# Consistent Factor Estimation in Dynamic Factor Models with Structural Instability

Brandon J. Bates

Harvard University

bbates@fas.harvard.edu

Mikkel Plagborg-Møller

Harvard University

plagborg@fas.harvard.edu

James H. Stock

Harvard University

james_stock@harvard.edu

Mark W. Watson

Princeton University

mwatson@princeton.edu

June 20, 2012

**Abstract**

This paper considers the estimation of approximate dynamic factor models when there is temporal instability in the factor loadings. We characterize the type and magnitude of instabilities under which the principal components estimator of the factors is consistent, and find that these instabilities can be larger than earlier theoretical calculations suggest. We further characterize the rate of convergence of the estimated factors as a function of the magnitude of the time variation in the factor loadings for general types of parameter instability, and provide numerical evidence that this consistency rate is tight in the special case of random walk parameter variation. We also discuss implications of these results for the robustness of regressions based on the estimated factors and of estimates of the number of factors in the presence of parameter instability.

## 1  Introduction

Dynamic factor models (DFMs) provide a flexible framework for simultaneously modeling a large number of macroeconomic time series.[1] In a DFM,

---

[1]The early work on DFMs considered a small number of time series. DFMs were introduced by Geweke (1977), and early low-dimensional applications include Sargent and Sims (1977), Engle and Watson (1981), Watson and Engle (1983), Sargent (1989) and Stock and Watson (1989). Work over the past fifteen years has focused on methods that

a potentially large number of observed time series variables are modeled as depending on a small number of unobserved factors, which account for the widespread co-movements of the observed series. Although there is now a large body of theory for the analysis of high-dimensional DFMs, nearly all of this theory has been developed for the case in which the DFM parameters are stable, in particular, in which there are no changes in the factor loadings (the coefficients on the factors); among the few exceptions are Stock and Watson (2002, 2009) and Breitung and Eickmeier (2011). This assumption of parameter stability, however, is at odds with broad evidence of time variation in many macroeconomic forecasting relations.

The goal of this paper is to characterize the type and magnitude of parameter instability that can be tolerated by a standard estimator of the factors, the principal components estimator, in a DFM when the coefficients of the model are unstable. In so doing, this paper contributes to a larger debate about how best to handle the instability that is widespread in macroeconomic forecasting relations. On the one hand, the conventional wisdom is that time series forecasts deteriorate when there are undetected structural breaks or unmodeled time-varying parameters, see for example Clements and Hendry (1998). This view underlies the large literatures on the detection of breaks and on models that incorporate breaks and time variation, for example by modeling the breaks as following a Markov process (Hamilton, 1989; Pesaran et al., 2006). In the context of DFMs, Breitung and Eickmeier (2011) show that a one-time structural break in the factor loadings has the effect of introducing new factors, so that estimation of the factors ignoring the break leads to estimating too many factors.

On the other hand, a few recent papers have provided evidence that sometimes it can be better to ignore parameter instability when forecasting. Pesaran and Timmermann (2005) point out that whether to use pre-break data for estimating an autoregression trades off an increase in bias against a reduction in estimator variance, and they provide empirical evidence supporting the use of pre-break data for forecasting. Pesaran and Timmermann (2007) go on to provide tools to help ascertain in practice whether pre-break data should be used for estimation of single-equation time series forecasting models. In DFMs, Stock and Watson (2009) provide empirical evidence using U.S. macroeconomic data from 1960-2007 that full-sample estimates of the factors are preferable to subsample estimates, despite clear evidence of a

---

facilitate the analysis of a large number of time series, see Forni et al. (2000) and Stock and Watson (2002) for early contributions. For recent contributions and discussions of this large literature see Bai and Ng (2008$a$), Eickmeier and Ziegler (2008), Chudik and Pesaran (2011) and Stock and Watson (2011).

break in many factor loadings around the beginning of the Great Moderation in 1984.

We therefore seek a precise theoretical understanding of the effect of instability in the factor loadings on the performance of principal components estimators of the factors. Specifically, we consider a DFM with $N$ variables observed for $T$ time periods and $r \ll N$ factors, where the $N \times r$ matrix of dynamic factor loadings $\Lambda$ can vary over time. We write this time variation so that $\Lambda$ at date $t$ equals its value at date 0, plus a deviation; that is, $\Lambda_t = \Lambda_0 + h_{NT}\xi_t$. The term $\xi_t$ is a possibly random disturbance, and $h_{NT}$ is a deterministic scalar sequence in $N$ and $T$ which sets the scale of the deviation. Using this framework and standard assumptions in the literature (Bai and Ng, 2002, 2006), we obtain general conditions on $h_{NT}$ under which the principal components estimates are mean square consistent for the space spanned by the true factors. We then specialize these general results to three leading cases: i.i.d. deviations of $\Lambda_t$ from $\Lambda_0$, random walk deviations that are independent across series, and an arbitrary one-time break that affects some or all of the series.

For the case in which $\Lambda_t$ is a vector of independent random walks, Stock and Watson (2002) showed that the factor estimates are consistent if $h_{NT} = T^{-1}$. By using a different method of proof (which builds on Bai and Ng, 2002), we are able to weaken this result considerably and show that the estimated factors are consistent if $h_{NT} = o(T^{-1/2})$. We further show that, if $h_{NT} = O(1/\min\{N^{1/4}T^{1/2}, T^{3/4}\})$, the estimated factors achieve the mean square consistency rate of $1/\min\{N, T\}$, a rate initially shown by Bai and Ng (2002) in the case of no time variation. Because $\xi_t$ in the random walk case is itself $O(t^{1/2})$, this means that deviations in the factor loadings on the order of $o_p(1)$ do not break the consistency of the principal components estimator. These rates are remarkable: as a comparison, if the factors were observed so an efficient test for time variation could be performed, the test would have nontrivial power against random walk deviations in a $h_{NT} \propto T^{-1}$ neighborhood of zero (e.g., Stock and Watson, 1998b) and would have power of one against parameter deviations of the magnitude tolerated by the principal components estimator. Intuitively, the reason that the principal component estimator can handle such large changes in the coefficients is that, if these shifts have limited dependence across series, their effect can be reduced, and eliminated asymptotically, by averaging across series.

We further provide the rate of mean square consistency as a function of $h_{NT}$, both in general and specialized to the random walk case. The resulting consistency rate function is nonlinear and reflects the tradeoff between the magnitude of the instability and, through the relative rate of $N/T$ as

3

$T$ increases, the amount of cross-sectional information that can be used to "average out" this instability. Although the bounds from which the consistency rate function is derived are tight using our method of proof, we cannot show that these rate function bounds provide necessary as well as sufficient conditions for consistency. We find numerically, however, that our theoretical consistency rate function matches Monte Carlo estimates of the rate functions, which suggests that no other method of proof could improve on these bounds (under our assumptions).

The rest of the paper proceeds as follows. Section 2 lays out the model, the assumptions, and the three special cases. Asymptotic results are provided in Section 3. Section 4 provides Monte Carlo results, and Section 5 concludes.

## 2 Model and assumptions

### 2.1 Basic model and intuition

The model and notation follow Bai and Ng (2002) closely. Denote the observed data by $X_{it}$ for $i = 1, \ldots, N$, $t = 1, \ldots, T$. It is assumed that the observed series are driven by a small number $r$ of unobserved common factors $F_{pt}$, $p = 1, \ldots, r$, such that

$$X_{it} = \lambda_{it}'F_t + e_{it}.$$

Here $\lambda_{it} \in \mathbb{R}^r$ is the possibly time-varying factor loading of data series $i$ at time $t$, while $e_{it}$ is an idiosyncratic error. Define $X_t = (X_{1t}, \ldots, X_{Nt})'$, $e_t = (e_{1t}, \ldots, e_{Nt})'$, $\Lambda_t = (\lambda_{1t}, \ldots, \lambda_{Nt})'$ and data matrices $X = (X_1, \ldots, X_T)'$, $F = (F_1, \ldots, F_T)'$. The initial factor loadings $\Lambda_0$ are fixed. We write the cumulative drift in the parameter loadings as

$$\Lambda_t - \Lambda_0 = h_{NT}\xi_t,$$

where $h_{NT}$ is a deterministic scalar that may depend on $N$ and $T$, while $\{\xi_t\}$ is a possibly degenerate random process of dimension $N \times r$ (in fact, it will be allowed to be a triangular array). Observe that

$$X_t = \Lambda_t F_t + e_t = \Lambda_0 F_t + e_t + w_t, \tag{1}$$

where $w_t = h_{NT}\xi_t F_t$. Our proof technique will be to treat $w_t$ as another error term in the factor model.

To establish some intuition for why estimation of the factors is possible despite structural instability, consider an independent random walk model

for the time variation in the factor loadings, so that $\xi_{it} = \xi_{i,t-1} + \zeta_{it}$, where $\zeta_{it}$ is i.i.d. across $i$ and $t$ with mean 0 and variance $\sigma_\zeta^2$, and suppose that $\Lambda_0$ is known. In addition, we look ahead to Assumption 2 and assume that $\Lambda_0'\Lambda_0/N \to D$, where $D$ has full rank. Because $\Lambda_0$ is known, we can consider the estimator $\hat{F}_t(\Lambda_0) = (\Lambda_0'\Lambda_0)^{-1}\Lambda_0'X_t$. From (1),

$$\hat{F}_t(\Lambda_0) = F_t + (\Lambda_0'\Lambda_0)^{-1}\Lambda_0'e_t + (\Lambda_0'\Lambda_0)^{-1}\Lambda_0'w_t,$$

so

$$\hat{F}_t(\Lambda_0) - F_t \approx D^{-1}N^{-1}\sum_{i=1}^{N}\lambda_{i0}e_{it} + D^{-1}N^{-1}\sum_{i=1}^{N}\lambda_{i0}w_{it}.$$

The first term does not involve time-varying factor loadings and under limited cross-sectional dependence it is $O_p(N^{-1/2})$. Using the definition of $w_t$, the second term can be written

$$D^{-1}N^{-1}\sum_{i=1}^{N}\lambda_{i0}w_{it} = D^{-1}\left(h_{NT}N^{-1}\sum_{i=1}^{N}\lambda_{i0}\xi_{it}\right)'F_t.$$

Since $F_t$ is $O_p(1)$, this second term is the same order as the first, $O_p(N^{-1/2})$, if $h_{NT}N^{-1}\sum_{i=1}^{N}\lambda_{i0}\xi_{it}$ is $O_p(N^{-1/2})$. Under the independent random walk model, $\xi_{it} = O_p(T^{-1/2})$, so

$$h_{NT}N^{-1}\sum_{i=1}^{N}\lambda_{i0}\xi_{it} = O_p(h_{NT}(T/N)^{1/2}),$$

which in turn is $O_p(N^{-1/2})$ if $h_{NT} = O(T^{-1/2})$. This informal reasoning suggests that the estimator $\hat{F}_t(\Lambda_0)$ satisfies $\hat{F}_t(\Lambda_0) = F_t + O_p(N^{-1/2})$ if $h_{NT} = cT^{-1/2}$.

In practice $\Lambda_0$ is not known so $\hat{F}_t(\Lambda_0)$ is not feasible. The principal components estimator of $F_t$ is $\hat{F}_t(\hat{\Lambda}^r)$, where $\hat{\Lambda}^r$ is the matrix of eigenvectors corresponding to the first $r$ eigenvalues of the sample second moment matrix of $X_t$. The calculations below suggest that the estimation of $\Lambda_0$ by $\hat{\Lambda}^r$ reduces the amount of time variation that can be tolerated in the independent random walk case; setting $h_{NT} = cT^{-1/2}$ results in an $o_p(1)$ mean square discrepancy between $\hat{F}_t(\hat{\Lambda}^r)$ and $F_t$.

## 2.2 Examples of structural instability

For concreteness, we highlight three special cases that will receive extra attention in the following analysis. In these examples, the scalar $h_{NT}$ is left unspecified for now. We will set the number of factors $r$ to 1 for ease of exposition.

**Example 1 (white noise).** All entries $\xi_{it}$ are i.i.d. across $i$ and $t$ with mean zero and $E(\xi_{it}^4) < \infty$. The factor loadings $\Lambda_t$ are then equal to the initial loading matrix $\Lambda_0$ plus uncorrelated noise.

**Example 2 (random walk).** Entries $\xi_{it}$ are given by $\xi_{it} = \sum_{s=1}^{t} \zeta_{is}$, where $\{\zeta_{is}\}$ is a random process that is i.i.d. across $i$ and $s$ with mean zero and $E(\zeta_{is}^4) < \infty$. In this example, the factor loadings evolve as cross-sectionally uncorrelated random walks.[2] Models of this type are often referred to as time-varying parameter models in the literature.

**Example 3 (single deterministic break).** Let $\bar{\tau} \in (0,1)$ be fixed and set $\kappa = [\bar{\tau}T]$, where $[\,\cdot\,]$ denotes the integer part. Let $\Delta \in \mathbb{R}^N$ be a shift parameter. We then define

$$\xi_t = \left\{ \begin{array}{ll} 0 & \text{for } t = 1,\ldots,\kappa \\ \Delta & \text{for } t = \kappa + 1, \ldots, T \end{array} \right. .$$

Such deterministic shifts have been extensively studied in the context of structural break tests in the linear regression model.

## 2.3 Principal components estimation

We are interested in the properties of the principal components estimator of the factors, where estimation is carried out as if the factor loadings were constant over time. Let $k$ denote the number of factors that are estimated. The principal component estimators of the loadings and factors are obtained by solving the minimization problem

$$V(k) = \min_{\Lambda^k, F^k} (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \lambda_i^{k'} F_t^k)^2,$$

where the supercripts on $\Lambda^k$ and $F^k$ signify that there are $k$ estimated factors. It is necessary to impose a normalization on the estimators to pin down the minimizers (see Bai and Ng, 2008b, for a thorough treatment). Such restrictions are innocuous since the unobserved true factors $F$ are only identifiable up to multiplication by a non-singular matrix. One estimator

---

[2]While conceptually clear, cross-sectional independence of the random walk innovations $\zeta_{it}$ is a stricter assumption than necessary for the subsequent treatment. It is straightforward to modify the example to allow $m$-dependence or exponentially decreasing correlation across $i$, and all the results below go through for these modifications.

of $F$ is obtained by first concentrating out $\Lambda^k$ and imposing the normalization $F^{k'}F^k/T = I_k$. The resulting estimator $\tilde{F}^k$ is given by $\sqrt{T}$ times the matrix of eigenvectors corresponding to the largest $k$ eigenvalues of the matrix $XX'$. A second estimator is obtained by first concentrating out $F^k$ and imposing the normalization $\Lambda^{k'}\Lambda^k/N = I_k$. This estimator equals $\bar{F}^k = X\bar{\Lambda}^k/N$, where $\bar{\Lambda}^k$ is $\sqrt{N}$ times the eigenvectors corresponding to the $k$ largest eigenvalues of $X'X$. Following Bai and Ng (2002), we use a rescaled estimator

$$\hat{F}^k = \bar{F}^k(\bar{F}^{k'}\bar{F}^k/T)^{1/2}$$

in the following.

## 2.4 Assumptions

Our assumptions on the factors, initial loadings and the idiosyncratic errors are the same as in Bai and Ng (2002). The matrix norm is chosen to be the Frobenius norm $\|A\| = [\mathrm{tr}(A'A)]^{1/2}$. The subscripts $i, j$ will denote cross-sectional indices, $s, t$ will denote time indices and $p, q$ will denote factor indices. $M \in (0, \infty)$ is a constant that is common to all the assumptions below. Finally, define $C_{NT} = \min\{N^{1/2}, T^{1/2}\}$. The following are Assumptions A–C in Bai and Ng (2002).

**Assumption 1** (Factors). *$E\|F_t\|^4 \le M$ and $T^{-1}\sum_{t=1}^{T} F_t F_t' \to \Sigma_F$ as $T \to \infty$ for some positive definite matrix $\Sigma_F$.*

**Assumption 2** (Initial factor loadings). *$\|\lambda_{i0}\| \le \bar{\lambda} < \infty$, and $\|\Lambda_0'\Lambda_0/N - D\| \to 0$ as $N \to \infty$ for some positive definite matrix $D \in \mathbb{R}^{r \times r}$.*

**Assumption 3** (Idiosyncratic errors). *The following conditions hold for all $N$ and $T$.*

1. *$E(e_{it}) = 0$, $E|e_{it}|^8 \le M$.*

2. *$\gamma_N(s, t) = E(e_s' e_t/N)$ exists for all $(s, t)$. $|\gamma_N(s, s)| \le M$ for all $s$, and $T^{-1}\sum_{s,t=1}^{T} |\gamma_N(s, t)| \le M$.*

3. *$\tau_{ij,ts} = E(e_{it}e_{js})$ exists for all $(i, j, s, t)$. $|\tau_{ij,tt}| \le |\tau_{ij}|$ for some $\tau_{ij}$ and for all $t$, while $N^{-1}\sum_{i,j=1}^{N} |\tau_{ij}| \le M$. Additionally,*

$$(NT)^{-1} \sum_{i,j=1}^{N} \sum_{s,t=1}^{T} |\tau_{ij,ts}| \le M.$$

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

As mentioned by Bai and Ng (2002), the above assumptions allow for weak cross-sectional and time dependence of the idiosyncratic errors. Note that the factors do not need to be stationary to satisfy Assumption 1.

The assumptions we need on the factor loading innovations $h_{NT}\xi_t$ are summarized below. For now we require the existence of three envelope functions that bound the rates, in terms of $N$ and $T$, at which certain sums of higher moments diverge. As we later state in Theorem 1, these rates determine the convergence rate of the principal components estimator of the factors.

**Assumption 4** (Factor loading innovations). *There exist envelope functions $Q_1(N,T)$, $Q_2(N,T)$ and $Q_3(N,T)$ such that the following conditions hold for all $N$, $T$ and factor indices $p_1, q_1, p_2, q_2 = 1, \ldots, r$.*

1. $\sup_{s,t\leq T}\sum_{i,j=1}^{N}|E(\xi_{isp_1}\xi_{jtq_1}F_{sp_1}F_{tq_1})| \leq Q_1(N,T)$.

2. $\sum_{s,t=1}^{T}\sum_{i,j=1}^{N}|E(\xi_{isp_1}\xi_{jsq_1}F_{sp_1}F_{sq_1}F_{tp_2}F_{tq_2})| \leq Q_2(N,T)$.

3. $\sum_{s,t=1}^{T}\sum_{i,j=1}^{N}|E(\xi_{isp_1}\xi_{jsq_1}\xi_{itp_2}\xi_{jtq_2}F_{sp_1}F_{sq_1}F_{tp_2}F_{tq_2})| \leq Q_3(N,T)$.

While consistency of the principal components estimator will require limited dependence between the factor loading innovations and the factors themselves, full independence is not necessary. This is empirically appealing, as it is reasonable to expect that breaks in the factor relationships may occur at times when the factors deviate substantially from their long-run means. That being said, we remark that if the processes $\{\xi_t\}$ and $\{F_t\}$ are assumed to be independent (and given Assumption 1), two sufficient conditions for Assumption 4 are that there exist envelope functions $\tilde{Q}_1(N,T)$ and $\tilde{Q}_3(N,T)$ such that for all factor indices,

$$\sup_{s,t\leq T}\sum_{i,j=1}^{N}|E(\xi_{isp_1}\xi_{jtq_1})| \leq \tilde{Q}_1(N,T) \tag{2}$$

and

$$\sum_{s,t=1}^{T}\sum_{i,j=1}^{N}|E(\xi_{isp_1}\xi_{jsq_1}\xi_{itp_2}\xi_{jtq_2})| \leq \tilde{Q}_3(N,T). \tag{3}$$

Under the above conditions, Assumption 4 holds with $Q_1(N,T) \propto \tilde{Q}_1(N,T)$, $Q_2(N,T) \propto T^2\tilde{Q}_1(N,T)$ and $Q_3(N,T) \propto \tilde{Q}_3(N,T)$.

**Examples (continued).** For Examples 1 and 2 (white noise and random walk), assume that $\{\xi_t\}$ and $\{F_t\}$ are independent.

In Example 1 (white noise), the supremum on the left-hand side of (2) reduces to $NE(\xi_{it}^2)$. By writing out terms, it may be verified that the quadruple sum in condition (3) is bounded by an $O(NT^2) + O(N^2T)$ expression. Consequently, Assumption 4 holds with $Q_1(N,T) = O(N)$, $Q_2(N,T) = O(NT^2)$ and $Q_3(N,T) = O(NT^2) + O(N^2T)$.

In Example 2 (random walk), due to cross-sectional i.i.d.-ness we obtain

$$
\sup_{s,t\leq T} \sum_{i=1}^{N} \sum_{j=1}^{N} |E(\xi_{is}\xi_{jt})| = N \sup_{s,t\leq T} |E(\xi_{is}\xi_{it})|
$$

$$
= N \sup_{s,t\leq T} \left| \sum_{u=1}^{s} \sum_{v=1}^{t} E(\zeta_{iu}\zeta_{iv}) \right|
$$

$$
= N \sup_{s,t\leq T} \min\{s,t\} E(\zeta_{i1}^2)
$$

$$
= O(NT),
$$

so Assumptions 4.1–4.2 hold with $Q_1(N,T) = O(NT)$ and $Q_2(N,T) = O(NT^3)$. A somewhat lengthier calculation gives that the quadruple sum in condition (3) is $O(N^2T^4)$ (a rate that cannot be improved upon), so Assumption 4.3 holds with $Q_3(N,T) = O(N^2T^4)$.

In Example 3 (single deterministic break), the supremum in inequality (2) evaluates as

$$
\sum_{i=1}^{N} |\Delta_i| \sum_{j=1}^{N} |\Delta_j|.
$$

Assume that $|\Delta_i| \leq M$ for some $M \in (0,\infty)$ that does not depend on $N$. We note for later reference that if $|\Delta_i| > 0$ for at most $O(N^{1/2})$ values of $i$, the expression above is $O(N)$. The same condition ensures that the left-hand side of condition (3) is $O(NT^2)$. Consequently, we can choose $Q_1(N,T) = O(N)$ and $Q_2(N,T) = Q_3(N,T) = O(NT^2)$ if at most $O(N^{1/2})$ series undergo a break.

Finally, rather than expanding the list of moment conditions in Assumption 4, we simply impose independence between the idiosyncratic errors and the other variables. It is possible to relax this assumption at the cost of added complexity.[3]

---

[3]Bai and Ng (2006) impose independence of $\{e_t\}$ and $\{F_t\}$ when providing inferential theory for regressions involving estimated factors.

**Assumption 5** (Independence)**.** *For all $(i, j, s, t)$, $e_{it}$ is independent of $(F_s, \xi_{js})$.*

# 3 Asymptotic theory

## 3.1 Consistent estimation of factors

Our main result provides the mean square convergence rate of the usual principal components estimator under Assumptions 1–5. After stating the general theorem, we give sufficient conditions that ensure the same convergence rate that Bai and Ng (2002) obtained in a setting with constant factor loadings.

**Theorem 1.** *Let Assumptions 1–5 hold. For any fixed $k$,*

$$T^{-1} \sum_{t=1}^{T} \|\hat{F}_t^k - H^{k'} F_t\|^2 = O_p(R_{NT})$$

*as $N, T \to \infty$, where*

$$R_{NT} = \max \left\{ \frac{1}{C_{NT}^2}, \frac{h_{NT}^2}{N^2} Q_1(N, T), \frac{h_{NT}^2}{N^2 T^2} Q_2(N, T), \frac{h_{NT}^4}{N^2 T^2} Q_3(N, T) \right\},$$

*and the $r \times k$ matrix $H^k$ is given by*

$$H^k = (\Lambda_0' \Lambda_0 / N)(F' \tilde{F}^k / T).$$

See the appendix for the proof. If $R_{NT} \to 0$ as $N, T \to \infty$, the theorem implies that the $r$-dimensional space spanned by the true factors is estimated consistently in mean square (averaging over time) as $N, T \to \infty$. While we do not discuss it here, a similar statement concerning pointwise consistency of the factors (Bai and Ng, 2002, p. 198) may be achieved by slightly modifying Assumptions 3–4.

We now give sufficient conditions on the envelope functions in Assumption 4 such that the principal components estimator achieves the same convergence rate as in Theorem 1 of Bai and Ng (2002). This rate, $C_{NT}^2$, turns out to be central for other results in the literature on DFMs (Bai and Ng, 2002, 2006). The following corollary is a straight-forward consequence of Theorem 1.

**Corollary 1.** *Under the assumptions of Theorem 1, and if additionally*

- $h_{NT}^2 Q_1(N,T) = O(N)$,

- $h_{NT}^2 Q_2(N,T) = O(NT^2)$,

- $h_{NT}^4 C_{NT}^2 Q_3(N,T) = O(N^2 T^2)$,

*it follows that, as $N, T \to \infty$,*

$$C_{NT}^2 \left( T^{-1} \sum_{t=1}^{T} \| \hat{F}_t^k - H^{k'} F_t \|^2 \right) = O_p(1).$$

**Examples (continued).** In Section 2.4 we computed the envelope functions $Q_1(N,T)$, $Q_2(N,T)$ and $Q_3(N,T)$ for our three examples. From these calculations we get that if $h_{NT} = 1$, the model in Example 1 (white noise) satisfies the conditions of Corollary 1. Hence, uncorrelated order-$O_p(1)$ white noise disturbances in the factor loadings do not affect the asymptotic performance of the principal components estimator.

Likewise, it follows from our calculations that the structural break process in Example 2 (random walk) satisfies the conditions of Corollary 1 if $h_{NT} = O(1/\min\{N^{1/4} T^{1/2}, T^{3/4}\})$. Moreover, a rate of $h_{NT} = o(T^{-1/2})$ is sufficient to achieve $R_{NT} = o(1)$ in Theorem 1, i.e., that the factor space is estimated consistently. This is a weaker rate requirement than the $O(T^{-1})$ scale factor imposed by Stock and Watson (2002).

For Example 3 (single deterministic break), Corollary 1 and our calculations in Section 2.4 yield that if we set $h_{NT} = 1$, the principal components estimator achieves the Bai and Ng (2002) convergence rate, provided at most $O(N^{1/2})$ series undergo a break. A fraction $O(N^{-1/2})$ of the series may therefore experience an order-$O(1)$, perfectly correlated shift in their factor loadings without affecting the asymptotic performance of the estimator.

## 3.2 Estimating the number of factors

Bai and Ng (2002) introduced a class of information criteria that consistently estimate the true number $r$ of factors when the factor loadings are constant through time. Lemma 2 of Amengual and Watson (2007) establishes that, under essentially the same assumptions as used by Bai and Ng, these information criteria remain consistent for $r$ when the data $X$ are measured with an additive error, i.e., if the researcher instead observes $\tilde{X} = X + b$ for an $T \times N$ error matrix $b$ that satisfies $(NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} b_{it}^2 = O_p(C_{NT}^{-2})$. By our decomposition (1) of $X_t$, time variation in the factor loadings may be

seen as contributing an extra error term $w_t$ to the usual terms $\Lambda_0 F_t + e_t$. It therefore follows from the Amengual and Watson result that for an appropriate choice of the scale factor $h_{NT}$, the information criteria will remain consistent under time variation. For example, it may be verified that in Example 2 (random walk), a scale factor $h_{NT} = O(T^{-1})$ will ensure that the $w_t$ term satisfies Amengual and Watson's condition. It is a topic for future research to determine whether this rate can be improved upon.

## 3.3 Diffusion index forecasting

As an application of Corollary 1, consider the diffusion index model of Stock and Watson (1998$a$, 2002) and Bai and Ng (2006). For ease of exposition we assume that the factors are the only explanatory variables, so the model is

$$y_{t+h} = \alpha' F_t + \varepsilon_{t+h}.$$

Here $y_{t+h}$ is the scalar random variable that we seek to forecast, while $\varepsilon_{t+h}$ is an idiosyncratic forecast error term that is independent of all other variables. We shall assume that the true number of factors $r$ is known. Because the true factors $F_t$ are not observable, one must forecast $y_{t+h}$ using the estimated factors $\hat{F}_t$. Does the sampling variability in $\hat{F}$ influence the precision and asymptotic normality of the feasible estimates of $\alpha$?

Let $\hat{F}$ be the principal components estimator with $k = r$ factors estimated and denote the $r \times r$ matrix $H^r$ from Theorem 1 by $H$. Define $\delta = H^{-1}\alpha$ (note that due to the factors being unobservable, $\alpha$ is only identified up to multiplication by a nonsingular matrix) and let $\hat{\delta}$ be the least squares estimator in the feasible diffusion index regression of $y_{t+h}$ on $\hat{F}_t$. Bai and Ng (2006) show that

$$\sqrt{T}(\hat{\delta} - \delta) = (T^{-1}\hat{F}'\hat{F})^{-1}T^{-1/2}\hat{F}'\varepsilon \\ - (T^{-1}\hat{F}'\hat{F})^{-1}[T^{-1/2}\hat{F}'(\hat{F} - FH)]H^{-1}\alpha, \tag{4}$$

where $\varepsilon = (\varepsilon_{1+h}, \ldots, \varepsilon_{T+h})'$. Under the assumptions of Corollary 1, the Cauchy-Schwarz inequality yields

$$\|T^{-1/2}\hat{F}'(\hat{F} - FH)\|^2 \leq T \left( T^{-1}\sum_{t=1}^{T}\|\hat{F}_t\|^2 \right) \left( T^{-1}\sum_{t=1}^{T}\|\hat{F}_t - H'F_t\|^2 \right) \\ = TO_p(1)O_p(C_{NT}^{-2}) \\ = O_p(1).$$

12

Similarly,

$$T^{-1/2}\hat{F}'\varepsilon = T^{-1/2}H'F'\varepsilon + T^{-1/2}(\hat{F} - FH)'\varepsilon = T^{-1/2}H'F'\varepsilon + O_p(1).$$

Suppose $T^{-1/2}F'\varepsilon = O_p(1)$, as implied by Assumption E in Bai and Ng (2006). It is easy to show that $H = O_p(1)$. The preceding calculations then suggest that $\hat{\delta} - \delta = O_p(T^{-1/2})$, i.e., the feasible diffusion regression estimator is consistent at the usual rate.

Going further, if $\alpha = 0$, which is often an interesting null hypothesis in applied work, the second term on the right-hand side of the decomposition (4) vanishes. Assume that $\{\varepsilon_{t+h}\}$ is independent of all other variables. Then, conditional on $\hat{F}$, the first term on the right-hand side of (4) will (under weak conditions) obey a central limit theorem, and so it seems reasonable to expect that $\hat{\delta}$ should be unconditionally asymptotically normally distributed under the null $H_0$: $\alpha = 0$. Bai and Ng (2006) prove that if the factor loadings are not subject to time variation, $\hat{\delta}$ will indeed be asymptotically normal, regardless of the true value of $\alpha$, as long as $\sqrt{T}/N \to 0$. We expect that a similar result can be proved formally in our framework but leave this for future research.

## 4    Simulations

### 4.1    Design

To illustrate our results and assess their finite sample validity we conduct a Monte Carlo simulation study. Stock and Watson (2002) assessed the performance of the principal components estimator when the factor loadings evolve as a random walk. We provide additional evidence on the necessary scale factor $h_{NT}$ for the random walk case (our Example 2). Second, we consider a different design in which a subset of the series undergo one large break in their factor loadings (an analog of Example 3).

The design follows Stock and Watson (2002) where possible:

$$X_{it} = \lambda'_{it}F_t + e_{it},$$

$$F_{tp} = \rho F_{t-1,p} + u_{tp},$$

$$(1 - aL)e_{it} = v_{it},$$

$$y_{t+1} = \sum_{q=1}^{r} F_{tq} + \varepsilon_{t+1},$$

where $i = 1, \ldots, N$, $t = 1, \ldots, T$, $p = 1, \ldots, r$, and the variables $\{u_{tp}\}$, $\{v_{it}\}$ and $\{\varepsilon_{t+1}\}$ are mutually independent i.i.d. standard normally distributed. The scalar $\rho$ is the common AR(1) coefficient for the $r$ factors, while $a$ is the AR(1) coefficient for the idiosyncratic errors.[4]

The initial values $F_0$ and $e_0$ for the factors and idiosyncratic errors are drawn from their respective stationary distributions. The initial factor loading matrix $\Lambda_0$ was chosen based on the population $R^2$ for the regression of $X_{i0} = \lambda'_{i0} F_0 + e_{i0}$ on $F_0$. Specifically, for each $i$ we draw a value $R_i^2$ uniformly at random from the interval $[0.1, 0.8]$. We then set $\lambda_{i0p} = \lambda^*(R_i^2) \bar{\lambda}_{i0j}$, where $\bar{\lambda}_{i0j}$ is i.i.d. standard normal and independent of all other disturbances. The scalar $\lambda^*(R_i^2)$ is given by the value for which $E[(\lambda'_{0i} F_0)^2]/E[X_{i0}^2] = R_i^2$, given the draw of $R_i^2$.[5]

We consider two different specifications for the evolution of factor weights over time. In the *random walk model* we set

$$\lambda_{itp} = \lambda_{i,t-1,p} + cT^{-3/4} \zeta_{itp},$$

$i = 1, \ldots, N$, $t = 1, \ldots, T$, $p = 1, \ldots, r$, where $c$ is a constant and the innovations $\zeta_{itp}$ are i.i.d. standard normal and independent of all other disturbances. Note that the $T^{3/4}$ rate is different from the rate of $T$ used by Stock and Watson (2002). We consider the choices $c = 0, 2, 5$ for our benchmark simulation. Observe that the standard deviation of $\lambda_{iTp} - \lambda_{i0p}$ is $cT^{-1/4}$, which for $T = 100$ and $c = 2$ equals 0.63, of the same magnitude as the standard deviation of $\lambda_{i0p}$ (e.g., for $a = \rho = 0$ and $r = 5$, we have $\lambda^*(0.45) = 0.40$).

In the *large break model*, we select a subset $J$ of size $[bN^{1/2}]$ from the integers $\{1, \ldots, N\}$, where $b$ is a constant. For $i \notin J$, we simply let $\lambda_{itp} = \lambda_{i0p}$ for all $t$. For $i \in J$, we set

$$\lambda_{itp} = \begin{cases} \lambda_{i0p} & \text{for } t \leq [0.5T] \\ \lambda_{i0p} + \Delta_p & \text{for } t > [0.5T] \end{cases}.$$

The shift $\Delta_p$ (which is the same for all $i \in J$) is distributed $\mathcal{N}(0, [\lambda^*(0.45)]^2)$, i.i.d. across $p = 1, \ldots, r$, so that the shift is of the same magnitude as the initial loading $\lambda_{i0p}$.[6] We set $b = 0, 4, 8$ in the benchmark simulations.

---

[4] We do not consider cross-sectional correlation of the idiosyncratic errors here.

[5] In this paper we have assumed that $\Lambda_0$ is fixed for simplicity. It is not difficult to verify that $\Lambda_0$ could instead be random, provided that it is independent of all other random variables, $N^{-1}\Lambda'_0\Lambda_0 \xrightarrow{p} D$ for an $r \times r$ non-singular matrix $D$, and $E\|\lambda_i\|^4 < M$, as in Bai and Ng (2006).

[6] This shift process satisfies Assumption 4 by essentially the same argument as was used for the deterministic break in Example 3.

| $T$ | 50 | 100 | 150 | 200 | 250 | 300 | 350 | 400 |
|---|---|---|---|---|---|---|---|---|
| $N$ | 125 | 250 | 375 | 500 | 625 | 750 | 875 | 1000 |
| $N^{-1/2}$ | 0.089 | 0.063 | 0.052 | 0.045 | 0.040 | 0.037 | 0.034 | 0.032 |

Table 1: Combinations of $T$ and $N$ used in the simulations.

The free parameters are $T$, $N$, $r$, $\rho$, $a$, $b$, $c$. We let $r = 5$ throughout and focus on sample sizes $T = 50, 100, 150, \ldots, 400$. We set $N = [dT]$, with $d = 2.5$ as in the baseline design of Stock and Watson (2002). Table 1 shows the $(T, N)$ combinations. Because the fraction of series that undergo a shift in the large break model is $[bN^{1/2}]/N \approx bN^{-1/2}$, the table also lists values of $N^{-1/2}$. For example, for $T = 100$, $N = 250$ and $b = 4$, about 25% of the series undergo a large structural break in their factor loadings.

We use the principal components estimator $\hat{F}$ described earlier to estimate the factors, assuming that the true number of factors $r$ is known. To evaluate the estimator's performance, we compute a trace $R^2$ statistic for the multivariate regression of $\hat{F}$ onto $F$,

$$
R^2_{\hat{F}, F} = \frac{\hat{E}\|P_F \hat{F}\|^2}{\hat{E}\|\hat{F}\|^2},
$$

where $\hat{E}$ denotes averaging over Monte Carlo repetitions and the projection matrix $P_F = F(F'F)^{-1}F'$. Corollary 1 states that this measure tends to 1 as $T \to \infty$. In each repetition we compute the feasible out-of-sample forecast $\hat{y}_{T+1|T} = \hat{\delta}'\hat{F}_T$, where $\hat{\delta}$ are the OLS coefficients in the regression of $y_{t+1}$ onto $\hat{F}_t$ for $t \leq T - 1$, as well as the infeasible forecast $\tilde{y}_{T+1|T} = \tilde{\delta}'F_T$, where $\tilde{\delta}$ is obtained by regressing $y_{t+1}$ onto the true factors $F_t$, $t \leq T - 1$. The closeness of the feasible and infeasible forecasts is measured by the statistic

$$
S^2_{\hat{y}, \tilde{y}} = 1 - \frac{\hat{E}(\hat{y}_{T+1|T} - \tilde{y}_{T+1|T})^2}{\hat{E}(\hat{y}^2_{T+1|T})}.
$$

The measures $R^2_{\hat{F}, F}$ and $S^2_{\hat{y}, \tilde{y}}$ were also used by Stock and Watson (2002).

## 4.2 Benchmark results

Our benchmark simulation sets $a = \rho = 0$ (no serial dependence for the factors or idiosyncratic errors). We perform 10,000 Monte Carlo repetitions. The results are summarized in Figure 1 (the random walk model) and Figure 2 (the large break model). Each figure has two panels. The top panel shows

the $R^2_{\hat{F},F}$ statistic as a function of the sample size $T$, for the three different choices of $c$ or $b$. Similarly, the bottom panel shows the $S^2_{\hat{y},\tilde{y}}$ statistic.

Figure 1 confirms that substantial random walk variation in the factor loadings, vanishing at rate $T^{3/4}$, does impact the precision of the principal components estimator, but the performance improves as $T$ increases, as suggested by Corollary 1. Both in terms of closeness of factor spaces and out-of-sample forecasts, setting $c = 2$ impacts the performance to about the same extent as introducing moderate serial correlation in the idiosyncratic errors (Stock and Watson, 2002, cf.). The $c = 5$ simulations give much worse results; however, $c = 5$ is extreme in sense that the time variation in the factor loadings is of larger magnitude than the initial loadings for the choices of $T$ that we consider here.

In Figure 2 we observe that the principal components estimator is also fairly robust to one-time, correlated large breaks. In fact, the feasible out-of-sample forecast is very close to the infeasible one, even when the sample size is small and about 50% of the series undergo a large shift in their factor loadings.

The results for $a = \rho = 0.5$ are very similar to the ones depicted in Figures 1–2. Because there does not appear to be any interesting interactions between serial dependence and time variation, we do not report the alternative simulations.

## 4.3 Rate of convergence

We now turn to the more detailed asymptotic rates stated in Theorem 1. Our method of proof suggests that it may not in general be possible to improve upon the $R_{NT}$ rate. To investigate this claim, we carry out two exercises. First, we execute a separate set of simulations in which $\lambda_{itp} - \lambda_{i,t-1,p} = cT^{-1/2}\zeta_{itp}$ for the random walk model, and the number of shifting series in the large break model is set to $[bN]$. These two rates both (just) violate the conditions for mean square consistency in Theorem 1. To make the results comparable to Figures 1–2, we scale down our choices of $b$ and $c$ so that the standard deviation of the random walk innovations and the number of shifting series approximately match at $T = 100$ under the two $(T, N)$ sequences. All other parameter choices are unchanged. See Figures 3–4 for the results. As hypothesized, for both models the trace $R^2$ statistic does not seem to improve systematically with the sample size.

Second, we construct a "rate frontier" that corresponds to the predictions of Theorem 1 for the special case of the random walk model. To this end, suppose we set $N = [T^\mu]$ and $h_{NT} = cT^{-\gamma}$. Using the formula for $R_{NT}$ and

the random walk calculations in Section 2.4, we obtain

$$R_{NT} = O(\max\{T^{-1}, T^{-\mu}, T^{1-2\gamma-\mu}, T^{2-4\gamma}\}) = O(T^{m(\mu,\gamma)}),$$

where

$$m(\mu,\gamma) = \max\{-1, -\mu, 1 - 2\gamma - \mu, 2 - 4\gamma\}.$$

This convergence rate exponent reflects the influence of the magnitude of the random walk deviations, as measured by $\gamma$, and the relative sizes of the cross-sectional and time dimensions, as measured by $\mu$. Evidently, increasing the number of available series relative to the sample size improves the convergence rate, but only up to a point. In the following we set $\mu = 1$ as in the previous simulations. Denote

$$m(\gamma) = m(1,\gamma) = \max\{-1, -2\gamma, 2 - 4\gamma\} = \max\{-1, 2 - 4\gamma\}.$$

We see that the dependence of the convergence rate on $\gamma$ is monotonic, as expected, but nonlinear. The flat profile of the trace $R^2$ statistic in Figure 3 is fully consistent with $m(1/2) = 0$.

These calculations have been carried out on the worst-case rate stated in Theorem 1. We have not been able to prove that the convergence rate $R_{NT}$ is sharp, in the sense that there exists a DFM satisfying Assumptions 1–5 that achieves the $R_{NT}$ rate. Instead, we provide simulation evidence suggesting that the independent random walk model (Example 2) achieves the stated bound. We maintain the simulation design described in Section 4.1 with $a = \rho = 0$ and $N = [2.5T]$, except that we set $h_{NT} = 5T^{-\gamma}$ and vary $\gamma$ over the range $0.25, 0.30, 0.35, \ldots, 1.50$. For each value of $\gamma$ and each sample size $T = 100, 200, 300, \ldots, 700$ we compute the statistic

$$\widehat{MSE}(\gamma, T) = T^{-1}(\hat{E}\|\hat{F}\|^2 - \hat{E}\|P_F\hat{F}\|^2),$$

where $\hat{E}$ denotes the average over 500 Monte Carlo repetitions. This statistic is a close analog of the mean square error that is the object of study in Theorem 1. Our theoretical results suggest that $\widehat{MSE}(\gamma, T)$ should grow or decay at rate $T^{m(\gamma)}$. We verify this by regressing, for each $\gamma$,

$$\log \widehat{MSE}(\gamma, T) = \text{constant}_\gamma + m_\gamma \log T,$$

using our seven observations $T = 100, 200, \ldots, 700$. Figure 5 plots the estimates $\hat{m}_\gamma$ against $\gamma$ along with the theoretical values $m(\gamma)$. The estimated rate frontier is strikingly close to the theoretical one, although some finite-sample issues remain for intermediate values of $\gamma$. This corroborates our conjecture that Theorem 1 provides sharp rates for the independent random walk case.

17

# 5 Discussion and conclusions

The theoretical results of Section 3 and the simulation study of Section 4 point towards a considerable amount of robustness of the principal components estimator of the factors when the factor loading matrix varies over time. Although we have not proved that the consistency rate function presented in Section 4 is tight, inspection of our proof does not suggest any room for improvement and moreover the Monte Carlo estimate of the rate function accords broadly with the theoretical rate function. This leads us to suspect that the rate function is tight, and in this sense represents an upper bound on the parameter instability that can be tolerated by the principal components estimator. The amount of such instability is quite large when calibrated for values of $N$ and $T$ typically used in applied work.

Our evidence concerning the robustness of the principal components estimator raise a tension with the results in Breitung and Eickmeier (2011), who in contrast suggest that undetected breaks in the factor loadings have the effect of increasing the dimension of the true factors. Indeed, we conjecture (but do not prove) that the principal components estimator will be consistent even under sequences of breaks for which the Breitung and Eickmeier (2011) test rejects. Pursuing this calculation would be of independent interest and would also return the large-dimensional discussion here to the bias-variance tradeoffs associated with ignoring breaks tackled in a low-dimensional setting by Pesaran and Timmermann (2005, 2007).

# A   Proof

To lighten the notation, we denote $\sum_i = \sum_{i=1}^N$ (the same for $j$) and $\sum_s = \sum_{s=1}^T$ (the same for $t$). A double sum $\sum_{i=1}^N \sum_{j=1}^N$ is denoted $\sum_{i,j}$.

**Proof of Theorem 1.**   We extend the proof of Theorem 1 in Bai and Ng (2002). By the definition of the estimator $\hat{F}^k$, we have $\hat{F}^k = (NT)^{-1} XX' \tilde{F}^k$, where $\tilde{F}^{k'} \tilde{F}^k / T = I_k$ (Bai and Ng, 2008$b$). Define $e = (e_1, \ldots, e_T)'$ and $w = (w_1, \ldots, w_T)'$. Since

$$XX' = F\Lambda_0' \Lambda_0 F' + F\Lambda_0'(e+w)' + (e+w)\Lambda_0 F' + (e+w)(e+w)',$$

we can write

$$\hat{F}_t^k - H^{k'} F_t = (NT)^{-1} \Big\{ \tilde{F}^{k'} F \Lambda_0' e_t + \tilde{F}^{k'} e\Lambda_0 F_t + \tilde{F}^{k'} ee_t + \tilde{F}^{k'} F\Lambda_0' w_t$$
$$+ \tilde{F}^{k'} w\Lambda_0 F_t + \tilde{F}^{k'} ww_t + \tilde{F}^{k'} ew_t + \tilde{F}^{k'} we_t \Big\}.$$

Label the eight terms on the right-hand side $A_{1t}, \ldots, A_{8t}$, respectively. By Loève's inequality,

$$T^{-1} \sum_t \|\hat{F}_t^k - H^{k'} F_t\|^2 \le 8 \sum_{n=1}^8 \left( T^{-1} \sum_t \|A_{nt}\|^2 \right). \tag{5}$$

Bai and Ng (2002) have shown that the terms corresponding to $n = 1, 2, 3$ are $O_p(C_{NT}^{-2})$ under Assumptions 1–3. We proceed to bound the remaining terms in probability.

We have

$$\|A_{4t}\|^2 \le \left( T^{-1} \sum_s \|\tilde{F}_s^k\|^2 \right) \left( T^{-1} \sum_s \|F_s\|^2 \right) \left\| N^{-1} \Lambda_0' w_t \right\|^2.$$

The first factor equals $\operatorname{tr}(\tilde{F}^{k'} \tilde{F}^k / T) = \operatorname{tr}(I_k) = k$. The second factor is $O_p(1)$ by Assumption 1. Also,

$$E \left\| \frac{\Lambda_0' w_t}{N} \right\|^2 \le N^{-2} \sum_{i,j} |E(w_{it} w_{jt}) \lambda_{i0}' \lambda_{j0}|$$
$$\le \bar{\lambda}^2 h_{NT}^2 N^{-2} \sum_{i,j} |E(\xi_{it} F_t \xi_{it} F_t)|$$
$$\le r^2 \bar{\lambda}^2 \sup_{p,q} h_{NT}^2 N^{-2} \sum_{i,j} |E(\xi_{itp} F_{tp} \xi_{itq} F_{tq})|$$
$$= O(h_{NT}^2 N^{-2} Q_1(N,T)),$$

uniformly in $t$, by Assumption 4.1. Hence,

$$T^{-1} \sum_t \|A_{4t}\|^2 = O_p(h_{NT}^2 N^{-2} Q_1(N, T)).$$

Similarly,

$$\|A_{5t}\|^2 \leq \left( T^{-1} \sum_s \|\tilde{F}_s^k\|^2 \right) \left( (N^2 T)^{-1} \sum_s (w_s' \Lambda_0 F_t)^2 \right),$$

where the first term is $O(1)$ and

$$(N^2 T)^{-1} E \sum_s (w_s' \Lambda_0 F_t)^2 \leq (N^2 T)^{-1} \sum_s \sum_{i,j} |E(w_{is} w_{js} \lambda_{i0}' F_t \lambda_{j0}' F_t)|$$

$$\leq r^4 \bar{\lambda}^2 \sup_{p_1, q_1, p_2, q_2} h_{NT}^2 (N^2 T)^{-1} \sum_s \sum_{i,j} |E(\xi_{isp_1} \xi_{jsq_1} F_{sp_1} F_{sq_1} F_{tp_2} F_{tq_2})|.$$

By summing over $t$, dividing by $T$ and using Assumption 4.2 we obtain

$$T^{-1} \sum_t \|A_{5t}\|^2 = O_p(h_{NT}^2 N^{-2} T^{-2} Q_2(N, T)).$$

For the sixth term,

$$E\|A_{6t}\|^2 \leq E\left\{ \left( T^{-1} \sum_s \|\tilde{F}_s^k\|^2 \right) \left( (N^2 T)^{-1} \sum_s (w_s' w_t)^2 \right) \right\}$$

$$= k(N^2 T)^{-1} \sum_s \sum_{i,j} E(w_{is} w_{it} w_{js} w_{jt})$$

$$\leq k r^4 \sup_{p_1, q_1, p_2, q_2} \frac{h_{NT}^4}{N^2 T} \sum_s \sum_{i,j} |E(\xi_{isp_1} \cdots \xi_{jtq_2} F_{sp_1} \cdots F_{tq_2})|,$$

where the last expectation contains the eight factors listed in Assumption 4.3. By Assumption 4.3, it follows that

$$T^{-1} \sum_t \|A_{6t}\|^2 = O_p(h_{NT}^4 N^{-2} T^{-2} Q_3(N, T)).$$

Regarding the seventh term, using Assumption 5,

$$E\|A_{7t}\|^2 \leq E\left\{\left(T^{-1}\sum_s \|\tilde{F}_s^k\|^2\right)\left((N^2T)^{-1}\sum_s (e_s'w_t)^2\right)\right\}$$

$$= k(N^2T)^{-1}\sum_s \sum_{i,j} E(e_{is}e_{js})E(w_{it}w_{jt})$$

$$\leq k(N^2T)^{-1}\sum_s \sum_{i,j}(E(e_{is}^2)E(e_{js}^2))^{1/2}|E(w_{it}w_{jt})|$$

$$\leq kr^2 M \sup_{p,q} h_{NT}^2(N^2T)^{-1}\sum_s \sum_{i,j}|E(\xi_{itp}\xi_{jtq}F_{tp}F_{tq})|$$

$$= O(h_{NT}^2 N^{-2}Q_1(N,T)),$$

uniformly in $t$. The second-to-last line uses $E(e_{it}^2) \leq M$, whereas the last follows from Assumption 4.1. We conclude that

$$T^{-1}\sum_t \|A_{7t}\|^2 = O_p(h_{NT}^2 N^{-2}Q_1(N,T)).$$

A similar argument gives

$$T^{-1}\sum_t \|A_{8t}\|^2 = O_p(h_{NT}^2 N^{-2}Q_1(N,T)).$$

We conclude that the right-hand side of inequality (5) is the sum of variables of four different stochastic orders: $O_p(C_{NT}^{-2})$, $O_p(h_{NT}^2 N^{-2}Q_1(N,T))$, $O_p(h_{NT}^2 N^{-2}T^{-2}Q_2(N,T))$ and $O_p(h_{NT}^4 N^{-2}T^{-2}Q_3(N,T))$. The statement of the theorem follows. □

# References

Amengual, D. and Watson, M. W. (2007), 'Consistent Estimation of the Number of Dynamic Factors in a Large N and T Panel', *Journal of Business and Economic Statistics* **25**(1), 91–96.

Bai, J. and Ng, S. (2002), 'Determining the Number of Factors in Approximate Factor Models', *Econometrica* **70**(1), 191–221.

Bai, J. and Ng, S. (2006), 'Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions', *Econometrica* **74**(4), 1133–1150.

Bai, J. and Ng, S. (2008*a*), 'Large Dimensional Factor Analysis', *Foundations and Trends in Econometrics* **3**(2), 89–163.

Bai, J. and Ng, S. (2008*b*), 'Recent Developments in Large Dimensional Factor Analysis'. Working paper.

Breitung, J. and Eickmeier, S. (2011), 'Testing for structural breaks in dynamic factor models', *Journal of Econometrics* **163**(1), 71–84.

Chudik, A. and Pesaran, M. H. (2011), 'Infinite-dimensional VARs and factor models', *Journal of Econometrics* **163**(1), 4–22.

Clements, M. and Hendry, D. (1998), *Forecasting Economic Time Series*, Cambridge University Press.

Eickmeier, S. and Ziegler, C. (2008), 'How Successful are Dynamic Factor Models at Forecasting Output and Inflation? A Meta-Analytic Approach', *Journal of Forecasting* **27**(3), 237–265.

Engle, R. and Watson, M. W. (1981), 'A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates', *Journal of the American Statistical Association* **76**(376), 774–781.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000), 'The Generalized Dynamic-Factor Model: Identification and Estimation', *Review of Economics and Statistics* **82**(4), 540–554.

Geweke, J. (1977), The Dynamic Factor Analysis of Economic Time Series, *in* D. J. Aigner and A. S. Goldberger, eds, 'Latent Variables in Socio-Economic Models', North-Holland.

Hamilton, J. (1989), 'A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle', *Econometrica* **57**(2), 357–384.

Pesaran, H. M., Pettenuzzo, D. and Timmermann, A. (2006), 'Forecasting Time Series Subject to Multiple Structural Breaks', *Review of Economic Studies* **73**(4), 1057–1084.

Pesaran, M. H. and Timmermann, A. (2005), 'Small sample properties of forecasts from autoregressive models under structural breaks', *Journal of Econometrics* **129**(1-2), 183–217.

Pesaran, M. H. and Timmermann, A. (2007), 'Selection of estimation window in the presence of breaks', *Journal of Econometrics* **137**(1), 134–161.

Sargent, T. (1989), 'Two Models of Measurements and the Investment Accelerator', *The Journal of Political Economy* **97**(2), 251–287.

Sargent, T. and Sims, C. A. (1977), Business Cycle Modeling Without Pretending to Have Too Much A Priori Economic Theory, *in* C. A. Sims, ed., 'New Methods in Business Research', Federal Reserve Bank of Minneapolis.

Stock, J. H. and Watson, M. W. (1989), New Indexes of Coincident and Leading Economic Indicators, *in* O. J. Blanchard and S. Fischer, eds, 'NBER Macroeconomics Annual 1989', Vol. 4, MIT Press, pp. 351–409.

Stock, J. H. and Watson, M. W. (1998*a*), 'Diffusion Indexes'. NBER Working Paper 6702.

Stock, J. H. and Watson, M. W. (1998*b*), 'Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model', *Journal of the American Statistical Association* **93**(441), 349–358.

Stock, J. H. and Watson, M. W. (2002), 'Forecasting Using Principal Components From a Large Number of Predictors', *Journal of the American Statistical Association* **97**(460), 1167–1179.

Stock, J. H. and Watson, M. W. (2009), Forecasting in Dynamic Factor Models Subject to Structural Instability, *in* D. F. Hendry, J. Castle and N. Shephard, eds, 'The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry', Oxford University Press, pp. 173–205.

Stock, J. H. and Watson, M. W. (2011), Dynamic Factor Models, *in* M. P. Clement and D. F. Hendry, eds, 'The Oxford Handbook of Economic Forecasting', Oxford University Press, pp. 35–59.

Watson, M. W. and Engle, R. F. (1983), 'Alternative Algorithms for the Estimation of Dynamic Factor, MIMIC and Varying Coefficient Regression Models', *Journal of Econometrics* **23**(3), 385–400.
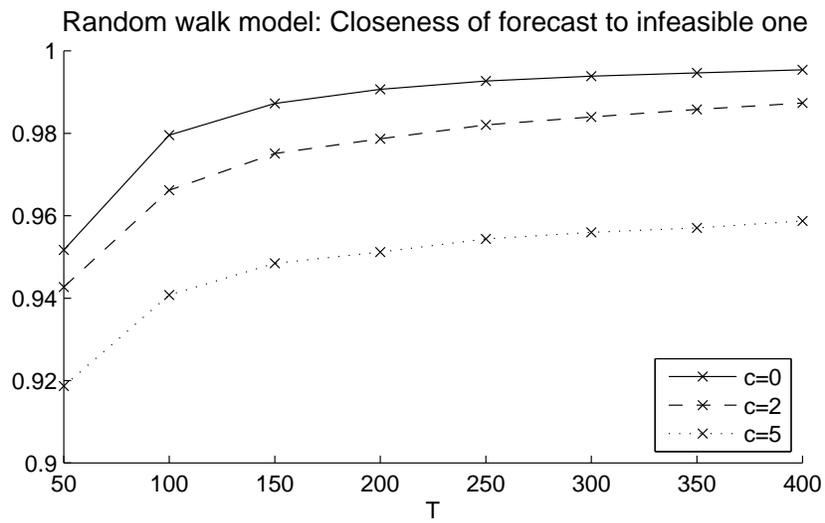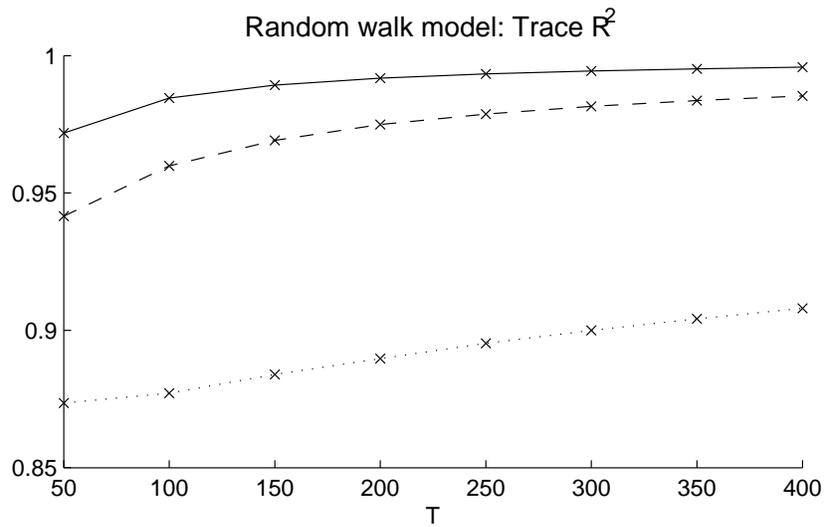
Figure 1: Simulation results for the random walk model, benchmark parameter and rate choices. Actual observations are marked with "x." The lines are piecewise linear interpolations.
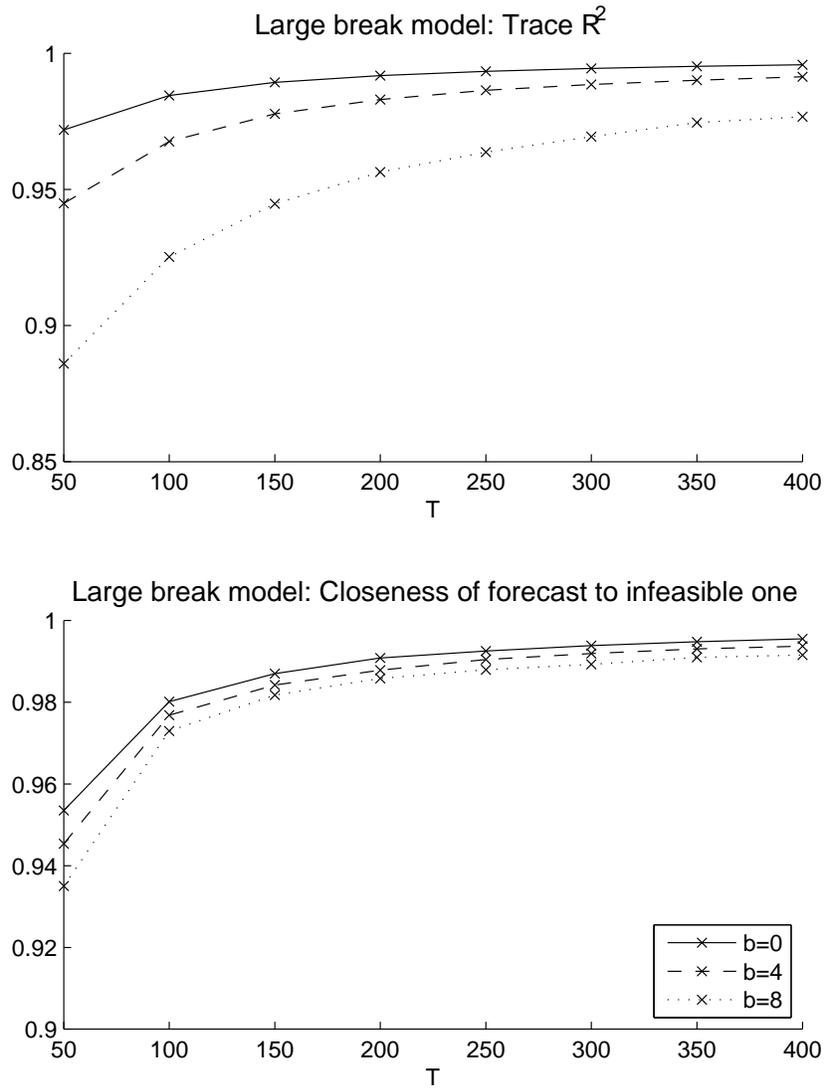
Figure 2: Simulation results for the large break model, benchmark parameter and rate choices.
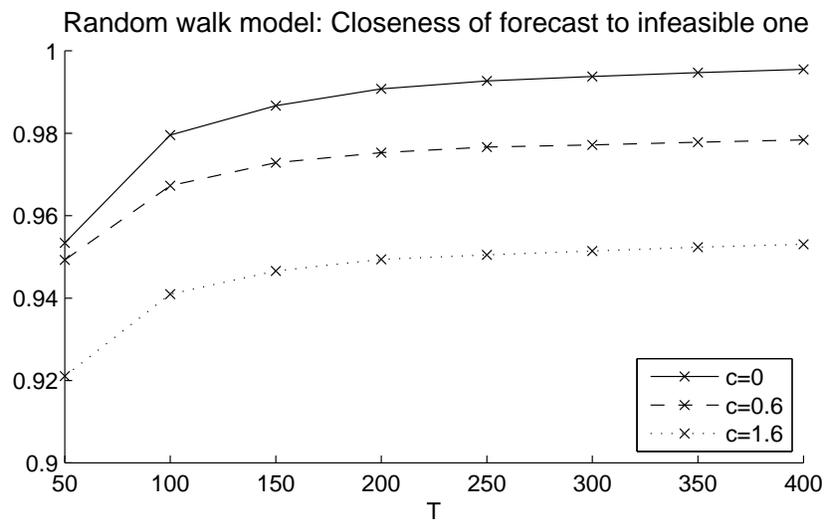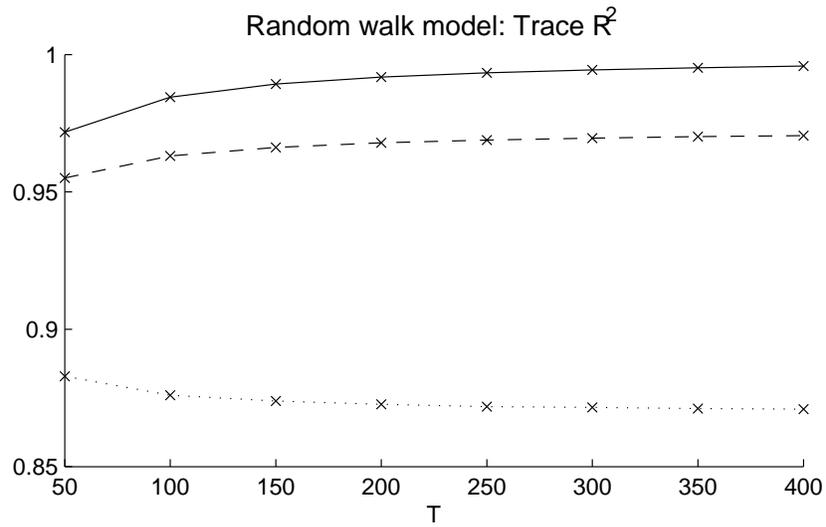
Figure 3: Simulation results for the random walk model, alternative rates.
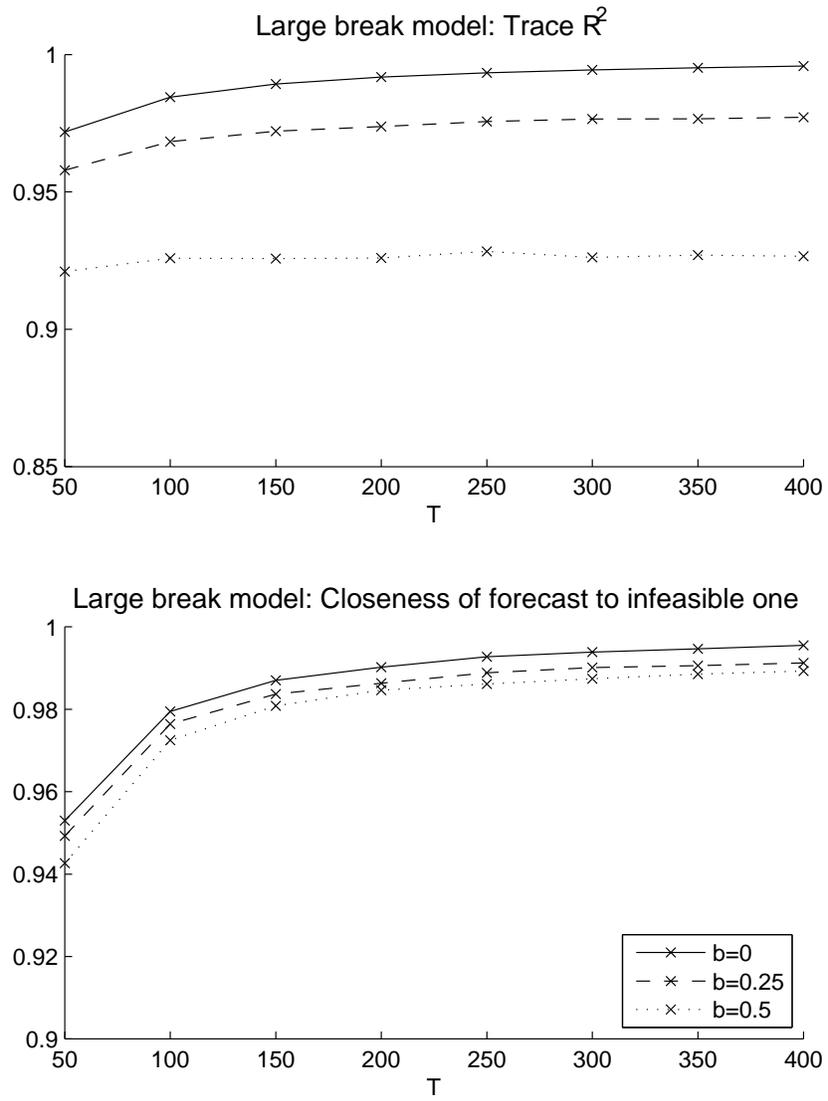
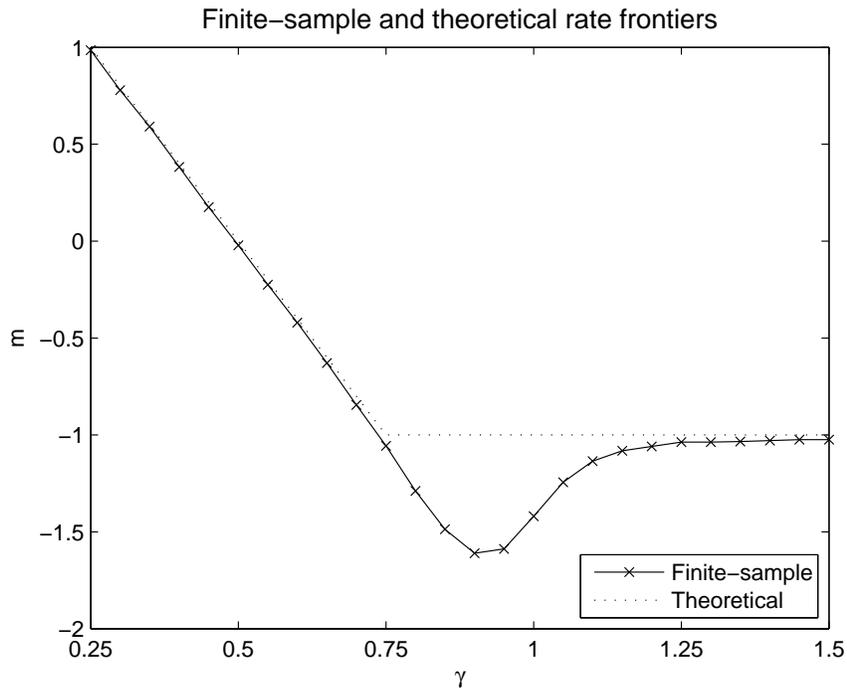Figure 4: Simulation results for the large break model, alternative rates.

Figure 5: Rate frontiers for the random walk model with $N = O(T)$ and $h_{NT} = O(T^{-\gamma})$. The solid line interpolates between the finite-sample rate exponent estimates $\hat{m}_\gamma$ (observations are marked with "x"), while the dotted line represents the theoretical rate exponent $m(\gamma)$.