if used in (B.46) establishes (63), as required.

**Online Supplement A: Data Sources and Calibration of Monte Carlo
Designs**

To

**Pricing errors in linear asset pricing models with strong, semi-strong, and
latent factors**

by

M. Hashem Pesaran
University of Southern California, and Trinity College, Cambridge

Ron P. Smith
Birkbeck, University of London

February 2023

S1 Introduction

This online supplement provides the details of data sources for the risk factors and the excess returns on securities used to calibrate the Monte Carlo designs and to carry out the empirical applications reported in Sections 4 and 6 of the main paper. Section S2 describes the data used to calibrate the Monte Carlo (MC) designs, and Section S3 provides the estimates that formed the basis of the calibration of the parameters of the MC designs. Section S4 provides evidence for the choice of 14 blocks used in calibration of error covariances in the Monte Carlo experiments. Section S5 describes the data for factors and excess returns used in the empirical applications, and Section S6 derives the relationship between pooled R^2 of return regressions and the factor strengths used in the discussion of the empirical results.

S2 Data used to calibrate the Monte Carlo designs

S-2.1 Factors

To calibrate the parameters of the three factor model used in the Monte Carlo experiments we used monthly Fama-French three factor data series over the long sample 1963m8-2021m12, downloaded from Kenneth French's webpage.¹ The factors are the market return minus the risk free rate, denoted by MKT, the value factor (high book to market minus low portfolios, HML) and the size factor (small minus big portfolios, SMB). The risk free rate is also downloaded from French's webpage. First order autoregressions, AR(1), were estimated for all the three factors using the full data set, 1963m8-2021m12. Then GARCH(1,1) models were then estimated on the residuals from the fitted AR(1) regressions.

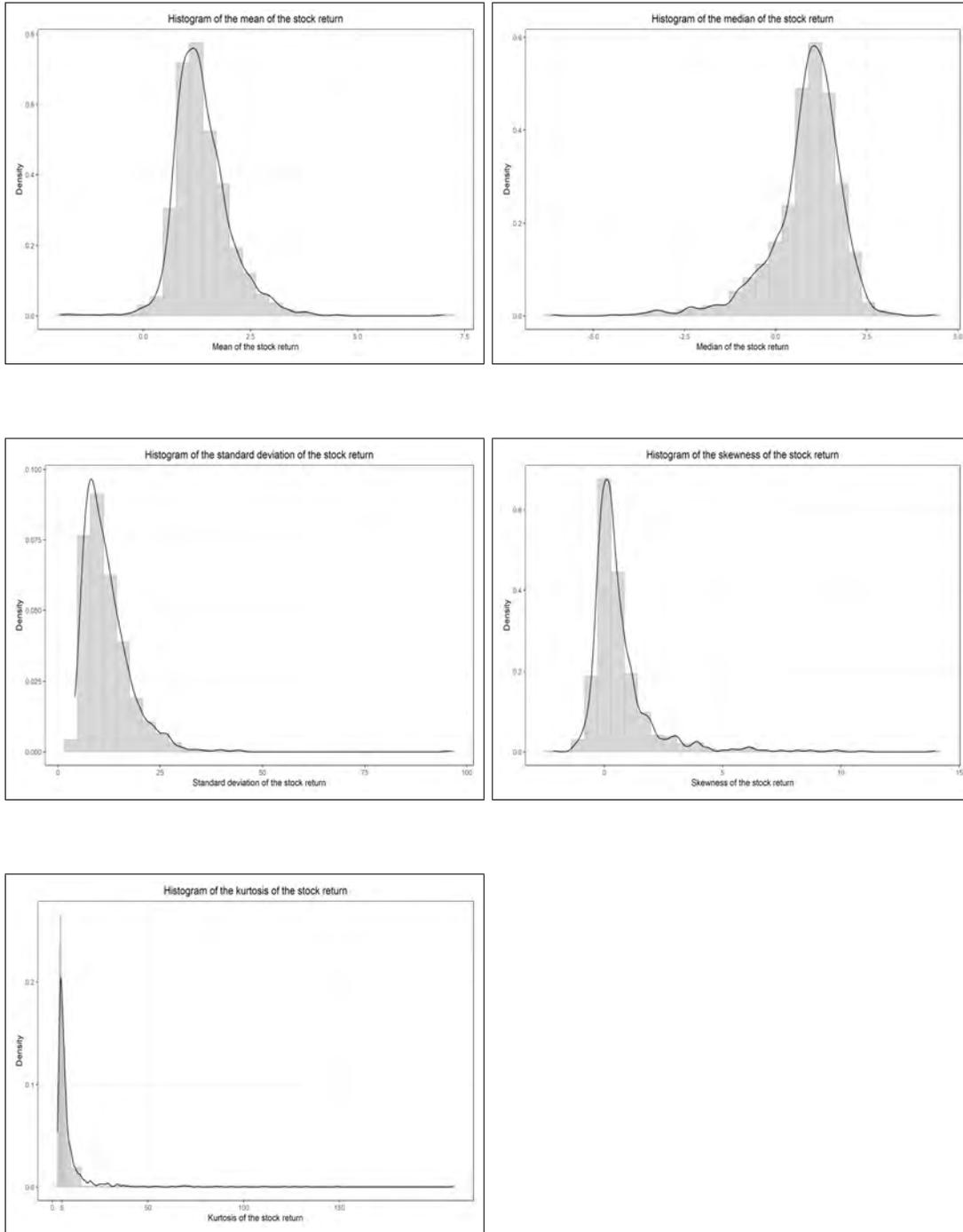
S-2.2 Excess returns

To calibrate the factor loadings and other parameters of the excess return regressions we used the shorter sample over the 20 years 2002m1 - 2021m12 ($T = 240$). Monthly returns (inclusive of dividend payment) over 2002m1 - 2021m12 ($T = 240$) for NYSE and NASDAQ stocks with share codes of 10 and 11 from CRSP were downloaded from Wharton Research Data Services and transformed to firm-specific excess returns using the risk free rate from French's webpage, and measured in percent, per month. Only stocks with data over the period 2002m1-2021m12 were used to arrive at a balanced panel with $T = 240$ monthly observations and a total number of $n = 1289$, securities.

¹See https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

To avoid extreme outliers influencing the estimates we first computed mean, median, standard deviation, skewness and kurtosis for each security over 2002m1 - 2021m12. Table DS1 reports the mean, standard deviations and interquartile range of these statistics over all the 1289 securities in our sample. The histograms of these summary statistics (mean, median, standard deviation, skewness and kurtosis) of the individual stock returns for the full sample over 2002m1 - 2021m12 are shown in Figure DS1. As can be seen from these summary statistics, there are outlier security returns with very large standard deviations. This is clear from the cross firm standard deviation of 13.97 which is much larger than the kurtosis of 9.24 (See Table DS1), resulting from a number of extreme outliers also seen from the long right tail of the histogram for the distribution of kurtosis across firms.

Figure DS1: Histogram and density function of the individual stock returns ($n = 1289$) over 2002m1-2021m12 ($T = 240$)



S-2.3 Stocks with the kurtosis less than or equal to 14 and less than 16

To reduce the influence of outlier returns on our results we considered dropping stock excess returns having kurtosis in excess of 14 and 16. Excluding stocks with kurtosis less than or equal to 14, resulted in a sample with $n = 1148$ securities, whilst if we use the cut off point of 16 we ended up with $n = 1175$ securities. The mean, median, standard deviation, skewness and kurtosis for each security over 2002m1 - 2021m12 of the whole and two sub-samples, (1289, 1148 and 1175). Then the average, standard deviation and interquartile range (IQR) for each summary statistic of the stocks from the two sub-sample are summarized respectively in Table DS1.

Table DS1: The average, standard deviation and interquartile range of the summary statistics of the individual stocks over 2002m1 - 2021m12 ($T = 240$)

	Average	Standard deviation	Interquartile range
<i>Panel A: All stocks ($n = 1289$)</i>			
stock.mean	1.3666	0.6593	0.7511
stock.median	0.8360	0.9970	0.9310
stock.standard deviation	11.9030	6.0082	6.5248
stock.skewness	0.6411	1.3766	0.9290
stock.kurtosis	9.2407	13.9786	4.0183
<i>Panel B: Kurtosis ≤ 14 ($n = 1148$)</i>			
stock.mean	1.3303	0.5898	0.7032
stock.median	0.9474	0.8816	0.8654
stock.standard deviation	10.8837	4.3944	5.6429
stock.skewness	0.2976	0.5996	0.7270
stock.kurtosis	5.9885	2.3932	2.7762
<i>Panel B: Kurtosis ≤ 16 ($n = 1175$)</i>			
stock.mean	1.3336	0.5922	0.7081
stock.median	0.9363	0.8878	0.8764
stock.standard deviation	10.9943	4.4695	5.6836
stock.skewness	0.3219	0.6285	0.7607
stock.kurtosis	6.1947	2.7222	2.9819

Histograms for mean, median, standard deviation, skewness and kurtosis of the individual stock returns for two sub-samples are shown in Figure DS2 for the sub-sample with the kurtosis less than or equal to 14 over 2002m1 - 2021m12 ($n = 1148$), and Figure DS3 are for the sub-sample with the kurtosis less than 16 over 2002m1 - 2021m12 ($n = 1175$).

Figure DS2: Histogram and density function of the individual stock returns with kurtosis less than or equal to 14 over 2002m1-2021m12 ($T = 240$) and $n = 1148$

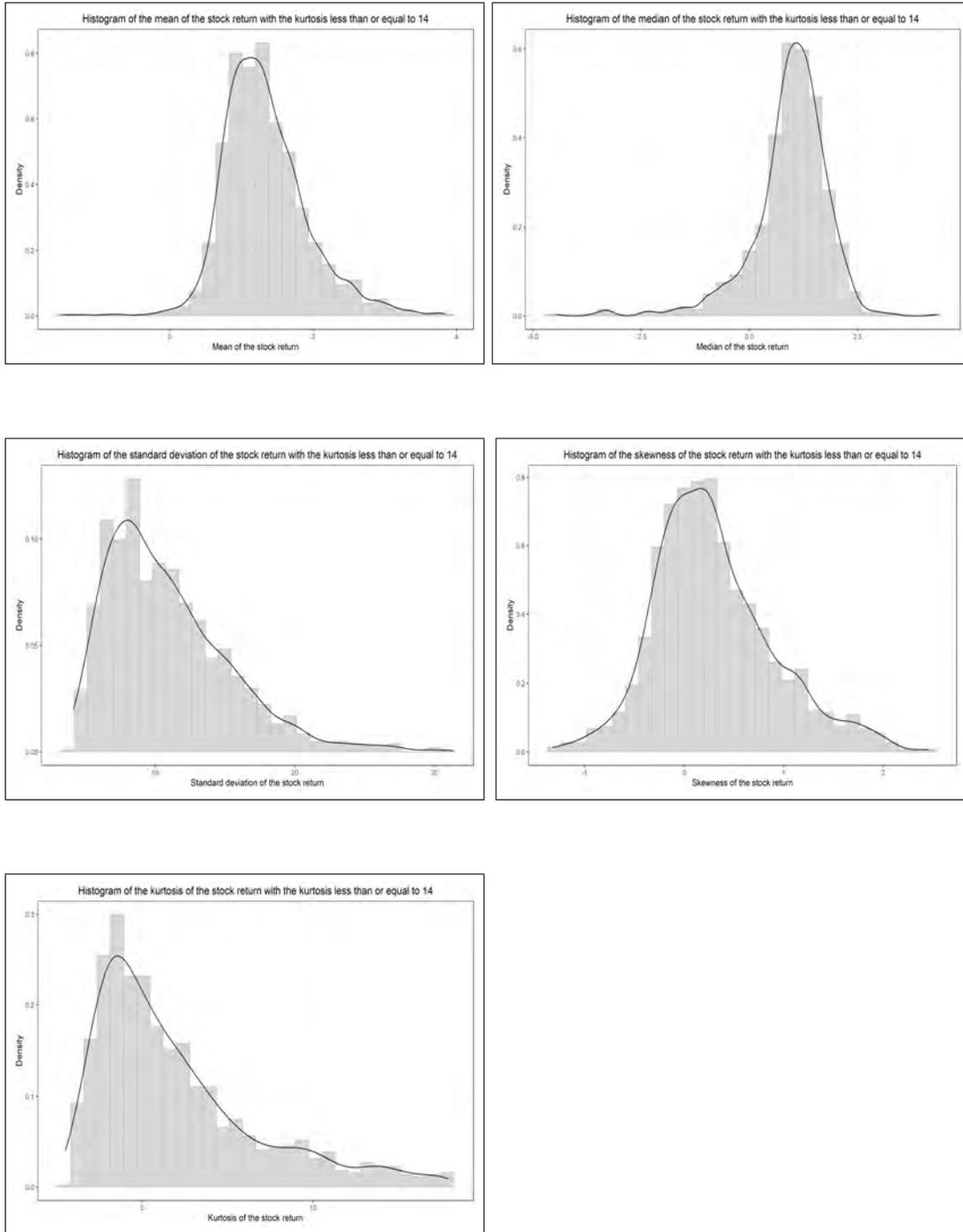
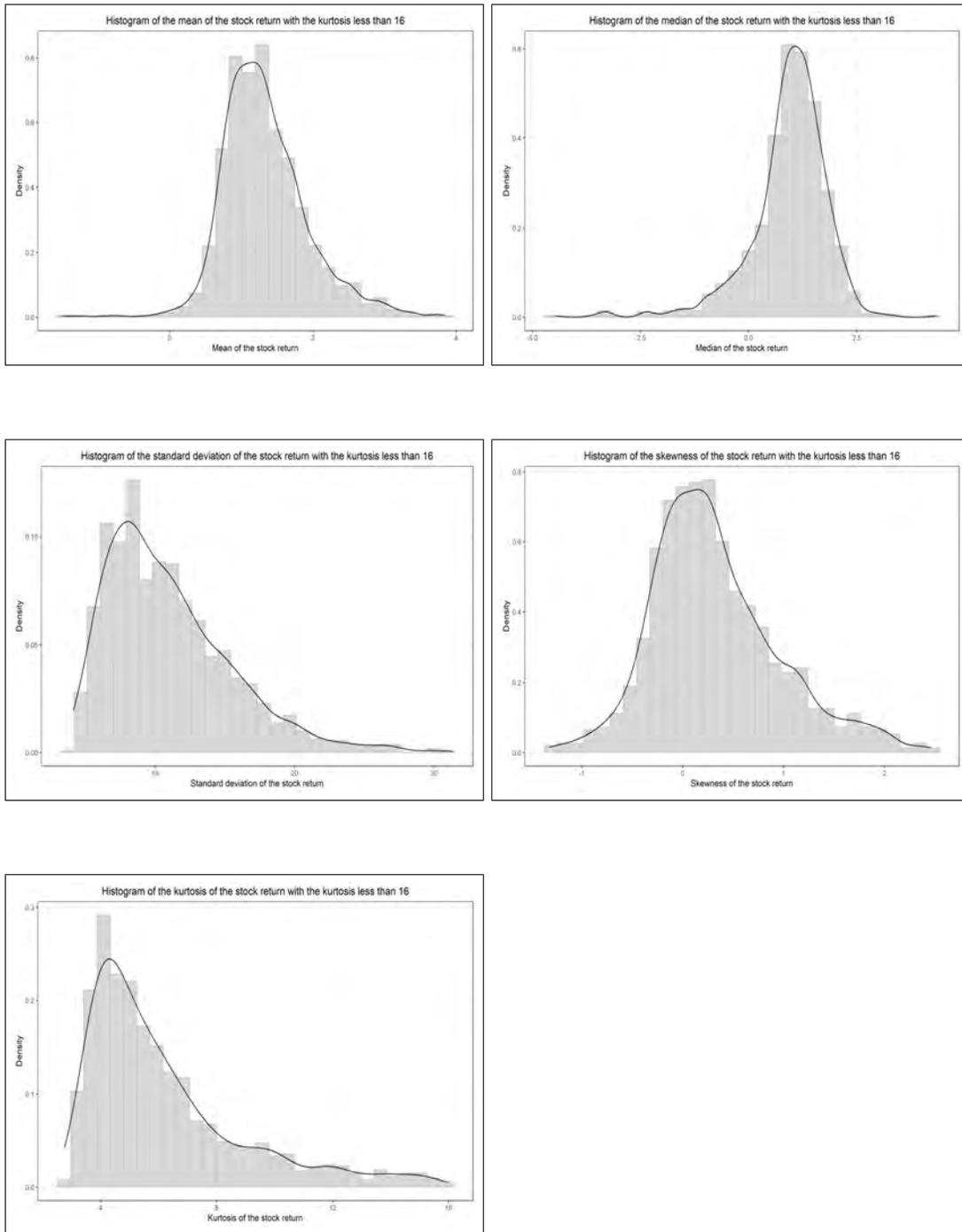


Figure DS3: Histogram and density function of the individual stock returns with kurtosis less than or equal to 16 over 2002m1-2021m12 ($T = 240$) and $n = 1148$



S3 The MC design and its calibration

S-3.1 Generation of returns

Excess returns, r_{it} , are generated as

$$\begin{aligned} r_{it} &= \mathbf{a}_i + \sum_{k=M,H,S} \beta_{ik} f_{kt} + u_{it}, \\ &= \mathbf{a}_i + \boldsymbol{\beta}'_i \mathbf{f}_t + u_{it}, \end{aligned} \quad (\text{S3.1})$$

for $i = 1, 2, \dots, n$; $t = 1, 2, \dots, T$, with

$$\mathbf{a}_i = c + \boldsymbol{\beta}'_i \boldsymbol{\phi}_0 + \eta_i, \quad (\text{S3.2})$$

where $\mathbf{f}_t = (f_{Mt}, f_{Ht}, f_{St})'$, and $\boldsymbol{\beta}_i = (\beta_{Mi}, \beta_{Hi}, \beta_{Si})'$, and $\boldsymbol{\phi}_0 = (\phi_M, \phi_H, \phi_S)'$.

S-3.2 Generation of factors

Factors are generated as first-order autoregressive, AR(1), processes with GARCH(1,1) effects:

$$f_{kt} = \mu_k(1 - \rho_k) + \rho_k f_{k,t-1} + (1 - \rho_k^2)^{1/2} \sigma_{kt} \zeta_{kt}, \quad (\text{S3.3})$$

$$\sigma_{kt}^2 = (1 - b_k - c_k) \sigma_k^2 + b_k \sigma_{k,t-1}^2 + c_k \sigma_{k,t-1}^2 \zeta_{k,t-1}^2, \quad (\text{S3.4})$$

for $k = M, H, S$, starting from $t = -49, \dots, 0, 1, 2, \dots, T$, with $f_{k,-50} = 0$ (and $\sigma_{k,-50} = 0$ in the case where $c_k \neq 0$) to minimize the effects of the initial values on the sample f_{kt} , $t = 1, 2, \dots, T$ used in the simulations.

The data generating process for the factors is calibrated using the full set of Fama-French three factor data set covering the period 1963m8-2021m12. The calibrated parameter values are $\boldsymbol{\mu} = (\mu_M, \mu_H, \mu_S)' = (0.59, 0.27, 0.23)'$, $\boldsymbol{\sigma} = (\sigma_M, \sigma_H, \sigma_S)' = (4.45, 2.86, 3.03)'$, and $\boldsymbol{\rho} = (\rho_M, \rho_H, \rho_S)' = (0.06, 0.17, 0.07)'$. Note that $\text{Var}(f_{kt}) = \sigma_k^2$. The parameters of (S3.2) are also estimated using the bias-corrected procedure and are set as $c = 0.83$ and $\boldsymbol{\phi}_0 = (-0.49, -0.35, 0.16)'$. To ensure that correlation across the three factors match the Fama-French data we generated $\boldsymbol{\zeta}_t = (\zeta_{Mt}, \zeta_{Ht}, \zeta_{St})'$ as $\boldsymbol{\zeta}_t = \mathbf{Q}_\zeta \boldsymbol{\omega}_t$, where \mathbf{Q}_ζ is the Cholesky factor of \mathbf{R}_ζ , the correlation matrix of \mathbf{f}_t given by

$$\mathbf{R}_\zeta = \begin{pmatrix} 1 & -0.21 & 0.28 \\ -0.21 & 1 & -0.02 \\ 0.27 & -0.02 & 1 \end{pmatrix}.$$

We consider both Gaussian and non-Gaussian errors and generate $\boldsymbol{\omega}_t$ as $IID(\mathbf{0}, \mathbf{I}_3)$, as well as a multivariate t with 5 degrees of freedom, namely $t(\mathbf{0}, \mathbf{I}_3, 5)$. The remaining

parameters are set as $b_k = c_k = 0$ to generate homoskedastic errors, and $b_k = 0.8$ and $c_k = 0.1$ for $k = M, H, S$ to generate GARCH effects.

S-3.3 Estimation of factor models

Consider the AR(1) processes with a GARCH(1,1) errors

$$f_{kt} = \mu_k(1 - \rho_k) + \rho_k f_{k,t-1} + (1 - \rho_k^2)^{1/2} \sigma_{kt} \zeta_{kt}, \quad (\text{S3.5})$$

$$\sigma_{kt}^2 = (1 - b_k - c_k)\sigma_k^2 + b_k \sigma_{k,t-1}^2 + c_k \sigma_{k,t-1}^2 \zeta_{k,t-1}^2, \quad (\text{S3.6})$$

where f_{kt} , $k = M, H, S$ and $t = 1, \dots, T$, denote the the values of three factors *MKT*, *HML*, *SMB* in month t , respectively. The estimates of GARCH parameters obtained using the sample 2002m1 - 2021m12 ($T = 240$) are summarized in Table DS2.

Table DS2: GARCH parameters for the models of three Fama-French factors for the sample over 2002m1 - 2021m12 ($T = 240$)

	$\hat{\mu}$	$\hat{\rho}$	$\hat{\sigma}$	\hat{b}	\hat{c}
<i>MKT</i>	0.8030 (0.3035)	0.0711 (0.0648)	4.5703 (·)	0.6781 (0.0854)	0.2395 (0.0627)
<i>HML</i>	-0.0805 (0.2194)	0.1816 (0.0639)	3.1513 (·)	0.7582 (0.0986)	0.1987 (0.0686)
<i>SMB</i>	0.1718 (0.1651)	-0.0267 (0.0649)	2.5815 (·)	0.8353 (0.1845)	0.0680 (0.0606)

The correlation matrix of three factors *MKT*, *HML*, *SMB* over 2001m1-2021m9 ($T = 240$) is

$$\begin{pmatrix} 1.00 & 0.20 & 0.35 \\ 0.20 & 1.00 & 0.35 \\ 0.35 & 0.35 & 1.00 \end{pmatrix}.$$

S-3.4 Factor loadings estimates

For each of the securities $i = 1, 2, \dots, 1175$ (with kurtosis below 16), and $t = 1, 2, \dots, 240$, OLS regressions excess returns y_{it} for security i was run on an intercept and the three

FF factors

$$y_{it} = r_{it} - r_t^f = a_i + \sum_{k \in \{M, H, S\}} \beta_{ik} f_{kt} + u_{it},$$

where r_{it} is the return of i^{th} security at time t , inclusive of dividend (if any), and r_t^f is the risk free rate. The sample mean and standard deviation of the excess return for each individual stock, denoted as \bar{y}_{iT} and $sd_{iT}(y)$ are computed as

$$\bar{y}_{iT} = T^{-1} \sum_{t=1}^T y_{it}, \quad (\text{S3.7})$$

$$sd_{iT}(y) = \sqrt{(T-1)^{-1} \sum_{t=1}^T (y_{it} - \bar{y}_{iT})^2}. \quad (\text{S3.8})$$

The estimates $\hat{a}_{i,T}$, $\hat{\beta}_{ik,T}$, $k = M, H, S$ are given by

$$\left(\hat{a}_{i,T}, \hat{\beta}_{iM,T}, \hat{\beta}_{iH,T}, \hat{\beta}_{iS,T} \right)' = (\mathbf{F}'_0 \mathbf{F}_0)^{-1} \mathbf{F}'_0 \mathbf{y}_{i\infty}, \quad (\text{S3.9})$$

where $\mathbf{F}_0 = (\boldsymbol{\tau}_T, \mathbf{F})$, $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_T)'$, $\mathbf{f}_t = (f_{Mt}, f_{Ht}, f_{St})'$ and $\mathbf{y}_{i\infty} = (y_{i1}, y_{i2}, \dots, y_{iT})'$. The standard error of the i^{th} regression, denoted as s_{iT} , is given by

$$\hat{\sigma}_{iT}^2 = (T - K - 1)^{-1} \sum_{t=1}^T \hat{u}_{it}^2, \quad (\text{S3.10})$$

where $\hat{u}_{it} = y_{it} - \hat{a}_{i,T} - \sum_{k \in \{M, H, S\}} \hat{\beta}_{ik,T} f_{kt}$. The coefficient of determination of the i^{th} regression, denoted by R_{iT}^2 , is given by

$$R_{iT}^2 = 1 - \frac{\sum_{t=1}^T \hat{u}_{it}^2}{\sum_{t=1}^T (y_{it} - \bar{y}_{iT})^2}. \quad (\text{S3.11})$$

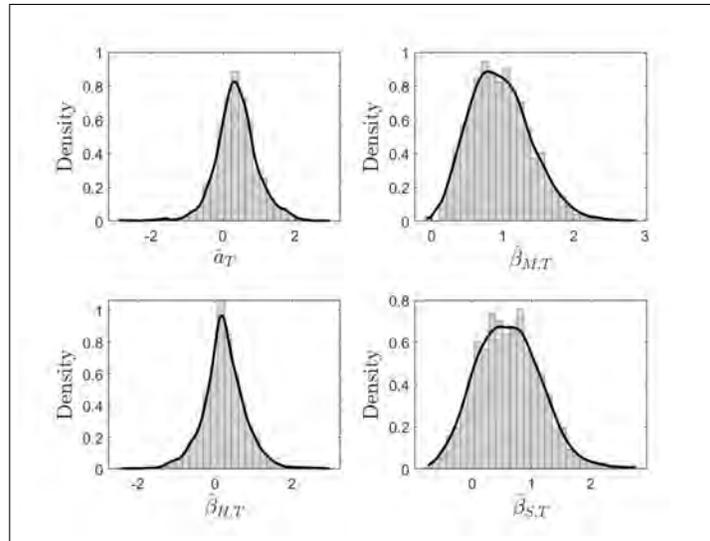
We compute the summary statistics: mean, median, standard deviation (S.D.), skewness, kurtosis, interquartile range, minimum, maximum for the sample mean and standard deviation of the excess returns, the estimates and the corresponding standard error, R -squared of the regressions over 2002m1 - 2021m12 ($T = 240$), for the $n = 1175$ securities: \bar{y}_{iT} , $sd_{iT}(y)$, $\hat{a}_{i,T}$, $\hat{\beta}_{iM,T}$, $\hat{\beta}_{iH,T}$, $\hat{\beta}_{iS,T}$, $\hat{\sigma}_{iT}^2$ and R_{iT}^2 for $i = 1, 2, \dots, 1175$, computed using (S3.7)-(S3.11). The results are summarized in Table DS3.

Table DS3: The summary statistics of the estimates, standard error and R-squared of the panel regression over 2002m1 - 2021m12 ($T = 240$) and $n = 1175$

	mean	median	S.D.	skewness	kurtosis	IQR	min	max
\bar{y}	1.2366	1.1572	0.5922	0.6219	4.9697	0.7085	-1.5971	3.7434
$sd(y)$	11.0006	10.1115	4.4695	1.1241	4.4498	5.6826	4.1706	31.3322
\hat{a}	0.3749	0.3562	0.5810	-0.1980	5.5804	0.6493	-2.8882	2.9736
$\hat{\beta}_M$	0.9714	0.9362	0.4279	0.5352	3.3349	0.5810	-0.0689	2.8590
$\hat{\beta}_H$	0.2235	0.2093	0.5276	-0.0916	4.7150	0.5915	-2.4498	2.9788
$\hat{\beta}_S$	0.6061	0.5790	0.5381	0.3552	3.2197	0.7495	-0.7504	2.7474
$\hat{\sigma}$	9.4123	8.2297	4.2794	1.3292	5.0599	5.3811	3.4523	30.6953
R^2	0.2840	0.2782	0.1383	0.2172	2.2941	0.2131	0.0050	0.6814

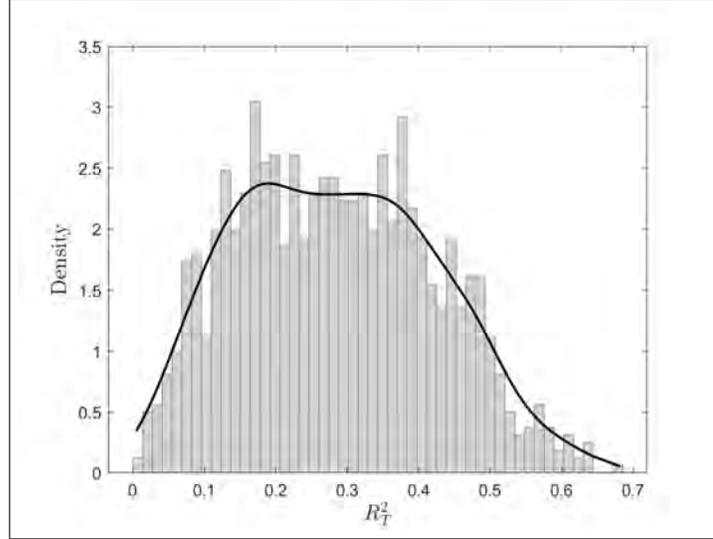
The histogram for the \hat{a}_T and the $\hat{\beta}_{k,T}$ for $k = M, H, S$ over 2002m1 - 2021m12 ($T = 240$), each using 1175 data points $\hat{a}_{i,T}$ and the $\hat{\beta}_{ik,T}$ for $k = M, H, S, i = 1, 2, \dots, 1175$, is shown in the Figure DS4.

Figure DS4: Histogram and density function of the coefficients of the panel regression over 2002m1-2021m12 ($T = 240$) and $n = 1175$



The histogram for the $R_{i,T}^2$ over 2002m1 - 2021m12 ($T = 240$), using 1175 data points $R_{i,T}^2$ for $i = 1, 2, \dots, 1175$, is shown in the Figure DS5.

Figure DS5: Histogram and density function of the R-squared of the panel regression over 2002m1-2021m12 ($T = 240$) and $n = 1775$



S-3.5 Calibrating the fit of return regressions

To see how κ controls the regression fit, note that the n return processes (S3.1) can be written more compactly in vector form as

$$\mathbf{r}_t = \mathbf{a} + \mathbf{B}\mathbf{f}_t + \mathbf{u}_t,$$

where $\mathbf{u}_t = \gamma g_t + \kappa \hat{\mathbf{S}}\boldsymbol{\varepsilon}_t$, $\mathbf{a} = (a_1, a_2, \dots, a_n)'$, with a_i given by $a_i = c + \boldsymbol{\beta}'_i \boldsymbol{\phi} + \eta_i$, and

$$\hat{\mathbf{S}} = \text{Diag}(\hat{\mathbf{S}}_b, b = 1, 2, \dots, B).$$

Overall, the DGP for the return regressions can be written compactly as

$$\mathbf{r}_t = c\boldsymbol{\tau}_n + \mathbf{B}(\mathbf{f}_t + \boldsymbol{\phi}) + \gamma g_t + \kappa \hat{\mathbf{S}}\boldsymbol{\varepsilon}_t + \boldsymbol{\eta}_n,$$

where $\boldsymbol{\eta}_n = (\eta_1, \eta_2, \dots, \eta_n)'$. We abstract from pricing errors and weak latent factor and $\eta_i = 0$, $g_t = 0$, and set κ such that the pooled R^2 (PR^2) of return regressions can be controlled to be around $R_0^2 = 0.30$. We have

$$PR_{nT}^2 = 1 - \frac{n^{-1}T^{-1} \sum_{t=1}^T \sum_{i=1}^n E(u_{it}^2)}{n^{-1}T^{-1} \sum_{t=1}^T \sum_{i=1}^n \text{Var}(r_{it})}.$$

$$\begin{aligned} \text{Var}(r_{it}) &= \text{Var}(\boldsymbol{\beta}'_i \mathbf{f}_t) + \text{Var}(u_{it}) = E(\boldsymbol{\beta}'_i \boldsymbol{\Sigma}_f \boldsymbol{\beta}_i) + E(u_{it}^2) \\ &= \text{Tr}[\boldsymbol{\Sigma}_f E(\boldsymbol{\beta}_i \boldsymbol{\beta}'_i)] + E(u_{it}^2). \end{aligned}$$

Denote the k^{th} element of $\boldsymbol{\beta}_i$ by β_{ik} , then if $\beta_{ik} \sim IIDN(\mu_{\beta_k}, \sigma_{\beta_k}^2)$, and β_{ik} are distributed independently over $k = 1, 2, \dots, K$, we have

$$E(\boldsymbol{\beta}_i \boldsymbol{\beta}'_i) = \text{Diag}(\mu_{\beta_k}^2 + \sigma_{\beta_k}^2, \text{ for } k = 1, 2, \dots, K).$$

Then,

$$\text{Var}(r_{it}) = \text{Tr}[\boldsymbol{\Sigma}_f E(\boldsymbol{\beta}_i \boldsymbol{\beta}'_i)] + E(u_{it}^2) = \sum_{k=1}^K \sigma_{fk}^2 (\mu_{\beta_k}^2 + \sigma_{\beta_k}^2) + E(u_{it}^2).$$

Also,

$$\sum_{i=1}^n E(u_{it}^2) = \text{Tr}[E(\mathbf{u}_t \mathbf{u}'_t)],$$

and (when $g_t = 0$) we have

$$E(\mathbf{u}_t \mathbf{u}'_t) = \mathbf{V}_u = \kappa^2 \hat{\mathbf{S}} E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}'_t) \hat{\mathbf{S}}' = \kappa^2 \hat{\mathbf{S}} \mathbf{V}_\varepsilon \hat{\mathbf{S}}',$$

where $\mathbf{V}_\varepsilon = \text{Diag}(\mathbf{V}_{b\varepsilon}^{(r)}, b = 1, 2, \dots, B)'$. Hence,

$$n^{-1} T^{-1} \sum_{t=1}^T \sum_{i=1}^n E(u_{it}^2) = n^{-1} \text{Tr}(\mathbf{V}_u) = \kappa^2 n^{-1} \text{Tr}(\hat{\mathbf{S}} \mathbf{V}_\varepsilon \hat{\mathbf{S}}')$$

and

$$PR_{nT}^2 = 1 - \frac{\kappa^2 n^{-1} \text{Tr}(\hat{\mathbf{S}} \mathbf{V}_\varepsilon \hat{\mathbf{S}}')}{\sum_{k=1}^K \sigma_{fk}^2 (\mu_{\beta_k}^2 + \sigma_{\beta_k}^2) + \kappa^2 n^{-1} \text{Tr}(\hat{\mathbf{S}} \mathbf{V}_\varepsilon \hat{\mathbf{S}}')}.$$

To achieve $\lim_{n \rightarrow \infty} PR_{nT}^2 = R_0^2$, we need to set (assuming all K factors are strong)

$$\kappa^2 = \frac{\sum_{k=1}^K \sigma_{fk}^2 (\mu_{\beta_k}^2 + \sigma_{\beta_k}^2)}{n^{-1} \text{Tr}(\hat{\mathbf{S}} \mathbf{V}_\varepsilon \hat{\mathbf{S}}')} \left(\frac{1 - R_0^2}{R_0^2} \right). \quad (\text{S3.12})$$

When there are no idiosyncratic error dependence, namely when $\hat{\mathbf{S}} = \mathbf{I}_n$, the above expression simplifies to

$$\kappa^2 = \frac{\sum_{k=1}^K \sigma_{fk}^2 (\mu_{\beta_k}^2 + \sigma_{\beta_k}^2)}{n^{-1} \text{Tr}(\mathbf{V}_\varepsilon)} \left(\frac{1 - R_0^2}{R_0^2} \right). \quad (\text{S3.13})$$

If we only use the market factor, we have

$$\kappa^2 = \frac{\sigma_M^2 (\mu_{\beta_M}^2 + \sigma_{\beta_M}^2)}{n^{-1} \text{Tr}(\mathbf{V}_\varepsilon)} \left(\frac{1 - R_0^2}{R_0^2} \right). \quad (\text{S3.14})$$

We expect that $n^{-1} \text{Tr}(\mathbf{V}_\varepsilon) \rightarrow 1$, if $E(\sigma_{ii}) = 1$, as under our DGP.

S-3.6 Estimation of FF factor strengths

Denote by $t_{ik,T} = \hat{\beta}_{ik,T} / \text{s.e.}(\hat{\beta}_{ik,T})$, the t-statistic corresponding to β_{ik} . The total number of factor loadings of factor k , that are statistically significant over $i = 1, 2, \dots, n$, $n = 1175$, is:

$$\hat{D}_{nT,k} = \sum_{i=1}^n \hat{d}_{ik,nT} = \sum_{i=1}^n \mathbf{1}[|t_{ik,T}| > c_p(n)],$$

where $\mathbf{1}(A) = 1$ if $A > 0$, and zero otherwise, and the critical value function that allows for the multiple testing nature of the problem, $c_p(n, \delta)$, is given by

$$c_p(n, \delta) = \Phi^{-1} \left(1 - \frac{p}{2n^\delta} \right), \quad (\text{S3.15})$$

where p is the nominal size, set, following Bailey, Kapetanios, and Pesaran (2021, BKP), as $p = 0.1$, $\delta > 0$ is the critical value exponent, set $\delta = 0.25$, and $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function of the standard normal distribution. Let $\hat{\pi}_{nT,k}$ be the fraction of significant loadings of factor k , and note that $\hat{\pi}_{nT,k} = \hat{D}_{nT,k}/n$. The strength of factor k , denoted by α_{k0} , for $k = 1, 2, \dots, K$, $K = 3$, is estimated by

$$\hat{\alpha}_k = \begin{cases} 1 + \frac{\ln \hat{\pi}_{nT,k}}{\ln n}, & \text{if } \hat{\pi}_{nT,k} > 0, \\ 0, & \text{if } \hat{\pi}_{nT,k} = 0. \end{cases} \quad (\text{S3.16})$$

The variance of the estimated strength of factor k is given by

$$\text{Var}(\hat{\alpha}_k) = (\ln n)^{-2} \psi_n(\alpha_{k0}),$$

where

$$\psi_n(\alpha_{k0}) = p(n - n^{\alpha_{k0}})n^{-\delta-2\alpha_{k0}} \left(1 - \frac{p}{n^\delta}\right).$$

So the standard error of the estimated strength of factor k can be computed by:

$$s.e.(\hat{\alpha}_k) = \frac{\sqrt{\psi_n(\hat{\alpha}_k)}}{\ln n}. \quad (\text{S3.17})$$

Mean and variance of the loadings associated with factor k are given by

$$\hat{\mu}_{\beta_{kT}}(\hat{\alpha}_k) = \frac{\sum_{i=1}^n \mathbf{1} [|t_{ikT}| > c_p(n)] \hat{\beta}_{ikT}}{\sum_{i=1}^n \mathbf{1} [|t_{ikT}| > c_p(n)]}, \quad (\text{S3.18})$$

$$\hat{\sigma}_{\beta_k}^2(\hat{\alpha}_k) = \frac{\sum_{i=1}^n \mathbf{1} [|t_{ikT}| > c_p(n)] \left(\hat{\beta}_{ikT} - \hat{\mu}_{\beta_{kT}}(\hat{\alpha}_k)\right)^2}{\sum_{i=1}^n \mathbf{1} [|t_{ikT}| > c_p(n)]}. \quad (\text{S3.19})$$

In the case where a factor is strong, namely $\alpha_k = 1$, then it must be that $\mathbf{1} [|t_{ikT}| > c_p(n)] = 1$ for all i . The estimated factor strengths $\hat{\alpha}_k$ and corresponding standard errors for $k = M, H, S$, using the sample over 2002m1 - 2021m12 ($T = 240$) and $n = 1175$ are reported in DS4.

Table DS4: Strength of three FF factors estimated over 2001m1-2021m9 ($T = 240$ and $n = 1175$)

	M	H	S
$\hat{\alpha}$	0.9941	0.8373	0.9023
	(0.0001)	(0.0014)	(0.0008)

Note: This table reports the estimates of the factor strength using (S3.16) and the standard errors that are reported in () using (S3.17), for three factors MKT, HML and SMB , using the sample over 2001m1-2021m9 ($T = 240$) and $n = 1175$, $K = 3$.

S-3.7 Estimates of spanning error coefficients for FF factors

Table DS5: The bias-corrected estimates of c_0 (intercept) and ϕ_M , ϕ_H , and ϕ_S for the sample 2002m1 - 2021m12 ($T = 240$) with $n = 1175$ securities

\hat{c}_0	$\hat{\phi}_M$	$\hat{\phi}_H$	$\hat{\phi}_S$
0.8423	-0.4808	-0.3259	0.1195
	(0.0971)	(0.0780)	(0.0846)

S4 Grouping of securities by their pair-wise correlations

The $T = 240$ sample ending in 2021 was used to estimate the pair-wise correlations of the residuals from the $n = 1168$ returns regressions using the Fama-French three factors. Then all the statistically insignificant correlations were set to zero. Significance was determined allowing for the multiple testing nature of the tests, using the critical value $c_p(n, \delta) = \Phi^{-1} \left(1 - \frac{p}{2n^\delta} \right)$, with $p = 0.05$ and $\delta = 2$, since the number of pair-wise correlations is of order $O(n^2)$. See also Bailey, Pesaran, and Smith (2019). For the majority of securities (668 out of the 1168), the pair-wise return correlations were not statistically significant. The securities with a relatively large number of non-zero correlations were either in the banking or energy related industries.

Initially, the securities were grouped using the two digit codes from the 1987 standard industrial classification (SIC 1987). But this gave too many groups, 62. Many of the groups had very few members: only one security for 3 out of the 62 groups and less than 10 for 36 groups. However, code 60 (banking) had 145 securities. Therefore, it was decided to work with the industrial classification based on a one digit level, and to aggregate the codes with a small number of securities, taking out two digit codes where there were large numbers in that code. We ended up with 14 contiguous groups ranging in size from 33 to 145. Average correlations were low overall, but the average absolute correlation of securities within the groups was around 10 times that with firms outside the group. See Table DS6, which gives averages of pair-wise correlations without thresholding. These estimates suggested that a block diagonal structure with 14 blocks was a reasonable characterization which is used in the Monte Carlo analysis. See Section 4 of the main paper.

Table DS6: Sector groupings by SIC codes and within and outside sector average pair-wise correlations.

	SIC codes	Number of stocks	Average correlations	
			within	outside
Agriculture & mining	0-17	51	0.0848	-0.0043
Food processing etc	20-27	82	0.0382	0.0074
Chemicals & refining	28-29	78	0.0234	0.0027
Metals	30-34	65	0.0439	0.0070
Machinery & equipment	35	72	0.0380	0.0037
Electrical Equipment	36	77	0.0601	0.0001
Transport equipment	37	33	0.0800	0.0105
Misc. manufacturing	38-39	78	0.0155	0.0027
Transport etc.	40-49	108	0.0810	0.0035
Wholesale & retail trade	50-59	122	0.0357	0.0034
Banking	60	145	0.1240	-0.0037
Other finance	61-67	98	0.0273	0.0061
Commercial Services	70-79	114	0.0149	0.0031
Professional Services	80-89	45	0.0172	0.0023
Total		1168		

Note: This table gives the average correlations within and between 14 groups selected based on one and two digits SIC codes. It shows the number of stocks in each sector, the average pair-wise correlation of stock returns within the sector as well as the average pair-wise correlations of returns of stocks in a given sector with those outside the sector.

S5 Data used in the empirical application

S-5.1 Security excess returns

Monthly returns (inclusive of dividends) for NYSE and NASDAQ stocks from CRSP with codes 10 and 11 were downloaded on July 2 2022 from Wharton Research Data Services. They were converted to excess returns by subtracting the risk free rate, which was taken from Kenneth French's data base. To obtain balanced panels of stock returns and factors only variables where there was data for the full sample under consideration were used. Excess returns are measured in percent per month. To avoid outliers influencing the results, stocks with a kurtosis greater than 16 were excluded.

Four main samples were considered, each had 20 years of data, $T = 240$, ending in 2015m12, 2017m12, 2019m12, 2021m12. Thus the earliest observation used is for

1996m1. Sub-samples of the main samples of size $T = 120$ and $T = 60$ ending at the same dates were also examined. Table DS7 gives averages of the summary statistics across the individual stocks for the various samples. These are very similar over these four periods. Mean returns were high and substantially greater than the median reflecting the skewness of returns, which was slightly less in the last period. Filtering out the stocks with very high kurtosis removed about 100 of the roughly 1200 stocks in each period and reduced mean return, standard deviation, skewness as well as kurtosis.

The 5 ($T = 60$) and 10 ($T = 120$) year sub-samples of the main samples, ending at the same dates, showed very similar patterns. Because of a requirement for a balanced panel, the shorter the sample the more stocks will be eligible for inclusion. Compared with around 1200 in the 20 year sample there were around 2,000 in the 10 year sub-sample and around 2500 in the 5 year sub-sample. Again filtering by kurtosis reduced the number of stocks by about 100. There is more variation in means and medians in the shorter sub-samples and the shorter the sample the lower the average kurtosis is.

Table DS7: Summary statistics for monthly returns in percent for NYSE and NASDAQ stocks code 10 and 11 for 20 year ($T = 240$), 10 year ($T = 120$) and 5 year ($T = 60$) samples ending at end of specified year

	All stocks				Stocks with kurtosis < 16			
End date	2015	2017	2019	2021	2015	2017	2019	2021
<i>Panel A: 20 year period, $T = 240$</i>								
Mean	1.38	1.35	1.30	1.37	1.33	1.28	1.25	1.33
Median	0.74	0.70	0.76	0.84	0.80	0.78	0.85	0.94
S.D.	12.44	12.61	12.22	11.90	11.60	11.64	11.31	10.99
Skewness	0.69	0.74	0.72	0.64	0.42	0.44	0.40	0.32
Kurtosis	8.68	9.08	9.22	9.24	6.10	6.29	6.29	6.19
n	1181	1243	1276	1289	1090	1132	1143	1175
<i>Panel B: 10 year period, $T = 120$</i>								
Mean	1.00	1.22	1.24	1.55	0.98	1.17	1.23	1.49
Median	0.52	0.70	0.78	0.89	0.59	0.77	0.84	1.01
S.D.	12.36	12.68	10.70	11.83	11.67	11.80	10.27	10.70
Skewness	0.48	0.50	0.44	0.50	0.29	0.30	0.35	0.28
Kurtosis	6.93	7.08	5.32	7.02	5.57	5.62	4.64	5.33
n	2045	2024	1925	1871	1929	1907	1873	1766
<i>Panel C: 5 year period, $T = 60$</i>								
Mean	0.98	1.39	0.82	1.60	0.96	1.35	0.80	1.47
Median	0.42	0.77	0.20	0.43	0.49	0.82	0.29	0.61
S.D.	10.89	10.62	11.88	14.91	10.41	10.15	11.28	13.28
Skewness	0.44	0.46	0.38	0.48	0.35	0.38	0.29	0.28
Kurtosis	4.84	4.74	4.68	6.21	4.34	4.29	4.20	5.04
n	2600	2425	2497	2512	2541	2373	2439	2388

Note: This table shows the average values of the summary statistics of individual stocks described in Section 3, and the number of stocks (n), for each sample.

S-5.2 Risk factors

For the empirical applications we combined the 5 Fama-French factors with the 207 factors of Chen and Zimmermann (2022), both downloaded on 6 July 2022. The available risk factors at the end of each of the four 240 months samples ending in 2015,

2017, 2019 and 2021 were then screened and factors whose correlations (in absolute value) with the market factor were larger than 0.70 were dropped. The basic idea was to remove factors that were closely correlated with market factor. However, the application of this filter only reduced the number of factors in the active set by around 9-11. See Table 4 of the main paper. The summary statistics for the factors in the active set are summarized in Table DS8. It reports mean, median, pair-wise correlation, standard deviation (S.D.), skewness and kurtosis of the statistics indicated in the sub-headings of the tables for the K factors included in the active set for samples of size $T = 240$ months ending in December of 2015, 2017, 2019 and 2021. The summary statistics reported for the "Mean" on the left panel of Table DS8 are based on the time series means of the individual factors, those under "Median" are based on the time series medians of the individual factors, and so on.

Table DS8: Summary statistics for mean, median, standard deviation, pairwise correlations, skewness and kurtosis of factors for $T = 240$ samples at the end of specified years (2015, 2017, 2019 and 2021)

End date	2015	2017	2019	2021	2015	2017	2019	2021
$m(\# \text{ Factors})$	190	191	190	178	190	191	190	178
	Mean				Median			
Mean	0.51	0.47	0.41	0.32	0.42	0.37	0.31	0.28
Median	0.45	0.39	0.33	0.26	0.31	0.28	0.22	0.20
S.D.	0.40	0.40	0.38	0.30	0.53	0.50	0.50	0.44
Skewness	1.88	2.50	2.28	0.76	1.59	1.89	1.58	1.38
Kurtosis	10.54	16.35	14.42	3.82	8.99	10.37	9.16	6.68
	Standard deviation				Pair-wise correlation			
Mean	3.97	3.90	3.78	3.33	0.22	0.22	0.22	0.21
Median	3.51	3.47	3.37	2.99	0.18	0.17	0.17	0.16
S.D.	2.21	2.16	2.12	1.68	0.17	0.17	0.17	0.17
Skewness	1.15	1.23	1.29	1.05	0.86	0.88	0.88	0.92
Kurtosis	4.25	4.69	4.96	4.17	2.88	2.93	2.93	2.98
	Skewness				Kurtosis			
Mean	0.14	0.21	0.23	0.04	8.51	8.79	9.42	7.03
Median	0.13	0.18	0.14	0.06	6.44	6.66	7.09	5.86
S.D.	1.15	1.21	1.27	1.01	5.76	6.55	7.12	5.28
Skewness	0.15	0.55	0.61	-1.03	2.34	3.18	2.96	5.01
Kurtosis	3.98	5.28	5.24	7.82	9.65	17.71	15.70	37.21

Note: The T=240 sample was used to select factors and only factors where the absolute correlation coefficient with the market factor is less than 0.70 are included.

S6 Pooled R squared and factor strengths

Lemma S6.1 Consider the factor model

$$r_{it} = a_i + \sum_{k=1}^K \beta_{ik} f_{kt} + u_{it} = a_i + \beta_i' \mathbf{f}_t + u_{it}, \text{ for } i = 1, 2, \dots, n; \quad t = 1, 2, \dots, T, \quad (\text{S6.1})$$

and consider the following adjusted pooled measure of fit

$$\overline{PR}^2 = 1 - \frac{\widehat{\sigma}_{nT}^2}{(nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T (r_{it} - \bar{r}_{i\cdot})^2}, \quad (\text{S6.2})$$

where $\widehat{\sigma}_{nT}^2$ is the bias-corrected estimator of $(nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T E(u_{it}^2) = n \sum_{i=1}^n \sigma_i^2 = \bar{\sigma}_n^2$, given by (44) $\bar{r}_{i\cdot} = T^{-1} \sum_{t=1}^T r_{it}$, and $\bar{u}_{i\cdot} = T^{-1} \sum_{t=1}^T u_{it}$. Then under Under Assumptions 4, 1 and 2 we have

$$\overline{PR}_{nT}^2 = \sum_{k=1}^K \ominus (n^{\alpha_k - 1}) + O_p \left(T^{-1/2} n^{-1 + \frac{\alpha_{max} + \alpha_\gamma}{2}} \right), \quad (\text{S6.3})$$

where α_k is the strength of factor f_{tk} , $\alpha_{max} = \max_k(\alpha_k)$, and α_γ is the strength of the missing factor.

Proof. Using (46) we first recall that

$$\widehat{\sigma}_{nT}^2 - \bar{\sigma}_n^2 = O_p \left(T^{-1/2} n^{-1/2} \right). \quad (\text{S6.4})$$

Now averaging (S6.1) over t and forming deviations of r_{it} from its time average, $\bar{r}_{i\cdot}$, we have (note that $\hat{\boldsymbol{\mu}}_T = T^{-1} \sum_{t=1}^T \mathbf{f}_t$)

$$r_{it} - \bar{r}_{i\cdot} = u_{it} - \bar{u}_{i\cdot} + \boldsymbol{\beta}'_i (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T).$$

Using this result we have

$$\begin{aligned} (nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T (r_{it} - \bar{r}_{i\cdot})^2 &= (nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T (u_{it} - \bar{u}_{i\cdot})^2 + n^{-1} \sum_{i=1}^n \boldsymbol{\beta}'_i \hat{\boldsymbol{\Sigma}}_f \boldsymbol{\beta}_i \\ &\quad - 2(nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T (u_{it} - \bar{u}_{i\cdot}) \boldsymbol{\beta}'_i (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T), \end{aligned} \quad (\text{S6.5})$$

where $\hat{\boldsymbol{\Sigma}}_f = T^{-1} \sum_{t=1}^T (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T) (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T)'$. For the first term we have

$$(nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T (u_{it} - \bar{u}_{i\cdot})^2 - \bar{\sigma}_n^2 = O_p(n^{-1/2} T^{-1/2}), \quad (\text{S6.6})$$

which follows from the proof of Theorem 2 by setting $\beta_i = \mathbf{0}$ and $\mathbf{f}_t = \mathbf{0}$ in Section B.2. For the cross product term we have

$$\begin{aligned}
& (nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T (u_{it} - \bar{u}_{i\circ}) \beta_i' (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T) \\
&= (nT)^{-1} \sum_{i=1}^n \beta_i' \sum_{t=1}^T (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T) (u_{it} - \bar{u}_{i\circ}) = (nT)^{-1} \sum_{i=1}^n \beta_i' \sum_{t=1}^T (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T) u_{it} \\
&- (nT)^{-1} \sum_{i=1}^n \bar{u}_{i\circ} \beta_i' \sum_{t=1}^T (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T) = (nT)^{-1} \sum_{i=1}^n \beta_i' \sum_{t=1}^T (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T) u_{it} = p_{nT}.
\end{aligned}$$

Also, using (27),

$$p_{nT} = (nT)^{-1} \sum_{i=1}^n \beta_i' \sum_{t=1}^T (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T) (\gamma_i g_t + v_{it}) = p_{1,nT} + p_{2,nT},$$

where

$$p_{1,nT} = \left(n^{-1} \sum_{i=1}^n \gamma_i \beta_i' \right) \left[T^{-1} \sum_{t=1}^T (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T) g_t \right],$$

and

$$p_{2,nT} = (nT)^{-1} \sum_{i=1}^n \beta_i' \sum_{t=1}^T (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T) v_{it}.$$

Under Assumption 3, $(\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T)$ and g_t are distributed independently and g_t are serially independent with $E(g_t) = 0$ and $E(g_t^2) = 1$, and it follows that

$$\text{Var} \left[T^{-1} \sum_{t=1}^T (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T) g_t \right] = E \left[T^{-2} \sum_{t=1}^T (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T)^2 E(g_t^2) \right] = T^{-1} E(T^{-1} \mathbf{F}' \mathbf{M}_T \mathbf{F}) = O(T^{-1}).$$

Hence, $T^{-1} \sum_{t=1}^T (\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T) g_t = O_p(T^{-1/2})$. Also (see (30) and (41))

$$\begin{aligned}
\left\| n^{-1} \sum_{i=1}^n \gamma_i \beta_i' \right\| &\leq n^{-1} \sum_{i=1}^n |\gamma_i| \|\beta_i\| = n^{-1} \sum_{i=1}^n |\gamma_i| (\beta_i' \beta_i)^{1/2} \\
&\leq \left(n^{-1} \sum_{i=1}^n |\gamma_i|^2 \right)^{1/2} \left(n^{-1} \sum_{i=1}^n \beta_i' \beta_i \right)^{1/2} = O_p \left(n^{\frac{-1+\alpha\gamma}{2}} \right) O_p \left(n^{\frac{-1+\alpha_{max}}{2}} \right) \\
&= O_p \left(n^{-1+\frac{\alpha_{max}+\alpha\gamma}{2}} \right).
\end{aligned}$$

Hence, $p_{1,nT} = O_p\left(T^{-1/2}n^{-1+\frac{\alpha_{max}+\alpha_\gamma}{2}}\right)$. Consider $p_{2,nT}$ and recall that under Assumption 4 v_{it} are serially independent, have zero means and are distributed independently of $(\mathbf{f}_t - \hat{\boldsymbol{\mu}}_T)$ and $\boldsymbol{\beta}_i$. Then $E(p_{2,nT}) = 0$ and

$$Var(p_{2,nT} | \mathbf{F}) = \frac{1}{nT} \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij,v} \boldsymbol{\beta}'_i \hat{\boldsymbol{\Sigma}}_f \boldsymbol{\beta}_j \right),$$

where $E(v_{it}v_{jt}) = \sigma_{ij,v}$. Also,

$$\left\| \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij,v} \boldsymbol{\beta}'_i \hat{\boldsymbol{\Sigma}}_f \boldsymbol{\beta}_j \right\| \leq (\sup_i \|\boldsymbol{\beta}_i\|)^2 \|\hat{\boldsymbol{\Sigma}}_f\| \left(\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n |\sigma_{ij,v}| \right),$$

and by assumption $\sup_i \|\boldsymbol{\beta}_i\| < C$, $E\|\hat{\boldsymbol{\Sigma}}_f\| < C$, and $n^{-1} \sum_{i=1}^n \sum_{j=1}^n |\sigma_{ij,v}| = O(1)$. Hence, $Var(p_{2,nT}) = O_p(n^{-1}T^{-1})$, and it follows that $p_{2,nT} = O_p(T^{-1/2}n^{-1/2})$, and overall (since $\alpha_\gamma < 1/2$ and $\alpha_{max} \leq 1$)

$$p_{nT} = O_p(T^{-1/2}n^{-1/2}) + O_p\left(T^{-1/2}n^{-1+\frac{\alpha_{max}+\alpha_\gamma}{2}}\right) = O_p\left(T^{-1/2}n^{-1+\frac{\alpha_{max}+\alpha_\gamma}{2}}\right). \quad (\text{S6.7})$$

Using (S6.6) and (S6.7) in (S6.5), we now have

$$(nT)^{-1} \sum_{i=1}^n \sum_{t=1}^T (r_{it} - \bar{r}_{io})^2 = \bar{\sigma}_n^2 + n^{-1} \sum_{i=1}^n \boldsymbol{\beta}'_i \hat{\boldsymbol{\Sigma}}_f \boldsymbol{\beta}_i + O_p\left(T^{-1/2}n^{-1+\frac{\alpha_{max}+\alpha_\gamma}{2}}\right).$$

Using this result and (S6.4) in (S6.2) yields

$$\overline{PR}_{nT}^2 = 1 - \frac{\bar{\sigma}_n^2 + O_p(T^{-1/2}n^{-1/2})}{\bar{\sigma}_n^2 + n^{-1} \sum_{i=1}^n \boldsymbol{\beta}'_i \hat{\boldsymbol{\Sigma}}_f \boldsymbol{\beta}_i + O_p(T^{-1/2}n^{-1/2}) + O_p\left(T^{-1/2}n^{-1+\frac{\alpha_{max}+\alpha_\gamma}{2}}\right)}.$$

Since $O_p(T^{-1/2}n^{-1/2})$ is dominated by $\left(T^{-1/2}n^{-1+\frac{\alpha_{max}+\alpha_\gamma}{2}}\right)$, we end up with

$$\overline{PR}_{nT}^2 = \frac{n^{-1} \sum_{i=1}^n \boldsymbol{\beta}'_i \hat{\boldsymbol{\Sigma}}_f \boldsymbol{\beta}_i / \bar{\sigma}_n^2 + O_p\left(T^{-1/2}n^{-1+\frac{\alpha_{max}+\alpha_\gamma}{2}}\right)}{1 + n^{-1} \sum_{i=1}^n \boldsymbol{\beta}'_i \hat{\boldsymbol{\Sigma}}_f \boldsymbol{\beta}_i / \bar{\sigma}_n^2 + O_p\left(T^{-1/2}n^{-1+\frac{\alpha_{max}+\alpha_\gamma}{2}}\right)}. \quad (\text{S6.8})$$

Hence, the order of \overline{PR}_{nT}^2 is governed by the pooled signal-to-noise ratio defined by

$$s_{nT}^2 = \frac{n^{-1} \sum_{i=1}^n \boldsymbol{\beta}'_i \hat{\boldsymbol{\Sigma}}_f \boldsymbol{\beta}_i}{\bar{\sigma}_n^2}.$$

However, under Assumption 1

$$\lambda_{\min}(\hat{\boldsymbol{\Sigma}}_f) \frac{n^{-1} \sum_{i=1}^n \boldsymbol{\beta}'_i \boldsymbol{\beta}_i}{\bar{\sigma}_n^2} \leq s_{nT}^2 \leq \lambda_{\max}(\hat{\boldsymbol{\Sigma}}_f) \frac{n^{-1} \sum_{i=1}^n \boldsymbol{\beta}'_i \boldsymbol{\beta}_i}{\bar{\sigma}_n^2}, \quad (\text{S6.9})$$

where $c < \lambda_{\min}(\hat{\boldsymbol{\Sigma}}_f) < \lambda_{\max}(\hat{\boldsymbol{\Sigma}}_f) < C$. Hence,

$$c \left(\frac{n^{-1} \sum_{i=1}^n \boldsymbol{\beta}'_i \boldsymbol{\beta}_i}{\bar{\sigma}_n^2} \right) \leq s_{nT}^2 \leq C \left(\frac{n^{-1} \sum_{i=1}^n \boldsymbol{\beta}'_i \boldsymbol{\beta}_i}{\bar{\sigma}_n^2} \right),$$

and it must be that

$$s_{nT}^2 = \ominus \left(n^{-1} \sum_{i=1}^n \boldsymbol{\beta}'_i \boldsymbol{\beta}_i \right) = \ominus \left[\sum_{k=1}^K \left(n^{-1} \sum_{i=1}^n \beta_{ik}^2 \right) \right].$$

Also, under Assumption 2, $n^{-1} \sum_{i=1}^n \beta_{ik}^2 = \ominus(n^{\alpha_k - 1})$. Hence,

$$s_{nT}^2 = \sum_{k=1}^K \ominus(n^{\alpha_k - 1}),$$

which in view of (S6.8) now yields (S6.3), as desired. ■

References

- Bailey, N., G. Kapetanios and M. H. Pesaran (2021). "Measurement of factor strength: theory and practice", *Journal of Applied Econometrics*, 36, 587-613.
- Bailey, N., M. H. Pesaran, and L.V. Smith (2019). "A multiple testing approach to the regularisation of sample correlation matrices", *Journal of Econometrics*, 208, 507-534.
- Chen, A. Y. and T. Zimmermann (2022). "Open source cross-sectional asset pricing", *Critical Finance Review*, 11, 207-264.

Online Supplement B: Monte Carlo Results

To

**"The role of pricing errors in linear asset pricing models with strong,
semi-strong, and latent factors"**

M. Hashem Pesaran

University of Southern California, and Trinity College, Cambridge

Ron P. Smith

Birkbeck, University of London

February 17, 2023

1 Introduction

This online supplement provides detailed Monte Carlo results for all experiments and risk factors. The summary tables below give the bias, RMSE and size ($\times 100$), for the DGP with one strong ($\alpha_M = 1$) and two semi-strong factors ($\alpha_H = 0.85$, $\alpha_S = 0.65$) for the two-step and the bias-corrected estimators of $\phi_0 = (\phi_M, \phi_H, \phi_S)'$, for different sample sizes. These are given for the twelve experimental designs listed in Table S-1 below. For experiments 8 and 9 we also report the results with larger values for the parameter of the pricing errors ($\alpha_\eta = 0.50$) and the spatial coefficients ($\rho_\varepsilon = 0.85$), denoted as Experiments 8a and 9a, respectively. Following each table the empirical power functions for the bias corrected estimator of ϕ_0 are displayed for different sample sizes. The threshold estimator of the covariance matrix described in Section 3.2 of the main paper is used in computing the standard errors of the tests.

2 List of Monte Carlo Experiments

The full list of Monte Carlo experiments is provided in Table S-1. Six designs, the odd numbered ones, have errors in the return equations that are Gaussian, six, the even numbered ones, have errors that are t distributed with 5 degrees of freedom. Designs 3 and 4 add GARCH effects in the factor errors to designs 1 and 2, respectively. Designs 5 and 6, add the pricing error, η_i , to designs 3 and 4, and designs 7 and 8 further add the missing factor, g_t , to the error of the return equations. Designs 1-8 have a diagonal covariance matrix for the idiosyncratic errors, v_{it} . Designs 9 and 10 introduce spatial errors in the idiosyncratic errors, v_{it} , and continue to allow for GARCH effects, pricing errors, and a missing factor. Designs 11 and 12 generate v_{it} with a block covariance matrix structure, instead of the spatial pattern assumed in designs 9 and 10. All experiments are implemented using $R = 2,000$ replications.

Table S-1: List of experimental designs and their parameter values

	Error distribution	GARCH effects	Pricing errors	Missing factor	Error covariance
1	Gaussian	$b_k = 0, c_k = 0$	No	No	$\sigma_{ij} = 0, i \neq j$
2	$t(5)$	$b_k = 0, c_k = 0$	No	No	$\sigma_{ij} = 0, i \neq j$
3	Gaussian	$b_k = 0.8, c_k = 0.1$	No	No	$\sigma_{ij} = 0, i \neq j$
4	$t(5)$	$b_k = 0.8, c_k = 0.1$	No	No	$\sigma_{ij} = 0, i \neq j$
5	Gaussian	$b_k = 0.8, c_k = 0.1$	$\alpha_\eta = 0.3$	No	$\sigma_{ij} = 0, i \neq j$
6	$t(5)$	$b_k = 0.8, c_k = 0.1$	$\alpha_\eta = 0.3$	No	$\sigma_{ij} = 0, i \neq j$
7	Gaussian	$b_k = 0.8, c_k = 0.1$	$\alpha_\eta = 0.3$	$\alpha_\gamma = 0.5$	$\sigma_{ij} = 0, i \neq j$
8	$t(5)$	$b_k = 0.8, c_k = 0.1$	$\alpha_\eta = 0.3$	$\alpha_\gamma = 0.5$	$\sigma_{ij} = 0, i \neq j$
9	Gaussian	$b_k = 0.8, c_k = 0.1$	$\alpha_\eta = 0.3$	$\alpha_\gamma = 0.5$	$\rho_\varepsilon = 0.5$
10	$t(5)$	$b_k = 0.8, c_k = 0.1$	$\alpha_\eta = 0.3$	$\alpha_\gamma = 0.5$	$\rho_\varepsilon = 0.5$
11	Gaussian	$b_k = 0.8, c_k = 0.1$	$\alpha_\eta = 0.3$	$\alpha_\gamma = 0.5$	Block
12	$t(5)$	$b_k = 0.8, c_k = 0.1$	$\alpha_\eta = 0.3$	$\alpha_\gamma = 0.5$	Block

Notes: t -distributed errors are denoted by $t(5)$, b_k and c_k are the parameters of the GARCH(1,1), α_η is the strength of the pricing errors, α_γ refers to the strength of the missing factor, $\sigma_{ij} = 0, i \neq j$ means that the error covariance is diagonal, ρ_ε is the coefficient of spatial error process, and "Block" means that the error covariance matrix is block diagonal. See the online supplement A for further details.

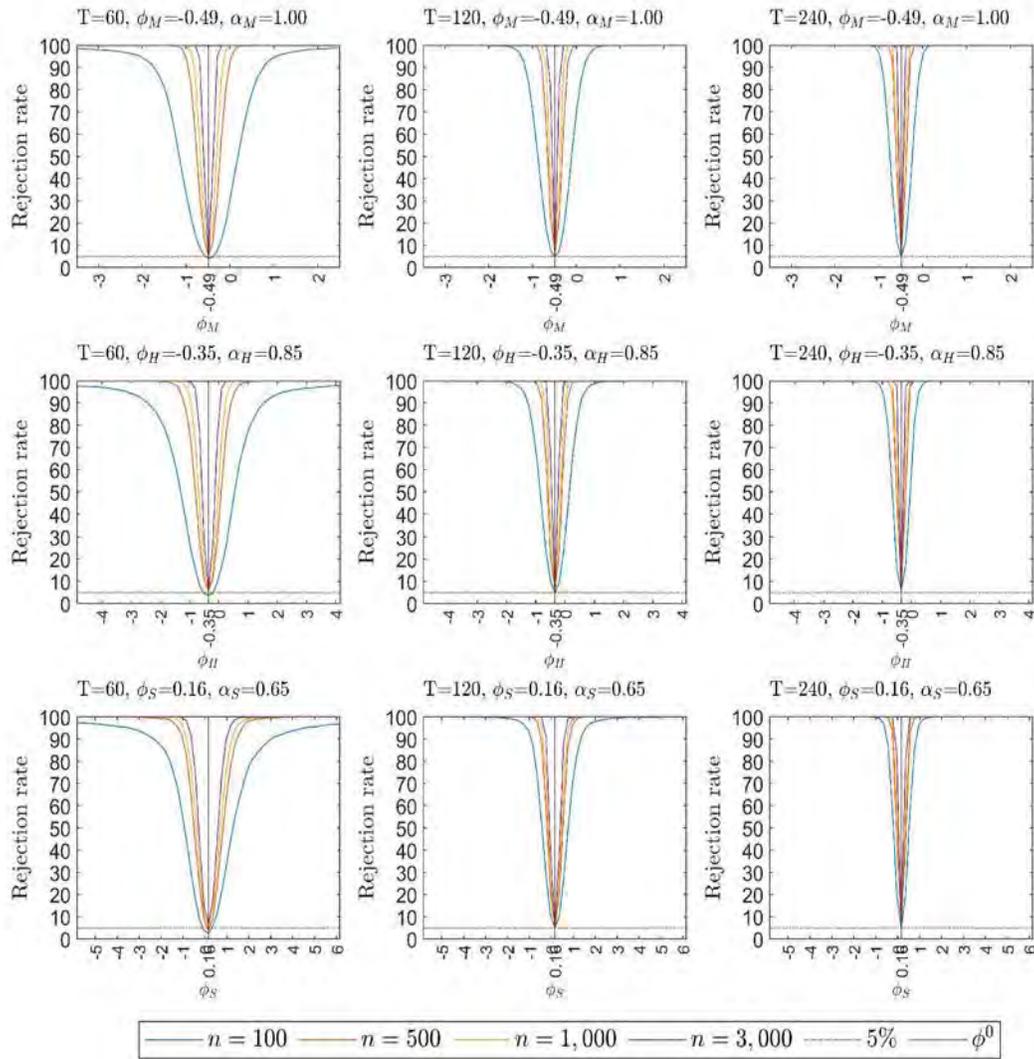
The simulation design, as presented in Table S-1, is aligned with the naming convention employed for tables and figures. Specifically, if a table is denoted as Table S-A-EX, it signifies that the table pertains to Experiment X when the GDP and the panel regressions correctly include one strong and two semi-strong factors. On the other hand, a table labeled as Table S-B-EX relates to Experiment X when the DGP includes one strong and two semi-strong factors, comparing the results when strong and semi-strong factors are included (correct specification) with the ones where only the strong factor is included (incorrect specification). Likewise, Table S-C-EX provides the summary results for Experiment X, when the DGP contains one strong and two weak factors, comparing the results to the case when strong and weak factors are included (correct specification) with the results obtained when the weak factors are excluded (incorrect specification). The aforementioned nomenclature also applies to the figures that present the empirical power functions.

Table S-A-E1: Bias, RMSE and size for the two-step and bias-corrected (BC) estimators of ϕ , for Experiment 1 with one strong and two semi-strong factors

	n	Bias(x100)		RMSE(x100)		Size(x100)		
		Two Step	BC	Two Step	BC	Two Step	BC	
$\phi_M = -0.49, \alpha_M = 1$	$T = 60$	100	0.51	-1.55	26.19	86.14	11.10	4.00
		500	-0.06	0.28	17.05	14.33	28.25	5.40
		1,000	-0.43	0.10	15.80	10.12	41.35	6.20
		3,000	-0.86	-0.01	14.77	5.56	60.90	5.70
	$T = 120$	100	0.24	1.84	17.47	91.35	7.70	4.70
		500	0.18	0.15	9.34	8.99	13.95	6.15
		1,000	-0.04	0.01	7.57	6.22	19.15	5.65
		3,000	-0.32	-0.06	6.27	3.49	36.65	4.95
	$T = 240$	100	0.24	-0.15	12.11	13.11	5.40	4.55
		500	0.31	0.09	5.75	5.75	8.00	4.85
		1,000	0.22	0.03	4.34	4.06	10.60	5.30
		3,000	0.07	-0.00	3.06	2.35	18.80	5.10
$\phi_H = -0.35, \alpha_H = 0.85$	$T = 60$	100	2.63	-0.89	29.26	140.65	17.55	3.65
		500	4.07	-1.16	25.10	24.87	45.65	5.90
		1,000	4.58	-0.12	25.19	18.25	57.25	6.30
		3,000	4.81	0.30	26.13	11.76	75.10	5.90
	$T = 120$	100	2.51	3.60	19.56	190.94	9.95	4.75
		500	2.97	-0.47	13.83	13.46	25.65	5.55
		1,000	3.34	0.02	13.68	10.20	39.85	6.00
		3,000	3.57	0.15	13.85	6.48	61.55	5.50
	$T = 240$	100	1.64	-0.50	13.44	16.67	7.75	6.15
		500	1.99	-0.39	7.87	8.01	12.95	5.55
		1,000	2.29	-0.15	7.28	6.04	23.05	5.40
		3,000	2.64	0.02	6.97	3.82	44.55	4.70
$\phi_S = 0.16, \alpha_S = 0.65$	$T = 60$	100	-20.09	1.57	37.57	256.48	24.55	2.60
		500	-23.78	1.80	37.42	37.75	56.20	4.05
		1,000	-25.54	1.05	38.85	32.14	68.60	5.45
		3,000	-28.27	0.35	42.26	24.96	82.80	7.00
	$T = 120$	100	-12.99	-6.16	25.50	378.75	16.25	4.90
		500	-16.99	0.65	23.97	19.49	46.85	4.80
		1,000	-18.80	0.56	25.41	16.40	62.60	5.55
		3,000	-22.08	0.44	28.91	12.73	80.55	5.75
	$T = 240$	100	-8.52	0.16	17.30	19.66	10.80	4.55
		500	-11.19	0.19	15.08	11.66	36.15	4.90
		1,000	-12.71	0.34	15.72	9.27	55.90	4.60
		3,000	-15.93	0.18	18.70	7.02	81.90	5.60

Notes: The DGP for Experiment 1 allows for Gaussian errors, no GARCH effects, without pricing errors, no missing factors, and without spatial/block error cross dependence. For further details of the experiments, see Table S-1.

Figure S-A-E1: Power functions of the bias-corrected estimators of ϕ_M , ϕ_H and ϕ_S for Experiment 1



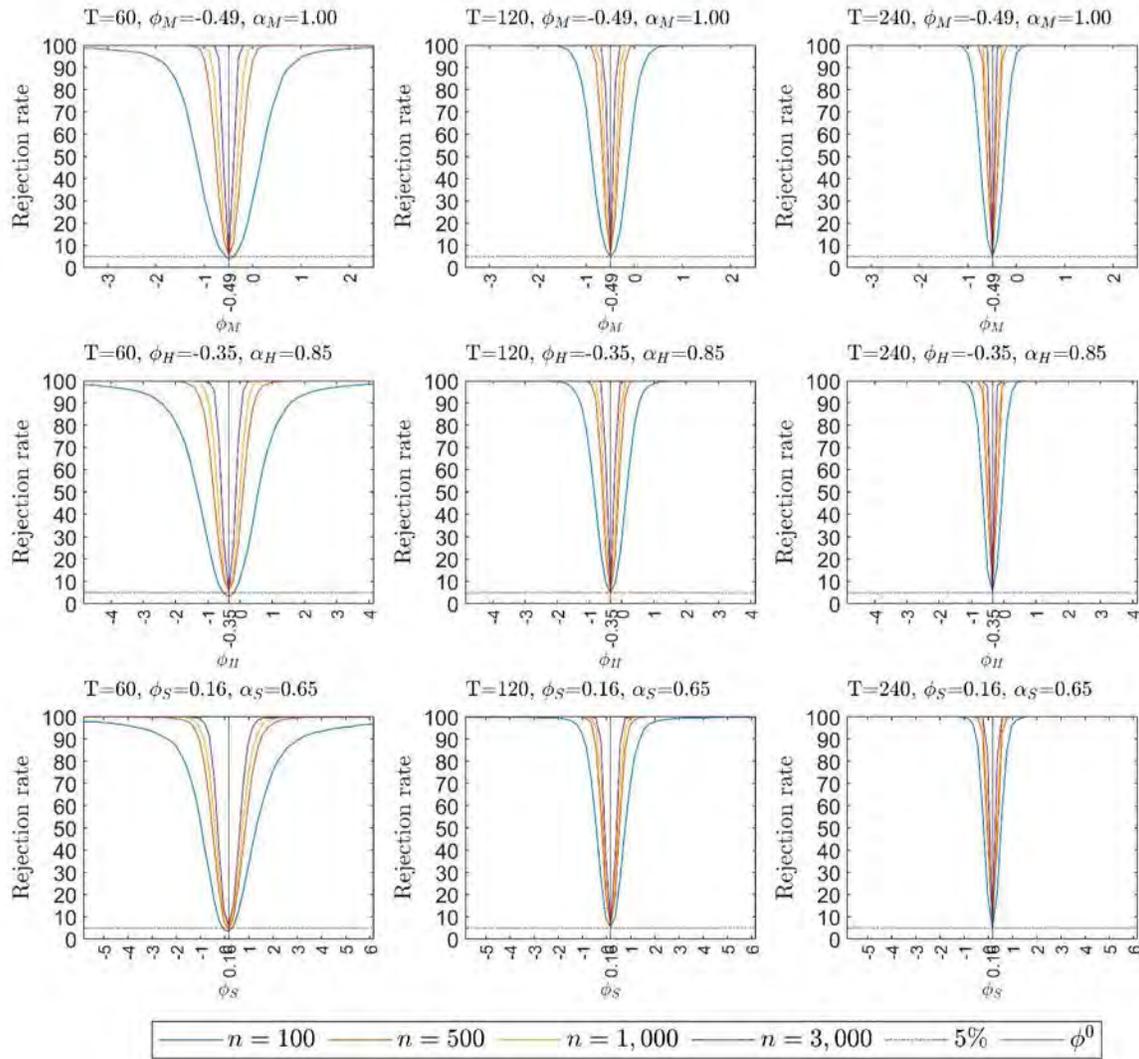
Note: See the notes to Table S-A-E1.

Table S-A-E2: Bias, RMSE and size for the two-step and bias-corrected (BC) estimators of ϕ , for Experiment 2 with one strong and two semi-strong factors

	n	Bias(x100)		RMSE(x100)		Size(x100)		
		Two Step	BC	Two Step	BC	Two Step	BC	
$\phi_M = -0.49, \alpha_M = 1$	T = 60	100	0.65	0.27	26.62	45.08	11.75	3.70
		500	-0.07	0.27	17.26	14.47	28.25	5.75
		1,000	-0.50	0.04	15.74	10.24	39.75	6.10
		3,000	-0.87	-0.05	14.80	5.72	60.70	5.90
	T = 120	100	0.42	0.02	17.50	20.45	7.05	4.90
		500	0.25	0.23	9.47	9.16	14.80	6.60
		1,000	-0.03	0.03	7.46	6.12	19.20	5.55
		3,000	-0.36	-0.09	6.31	3.52	36.70	5.65
	T = 240	100	0.27	-0.14	12.17	13.18	6.20	5.05
		500	0.34	0.11	5.84	5.81	8.15	5.05
		1,000	0.22	0.04	4.34	4.07	10.40	5.55
		3,000	0.04	-0.03	3.07	2.36	19.15	5.25
$\phi_H = -0.35, \alpha_H = 0.85$	T = 60	100	2.88	-2.77	29.60	81.53	17.70	3.25
		500	4.18	-0.95	25.36	25.93	44.85	5.80
		1,000	4.62	-0.02	25.27	18.42	57.40	5.00
		3,000	4.75	0.13	26.17	12.23	74.80	5.95
	T = 120	100	2.49	-0.71	19.67	28.58	9.70	4.85
		500	3.17	-0.13	13.89	13.46	25.75	5.10
		1,000	3.30	-0.05	13.75	10.34	40.05	6.00
		3,000	3.52	0.08	13.84	6.61	62.40	6.05
	T = 240	100	1.75	-0.39	13.41	16.61	7.70	5.65
		500	2.08	-0.27	7.96	8.01	13.75	5.10
		1,000	2.23	-0.23	7.22	6.08	22.20	5.15
		3,000	2.64	0.02	6.95	3.83	43.50	4.60
$\phi_S = 0.16, \alpha_S = 0.65$	T = 60	100	-20.16	5.40	37.93	125.25	24.85	3.60
		500	-23.88	1.79	37.72	40.47	55.30	4.25
		1,000	-25.44	1.10	38.75	33.04	67.85	5.35
		3,000	-28.24	0.82	42.24	25.93	82.05	6.45
	T = 120	100	-13.00	2.26	25.41	38.09	16.65	5.40
		500	-17.08	0.48	24.21	20.35	46.35	5.55
		1,000	-18.80	0.53	25.43	16.73	62.75	5.50
		3,000	-22.12	0.38	28.88	12.87	79.90	5.70
	T = 240	100	-8.38	0.35	17.38	19.88	10.75	4.90
		500	-11.10	0.33	14.99	11.64	35.40	5.35
		1,000	-12.66	0.43	15.60	9.15	55.40	4.55
		3,000	-15.88	0.29	18.65	6.97	80.75	4.95

Notes: The DGP for Experiment 2 allows for $t(5)$ distributed errors, no GARCH effects, without pricing errors, no missing factors, and without spatial/block error cross dependence. For further details of the experiments, see Table S-1.

Figure S-A-E2: Empirical Power Functions, experiment 2, for coefficient of the semi-strong factors



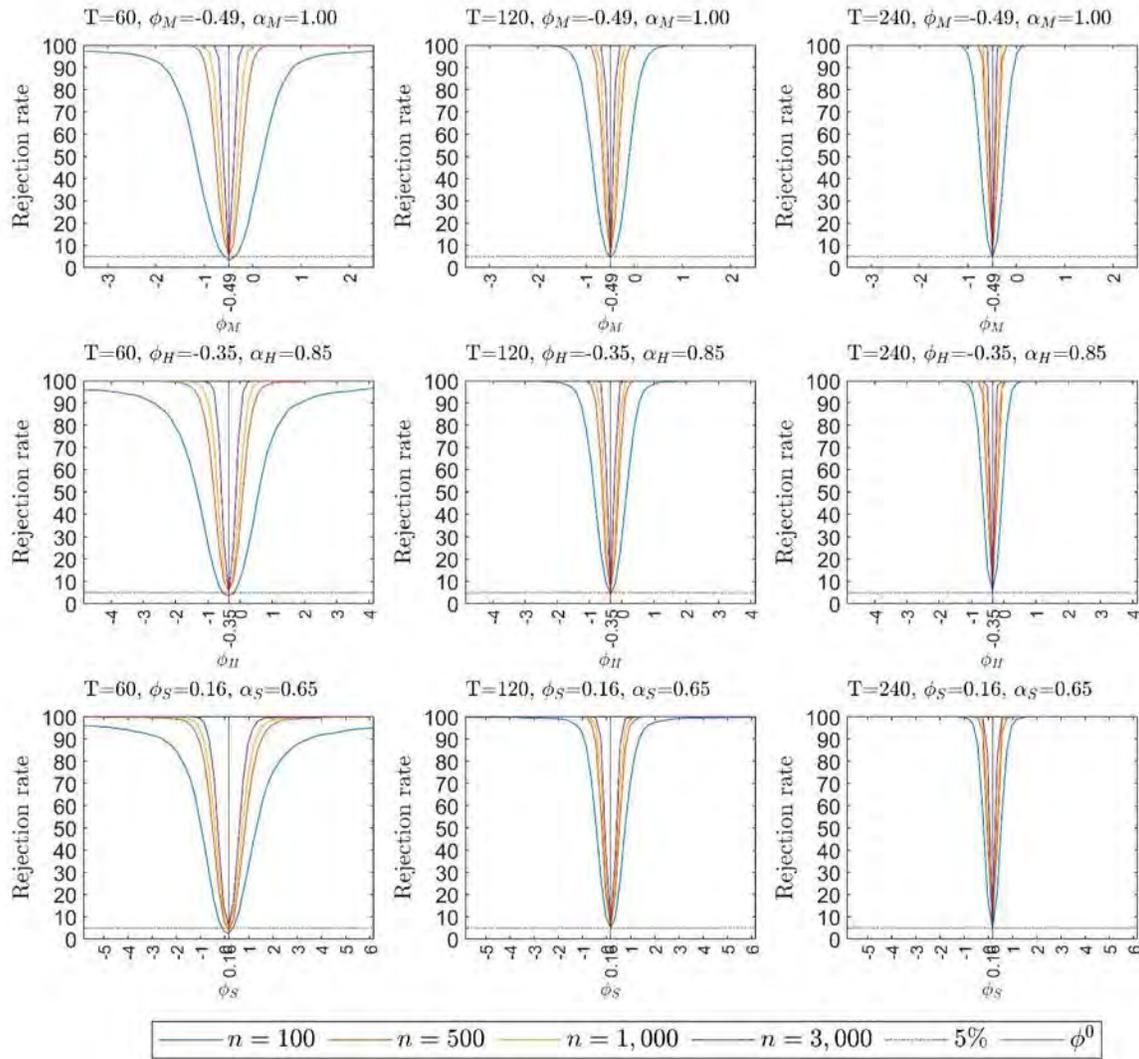
Note: See the notes to Table S-A-E2.

Table S-A-E3: Bias, RMSE and size for the two-step and bias-corrected (BC) estimators of ϕ , for Experiment 3 with one strong and two semi-strong factors

	n	Bias(x100)		RMSE(x100)		Size(x100)		
		Two Step	BC	Two Step	BC	Two Step	BC	
$\phi_M = -0.49, \alpha_M = 1$	$T = 60$	100	0.51	-6.00	26.18	227.89	11.05	3.40
		500	-0.14	0.29	17.03	14.71	28.20	5.65
		1,000	-0.51	0.09	15.81	10.33	40.30	6.35
		3,000	-0.92	0.00	14.83	5.65	61.15	5.65
	$T = 120$	100	0.21	-0.19	17.45	20.63	7.85	4.50
		500	0.13	0.15	9.35	9.05	13.80	6.25
		1,000	-0.09	0.02	7.61	6.25	19.95	5.75
		3,000	-0.37	-0.05	6.31	3.50	36.65	4.90
	$T = 240$	100	0.25	-0.13	12.15	13.17	5.35	4.45
		500	0.31	0.10	5.77	5.78	8.00	5.20
		1,000	0.21	0.05	4.36	4.07	11.15	5.25
		3,000	0.06	0.00	3.07	2.36	18.80	5.20
$\phi_H = -0.35, \alpha_H = 0.85$	$T = 60$	100	2.67	-7.85	28.98	155.63	17.70	3.80
		500	4.06	-1.08	24.84	26.10	45.75	6.05
		1,000	4.57	-0.15	24.93	18.93	56.85	6.35
		3,000	4.80	0.28	25.82	12.24	75.55	5.40
	$T = 120$	100	2.51	-0.70	19.41	28.86	10.05	4.25
		500	2.95	-0.46	13.76	13.63	25.90	5.55
		1,000	3.31	0.05	13.56	10.28	39.70	5.70
		3,000	3.52	0.17	13.70	6.56	61.60	5.30
	$T = 240$	100	1.64	-0.53	13.42	16.73	7.65	6.15
		500	2.02	-0.38	7.86	8.04	13.35	5.65
		1,000	2.32	-0.13	7.24	6.05	22.35	5.45
		3,000	2.65	0.03	6.93	3.82	44.45	4.35
$\phi_S = 0.16, \alpha_S = 0.65$	$T = 60$	100	-20.59	25.62	37.71	892.51	25.80	2.50
		500	-24.12	1.65	37.45	40.44	57.75	3.90
		1,000	-25.89	1.16	38.88	34.07	68.95	5.40
		3,000	-28.50	0.41	42.22	26.13	82.65	6.70
	$T = 120$	100	-13.31	2.54	25.72	37.94	16.80	4.95
		500	-17.21	0.73	24.19	19.85	47.55	5.25
		1,000	-19.04	0.61	25.61	16.70	63.75	5.65
		3,000	-22.27	0.51	29.04	12.99	80.65	5.95
	$T = 240$	100	-8.63	0.19	17.37	19.69	10.70	4.55
		500	-11.31	0.21	15.22	11.72	36.85	5.15
		1,000	-12.87	0.33	15.91	9.31	56.35	4.75
		3,000	-16.07	0.17	18.87	7.04	81.75	5.60

Notes: The DGP for Experiment 3 allows for Gaussian errors, with GARCH effects, without pricing errors, no missing factors, and without spatial/block error cross dependence. For further details of the experiments, see Table S-1.

Figure S-A-E3: Empirical Power Functions, experiment 3, for coefficient of the semi-strong factors



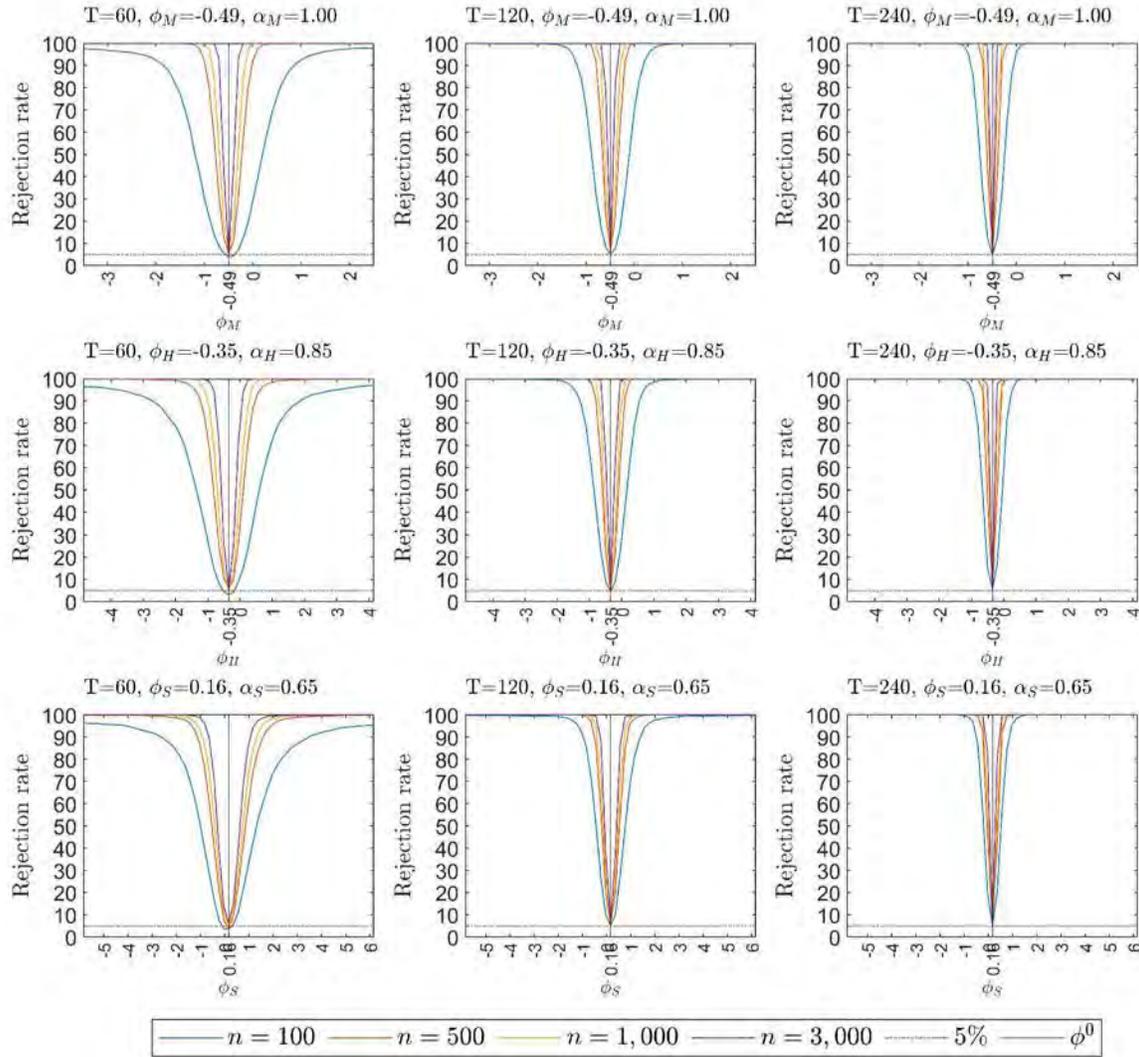
Note: See the notes to Table S-A-E3.

Table S-A-E4: Bias, RMSE and size for the two-step and bias-corrected (BC) estimators of ϕ , for Experiment 4 with one strong and two semi-strong factors

	n	Bias(x100)		RMSE(x100)		Size(x100)		
		Two Step	BC	Two Step	BC	Two Step	BC	
$\phi_M = -0.49, \alpha_M = 1$	$T = 60$	100	0.59	0.17	26.59	107.83	12.15	3.75
		500	-0.16	0.23	17.26	14.75	28.05	5.65
		1,000	-0.57	0.05	15.76	10.43	39.95	6.10
		3,000	-0.93	-0.04	14.85	5.82	60.70	5.95
	$T = 120$	100	0.37	-0.86	17.48	44.37	7.05	4.95
		500	0.20	0.21	9.49	9.24	15.15	6.45
		1,000	-0.08	0.03	7.50	6.16	19.45	4.90
		3,000	-0.40	-0.08	6.35	3.53	36.70	5.85
	$T = 240$	100	0.27	-0.13	12.20	13.22	6.35	4.85
		500	0.34	0.12	5.86	5.83	8.80	5.10
		1,000	0.22	0.05	4.36	4.08	10.60	5.60
		3,000	0.03	-0.02	3.08	2.36	19.15	5.25
$\phi_H = -0.35, \alpha_H = 0.85$	$T = 60$	100	2.92	-5.32	29.30	146.82	17.95	3.15
		500	4.18	-0.76	25.12	27.25	44.65	5.80
		1,000	4.61	-0.06	25.02	19.04	56.50	4.85
		3,000	4.73	0.12	25.86	12.64	74.85	5.60
	$T = 120$	100	2.50	-3.13	19.53	113.15	9.75	4.55
		500	3.16	-0.11	13.84	13.64	25.65	5.25
		1,000	3.27	-0.02	13.64	10.41	40.10	6.05
		3,000	3.47	0.10	13.70	6.67	60.75	6.25
	$T = 240$	100	1.75	-0.41	13.39	16.66	7.85	5.75
		500	2.12	-0.25	7.97	8.07	13.90	5.05
		1,000	2.25	-0.22	7.17	6.09	21.70	5.40
		3,000	2.65	0.02	6.92	3.85	43.35	4.70
$\phi_S = 0.16, \alpha_S = 0.65$	$T = 60$	100	-20.62	12.61	38.03	267.24	25.90	3.60
		500	-24.23	1.81	37.80	44.22	56.50	3.85
		1,000	-25.79	1.07	38.78	34.41	68.90	5.45
		3,000	-28.50	0.96	42.21	27.28	81.45	6.20
	$T = 120$	100	-13.32	7.33	25.66	214.62	16.60	5.25
		500	-17.29	0.59	24.43	20.72	46.95	5.50
		1,000	-19.05	0.56	25.63	16.96	62.30	5.80
		3,000	-22.32	0.41	29.03	13.12	80.50	5.85
	$T = 240$	100	-8.49	0.40	17.46	19.99	10.75	5.10
		500	-11.22	0.35	15.14	11.73	36.10	5.20
		1,000	-12.80	0.44	15.77	9.19	55.95	4.40
		3,000	-16.02	0.29	18.80	6.99	81.40	5.00

Notes: The DGP for Experiment 4 allows for t(5) distributed errors, with GARCH effects, without pricing errors, no missing factors, and without spatial/block error cross dependence. For further details of the experiments, see Table S-1.

Figure S-A-E4: Empirical Power Functions, experiment 4, for coefficient of the semi-strong factors



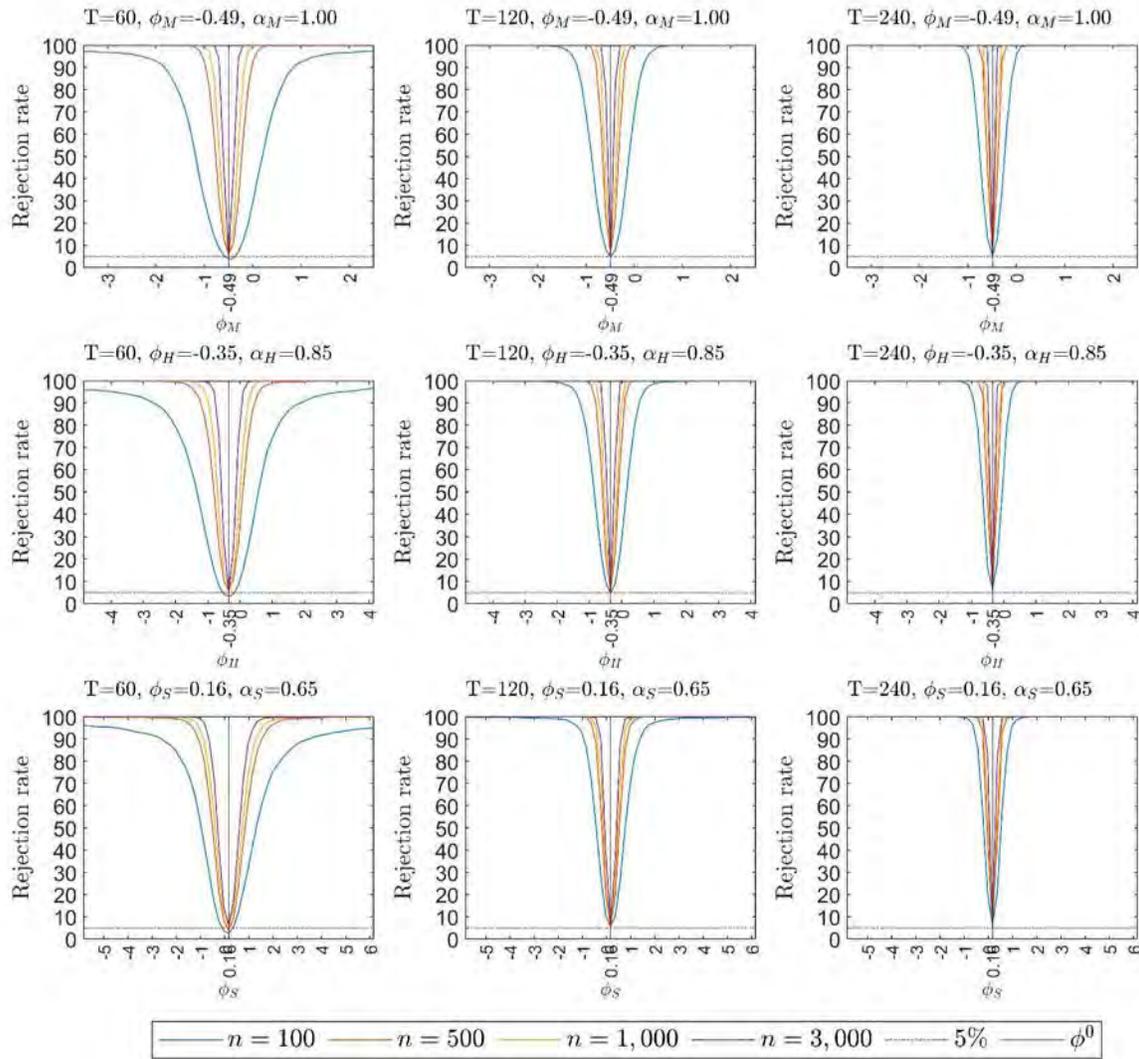
Note: See the notes to Table S-A-E4.

Table S-A-E5: Bias, RMSE and size for the two-step and bias-corrected (BC) estimators of ϕ , for Experiment 5 with one strong and two semi-strong factors

	n	Bias(x100)		RMSE(x100)		Size(x100)		
		Two Step	BC	Two Step	BC	Two Step	BC	
$\phi_M = -0.49, \alpha_M = 1$	$T = 60$	100	0.42	-7.58	26.27	285.21	11.35	3.75
		500	-0.15	0.28	17.05	14.79	28.10	5.80
		1,000	-0.49	0.10	15.81	10.35	40.35	6.25
		3,000	-0.92	-0.00	14.83	5.65	61.10	5.75
	$T = 120$	100	0.06	-0.35	17.63	20.82	7.85	4.95
		500	0.13	0.14	9.40	9.15	14.85	6.45
		1,000	-0.07	0.04	7.60	6.28	20.00	5.90
		3,000	-0.37	-0.06	6.32	3.50	37.00	4.80
	$T = 240$	100	0.08	-0.31	12.48	13.53	6.30	5.35
		500	0.31	0.10	5.83	5.84	8.35	5.35
		1,000	0.23	0.07	4.38	4.11	10.95	5.20
		3,000	0.06	0.00	3.08	2.36	18.75	5.25
$\phi_H = -0.35, \alpha_H = 0.85$	$T = 60$	100	2.58	-8.49	28.97	174.83	18.20	3.30
		500	4.10	-1.01	24.82	26.10	45.25	6.15
		1,000	4.58	-0.12	24.96	18.91	57.15	6.50
		3,000	4.80	0.28	25.82	12.25	75.60	5.60
	$T = 120$	100	2.44	-0.79	19.68	28.94	10.35	4.70
		500	2.98	-0.43	13.78	13.79	25.75	5.60
		1,000	3.32	0.07	13.57	10.29	40.00	6.05
		3,000	3.52	0.17	13.70	6.56	61.85	5.45
	$T = 240$	100	1.57	-0.63	13.97	17.37	8.25	6.75
		500	2.04	-0.35	7.92	8.17	13.85	5.75
		1,000	2.33	-0.12	7.28	6.09	22.70	5.50
		3,000	2.66	0.04	6.94	3.83	44.70	4.55
$\phi_S = 0.16, \alpha_S = 0.65$	$T = 60$	100	-20.57	31.92	37.88	1112.05	25.45	2.55
		500	-24.15	1.61	37.49	40.53	57.90	3.95
		1,000	-25.88	1.16	38.88	34.12	68.80	5.70
		3,000	-28.51	0.40	42.23	26.13	82.65	6.60
	$T = 120$	100	-13.21	2.55	26.09	37.89	16.70	5.50
		500	-17.24	0.70	24.27	20.01	47.90	5.40
		1,000	-19.03	0.62	25.61	16.77	63.50	5.85
		3,000	-22.26	0.53	29.04	13.01	80.55	6.05
	$T = 240$	100	-8.55	0.29	18.03	20.67	11.85	6.25
		500	-11.35	0.16	15.31	11.89	37.35	5.85
		1,000	-12.87	0.33	15.91	9.37	56.40	4.75
		3,000	-16.07	0.18	18.86	7.07	81.75	5.65

Notes: The DGP for Experiment 5 allows for Gaussian errors, with GARCH effects, with pricing errors ($\alpha_\eta = 0.3$), no missing factors, and without spatial/block error cross dependence. For further details of the experiments, see Table S-1.

Figure S-A-E5: Empirical Power Functions, experiment 5, for coefficient of the semi-strong factors



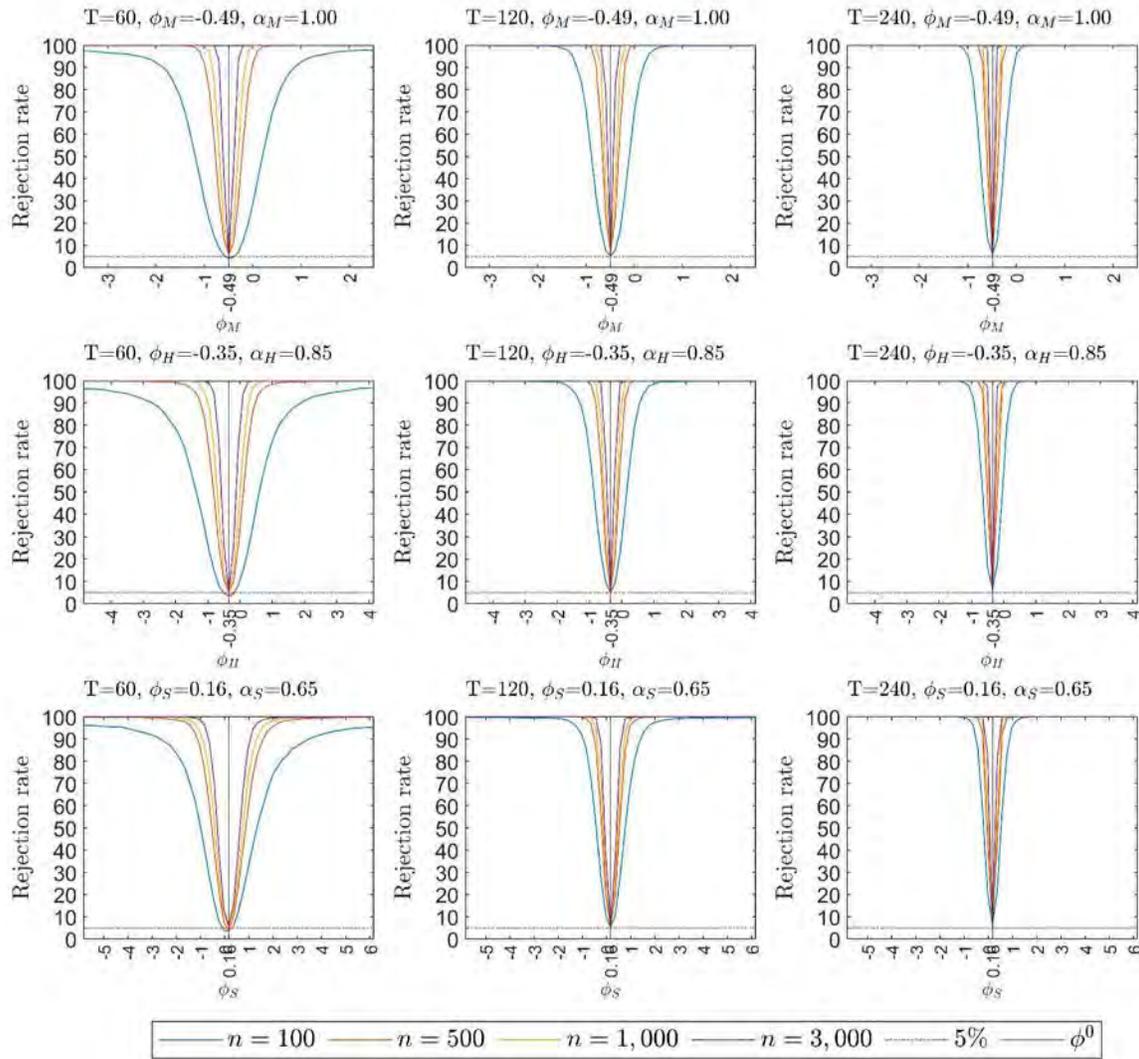
Note: See the notes to Table S-A-E5.

Table S-A-E6: Bias, RMSE and size for the two-step and bias-corrected (BC) estimators of ϕ , for Experiment 6 with one strong and two semi-strong factors

	n	Bias(x100)		RMSE(x100)		Size(x100)		
		Two Step	BC	Two Step	BC	Two Step	BC	
$\phi_M = -0.49, \alpha_M = 1$	$T = 60$	100	0.49	0.05	26.69	103.65	12.25	3.95
		500	-0.17	0.24	17.27	14.81	28.25	5.95
		1,000	-0.56	0.06	15.75	10.44	39.60	6.00
		3,000	-0.93	-0.05	14.85	5.82	60.75	5.80
	$T = 120$	100	0.22	-0.76	17.65	33.47	7.45	5.15
		500	0.20	0.22	9.54	9.32	15.30	6.95
		1,000	-0.06	0.05	7.49	6.18	19.50	5.30
		3,000	-0.40	-0.09	6.36	3.54	37.05	5.70
	$T = 240$	100	0.11	-0.30	12.57	13.63	7.55	6.10
		500	0.33	0.11	5.91	5.88	8.60	5.10
		1,000	0.24	0.07	4.38	4.13	10.30	5.80
		3,000	0.03	-0.03	3.09	2.37	19.25	5.45
$\phi_H = -0.35, \alpha_H = 0.85$	$T = 60$	100	2.83	-5.32	29.34	144.70	18.05	3.40
		500	4.22	-0.67	25.10	27.28	44.70	5.45
		1,000	4.62	-0.03	25.04	19.02	56.30	4.80
		3,000	4.73	0.12	25.86	12.64	74.70	5.65
	$T = 120$	100	2.42	-2.45	19.82	78.69	11.40	4.90
		500	3.19	-0.06	13.87	13.81	25.85	5.45
		1,000	3.28	-0.00	13.65	10.44	40.80	6.35
		3,000	3.47	0.11	13.70	6.67	61.05	6.10
	$T = 240$	100	1.68	-0.51	13.94	17.33	8.35	6.70
		500	2.15	-0.22	8.05	8.22	14.70	5.60
		1,000	2.26	-0.20	7.23	6.14	21.75	5.65
		3,000	2.65	0.03	6.93	3.86	43.85	5.00
$\phi_S = 0.16, \alpha_S = 0.65$	$T = 60$	100	-20.59	12.67	38.25	257.83	25.55	3.70
		500	-24.27	1.73	37.83	44.57	56.60	3.95
		1,000	-25.78	1.10	38.79	34.52	68.85	5.30
		3,000	-28.50	0.95	42.22	27.27	81.40	6.05
	$T = 120$	100	-13.22	5.85	26.06	146.46	16.95	5.40
		500	-17.32	0.55	24.50	20.89	47.05	5.55
		1,000	-19.04	0.58	25.64	17.05	62.35	5.95
		3,000	-22.32	0.42	29.03	13.16	80.25	5.90
	$T = 240$	100	-8.41	0.50	18.09	20.93	12.15	6.40
		500	-11.25	0.31	15.25	11.95	36.35	5.70
		1,000	-12.80	0.44	15.77	9.26	55.35	5.05
		3,000	-16.02	0.29	18.80	7.02	81.00	5.10

Notes: The DGP for Experiment 6 allows for t(5) distributed errors, with GARCH effects, with pricing errors ($\alpha_\eta = 0.3$), no missing factors, and without spatial/block error cross dependence. For further details of the experiments, see Table S-1.

Figure S-A-E6: Empirical Power Functions, experiment 6, for coefficient of the semi-strong factors



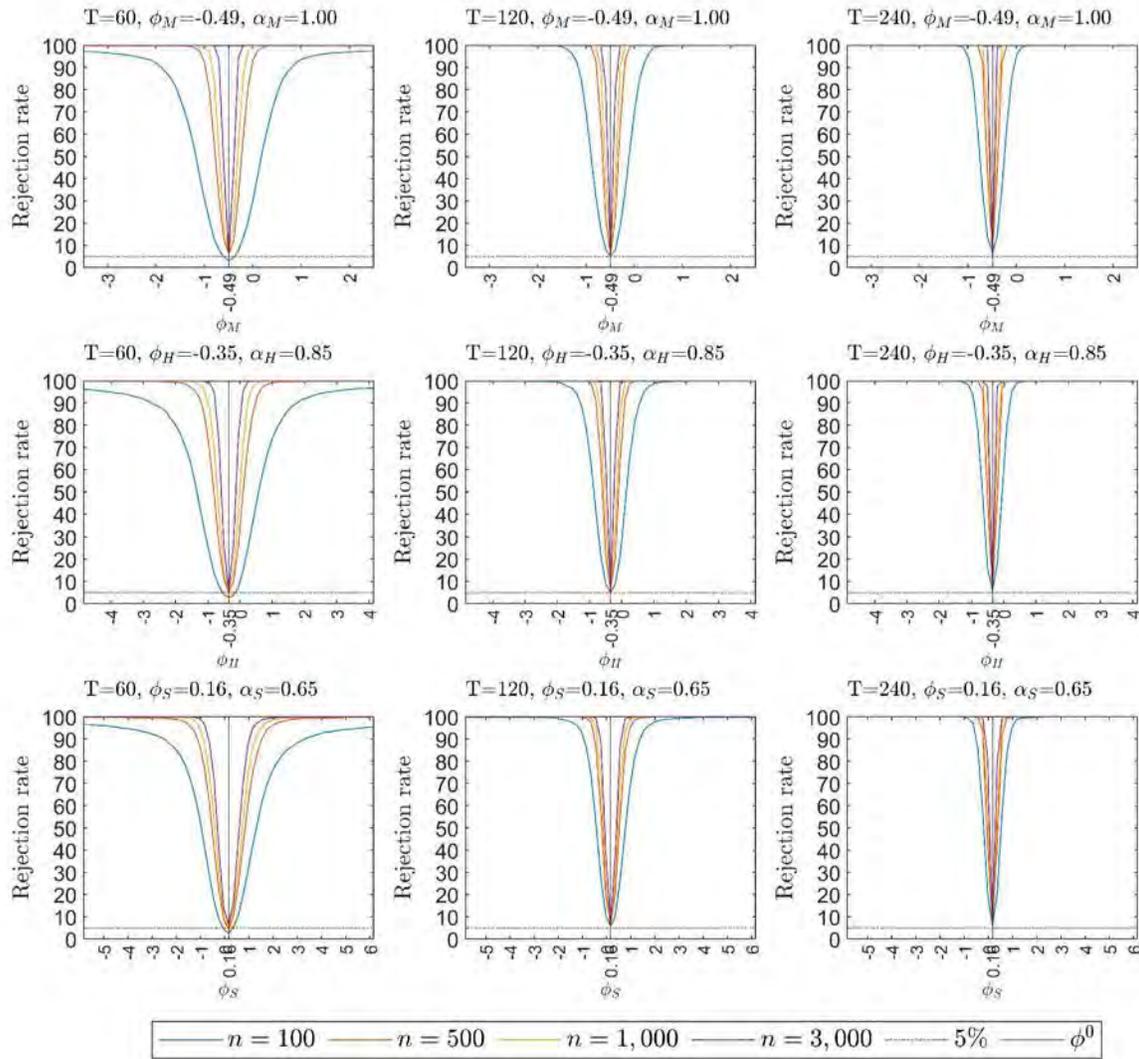
Note: See the notes to Table S-A-E6.

Table S-A-E7: Bias, RMSE and size for the two-step and bias-corrected (BC) estimators of ϕ , for Experiment 7 with one strong and two semi-strong factors

	n	Bias(x100)		RMSE(x100)		Size(x100)		
		Two Step	BC	Two Step	BC	Two Step	BC	
$\phi_M = -0.49, \alpha_M = 1$	$T = 60$	100	-0.33	-14.31	26.54	555.11	11.80	3.20
		500	-0.50	-0.18	17.67	15.13	29.80	6.40
		1,000	-0.73	-0.15	15.95	10.41	42.20	6.25
		3,000	-0.91	0.04	14.87	5.81	61.20	5.90
	$T = 120$	100	-0.13	-0.49	17.44	20.30	7.55	5.05
		500	0.01	0.01	9.49	8.86	14.20	5.85
		1,000	-0.11	0.04	7.74	6.31	20.50	6.60
		3,000	-0.27	0.08	6.40	3.60	36.65	5.65
	$T = 240$	100	-0.00	-0.39	12.82	14.02	7.30	6.75
		500	0.15	-0.08	6.14	6.13	9.85	6.95
		1,000	0.14	-0.03	4.52	4.22	11.90	5.80
		3,000	0.07	0.01	3.05	2.32	18.30	5.15
$\phi_H = -0.35, \alpha_H = 0.85$	$T = 60$	100	3.05	-222.60	29.40	9435.95	17.55	2.70
		500	4.42	-0.75	25.03	26.57	46.25	4.95
		1,000	4.43	-0.56	25.18	19.01	58.75	5.45
		3,000	4.63	-0.15	25.85	12.14	76.15	5.75
	$T = 120$	100	2.45	-0.89	19.84	28.94	10.55	4.95
		500	3.24	-0.00	13.84	13.44	26.75	5.95
		1,000	3.30	0.05	13.47	9.91	41.30	5.20
		3,000	3.29	-0.25	13.67	6.56	62.45	5.75
	$T = 240$	100	1.67	-0.55	13.68	16.80	7.45	6.10
		500	2.28	-0.04	8.00	8.00	13.35	5.50
		1,000	2.42	-0.02	7.22	6.00	22.60	5.10
		3,000	2.54	-0.13	6.85	3.82	44.00	5.25
$\phi_S = 0.16, \alpha_S = 0.65$	$T = 60$	100	-20.23	62.14	37.82	2347.80	25.15	2.70
		500	-24.55	1.18	37.58	46.58	59.25	4.40
		1,000	-26.04	0.76	38.91	33.25	69.25	5.25
		3,000	-28.68	0.06	42.34	26.75	82.30	6.90
	$T = 120$	100	-12.90	2.52	25.93	38.03	16.35	5.70
		500	-17.48	0.44	24.54	21.10	50.00	6.00
		1,000	-19.25	0.09	25.81	16.81	63.35	5.45
		3,000	-22.34	0.29	29.12	13.12	80.15	6.15
	$T = 240$	100	-8.13	0.84	17.77	20.91	11.30	5.65
		500	-11.31	0.30	15.26	11.79	37.05	5.70
		1,000	-12.84	0.38	15.96	9.45	56.40	5.15
		3,000	-16.09	0.21	18.91	7.26	80.95	6.30

Notes: The DGP for Experiment 7 allows for Gaussian errors, with GARCH effects, with pricing errors ($\alpha_\eta = 0.3$), with one weak missing factor ($\alpha_\gamma = 0.5$), and without spatial/block error cross dependence. For further details of the experiments, see S-1.

Figure S-A-E7: Empirical Power Functions, experiment 7 for coefficient of the semi-strong factors



Note: See the notes to Table S-A-E7.

Table S-A-E8: Bias, RMSE and size for the two-step and bias-corrected (BC) estimators of ϕ , for Experiment 8 with one strong and two semi-strong factors

	n	Bias(x100)		RMSE(x100)		Size(x100)		
		Two Step	BC	Two Step	BC	Two Step	BC	
$\phi_M = -0.49, \alpha_M = 1$	$T = 60$	100	-0.27	9.06	27.12	251.34	11.45	3.75
		500	-0.74	-0.67	17.81	16.71	29.75	6.30
		1,000	-0.64	0.01	16.06	10.38	40.90	5.70
		3,000	-0.85	0.12	14.90	5.96	59.65	6.15
	$T = 120$	100	0.04	-0.32	17.49	20.53	7.35	4.75
		500	-0.06	-0.04	9.49	8.89	13.95	5.80
		1,000	0.01	0.19	7.78	6.40	20.90	6.10
		3,000	-0.24	0.14	6.39	3.63	37.25	5.90
	$T = 240$	100	0.09	-0.29	12.84	14.07	7.10	6.35
		500	0.20	-0.01	6.08	6.06	9.50	6.05
		1,000	0.17	0.01	4.52	4.25	11.55	6.30
		3,000	0.06	0.01	3.07	2.35	18.15	5.15
$\phi_H = -0.35, \alpha_H = 0.85$	$T = 60$	100	3.41	39.71	29.84	1913.29	17.65	3.15
		500	4.52	-1.03	25.21	30.79	44.70	4.95
		1,000	4.47	-0.47	25.14	19.34	57.65	5.85
		3,000	4.59	-0.21	25.88	12.19	75.20	5.45
	$T = 120$	100	2.68	-0.54	20.19	28.91	11.20	5.05
		500	3.38	0.31	13.97	13.39	26.60	5.35
		1,000	3.36	0.14	13.38	9.90	40.55	4.90
		3,000	3.33	-0.18	13.64	6.52	61.95	5.20
	$T = 240$	100	1.77	-0.40	13.95	17.05	7.45	6.00
		500	2.38	0.12	8.01	7.92	14.30	4.70
		1,000	2.58	0.20	7.25	5.99	23.15	4.75
		3,000	2.54	-0.13	6.81	3.84	42.85	4.95
$\phi_S = 0.16, \alpha_S = 0.65$	$T = 60$	100	-20.01	-141.72	37.89	5932.49	25.40	3.15
		500	-24.52	3.22	37.63	77.87	57.05	4.45
		1,000	-26.14	0.57	39.14	34.26	69.30	5.10
		3,000	-28.54	0.84	42.33	26.81	81.00	5.85
	$T = 120$	100	-13.01	2.45	26.01	36.96	16.80	5.60
		500	-17.52	0.34	24.66	20.94	48.35	6.55
		1,000	-19.32	-0.11	25.88	16.81	64.05	5.80
		3,000	-22.35	0.26	29.14	13.22	79.70	6.20
	$T = 240$	100	-8.14	0.79	17.57	20.56	11.15	5.85
		500	-11.42	0.12	15.36	11.78	37.65	5.55
		1,000	-13.01	0.12	16.02	9.32	55.70	5.20
		3,000	-16.10	0.20	18.91	7.40	80.10	6.70

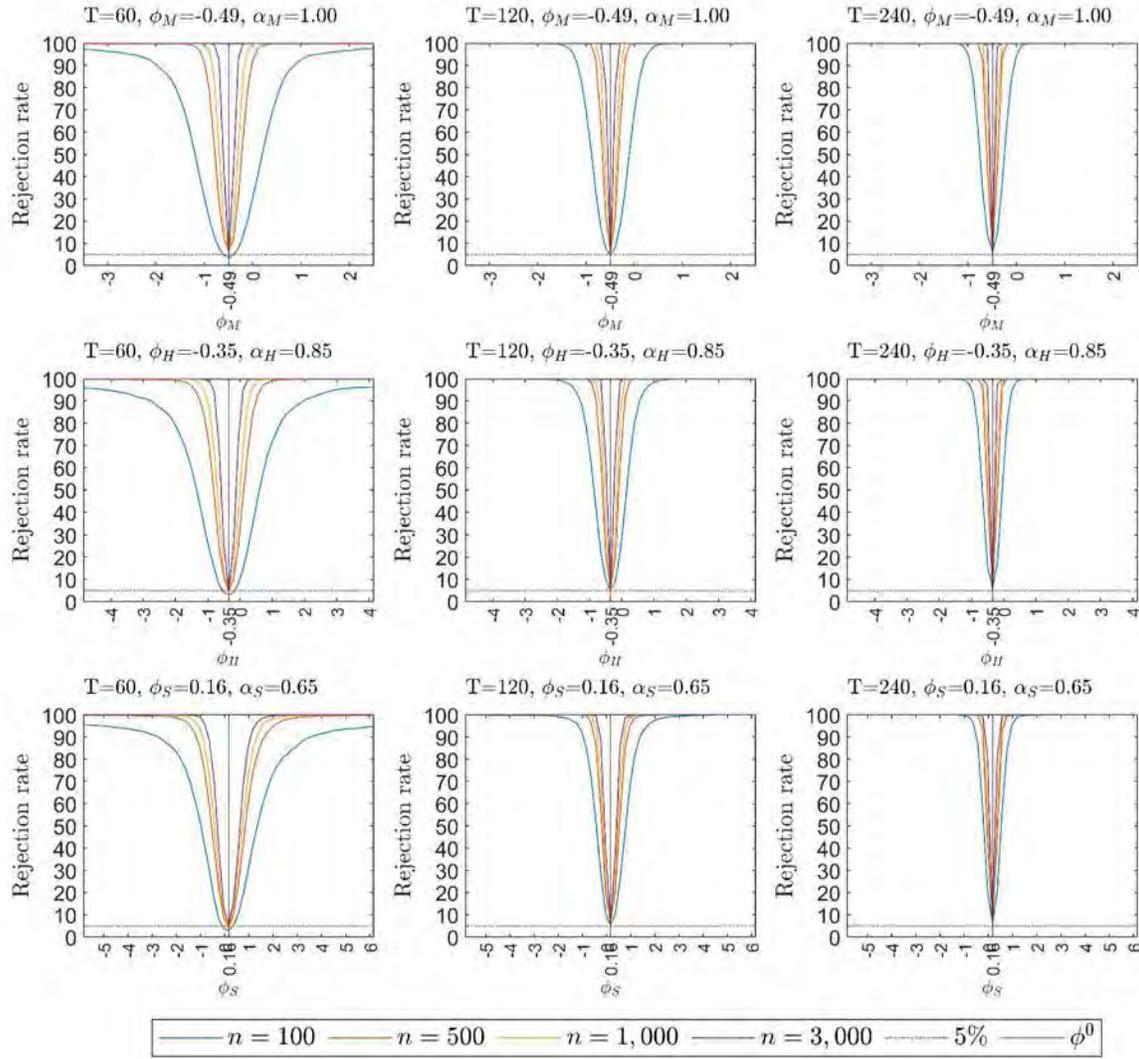
Notes: The DGP for Experiment 8 allows for t(5) distributed errors, with GARCH effects, with pricing errors ($\alpha_\eta = 0.3$), with one weak missing factor ($\alpha_\gamma = 0.5$), and without spatial/block error cross dependence. For further details of the experiments, see S-1.

Table S-A-E8a: Bias, RMSE and size for the two-step and bias-corrected (BC) estimators of ϕ for Experiment 8 with one strong and two semi-strong factors

	n	Bias(x100)		RMSE(x100)		Size(x100)		
		Two Step	BC	Two Step	BC	Two Step	BC	
$\phi_M = -0.49, \alpha_M = 1$	$T = 60$	100	-0.16	12.00	27.46	356.79	12.85	4.05
		500	-0.77	-0.71	17.89	16.74	29.55	6.50
		1,000	-0.66	-0.02	16.09	10.45	40.95	5.70
		3,000	-0.86	0.10	14.91	5.99	59.85	6.20
	$T = 120$	100	0.19	-0.16	18.02	21.26	8.25	5.30
		500	-0.09	-0.08	9.66	9.09	14.80	6.00
		1,000	-0.01	0.16	7.84	6.51	20.95	6.70
		3,000	-0.25	0.12	6.40	3.65	38.10	6.35
	$T = 240$	100	0.27	-0.10	13.79	15.14	8.85	8.70
		500	0.19	-0.03	6.34	6.36	11.05	7.05
		1,000	0.14	-0.03	4.63	4.40	12.05	7.50
		3,000	0.05	-0.00	3.09	2.39	18.85	5.70
$\phi_H = -0.35, \alpha_H = 0.85$	$T = 60$	100	3.10	55.07	29.92	2828.62	18.85	3.10
		500	4.53	-0.98	25.32	30.87	45.10	5.45
		1,000	4.46	-0.51	25.19	19.42	58.20	5.45
		3,000	4.59	-0.21	25.88	12.26	75.20	5.75
	$T = 120$	100	2.39	-0.98	20.59	29.45	11.90	5.85
		500	3.39	0.32	14.17	13.75	27.35	5.45
		1,000	3.34	0.10	13.45	10.09	40.55	5.00
		3,000	3.33	-0.18	13.63	6.58	62.30	5.20
	$T = 240$	100	1.52	-0.70	14.82	18.29	10.60	8.55
		500	2.38	0.12	8.27	8.27	15.65	6.05
		1,000	2.55	0.15	7.35	6.23	23.20	5.65
		3,000	2.54	-0.13	6.83	3.90	42.75	5.95
$\phi_S = 0.16, \alpha_S = 0.65$	$T = 60$	100	-19.86	-210.04	38.24	8784.76	26.15	3.40
		500	-24.49	3.17	37.68	75.99	56.80	4.95
		1,000	-26.11	0.71	39.11	34.49	69.05	4.95
		3,000	-28.55	0.78	42.34	26.88	81.20	6.05
	$T = 120$	100	-12.88	2.68	26.48	38.62	17.60	5.85
		500	-17.41	0.56	24.62	21.19	47.90	6.25
		1,000	-19.28	-0.02	25.86	17.01	63.85	6.00
		3,000	-22.36	0.23	29.18	13.36	79.65	6.75
	$T = 240$	100	-8.03	0.96	18.42	22.23	13.25	7.35
		500	-11.34	0.24	15.47	12.20	36.80	6.50
		1,000	-12.95	0.21	16.03	9.63	56.70	6.20
		3,000	-16.10	0.19	18.96	7.59	80.25	7.25

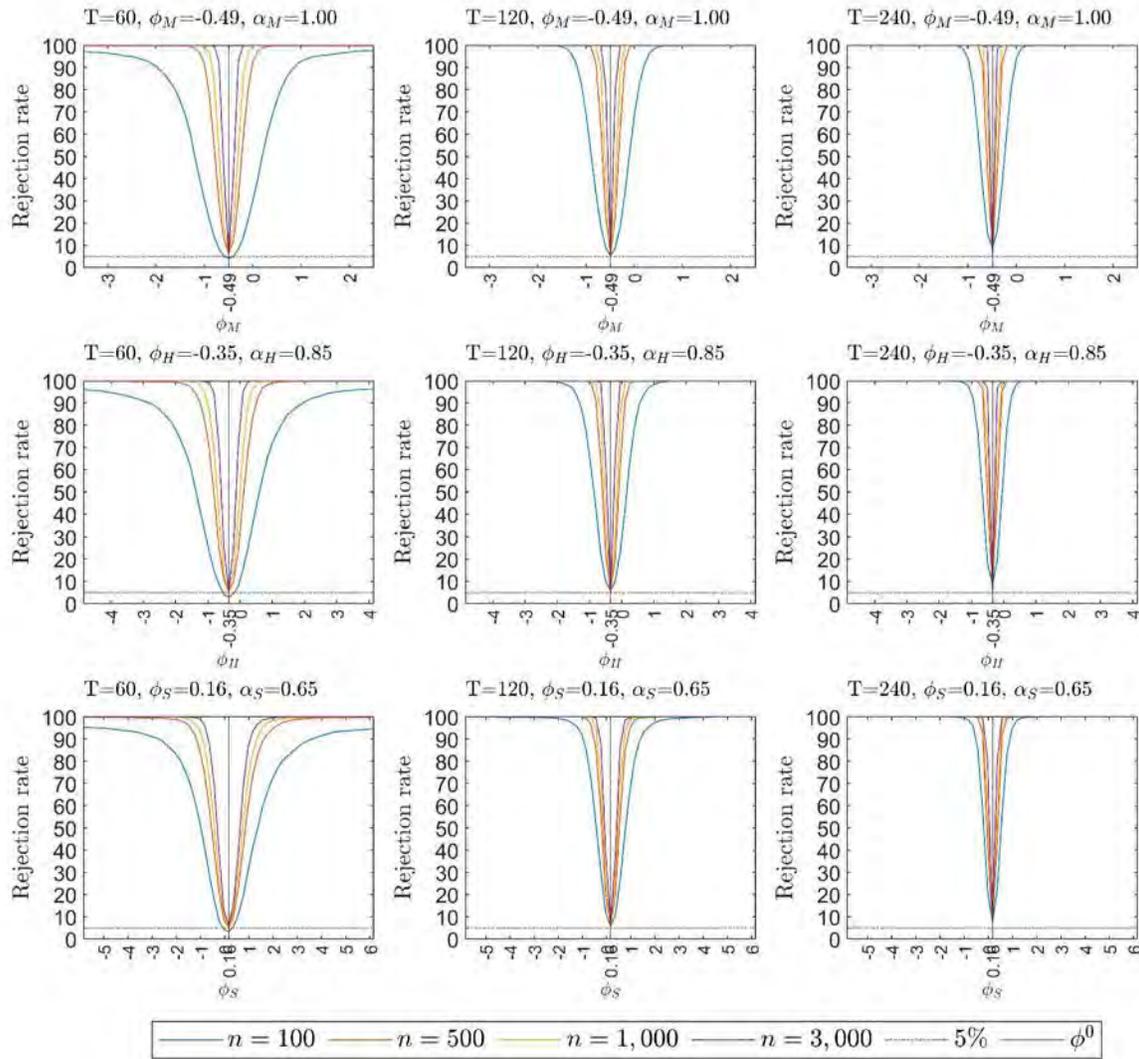
Notes: The DGP for Experiment 8 allows for t(5) distributed errors, with GARCH effects, with pricing errors ($\alpha_\eta = 0.5$), with one weak missing factor ($\alpha_\gamma = 0.5$), and without spatial/block error cross dependence. For further details of the experiments, see S-1.

Figure S-A-E8: Empirical Power Functions, experiment 8 for coefficient of the semi-strong factors



Note: See the notes to Table S-A-E8.

Figure S-A-E8a: Empirical Power Functions, experiment 8a for coefficient of the semi-strong factors



Note: See the notes to Table S-A-E8a.

Table S-A-E9: Bias, RMSE and size for the two-step and bias-corrected (BC) estimators of ϕ , for Experiment 9 with one strong and two semi-strong factors

	n	Bias(x100)		RMSE(x100)		Size(x100)		
		Two Step	BC	Two Step	BC	Two Step	BC	
$\phi_M = -0.49, \alpha_M = 1$	T = 60	100	-0.51	-15.47	27.16	662.68	11.80	3.40
		500	-0.55	1.29	18.02	84.39	31.25	7.75
		1,000	-0.85	-0.47	16.24	11.31	43.00	8.50
		3,000	-0.99	-0.16	14.97	6.25	61.30	7.75
	T = 120	100	-0.17	-0.59	17.92	21.49	7.15	4.50
		500	0.03	-0.01	9.70	9.20	13.95	6.15
		1,000	-0.17	-0.07	7.88	6.53	20.35	6.30
		3,000	-0.33	-0.01	6.45	3.71	36.90	6.50
	T = 240	100	-0.12	-0.54	12.73	14.02	6.80	6.85
		500	0.13	-0.12	6.18	6.19	9.65	6.55
		1,000	0.11	-0.08	4.57	4.30	11.75	6.60
		3,000	0.06	-0.00	3.09	2.37	17.90	5.45
$\phi_H = -0.35, \alpha_H = 0.85$	T = 60	100	2.80	3.36	30.96	683.86	19.20	3.90
		500	4.18	2.34	25.47	189.26	47.10	8.65
		1,000	4.41	-0.83	25.43	22.18	59.30	9.35
		3,000	4.63	-0.28	25.90	14.12	76.45	10.60
	T = 120	100	2.27	-1.02	20.81	33.62	10.25	4.50
		500	3.04	-0.39	14.24	14.82	24.60	5.70
		1,000	3.17	-0.24	13.77	11.21	38.65	6.10
		3,000	3.24	-0.40	13.74	7.47	60.05	7.75
	T = 240	100	1.62	-0.61	14.29	17.94	7.20	5.75
		500	2.25	-0.09	8.26	8.60	13.10	5.35
		1,000	2.36	-0.10	7.47	6.57	21.10	5.65
		3,000	2.57	-0.10	6.99	4.27	41.70	5.85
$\phi_S = 0.16, \alpha_S = 0.65$	T = 60	100	-19.57	38.19	40.89	1634.08	26.25	4.75
		500	-24.60	-17.77	38.37	944.75	59.35	10.05
		1,000	-26.04	1.99	39.30	41.96	70.30	11.50
		3,000	-28.69	0.77	42.55	32.04	82.85	11.70
	T = 120	100	-12.73	4.51	29.32	56.35	14.45	5.50
		500	-17.45	0.82	25.67	26.62	44.50	8.70
		1,000	-19.16	0.54	26.31	20.96	58.65	8.40
		3,000	-22.27	0.58	29.33	16.06	77.55	8.95
	T = 240	100	-8.13	1.07	21.26	26.72	10.60	6.50
		500	-11.34	0.35	16.58	15.28	31.25	7.40
		1,000	-12.81	0.50	16.52	11.98	46.35	6.40
		3,000	-16.03	0.35	19.08	9.09	75.35	7.25

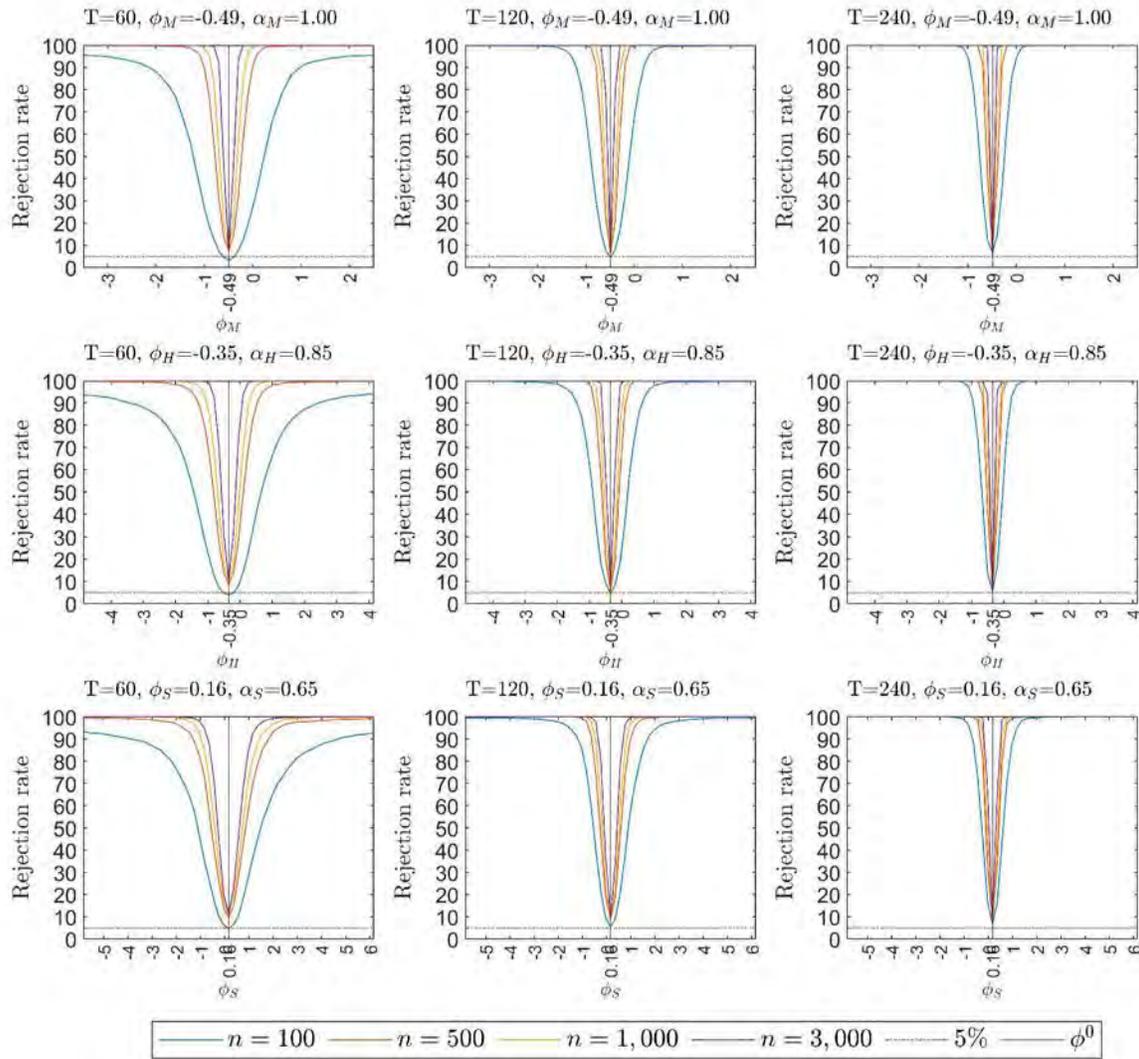
Notes: The DGP for Experiment 9 allows for Gaussian errors, with GARCH effects, with pricing errors ($\alpha_\eta = 0.3$), with one weak missing factor ($\alpha_\gamma = 0.5$), and with spatial error cross dependence ($\rho_\varepsilon = 0.5$). For further details of the experiments, see Table S-1.

Table S-A-E9a: Bias, RMSE and size for the two-step and bias-corrected (BC) estimators of ϕ for Experiment 9 with one strong and two semi-strong factors

	n	Bias(x100)		RMSE(x100)		Size(x100)		
		Two Step	BC	Two Step	BC	Two Step	BC	
$\phi_M = -0.49, \alpha_M = 1$	$T = 60$	100	-0.48	24.91	29.91	1802.71	10.25	3.05
		500	-0.58	-1.22	19.37	92.46	26.95	6.75
		1,000	-0.92	-0.95	17.00	17.30	37.20	8.60
		3,000	-1.03	-0.35	15.24	8.41	55.50	8.65
	$T = 120$	100	-0.02	-1.97	19.95	35.08	8.60	6.20
		500	0.03	-0.10	10.58	10.55	13.75	6.10
		1,000	-0.24	-0.22	8.38	7.44	17.85	6.55
		3,000	-0.40	-0.12	6.67	4.26	33.75	6.90
	$T = 240$	100	0.02	-0.44	13.81	15.39	8.65	7.40
		500	0.12	-0.14	6.62	6.70	10.25	6.35
		1,000	0.06	-0.15	4.86	4.72	10.85	7.45
		3,000	0.04	-0.03	3.26	2.62	16.95	6.45
$\phi_H = -0.35, \alpha_H = 0.85$	$T = 60$	100	1.63	29.05	36.11	1916.43	14.35	1.95
		500	3.76	1.48	27.36	303.36	36.10	7.25
		1,000	4.26	-1.84	26.21	39.83	48.45	8.20
		3,000	4.56	-0.62	26.15	22.78	69.25	11.05
	$T = 120$	100	1.79	-3.12	24.84	50.47	9.30	4.50
		500	2.85	-0.87	16.01	19.66	18.60	5.60
		1,000	3.03	-0.57	14.77	15.08	28.80	6.90
		3,000	3.13	-0.68	14.02	10.07	49.45	8.40
	$T = 240$	100	1.21	-1.21	17.34	22.52	8.30	6.75
		500	2.22	-0.15	9.58	10.83	11.00	5.95
		1,000	2.27	-0.24	8.46	8.56	18.20	6.15
		3,000	2.60	-0.08	7.42	5.55	33.25	7.10
$\phi_S = 0.16, \alpha_S = 0.65$	$T = 60$	100	-18.49	110.60	50.00	7440.55	17.75	2.35
		500	-24.61	18.11	41.10	1183.88	44.00	6.75
		1,000	-26.08	6.93	40.81	118.48	58.05	8.75
		3,000	-28.71	2.78	43.19	66.65	75.45	11.30
	$T = 120$	100	-12.19	13.59	37.47	192.49	10.25	4.45
		500	-17.28	2.13	28.78	40.29	30.50	8.05
		1,000	-19.06	1.36	27.98	31.28	44.05	8.90
		3,000	-22.25	1.02	30.06	23.69	66.95	8.90
	$T = 240$	100	-8.15	1.74	28.35	38.76	8.85	5.65
		500	-11.28	0.68	19.74	22.51	19.25	7.65
		1,000	-12.72	0.82	18.22	17.60	31.20	7.15
		3,000	-15.99	0.51	19.76	13.26	62.25	8.00

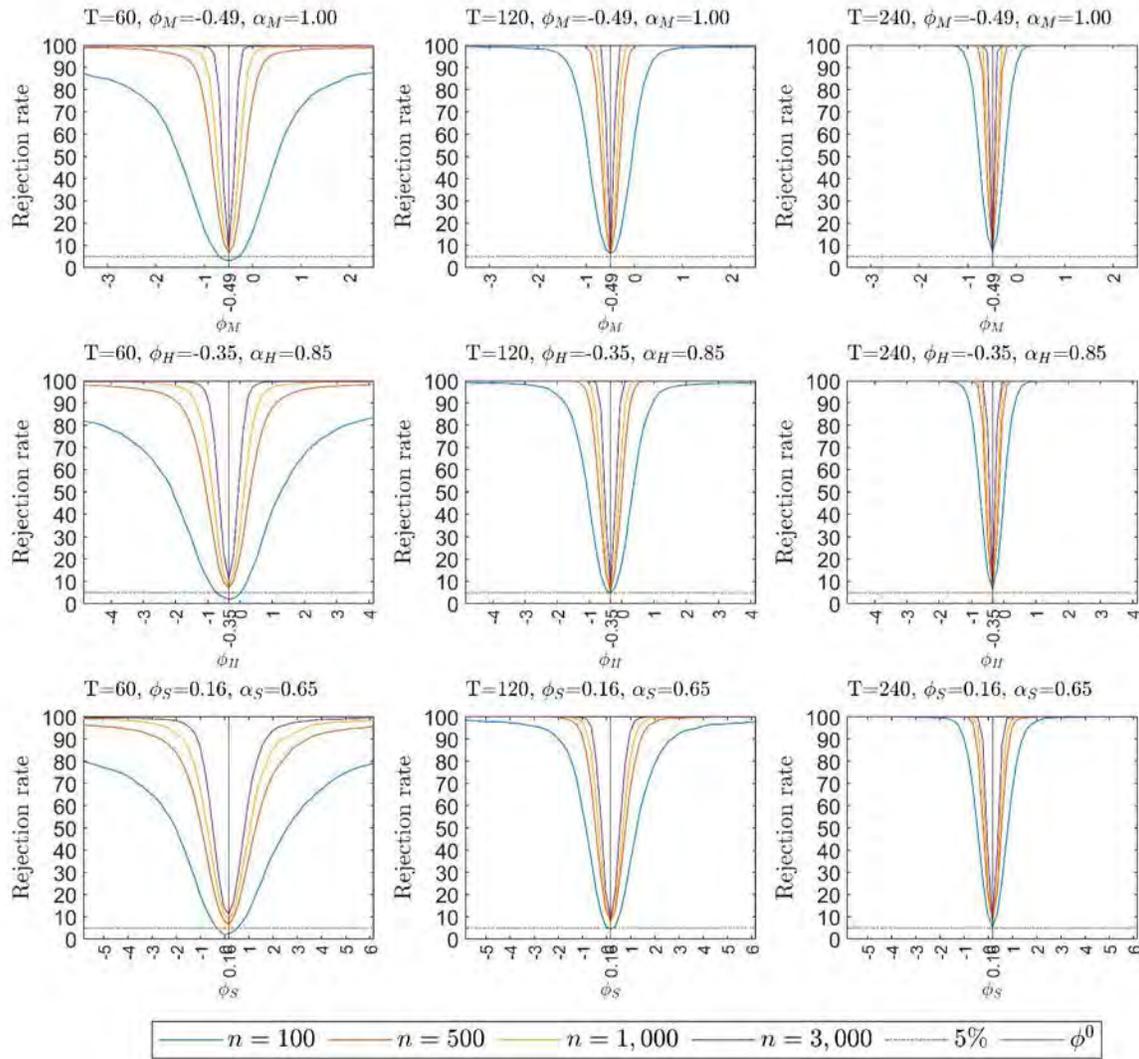
Notes: The DGP for Experiment 9 allows for Gaussian errors, with GARCH effects, with pricing errors ($\alpha_\eta = 0.5$), with one weak missing factor ($\alpha_\gamma = 0.5$), and with spatial error cross dependence ($\rho_\varepsilon = 0.85$). For further details of the experiments, see Table S-1.

Figure S-A-E9: Empirical Power Functions, experiment 9 for coefficient of the semi-strong factors



Note: See the notes to Table S-A-E9.

Figure S-A-E9a: Empirical Power Functions, experiment 9a for coefficient of the semi-strong factors



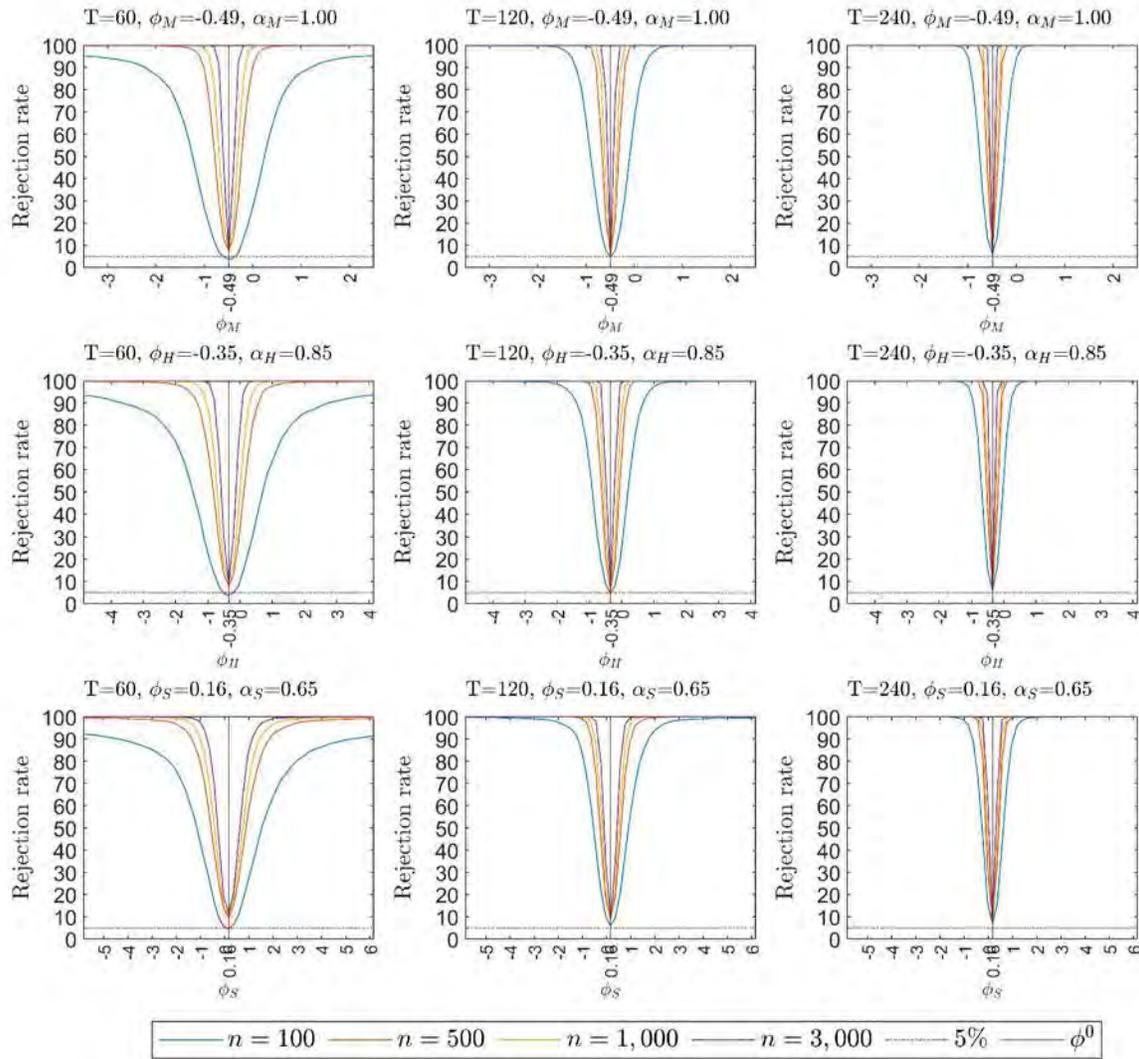
Note: See the notes to Table S-A-E9a.

Table S-A-E10: Bias, RMSE and size for the two-step and bias-corrected (BC) estimators of ϕ , for Experiment 10 with one strong and two semi-strong factors

	n	Bias(x100)		RMSE(x100)		Size(x100)		
		Two Step	BC	Two Step	BC	Two Step	BC	
$\phi_M = -0.49, \alpha_M = 1$	$T = 60$	100	-0.39	-1.61	27.55	151.91	11.70	3.75
		500	-0.82	-0.69	18.11	17.07	30.40	7.70
		1,000	-0.73	-0.31	16.33	11.30	42.50	7.05
		3,000	-0.92	-0.06	14.97	6.36	60.05	8.00
	$T = 120$	100	0.02	-1.86	18.04	69.05	7.65	4.80
		500	-0.07	-0.09	9.78	9.29	15.05	6.10
		1,000	0.01	0.15	7.90	6.62	19.70	6.75
		3,000	-0.29	0.05	6.42	3.72	37.55	5.80
	$T = 240$	100	0.01	-0.39	12.87	14.17	6.30	6.35
		500	0.17	-0.05	6.13	6.15	9.25	6.05
		1,000	0.17	0.00	4.55	4.30	11.30	6.40
		3,000	0.06	0.00	3.12	2.40	17.65	5.75
$\phi_H = -0.35, \alpha_H = 0.85$	$T = 60$	100	3.25	-8.09	31.16	504.30	19.35	3.65
		500	4.29	-1.32	25.64	33.61	46.05	8.10
		1,000	4.50	-0.68	25.36	22.95	58.25	9.70
		3,000	4.57	-0.37	25.89	14.12	76.05	9.50
	$T = 120$	100	2.48	-1.13	21.16	32.42	11.15	4.50
		500	3.15	-0.09	14.31	14.78	25.55	5.70
		1,000	3.25	-0.07	13.66	11.19	38.65	5.55
		3,000	3.31	-0.26	13.70	7.36	60.50	6.70
	$T = 240$	100	1.78	-0.40	14.58	18.18	8.25	6.05
		500	2.35	0.08	8.31	8.60	13.35	5.20
		1,000	2.53	0.12	7.51	6.58	21.05	5.00
		3,000	2.57	-0.10	6.96	4.28	41.35	6.00
$\phi_S = 0.16, \alpha_S = 0.65$	$T = 60$	100	-19.35	25.35	40.86	579.34	25.25	4.60
		500	-24.56	2.56	38.40	68.65	57.50	9.85
		1,000	-26.17	1.78	39.59	43.66	69.65	10.70
		3,000	-28.53	1.55	42.55	32.38	81.95	11.90
	$T = 120$	100	-12.71	10.19	29.68	304.30	14.70	6.05
		500	-17.42	0.86	25.76	26.78	43.95	8.75
		1,000	-19.30	0.15	26.41	20.74	59.65	8.50
		3,000	-22.31	0.47	29.35	16.12	77.35	8.95
	$T = 240$	100	-8.09	1.05	21.11	26.50	10.55	6.65
		500	-11.52	0.08	16.67	15.20	30.35	7.15
		1,000	-12.98	0.22	16.56	11.84	47.50	6.80
		3,000	-16.05	0.32	19.11	9.25	75.50	8.00

Notes: The DGP for Experiment 10 allows for t(5) distributed errors, with GARCH effects, with pricing errors ($\alpha_\eta = 0.3$), with one weak missing factor ($\alpha_\gamma = 0.5$), and with spatial error cross dependence ($\rho_\varepsilon = 0.5$). For further details of the experiments, see Table S-1.

Figure S-A-E10: Empirical Power Functions, experiment 10 for coefficient of the semi-strong factors



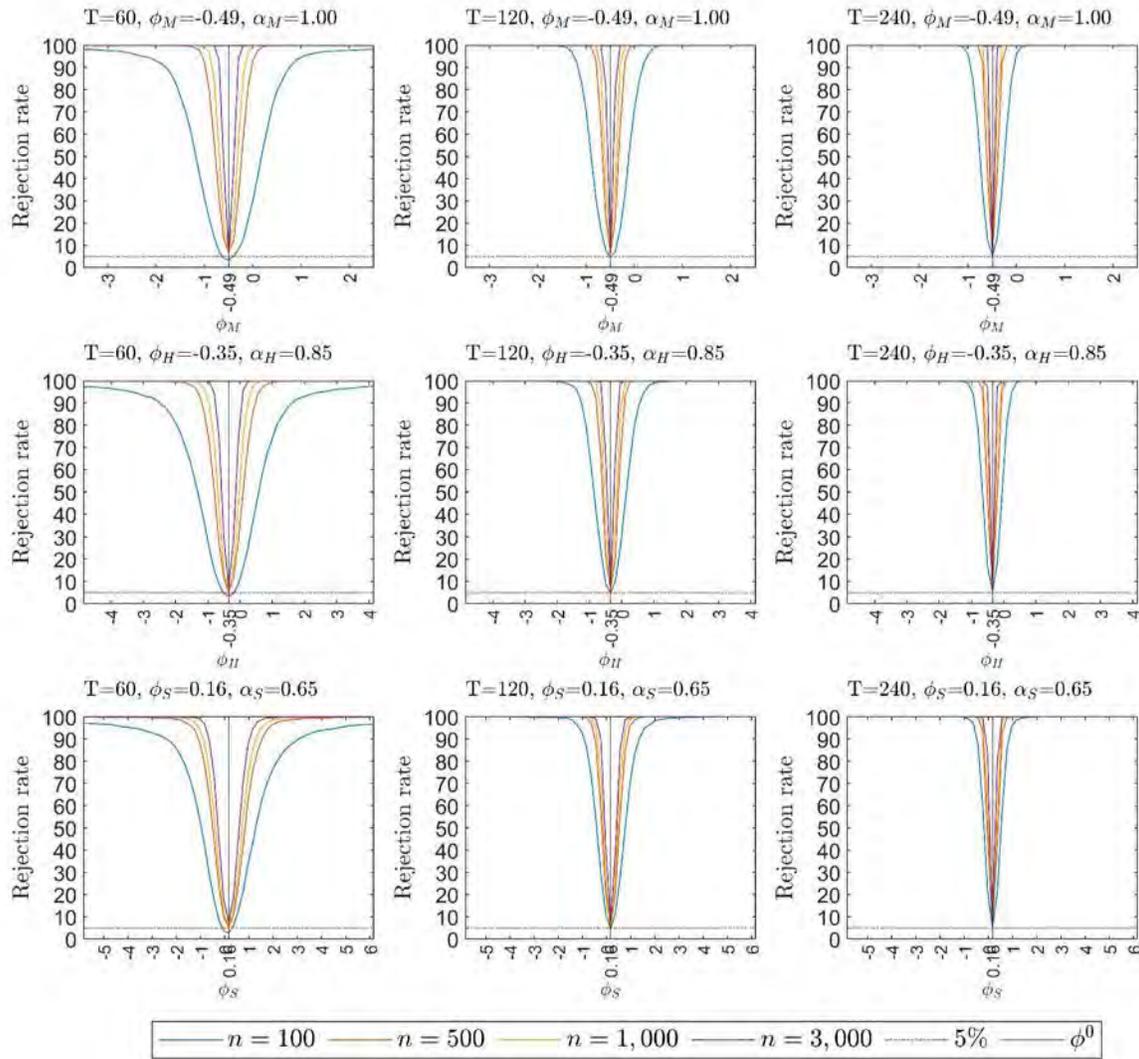
Note: See the notes to Table S-A-E10.

Table S-A-E11: Bias, RMSE and size for the two-step and bias-corrected (BC) estimators of ϕ , for Experiment 11 with one strong and two semi-strong factors

	n	Bias(x100)		RMSE(x100)		Size(x100)		
		Two Step	BC	Two Step	BC	Two Step	BC	
$\phi_M = -0.49, \alpha_M = 1$	T = 60	100	-0.05	-0.72	26.31	47.59	11.30	3.40
		500	-0.33	-0.07	17.63	15.12	29.40	6.70
		1,000	-0.71	-0.22	16.06	10.46	41.95	6.35
		3,000	-0.85	0.05	14.87	5.91	60.85	7.00
	T = 120	100	0.10	-0.46	17.08	19.81	7.15	4.55
		500	0.13	0.10	9.51	8.93	13.90	6.40
		1,000	-0.11	-0.01	7.80	6.28	20.65	6.15
		3,000	-0.21	0.09	6.41	3.60	37.25	6.20
	T = 240	100	0.20	-0.24	12.33	13.43	6.15	5.30
		500	0.24	0.01	6.12	6.11	10.05	6.75
		1,000	0.13	-0.07	4.52	4.19	11.75	5.85
		3,000	0.11	0.02	3.05	2.32	18.35	4.75
$\phi_H = -0.35, \alpha_H = 0.85$	T = 60	100	2.50	-2.25	29.66	103.01	18.75	3.35
		500	3.74	-0.07	24.90	25.61	45.55	5.45
		1,000	3.60	-0.52	24.93	18.83	59.20	6.05
		3,000	3.98	0.01	25.84	12.59	75.45	7.90
	T = 120	100	1.81	-0.44	19.83	27.88	10.00	4.95
		500	2.49	0.09	13.69	13.12	26.00	6.15
		1,000	2.45	0.01	13.32	10.12	40.65	6.00
		3,000	2.63	0.01	13.61	6.86	62.05	7.20
	T = 240	100	0.97	-0.57	13.36	16.38	6.90	5.55
		500	1.60	-0.05	7.76	8.01	13.50	5.40
		1,000	1.72	0.01	6.95	5.93	20.35	5.20
		3,000	1.85	-0.07	6.58	3.89	41.35	5.95
$\phi_S = 0.16, \alpha_S = 0.65$	T = 60	100	-20.60	-0.57	38.15	212.41	27.80	2.75
		500	-23.77	2.07	36.92	42.46	58.25	5.55
		1,000	-25.88	1.00	39.00	34.14	69.55	6.15
		3,000	-28.31	0.81	41.99	28.50	82.20	7.90
	T = 120	100	-12.85	2.34	26.13	35.00	17.60	4.80
		500	-16.99	0.79	23.96	20.38	47.50	5.65
		1,000	-19.15	0.29	25.77	16.81	64.20	5.85
		3,000	-21.97	0.81	28.73	13.50	80.80	7.00
	T = 240	100	-7.60	1.18	17.18	20.04	10.50	5.15
		500	-10.88	0.56	14.80	11.65	35.40	5.55
		1,000	-12.61	0.61	15.89	9.46	55.65	5.80
		3,000	-15.72	0.52	18.57	7.12	80.50	5.35

Notes: The DGP for Experiment 11 allows for Gaussian errors, with GARCH effects, with pricing errors ($\alpha_\eta = 0.3$), with one weak missing factor ($\alpha_\gamma = 0.5$), and with block error cross dependence. For further details of the experiments, see Table S-1.

Figure S-A-E11: Empirical Power Functions, experiment 11 for coefficient of the semi-strong factors



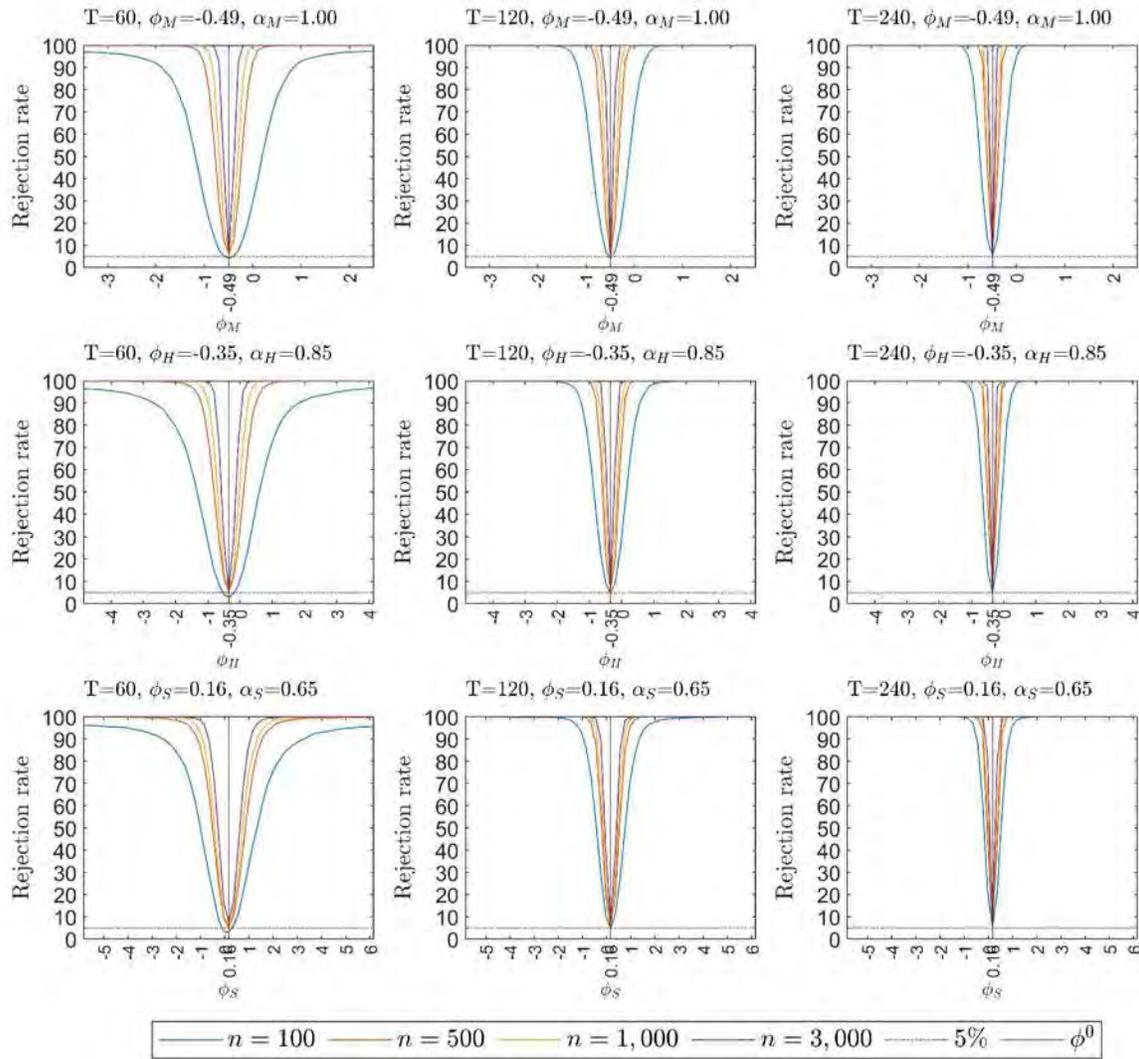
Note: See the notes to Table S-A-E11.

Table S-A-E12: Bias, RMSE and size for the two-step and bias-corrected (BC) estimators of ϕ , for Experiment 12 with one strong and two semi-strong factors

	n	Bias(x100)		RMSE(x100)		Size(x100)		
		Two Step	BC	Two Step	BC	Two Step	BC	
$\phi_M = -0.49, \alpha_M = 1$	T = 60	100	0.09	-0.14	27.00	129.71	11.65	4.10
		500	-0.57	-0.25	17.72	15.76	29.85	6.10
		1,000	-0.60	-0.04	16.17	10.48	41.65	5.65
		3,000	-0.80	0.10	14.89	6.07	60.25	6.75
	T = 120	100	0.22	-0.35	17.19	20.06	6.90	4.15
		500	0.07	0.06	9.47	8.91	13.70	5.40
		1,000	0.04	0.16	7.81	6.35	20.50	6.00
		3,000	-0.18	0.14	6.40	3.62	37.75	5.60
	T = 240	100	0.30	-0.13	12.34	13.45	6.50	5.90
		500	0.30	0.07	6.02	6.01	8.90	6.30
		1,000	0.17	-0.02	4.50	4.21	11.40	5.80
		3,000	0.10	0.02	3.09	2.35	18.15	5.05
$\phi_H = -0.35, \alpha_H = 0.85$	T = 60	100	2.62	9.84	29.60	432.72	18.55	2.75
		500	3.90	0.37	25.08	28.76	45.25	5.35
		1,000	3.57	-0.62	24.92	19.22	58.20	5.95
		3,000	3.94	-0.10	25.92	12.64	74.50	6.90
	T = 120	100	1.82	-0.38	19.87	27.99	9.25	4.65
		500	2.59	0.31	13.82	13.27	25.70	5.85
		1,000	2.45	0.02	13.31	10.26	39.70	5.95
		3,000	2.70	0.11	13.61	6.85	61.20	7.15
	T = 240	100	0.92	-0.65	13.61	16.79	7.15	4.90
		500	1.65	0.02	7.79	8.06	12.50	5.05
		1,000	1.77	0.07	7.03	6.03	21.70	5.55
		3,000	1.87	-0.05	6.57	3.93	41.05	6.05
$\phi_S = 0.16, \alpha_S = 0.65$	T = 60	100	-20.25	1.01	38.24	407.11	26.15	3.00
		500	-23.75	1.77	36.94	55.06	55.90	5.30
		1,000	-26.03	0.80	39.26	35.51	69.75	6.50
		3,000	-28.18	1.33	41.98	27.55	81.60	7.85
	T = 120	100	-13.05	2.21	26.45	36.34	18.15	5.05
		500	-16.97	0.78	24.00	20.16	47.40	5.40
		1,000	-19.18	0.24	25.77	17.03	64.35	6.10
		3,000	-21.96	0.84	28.73	13.65	80.55	7.20
	T = 240	100	-7.73	1.03	17.33	20.17	11.10	5.40
		500	-10.89	0.50	14.85	11.67	34.35	5.40
		1,000	-12.73	0.44	15.92	9.49	54.85	5.60
		3,000	-15.75	0.49	18.58	7.25	80.80	5.65

Notes: The DGP for Experiment 12 allows for t(5) distributed errors, with GARCH effects, with pricing errors ($\alpha_\eta = 0.3$), with one weak missing factor ($\alpha_\gamma = 0.5$), and with block error cross dependence. For further details of the experiments, see S-1.

Figure S-A-E12: Empirical Power Functions, experiment 12 for coefficient of the semi-strong factors



Note: See the notes to Table S-A-E12.

2.1 Estimators of ϕ for one strong and two semi-strong factors, threshold estimator of the covariance matrix with and without misspecification

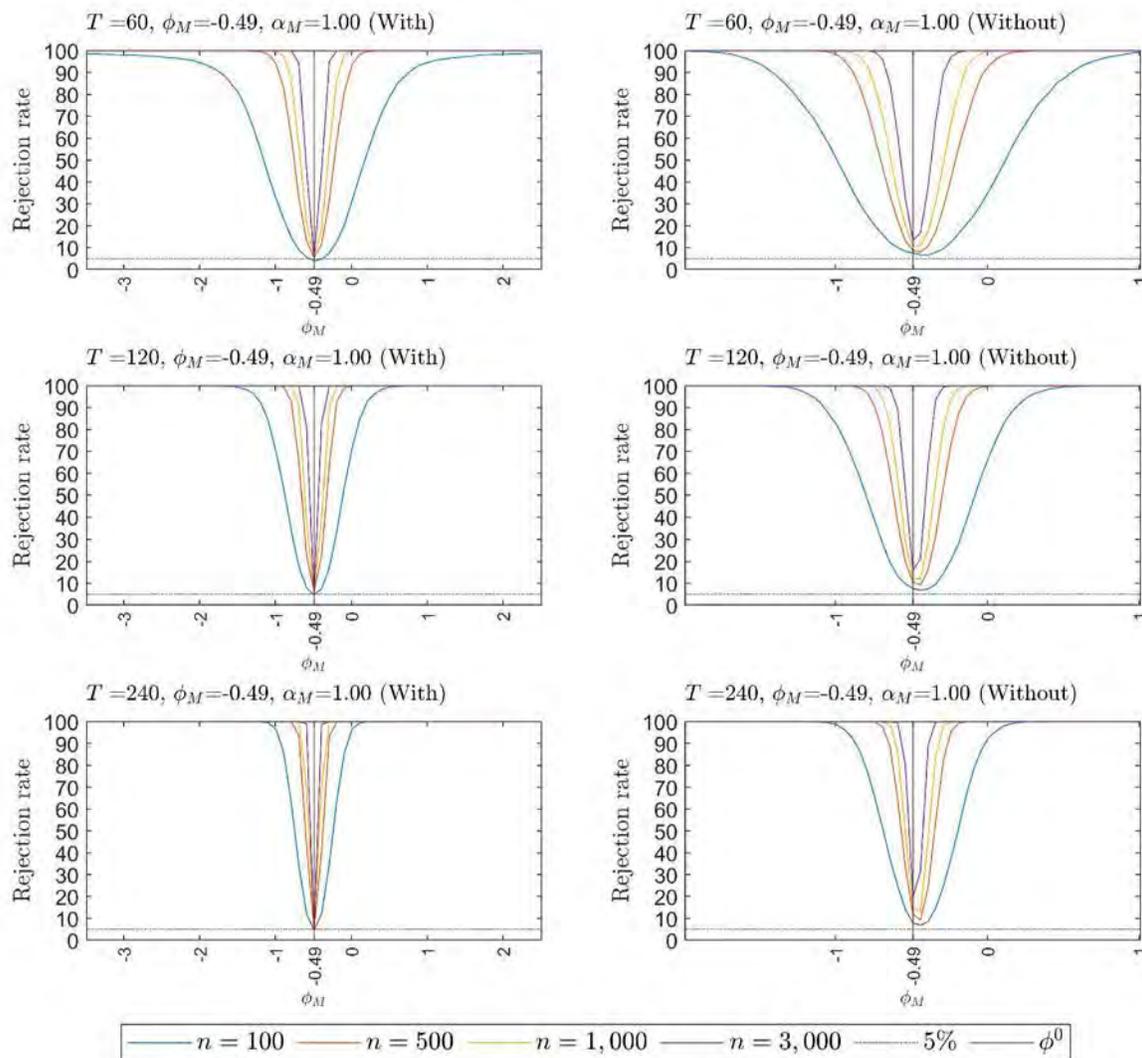
This subsection compares the estimators of ϕ for one strong and two semi-strong factors with and without misspecification. Under the misspecification, only strong estimators ϕ_M are considered in the estimation process. Each table contains the results of three of the 12 experiments described in Table S-1, and corresponding empirical power functions follow.

Table S-B-E1-3: Bias, RMSE and size for the bias-corrected estimators of $\phi_M = -0.49$, $\alpha_M = 1$ with and without semi-strong factors ($\alpha_H = 0.85$, $\alpha_S = 0.65$) included in the regression for the cases of experiments 1, 2 and 3

Experiment 1	n	Bias(x100)		RMSE(x100)		Size(x100)		
		With	Without	With	Without	With	Without	
		semi-strong factors		semi-strong factors		semi-strong factors		
$T = 60$	100	-1.55	5.63	86.14	31.56	4.00	7.60	
	500	0.28	3.72	14.33	14.60	5.40	8.80	
	1,000	0.10	2.96	10.12	11.02	6.20	10.85	
	3,000	-0.01	1.96	5.56	6.98	5.70	13.50	
$T = 120$	100	1.84	4.96	91.35	20.49	4.70	7.50	
	500	0.15	3.36	8.99	9.93	6.15	10.50	
	1,000	0.01	2.72	6.22	7.40	5.65	12.15	
	3,000	-0.06	1.88	3.49	4.71	4.95	15.80	
$T = 240$	100	-0.15	5.00	13.11	14.46	4.55	7.90	
	500	0.09	3.33	5.75	7.03	4.85	11.90	
	1,000	0.03	2.73	4.06	5.33	5.30	14.40	
	3,000	-0.00	1.95	2.35	3.52	5.10	19.95	
Experiment 2	$T = 60$	100	0.27	6.04	45.08	32.20	3.70	7.25
		500	0.27	3.69	14.47	14.60	5.75	7.75
		1,000	0.04	2.85	10.24	11.07	6.10	11.10
		3,000	-0.05	1.94	5.72	7.03	5.90	14.45
	$T = 120$	100	0.02	5.24	20.45	20.58	4.90	8.30
		500	0.23	3.46	9.16	10.04	6.60	10.35
		1,000	0.03	2.73	6.12	7.37	5.55	11.35
		3,000	-0.09	1.83	3.52	4.70	5.65	16.10
	$T = 240$	100	-0.14	5.08	13.18	14.44	5.05	8.75
		500	0.11	3.39	5.81	7.06	5.05	11.65
		1,000	0.04	2.74	4.07	5.35	5.55	15.50
		3,000	-0.03	1.92	2.36	3.51	5.25	20.55
Experiment 3	$T = 60$	100	-6.00	5.70	227.89	31.96	3.40	7.55
		500	0.29	3.68	14.71	14.81	5.65	9.10
		1,000	0.09	2.94	10.33	11.25	6.35	11.55
		3,000	0.00	1.96	5.65	7.18	5.65	14.55
	$T = 120$	100	-0.19	4.89	20.63	20.51	4.50	7.45
		500	0.15	3.32	9.05	9.98	6.25	10.65
		1,000	0.02	2.68	6.25	7.46	5.75	12.15
		3,000	-0.05	1.85	3.50	4.75	4.90	15.20
	$T = 240$	100	-0.13	4.96	13.17	14.48	4.45	7.90
		500	0.10	3.32	5.78	7.05	5.20	11.85
		1,000	0.05	2.72	4.07	5.34	5.25	14.10
		3,000	0.00	1.93	2.36	3.53	5.20	19.90

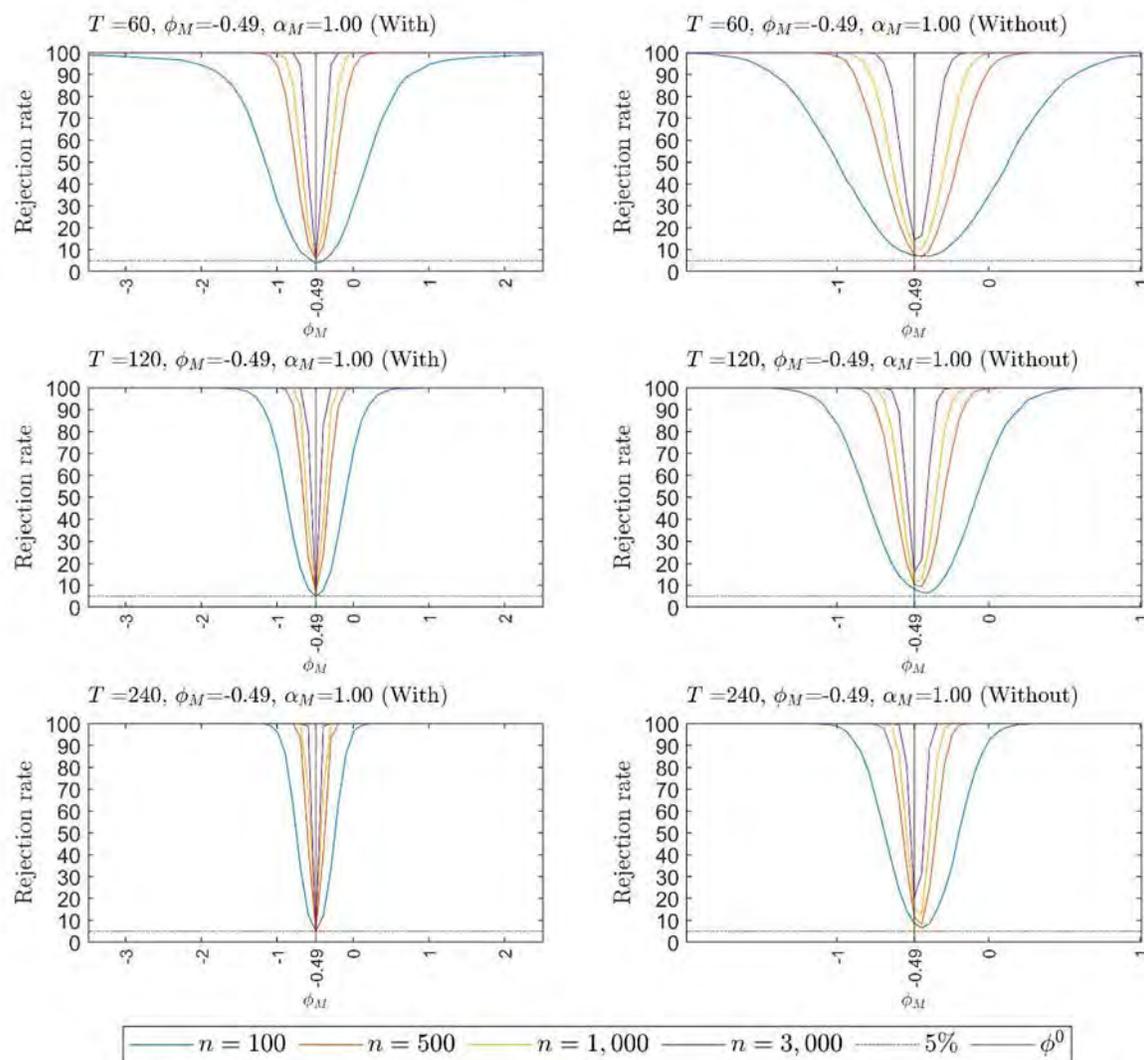
Notes: The DGP includes one strong $\alpha_M = 1$ and two semi-strong ($\alpha_H = 0.85$, $\alpha_S = 0.65$) factors, the regression with the two semi-strong factors includes them, the regression without excludes them. For further details of the experiments, see S-1.

Figure S-B-E1: Empirical Power Functions, experiment 1, for coefficient of the ϕ_M factor with and without misspecification



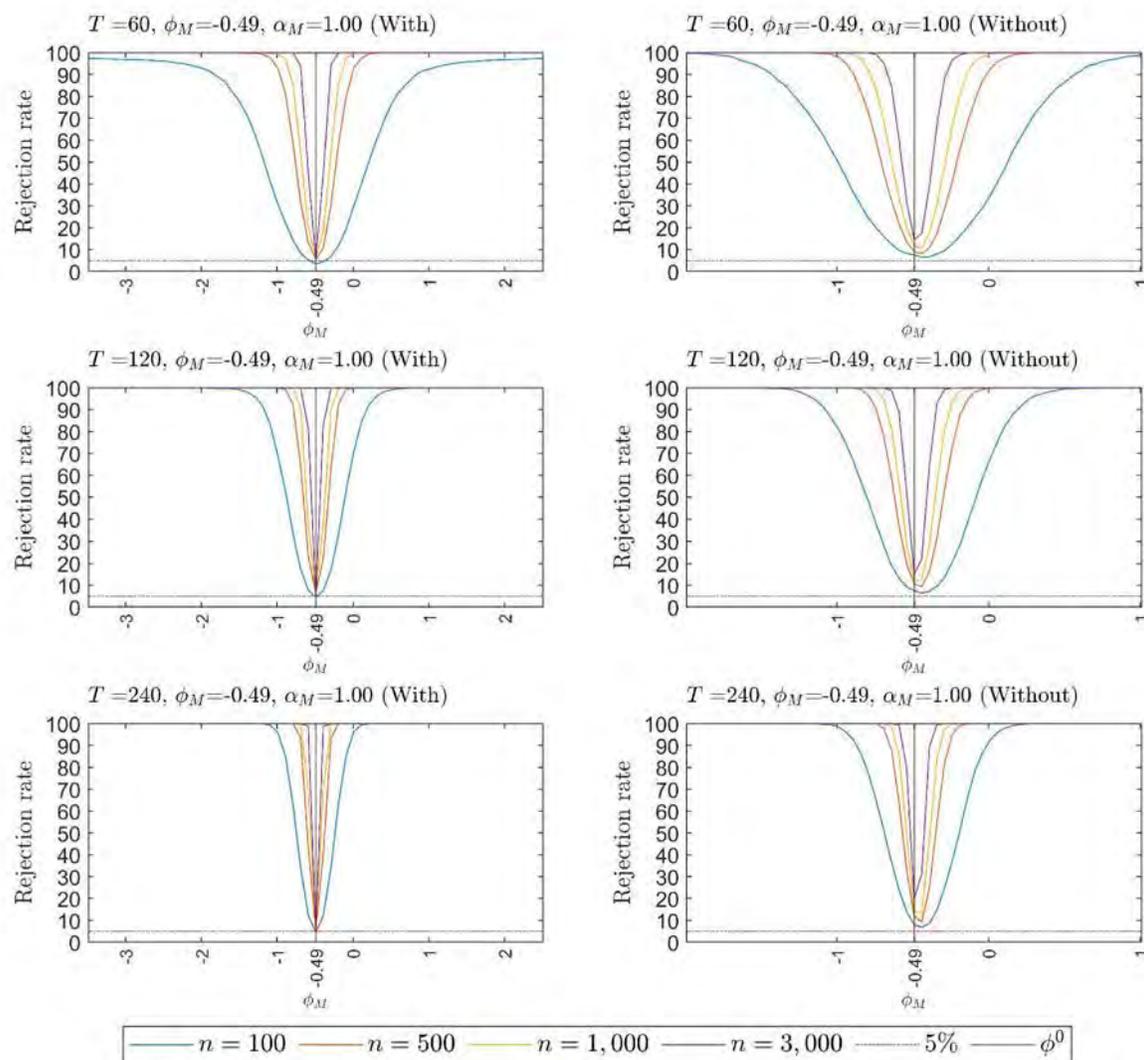
Note: See the notes to Table S-B-E1-3.

Figure S-B-E2: Empirical Power Functions, experiment 2, for coefficient of the ϕ_M factor with and without misspecification



Note: See the notes to Table S-B-E1-3.

Figure S-B-E3: Empirical Power Functions, experiment 3, for coefficient of the ϕ_M factor with and without misspecification



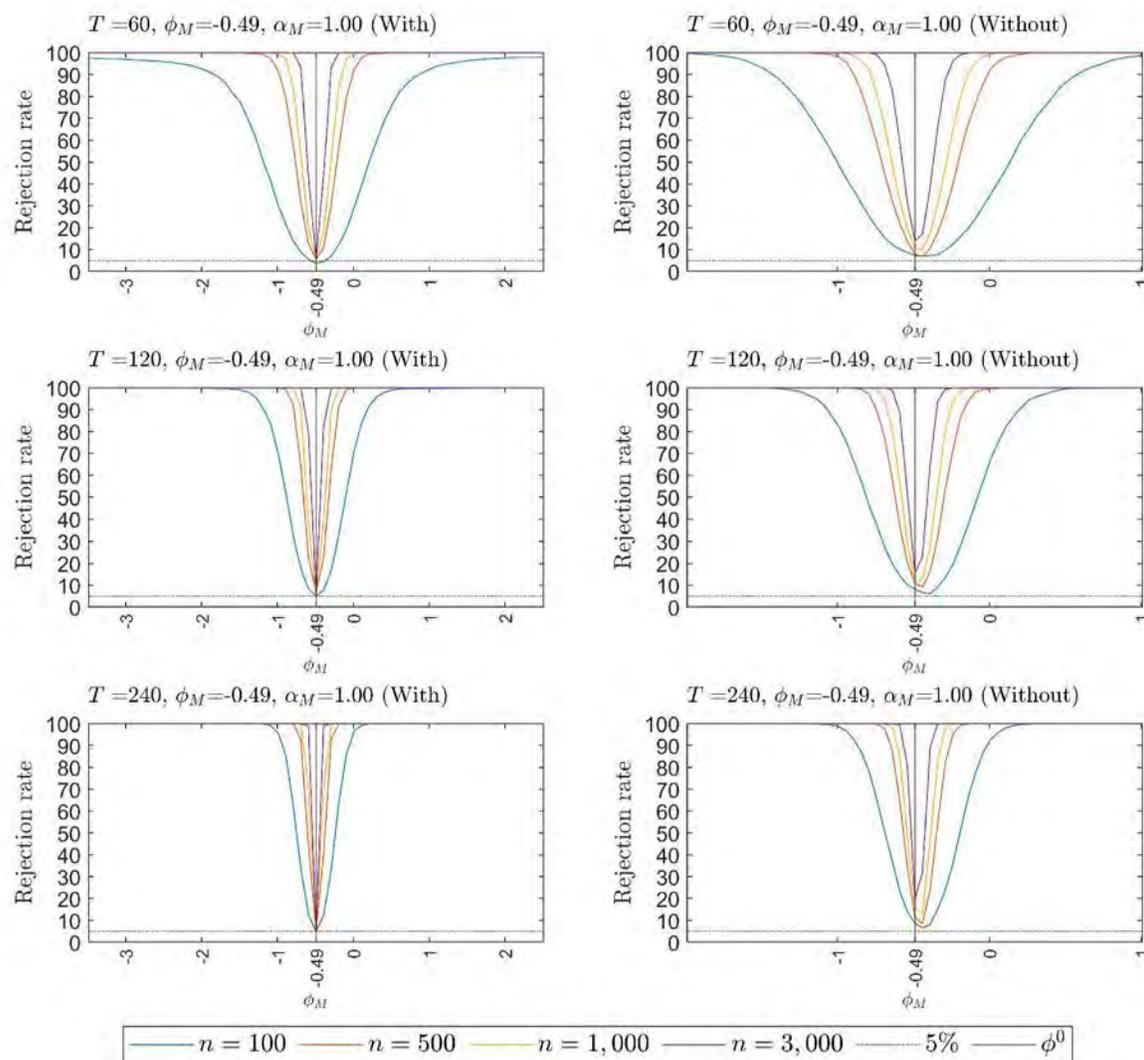
Note: See the notes to Table S-B-E1-3.

Table S-B-E4-6: Bias, RMSE and size for the bias-corrected estimators of $\phi_M = -0.49$, $\alpha_M = 1$ with and without semi-strong factors ($\alpha_H = 0.85$, $\alpha_S = 0.65$) included in the regression for the cases of experiments 4, 5 and 6

Experiment 4	n	Bias(x100)		RMSE(x100)		Size(x100)		
		With	Without	With	Without	With	Without	
		semi-strong factors		semi-strong factors		semi-strong factors		
T = 60	100	0.17	6.08	107.83	32.68	3.75	7.25	
	500	0.23	3.64	14.75	14.88	5.65	7.85	
	1,000	0.05	2.84	10.43	11.29	6.10	10.85	
	3,000	-0.04	1.94	5.82	7.21	5.95	14.15	
T = 120	100	-0.86	5.15	44.37	20.62	4.95	8.10	
	500	0.21	3.42	9.24	10.11	6.45	10.75	
	1,000	0.03	2.69	6.16	7.41	4.90	11.05	
	3,000	-0.08	1.81	3.53	4.75	5.85	15.55	
T = 240	100	-0.13	5.03	13.22	14.46	4.85	8.70	
	500	0.12	3.37	5.83	7.07	5.10	11.15	
	1,000	0.05	2.72	4.08	5.36	5.60	15.50	
	3,000	-0.02	1.90	2.36	3.52	5.25	20.00	
Experiment 5	T = 60	100	-7.58	5.70	285.21	31.96	3.75	7.55
		500	0.28	3.68	14.79	14.86	5.80	9.05
		1,000	0.10	2.92	10.35	11.25	6.25	11.80
		3,000	-0.00	1.97	5.65	7.19	5.75	14.70
	T = 120	100	-0.35	4.89	20.82	20.51	4.95	7.45
		500	0.14	3.32	9.15	10.03	6.45	10.30
		1,000	0.04	2.67	6.28	7.48	5.90	12.20
		3,000	-0.06	1.86	3.50	4.76	4.80	15.45
	T = 240	100	-0.31	4.96	13.53	14.48	5.35	7.90
		500	0.10	3.31	5.84	7.08	5.35	11.50
		1,000	0.07	2.71	4.11	5.36	5.20	14.10
		3,000	0.00	1.94	2.36	3.53	5.25	20.30
Experiment 6	T = 60	100	0.05	6.08	103.65	32.68	3.95	7.25
		500	0.24	3.63	14.81	14.93	5.95	8.25
		1,000	0.06	2.82	10.44	11.30	6.00	11.00
		3,000	-0.05	1.95	5.82	7.22	5.80	14.40
	T = 120	100	-0.76	5.15	33.47	20.62	5.15	8.10
		500	0.22	3.42	9.32	10.16	6.95	10.65
		1,000	0.05	2.68	6.18	7.43	5.30	11.25
		3,000	-0.09	1.82	3.54	4.75	5.70	15.40
	T = 240	100	-0.30	5.03	13.63	14.46	6.10	8.70
		500	0.11	3.37	5.88	7.10	5.10	11.85
		1,000	0.07	2.71	4.13	5.38	5.80	15.50
		3,000	-0.03	1.91	2.37	3.52	5.45	20.40

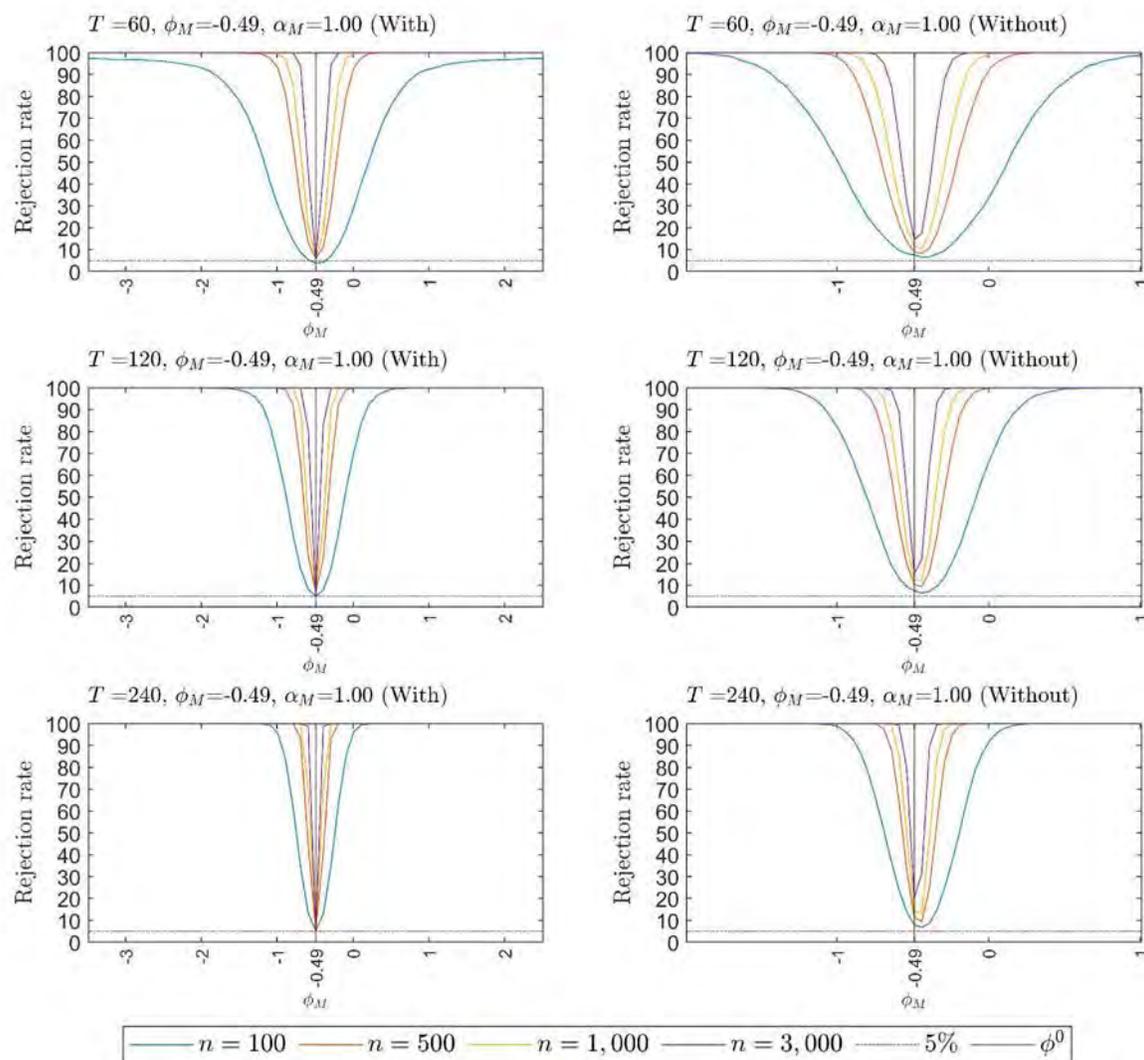
Notes: The DGP includes one strong $\alpha_M = 1$ and two semi-strong ($\alpha_H = 0.85$, $\alpha_S = 0.65$) factors, the regression with the two semi-strong factors includes them, the regression without excludes them. For further details of the experiments, see S-1.

Figure S-B-E4: Empirical Power Functions, experiment 4, for coefficient of the ϕ_M factor with and without misspecification



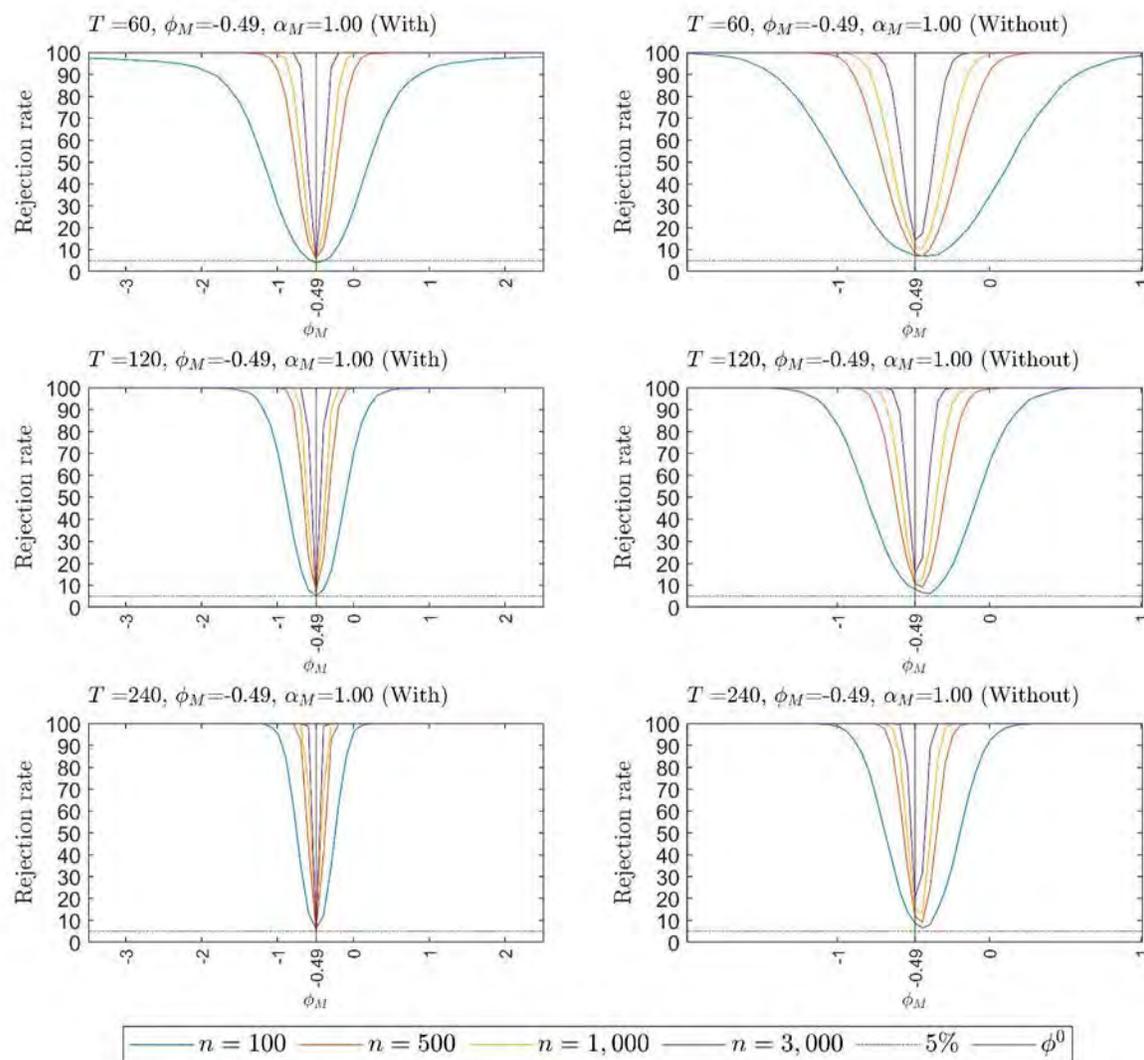
Note: See the notes to Table S-B-E4-6.

Figure S-B-E5: Empirical Power Functions, experiment 5, for coefficient of the ϕ_M factor with and without misspecification



Note: See the notes to Table S-B-E4-6.

Figure S-B-E6: Empirical Power Functions, experiment 6, for coefficient of the ϕ_M factor with and without misspecification



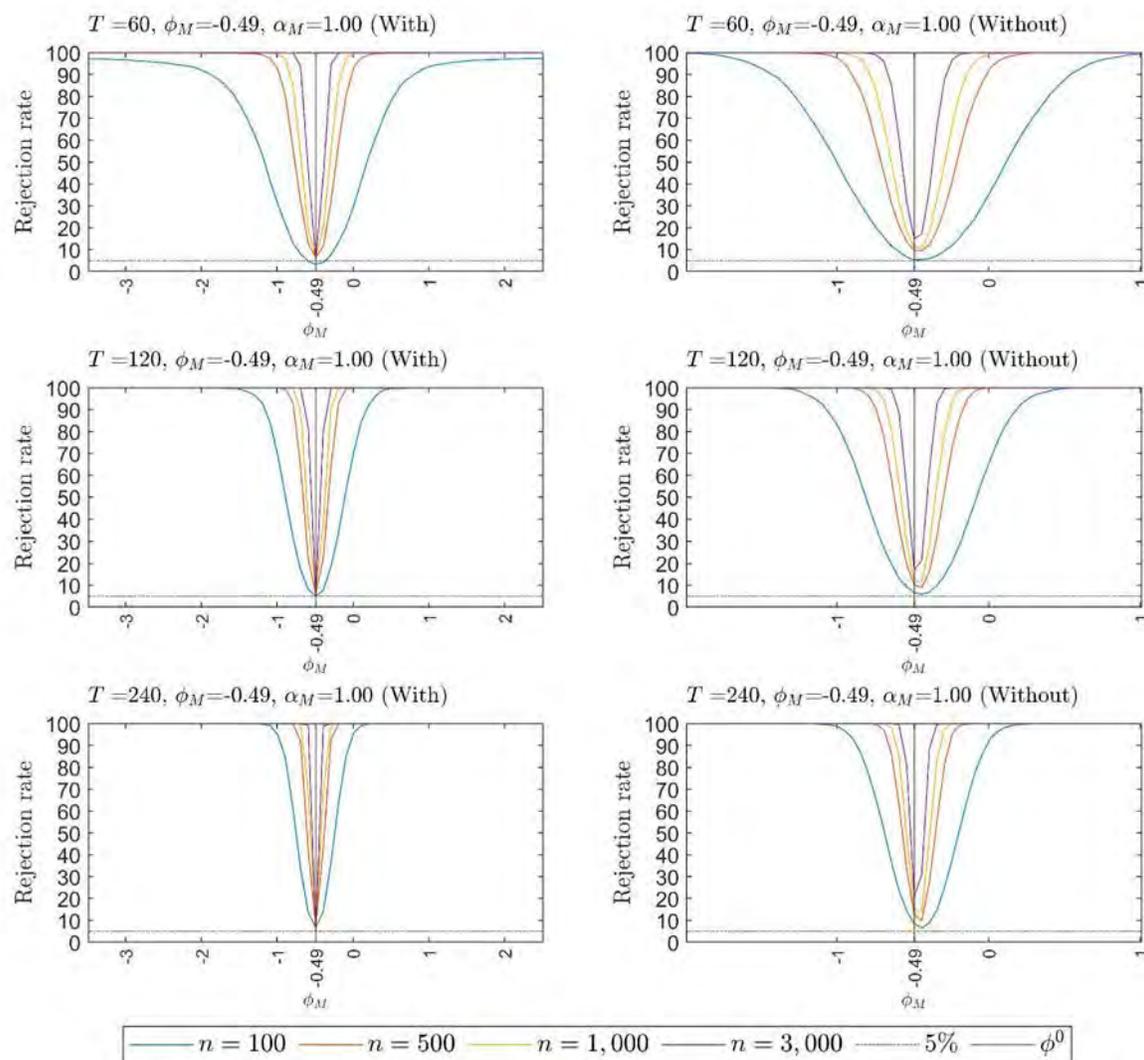
Note: See the notes to Table S-B-E4-6.

Table S-B-E7-9: Bias, RMSE and size for the bias-corrected estimators of $\phi_M = -0.49$, $\alpha_M = 1$ with and without semi-strong factors ($\alpha_H = 0.85$, $\alpha_S = 0.65$) included in the regression for the cases of experiments 7, 8 and 9

Experiment 7	n	Bias(x100)		RMSE(x100)		Size(x100)		
		With	Without	With	Without	With	Without	
		semi-strong factors		semi-strong factors		semi-strong factors		
T = 60	100	-14.31	4.76	555.11	31.07	3.20	5.35	
	500	-0.18	3.25	15.13	15.20	6.40	9.65	
	1,000	-0.15	2.62	10.41	11.29	6.25	11.35	
	3,000	0.04	1.99	5.81	7.17	5.90	14.95	
T = 120	100	-0.49	4.70	20.30	19.98	5.05	6.55	
	500	0.01	3.15	8.86	9.80	5.85	10.15	
	1,000	0.04	2.62	6.31	7.45	6.60	11.85	
	3,000	0.08	1.99	3.60	4.80	5.65	17.05	
T = 240	100	-0.39	4.91	14.02	14.48	6.75	8.50	
	500	-0.08	3.13	6.13	7.18	6.95	12.85	
	1,000	-0.03	2.67	4.22	5.45	5.80	15.15	
	3,000	0.01	1.97	2.32	3.50	5.15	20.90	
Experiment 8	T = 60	100	9.06	4.80	251.34	31.95	3.75	6.85
		500	-0.67	2.88	16.71	15.13	6.30	9.95
		1,000	0.01	2.75	10.38	11.32	5.70	10.25
		3,000	0.12	2.09	5.96	7.25	6.15	14.70
	T = 120	100	-0.32	4.83	20.53	20.18	4.75	6.25
		500	-0.04	3.05	8.89	9.73	5.80	10.20
		1,000	0.19	2.76	6.40	7.60	6.10	12.70
		3,000	0.14	2.01	3.63	4.81	5.90	17.05
	T = 240	100	-0.29	5.00	14.07	14.58	6.35	9.40
		500	-0.01	3.19	6.06	7.15	6.05	12.05
		1,000	0.01	2.68	4.25	5.46	6.30	15.90
		3,000	0.01	1.96	2.35	3.52	5.15	20.95
Experiment 9	T = 60	100	-15.47	4.04	662.68	33.36	3.40	6.95
		500	1.29	2.84	84.39	15.74	7.75	11.40
		1,000	-0.47	2.26	11.31	11.50	8.50	12.25
		3,000	-0.16	1.69	6.25	7.20	7.75	15.35
	T = 120	100	-0.59	4.09	21.49	21.20	4.50	7.00
		500	-0.01	2.95	9.20	9.95	6.15	9.15
		1,000	-0.07	2.40	6.53	7.37	6.30	10.80
		3,000	-0.01	1.74	3.71	4.68	6.50	14.30
	T = 240	100	-0.54	4.06	14.02	14.60	6.85	8.50
		500	-0.12	2.85	6.19	7.11	6.55	10.80
		1,000	-0.08	2.42	4.30	5.27	6.60	13.35
		3,000	-0.00	1.75	2.37	3.37	5.45	17.50

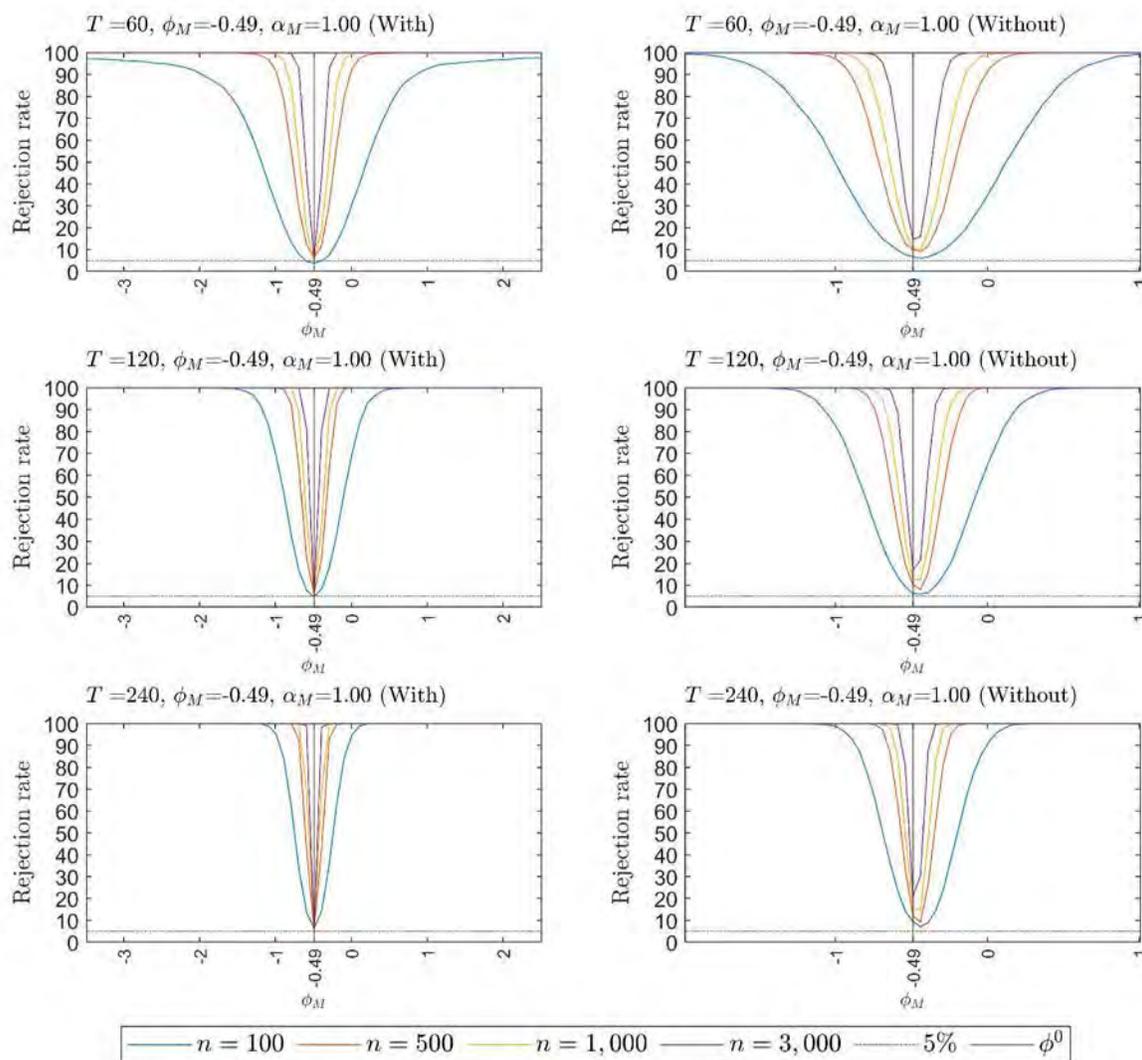
Notes: The DGP includes one strong $\alpha_M = 1$ and two semi-strong ($\alpha_H = 0.85$, $\alpha_S = 0.65$) factors, the regression with the two semi-strong factors includes them, the regression without excludes them. For further details of the experiments, see S-1.

Figure S-B-E7: Empirical Power Functions, experiment 7, for coefficient of the ϕ_M factor with and without misspecification



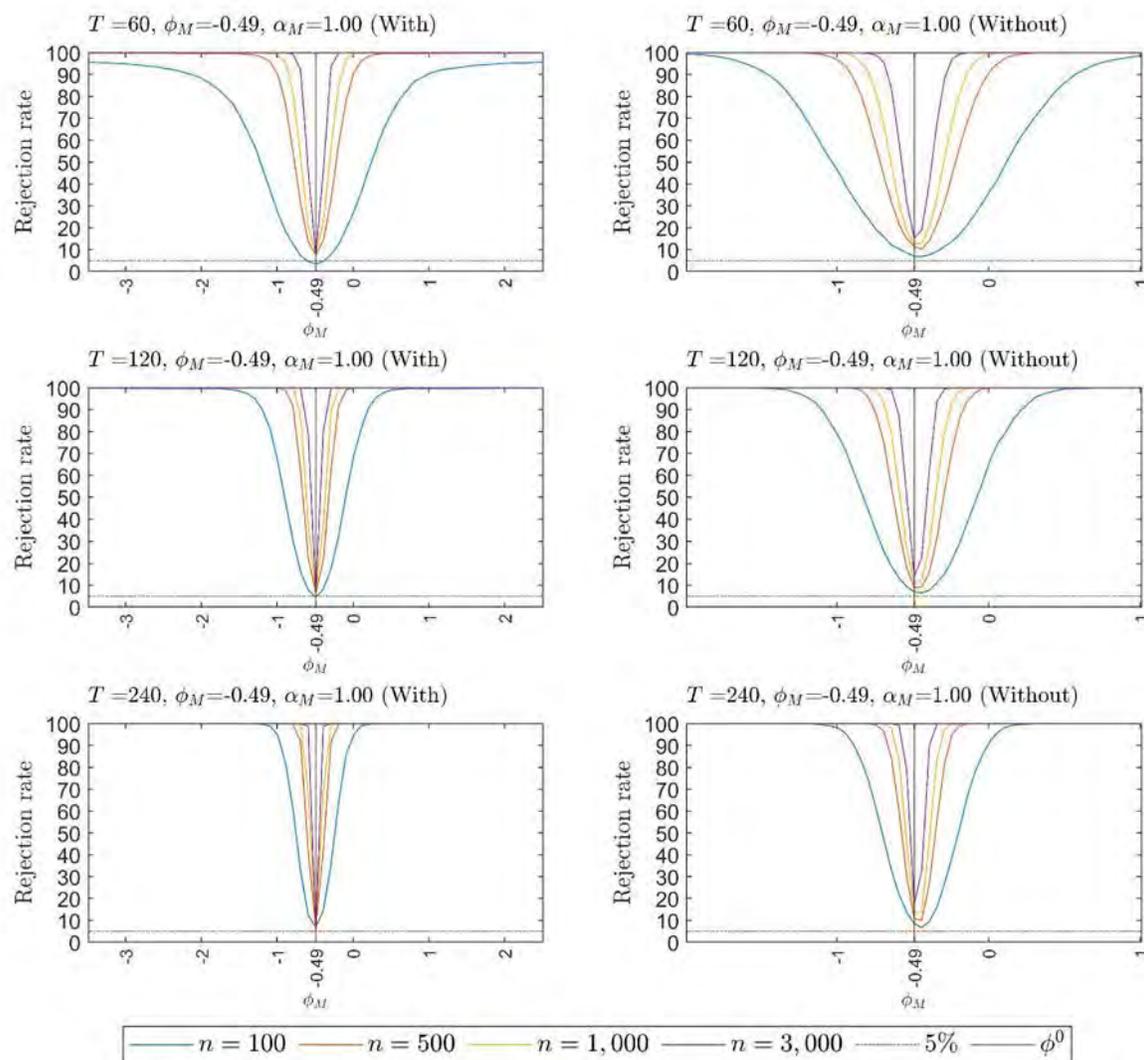
Note: See the notes to Table S-B-E7-9.

Figure S-B-E8: Empirical Power Functions, experiment 8, for coefficient of the ϕ_M factor with and without misspecification



Note: See the notes to Table S-B-E7-9.

Figure S-B-E9: Empirical Power Functions, experiment 9, for coefficient of the ϕ_M factor with and without misspecification



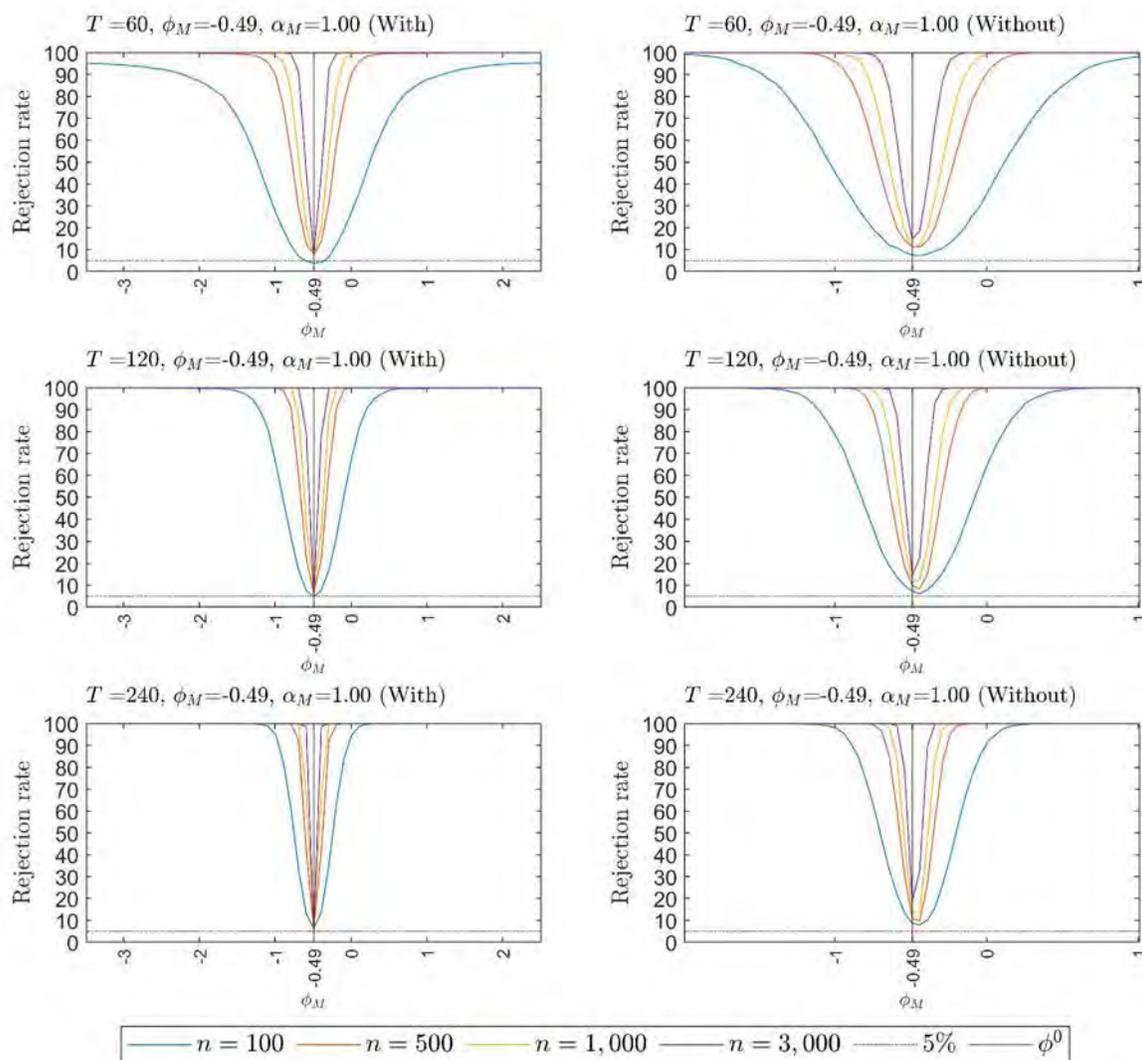
Note: See the notes to Table S-B-E7-9.

Table S-B-E10-12: Bias, RMSE and size for the bias-corrected estimators of $\phi_M = -0.49$, $\alpha_M = 1$ with and without semi-strong factors ($\alpha_H = 0.85$, $\alpha_S = 0.65$) included in the regression for the cases of experiments 10, 11 and 12

Experiment 10	n	Bias(x100)		RMSE(x100)		Size(x100)		
		With	Without	With	Without	With	Without	
		semi-strong factors		semi-strong factors		semi-strong factors		
T = 60	100	-1.61	4.07	151.91	33.92	3.75	7.45	
	500	-0.69	2.46	17.07	15.76	7.70	11.35	
	1,000	-0.31	2.39	11.30	11.47	7.05	10.90	
	3,000	-0.06	1.79	6.36	7.31	8.00	14.95	
T = 120	100	-1.86	4.22	69.05	21.53	4.80	7.25	
	500	-0.09	2.81	9.29	9.92	6.10	9.85	
	1,000	0.15	2.59	6.62	7.51	6.75	11.40	
	3,000	0.05	1.76	3.72	4.69	5.80	14.55	
T = 240	100	-0.39	4.20	14.17	14.82	6.35	8.60	
	500	-0.05	2.87	6.15	7.09	6.05	10.80	
	1,000	0.00	2.46	4.30	5.28	6.40	13.65	
	3,000	0.00	1.75	2.40	3.40	5.75	18.80	
Experiment 11	T = 60	100	-0.72	4.82	47.59	31.18	3.40	5.90
		500	-0.07	3.31	15.12	15.30	6.70	10.30
		1,000	-0.22	2.56	10.46	11.28	6.35	11.45
		3,000	0.05	1.96	5.91	7.26	7.00	16.00
	T = 120	100	-0.46	4.75	19.81	20.00	4.55	6.60
		500	0.10	3.25	8.93	9.83	6.40	10.00
		1,000	-0.01	2.59	6.28	7.41	6.15	11.95
		3,000	0.09	1.99	3.60	4.83	6.20	16.70
	T = 240	100	-0.24	4.94	13.43	14.51	5.30	8.45
		500	0.01	3.22	6.11	7.23	6.75	12.95
		1,000	-0.07	2.64	4.19	5.42	5.85	14.65
		3,000	0.02	1.97	2.32	3.50	4.75	20.55
Experiment 12	T = 60	100	-0.14	4.87	129.71	32.05	4.10	6.75
		500	-0.25	2.94	15.76	15.19	6.10	9.45
		1,000	-0.04	2.71	10.48	11.34	5.65	10.40
		3,000	0.10	2.05	6.07	7.34	6.75	15.20
	T = 120	100	-0.35	4.86	20.06	20.21	4.15	6.35
		500	0.06	3.16	8.91	9.76	5.40	10.35
		1,000	0.16	2.76	6.35	7.57	6.00	12.75
		3,000	0.14	2.01	3.62	4.82	5.60	16.95
	T = 240	100	-0.13	5.03	13.45	14.61	5.90	9.25
		500	0.07	3.27	6.01	7.18	6.30	12.40
		1,000	-0.02	2.66	4.21	5.41	5.80	15.70
		3,000	0.02	1.96	2.35	3.52	5.05	20.65

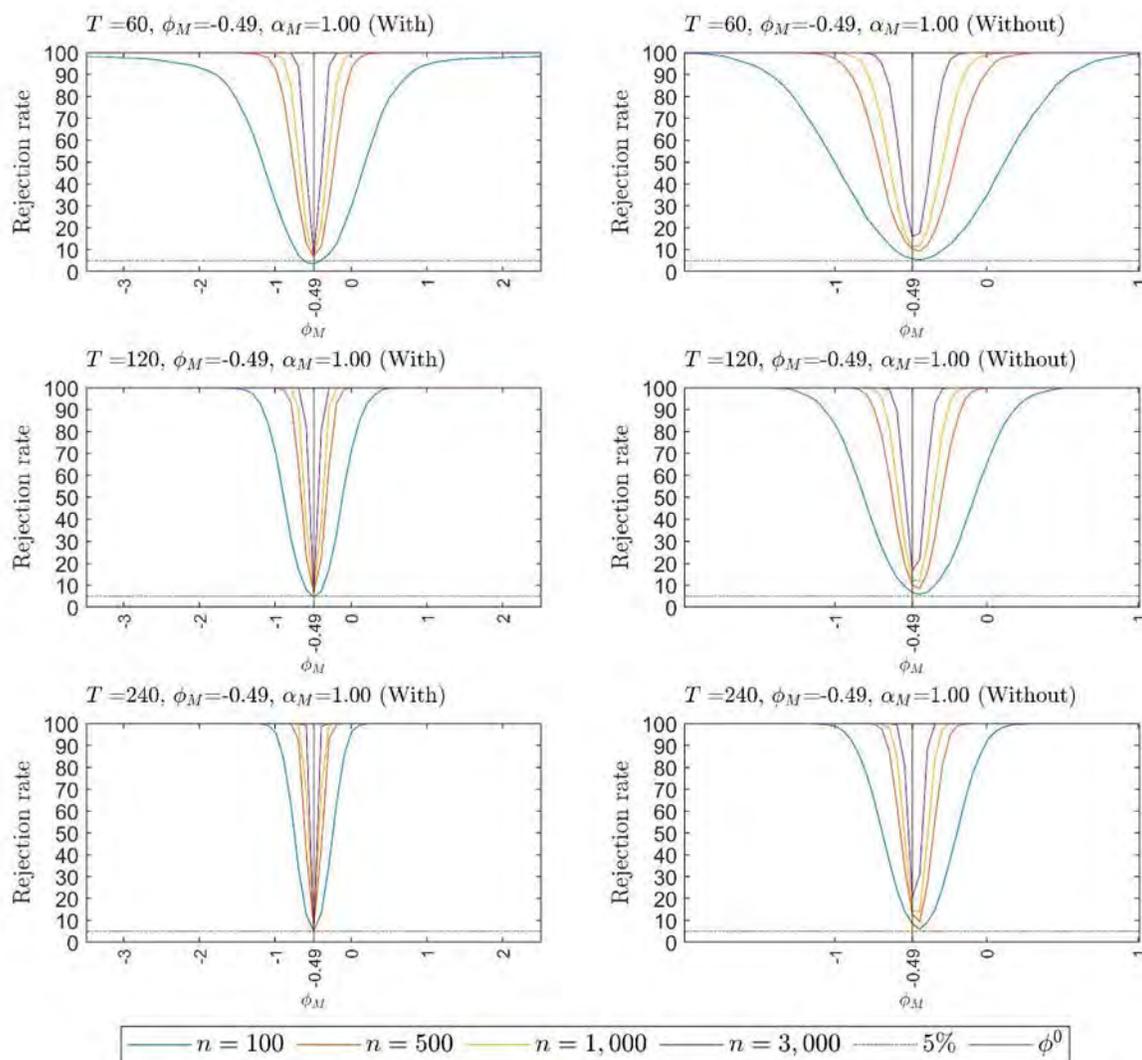
Notes: The DGP includes one strong $\alpha_M = 1$ and two semi-strong ($\alpha_H = 0.85$, $\alpha_S = 0.65$) factors, the regression with the two semi-strong factors includes them, the regression without excludes them. For further details of the experiments, see S-1.

Figure S-B-E10: Empirical Power Functions, experiment 10, for coefficient of the ϕ_M factor with and without misspecification



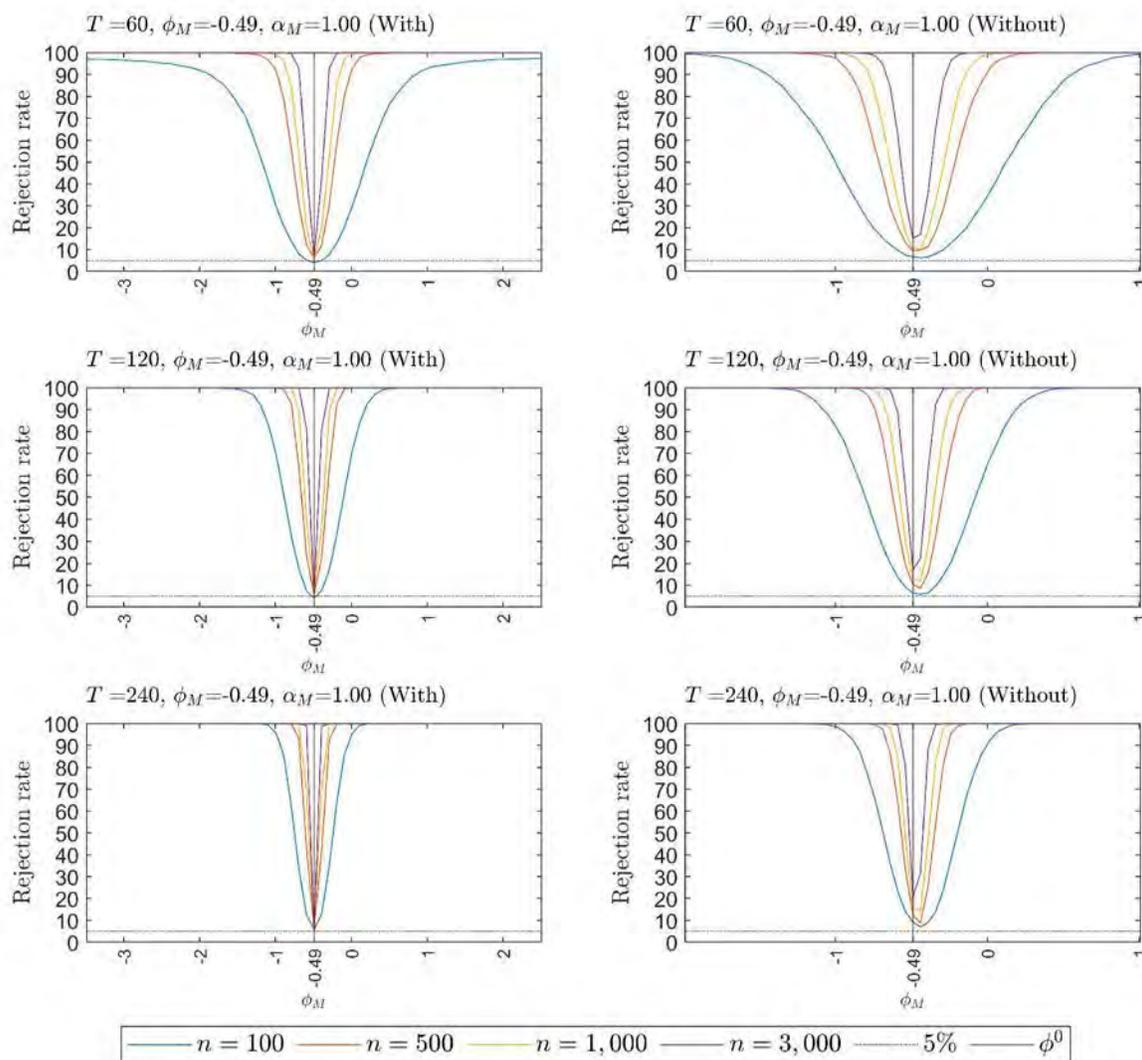
Note: See the notes to Table S-B-E10-12.

Figure S-B-E11: Empirical Power Functions, experiment 11, for coefficient of the ϕ_M factor with and without misspecification



Note: See the notes to Table S-B-E10-12.

Figure S-B-E12: Empirical Power Functions, experiment 12, for coefficient of the ϕ_M factor with and without misspecification



Note: See the notes to Table S-B-E10-12.

2.2 Estimators of ϕ for one strong and two weak factors, threshold estimator of the covariance matrix with and without misspecification

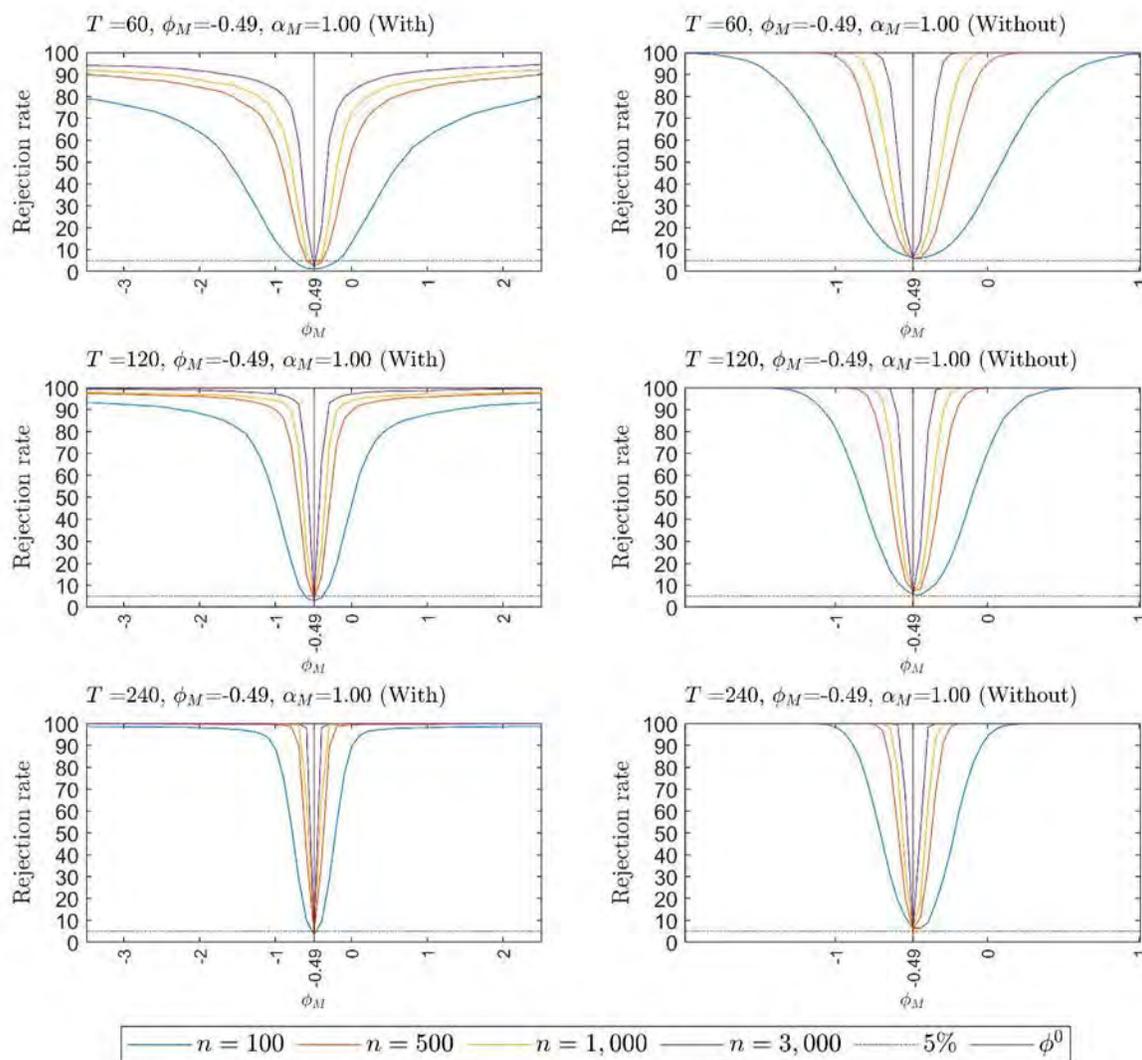
This subsection compares the estimators of ϕ for one strong and two weak factors with and without misspecification. Under the misspecification, only strong estimators ϕ_M are considered in the estimation process. Each table contains the results of three of the 12 experiments described in Table S-1, and corresponding empirical power functions follow.

Table S-C-E1-3: Bias, RMSE and size for the bias-corrected estimators of $\phi_M = -0.49$, $\alpha_M=1$ with and without weak factors included in the regression for the cases of experiments 1,2 and 3

Experiment 1	n	Bias(x100)		RMSE(x100)		Size(x100)	
		With	Without	With	Without	With	Without
		weak factors		weak factors		weak factors	
T = 60	100	22.05	3.67	454.21	30.53	1.15	6.10
	500	-10.55	1.66	332.83	13.13	2.25	6.15
	1,000	-1.33	1.10	382.27	9.48	1.65	6.65
	3,000	-19.15	0.58	932.38	5.38	2.00	6.20
T = 120	100	6.52	3.04	237.70	19.49	2.90	5.80
	500	0.73	1.37	35.05	8.75	3.20	7.30
	1,000	4.61	0.94	152.17	6.12	3.60	6.50
	3,000	0.02	0.50	6.94	3.42	2.90	5.00
T = 240	100	8.56	2.91	385.16	13.50	3.45	6.45
	500	0.22	1.24	7.22	5.80	4.00	5.80
	1,000	0.06	0.90	4.23	4.14	4.00	6.50
	3,000	0.01	0.51	2.41	2.38	4.90	5.25
Experiment 2 T = 60	100	-13.53	4.08	774.85	31.16	1.40	6.00
	500	-11.15	1.63	755.67	13.33	2.10	5.65
	1,000	10.51	1.01	256.15	9.61	2.00	6.55
	3,000	-0.10	0.56	71.71	5.50	2.25	6.60
T = 120	100	468.29	3.25	20351.06	19.57	2.25	6.60
	500	5.00	1.46	141.50	8.89	3.85	7.55
	1,000	-0.54	0.97	16.07	6.06	2.80	5.75
	3,000	0.00	0.45	9.71	3.45	3.35	5.45
T = 240	100	-1.38	2.96	67.20	13.49	4.15	7.50
	500	0.04	1.28	6.83	5.85	4.15	5.60
	1,000	0.11	0.91	4.59	4.15	4.65	7.15
	3,000	-0.03	0.48	2.46	2.39	5.15	6.10
Experiment 3 T = 60	100	11.17	3.75	686.97	30.84	1.00	6.25
	500	-10.50	1.65	377.36	13.28	2.00	6.15
	1,000	-4.65	1.09	120.83	9.59	1.50	6.15
	3,000	-4.51	0.59	134.45	5.45	1.90	6.25
T = 120	100	5.69	3.01	267.53	19.52	2.65	5.65
	500	-5.01	1.36	140.66	8.79	3.15	7.45
	1,000	-0.91	0.94	23.35	6.14	3.35	6.45
	3,000	-0.35	0.50	20.92	3.43	2.70	4.75
T = 240	100	1.04	2.89	38.32	13.51	3.45	6.35
	500	0.32	1.24	8.13	5.82	3.95	5.90
	1,000	0.08	0.90	4.42	4.14	4.00	6.45
	3,000	0.04	0.51	2.59	2.39	4.85	5.45

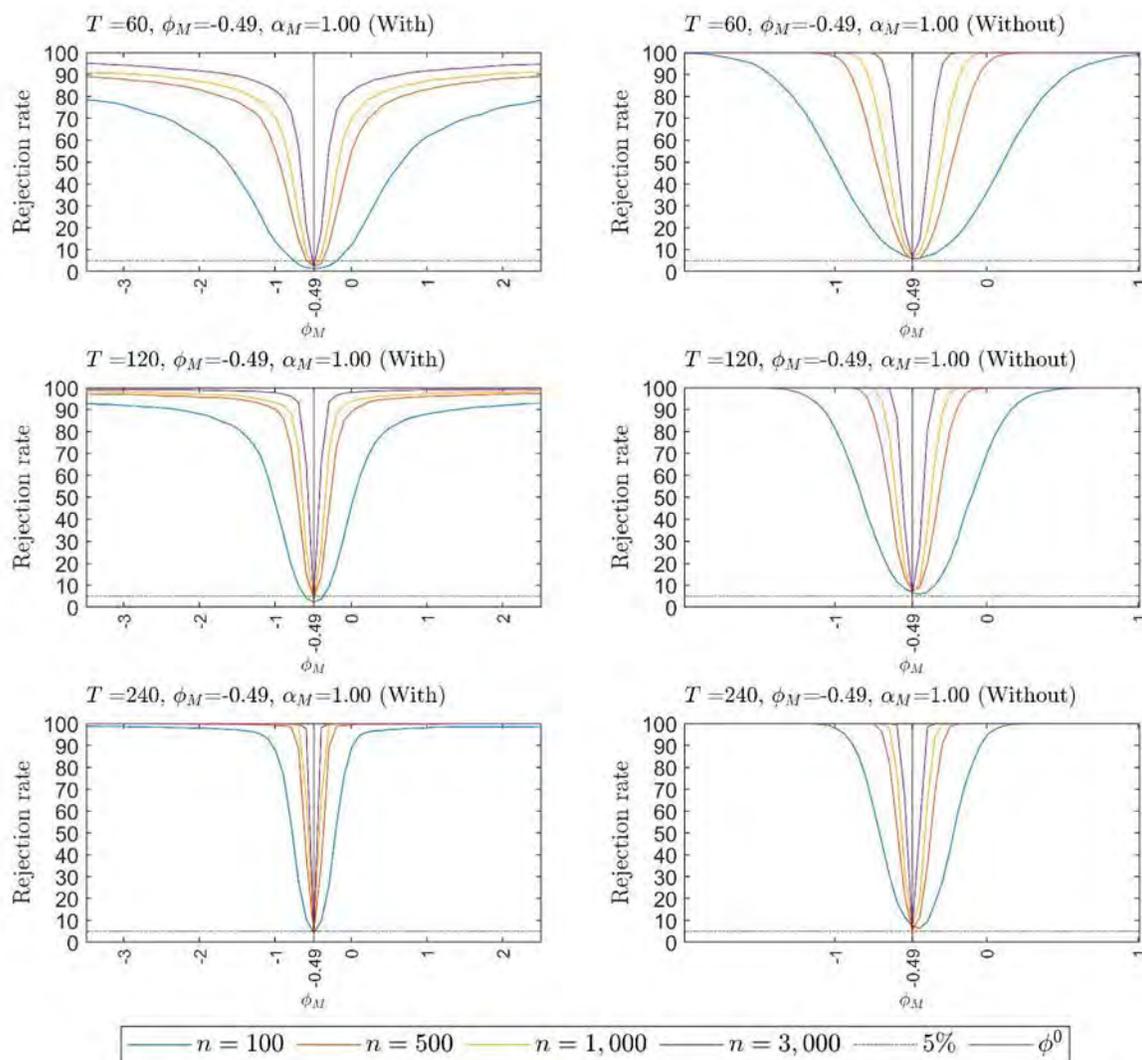
Notes: The DGP includes one strong $\alpha_M = 1$ and two weak ($\alpha_H = \alpha_S = 0.5$) factors, the regression with weak factors includes them, the regression without excludes them. For further details of the experiments, see Table S-1.

Figure S-C-E1: Empirical Power Functions, experiment 1, for coefficient of the ϕ_M factor with and without misspecification



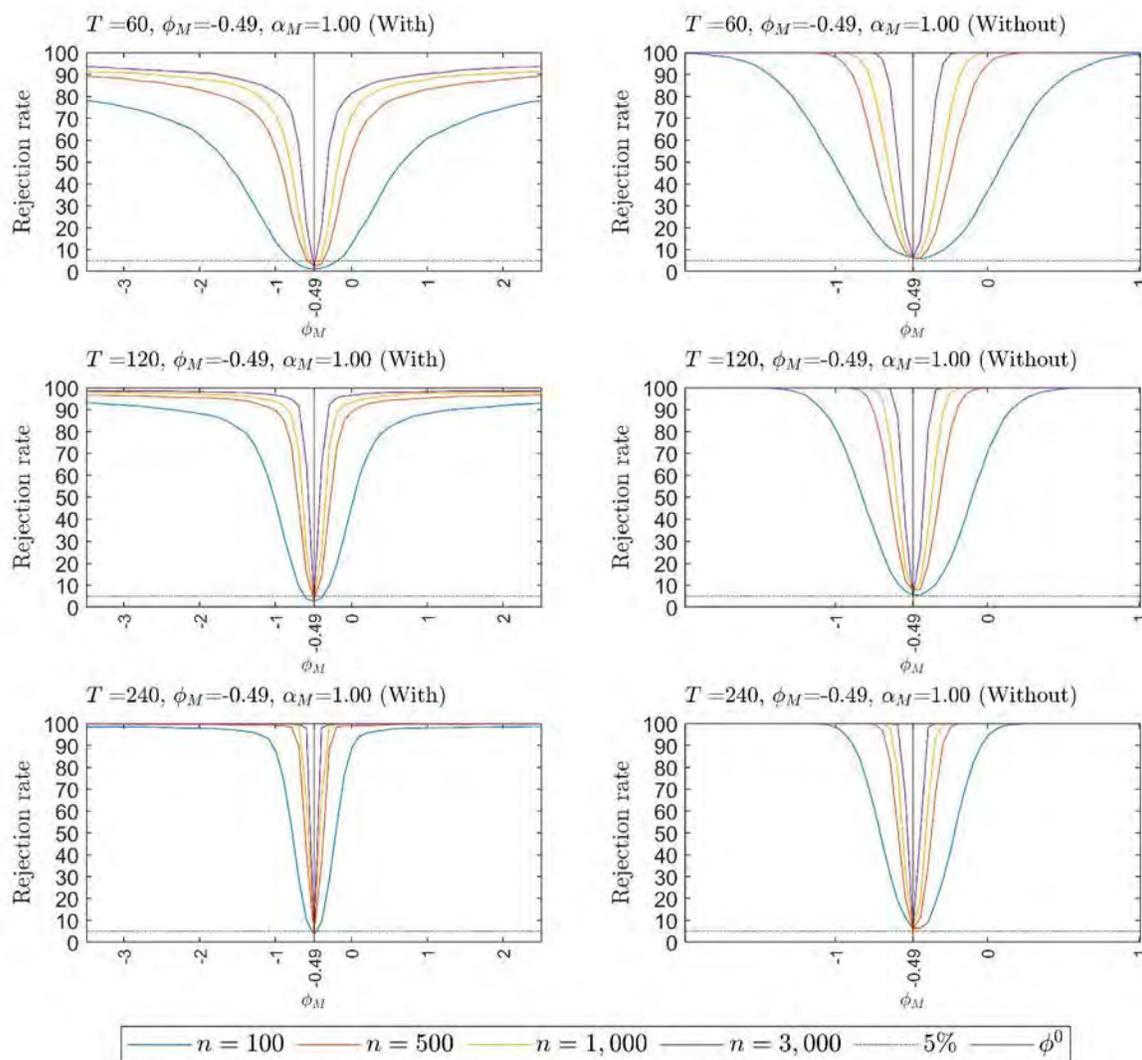
Note: See the notes to Table S-C-E1-3.

Figure S-C-E2: Empirical Power Functions, experiment 2, for coefficient of the ϕ_M factor with and without misspecification



Note: See the notes to Table S-C-E1-3.

Figure S-C-E3: Empirical Power Functions, experiment 3, for coefficient of the ϕ_M factor with and without misspecification



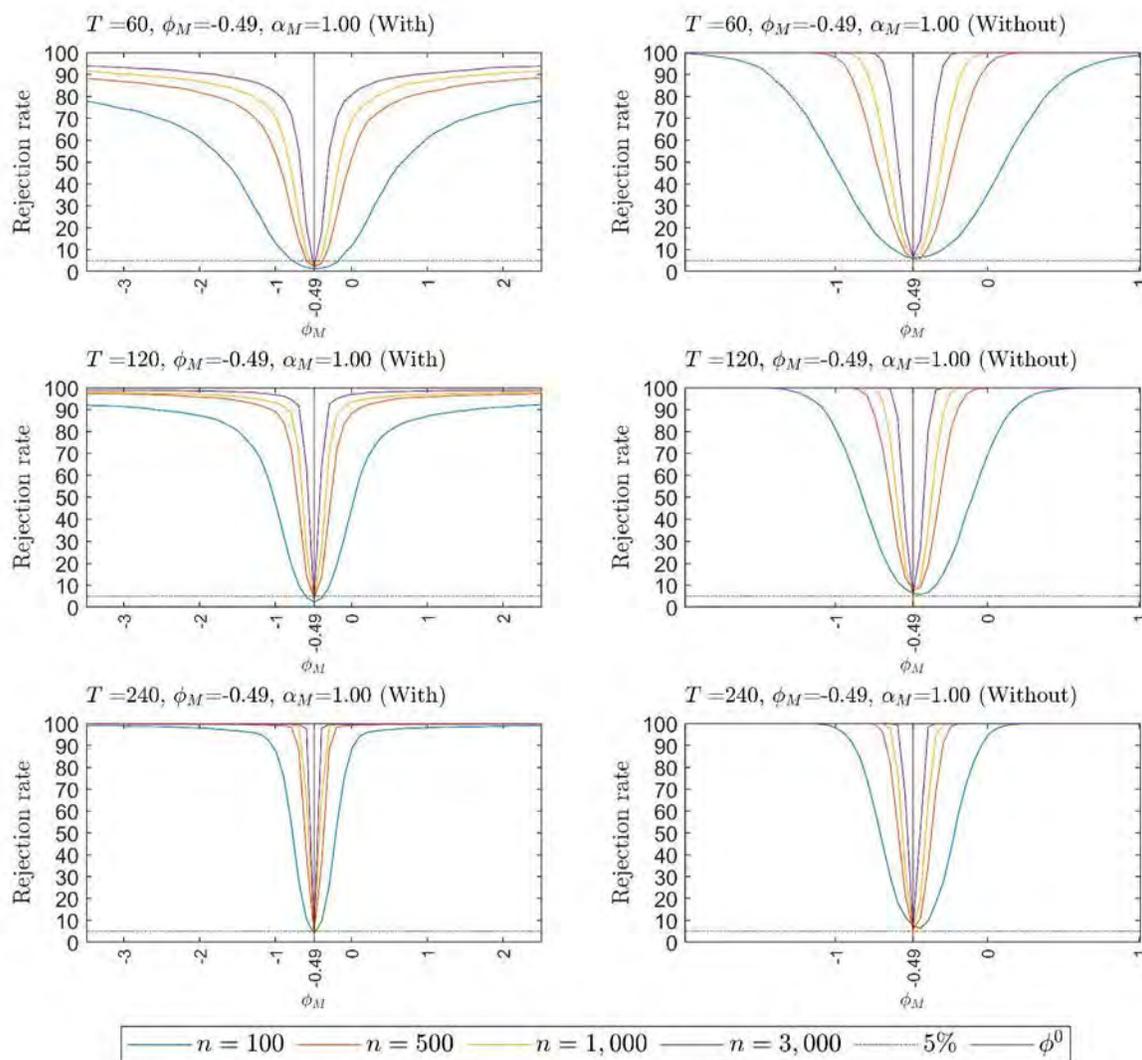
Note: See the notes to Table S-C-E1-3.

Table S-C-E4-6: Bias, RMSE and size for the bias-corrected estimators of $\phi_M = -0.49$, $\alpha_M=1$ with and without weak factors included in the regression for the cases of experiments 4, 5 and 6

Experiment 4	n	Bias(x100)		RMSE(x100)		Size(x100)	
		With	Without	With	Without	With	Without
		weak factors		weak factors		weak factors	
T = 60	100	-24.42	4.12	1165.53	31.54	1.35	6.15
	500	-11.84	1.61	302.88	13.49	1.95	5.55
	1,000	5.16	1.01	229.86	9.72	1.90	6.75
	3,000	-8.17	0.57	450.22	5.57	2.05	6.60
T = 120	100	-4.28	3.22	452.39	19.63	2.30	6.30
	500	-0.54	1.44	26.55	8.94	3.80	7.85
	1,000	0.98	0.96	31.58	6.08	2.75	5.80
	3,000	-0.17	0.45	8.84	3.46	3.25	5.40
T = 240	100	-0.11	2.94	31.04	13.51	3.95	7.55
	500	-0.50	1.28	27.63	5.85	4.05	5.90
	1,000	-0.05	0.91	5.23	4.16	4.65	6.90
	3,000	0.02	0.48	3.36	2.39	4.85	6.05
Experiment 5							
T = 60	100	12.72	3.75	679.29	30.84	0.95	6.25
	500	-10.07	1.66	364.25	13.34	2.10	6.15
	1,000	-5.01	1.07	122.63	9.60	1.45	5.90
	3,000	-4.43	0.60	133.49	5.47	1.90	6.25
T = 120	100	6.87	3.01	277.10	19.52	2.55	5.65
	500	-4.67	1.36	127.57	8.85	3.65	7.60
	1,000	-0.85	0.92	24.13	6.16	3.60	6.40
	3,000	-0.35	0.51	20.50	3.44	2.65	5.20
T = 240	100	0.54	2.89	31.90	13.51	4.15	6.35
	500	0.29	1.25	7.91	5.86	3.85	6.00
	1,000	0.10	0.89	4.44	4.16	4.30	6.75
	3,000	0.04	0.52	2.58	2.40	5.20	5.45
Experiment 6							
T = 60	100	-27.78	4.12	905.91	31.54	1.45	6.15
	500	-11.62	1.60	305.30	13.55	2.05	5.95
	1,000	5.28	0.99	231.07	9.73	1.65	6.50
	3,000	-7.16	0.58	412.98	5.59	2.05	6.95
T = 120	100	-2.66	3.22	451.26	19.63	2.30	6.30
	500	-0.54	1.44	26.52	9.00	4.15	7.85
	1,000	0.92	0.94	30.26	6.10	2.90	6.30
	3,000	-0.18	0.46	8.91	3.47	3.25	5.55
T = 240	100	-0.18	2.94	29.74	13.51	5.10	7.55
	500	-0.49	1.29	27.08	5.89	4.40	6.30
	1,000	-0.02	0.90	5.12	4.17	5.20	6.80
	3,000	0.03	0.49	3.47	2.40	4.50	6.45

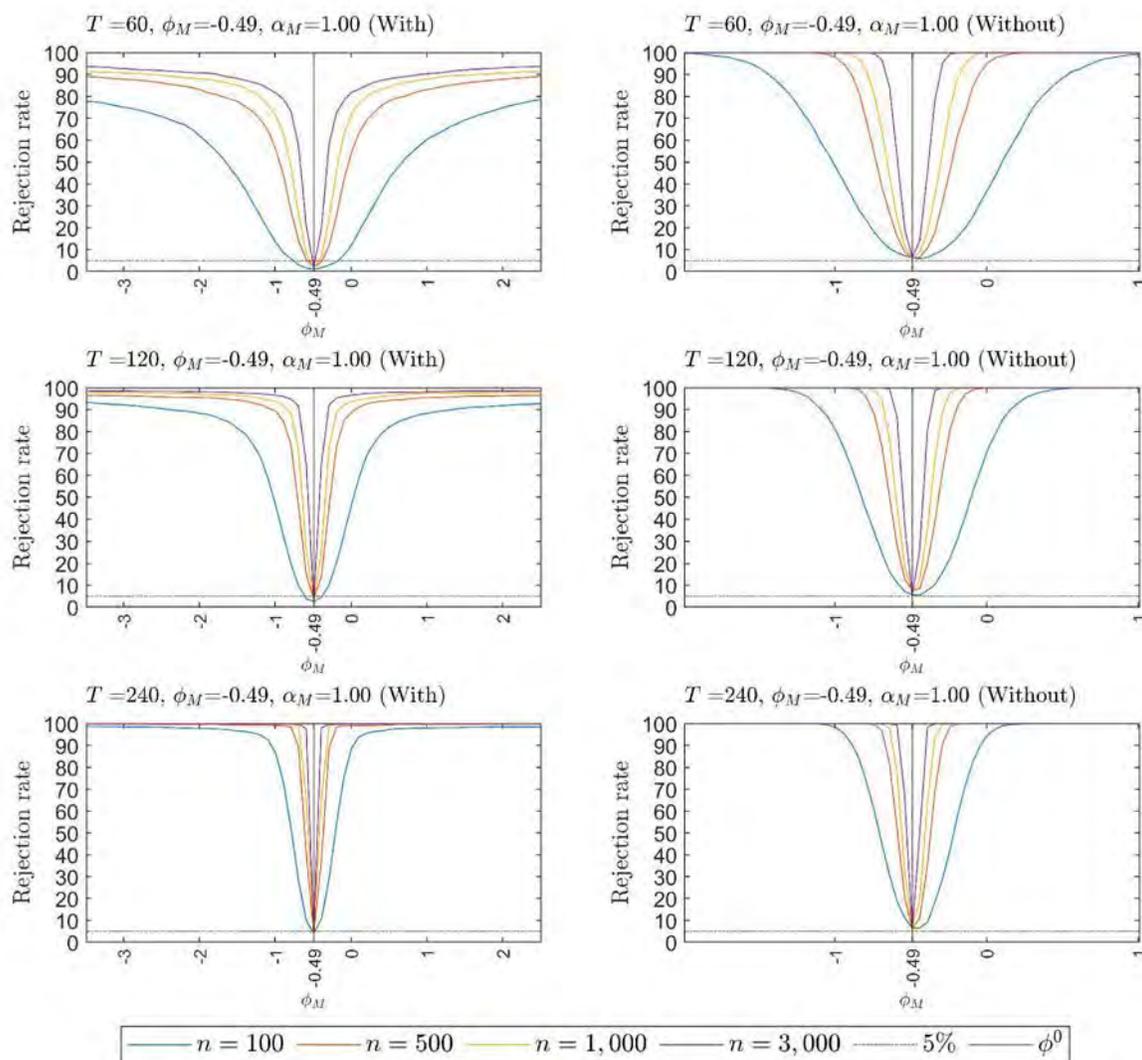
Notes: The DGP includes one strong $\alpha_M = 1$ and two weak ($\alpha_H = \alpha_S = 0.5$) factors, the regression with weak factors includes them, the regression without excludes them. For further details of the experiments, see Table S-1.

Figure S-C-E4: Empirical Power Functions, experiment 4, for coefficient of the ϕ_M factor with and without misspecification



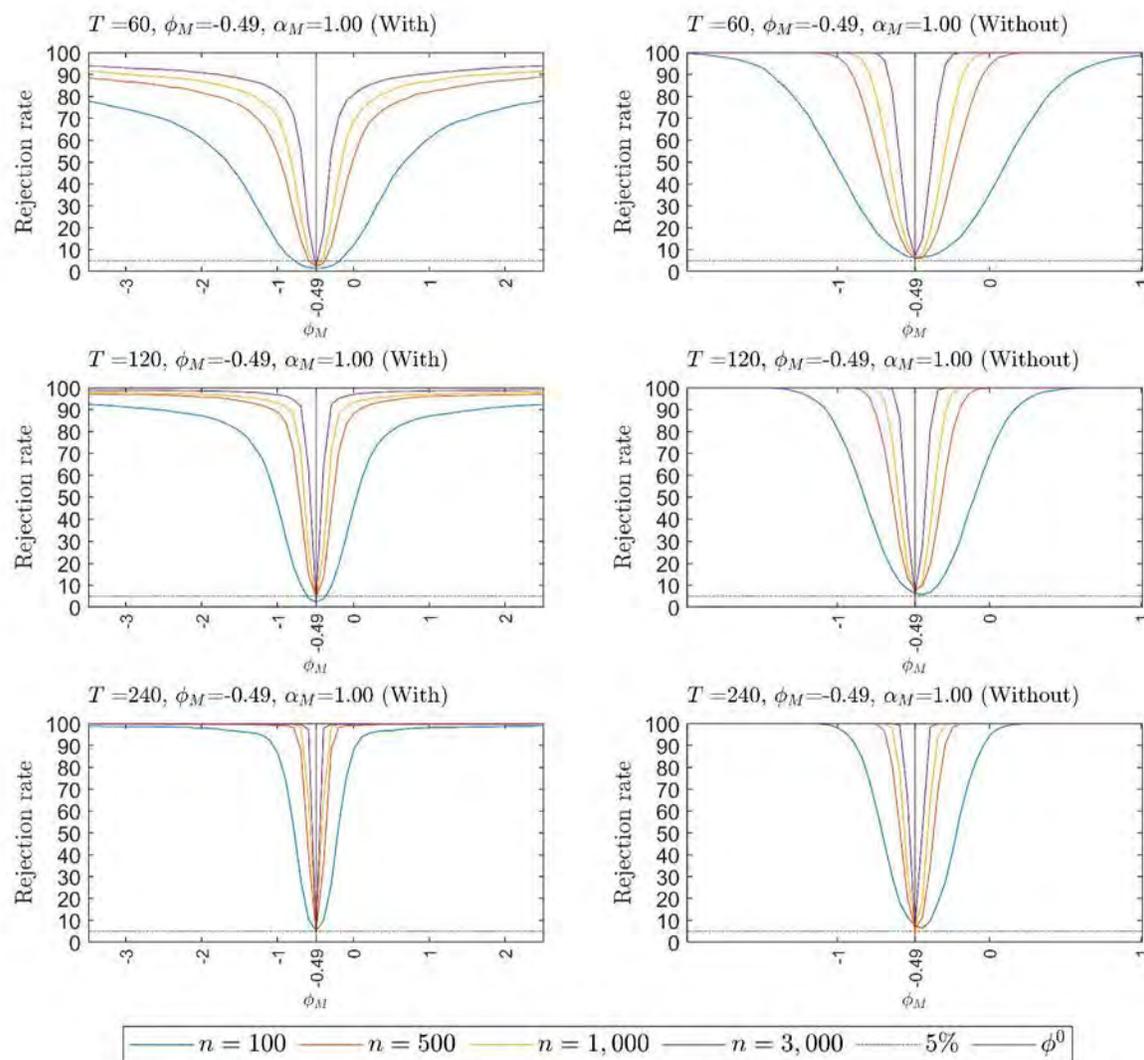
Note: See the notes to Table S-C-E4-6.

Figure S-C-E5: Empirical Power Functions, experiment 5, for coefficient of the ϕ_M factor with and without misspecification



Note: See the notes to Table S-C-E4-6.

Figure S-C-E6: Empirical Power Functions, experiment 6, for coefficient of the ϕ_M factor with and without misspecification



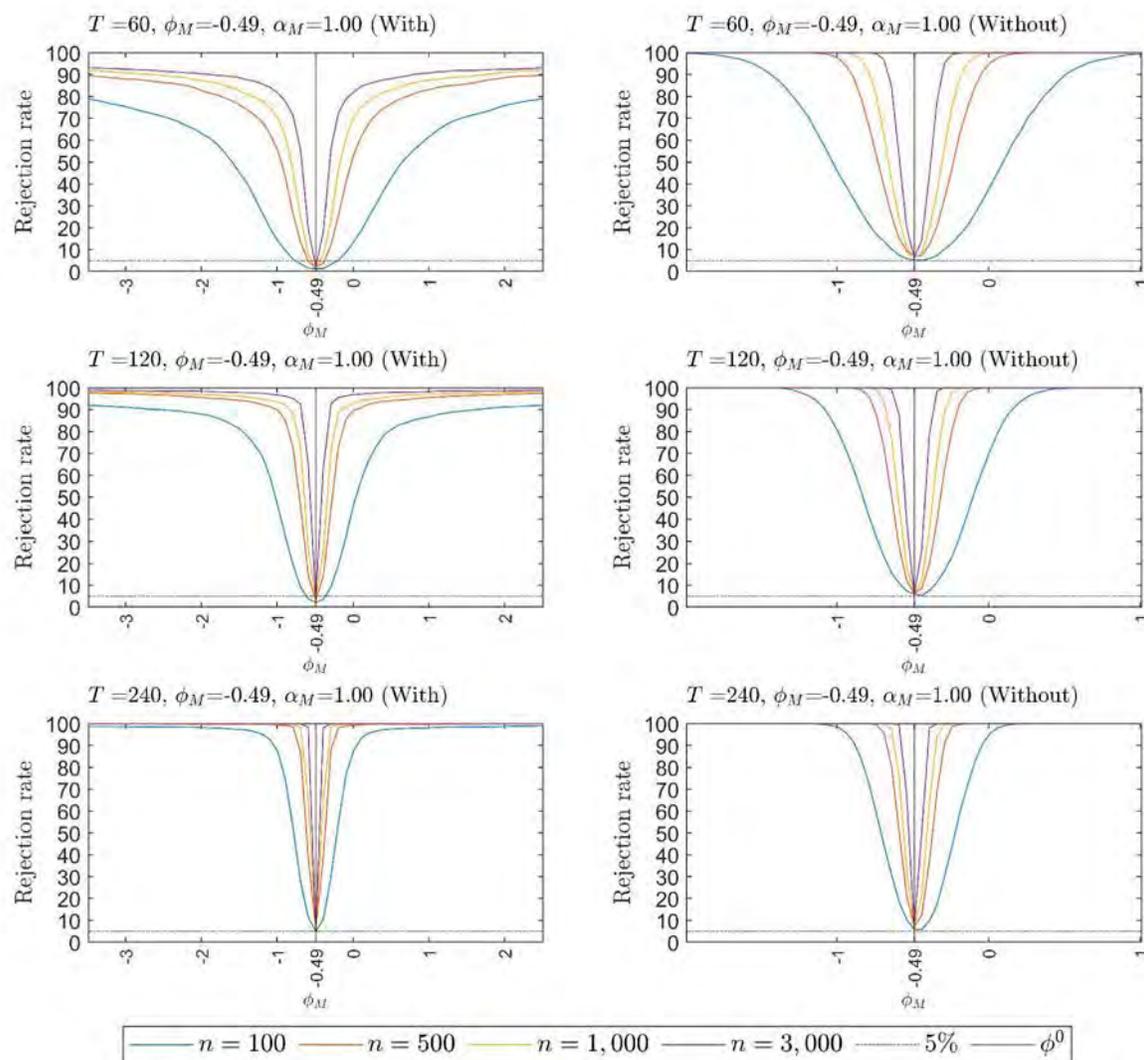
Note: See the notes to Table S-C-E4-6.

Table S-C-E7-9: Bias, RMSE and size for the bias-corrected estimators of $\phi_M = -0.49$, $\alpha_M=1$ with and without weak factors included in the regression for the cases of experiments 7, 8 and 9

Experiment 7	n	Bias(x100)		RMSE(x100)		Size(x100)	
		With weak factors	Without	With weak factors	Without	With weak factors	Without
T = 60	100	3.08	2.82	859.39	30.18	1.30	5.35
	500	24.88	1.20	840.21	13.62	1.85	6.65
	1,000	0.24	0.84	79.72	9.64	1.35	6.25
	3,000	-4.91	0.61	528.52	5.51	2.20	6.40
T = 120	100	4.42	2.67	216.43	19.15	2.10	6.00
	500	1.74	1.23	147.10	8.59	2.95	6.05
	1,000	-3.59	0.89	154.11	6.12	3.75	5.90
	3,000	-0.84	0.61	41.40	3.52	3.75	5.75
T = 240	100	-0.94	2.74	22.50	13.27	4.95	5.80
	500	-0.83	1.11	38.11	6.04	6.15	7.15
	1,000	-0.06	0.84	4.85	4.25	4.95	7.00
	3,000	-0.01	0.55	2.46	2.38	4.70	6.50
Experiment 8							
T = 60	100	-18.57	2.85	1090.60	31.04	1.30	5.70
	500	3.08	0.84	248.52	13.70	1.90	6.10
	1,000	11.56	0.98	511.88	9.66	1.60	5.60
	3,000	-0.24	0.71	45.00	5.66	2.05	6.50
T = 120	100	15.36	2.83	488.96	19.33	2.40	5.00
	500	-8.41	1.12	424.57	8.53	3.55	5.70
	1,000	-1.21	1.04	42.58	6.25	3.65	6.25
	3,000	-0.68	0.64	43.26	3.56	3.60	6.00
T = 240	100	-4.70	2.85	184.48	13.30	5.10	6.30
	500	0.07	1.17	7.15	5.98	5.45	6.30
	1,000	-0.07	0.85	4.44	4.27	5.10	6.60
	3,000	0.01	0.54	2.41	2.40	4.80	6.70
Experiment 9							
T = 60	100	16.90	1.71	1051.65	32.31	1.35	6.65
	500	-10.51	0.95	739.68	14.35	2.35	7.50
	1,000	-2.96	0.53	160.31	10.26	2.65	8.75
	3,000	-0.09	0.43	106.32	5.80	2.70	8.35
T = 120	100	61.02	1.74	2471.10	20.37	2.05	6.25
	500	-9.78	1.08	365.78	8.87	2.70	6.00
	1,000	0.35	0.73	77.55	6.36	3.00	7.30
	3,000	1.23	0.46	48.78	3.59	3.85	6.55
T = 240	100	0.51	1.73	52.46	13.89	4.55	6.90
	500	-0.22	0.91	8.62	6.08	5.75	6.40
	1,000	-0.20	0.71	7.44	4.27	5.50	7.15
	3,000	-0.04	0.43	2.52	2.37	4.30	5.60

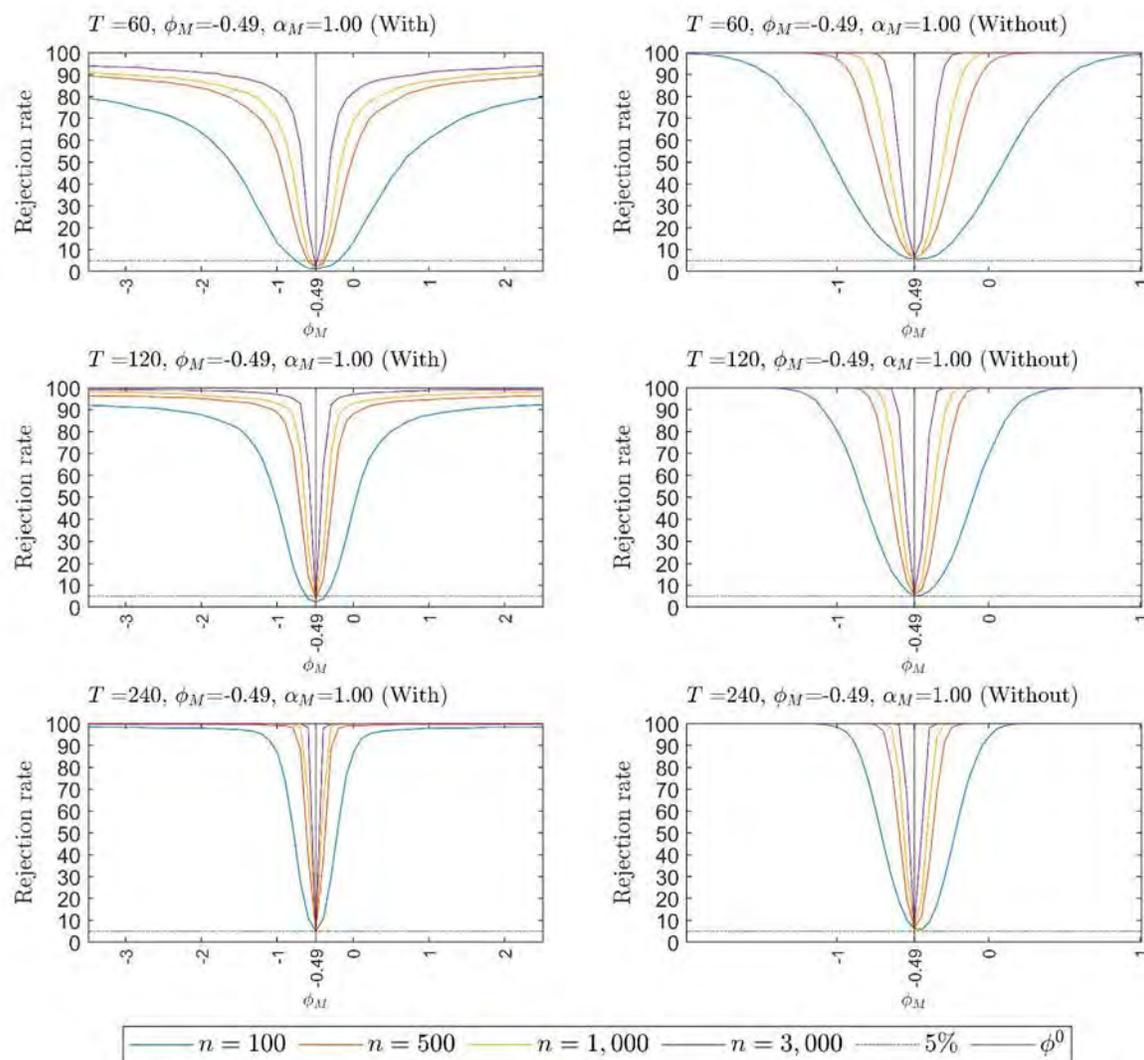
Notes: The DGP includes one strong $\alpha_M = 1$ and two weak ($\alpha_H = \alpha_S = 0.5$) factors, the regression with weak factors includes them, the regression without excludes them. For further details of the experiments, see Table S-1.

Figure S-C-E7: Empirical Power Functions, experiment 7, for coefficient of the ϕ_M factor with and without misspecification



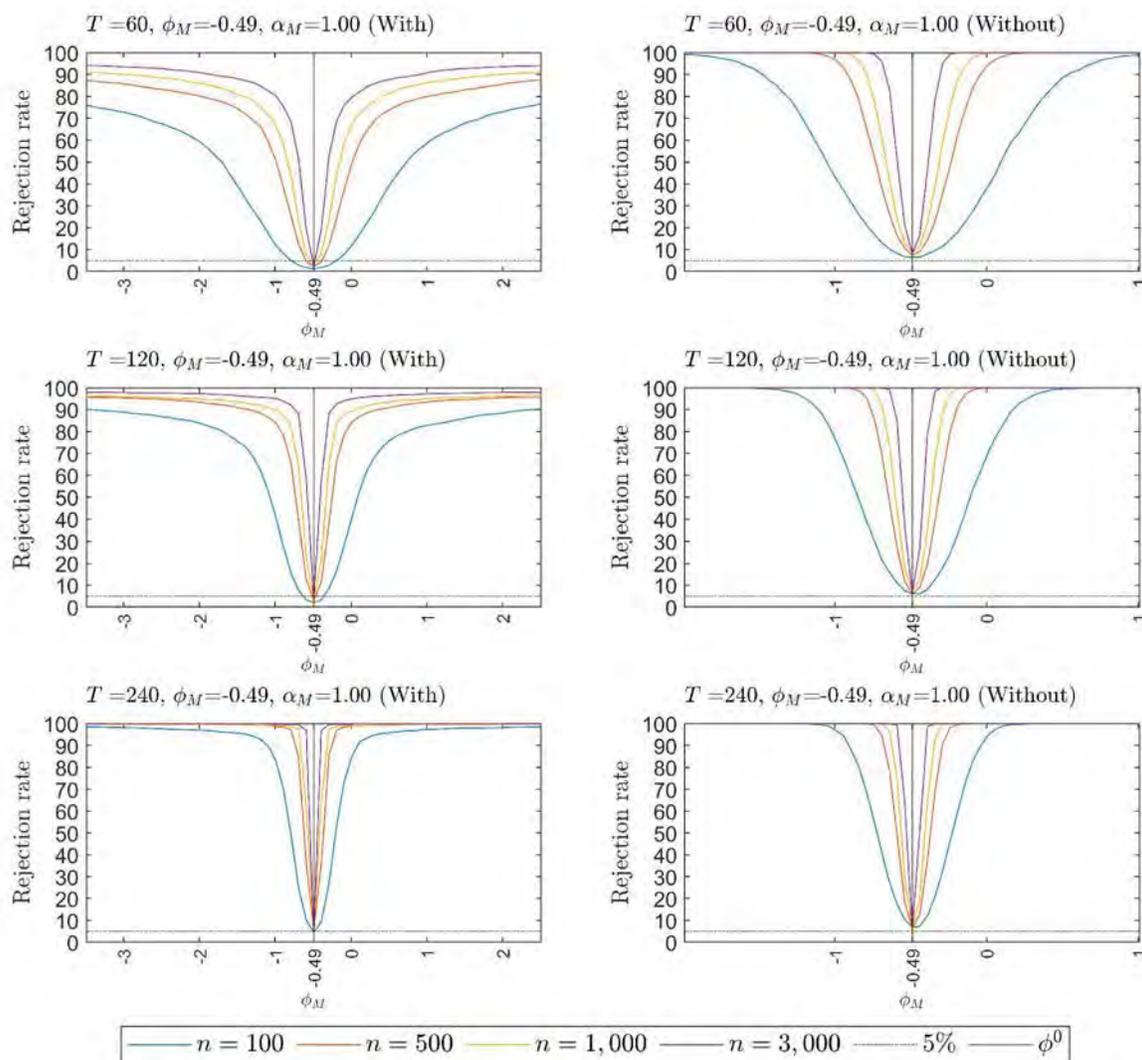
Note: See the notes to Table S-C-E7-9.

Figure S-C-E8: Empirical Power Functions, experiment 8, for coefficient of the ϕ_M factor with and without misspecification



Note: See the notes to Table S-C-E7-9.

Figure S-C-E9: Empirical Power Functions, experiment 9, for coefficient of the ϕ_M factor with and without misspecification



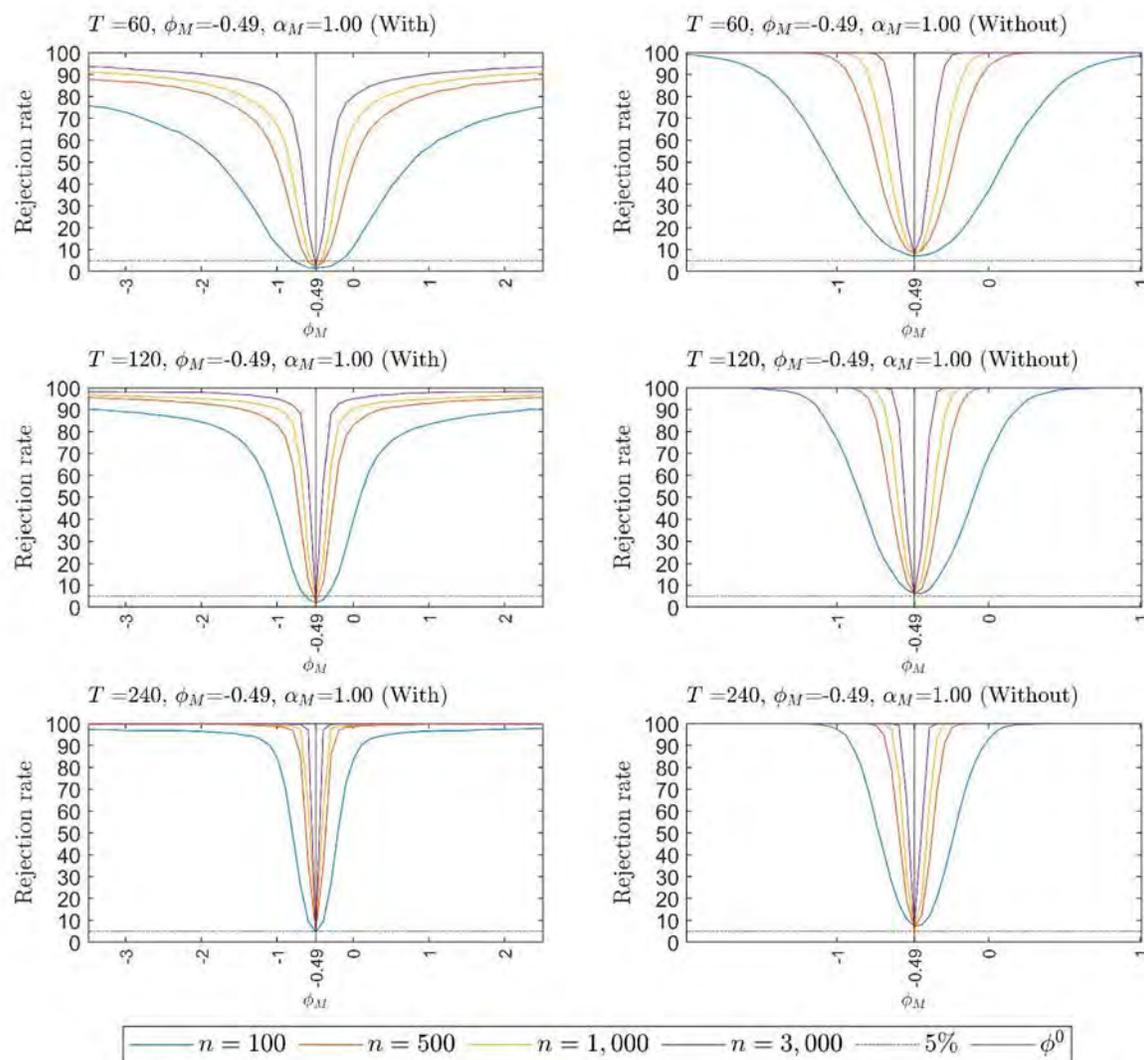
Note: See the notes to Table S-C-E7-9.

Table S-C-E10-12: Bias, RMSE and size for the bias-corrected estimators of $\phi_M = -0.49$, $\alpha_M=1$ with and without weak factors included in the regression for the cases of experiments 10, 11 and 12

Experiment	n	Bias(x100)		RMSE(x100)		Size(x100)		
		With weak factors	Without	With weak factors	Without	With weak factors	Without	
Experiment 10	T = 60	100	-1.30	1.80	581.15	33.08	1.55	6.95
		500	19.94	0.54	738.70	14.45	2.10	8.20
		1,000	1.31	0.69	76.46	10.23	1.85	7.35
		3,000	1.02	0.53	333.48	5.91	2.70	8.30
	T = 120	100	27.62	1.91	810.64	20.78	1.90	6.30
		500	-4.30	0.95	195.46	8.90	2.40	6.25
		1,000	7.68	0.93	212.90	6.46	3.45	7.25
		3,000	-0.46	0.49	34.61	3.60	3.20	5.50
	T = 240	100	0.01	1.87	109.19	14.08	4.60	7.30
		500	0.27	0.94	19.59	6.08	5.20	6.50
		1,000	0.41	0.76	21.39	4.28	5.50	6.55
		3,000	-0.04	0.76	2.57	2.40	4.70	7.10
Experiment 11	T = 60	100	17.29	2.89	696.53	30.28	1.45	5.40
		500	-5.92	1.27	510.19	13.71	2.00	6.80
		1,000	45.15	0.78	2091.69	9.66	2.00	6.45
		3,000	2.77	0.59	75.60	5.61	2.65	7.05
	T = 120	100	4.02	2.73	135.30	19.16	2.55	6.20
		500	0.44	1.34	24.35	8.61	3.50	5.60
		1,000	10.16	0.86	415.98	6.09	3.85	6.10
		3,000	0.10	0.62	7.88	3.54	4.65	6.70
	T = 240	100	-0.31	2.77	14.31	13.28	4.10	6.15
		500	0.07	1.19	6.78	6.06	5.45	6.90
		1,000	-0.06	0.81	4.40	4.23	5.45	6.65
		3,000	0.01	0.55	2.37	2.38	4.65	6.50
Experiment 12	T = 60	100	0.33	2.93	531.55	31.13	1.35	5.75
		500	-16.77	0.90	496.29	13.75	1.70	5.60
		1,000	17.22	0.93	766.48	9.68	1.65	6.15
		3,000	-0.91	0.68	162.80	5.75	2.40	7.00
	T = 120	100	-11.47	2.88	529.74	19.36	2.55	5.15
		500	-16.06	1.23	555.75	8.55	3.55	5.80
		1,000	0.99	1.03	20.02	6.22	3.50	6.50
		3,000	-2.21	0.64	100.80	3.57	4.00	5.90
	T = 240	100	-0.27	2.88	19.39	13.31	4.20	6.30
		500	0.10	1.24	6.30	5.99	5.35	6.25
		1,000	-0.05	0.83	4.37	4.23	5.10	6.70
		3,000	0.01	0.54	2.39	2.41	4.95	6.50

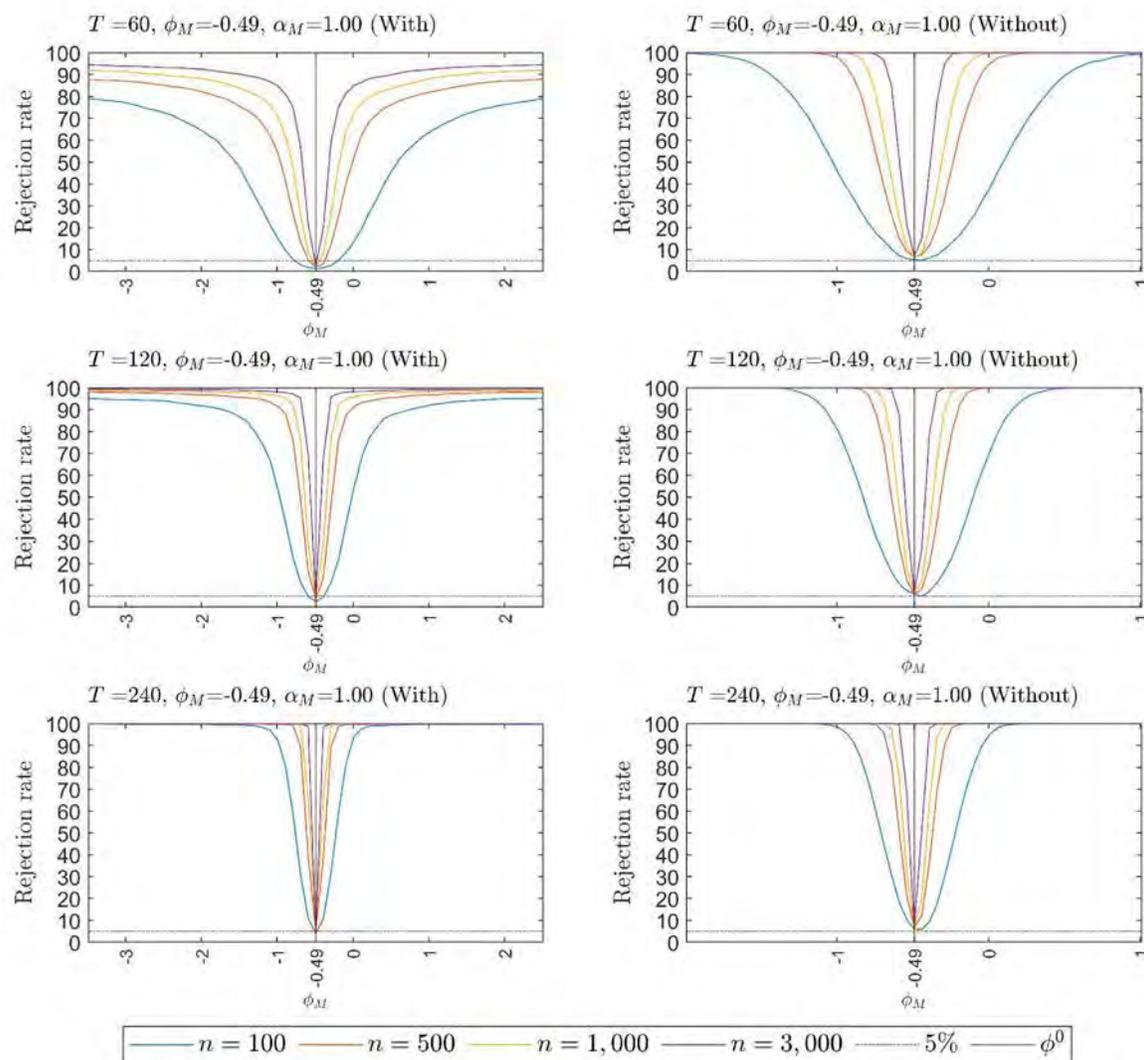
Notes: The DGP includes one strong $\alpha_M = 1$ and two weak ($\alpha_H = \alpha_S = 0.5$) factors, the regression with weak factors includes them, the regression without excludes them. For further details of the experiments, see Table S-1.

Figure S-C-E10: Empirical Power Functions, experiment 10, for coefficient of the ϕ_M factor with and without misspecification



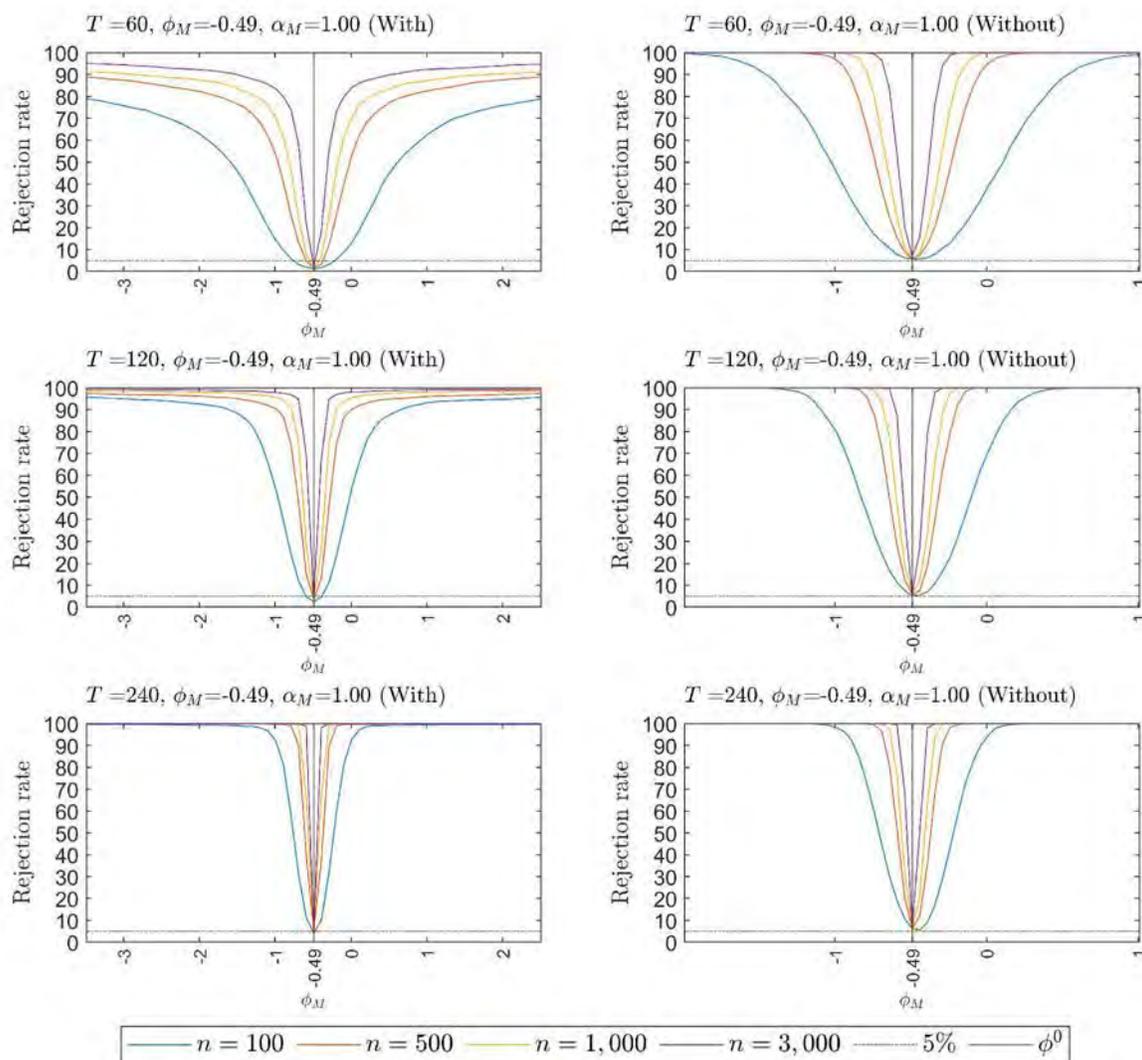
Note: See the notes to Table S-C-E10-12.

Figure S-C-E11: Empirical Power Functions, experiment 11, for coefficient of the ϕ_M factor with and without misspecification



Note: See the notes to Table S-C-E10-12.

Figure S-C-E12: Empirical Power Functions, experiment 12, for coefficient of the ϕ_M factor with and without misspecification



Note: See the notes to Table S-C-E10-12.