Bootstrapping factor models with cross sectional dependence

Sílvia Gonçalves and Benoit Perron

University of Western Ontario and Université de Montréal, CIREQ, CIRANO

April 4, 2016

Abstract

We consider bootstrap methods for factor-augmented regressions with cross sectional dependence among idiosyncratic errors. This is important to capture the bias of the OLS estimator derived recently by Gonçalves and Perron (2014). We first show that a common approach of resampling cross sectional vectors over time is invalid in this context because it induces a zero bias. We then propose the cross-sectional dependent (CSD) bootstrap where bootstrap samples are obtained by taking a random vector and multiplying it by the square root of a consistent estimator of the covariance matrix of the idiosyncratic errors. We show that if the covariance matrix estimator is consistent in the spectral norm, then the CSD bootstrap is consistent, and we verify this condition for the thresholding estimator of Bickel and Levina (2008). Finally, we apply our new bootstrap procedure to forecasting inflation using convenience yields as recently explored by Gospodinov and Ng (2013).

Keywords: factor model, bootstrap, asymptotic bias.

1 Introduction

Factor-augmented regressions, involving factors estimated from a large panel data set, have become popular in empirical macroeconomics. Recent applications include forecasting with diffusion indices (Stock and Watson, 2002), predicting excess stock returns (Ludvigson and Ng, 2007, Neely, Rapach, Tu and Zhou, 2015), predicting bond yields (Ludvigson and Ng, 2009 and 2011), modeling commodity convenience yields (Gospodinov and Ng, 2013), and analyzing spillovers across banks using credit default swap spreads (Eichengreen et al. 2012).

Inference in these models is complicated by the need to account for the preliminary estimation of the factors. Bai and Ng (2006) derived the asymptotic distribution of the ordinary least squares (OLS) estimator in this context and showed that one can neglect the effect of factor estimation uncertainty if $\sqrt{T}/N \to 0$ where N is the cross sectional dimension of the panel from which factors are extracted and T is the time series dimension. Recently, Gonçalves and Perron (2014) have relaxed this condition and shown that a bias appears in the asymptotic distribution when $\sqrt{T}/N \to c$ and $c \neq 0$. They proposed a wild bootstrap method that removes this bias and outperforms the asymptotic approach of Bai and Ng (2006).

The expression for the bias obtained by Gonçalves and Perron (2014) depends, among other things, on the cross sectional dependence of the idiosyncratic errors in the factor model. Unfortunately, the wild bootstrap method that they proposed destroys cross sectional dependence and is only valid when no cross sectional dependence among idiosyncratic errors is present.

This paper analyzes the issue of bootstrapping with general cross sectional dependence among the idiosyncratic errors. Proposing a bootstrap method that is robust to cross sectional dependence without making parametric assumptions is a much harder task than for time series dependence. The main reason is that, contrary to the time dimension, no natural ordering among the variables needs to exist. This makes it harder to apply blocking methods, which are often used to capture time series dependence of unknown form.

We make two important contributions. First, we show that a common approach of resampling vectors containing all the cross sectional variables only in the time series dimension as a way to preserve cross sectional dependence, while valid in other contexts (see for example Gonçalves (2011) for inference in a linear panel data model with fixed effects), is invalid in this context. The reason is that it induces a zero bias in the bootstrap asymptotic distribution by not reproducing the uncertainty associated with the estimation of the factors. In particular, resampling only in the time series dimension implies that the bootstrap variance of the cross sectional average of the estimated panel factor scores is equal to the empirical time series variance of this cross sectional average, which is zero by the first order conditions that define the principal components estimator of the factors and the factor loadings.

Our second contribution is to propose a solution, which we call the cross sectional dependent (CSD) bootstrap, where bootstrap samples are obtained by taking a random vector of dimension $N \times 1$ with mean 0 and identity covariance matrix and multiplying it by the square root of a consistent estimator of the covariance matrix of the idiosyncratic errors. We show that if the covariance matrix estimator is consistent in the spectral norm, then the CSD bootstrap is consistent, and we verify this condition for the thresholding estimator of Bickel and Levina (2008). Other covariance matrix estimators could be used to implement the cross sectional dependent bootstrap, and one could check that our general sufficient conditions are satisfied to ensure validity of the bootstrap.

We apply our new bootstrap procedure to forecasting inflation using convenience yields as recently explored by Gospodinov and Ng (2013). We find that our intervals are tighter than those reported in Gospodinov and Ng and that they are shifted to account for the presence of bias. We also see a difference in the center of these intervals relative to the wild bootstrap of Gonçalves and Perron (2014) because of the effect of cross sectional dependence on the bias.

The remainder of the paper is organized as follows. Section 2 introduces our model and shows that bootstrap methods that have been used in this context will not replicate factor estimation uncertainty. Section 3 presents our solution and a set of high level conditions under which the bootstrap is valid, and we check these conditions for the thresholding estimator. Section 4 presents our simulation experiments, while Section 5 presents our empirical illustration to forecasting inflation with convenience yields. Finally, Section 6 concludes. We also provide two appendices: our assumptions are in Appendix A and mathematical proofs appear in Appendix B.

For any matrix A, we let $\rho(A) = \max_{\|x\|=1} \|Ax\|$ denote the operator (or spectral) norm of A, where $\|Ax\| = (x'A'Ax)^{1/2}$ is the Euclidean vector norm of the vector Ax. When A is symmetric, $\rho(A)$ is equal to the maximum eigenvalue of A, in absolute value. Similarly, we let $\|A\|$ denote its Frobenius norm defined as $\|A\| = (trace\,(A'A))^{1/2}$.

2 Why cross sectional dependence matters for inference

2.1 Setup and review of existing results

We consider the following regression model

$$y_{t+1} = \alpha' F_t + \beta' W_t + \varepsilon_{t+1},\ t = 1, \ldots, T-1. \tag{1}$$

The q observed regressors, typically a constant and lags of y_t, are contained in W_t. The r unobserved regressors F_t are the common factors in the following panel factor model,

$$X_{it} = \lambda_i' F_t + e_{it}, \quad i = 1, \ldots, N,\ t = 1, \ldots, T, \tag{2}$$

where the $r \times 1$ vector λ_i contains the factor loadings and e_{it} is an idiosyncratic error term. In matrix form, we can write (2) as

$$X = F\Lambda' + e,$$

where X is an observed data matrix of size $T \times N$, $F = (F_1, \ldots, F_T)'$ is a $T \times r$ matrix of random factors, with r the number of common factors, $\Lambda = (\lambda_1, \ldots, \lambda_N)'$ is a fixed $N \times r$ matrix, and e is $T \times N$. Throughout, we consider the number of factors r as given. Forecasting horizons greater than 1 could be considered, but this would generally entail serial correlation in the regression errors. Because our focus is on bootstrap inference under cross sectional dependence, we focus on one-step ahead forecast horizons only.

Estimation proceeds in two steps. Given X, we estimate F and Λ with the method of principal components. In particular, F is estimated with the $T \times r$ matrix $\tilde{F} = \begin{pmatrix} \tilde{F}_1 & \ldots & \tilde{F}_T \end{pmatrix}'$ composed of $\sqrt{T}$ times the eigenvectors corresponding to the r largest eigenvalues of of XX'/TN (arranged in decreasing order), where the normalization $\frac{\tilde{F}'\tilde{F}}{T} = I_r$ is used. The matrix containing the estimated loadings is then $\tilde{\Lambda} = \left(\tilde{\lambda}_1, \ldots, \tilde{\lambda}_N\right)' = X'\tilde{F}\left(\tilde{F}'\tilde{F}\right)^{-1} = X'\tilde{F}/T$. As is well known in this literature, the principal components $\tilde{F}_t$ can only consistently estimate a transformation of the true factors F_t, given by HF_t, where H is a rotation matrix defined as

$$H = \tilde{V}^{-1} \frac{\tilde{F}'F}{T} \frac{\Lambda'\Lambda}{N}, \tag{3}$$

where $\tilde{V}$ is the $r \times r$ diagonal matrix containing on the main diagonal the r largest eigenvalues of XX'/NT, in decreasing order, see Bai (2003). In the following we let $H_0 \equiv p\lim H$ and $V \equiv p\lim \tilde{V}$. In the second step, we run an OLS regression of y_{t+1} on $\hat{z}_t = \begin{pmatrix} \tilde{F}_t' & W_t' \end{pmatrix}'$ and obtain $\hat{\delta} = \left(\hat{\alpha}', \hat{\beta}'\right)'$.

The asymptotic properties of $\hat{\delta}$ as well as those of the corresponding prediction intervals were studied by Bai and Ng (2006) under standard assumptions in this literature that allow for weak cross sectional and serial dependence in the idiosyncratic error term (cf. Assumptions 1-5 in Appendix A, which are similar to the assumptions in Bai (2003) and Bai and Ng (2006)). In particular, Bai and Ng (2006) showed that when $\sqrt{T}/N \to 0$, $\hat{\delta}$ is asymptotically distributed as a normal random vector with mean 0 and covariance matrix $\Sigma_\delta = (\Phi_0')^{-1}\Sigma_{zz}^{-1}\Omega\Sigma_{zz}^{-1}\Phi_0^{-1}$, where $\Phi_0 \equiv diag(H_0, I_q)$, $\Sigma_{zz} = p\lim \frac{1}{T}\sum_{t=1}^{T} z_t z_t'$, with $z_t = (F_t', W_t')'$, and $\Omega = \lim_{T\to\infty} Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1} z_t \varepsilon_{t+1}\right)$. This covariance matrix is of the usual sandwich form and does not reflect the added factors estimation uncertainty caused by replacing the true latent factors by their estimates, implying that cross sectional dependence in e_{it} is asymptotically irrelevant when $\sqrt{T}/N \to 0$.

More recently, Gonçalves and Perron (2014) showed that if instead $\sqrt{T}/N \to c \neq 0$, then

$$\sqrt{T}\left(\hat{\delta} - \delta\right) \to^d N\left(-c\Delta_\delta, \Sigma_\delta\right),$$

where $\delta \equiv \left(\begin{array}{cc} \alpha' H^{-1} & \beta' \end{array}\right)'$ and Δ_δ is a bias term that depends on the cross sectional dependence of e_{it}. Specifically, Δ_δ is given by

$$\Delta_\delta = \left(p\lim \frac{1}{T}\sum_{t=1}^{T} \hat{z}_t \hat{z}_t'\right)^{-1} \left(\begin{array}{c} \Sigma_{\tilde{F}} + V\Sigma_{\tilde{F}}V^{-1} \\ \Sigma_{W\tilde{F}}V\Sigma_{\tilde{F}}V^{-1} \end{array}\right) p\lim\left(\hat{\alpha}\right),$$

where $p\lim(\hat{\alpha}) = H_0^{-1\prime}\alpha$, $\Sigma_{W\tilde{F}} \equiv p\lim\left(\frac{W'\tilde{F}}{T}\right)$, and

$$\Sigma_{\tilde{F}} = V^{-1}Q\Gamma Q'V^{-1} = \lim_{N,T\to\infty} \frac{1}{T}\sum_{t=1}^{T} V^{-1}Q\Gamma_t Q'V^{-1},$$

where $Q = H_0'^{-1}$ and $\Gamma = \lim_{N,T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\Gamma_t$ (see Theorem 2.1 of Gonçalves and Perron (2014)). From Bai (2003), for a given t, $V^{-1}Q\Gamma_t Q'V^{-1}$ is the asymptotic variance-covariance matrix of $\sqrt{N}\left(\tilde{F}_t - HF_t\right)$ and therefore we can interpret $\Sigma_{\tilde{F}}$ as the time average of the asymptotic variance-covariance matrix of the factors estimation error. Since

$$\Gamma_t \equiv Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it}\right) = Var\left(\frac{\Lambda' e_t}{\sqrt{N}}\right)$$

is a function of the cross sectional dependence of e_{it}, inference on δ requires that we account for idiosyncratic error cross sectional dependence when $\sqrt{T}/N \to c \neq 0$. One approach is to rely on the asymptotic normal approximation together with consistent estimators of the bias term Δ_δ and the covariance matrix Σ_δ. Ludvigson and Ng (2009) proposed such an approach building on the cross sectional HAC estimator of Γ_t proposed by Bai and Ng (2006).

Another approach is to use the bootstrap. In particular, Gonçalves and Perron (2014) proposed a general residual-based bootstrap method that requires resampling the residuals of the regression model $\hat{\varepsilon}_{t+1}$ and those of the panel factor model $\tilde{e}_{it}$. Specifically, let $\left\{e_t^* = (e_{1t}^*, \ldots, e_{Nt}^*)'\right\}$ denote a bootstrap sample from $\left\{\tilde{e}_t = X_t - \tilde{\Lambda}\tilde{F}_t\right\}$ and $\left\{\varepsilon_{t+1}^*\right\}$ a bootstrap sample from $\left\{\hat{\varepsilon}_{t+1} = y_{t+1} - \hat{\alpha}'\tilde{F}_t - \hat{\beta}'W_t\right\}$. The bootstrap DGP for $\left(y_{t+1}^*, X_t^{*\prime}\right)$ is given by

$$\begin{aligned} y_{t+1}^* &= \hat{\alpha}'\tilde{F}_t + \hat{\beta}'W_t + \varepsilon_{t+1}^*, \quad t = 1, \ldots, T-1, \\ X_t^* &= \tilde{\Lambda}\tilde{F}_t + e_t^*, \quad t = 1, \ldots, T. \end{aligned}$$

Estimation in the bootstrap proceeds in two stages as in the sample. First, we estimate the factors by the method of principal components using the bootstrap panel data set $\{X_t^*\}$. Second, we run a regression of y_{t+1}^* on the bootstrap estimated factors and on the fixed observed regressors W_t. Let $\hat{\delta}^*$ denote the bootstrap OLS estimator.

Gonçalves and Perron (2014) showed that under a set of high level conditions on $\left\{e_t^*, \varepsilon_{t+1}^*\right\}$ (which we collect in Appendix A for convenience), a rotated version of $\hat{\delta}^*$ given by $\Phi^{*\prime}\hat{\delta}^*$, where $\hat{\Phi}^* = diag\left(H^*, I_q\right)$ and H^* is the bootstrap analogue of H, is also asymptotically normally distributed. In particular, under Conditions A* through D*,

$$\sqrt{T}\left(\Phi^{*\prime}\hat{\delta}^* - \hat{\delta}\right) \rightarrow^{d^*} N\left(-c\Delta_\delta^*, \Sigma_\delta^*\right),$$

where Δ_δ^* and Σ_δ^* are the bootstrap analogues of Δ_δ and Σ_δ. Here and throughout, we write $T^* \rightarrow^{d^*} D$, in probability, if conditional on a sample with probability that converges to one, the bootstrap statistic T^* weakly converges to the distribution D under P^*, i.e. $E^*\left(f\left(T_{NT}^*\right)\right) \rightarrow^P E\left(f\left(D\right)\right)$ for all bounded and uniformly continuous functions f, where P^* denotes the bootstrap probability measure induced by the resampling, conditional on the original sample.

As explained by Gonçalves and Perron (2014), the need for rotation is due to the fact that the bootstrap estimated factors estimate only a rotation of the "latent" factors driving the bootstrap DGP. In particular, $\tilde{F}_t^*$ consistently estimates $H^*\tilde{F}_t$ and not $\tilde{F}_t$. However, and contrary to Φ, Φ^* is observed so that the rotation of $\hat{\delta}^*$ is feasible.

The consistency of the bootstrap distribution requires that the bootstrap matches the bias term and the covariance matrix, i.e. that $\Delta_\delta^* = \Delta_\delta$ and $\Sigma_\delta^* = \Sigma_\delta$. To ensure that $\Sigma_\delta^* = \Sigma_\delta$, Gonçalves and Perron (2014) imposed Condition E*, which requires that $\Omega^* = Var^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^T \tilde{z}_t\varepsilon_{t+1}^*\right)$ converges in probability to $\Phi_0\Omega\Phi_0'$. This condition is satisfied if we choose ε_{t+1}^* as to replicate the serial dependence and heterogeneity properties of ε_{t+1}. Under a standard m.d.s. assumption on ε_{t+1}, a natural way of bootstrapping ε_{t+1}^* is to apply the wild bootstrap, i.e. let $\varepsilon_{t+1}^* = \hat{\varepsilon}_{t+1} \cdot v_{t+1}$ with $v_{t+1} \sim$ i.i.d $(0, 1)$, as proposed by Gonçalves and Perron (2014). We will maintain the m.d.s. assumption on ε_{t+1} (cf. Assumption 5) and therefore we will also rely on the wild bootstrap to generate ε_{t+1}^* in this paper.

To ensure that $\Delta_\delta^* = \Delta_\delta$, it is important that we choose e_t^* such that

$$\Gamma^* = \frac{1}{T}\sum_{t=1}^T Var^*\left(\frac{1}{\sqrt{N}}\tilde{\Lambda}'e_t^*\right)$$

converges in probability to $Q\Gamma Q'$ (this is Condition F*). Since Γ is a function of the cross sectional dependence of e_{it}, this requires that we choose e_{it}^* in a way that replicates the cross sectional properties (dependence and heteroskedasticity across i) of e_{it}. By assuming away cross sectional dependence (cf. their Assumption 8), Gonçalves and Perron (2014) showed the validity of a wild bootstrap for inference on δ, whereby $e_{it}^* = \tilde{e}_{it}\eta_{it}$, with $\eta_{it} \sim$ i.i.d.$(0,1)$. Our goal in this paper is to relax this assumption and propose a bootstrap method for e_{it}^* for which $p\lim \Gamma^* = Q\Gamma Q'$ under general cross sectional dependence.

Proposing a bootstrap method that is robust to cross sectional dependence is a much harder task than proposing a method that handles serial dependence. The main reason is that contrary to the time dimension, no natural ordering among the variables needs to exist in the cross sectional dimension. This makes it harder for instance to apply block bootstrap methods, which are often used to capture time series dependence of unknown form.

For panel data, where both the time series and the cross sectional dimensions exist, one common way of preserving the dependence in one dimension is to resample only in the other dimension. The intuition is that by not resampling in one particular dimension, we do not destroy the dependence along this dimension. This idea was recently used by Gonçalves (2011) to propose a blocks bootstrap method that is asymptotically valid for the fixed effects OLS estimator in a panel linear regression model. By applying the moving blocks bootstrap in the time series dimension to the full vector of variables available in each period, this method was shown to be robust to serial and cross sectional dependence of unknown form. A similar idea was used (but without a theoretical justification) by Ludvigson and Ng (2007, 2009, 2011) and Gospodinov and Ng (2013) when testing for predictability using factor augmented regressions.

As we will show next, using a bootstrap that only resamples in the time dimension (and leaves the cross sectional dimension untouched) is in general not valid in the context of factor-augmented regression models.

2.2 Failure of bootstrap methods that only resample in the time dimension

Suppose that we resample the entire $N \times 1$ vector of residuals $\tilde{e}_t = (\tilde{e}_{1t}, \ldots, \tilde{e}_{Nt})'$ only in the time series dimension. In particular, for simplicity, suppose that we let

$$e_t^* \sim \text{ i.i.d. } \left\{\tilde{e}_t - \overline{\tilde{e}}\right\}, \tag{4}$$

where $\overline{\tilde{e}} = \frac{1}{T}\sum_{t=1}^T \tilde{e}_t$ is the time series average of $\tilde{e}_t$; resampling the recentered vector of residuals ensures that $E^*\left(e_t^*\right) = 0$.

The following result shows that generating e_t^* as in (4) implies a zero Γ^*.

Proposition 2.1 *Suppose* $e_t^* \sim$ *i.i.d.* $\left\{\tilde{e}_t - \overline{\tilde{e}}\right\}$ *for* $t = 1, \ldots, T$. *Then* $\Gamma^* = 0$ *for any* N, T.

The proof of Proposition 2.1 follows trivially from the first order conditions that define $\tilde{\Lambda}$. In particular,

$$\Gamma^* = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N}\tilde{\Lambda}' Var^*\left(e_t^*\right)\tilde{\Lambda},$$

where

$$Var^*\left(e_t^*\right) = E^*\left(e_t^* e_t^{*\prime}\right) = \frac{1}{T}\sum_{t=1}^{T}\left(\tilde{e}_t - \bar{\tilde{e}}\right)\left(\tilde{e}_t - \bar{\tilde{e}}\right)' = \frac{\tilde{e}'\tilde{e}}{T} - \frac{\tilde{e}'\iota\iota'\tilde{e}}{T},$$

and where $\tilde{e}$ is a $T \times N$ matrix with rows given by $\tilde{e}_t'$ and $\iota = (1, \ldots, 1)'$ is $T \times 1$. It follows that

$$\Gamma^* = \frac{1}{N}\tilde{\Lambda}' Var^*\left(e_t^*\right)\tilde{\Lambda} = \frac{1}{NT}\left(\tilde{\Lambda}'\tilde{e}'\tilde{e}\tilde{\Lambda} - \tilde{\Lambda}'\tilde{e}'\iota\iota'\tilde{e}\tilde{\Lambda}\right) = 0$$

since $\tilde{\Lambda}'\tilde{e}' = 0$ by the first order conditions that define $\left(\tilde{F}, \tilde{\Lambda}\right)$. Notice that this result holds for any possible value of (N, T).

The main implication of Proposition 2.1 is that the i.i.d. bootstrap distribution is centered at zero (i.e. $\Delta_\delta^* = 0$ because Δ_{δ^*} is a linear function of Γ^* and $\Gamma^* = 0$). Since the OLS estimator is asymptotically biased when the cross sectional dimension is relatively small compared to the time series dimension (i.e. when $\sqrt{T}/N \to c \neq 0$), the i.i.d. bootstrap does not replicate this important feature of the OLS distribution. Note that this failure of the i.i.d. bootstrap holds irrespectively of whether cross sectional dependence exists or not. The problem is not that the i.i.d. bootstrap does not capture cross sectional dependence. Rather the problem is that it induces a zero bias term which should be there even under cross sectional independence as long as $-c\Delta_\delta \neq 0$ (i.e. as long as $c \neq 0$ and $p\lim \hat{\alpha} = H_0^{-1\prime}\alpha \neq 0$).

Although Proposition 2.1 considers the special case of a bootstrap method that resamples residuals in an i.i.d. fashion in the time dimension, the result extends to any bootstrap method that only resamples in the time dimension. To see this, let $\{\tau_t : t = 1, \ldots, T\}$ denote a sequence of random indices taking values on $\{1, \ldots, T\}$. We can think of any time series bootstrap that does not resample in the cross sectional dimension as letting $e_t^* = \tilde{e}_{\tau_t}$, for $t = 1, \ldots, T$. For instance, for the i.i.d. bootstrap analyzed above, τ_t is a sequence of i.i.d. random variables uniformly distributed on $\{1, 2, \ldots, T\}$. For the moving blocks bootstrap with block size equal to b, $\{\tau_t, t = 1, \ldots, T\} = \{I_1 + 1, \ldots, I_1 + b, I_2 + 1, \ldots, I_2 + b, \ldots\}$, where I_j are i.i.d. uniform on $\{0, 1, \ldots, T - b\}$. It follows that

$$E^*\left(e_t^*\right) = \sum_{t=1}^{T} w_t \tilde{e}_t \text{ and } E^*\left(e_t^* e_t^{*\prime}\right) = \sum_{t=1}^{T} w_t \tilde{e}_t \tilde{e}_t'$$

for some sequence of weights w_t such that $\sum_{t=1}^{T} w_t = 1$. This sequence is specific to the particular bootstrap method being used, for instance $w_t = 1/T$ for all t for the i.i.d. bootstrap. See Gonçalves and White (2002) for the formula that defines w_t for the moving blocks bootstrap and the stationary bootstrap. In any case, for a given sequence w_t associated with a particular bootstrap, we have that

$$Var^*\left(e_t^*\right) = \sum_{t=1}^{T} w_t \tilde{e}_t \tilde{e}_t' - \left(\sum_{t=1}^{T} w_t \tilde{e}_t\right)\left(\sum_{t=1}^{T} w_t \tilde{e}_t\right)',$$

implying that

$$\Gamma^* = \frac{1}{N}\sum_{t=1}^{T} w_t \tilde{\Lambda}'\tilde{e}_t\tilde{e}_t'\tilde{\Lambda} - \frac{1}{N}\left(\sum_{t=1}^{T} w_t \tilde{\Lambda}'\tilde{e}_t\right)\left(\sum_{t=1}^{T} w_t \tilde{\Lambda}'\tilde{e}_t\right)',$$

which is zero given that $\tilde{\Lambda}'\tilde{e}_t = 0$ for each $t = 1, \ldots, T$.

3 A new cross sectional dependence robust bootstrap method

In this section, we propose a new bootstrap method for factor models that is consistent under cross sectional dependence. Following Bai and Ng (2006), we impose the following assumption, which is a strengthening of Assumptions 1-5 in Appendix A.

Assumption CS $\Sigma \equiv E\left(e_t e_t'\right) = \left(\sigma_{ij}\right)_{i,j=1,\ldots,N}$ for all t, i, j and is such that $\lambda_{\min}\left(\Sigma\right) > c_1$ and $\lambda_{\max}\left(\Sigma\right) < c_2$ for some positive constants c_1 and c_2.

Under Assumption CS, the $N \times N$ covariance matrix of the idiosyncratic errors e_t is time invariant and has eigenvalues that are bounded and bounded away from zero, uniformly in N. The boundedness assumption on the maximum eigenvalue of Σ is standard in the approximate factor model literature (cf. Chamberlain and Rothschild (1983), Bai (2003) and Bai and Ng (2006), among many others), allowing for weak cross sectional dependence of unknown form. The time series stationarity assumption is less standard but has been used by Bai and Ng (2006) to propose a consistent estimator of $\Gamma = Var\left(N^{-1/2}\Lambda' e_t\right)$ when there is weak cross sectional dependence in e_{it}. As they explain, the main intuition is that if covariance stationarity holds we can use the time series observations on $\{\tilde{e}_{it}\}$ to consistently estimate the cross section correlations σ_{ij} and hence Σ. Here we rely on this same idea to propose a consistent bootstrap method that is robust to cross sectional dependence and that at the same time does not yield a zero Γ^*.

3.1 The cross sectional dependent bootstrap for factor models

Let $\tilde{\Sigma} = \left(\tilde{\sigma}_{ij}\right)_{i,j=1,\ldots,N}$ denote an estimator of Σ. The cross sectional dependence robust bootstrap algorithm is as follows.

CSD bootstrap algorithm

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

   $$X_t^* = \tilde{\Lambda}\tilde{F}_t + e_t^*,$$

   where $\left\{e_t^* = \left(e_{1t}^*, \ldots, e_{Nt}^*\right)'\right\}$ is such that

   $$e_t^* = \tilde{\Sigma}^{1/2}\eta_t, \text{ where } \eta_t \text{ is i.i.d. } \left(0, I_N\right) \text{ over } t \tag{5}$$

   and the elements of $\eta_t = \left(\eta_{1t}, \ldots, \eta_{Nt}\right)'$ are mutually independent for given t. $\tilde{\Sigma}^{1/2}$ is the square root matrix of $\tilde{\Sigma}$.

2. Estimate the bootstrap factors $\tilde{F}^*$ and the bootstrap loadings $\tilde{\Lambda}^*$ using X^*.

3. For $t = 1, \ldots, T-1$, let

$$y^*_{t+1} = \hat{\alpha}'\tilde{F}_t + \hat{\beta}'W_t + \varepsilon^*_{t+1},$$

   where the error term ε^*_{t+1} is a wild bootstrap resampled version of $\hat{\varepsilon}_{t+1}$, i.e.

$$\varepsilon^*_{t+1} = \hat{\varepsilon}_{t+1} v_{t+1},$$

   where the external random variable v_{t+1} is i.i.d. $(0, 1)$ and is independent of η_{it}.

4. Regress y^*_{t+1} generated in 3. on the estimated bootstrap factors and the fixed regressors $\hat{z}^*_t = \left(\tilde{F}^{*\prime}_t, W'_t\right)'$. This yields the bootstrap OLS estimators $\hat{\delta}^*$.

In step 2, we generate ε^*_{t+1} using a wild bootstrap, which is appropriate under our martingale difference sequence assumption (cf. Assumption 5(a)). When forecasting over longer horizons, Djogbenou et al. (2015) show that applying a combination of the moving blocks and the wild bootstrap corrects for serial dependence in ε_{t+h} when $h > 1$.

The following result proves the asymptotic validity of the CSD bootstrap under a convergence condition on the spectral norm of $\tilde{\Sigma} - \Sigma$.

Theorem 3.1 *Suppose Assumptions 1-5 strengthened by Assumption CS hold and we implement the CSD bootstrap with $\tilde{\Sigma}$ such that*

$$\rho\left(\tilde{\Sigma} - \Sigma\right) \to^P 0. \tag{6}$$

Then, if $\sqrt{T}/N \to c$, with $0 \leq c < \infty$,

$$\sup_{x \in \mathbb{R}^p} \left| P^*\left(\sqrt{T}\left(\Phi^{*\prime}\hat{\delta}^* - \hat{\delta}\right) \leq x\right) - P\left(\sqrt{T}\left(\hat{\delta} - \delta\right) \leq x\right)\right| \to^P 0,$$

where $\Phi^* = diag\left(H^*, I_q\right)$ and H^* is the bootstrap analogue of H.

Theorem 3.1 shows that a sufficient condition for the CSD bootstrap to be asymptotically valid when Assumptions 1-5 strengthened by Assumption CS hold and $\sqrt{T}/N \to c$ is that $\tilde{\Sigma}$ is consistent towards Σ under the spectral norm ρ. This condition is used to show that

$$\Gamma^* \equiv Var^*\left(\frac{1}{\sqrt{N}}\tilde{\Lambda}'e^*_t\right) = \frac{1}{N}\tilde{\Lambda}'Var^*\left(e^*_t\right)\tilde{\Lambda} = \frac{1}{N}\tilde{\Lambda}'\tilde{\Sigma}\tilde{\Lambda} \xrightarrow{P} Q\Gamma Q',$$

thus verifying Condition F* in Gonçalves and Perron (2014). Together with Assumptions 1-5, (6) suffices to show that the remaining high level conditions for bootstrap validity (cf. Conditions A*-E* in Appendix A) hold. Note that consistency of $\tilde{\Sigma}$ towards Σ under the spectral norm ensures that asymptotically, as $N, T \to \infty$, all eigenvalues of $\tilde{\Sigma}$ converge to the corresponding eigenvalues of Σ. Since the latter are bounded away from zero by Assumption CS, it follows that $\tilde{\Sigma}$ is asymptotically nonsingular. This is sufficient to guarantee that $p\lim \Gamma^* \neq 0$.

In order to implement the CSD bootstrap method, we need to choose $\tilde{\Sigma}$. One natural choice could be the sample covariance matrix given by

$$\tilde{\Sigma} = (\hat{\sigma}_{ij}), \quad \text{with} \quad \hat{\sigma}_{ij} = \frac{1}{T}\sum_{t=1}^{T} \tilde{e}_{it}\tilde{e}_{jt}.$$

This choice is not a good choice, for two reasons. First, it is well known that the sample covariance matrix can behave poorly, especially when the cross sectional dimension is larger than the time series dimension (in particular, it is not consistent in the spectral norm). The second reason, specific to our context, is that it also induces a zero bias term in the bootstrap distribution by implying $\Gamma^* = 0$, just as the i.i.d. bootstrap analyzed in the previous section. Indeed,

$$\Gamma^* = Var^*\left(\frac{1}{\sqrt{N}}\tilde{\Lambda}' e_t^*\right) = \frac{1}{N}\tilde{\Lambda}' Var^*\left(e_t^*\right)\tilde{\Lambda} = \frac{1}{N}\tilde{\Lambda}'\tilde{\Sigma}\tilde{\Lambda},$$

where

$$\tilde{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} \tilde{e}_t \tilde{e}_t'.$$

Therefore,

$$\Gamma^* = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{N}}\tilde{\Lambda}'\tilde{e}_t\right)\left(\frac{1}{\sqrt{N}}\tilde{\Lambda}'\tilde{e}_t\right)' = 0,$$

by the first order conditions defining $\tilde{\Lambda}$ and $\tilde{F}$. Hence, the CSD bootstrap implemented with the sample covariance matrix leads to a distribution centered at zero, which is only correct when either $\alpha = 0$ or $\sqrt{T}/N \to 0$.

3.2 A CSD bootstrap based on thresholding

In order to avoid a zero Γ^* matrix, some regularization of $\tilde{\Sigma} = \tilde{e}'\tilde{e}/T$ is needed. Our approach in this paper is to use thresholding. The main idea is that rather than using the sample covariance to estimate all the off-diagonal elements of Σ, we keep only those that exceed a given threshold. If the covariance matrix Σ is sparse in the sense that most of its off-diagonal elements are zero, thresholding allows for consistent estimation of Σ even when its dimension is very large (and potentially larger than T). Although there are other regularization methods we could use, thresholding has the advantage that it is invariant to variable permutations. This is particularly useful when no natural ordering exists among the variables. If a distance metric on the indices (i, j) is available and we can order variables according to this measure, then banding/tapering is another regularization method that could be used.

More specifically, we follow Bickel and Levina (2008) and consider the following estimator of Σ:

$$\tilde{\Sigma} = [\tilde{\sigma}_{ij}],$$

with

$$\tilde{\sigma}_{ij} = \begin{cases} \hat{\sigma}_{ij} & i = j \\ \hat{\sigma}_{ij} 1\left(\hat{\sigma}_{ij} \geq \omega\right) & i \neq j \end{cases}, \text{ with } \hat{\sigma}_{ij} = \frac{1}{T}\sum_{t=1}^{T}\tilde{e}_{it}\tilde{e}_{jt},$$

and where $\omega \equiv \omega_{NT}$ is a threshold value that we need to specify. This form of thresholding is known in the literature as a "universal hard thresholding rule": "universal" because we apply a single threshold level to all the entries of $\tilde{\Sigma}$ and "hard" because we use the indicator function as a thresholding function ($s(z) = z1(|z| > \omega)$). Rothman, Levina and Zhu (2009) relax this assumption by allowing for more general thresholding functions such as the soft thresholding where $s(z) = sgn(z)(|z| - \omega)_+$, where $(\cdot)_+$ denotes the positive part. Cai and Liu (2011) propose a generalization of the universal thresholding rule that adapts to the amount of variability of the entries by using different threshold values (this is the so-called adaptive thresholding method). Whereas these papers apply thresholding to a sample covariance matrix that is obtained from a set of observed variables, here we apply thresholding to the estimated residuals of a factor model. Hence, our paper is more closely related to Fan, Liao and Mincheva (2011, 2013), who consider the adaptive thresholding approach with estimated residuals. In particular, we draw heavily on Fan, Liao and Mincheva (2013), where the "Principal orthogonal complement thresholding estimator" (POET) for factor models was proposed. Contrary to Fan, Liao and Mincheva (2013), here we focus on the universal hard thresholding rule rather than the adaptive thresholding function. The main reason for doing so is that this allows us to dispense with the assumption that $\sqrt{T}/N \to 0$, which is used by Fan, Liao and Mincheva (2013) to prove consistency of the POET estimator (see their Theorem 1). This assumption is too restrictive for our purposes because it implies that factors estimation uncertainty does not matter for inference. As we show next, the estimator given above is consistent for Σ under the spectral norm with the assumption that $\sqrt{T}/N \to c \neq 0$, where we expect the gains from bootstrapping to be larger.

The choice of the threshold ω_{NT} that ensures the convergence of $\tilde{\Sigma}$ towards Σ under the spectral norm depends on the convergence rate of $\max_{i,j} |\hat{\sigma}_{ij} - \sigma_{ij}|$ (cf. Lemma B.1). The following additional assumptions allow us to derive this rate.

Assumption TS As $N, T \to \infty$ such that $\log N/T \to 0$,

(a) $\max_{i,j \leq N} \left| \frac{1}{T} \sum_{t=1}^{T} e_{it} e_{jt} - \sigma_{ij} \right| = O_P\left(\sqrt{\frac{\log N}{T}} \right)$.

(b) $\max_{i \leq N} \left\| \frac{1}{T} \sum_{t=1}^{T} F_t e_{it} \right\| = O_P\left(\sqrt{\frac{\log N}{T}} \right)$.

Assumption TS (a) is a high level assumption which requires the time average of $e_{it}e_{jt}$ to converge to σ_{ij} at the indicated rate uniformly in $i, j = 1, \ldots, N$. This rate is implied by more primitive assumptions that include stationary mixing conditions as well as exponential-type tail conditions on the unobserved errors $e_t \equiv (e_{1t}, \ldots, e_{Nt})'$. Similarly, part (b) of Assumption TS is implied by stationarity, strong mixing and exponential tail conditions on the latent factors. See Lemma 4 of Fan, Liao and Mincheva (2013) for a specific set of regularity conditions that imply Assumption TS.

Given Assumptions 1-5 strengthened by Assumptions CS and TS, and using results from Fan, Liao and Mincheva (2011, 2013), we can easily show that $\max_{i,j} |\hat{\sigma}_{ij} - \sigma_{ij}| = O_P\left(\frac{1}{\sqrt{N}} + \sqrt{\frac{\log N}{T}} \right)$

(cf. Lemma B.2). The $1/\sqrt{N}$ term captures the effect of factors estimation uncertainty whereas the second term is the optimal uniform rate of convergence of the sample covariances between e_{it} and e_{jt}. Following Fan, Liao and Mincheva (2013), we set ω_{NT} equal to the maximum estimation error in $\hat{\sigma}_{ij}$, i.e. we set

$$\omega_{NT} = C\left(\frac{1}{\sqrt{N}} + \sqrt{\frac{\log N}{T}}\right),$$

where $C > 0$ is a sufficiently large constant[1]. With this choice of ω_{NT}, we can prove the consistency of $\tilde{\Sigma}$ towards Σ provided Σ is sufficiently sparse.

To characterize the sparsity of Σ, we follow Fan, Liao and Mincheva (2011) and impose an upper bound restriction on the maximum number of non-zero elements of Σ across rows,

$$m_N \equiv \max_{i \leq N} \sum_{j=1}^{N} 1\left(\sigma_{ij} \neq 0\right).$$

In particular, although we allow m_N to grow to infinity[2], we require

$$m_N = o\left(\min\left(\sqrt{N}, \sqrt{\frac{T}{\log N}}\right)\right).$$

Assuming a sparse covariance matrix for the idiosyncratic errors of a panel factor model is a very natural assumption since it is consistent with the idea that the common factors capture most of the dependence in the observable variables X_{it} and any residual cross sectional dependence that is left is weak (as postulated by the approximate factor panel model of Chamberlain and Rothschild (1983)). Two recent papers in econometrics that have relied on a similar sparsity assumption are Gagliardini, Ossola and Scaillet (2015) and Fan, Liao and Yao (2015). Their testing problem is different from ours and they do not rely on the bootstrap.

Our main result is as follows.

Theorem 3.2 *Suppose Assumptions 1-5 strengthened by Assumptions CS and TS hold and* $\sqrt{T}/N \to c$, *with* $0 \leq c < \infty$, *as* $N, T \to \infty$. *If* $\log N = o(T)$ *and* $m_N = o\left(\min\left(\sqrt{N}, \sqrt{\frac{T}{\log N}}\right)\right)$, *then the conclusions of Theorem 3.1 hold for the thresholding CSD bootstrap method based on a threshold value* $\omega_{NT} = C\left(\frac{1}{\sqrt{N}} + \sqrt{\frac{\log N}{T}}\right)$.

The condition $\log N/T \to 0$ is needed to consistently estimate the high dimensional covariance matrix Σ under the spectral norm. This condition is implied by the condition that $\sqrt{T}/N \to c$ when $c \neq 0$ (since then N is proportional to $\sqrt{T}$ and $\log T/T \to 0$). When $c = 0$, the two conditions are consistent with each other and we can interpret the former as imposing an upper bound on N as a function of T whereas $\sqrt{T}/N \to 0$ imposes a lower bound.

[1] In the simulations, we use cross-validation to choose C.

[2] Note that this is consistent with Assumption CS, which requires that $\lambda_{\max}(\Sigma)$ by uniformly bounded above. Indeed, we can show that $\rho(\Sigma) = \lambda_{\max}(\Sigma) \leq \max_i \sum_j |\sigma_{ij}| \leq m_N M$ under the assumption that $|\sigma_{ij}| \leq M$ since for any symmetric matrix A, $\rho(A) \leq \max_i \sum_j |a_{ij}|$.

4 Monte Carlo results

In this section, we report results from a simulation experiment documenting the properties of bootstrap procedures in factor-augmented regressions. The data-generating process (DGP) is the same as DGP 6 in Gonçalves and Perron (2014). We consider the single factor model:

$$y_{t+1} = \alpha F_t + \varepsilon_{t+1}, \tag{7}$$

with $\alpha = 1$ and where F_t is drawn from a standard normal distribution independently over time. The regression error ε_{t+1} is normally distributed but heteroskedastic with variance $\frac{F_t^2}{3}$. The rescaling is done to make the asymptotic variance of $\sqrt{T}\left(\hat{\alpha} - \alpha H_0^{-1}\right)$ equal to 1. Because our DGP satisfies condition PC1 in Bai and Ng (2013), H_0 is ± 1, so that we identify the parameter up to sign.

The $(T \times N)$ matrix of panel variables is generated as:

$$X_{it} = \lambda_i F_t + \theta e_{it}, \tag{8}$$

where λ_i is drawn from a $U[0,1]$ distribution (independent across i). The variance of e_{it} is drawn from $U[.5, 1.5]$, and cross sectional dependence is similar to the design in Bai and Ng (2006):

$$corr\left(e_{it} e_{jt}\right) = \begin{cases} .5^{|i-j|} \text{ if } |i-j| \leq 5 \\ 0 \text{ otherwise} \end{cases}.$$

This makes m_N fixed at 11 and thus the rate restriction on m_N is satisfied. The parameter $\theta = \sqrt{\frac{.333}{.817}}$ is used to rescale e_{it} to obtain $\Gamma = 1/3$ and make the results comparable to those in Gonçalves and Perron (2014). We report results from experiments based on 5000 replications with $B = 399$ bootstrap repetitions. We consider three values for N (50, 100, and 200) and T (50, 100, and 200).

In a second experiment, we consider the case where we change the order of the columns of the X matrix, so that the covariance matrix of the idiosyncratic errors is no longer Toeplitz (we call this DGP 2). This will allow us to look at the robustness of estimators of Γ to arbitrary changes in the ordering of the data.

We concentrate on inference about α in (7). We report mean bias and coverage rates for asymptotic and bootstrap equal-tailed percentile t confidence intervals at a nominal level of 95%. We report results for three asymptotic methods and four bootstrap methods. The asymptotic methods are the standard OLS estimator and two bias-corrected estimators that plug in estimates of Γ to estimate the bias. The two estimators we consider are the thresholding estimator of Bickel and Levina (2008), denoted BL, using their suggested cross-validation procedure to choose the threshold, and the CS-HAC estimator of Bai and Ng (2006):

$$\hat{\Gamma}_{CS-HAC} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \tilde{\lambda}_i \tilde{\lambda}_j \frac{1}{T} \sum_{t=1}^{T} \tilde{e}_{it} \tilde{e}_{jt}$$

with $n = \min\left(\sqrt{N}, \sqrt{T}\right)$.

The four bootstrap methods are the CSD bootstrap based on the Bickel and Levina (2008) thresholding estimator, the wild bootstrap (WB) estimator of Gonçalves and Perron (2014), the bootstrap that resamples vectors independently over time, and the CSD bootstrap based on the empirical covariance matrix:

$$\tilde{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} \tilde{e}_t \tilde{e}_t'.$$

One issue in implementing the bootstrap with the thresholding estimator is that the estimator is not necessarily positive definite. This is a finite-sample issue as it converges asymptotically to a positive definite matrix. We followed McMurry and Politis (2010) and replaced the small eigenvalues of the estimated matrix by a small positive number (10^{-6}).

The results are presented in Table 1. The top panel of the table includes results for DGP 1 (with a Toeplitz covariance matrix of the idiosyncratic errors), while the bottom panel reports results for DGP 2 (with the reshuffled columns of X). In each panel, the top portion refers to bias results, and the bottom portion to coverage rates.

The bias of the OLS estimator is reported along the top row of the table. These numbers are identical to those in Gonçalves and Perron (2014): the bias is negative, it is reduced for large values of N, and it is little affected by changes in T. The two rows below report the mean estimated bias with each of the estimators of the covariance matrix. Note that we do not report results for the empirical covariance matrix as the estimate is exactly 0 by the first-order conditions as discussed above. The CS-HAC seems to have a slight advantage over the BL estimator, but this advantage disappears for larger N and T.

The cross sectional dependent bootstrap using the threshold estimator reproduces the Bickel and Levina bias results above. As expected, the wild bootstrap does not perform as well, but it does capture a large fraction of the bias as it includes the diagonal elements of $\tilde{\Sigma}$ in the computation of Γ^*. Finally, we see that the two methods that lead to a zero Γ^*, the empirical covariance matrix and i.i.d. resampling over time, do not capture the bias, though the mean is not quite 0.

The results for coverage rates mirror those for the bias. The first row reports coverage rates for the OLS estimator based on the asymptotic theory of Bai and Ng that assumes the absence of a bias. We see that the presence of this bias leads to severe under coverage, for example a 95% confidence intervals for $N = T = 50$ only includes the true value of the parameter in 70.5% of the replications. These distortions are reduced as N increases as the bias of the estimator is reduced.

The bias-correction methods are partially successful in providing accurate coverage. For example, for $N = T = 50$, the coverage rates are 77.9% and 79.8% compared to 70.5% for the OLS estimator. As N increases, the difference between the bias-correction methods and OLS disappears as the bias converges towards 0. Again, the CS-HAC seems to dominate, though this advantage disappears as N and T increase.

The cross sectional bootstrap with the thresholding covariance matrix improves on these coverage rates. Still for $N = T = 50$, we obtain a coverage rate of 87.9%, the highest among all methods considered. The wild bootstrap of Gonçalves and Perron (2014) corrects some of the distortions as it captures part of the bias term. Finally, the two invalid methods behave in a way similar to the OLS estimator since they do not capture the bias term at all.

The bottom panel of Table 1 presents the results for DGP 2. Most of the results are identical to those of the upper panel, showing the robustness of these methods to random reshuffling of the data. The only results that are markedly different are those for CS-HAC as it depends on the ordering of the data. This method is much less effective in estimating the bias and in correcting the undercoverage of the confidence intervals based on the OLS estimator. On the other hand, bias correction with the Bickel and Levina estimator works just as well as before. Similarly, the cross sectional dependent bootstrap with the thresholding covariance matrix works just as before.

5 Empirical application

In this section, we apply our cross sectional dependent bootstrap to the problem of forecasting inflation using convenience yields on commodities as explored recently by Gospodinov and Ng (2013), GN henceforth. The data is the same as GN[3] and contains convenience yields and HP-filtered log prices for $N = 23$ commodities divided into 6 categories: Foodstuffs, Grains and Oilseeds, Industrials, Livestock and Meats, Metals, and Energy. The variable to be forecast is monthly CPI inflation (all items, urban consumers, seasonally adjusted), and several observable variables are also included as predictors: detrended oil prices and a lag, $q_{oil,t}$ and $q_{oil,t-1}$, respectively, the 3-month T-bill rate i_t, the log change of trade-weighted USD exchange rate Δx_t, and the deviation of the unemployment rate from the HP trend. Readers interested in more details regarding the data should consult the online appendix to GN.

We consider forecasting inflation using GN's augmented model which includes two factors (or principal components) extracted from convenience yields, two factors extracted from detrended commodity prices, two autoregressive components, and the observable predictors:

$$\Delta p_{t+1} = \beta_0 + \alpha_1 \tilde{F}_{1,t} + \alpha_2 \tilde{F}_{2,t} + \alpha_3 \tilde{G}_{1,t} + \alpha_4 \tilde{G}_{2,t} + \beta' W_t + \varepsilon_{t+1},$$

where $\Delta p_{t+1} = \log\left(\frac{p_{t+1}}{p_t}\right)$ is monthly inflation and $\tilde{F}_{j,t}$ is a factor summarizing the dynamics of the convenience yields for $j = 1, 2$, $\tilde{G}_{i,t}$ is a factor summarizing the dynamics of detrended commodity prices for $i = 1, 2$, and W_t is a vector containing the observable predictors (including the lags of Δp_{t+1}). This specification corresponds to the second column in their Table 1.

Table 2 reports the point estimates along with four sets of 90% confidence intervals. The point estimates are identical to those reported in GN. The first confidence interval is based on asymptotic theory and is obtained by adding and subtracting 1.645 times the heteroskedasticity-robust standard error to the estimate. The other three intervals are computed using the bootstrap. The first bootstrap

[3] We thank Nikolay Gospodinov for providing us this data.

intervals are based on the above cross sectional dependent bootstrap using the thresholding estimator of the covariance matrix of the idiosyncratic errors with data-based threshold. For this purpose, we apply the CSD bootstrap to the stacked vectors of idiosyncratic residuals for the panel of convenience yields and detrended commodity prices. The second confidence interval uses the diagonal of the estimated covariance matrix of the stacked errors and corresponds to the wild bootstrap of Gonçalves and Perron (2014). Differences between these two intervals measure the importance of allowing for cross sectional dependence. Finally, the last row reports intervals based on the GN block bootstrap, and they differ slightly from those reported in their paper for two reasons: first, we use heteroskedasticity-robust standard errors instead of standard errors robust to serial correlation, and second we use a block size of 1 instead of 4. These two modifications are justified by the fact that our forecast horizon is 1 and that ε_{t+1} should not have serial correlation under correct specification.

We now discuss the results. Overall, our confidence intervals tend to be narrower than those in GN. Second, our intervals for the coefficients on the factors, $\alpha_1, \ldots, \alpha_4$, are shifted relative to those in GN. For example, for the first convenience yield factor, the GN 90% confidence interval is $[.00, .18]$ which is roughly centered at the point estimate of .085. Instead, our interval is $[.06, .19]$, reflecting the presence of a downward bias in the OLS estimator due to factor estimation. This is also the instance where allowing for cross sectional dependence among idiosyncratic errors seems most important. The interval based on the wild bootstrap (which assumes no dependence) is $[.09, .24]$. The estimated bias is so large that the confidence interval lies completely to the right of the point estimate.

For the observable predictors, bias does not seem to be very important. This is reflected in the fact that the CSD and wild bootstrap intervals are similar and centered at the point estimates.

6 Conclusion

In this paper, we consider the bootstrap for factor models where the idiosyncratic errors are correlated in the cross-section. We show that some natural approaches fail in this context because they lead to a singular bootstrap covariance matrix for the estimated factors, inducing a zero bias in the bootstrap distribution of the estimated coefficients. Instead, we propose a solution based on a consistent estimator of the covariance matrix of the idiosyncratic errors. We show that if we use a hard threshold estimator, we can obtain bootstrap consistency for inference on the parameters in a factor-augmented regression.

It would be interesting to see whether this approach can be generalized to other estimators of the covariance matrix of the idiosyncratic errors. We also think that it would also be interesting to extend this approach to more general models with cross-sectional dependence and with time heterogeneity.

A Appendix A: Assumptions and bootstrap high level conditions

The following set of assumptions is standard in the literature on factor models, see Bai (2003), Bai and Ng (2006) and Gonçalves and Perron (2014). We let $z_t = \begin{pmatrix} F_t' & W_t' \end{pmatrix}'$, where z_t is $p \times 1$, with $p = r + q$.

Assumption 1

(a) $E\|F_t\|^4 \leq M$ and $\frac{1}{T}\sum_{t=1}^{T} F_t F_t' \to^P \Sigma_F > 0$, where Σ_F is a non-random $r \times r$ matrix.

(b) The factor loadings λ_i are deterministic such that $\|\lambda_i\| \leq M$ and $\Lambda'\Lambda/N \to \Sigma_\Lambda > 0$.

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

Assumption 2

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

(b) $E(e_{it}e_{js}) = \sigma_{ij,ts}$, $|\sigma_{ij,ts}| \leq \bar{\sigma}_{ij}$ for all (t,s) and $|\sigma_{ij,ts}| \leq \tau_{ts}$ for all (i,j) such that $\frac{1}{N}\sum_{i,j=1}^{N} \bar{\sigma}_{ij} \leq M$, $\frac{1}{T}\sum_{t,s=1}^{T} \tau_{ts} \leq M$, and $\frac{1}{NT}\sum_{t,s,i,j} |\sigma_{ij,ts}| \leq M$.

(c) For every (t,s), $E\left|N^{-1/2}\sum_{i=1}^{N}(e_{it}e_{is} - E(e_{it}e_{is}))\right|^4 \leq M$.

Assumption 3

(a) $E\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} F_t e_{it}\right\|^2\right) \leq M$, where $E(F_t e_{it}) = 0$ for all (i,t).

(b) For each t, $E\left\|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T}\sum_{i=1}^{N} z_s (e_{it}e_{is} - E(e_{it}e_{is}))\right\|^2 \leq M$, where $z_s = \begin{pmatrix} F_s' & W_s' \end{pmatrix}'$.

(c) $E\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T} z_t e_t' \Lambda\right\|^2 \leq M$, where $E(z_t \lambda_i' e_{it}) = 0$ for all (i,t).

(d) $E\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \lambda_i e_{it}\right\|^2\right) \leq M$, where $E(\lambda_i e_{it}) = 0$ for all (i,t).

(e) As $N, T \to \infty$, $\frac{1}{TN}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j' e_{it} e_{jt} - \Gamma \to^P 0$, where $\Gamma \equiv \lim_{N,T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \Gamma_t > 0$, and $\Gamma_t \equiv Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \lambda_i e_{it}\right)$.

Assumption 4

(a) For each t, $E\left|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T-1}\sum_{i=1}^{N} \varepsilon_{s+1}(e_{it}e_{is} - E(e_{it}e_{is}))\right|^2 \leq M$.

(b) $E\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T-1}\sum_{i=1}^{N} \lambda_i e_{it} \varepsilon_{t+1}\right\|^2 \leq M$, where $E(\lambda_i e_{it} \varepsilon_{t+1}) = 0$ for all (i,t).

Assumption 5

(a) $E(\varepsilon_{t+1}|y_t, F_t, y_{t-1}, F_{t-1}, \ldots) = 0$, $E|\varepsilon_{t+1}|^2 < M$, and F_t and ε_t are independent of the idiosyncratic errors e_{is} for all (i,s,t).

(b) $E\|z_t\|^4 \leq M$ and $\frac{1}{T}\sum_{t=1}^{T} z_t z_t' \to^P \Sigma_{zz} > 0.$

(c) As $T \to \infty$, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1} z_t \varepsilon_{t+1} \to^d N(0, \Omega)$, where $E\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1} z_t \varepsilon_{t+1}\right\|^2 < M$, and $\Omega \equiv \lim_{T\to\infty} Var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1} z_t \varepsilon_{t+1}\right) > 0.$

Next, we review the bootstrap high level conditions proposed by GP (2014). As usual in the bootstrap literature, we use P^* to denote the bootstrap probability measure, conditional on the original sample (defined on a given probability space $(\Omega, \mathcal{F}, P)$). Because the sample depends on N and T, as well as on the given sample realization ω, P^* is a random measure that depends on N, T and ω and we should write $P^*_{NT,\omega}$. However, for simplicity, we omit the indices in P^*. Similarly, we omit the indices NT when referring to the bootstrap samples $\left\{e^*_{it}, \varepsilon^*_{t+h}\right\}$. For any bootstrap statistic T^*_{NT}, we write $T^*_{NT} = o_{P^*}(1)$, in probability, or $T^*_{NT} \to^{P^*} 0$, in probability, when for any $\delta > 0$, $P^*(|T^*_{NT}| > \delta) = o_P(1)$. We write $T^*_{NT} = O_{P^*}(1)$, in probability, when for all $\delta > 0$ there exists $M_\delta < \infty$ such that $\lim_{N,T\to\infty} P\left[P^*(|T^*_{NT}| > M_\delta) > \delta\right] = 0$. Finally, we write $T^*_{NT} \to^{d^*} D$, in probability, if conditional on a sample with probability that converges to one, T^*_{NT} weakly converges to the distribution D under P^*, i.e. $E^*(f(T^*_{NT})) \to^P E(f(D))$ for all bounded and uniformly continuous functions f.

Condition A* (weak time series and cross section dependence in e^*_{it})

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

(b) $\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} |\gamma^*_{st}|^2 = O_P(1)$, where $\gamma^*_{st} = E^*\left(\frac{1}{N}\sum_{i=1}^{N} e^*_{it} e^*_{is}\right)$.

(c) $\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} E^*\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(e^*_{it}e^*_{is} - E^*(e^*_{it}e^*_{is})\right)\right|^2 = O_P(1)$.

Condition B* (weak dependence among $\hat{z}_t, \tilde{\lambda}_i$, and e^*_{it})

(a) $\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} \tilde{F}_s \tilde{F}_t' \gamma^*_{st} = O_P(1)$.

(b) $\frac{1}{T}\sum_{t=1}^{T} E^*\left\|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T}\sum_{i=1}^{N} \hat{z}_s\left(e^*_{it}e^*_{is} - E^*(e^*_{it}e^*_{is})\right)\right\|^2 = O_P(1)$, where $\hat{z}_s = \begin{pmatrix} \tilde{F}_s' & W_s' \end{pmatrix}'$.

(c) $E^*\left\|\frac{1}{\sqrt{TN}}\sum_{t=1}^{T}\sum_{i=1}^{N} \hat{z}_t \tilde{\lambda}_i' e^*_{it}\right\|^2 = O_P(1)$.

(d) $\frac{1}{T}\sum_{t=1}^{T} E^*\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N} \tilde{\lambda}_i e^*_{it}\right\|^2 = O_P(1)$.

(e) $\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\tilde{\Lambda}' e^*_t}{\sqrt{N}}\right)\left(\frac{e^{*\prime}_t \tilde{\Lambda}}{\sqrt{N}}\right) - \Gamma^* = o_{P^*}(1)$, in probability, where $\Gamma^* \equiv \frac{1}{T}\sum_{t=1}^{T} Var^*\left(\frac{1}{\sqrt{N}}\tilde{\Lambda}' e^*_t\right) > 0.$

Condition C* (weak dependence between e^*_{it} and ε^*_{t+1})

(a) $\frac{1}{T}\sum_{t=1}^{T} E^*\left|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T-1}\sum_{i=1}^{N} \varepsilon^*_{s+1}\left(e^*_{it}e^*_{is} - E^*(e^*_{it}e^*_{is})\right)\right|^2 = O_P(1)$.

(b) $E^*\left\|\frac{1}{\sqrt{TN}}\sum_{t=1}^{T-1}\sum_{i=1}^{N} \tilde{\lambda}_i e^*_{it} \varepsilon^*_{t+1}\right\|^2 = O_P(1)$, where $E\left(e^*_{it}\varepsilon^*_{t+1}\right) = 0$ for all (i, t).

(c) $\frac{1}{T}\sum_{t=1}^{T-1}\sum_{s=1}^{T}\tilde{F}_s\varepsilon_{t+1}^{*}\gamma_{st}^{*}=O_{P^*}(1)$, in probability.

Condition D* (bootstrap CLT)

(a) $E^*\left(\varepsilon_{t+1}^*\right)=0$ and $\frac{1}{T}\sum_{t=1}^{T-1}E^*\left|\varepsilon_{t+1}^*\right|^2=O_P(1)$.

(b) $\Omega^{*-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\hat{z}_t\varepsilon_{t+1}^*\to^{d^*}N(0,I_p)$, in probability, where $E^*\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\hat{z}_t\varepsilon_{t+1}^*\right\|^2=O_P(1)$, and $\Omega^*\equiv Var^*\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T-1}\hat{z}_t\varepsilon_{t+1}^*\right)>0$.

Condition E*. $p\lim\Omega^*=\Phi_0\Omega\Phi_0'$.

Condition F*. $p\lim\Gamma^*=Q\Gamma Q'$.

B Appendix B: Proofs

First, we provide two auxiliary lemmas, followed by their proofs. Finally, we prove Theorems 3.1 and 3.2.

Lemma B.1 *Suppose that the two following conditions hold:*

(a) $\max_{i,j\leq N}|\hat{\sigma}_{ij}-\sigma_{ij}|=O_P(b_{NT})$ *for some sequence* $b_{NT}\to 0$.

(b) $\omega\equiv\omega_{NT}=Cb_{NT}\to 0$ *for some sufficiently large constant* $C>0$. *It follows that*

$$\rho\left(\tilde{\Sigma}-\Sigma\right)=O_P(m_N\omega_{NT})=o_P(1),$$

if ω_{NT} *is such that* $m_N\omega_{NT}=o(1)$.

Lemma B.2 *Under Assumptions 1-5 strengthened by Assumptions CS and TS, we have that* $\max_{i,j\leq N}|\hat{\sigma}_{ij}-\sigma_{ij}|=O_P\left(\frac{1}{\sqrt{N}}+\sqrt{\frac{\log N}{T}}\right)$.

Proof of Lemma B.1. Noting that for any symmetric matrix A, $\rho(A)\leq\max_{i\leq N}\sum_j|a_{ij}|$, it follows that

$$\rho\left(\tilde{\Sigma}-\Sigma\right)\leq\max_{i\leq N}\sum_{j=1}^{N}|\tilde{\sigma}_{ij}-\sigma_{ij}|\leq\max_{i\leq N}\sum_{j=1}^{N}|\hat{\sigma}_{ij}-\sigma_{ij}|1(|\hat{\sigma}_{ij}|>\omega)+\max_{i\leq N}\sum_{j=1}^{N}|\sigma_{ij}|1(|\hat{\sigma}_{ij}|\leq\omega),$$

where we let $\omega\equiv\omega_{NT}$ (and we also write $b\equiv b_{NT}$). Given (a), there exists a constant $C_1>0$ such that $P\left(\bar{A}\right)\to 0$, where

$$A=\left\{\max_{i,j\leq N}|\hat{\sigma}_{ij}-\sigma_{ij}|\leq C_1 b\right\}.$$

Moreover, conditional on A, $\omega<|\hat{\sigma}_{ij}|\leq\max_{i,j\leq N}|\hat{\sigma}_{ij}-\sigma_{ij}|+|\sigma_{ij}|\leq C_1 b+|\sigma_{ij}|$, implying that $|\sigma_{ij}|>\omega-C_1 b=(C-C_1)b>C'\omega$, for some positive constant C' (given condition (b), it suffices to choose $C > C_1$). Similarly, given A, the event $|\hat{\sigma}_{ij}| \leq \omega$ implies $|\sigma_{ij}| \leq C''\omega$ for some constant $C'' > 0$). Thus, conditional on A, we have that

$$\begin{aligned} \rho\left(\tilde{\Sigma} - \Sigma\right) &\leq \max_{i,j\leq N} |\hat{\sigma}_{ij} - \sigma_{ij}| \sum_{j=1}^{N} 1\left(|\sigma_{ij}| > C'\omega\right) + \max_{i\leq N} \sum_{j=1}^{N} |\sigma_{ij}| \, 1\left(|\sigma_{ij}| \leq C''\omega\right) \\ &\leq C_1 b.m_N + C''\omega.m_N \leq C'''m_N\omega, \end{aligned}$$

for some sufficiently large constant $C''' > 0$, given that ω and b are of the same order of magnitude by condition b). Since $P\left(\bar{A}\right) \to 0$, it follows that $P\left(\rho\left(\tilde{\Sigma} - \Sigma\right) > C'''m_N\omega\right) \to 0$, proving the result. ■

Proof of Lemma B.2. The proof of this result follows the proof of Lemma A.3 of Fan et al. (2011). In particular, by the triangle inequality,

$$\max_{i,j\leq N} |\hat{\sigma}_{ij} - \sigma_{ij}| \leq \max_{i,j\leq N} \left| T^{-1} \sum_{t=1}^{T} e_{it}e_{jt} - \sigma_{ij} \right| + \max_{i,j\leq N} \left| T^{-1} \sum_{t=1}^{T} \left(\tilde{e}_{it}\tilde{e}_{jt} - e_{it}e_{jt}\right) \right|.$$

The first term is $O_P\left(\sqrt{\log N/T}\right)$ by Assumption TS(a) whereas the second term is of the same order of magnitude as $\sqrt{\max_{i\leq N} T^{-1} \sum_{t=1}^{T} \left(\tilde{e}_{it} - e_{it}\right)^2}$. To derive this order, note that given the definitions of $\tilde{e}_{it}$ and e_{it},

$$\begin{aligned} \max_{i\leq N} T^{-1} \sum_{t=1}^{T} \left(\tilde{e}_{it} - e_{it}\right)^2 &\leq 2 \max_{i\leq N} \left\|\lambda_i' H^{-1}\right\|^2 \frac{1}{T} \sum_{t=1}^{T} \left\|\tilde{F}_t - HF_t\right\|^2 + 2 \max_{i\leq N} \left\|\tilde{\lambda}_i - H^{-1\prime}\lambda_i\right\|^2 \frac{1}{T} \sum_{t=1}^{T} \left\|\tilde{F}_t\right\|^2 \\ &= O_P\left(1/\delta_{NT}^2\right) + O_P\left(1/N + \log N/T\right) = O_P\left(\frac{1}{N} + \frac{\log N}{T}\right), \end{aligned}$$

where $\delta_{NT}^2 = \min\left(N, T\right)$. The indicated orders of magnitude follow from

$$\frac{1}{T} \sum_{t=1}^{T} \left\|\tilde{F}_t - HF_t\right\|^2 = O_P\left(1/\delta_{NT}^2\right)$$

and

$$\max_{i\leq N} \left\|\tilde{\lambda}_i - H^{-1\prime}\lambda_i\right\|^2 = O_P\left(\frac{1}{N} + \frac{\log N}{T}\right).$$

To see the last result, note that we can write

$$\tilde{\lambda}_i - H^{-1\prime}\lambda_i = \frac{1}{T} HF' \underline{e}_i - \frac{1}{T} \tilde{F}' \left(\tilde{F} - FH'\right) H^{-1\prime}\lambda_i + \frac{1}{T} \left(\tilde{F} - FH'\right) \underline{e}_i \equiv I_{1i} + I_{2i} + I_{3i},$$

where $\underline{e}_i = \left(e_{i1}, \ldots, e_{iT}\right)'$. Thus, we can bound $\max_{i\leq N} \left\|\tilde{\lambda}_i - H^{-1\prime}\lambda_i\right\|^2$ by $\max_{i\leq N} \|I_{ki}\|^2$ for $k = 1, 2, 3$. Starting with the first term, note that

$$\max_{i\leq N} \|I_{1i}\|^2 \leq \|H\|^2 \max_{i\leq N} \left\| \frac{1}{T} \sum_{t=1}^{T} F_t e_{it} \right\|^2 = O_P\left(\frac{\log N}{T}\right),$$

given Assumption TS(b). For the second term,

$$\begin{aligned}\max_{i\leq N}\|I_{2i}\|^2 &\leq \max_{i\leq N}\left\|\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t\left(\tilde{F}_t-HF_t\right)'H^{-1\prime}\lambda_i\right\|^2\\ &\leq \max_{i\leq N}\left\|H^{-1\prime}\lambda_i\right\|^2\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t\right\|^2\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t-HF_t\right\|^2=O_P\left(\frac{1}{\delta_{NT}^2}\right),\end{aligned}$$

whereas for the third term

$$\max_{i\leq N}\|I_{3i}\|^2\leq\max_{i\leq N}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\tilde{F}_t-HF_t\right)e_{it}\right\|^2\leq\frac{1}{T}\sum_{t=1}^{T}\left\|\tilde{F}_t-HF_t\right\|^2\max_{i\leq N}\left(\frac{1}{T}\sum_{t=1}^{T}e_{it}^2\right)=O_P\left(\frac{1}{\delta_{NT}^2}\right).$$

■

Proof of Theorem 3.1. We verify Conditions A* through F* for the CSD bootstrap scheme. Condition A*(a) follows by the fact that $E^*(\eta_t)=0$, whereas (b) is implied by $tr\left(\tilde{\Sigma}\right)/N=O_P(1)$. Indeed, using the properties of the trace operator and the definition of $e_t^*=\tilde{\Sigma}^{1/2}\eta_t$ with $\tilde{\Sigma}^{1/2}\tilde{\Sigma}^{1/2\prime}=\tilde{\Sigma}$, where $\eta_t\sim$ i.i.d.$(0,I_N)$,

$$\gamma_{s,t}^*=\frac{1}{N}E^*\left(tr\left(e_t^*e_s^{*\prime}\right)\right)=tr\left(E^*\left(\frac{1}{N}e_t^*e_s^{*\prime}\right)\right)=\frac{1}{N}tr\left(\tilde{\Sigma}^{1/2}E^*\left(\eta_t\eta_s'\right)\tilde{\Sigma}^{1/2\prime}\right)=\frac{1}{N}tr\left(\tilde{\Sigma}\right)1\left(t=s\right),$$

which implies that

$$\frac{1}{T}\sum_{t,s=1}^{T}\left|\gamma_{s,t}^*\right|^2=\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{N}tr\left(\tilde{\Sigma}\right)\right)^2=\left(\frac{1}{N}tr\left(\tilde{\Sigma}\right)\right)^2.$$

This is $O_P(1)$ provided $\frac{1}{N}tr\left(\tilde{\Sigma}\right)=O_P(1)$. Since for any matrix A, $tr(A)\leq\rho(A)\,rank(A)$, it follows that

$$\frac{1}{N}tr\left(\tilde{\Sigma}\right)\leq\frac{1}{N}\rho\left(\tilde{\Sigma}\right)rank\left(\tilde{\Sigma}\right)\leq\rho\left(\tilde{\Sigma}\right),$$

given that $rank\left(\tilde{\Sigma}\right)\leq N$. By the triangle inequality for matrix norms, $\rho\left(\tilde{\Sigma}\right)\leq\rho\left(\tilde{\Sigma}-\Sigma\right)+\rho(\Sigma)=o_P(1)+O(1)=O_P(1)$, since $\rho\left(\tilde{\Sigma}-\Sigma\right)\to^P 0$ and $\rho(\Sigma)=\lambda_{\max}(\Sigma)=O(1)$ by Assumption CS. For A*(c), from Gonçalves and Perron (2014), it suffices to show that

$$\frac{1}{N}\sum_{i,j=1}^{N}Cov^*\left(e_{it}^*e_{is}^*,e_{jt}^*e_{js}^*\right)=O_P(1),\tag{9}$$

uniformly in (t,s). Letting $e_{it}^*=a_i'\eta_t=\sum_{l=1}^{N}a_{il}\eta_{lt}$, where a_i' denotes the ith row of $\tilde{\Sigma}^{1/2}$, we can write

$$Cov^*\left(e_{it}^*e_{is}^*,e_{jt}^*e_{js}^*\right)=Cov^*\left(a_i'\eta_t\eta_s'a_i,a_j'\eta_t\eta_s'a_j\right)=\sum_{l_1,l_2,l_3,l_4=1}^{N}a_{il_1}a_{il_2}a_{jl_3}a_{jl_4}Cov^*\left(\eta_{l_1t}\eta_{l_2s},\eta_{l_3t}\eta_{l_4s}\right).$$

Using the assumption that the N elements of η_t are mutually independent with mean zero and variance one for each t and the independence of η_t from η_s for $t\neq s$, we can evaluate $Cov^*\left(\eta_{l_1t}\eta_{l_2s},\eta_{l_3t}\eta_{l_4s}\right)$ for all possible combinations of l_1,l_2,l_3,l_4 and t,s. For instance, when $l_1=l_2=l_3=l_4=l$, we have that

$$Cov^*\left(\eta_{l_1t}\eta_{l_2s},\eta_{l_3t}\eta_{l_4s}\right)=Cov^*\left(\eta_{lt}\eta_{ls},\eta_{lt}\eta_{ls}\right)=E^*\left(\eta_{lt}^2\eta_{ls}^2\right)-E\left(\eta_{lt}\eta_{ls}\right)^2,$$

which is equal to $E^*\left(\eta_{lt}^4\right)-1$ when $t=s$ and is equal to 1 when $t\neq s$. Thus, the contribution of this term to (9) is bounded by

$$C\frac{1}{N}\sum_{i,j=1}^{N}\sum_{l=1}^{N}a_{il}^2a_{jl}^2=C\frac{1}{N}\sum_{l=1}^{N}\left(\sum_{i=1}^{N}a_{il}^2\right)^2=C\frac{1}{N}\sum_{l=1}^{N}\left(A_l'A_l\right)^2$$

where $C\geq\max\left(E^*\left(\eta_{lt}^4\right)-1,1\right)$ and $A_l=(a_{1l},\ldots,a_{Nl})'$ is the l^{th} column of $\tilde{\Sigma}^{1/2}$. Proceeding this way, we can show that $Cov^*\left(\eta_{l_1t}\eta_{l_2s},\eta_{l_3t}\eta_{l_4s}\right)\neq 0$ only if two $l_1=l_3\neq l_2=l_4$ or $l_1=l_4\neq l_2=l_3$, implying that he only other contribution to the sum in (9) comes from a term that is bounded by

$$C\frac{1}{N}\sum_{i,j=1}^{N}\sum_{l\neq k}^{N}a_{il}a_{i_k}a_{jl}a_{jk}=C\frac{1}{N}\sum_{l\neq k}^{N}\left(\sum_{i=1}^{N}a_{il}a_{i_k}\right)\left(\sum_{j=1}^{N}a_{jl}a_{jk}\right)=C\frac{1}{N}\sum_{l\neq k}^{N}\left(A_l'A_k\right)^2,$$

from some constants C. Thus, we can bound (9) by

$$C\frac{1}{N}\sum_{l,k=1}^{N}\left(A_l'A_k\right)^2=C\frac{1}{N}\left\|\tilde{\Sigma}\right\|^2,$$

given that $\tilde{\Sigma}=\tilde{\Sigma}^{1/2}\tilde{\Sigma}^{1/2\prime}=\sum_{l=1}^{N}A_lA_l'$ and

$$\left\|\tilde{\Sigma}\right\|^2=tr\left(\tilde{\Sigma}'\tilde{\Sigma}\right)=tr\left(\sum_{k,l=1}^{N}A_lA_l'A_kA_k'\right)=\sum_{k,l=1}^{N}\left(A_l'A_k\right)tr\left(A_lA_k'\right)=\sum_{k,l=1}^{N}\left(A_l'A_k\right)^2.$$

Since $\left\|\tilde{\Sigma}\right\|\leq\rho\left(\tilde{\Sigma}\right)\sqrt{rank\left(\tilde{\Sigma}\right)}\leq\rho\left(\tilde{\Sigma}\right)\sqrt{N}$, it follows that (9) is bounded by $C\left(\rho\left(\tilde{\Sigma}\right)\right)^2=O_P(1)$ given that $\rho\left(\tilde{\Sigma}-\Sigma\right)\to^P 0$ and $\rho(\Sigma)=\lambda_{\max}(\Sigma)=O(1)$. For Condition B*(a), note that $\gamma_{s,t}^*=0$ for $t\neq s$, whereas $\gamma_{t,t}^*=\frac{1}{N}tr\left(\tilde{\Sigma}\right)$, which implies that

$$\frac{1}{T}\sum_{s,t=1}^{T}\tilde{F}_s\tilde{F}_t'\gamma_{s,t}^*=\frac{1}{T}\sum_{t=1}^{T}\tilde{F}_t\tilde{F}_t'\frac{1}{N}tr\left(\tilde{\Sigma}\right)=\frac{\tilde{F}'\tilde{F}}{T}\frac{1}{N}tr\left(\tilde{\Sigma}\right)=\frac{1}{N}tr\left(\tilde{\Sigma}\right)=O_P(1),$$

as shown above (note that $\frac{\tilde{F}'\tilde{F}}{T}=I_r$). To check Condition B*(b), for given t, we can write

$$E^*\left\|\frac{1}{\sqrt{TN}}\sum_{s=1}^{T}\sum_{i=1}^{N}\hat{z}_s\left(e_{it}^*e_{is}^*-E^*\left(e_{it}^*e_{is}^*\right)\right)\right\|^2=\frac{1}{T}\sum_{s,l=1}^{T}\tilde{z}_s'\tilde{z}_l\Delta_{t,l,s},$$

where

$$\Delta_{t,l,s}\equiv\frac{1}{N}\sum_{i,j=1}^{N}Cov^*\left(e_{it}^*e_{is}^*,e_{jt}^*e_{jl}^*\right).$$

Using the same approach as when verifying A*(c), we can show that $\Delta_{t,l,s} = 0$ whenever $l \neq s$, whereas for $l = s$, $\Delta_{t,l,l} \leq \left\| \tilde{\Sigma} \right\|^2 / N$ uniformly in t. Therefore,

$$\frac{1}{T} \sum_{s,l=1}^{T} \tilde{z}_s' \tilde{z}_l \Delta_{t,l,s} = \left(\frac{1}{T} \sum_{l=1}^{T} \tilde{z}_l' \tilde{z}_l \right) \frac{\left\| \tilde{\Sigma} \right\|^2}{N} \leq O_P(1) \rho \left(\tilde{\Sigma} \right)^2 .$$

Thus, B*(b) is implied by the condition that $\rho \left(\tilde{\Sigma} \right) = O_P(1)$, which follows under Assumption CS and the condition that $\rho \left(\tilde{\Sigma} - \Sigma \right) \to^P 0$. For B*(c), using the properties of the trace operator and the definition of the Frobenius norm, we get that

$$E^* \left\| \frac{1}{\sqrt{TN}} \sum_{t=1}^{T} \sum_{i=1}^{N} \tilde{z}_t \tilde{\lambda}_i' e_{it}^* \right\|^2 = E^* \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tilde{z}_t \left(\frac{e_t^{*\prime} \tilde{\Lambda}}{\sqrt{N}} \right) \right\|^2 = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} tr \left(\tilde{z}_s \tilde{z}_t' \right) E^* \left(\frac{e_s^{*\prime} \tilde{\Lambda}}{\sqrt{N}} \frac{\tilde{\Lambda}' e_t^*}{\sqrt{N}} \right),$$

where

$$E^* \left(\frac{e_s^{*\prime} \tilde{\Lambda}}{\sqrt{N}} \frac{\tilde{\Lambda}' e_t^*}{\sqrt{N}} \right) = E^* \left(tr \left(\frac{\tilde{\Lambda}' e_t^*}{\sqrt{N}} \frac{e_s^{*\prime} \tilde{\Lambda}}{\sqrt{N}} \right) \right) = \frac{1}{N} tr \left[\tilde{\Lambda}' E^* \left(e_t^* e_s^{*\prime} \right) \tilde{\Lambda} \right] = \begin{cases} 0 & , t \neq s \\ \frac{1}{N} tr \left(\tilde{\Lambda}' \hat{\Sigma} \tilde{\Lambda} \right) \equiv tr \left(\tilde{\Gamma} \right) & , t = s. \end{cases}$$

Hence,

$$E^* \left\| \frac{1}{\sqrt{TN}} \sum_{t=1}^{T} \sum_{i=1}^{N} \tilde{z}_t \tilde{\lambda}_i' e_{it}^* \right\|^2 = \frac{1}{T} \sum_{t=1}^{T} tr \left(\tilde{z}_t \tilde{z}_t' \right) tr \left(\tilde{\Gamma} \right) = \frac{1}{T} \sum_{t=1}^{T} \left\| \tilde{z}_t \right\|^2 tr \left(\tilde{\Gamma} \right) = O_P(1) \, tr \left(\tilde{\Gamma} \right),$$

which is $O_P(1)$ if $tr \left(\tilde{\Gamma} \right) = O_P(1)$. This condition is implied by Condition F* (which we will verify later) and the fact that $Q \Gamma Q' = O_P(1)$ under our assumptions. Similarly, we can easily show that B*(d) is equivalent to the requirement that $tr \left(\tilde{\Gamma} \right) = O_P(1)$. For B*(e), following Gonçalves and Perron (2014), it suffices to show that

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

Using $e_{it}^* = a_i' \eta_t$ and the independence of η_t over time, we can show that $Cov^* \left(e_{it}^* e_{jt}^*, e_{ls}^* e_{ks}^* \right) = 0$ when $t \neq s$, whereas when $t = s$,

$$\begin{aligned} Cov^* \left(e_{it}^* e_{jt}^*, e_{lt}^* e_{kt}^* \right) &= Cov^* \left(a_i' \eta_t \eta_t' a_j, a_l' \eta_t \eta_t' a_k \right) \\ &= Cov^* \left(\sum_{m_1, m_2 = 1}^{N} a_{im_1} a_{jm_2} \eta_{m_1 t} \eta_{m2t}, \sum_{m_3, m_4 = 1}^{N} a_{lm_3} a_{km_4} \eta_{m_3 t} \eta_{m4t} \right) \\ &= \sum_{m_1, m_2, m_3, m_4 = 1}^{N} a_{im_1} a_{jm_2} a_{lm3} a_{km4} Cov^* \left(\eta_{m_1 t} \eta_{m_2 t}, \eta_{m_3 t} \eta_{m_4 t} \right) \\ &\leq C \sum_{m,n=1}^{N} a_{im} a_{jn} a_{lm} a_{kn} = C \left(\sum_{m=1}^{N} a_{im} a_{lm} \right) \left(\sum_{n=1}^{N} a_{jn} a_{kn} \right) = C \left(a_i' a_l \right) \left(a_j' a_k \right), \end{aligned}$$

given that $Cov^*\left(\eta_{m_1t}\eta_{m_2t}, \eta_{m_3t}\eta_{m_4t}\right) = 0$ whenever more than two indices are equal to each other. Given that a_i' is the i^{th} row of $\tilde{\Sigma}^{1/2}$ and that $\tilde{\Sigma} = \tilde{\Sigma}^{1/2}\tilde{\Sigma}^{1/2\prime}$, we can see that $a_i'a_l = \tilde{\sigma}_{il}$, implying that

$$Var^*\left(A^*\right) \leq C\frac{1}{T^2}\sum_{t=1}^{T}\frac{1}{N^2}\sum_{i,j,k,l=1}^{N}\tilde{\lambda}_i\tilde{\lambda}_j\tilde{\lambda}_l\tilde{\lambda}_k\tilde{\sigma}_{il}^2\tilde{\sigma}_{jk}^2 = \frac{1}{T}\left(\frac{1}{N}\sum_{i,l=1}^{N}\tilde{\lambda}_i\tilde{\lambda}_l\tilde{\sigma}_{il}\right)^2 = O_P\left(\frac{1}{T}\right) = o_P\left(1\right),$$

since the term in parenthesis is equal to $\tilde{\Gamma} = O_P\left(1\right)$. Next, we verify Condition C*. Using the independence between ε_{t+1}^* and e_t^*, we can show that part a) is implied by the condition that $\left\|\tilde{\Sigma}\right\|^2/N \leq \rho\left(\tilde{\Sigma}\right) = O_P\left(1\right)$, part b) follows by the condition that $tr\left(\tilde{\Gamma}\right) = O_P\left(1\right)$ and part c) follows by $tr\left(\tilde{\Sigma}/N\right) \leq \rho\left(\tilde{\Sigma}\right) = O_P\left(1\right)$, which are implied by the convergence condition $\rho\left(\tilde{\Sigma} - \Sigma\right) \to 0$ and Condition F*, which we show next. Note that Conditions D* and E* are satisfied under Assumptions 1-5, as shown by Gonçalves and Perron (2014) (since the CSD bootstrap algorithm utilizes the same procedure as theirs to generate ε_{t+1}^* and Condition D* and E* only involve these bootstrap residuals).

To conclude the proof, we show Condition F*. Letting $\bar{\Gamma} = \frac{\Lambda'\tilde{\Sigma}\Lambda}{N}$, we have that

$$\tilde{\Gamma} - Q\Gamma Q' = \tilde{\Gamma} - Q\bar{\Gamma}Q' + Q\left(\bar{\Gamma} - \Gamma\right)Q' \equiv A_1 + A_2.$$

We can write

$$A_2 = Q\left(\bar{\Gamma} - \Gamma\right)Q' = Q\frac{1}{N}\Lambda'\left(\tilde{\Sigma} - \Sigma\right)\Lambda Q',$$

implying that

$$\left\|A_2\right\| \leq \left\|Q\right\|^2\left\|\Lambda/\sqrt{N}\right\|^2\rho\left(\tilde{\Sigma} - \Sigma\right) \to^P 0$$

given that $Q = H^{-1\prime} = O_P\left(1\right)$, $\left\|\Lambda/\sqrt{N}\right\| = O\left(1\right)$ and $\rho\left(\tilde{\Sigma} - \Sigma\right) \to^P 0$. For A_1, adding and subtracting appropriately yields

$$A_1 = a_1 + a_2 + a_2', \text{ where}$$

$$a_1 \equiv \frac{1}{N}\left(\tilde{\Lambda} - \Lambda H^{-1}\right)'\tilde{\Sigma}\left(\tilde{\Lambda} - \Lambda H^{-1}\right) \quad \text{and} \quad a_2 \equiv \frac{1}{N}H^{-1\prime}\Lambda'\tilde{\Sigma}\left(\tilde{\Lambda} - \Lambda H^{-1}\right).$$

We have that

$$\left\|a_1\right\| \leq \left\|\frac{1}{\sqrt{N}}\left(\tilde{\Lambda} - \Lambda H^{-1}\right)\right\|^2\rho\left(\tilde{\Sigma}\right) = o_P\left(1\right)O_P\left(1\right),$$

since the first factor is equal to $\frac{1}{N}\sum_{i=1}^{N}\left\|\tilde{\lambda}_i - H^{-1\prime}\lambda_i\right\|^2 = O_P\left(1/\delta_{NT}^2\right) = o_P\left(1\right)$ under Assumptions 1-5 whereas $\rho\left(\tilde{\Sigma}\right) = O_P\left(1\right)$ given that we choose $\tilde{\Sigma}$ such that $\rho\left(\tilde{\Sigma} - \Sigma\right) \to^P 0$. For a_2,

$$\left\|a_2\right\| \leq \left\|H^{-1}\right\|\left\|\Lambda/\sqrt{N}\right\|\rho\left(\tilde{\Sigma}\right)\left(\left\|\frac{1}{\sqrt{N}}\left(\tilde{\Lambda} - \Lambda H^{-1}\right)\right\|^2\right)^{1/2} = o_P\left(1\right),$$

using the same arguments as for a_1. ■

Proof of Theorem 3.2. Given Theorem 3.1, it suffices to show that $\rho\left(\tilde{\Sigma} - \Sigma\right) \to^P 0$, which is implied by Lemmas B.1 and B.2 given our assumptions. ■

References

[1] Cai, T. and W. Liu, 2011. "Adaptive thresholding for sparse covariance matrix estimation", *Journal of the American Statistical Association*, Vol. 106, 672-684.

[2] Chamberlain, G., and M. Rothschild, 1983. "Arbitrage, Factor Structure and Mean–Variance Analysis in Large Asset Markets," *Econometrica*, 51, 1281-1304.

[3] Bai, J., 2003. "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135-172.

[4] Bai, J. and S. Ng, 2006. "Confidence Intervals for Diffusion Index Forecasts and Inference with Factor-augmented Regressions," *Econometrica*, 74, 1133-1150.

[5] Bai, J. and S. Ng, 2013. "Principal Components Estimation and Identification of the Factors", *Journal of Econometrics*, 176, 18-29.

[6] Bickel, P. and E. Levina, 2008. "Covariance regularization by thresholding", *Annals of Statistics*, Vol. 36, 2577-2604.

[7] Eichengreen, B. A. Mody, M. Nedeljkovic, and L. Sarno, 2012. "How the Subprime Crisis Went Global: Evidence from Bank Credit Default Swap Spreads", *Journal of International Money and Finance*, 31, 1299-1318.

[8] Fan, J., Liao, Y. and M. Mincheva, 2011. "High-dimensional covariance matrix estimation in approximate factor models", *Annals of Statistics*, Vol. 39, 3320-3356.

[9] Fan, J., Liao, Y. and M. Mincheva, 2013. "Large covariance estimation by thresholding principal orthogonal complements", *Journal of the Royal Statistical Society*, Series B, 75, 603-690.

[10] Fan, J., Liao, Y. and J. Yao, 2015. "Power enhancement in high-dimensional cross-sectional tests", *Econometrica*, Vol. 83, 1497-1541.

[11] Gagliardini, P., E. Ossola and O. Scaillet, 2015. "Time-varying risk premium in large cross sectional equity datasets", *Econometrica*, forthcoming.

[12] Gonçalves. S., 2011. " The moving blocks bootstrap for panel linear regression models with individual fixed effects", *Econometric Theory*, 27, 1048-1082.

[13] Gonçalves, S. and B. Perron, 2014. "Bootstrapping Factor-Augmented Regressions", *Journal of Econometrics*, 182, 156-173.

[14] Gospodinov, N. and S. Ng, 2013. "Commodity Prices, Convenience Yields, and Inflation", *Review of Economics and Statistics*, 95, 206–219.

[15] Ludvigson, S. and S. Ng, 2007. “The empirical risk return relation: a factor analysis approach,” *Journal of Financial Economics*, 83, 171-222.

[16] Ludvigson, S. and S. Ng, 2009. “Macro factors in bond risk premia,” *Review of Financial Studies*, 22, 5027-5067.

[17] Ludvigson, S. and S. Ng, 2011. “A factor analysis of bond risk premia," in Ullah, A. and D. Giles eds., Handbook of Empirical Economics and Finance, Chapman and Hall, 313-372.

[18] McMurry, T. and D. Politis, 2010. “Banded and tapered estimates for autocovariance matrices and the linear process bootstrap”, *Journal of Time Series Analysis*, vol. 31, 471-482, 2010.

[19] Neely, C., D. Rapach, J. Tu and , and G. Zhou, 2015. “Forecasting the equity risk premium: the role of technical indicators”, *Management Science*, 60, 1772-1791.

[20] Rothman, A., E. Levina and J. Zhu, 2009. “Generalized thresholding of large covariance matrices”, *Journal of the American Statistical Association*, 97, 104, 177-186.

[21] Stock, J. H. and M. Watson, 2002. “Forecasting using Principal Components from a Large Number of Predictors,” *Journal of the American Statistical Association*, 97, 1167-1179.

Table 1: Bias and coverage rate of 95% CIs for delta

		N = 50			N = 100			N = 200		
		T = 50	T = 100	T = 200	T = 50	T = 100	T = 200	T = 50	T = 100	T = 200
						Bias				
	bias	-0.13	-0.12	-0.11	-0.08	-0.07	-0.07	-0.05	-0.04	-0.04
	BL	-0.04	-0.06	-0.07	-0.03	-0.03	-0.04	-0.01	-0.02	-0.02
	CS-HAC	-0.07	-0.07	-0.07	-0.04	-0.04	-0.04	-0.02	-0.02	-0.02
	CSD	-0.06	-0.06	-0.07	-0.04	-0.04	-0.04	-0.03	-0.03	-0.03
	WB	-0.05	-0.05	-0.04	-0.04	-0.03	-0.03	-0.03	-0.02	-0.02
	iid over time	-0.02	-0.01	-0.01	-0.02	-0.01	-0.01	-0.02	-0.01	-0.01
	empirical	-0.02	-0.01	-0.01	-0.02	-0.01	-0.01	-0.02	-0.01	-0.01
DGP 1						Coverage rate				
alpha = 1										
CS	OLS	70.5	64.9	52.8	81.5	83.0	79.8	87.6	89.4	89.0
	BC - BL	77.9	79.4	81.4	83.7	88.3	89.0	88.6	90.7	91.5
	BC- CS-HAC	79.8	80.8	80.6	83.9	88.5	88.4	88.9	90.6	91.5
	CSD	87.9	87.6	88.5	90.2	92.7	92.1	92.1	93.7	93.6
	WB	87.0	83.1	78.2	89.9	91.4	89.0	92.1	93.1	92.6
	iid over time	79.1	70.6	56.7	87.5	87.7	82.4	91.4	92.2	90.9
	empirical	79.8	70.9	57.0	87.2	87.7	82.4	91.4	92.3	91.0
						Bias				
	bias	-0.13	-0.12	-0.11	-0.08	-0.07	-0.07	-0.05	-0.04	-0.03
	BL	-0.04	-0.06	-0.07	-0.03	-0.03	-0.04	-0.01	-0.02	-0.02
	CS-HAC	-0.03	-0.03	-0.03	-0.02	-0.02	-0.02	-0.01	-0.01	-0.01
	WB	-0.05	-0.05	-0.04	-0.04	-0.03	-0.03	-0.03	-0.02	-0.02
	iid over time	-0.02	-0.01	-0.01	-0.02	-0.01	-0.01	-0.02	-0.01	-0.01
	empirical	-0.02	-0.01	-0.01	-0.02	-0.01	-0.01	-0.02	-0.01	-0.01
	CSD	-0.06	-0.06	-0.07	-0.04	-0.04	-0.04	-0.03	-0.03	-0.03
DGP 2						Coverage rate				
alpha = 1										
CS	OLS	70.7	65.1	52.8	81.1	83.2	79.5	87.2	89.5	89.9
	BC - BL	78.1	80.1	81.2	83.9	88.3	88.8	87.9	90.7	92.5
	BC- CS-HAC	76.6	74.9	68.9	83.2	86.7	85.4	87.9	90.3	91.5
reshuffled X										
	CSD	87.5	87.6	88.3	89.9	92.2	91.9	92.1	93.2	94.0
	WB	86.9	83.6	77.9	89.9	91.4	89.0	92.2	92.9	93.3
	iid over time	79.8	70.9	56.4	87.5	87.4	82.0	91.5	92.0	91.7
	empirical	79.9	71.3	56.6	87.4	87.7	82.1	91.3	92.0	91.8

Each part of the table reports estimates of the bias in the estimation of α and the associated coverage rate of 95% confidence intervals for two asymptotic and 4 bootstrap methods. The asymptotic methods are the OLS estimator and two bias-corrected estimators obtained by plugging in the Bickel-Levina (2008) or CS-HAC of Bai and Ng (2006) estimators. The four bootstrap methods are the cross-sectional bootstrap with Bickel-Levina (2008) estimator, he wild bootstrap of Gonçalves and Perron (2014), the bootstrap that resamples vectors indepedently over time, and the cross-sectional bootstrap using the empirical covariance matrix. All results are based on 5000 replications and B=399 bootstraps.

Table 2. Estimation results for augmented inflation equation

pccy_1		0.085		qoil_t		1.184	
	Asymptotic	0.03	0.14		Asymptotic	0.91	1.46
	CSD	0.06	0.19		CSD	0.93	1.49
	wild bootstrap	0.09	0.24		wild bootstrap	0.93	1.49
	GN	0.00	0.18		GN	0.90	1.47
pccy_2		-0.087		qoil_t-1		-1.095	
	Asymptotic	-0.13	-0.04		Asymptotic	-1.37	-0.82
	CSD	-0.21	-0.10		CSD	-1.39	-0.82
	wild bootstrap	-0.23	-0.11		wild bootstrap	-1.40	-0.82
	GN	-0.19	-0.02		GN	-1.36	-0.83
pcq_1		-0.050		i_t		0.021	
	Asymptotic	-0.09	-0.01		Asymptotic	0.01	0.03
	CSD	-0.13	-0.05		CSD	0.02	0.03
	wild bootstrap	-0.14	-0.06		wild bootstrap	0.02	0.04
	GN	-0.11	0.01		GN	0.01	0.03
pcq_2		0.039		Δx		0.008	
	Asymptotic	-0.02	0.10		Asymptotic	0.00	0.02
	CSD	0.01	0.15		CSD	0.00	0.02
	wild bootstrap	0.02	0.16		wild bootstrap	0.00	0.02
	GN	-0.03	0.10		GN	0.00	0.02
Δp_t		0.177		gap		-0.010	
	Asymptotic	0.07	0.28		Asymptotic	-0.08	0.06
	CSD	0.04	0.27		CSD	-0.06	0.09
	wild bootstrap	0.05	0.27		wild bootstrap	-0.06	0.09
	GN	0.08	0.29		GN	-0.09	0.07
Δp_t-1		-0.138				0.149	
	Asymptotic	-0.26	-0.01	constant	Asymptotic	0.10	0.20
	CSD	-0.28	-0.01		CSD	0.09	0.19
	wild bootstrap	-0.28	-0.01		wild bootstrap	0.08	0.18
	GN	-0.25	-0.04		GN	0.08	0.22

Each panel contains the results for a parameter of the basic inflation equation. The first line gives the OLS point estimates, and the following lines are 90% confidence intervals obtained using the Bai and Ng (2006) asymptotic theory, the cross-sectional bootstrap with hard threshold estimator, the wild bootstrap, and the Gospodinov and Ng (2013) boostrap with block size equal to 1. All bootstrap intervals are equal-tailed and were obtained with 4,999 bootstrap replications.