Consistent Factor Estimation in Dynamic Factor Models with Structural Instability*

Brandon J. Bates
BlackRock, Inc.
brandon.bates@blackrock.com

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

December 28, 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 also discuss implications of our 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. Simulations calibrated to an empirical application indicate that instability in the factor loadings has a limited impact on estimation of the factor space and diffusion index forecasting, whereas estimation of the number of factors is more substantially affected.

1 Introduction

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

*We thank Gary Chamberlain, Herman van Dijk, Anna Mikusheva, Allan Timmermann and two anonymous referees for helpful comments.

[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 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), Eickmeier and Ziegler (2008), Chudik and Pesaran (2011) and Stock and Watson (2011).

broad evidence of time variation in many macroeconomic forecasting relations. Recently, a number of empirical DFM papers have explicitly allowed for structural instability, e.g., Banerjee et al. (2008), Stock and Watson (2009), Eickmeier et al. (2011) and Korobilis (forthcoming). However, theoretical guidance remains scant.

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 supply empirical evidence supporting the use of pre-break data for forecasting. Pesaran and Timmermann (2007) develop 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 an empirical example using U.S. macroeconomic data from 1960–2007 in which full-sample estimates of the factors are preferable to subsample estimates, despite clear evidence of a 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 Λ can vary over time. We write this time variation so that Λ at date t equals its value at date 0, plus a deviation; that is, $\Lambda_t = \Lambda_0 + h_{NT}\xi_t$. The term ξ_t is a possibly random disturbance, and h_{NT} is a deterministic scalar sequence in N and T which governs the scale of the deviation. Using this framework and standard assumptions in the literature (Bai and Ng, 2002, 2006*a*), 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 Λ_t from Λ_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 Λ_t is a vector of independent random walks, Stock and Watson (2002) showed that the factor estimates are consistent if $h_{NT} = O(T^{-1})$. 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 established by Bai and Ng (2002) in the case of no time variation. Because the elements of ξ_t in the random walk case are themselves $O_p(t^{1/2})$, 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, 1998*b*) and would have power of one against parameter deviations of the magnitude tolerated by the principal components estimator. Intuitively, the reason that the principal components 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 N/T as T increases, the amount of cross-sectional information that can be used to "average out" this instability. To elaborate on the practical implications of the theory, we conduct a simulation study calibrated to the Stock and Watson (2009) dataset. The results confirm that the principal components estimator and derived diffusion index forecasts are robust to empirically relevant degrees of temporal instability in the factor loadings, although the precise quantitative conclusions depend on the assumed type of structural instability and the persistence of the factors. Interestingly, the robustness obtains even though the Bai and Ng (2002) information criterion estimator of the rank of the factor space appears to be asymptotically biased for some of our parametrizations.

The rest of the paper proceeds as follows. Section 2 lays out the model, the assumptions, and the three special cases. Our main result on consistency of the principal components estimator is presented in Section 3. Rank selection and diffusion index forecasting are discussed in Section 4. Section 5 provides Monte Carlo results, and Section 6 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, fixed 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 series i at time t, $F_t = (F_{1t}, \ldots, F_{rt})'$, and e_{it} is an idiosyncratic error. Define vectors $X_t = (X_{1t}, \ldots, X_{Nt})'$, $e_t = (e_{1t}, \ldots, e_{Nt})'$, $\Lambda_t = (\lambda_{1t}, \ldots, \lambda_{Nt})'$ and data matrices $X = (X_1, \ldots, X_T)'$, $F = (F_1, \ldots, F_T)'$. The initial factor loadings Λ_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 $N \times r$, $\xi_t = (\xi_{1t}, \ldots, \xi_{Nt})'$ (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.[2]

To establish some intuition for why estimation of the factors is possible despite structural instability, let the number of factors be $r = 1$ and 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 ζ_{it} is i.i.d. across i and t with mean 0 and variance σ_ζ^2, and suppose that Λ_0 is known. In addition, we look ahead to Assumption 2 and assume that $\Lambda_0'\Lambda_0/N \to D > 0$. Because Λ_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 Λ_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 Λ_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 continue to set the number of factors r to 1 for ease of exposition.

[2] As pointed out by our referees, a straight-forward approach would be to treat $e_t^* = e_t + w_t$ as a catch-all error term and provide conditions on h_{NT} and ξ_t such that e_t^* satisfies Assumption C in Bai and Ng (2002). Some of the examples below could be handled this way. However, in the case of random walk factor loadings, applying the Bai and Ng assumption to e_t^* would restrict the temporal dependence of ξ_t more severely than required by our Theorem 1 (cf. Assumption 3.2 below).

Example 1 (white noise). All entries ξ_{it} are i.i.d. across i and t with mean zero and $E(\xi_{it}^4) < \infty$. The factor loadings Λ_t are then equal to the initial loading matrix Λ_0 plus uncorrelated noise.[3]

Example 2 (random walk). Entries ξ_{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.[4] Models of this type are often referred to as time-varying parameter models in the literature. DFMs with time-varying parameters have recently received attention in the empirical macro literature, cf. Eickmeier et al. (2011), Korobilis (forthcoming) and references therein.

Example 3 (single large 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

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

Breitung and Eickmeier (2011) demonstrate that a structurally unstable model of this kind may equivalently be written as a stable DFM with $2r$ dynamic factors. Deterministic parameter shifts have also 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 components estimators of the loadings and factors are obtained by solving the minimization problem

$$V(k) = \min_{\Lambda^k, F^k} (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \lambda_i^{k\prime} F_t^k)^2, \tag{2}$$

where the supercripts on Λ^k and F^k signify that there are k estimated factors. It is necessary to impose a normalization on the estimators to uniquely define the minimizers (see Bai and Ng, 2008, 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 of F is obtained by first concentrating out Λ^k and imposing the normalization $F^{k\prime}F^k/T = I_k$. 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 $\Lambda^{k\prime}\Lambda^k/N = I_k$. 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

$$\hat{F}^k = \bar{F}^k(\bar{F}^{k\prime}\bar{F}^k/T)^{1/2}$$

in the following.

[3] As is clear from the subsequent calculations, our conclusions remain true if the disturbances are weakly dependent in the temporal and cross-sectional dimensions. In the interest of clarity we focus on the i.i.d. case.

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

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 $\|A\| = [\text{tr}(A'A)]^{1/2}$. 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). *$E\|F_t\|^4 \leq M$ and $T^{-1}\sum_{t=1}^{T} F_t F_t' \xrightarrow{p} \Sigma_F$ as $T \to \infty$ for some positive definite matrix Σ_F.*

Assumption 2 (Initial factor loadings). *$\|\lambda_{i0}\| \leq \bar{\lambda} < \infty$, 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. *$E(e_{it}) = 0$, $E|e_{it}|^8 \leq M$.*
2. *$\gamma_N(s,t) = E(e_s' e_t/N)$ exists for all (s,t). $|\gamma_N(s,s)| \leq M$ for all s, and $T^{-1}\sum_{s,t=1}^{T} |\gamma_N(s,t)| \leq M$.*
3. *$\tau_{ij,ts} = E(e_{it}e_{js})$ exists for all (i,j,s,t). $|\tau_{ij,tt}| \leq |\tau_{ij}|$ for some τ_{ij} and for all t, while $N^{-1}\sum_{i,j=1}^{N} |\tau_{ij}| \leq M$. In addition, $(NT)^{-1}\sum_{i,j=1}^{N}\sum_{s,t=1}^{T} |\tau_{ij,ts}| \leq M$.*
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 temporal 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. Their interpretation will be made clear in examples below. 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,

$$\sup_{s,t\leq T}\sum_{i,j=1}^{N}|E(\xi_{isp_1}\xi_{jtq_1})|\leq\tilde{Q}_1(N,T) \tag{3}$$

and

$$\sum_{s,t=1}^{T}\sum_{i,j=1}^{N}|E(\xi_{isp_1}\xi_{jsq_1}\xi_{itp_2}\xi_{jtq_2})|\leq\tilde{Q}_3(N,T). \tag{4}$$

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

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.[5]

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

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 (3) reduces to $NE(\xi_{it}^2)$. By writing out terms, it may be verified that the quadruple sum in condition (4) 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

$$\begin{aligned}\sup_{s,t\leq T}\sum_{i=1}^{N}\sum_{j=1}^{N}|E(\xi_{is}\xi_{jt})| &= N\sup_{s,t\leq T}|E(\xi_{is}\xi_{it})|\\ &= N\sup_{s,t\leq T}\min\{s,t\}E(\zeta_{i1}^2)\\ &= O(NT),\end{aligned}$$

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 (4) is $O(N^2T^4)$, so Assumption 4.3 holds with $Q_3(N,T)=O(N^2T^4)$.

In Example 3 (single large break), the supremum in inequality (3) 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 (4) 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.

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

3 Consistent estimation of the factor space

3.1 Main result

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

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

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*a*). The following corollary is a straight-forward consequence of Theorem 1.

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

- $h_{NT}^2 Q_1(N,T) = O(N)$,
- $h_{NT}^2 Q_2(N,T) = O(NT^2)$,
- $h_{NT}^4 C_{NT}^2 Q_3(N,T) = O(N^2T^2)$,

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 note 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 consistency 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).[6] To elaborate on the convergence rate in Theorem 1, suppose we set $N=[T^{\mu}]$ and $h_{NT}=cT^{-\gamma}$, $\mu,\gamma\geq 0$. Using the formula for R_{NT} and the random walk calculations in Section 2.4, we obtain

$$R_{NT}=O(\max\{T^{-1},T^{-\mu},T^{1-2\gamma-\mu},T^{2-4\gamma}\})=O(T^{m(\mu,\gamma)}), \tag{5}$$

where

$$m(\mu,\gamma)=\max\{-1,-\mu,1-2\gamma-\mu,2-4\gamma\}=\max\{-1,-\mu,2-4\gamma\}. \tag{6}$$

This convergence rate exponent reflects the influence of the magnitude of the random walk deviations, as measured by γ, and the relative sizes of the cross-sectional and temporal dimensions, as measured by μ. Evidently, increasing the number of available series relative to the sample size improves the worst-case convergence rate, but only up to a point. The dependence of the convergence rate on γ is monotonic, as expected, but nonlinear.

For Example 3 (single large 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 consistency of the estimator.

3.2 Detailed calculations for special cases

Theorem 1 shows the convergence rate of the principal components estimator but does not offer any information on the constant of proportionality, which in general will depend on the size of the various moments in Assumptions 1–4. In this subsection we consider examples in which we can say more about the speed of convergence.

For analytical tractability, we assume in this subsection that the initial factor loadings Λ_0 are 0 and the true number of factors r is 1. When $\Lambda_0=0$, the matrix H^k in Theorem 1 is equal to zero, so that consistency of the principal components estimator hinges on how fast the norm of $\hat{F}_t^k$ tends to zero in mean square.[7] As shown in the appendix, when $\Lambda_0=0$ and $r=1$,

$$T^{-1}\sum_{t=1}^{T}\|\hat{F}_t^k-H^{k\prime}F_t\|^2=(NT)^{-2}\sum_{l=1}^{k}\omega_l^2,$$

were ω_l is the l-th largest eigenvalue of the $T\times T$ matrix XX'.

[6] Empirical implementations of principal components estimation of structurally unstable DFMs, such as Eickmeier et al. (2011) and Korobilis (forthcoming), rely on robustness of the estimator to small degrees of instability. Our theorem shows that the asymptotically allowable amount of instability is larger than hitherto assumed.

[7] Note that while $\Lambda_0=0$ violates Assumption 2, the proof of Theorem 1 does not rely on the matrix $D=\operatorname{plim}\Lambda_0'\Lambda_0/N$ being positive definite.

Example 1 (white noise, continued). Suppose the single factor is identically 1 ($F_t \equiv 1$), and N and T tend to infinity at the relative rate $\theta = \lim_{N\to\infty} T/N$, $\theta \in (0, \infty)$. Let the idiosyncratic errors e_{it} be i.i.d. across i and t with $E(e_{it}^2) = \sigma_e^2$. Denote $\sigma_\xi^2 = E(\xi_{it}^2)$. The appendix shows that if the number of estimated factors is $k = 1$, then

$$T^{-1}\sum_{t=1}^{T} \|\hat{F}_t^k - H^{k\prime}F_t\|^2 = T^{-2}(\sigma_e^2 + h_{NT}^2\sigma_\xi^2)^2(1+\sqrt{\theta})^4(1+o_p(1)). \tag{7}$$

When $h_{NT} = 1$, the right-hand side quantity is $O_p(T^{-2})$, which is stronger than the $O_p(C_{NT}^{-2})$ rate bound in Theorem 1. Introducing cross-sectional and temporal dependence in the idiosyncratic errors causes the left-hand side above to achieve the worst-case rate asymptotically, as noted by Bai and Ng (2002, pp. 199–200). According to the expression on the right-hand side of equation (7), h_{NT}^2 measures the importance of the factor loading disturbance variance relative to the idiosyncratic error variance. Furthermore, for given T, the mean square error of the principal components estimator increases with the ratio $\theta \approx T/N$.

Example 2 (random walk, continued). Suppose that the idiosyncratic errors are cross-sectionally i.i.d. Denote $\sigma_\zeta^2 = E(\zeta_{it}^2)$. If the number of estimated factors is $k = 1$, we show in the appendix that

$$E\left(T^{-1}\sum_{t=1}^{T} \|\hat{F}_t^k - H^{k\prime}F_t\|^2\right) \geq \left\{T^{-2}\sum_{s,t=1}^{T}\left[\gamma_N(s,t) + h_{NT}^2\sigma_\zeta^2 \min\{s,t\}E(F_sF_t)\right]\right\}^2, \tag{8}$$

where $\gamma_N(s,t)$ is defined in Assumption 3. This lower bound on the expectation of the mean square error of the principal components estimator complements the upper rate bound in Theorem 1. The expression reinforces the intuition that the factor space will be poorly estimated in models with persistent errors (here e_{it} and $h_{NT}\xi_{it}'F_t$).

Without prior knowledge about the factor process, a conservative benchmark sets $E(F_sF_t) = O(1)$. Note that $\sum_{s,t=1}^{T}\min\{s,t\} = \frac{1}{3}T^3 + O(T^2)$, and $\sum_{s,t=1}^{T}\gamma_N(s,t) = O(T)$ by Assumption 3. If $h_{NT} \geq T^{-1}$ asymptotically, the right-hand side of inequality (8) is then of order $h_{NT}^4T^2$. Together with Theorem 1, this establishes that there exist constants $\underline{C}, \overline{C} > 0$ such that

$$\underline{C} \leq (h_{NT}^2T)^{-2}E\left(T^{-1}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k\prime}F_t\|^2\right) \leq \overline{C}\max\{(h_{NT}^2TC_{NT})^{-2}, 1\}$$

for sufficiently large N and T.[8] The maximum on the right-hand side above tends to 1 as long as $h_{NT} \geq (TC_{NT})^{-1/2} = 1/\min\{N^{1/4}T^{1/2}, T^{3/4}\}$ asymptotically.[9] Thus, unless we have special knowledge about the factor process, we generically need $h_{NT} = o(T^{-1/2})$ for mean square consistency of the factors, while $h_{NT} = O(1/\min\{N^{1/4}T^{1/2}, T^{3/4}\})$ is generically necessary to achieve the Bai and Ng (2002) convergence rate.

[8] The rate bound in Theorem 1 is in probability, but the proof given in the appendix shows that the bound holds in expectation as well.

[9] Recall that such rates for h_{NT} are exactly the ones we are most interested in, since any faster rate of decay for h_{NT} will lead to $R_{NT} = C_{NT}^{-2}$ in Theorem 1.

Example 3 (single large break, continued). Here we consider a limiting case with $e_{it} \equiv 0$, so that all the variance in the observed data is due to structural instability. Suppose the single factor F_t satisfies $(T-\kappa)^{-1}\sum_{t=\kappa+1}^{T} F_t^2 \xrightarrow{p} \tilde{\Sigma}_F$ as $T \to \infty$. Then, regardless of the number of estimated factors k,

$$T^{-1}\sum_{t=1}^{T} \|\hat{F}_t^k - H^{k\prime}F_t\|^2 = \frac{h_{NT}^4}{N^2}\|\Delta\|^4(1-\bar{\tau})^2\tilde{\Sigma}_F^2(1+o_p(1)), \tag{9}$$

as shown in the appendix. The result indicates that the mean square error of the principal components estimator is larger the smaller is $\bar{\tau}$ (the break fraction), the larger is $\tilde{\Sigma}_F$ (the post-break factor second moment), and the larger is $\|\Delta\|$ (the size of the break vector). Note that if the elements of Δ are uniformly bounded, $|\Delta_i| \leq M$, then $\|\Delta\|^2$ is on the order of the number of series undergoing a break. Denote this number by B_{NT}. The right-hand side above is then $O_p\left((h_{NT}^2 B_{NT}/N)^2\right)$, which is also the rate stated in the bound in Theorem 1, provided that $h_{NT} = 1$.

4 Rank selection and diffusion index forecasting

4.1 Estimating the number of factors

Bai and Ng (2002, 2006*b*) introduce a class of information criteria that consistently estimate the true number r of factors when the factor loadings are constant through time. Specifically, define the two classes of criteria

$$PC(k) = V(k) + kg(N,T), \quad IC(k) = \log V(k) + kg(N,T), \tag{10}$$

where $V(k)$ is the sum of squared residuals defined by (2), and $g(N,T)$ is a deterministic function satisfying $g(N,T) \to 0$, $C_{NT}^2 g(N,T) \to \infty$ as $N,T \to \infty$. Let $k_{\max} \geq r$ be an upper bound on the estimated rank. With constant factor loadings, a consistent estimate of r is then given by either $\hat{k} = \arg\min_{0\leq k\leq k_{\max}} PC(k)$ or $\hat{k} = \arg\min_{0\leq k\leq k_{\max}} IC(k)$.

Lemma 2 of Amengual and Watson (2007) establishes that 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 a $T \times N$ 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 $\Lambda_0 F_t + e_t$. The following result is therefore a direct consequence of Lemma 2 of Amengual and Watson (2007) and Markov's inequality.

Observation 1. *Let assumptions (A1)–(A9) in Amengual and Watson (2007) hold. If in addition*

$$h_{NT}^2 \sum_{i=1}^{N}\sum_{t=1}^{T} E[(\xi_{it}'F_t)^2] = O(\max\{N,T\}), \tag{11}$$

then $\arg\min_{0\leq k\leq k_{\max}} PC(k) \xrightarrow{p} r$ *and* $\arg\min_{0\leq k\leq k_{\max}} IC(k) \xrightarrow{p} r$ *as* $N,T \to \infty$.

In the interest of brevity we do not state the precise Amengual and Watson conditions here but remark that they are very similar to our Assumptions 1–3 and 5. We now comment on how the sufficient condition in Observation 1 bears on our three examples of structural breaks. The finite-sample performance of the information criteria will be explored in Section 5.

Examples (continued). If $r = 1$ and ξ_{it} is independent of F_t, the left-hand side of condition (11) is of order $h_{NT}^2 \sum_{i=1}^N \sum_{t=1}^T E(\xi_{it}^2)$. In Example 1 (white noise), $\sum_{i=1}^N \sum_{t=1}^T E(\xi_{it}^2) = O(NT)$, so condition (11) holds if $h_{NT} = O(C_{NT}^{-1})$. The white noise disturbances must therefore vanish asymptotically, albeit slowly, for the Amengual and Watson (2007) result to ensure consistent estimation of the factor rank.

For Example 2 (random walk), $\sum_{i=1}^N \sum_{t=1}^T E(\xi_{it}^2) = O(NT^2)$, implying that we need $h_{NT} = O(1/\min\{T, (NT)^{1/2}\})$ to fulfill condition (11). In particular, the Stock and Watson (2002) assumption $h_{NT} = O(1/T)$ admits consistent estimation of the true number of factors using the Bai and Ng (2002) information criteria.

In Example 3 (single large break), we set $h_{NT} = 1$ as before. If $(T-\kappa)^{-1} \sum_{t=\kappa+1}^T E(F_t^2) = O(1)$, we get $\sum_{i=1}^N \sum_{t=1}^T E(\xi_{it}^2) = O(T\|\Delta\|^2)$, so $\|\Delta\|^2 = O(\max\{N/T, 1\})$ is needed to satisfy condition (11). As previously explained, if the elements of Δ are uniformly bounded, $\|\Delta\|^2$ is on the order of the number B_{NT} of series undergoing a break at time $t = \kappa + 1$. The fraction B_{NT}/N of series undergoing a break must therefore be of order at most C_{NT}^{-2} for the Amengual and Watson (2007) result to apply. The conclusion that large breaks are more problematic for rank estimation than for mean square consistency is not surprising given Breitung and Eickmeier's (2011) insight that the large break model (with non-vanishing break parameter) is equivalent to a DFM with $2r$ factors.

In summary, in all three of our examples we need more stringent assumptions on h_{NT} in order to ensure consistent estimation of r than we did for consistency of the principal components estimator. It is a topic for future research to determine whether these tentative results can be improved upon.

4.2 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*a*). 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 ε_{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 α?

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, α 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*a*) show that

$$\sqrt{T}(\hat{\delta} - \delta) = (T^{-1}\hat{F}'\hat{F})^{-1}T^{-1/2}\hat{F}'\varepsilon - (T^{-1}\hat{F}'\hat{F})^{-1}[T^{-1/2}\hat{F}'(\hat{F} - FH)]H^{-1}\alpha, \tag{12}$$

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

yields

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

Similarly,

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

Suppose $T^{-1/2}F'\varepsilon = O_p(1)$, as implied by Assumption E in Bai and Ng (2006*a*). It is easy to show that $H = O_p(1)$. Provided $T = O(N)$, we thus obtain $\hat{\delta}-\delta = O_p(T^{-1/2})$, i.e., under the conditions of Corollary 1, the feasible diffusion regression estimator is consistent at the usual rate. The restrictions on h_{NT} for the three examples are discussed immediately following Corollary 1.[10]

5 Simulations

5.1 Design

To illustrate our results and assess their finite sample validity we conduct a Monte Carlo simulation study. Stock and Watson (2002) and Eickmeier et al. (2011) numerically evaluate the performance of the principal components estimator when the factor loadings evolve as random walks, and Banerjee et al. (2008) focus in particular on the effect of time variation in short samples.[11] We provide additional evidence on the necessary scale factor h_{NT} for the random walk case (our Example 2). Moreover, we consider data generating processes (DGPs) in which the factor loadings are subject to white noise disturbances (as in Example 1), as well as DGPs for which a subset of the series undergo one large break in their factor loadings (an analog of Example 3).

The design broadly follows Stock and Watson (2002):

$$X_{it} = \lambda_{it}'F_t + e_{it}, \quad F_{tp} = \rho F_{t-1,p} + u_{tp}, \quad (1-aL)e_{it} = v_{it}, \quad y_{t+1} = \sum_{q=1}^{r} F_{tq} + \varepsilon_{t+1},$$

where $i = 1,\ldots,N$, $t = 1,\ldots,T$, $p = 1,\ldots,r$. The processes $\{u_{tp}\}$, $\{v_{it}\}$ and $\{\varepsilon_{t+1}\}$ are mutually independent, with u_{tp} and ε_{t+1} being i.i.d. standard normally distributed. To capture cross-sectional dependence of the idiosyncratic errors, we let $v_t = (v_{1t},\ldots,v_{Nt})'$ be i.i.d. normally distributed with covariance matrix $\Omega = (\beta^{|i-j|})_{ij}$, as in Amengual and Watson (2007). The

[10] If $\alpha = 0$, which is often an interesting null hypothesis in applied work, the second term on the right-hand side of the decomposition (12) 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 (12) will (under weak conditions) obey a central limit theorem, and so $\hat{\delta}$ should be unconditionally asymptotically normally distributed under the null $H_0\colon \alpha = 0$. Bai and Ng (2006*a*) 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 α, 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.

[11] The Eickmeier et al. (2011) Monte Carlo study appears in the updated version of their paper dated October 15, 2012.

scalar ρ is the common AR(1) coefficient for the r factors, while a is the AR(1) coefficient for the idiosyncratic errors.

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 Λ_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 $[0, 0.8]$. 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.[12] The scalar $\lambda^*(R_i^2)$ is given by the value for which $E[(\lambda'_{0i}F_0)^2|R_i^2]/E[X_{i0}^2|R_i^2] = R_i^2$, given the draw of R_i^2.[13]

We consider three different specifications for the evolution of factor loadings over time. In the *white noise model* the loadings are given by

$$\lambda_{itp} = \lambda_{i0p} + d\xi_{itp},$$

$i = 1, \ldots, N$, $t = 1, \ldots, T$, $p = 1, \ldots, r$, where d is a constant and the disturbances ξ_{itp} are i.i.d. standard normal and independent of all other disturbances. Note that the standard deviation of $\lambda_{itp} - \lambda_{i0p}$ is d for all t.

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 ζ_{itp} are i.i.d. standard normal and independent of everything. Note that the $T^{3/4}$ rate is different from the rate of T used by Stock and Watson (2002) and Banerjee et al. (2008). In our design, the standard deviation of $\lambda_{iTp} - \lambda_{i0p}$ is $cT^{-1/4}$.

In the *large break model* we select a subset J of size $[bN^{1/2}]$ uniformly at random 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 Δ_p (which is the same for all $i \in J$) is distributed $\mathcal{N}(0, [\lambda^*(0.4)]^2)$, i.i.d. across $p = 1, \ldots, r$, so that the shift is of the same magnitude as the initial loading λ_{i0p}.[14] The fraction of series that undergo a shift in the large break model is $[bN^{1/2}]/N \approx bN^{-1/2}$.

The principal components estimator $\hat{F}^k$ described earlier is used to estimate the factors. Estimation of the factor rank r is done using the "IC_{p2}" information criterion of Bai and Ng (2002) with a maximum rank of $r_{\max} = 10$, and, for simplicity, a minimum estimated rank of 1. The criterion is of the IC type in definition (10) with $g(N,T) = (\log C_{NT}^2)(N+T)/(NT)$. We also consider principal components estimates that impose the true rank $k = r$. To evaluate the principal components estimator's performance, we compute a trace R^2 statistic for the multivariate regression of $\hat{F}$ onto F,

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

[12] We assumed above that Λ_0 is fixed for simplicity. It is not difficult to verify that Λ_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*a*).

[13] Specifically, $[\lambda^*(R_i^2)]^2 = \frac{1-\rho^2}{r(1-a^2)}\frac{R_i^2}{1-R_i^2}$.

[14] This shift process satisfies Assumption 4 with envelope functions of the same order as was used for the deterministic break in Example 3.

where $\hat{E}$ denotes averaging over Monte Carlo repetitions and $P_F = F(F'F)^{-1}F'$. 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} on 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).

5.2 Calibration

The free parameters are T, N, r, ρ, a, β, b, c and d. We set $r = 5$ throughout. In line with Stock and Watson (2002) and Amengual and Watson (2007), we consider $\rho = 0, 0.9$, $a = 0, 0.5$ and $\beta = 0, 0.5$.

To guide our choice of the crucial parameters b, c and d, we turn to the empirical analysis of Stock and Watson (2009). They fit a DFM to 144 quarterly U.S. macroeconomic time series from 1959 to 2006, splitting the sample at the first quarter of 1984. Using their results, we compare the pre- and post-break estimated factor loadings. The ratio of the mean square changes in the factor loadings to the mean square pre-break factor loadings is 0.21. Assuming that the break date and factor loadings are known, the corresponding ratio in our large break DGP is

$$\frac{(Nr)^{-1}\sum_{i=1}^N\sum_{p=1}^r \Delta_p^2}{(Nr)^{-1}\sum_{i=1}^N\sum_{p=1}^r \lambda_{i0p}^2} = \frac{bN^{-1/2}[\lambda^*(0.4)]^2}{\int_0^{0.8}[\lambda^*(x)]^2dx/0.8} + o_p(1) = 0.66bN^{-1/2} + o_p(1),$$

regardless of the values of r, ρ, a and β. For $N = 144$ series, the value of the parameter b that brings the theoretical ratio in line with the observed one in the Stock and Watson (2009) dataset is $b = \sqrt{144} \cdot 0.21/0.66 = 3.7$. While we have ignored estimation error, it therefore seems empirically relevant to consider large break DGPs with a b of this magnitude. We pick $b = 3.5$ to be our benchmark value, which for $N = 100$ implies that $bN^{-1/2} = 35\%$ of the loadings undergo a break (for $N = 200$ and $N = 400$ the fraction is 25% and 18%, respectively). To stress test our conclusions, we also examine the extreme choice $b = 7$.

When calibrating the values of c and d, we take the following steps. Focusing on the parametrization $N = T = 200$ and $a = \beta = \rho = 0$, we first record the trace R^2 statistics for the large break DGPs with $b = 3.5$ and $b = 7$, respectively. We then determine round values of c and d such that the corresponding trace R^2 statistics for the random walk and white noise DGPs approximately match the above-mentioned two figures for the large break model. This yields $c = 2, 3.5$ and $d = 0.4, 0.7$. To compare the time variation with the scale of the initial factor loadings, note that with $a = \beta = \rho = 0$ and $r = 5$, the unconditional standard deviation of each initial factor loading is $\sqrt{\int_0^{0.8}[\lambda^*(x)]^2dx/0.8} = 0.45$. Because d is the standard deviation of $\lambda_{itp} - \lambda_{i0p}$ in the white noise model, the choice $d = 0.4$ creates fluctuations of about the same magnitude as the initial factor loadings. Similarly, the standard deviation of $\lambda_{iTp} - \lambda_{i0p}$ in the random walk model equals $cT^{-1/4} = 0.53$ for $T = 200$ and $c = 2$. In the appendix we show that this amount of random walk

parameter variation is of the same magnitude as the estimates for U.S. data presented in Eickmeier et al. (2011), while our $c = 3.5$ parametrization exhibits substantially more instability.[15]

5.3 Results

We perform 5,000 Monte Carlo repetitions for each DGP. To graphically illustrate the convergence properties of the principal components estimator, we first focus on the baseline set-up with $a = \beta = \rho = 0$, $N = T$ and $k = r$ (the true number of factors is known). We run simulations for a fine grid of T values, $T = 50, 100, 150, \ldots, 400$. The results are plotted in Figures 1–3, corresponding to the white noise, random walk and large break models, respectively. 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 b, c or d. Similarly, the bottom panel shows the $S^2_{\hat{y},\tilde{y}}$ statistic. All figures confirm that, while time variation in the factor loadings, vanishing at the appropriate rate, does impact the precision of the principal components estimator, the performance improves as T increases, both in absolute terms and relative to the no-instability benchmark.

Tables 1–3 display a more comprehensive range of simulation results for the white noise, random walk and large break models, respectively. As explained above, we consider two values each for the instability parameters b, c and d, and each table compares those results to the no-instability benchmark ($b = c = d = 0$). The columns marked "$k = r$" impose knowledge of the true number of factors, while the columns marked "IC" correspond to simulations in which the factor rank is estimated using an information criterion. $\hat{E}(\hat{k})$ denotes the average estimated rank. We focus on dataset dimensions that are especially relevant for macroeconomic analyses with quarterly data, namely $T = 50, 100, 200$ and N either equal to, half of or double the value of T.

Our first set of simulations has $a = \beta = \rho = 0$, i.e., no serial or cross-sectional dependence in the factors or idiosyncratic errors. For the empirically calibrated amount of instability (the middle five columns in each table), the $R^2_{\hat{F},F}$ and $S^2_{\hat{y},\tilde{y}}$ statistics are close to the no-instability benchmark as long as $N \geq T \geq 100$. The average estimated rank is also close to the truth $r = 5$ in these cases. Throughout Table 3 and Figure 2, the large break model does remarkably well in terms of the closeness $S^2_{\hat{y},\tilde{y}}$ of the feasible and infeasible forecasts, even when a majority of factor loadings undergo a break. As Figure 1 already demonstrated, the white noise model gives comparatively poor results for small T and when $N < T$, as predicted by our $\Lambda_0 = 0$ calculation, cf. expression (7). Increasing the amount of structural instability to extreme values (the right-most five columns of each table) substantially affects the results, more so than the introduction of moderate serial and cross-sectional correlation. The white noise model fares particularly poorly for $d = 0.7$, except when $N > T \geq 200$, and the estimated factor rank tends to severely undershoot the target for small sample sizes, as the common component in the data is diluted by the loading disturbances. For the random walk DGP, while the average estimated rank is hardly affected by moving from $c = 0$ to $c = 2$, extreme structural instability $c = 3.5$ does lead to significant deterioration of the performance of the information criterion; the continual evolution of the factor loadings over time causes overestimation of the number of common factors. For the large break model the information criterion does much better, although it overshoots somewhat, as established by Breitung and Eickmeier (2011).

We consider separately the effects of introducing serial ($a = 0.5$) or cross-sectional ($\beta = 0.5$)

[15] For $T \geq 67$ our worst-case random walk DGP, $c = 3.5$, exhibits more time variation in factor loadings than any of the parametrizations considered by Stock and Watson (2002) and Banerjee et al. (2008).

dependence in the idiosyncratic errors. Moderate serial correlation in the errors is clearly a second-order issue.[16] Exponentially decreasing cross-sectional correlation of the above-mentioned magnitude has only a slightly larger impact. Furthermore, there appears to be no interesting interaction between dependence in the idiosyncratic errors and instability in the factor loadings.

Introducing persistence in the factors ($\rho = 0.9$) dramatically worsens the results for the white noise DGP. For the empirically calibrated amount of instability, $d = 0.4$, the $R^2_{\hat{F},F}$ and $S^2_{\hat{y},\tilde{y}}$ statistics are unacceptably poor, except perhaps for large sample sizes, and the estimated rank is much too low. For the random walk model, factor persistence has a more moderate, but still noticeable, effect. It causes overestimation of the number of factors, which only becomes worse as the sample size increases, and the convergence to 1 of the $R^2_{\hat{F},F}$ and $S^2_{\hat{y},\tilde{y}}$ statistics is not evident for $T \leq 200$.[17] However, the absolute impact of the factor loading instability is not alarming, even for $c = 3.5$, unless consistent estimation of r is viewed as a goal in and of itself. In contrast to the first two models, the large break model does not exhibit noticeable sensitivity to the persistence of the factors. Since serial correlation in the factors tends to bias downward the estimate of the factor rank, it actually partially corrects for the upward bias induced by the one-time loading break.[18]

The last seven rows in the tables display results for the most empirically relevant case in which the factors are persistent and the idiosyncratic errors are both serially and cross-sectionally correlated ($a = \beta = 0.5, \rho = 0.9$). As expected based on the discussion above, these figures are similar to those for $a = \beta = 0, \rho = 0.9$, and we find no interesting compounding effects of the various departures from the baseline parametrization.

We summarize the findings of the simulation study as follows.

- Empirically calibrated structural instability of the random walk or large break variety does not, on average, markedly impact the estimation of the factor space or diffusion index forecasts. Increasing the temporal instability by an order of magnitude does not overturn this conclusion.

- The impact of white noise disturbances is a lot more sensitive to the sample size, to the ratio of N to T (higher is better), and to the persistence of the factors (lower is better). The numbers in Table 1 arguably overstate this sensitivity, since d was calibrated based on a setting with $\rho = 0$ and $N = T = 200$, which is relatively favorable for the white noise model. In a sense, Table 1 documents how well the principal components estimator deals with substantial white noise disturbances when the sample size and relative dimension N/T are both large.

- The correlation structure of the idiosyncratic errors is not an important concern in the exponential design we consider here. We have also tried the linearly decreasing correlation structure of Bai and Ng (2002, section 6). As expected, such a set-up yields worse convergence rates than those exhibited in Table 1–3, although the results are sensitive to the choice of correlation parameters.

- Estimation of the factor rank r is governed by somewhat different forces than estimation of the factor space or diffusion index forecasting, as we anticipated in Section 4.1. Relative

[16] In fact, relative to the i.i.d. benchmark, the $a = 0.5$ results are somewhat better in cases in which the estimated rank is much too low.

[17] In unreported simulations, we have confirmed that these statistics do begin to improve for larger values of T.

[18] In Table 3, the large break model often performs better for $\rho = 0.9$ than for $\rho = 0$. The reason is that the denominators in the $R^2_{\hat{F},F}$ and $S^2_{\hat{y},\tilde{y}}$ statistics tend to increase with the persistence of the factors. For the two other models, the detrimental impact on the numerators outweigh this effect.

to the no-instability benchmark, the Bai and Ng (2002) information criterion estimator is generally biased downward in the white noise model, whereas it is biased upward in the large break model and (especially) the random walk model. In the latter two models, there is no indication that this bias vanishes as $N, T \to \infty$ for the choices of h_{NT} and $\|\Delta\|^2$ that we have considered here. However, overestimation of r is not a problem, on average, for diffusion index forecasting.

5.4 Rate of convergence

We now turn to the more detailed asymptotic rates stated in Theorem 1. Our method of proof and the calculations in Section 3.2 suggest that it may not in general be possible to improve upon the R_{NT} rate for our three examples of break processes. To investigate this claim, we carry out two exercises. First, we set $N = T$ and execute a separate set of simulations in which $\lambda_{itp} - \lambda_{i0p} = dT^{1/4}\xi_{itp}$ for the white noise model, $\lambda_{itp} - \lambda_{i,t-1,p} = cT^{-1/2}\zeta_{itp}$ for the random walk model, and the number of shifting series in the large break model is set to $[bN]$. These three rates all (just) violate the conditions for mean square consistency in Theorem 1. To make the results comparable to Figures 2–3, we scale down our choices of b, c and d so that the amount of time variation in the two experiments coincide for $T = 200$. All other parameters are unchanged. See Figures 4–6 for the results. As hypothesized, for the random walk and large break models the trace R^2 curve flattens out for large T, instead of converging with the no-instability curve as in Figures 2–3. For the white noise model, convergence seems to still obtain with $h_{NT} = dT^{1/4}$.[19] It would be interesting to explore whether temporal or cross-sectional dependence in the disturbances ξ_{it} would make Theorem 1 tight also for the white noise model.

Second, we construct a "rate frontier" that corresponds to the predictions of Theorem 1 for the special case of the random walk model, which is the break process that has received most attention in the literature. Consider the explicit rate expression (5)–(6) for the random walk model under the assumptions $N = [T^\mu]$ and $h_{NT} = cT^{-\gamma}$. In the following we set $\mu = 1$ so that the rate exponent (6) reduces to

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

The flat profile of the trace R^2 statistic in Figure 5 is fully consistent with $m(1/2) = 0$. These calculations pertain to the worst-case rate stated in Theorem 1. While Section 3.2 showed suggestive calculations for the special case $\Lambda_0 = 0$, we have not been able to prove that the convergence rate R_{NT} is sharp, in the sense that a generic DFM with random walk factor loadings that satisfies Assumptions 1–5 achieves the R_{NT} rate. Instead, we provide simulation evidence indicating that the independent random walk model achieves the stated bound. We maintain the simulation design described in Section 5.1 with $a = \beta = \rho = 0$ and $N = T$, except that we set $h_{NT} = 5T^{-\gamma}$ and vary γ over the range $0.25, 0.30, 0.35, \ldots, 1.50$. For each value of γ and each sample size $T = 200, 300, \ldots, 700$ 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

[19] This is consistent with the calculations in Section 3.2, which showed that $h_{NT} = o(T^{1/2})$ is necessary and sufficient for mean square consistency when $\Lambda_0 = 0$, $F_t \equiv 1$, $k = r = 1$ and $T/N \to \theta \in (0, \infty)$, cf. equation (7).

that $\widehat{MSE}(\gamma, T)$ should grow or decay at rate $T^{m(\gamma)}$. We verify this by regressing, for each γ,

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

using our six observations $T = 200, \ldots, 700$. Figure 7 plots the estimates $\hat{m}_\gamma$ against γ 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 γ. This corroborates our conjecture that Theorem 1 provides sharp rates for the independent random walk case.

6 Discussion and conclusions

The theoretical results of Section 3 and the simulation study of Section 5 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 3.1 is tight in a formal sense, inspection of our proof, as well as calculations for special cases and Monte Carlo evidence, do not suggest any room for improvement, particularly for the random walk and large break models. In this sense our rate function 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 to values of N and T typically used in applied work, which is reassuring for the nascent empirical research agenda that allows for structural instability in estimation of DFMs (Eickmeier et al., 2011; Korobilis, forthcoming).

Our evidence concerning the robustness of the principal components estimator raises a tension with the results in Breitung and Eickmeier (2011), who stress the harmful effect of undetected factor loading breaks on rank estimation. Our simulations show that diffusion index forecasting using principal components estimates can be effective even when the rank of the factor space is not estimated consistently. Indeed, we conjecture (but do not prove) that the principal components estimator and feasible diffusion index regression will be consistent under sequences of breaks for which the Breitung and Eickmeier (2011) test rejects. Furthermore, our simulations indicate that the direction of the rank estimation bias depends on the type of structural instability. Sorting out the relative importance of these countervailing forces for the sampling distribution of forecasts 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).

In some applications, such as with data on asset returns, accurate estimation of the number of factors is of direct concern. Our results suggest that the allowable amount of structural instability in these cases is smaller than for forecasting purposes. More work is needed to establish necessary conditions for consistent rank estimation, and if necessary, to develop rank estimators that are more robust to different types of instability.

A Appendix

A.1 Proof of Theorem 1

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 $\tilde{F}^{k\prime}\tilde{F}^k/T = I_k$ (Bai and Ng, 2008). 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

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

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

$$T^{-1}\sum_t \|\hat{F}_t^k - H^{k\prime}F_t\|^2 \le 8\sum_{n=1}^8 \left(T^{-1}\sum_t \|A_{nt}\|^2\right). \tag{13}$$

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 $\mathrm{tr}(\tilde{F}^{k\prime}\tilde{F}^k/T) = \mathrm{tr}(I_k) = k$. The second factor is $O_p(1)$ by Assumption 1. Also,

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

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,

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

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

$$(N^2T)^{-1}E\sum_s(w_s'\Lambda_0F_t)^2 \le (N^2T)^{-1}\sum_s\sum_{i,j}|E(w_{is}w_{js}\lambda_{i0}'F_t\lambda_{j0}'F_t)|$$
$$\le r^4\bar{\lambda}^2\sup_{p_1,q_1,p_2,q_2}h_{NT}^2(N^2T)^{-1}\sum_s\sum_{i,j}|E(\xi_{isp_1}\xi_{jsq_1}F_{sp_1}F_{sq_1}F_{tp_2}F_{tq_2})|.$$

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

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

For the sixth term,

$$E\|A_{6t}\|^2 \le E\left\{\left(T^{-1}\sum_s\|\tilde{F}_s^k\|^2\right)\left((N^2T)^{-1}\sum_s(w_s'w_t)^2\right)\right\}$$
$$= k(N^2T)^{-1}\sum_s\sum_{i,j}E(w_{is}w_{it}w_{js}w_{jt})$$
$$\le kr^4\sup_{p_1,q_1,p_2,q_2}\frac{h_{NT}^4}{N^2T}\sum_s\sum_{i,j}|E(\xi_{isp_1}\xi_{jsq_1}\xi_{itp_2}\xi_{jtq_2}F_{sp_1}F_{sq_1}F_{tp_2}F_{tq_2})|.$$

By Assumption 4.3, it follows that

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

Regarding the seventh term, using Assumption 5,

$$E\|A_{7t}\|^2 \le E\left\{\left(T^{-1}\sum_s\|\tilde{F}_s^k\|^2\right)\left((N^2T)^{-1}\sum_s(e_s'w_t)^2\right)\right\}$$
$$= k(N^2T)^{-1}\sum_s\sum_{i,j}E(e_{is}e_{js})E(w_{it}w_{jt})$$
$$\le k(N^2T)^{-1}\sum_s\sum_{i,j}(E(e_{is}^2)E(e_{js}^2))^{1/2}|E(w_{it}w_{jt})|$$
$$\le kr^2M\sup_{p,q}h_{NT}^2(N^2T)^{-1}\sum_s\sum_{i,j}|E(\xi_{itp}\xi_{jtq}F_{tp}F_{tq})|$$
$$= O(h_{NT}^2N^{-2}Q_1(N,T)),$$

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

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

A similar argument gives

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

We conclude that the right-hand side of inequality (13) is the sum of variables of four stochastic orders: $O_p(C_{NT}^{-2})$, $O_p(h_{NT}^2N^{-2}Q_1(N,T))$, $O_p(h_{NT}^2N^{-2}T^{-2}Q_2(N,T))$ and $O_p(h_{NT}^4N^{-2}T^{-2}Q_3(N,T))$. The statement of the theorem follows. □

A.2 Detailed calculations for the case $\Lambda_0 = 0$

Using the definitions of $\hat{F}^k$ and H^k, we get

$$T^{-1}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k\prime}F_t\|^2 = \text{tr}\left\{(\hat{F}^k - FH^k)(\hat{F}^k - FH^k)'\right\}$$
$$= N^{-2}T^{-3}\text{tr}\left\{\tilde{F}^{k\prime}(XX' - F\Lambda_0'\Lambda_0F')(XX' - F\Lambda_0'\Lambda_0F')'\tilde{F}^k\right\}.$$

Let $\Lambda_0 = 0$. By definition, $\tilde{F}^k$ equals $\sqrt{T}$ times the $T\times k$ matrix whose columns are the eigenvectors of XX' corresponding to its k largest eigenvalues. That is, if we write $(XX')R = RC$, where R is the orthogonal matrix of eigenvectors and C the diagonal matrix of eigenvalues (in descending order), we have $\sqrt{T}R = (\tilde{F}^k, \breve{F}^k)$ for a $T\times(T-k)$ matrix $\breve{F}^k$ that satisfies $\tilde{F}^{k\prime}\breve{F}^k = 0$. Observe that

$$\tilde{F}^k = \sqrt{T}R(I_k, 0_{k\times(T-k)})',$$

so

$$(XX')\tilde{F}^k = \sqrt{T}(XX')R(I_k, 0_{k\times(T-k)})' = \sqrt{T}RC(I_k, 0_{k\times(T-k)})',$$

and

$$\tilde{F}^{k\prime}(XX')(XX')\tilde{F}^k = T(I_k, 0_{k\times(T-k)})CR'RC(I_k, 0_{k\times(T-k)})'$$
$$= TC_k^2,$$

where $C_k = (I_k, 0_{k\times(T-k)})C(I_k, 0_{k\times(T-k)})'$ denotes the diagonal matrix containing the k largest eigenvalues $\omega_1, \ldots, \omega_k$ of XX'. Hence,

$$T^{-1}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k\prime}F_t\|^2 = (NT)^{-2}\text{tr}\{C_k^2\}$$
$$= (NT)^{-2}\sum_{l=1}^{k}\omega_l^2. \tag{14}$$

Example 1 (white noise, continued). Under the assumptions in the main text, the $T\times N$ data matrix X has elements $x_{it} = e_{it} + h_{NT}\xi_{it}$ that are i.i.d. across i and t with mean 0 and variance $\Omega_{NT} = \sigma_e^2 + h_{NT}^2\sigma_\xi^2$. Let Z be a $T\times N$ matrix with elements $z_{it} = x_{it}/\sqrt{\Omega_{NT}}$. Then z_{it} is i.i.d. across i and t with mean zero and unit variance. Let $\tilde{\omega}_1$ denote the largest eigenvalue of the sample covariance matrix $N^{-1}ZZ'$. By Theorem 5.8 of Bai and Silverstein (2009), $\tilde{\omega}_1 \overset{a.s.}{\to} (1+\sqrt{\theta})^2$. Because the largest eigenvalue of $N^{-1}XX'$ satisfies $\omega_1 = \Omega_{NT}\tilde{\omega}_1$, the result (7) follows from (14).

Example 2 (random walk, continued). Let 1_T denote the T-vector of ones. Setting $k = 1$ in equation (14), we obtain

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k\prime}F_t\|^2 &= (NT)^{-2}\left(\max_{v\in\mathbb{R}^T}\frac{v'XX'v}{v'v}\right)^2 \\ &\geq (NT)^{-2}\left(\frac{1_T'XX'1_T}{T}\right)^2 \\ &= \frac{1}{N^2T^4}\left[\sum_{i=1}^{N}\left(\sum_{t=1}^{T}x_{it}\right)^2\right]^2. \end{aligned}$$

Jensen's inequality and cross-sectional i.i.d.-ness of x_{it} implies

$$E\left[\sum_{i=1}^{N}\left(\sum_{t=1}^{T}x_{it}\right)^2\right]^2 \geq \left[NE\left(\sum_{t=1}^{T}x_{it}\right)^2\right]^2,$$

so inequality (8) follows.

Example 3 (single large break, continued). In the large break model, $w_{it} = \Delta_i F_t \mathbf{1}_{\{t\geq\kappa+1\}}$. Denote the last $(T-\kappa)$ elements of the T-vector F by $F_{\kappa+1:T}$. Then we can write

$$w = (w_1,\ldots,w_T)' = h_{NT}\begin{pmatrix} 0_{\kappa\times N} \\ F_{\kappa+1:T}\otimes\Delta' \end{pmatrix},$$

so that

$$ww' = h_{NT}^2\begin{pmatrix} 0_{\kappa\times\kappa} & 0_{\kappa\times(T-\kappa)} \\ 0_{(T-\kappa)\times\kappa} & (F_{\kappa+1:T}F_{\kappa+1:T}')\|\Delta\|^2 \end{pmatrix}.$$

It follows that the eigenvalues of ww' are 0 (with multiplicity κ) along with $h_{NT}^2\|\Delta\|^2$ times the $(T-\kappa)$ eigenvalues of $F_{\kappa+1:T}F_{\kappa+1:T}'$. But the eigenvalues of $F_{\kappa+1:T}F_{\kappa+1:T}'$ are just $\|F_{\kappa+1:T}\|^2$ (with multiplicity 1) and 0 (with multiplicity $T-\kappa-1$). The k largest eigenvalues $\omega_1,\ldots,\omega_k$ of $XX' = ww'$ are therefore

$$\omega_1 = h_{NT}^2\|\Delta\|^2\|F_{\kappa+1:T}\|^2, \quad \omega_2 = \omega_3 = \cdots = \omega_k = 0.$$

Consequently, regardless of the number of estimated factors k,

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}\|\hat{F}_t^k - H^{k\prime}F_t\|^2 &= (NT)^{-2}\sum_{l=1}^{k}\omega_l^2 \\ &= \frac{h_{NT}^4}{(NT)^2}\|\Delta\|^4\|F_{\kappa+1:T}\|^4 \\ &= \frac{h_{NT}^4}{N^2}\|\Delta\|^4(1-\bar{\tau})^2\left(\frac{1}{T-\kappa}\sum_{t=\kappa+1}^{T}F_t^2 + o_p(1)\right)^2 \end{aligned}$$

where the last equality uses $T-\kappa = (1-\bar{\tau})T(1+o(1))$. Expression (9) follows.

A.3 Comparison of our Monte Carlo calibration with Eickmeier et al. (2011)

Eickmeier et al. (2011) use a two-step maximum likelihood procedure to estimate a five-factor DFM with time-varying parameters on quarterly U.S. data from 1972 to 2007. As in some of our simulations, the factor loadings in their model evolve as independent random walks. From their smoothed estimates of the factor loading paths (restricting attention to the paths that exhibit non-negligible time variation) one obtains a median standard deviation of the innovations equal to 0.0165 for loadings on the first factor, which has the largest median loading innovation standard deviation of the five factors. Because their sample size is $T = 140$, the random walk specification implies a median standard deviation of $\lambda_{iT1} - \lambda_{i01}$ of about 0.20; the 95th percentile of the implied standard deviation of $\lambda_{iT1} - \lambda_{i01}$ is about 0.75. The 5–95 percentile range of estimated initial factor loadings is $[-0.87, 0.28]$.[20] As explained in the main text, in our random walk design with $a = \beta = \rho = 0$, $c = 2$ and $T = 200$, the standard deviation of $\lambda_{iT1} - \lambda_{i01}$ is 0.53 for all i, while the 5–95 percentile range for initial factor loadings is $[-0.74, 0.74]$. Our $c = 2$ calibration is therefore similar to the Eickmeier et al. (2011) estimated amount of factor loading time variation in U.S. data, while our $c = 3.5$ simulations appear to exhibit substantially more instability.

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

Bai, J. and Ng, S. (2006*b*), 'Determining the number of factors in approximate factor models, Errata'. Manuscript, Columbia University.

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

Bai, Z. and Silverstein, J. (2009), *Spectral Analysis of Large Dimensional Random Matrices*, Springer Series in Statistics, Springer.

Banerjee, A., Marcellino, M. and Masten, I. (2008), Forecasting Macroeconomic Variables Using Diffusion Indexes in Short Samples with Structural Change, *in* D. E. Rapach and M. E. Wohar, eds, 'Forecasting in the Presence of Structural Breaks and Model Uncertainty (Frontiers of Economics and Globalization, Volume 3)', Emerald Group Publishing Limited, chapter 4, pp. 149–194.

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.

[20] We are grateful to Wolfgang Lemke and Massimiliano Marcellino for helping us obtain these figures.

Eickmeier, S., Lemke, W. and Marcellino, M. (2011), 'Classical time-varying FAVAR models – estimation, forecasting and structural analysis'. Deutsche Bundesbank Discussion Paper, No. 04/2011.

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.

Korobilis, D. (forthcoming), 'Assessing the Transmission of Monetary Policy Using Time-varying Parameter Dynamic Factor Models', *Oxford Bulletin of Economics and Statistics* .

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.

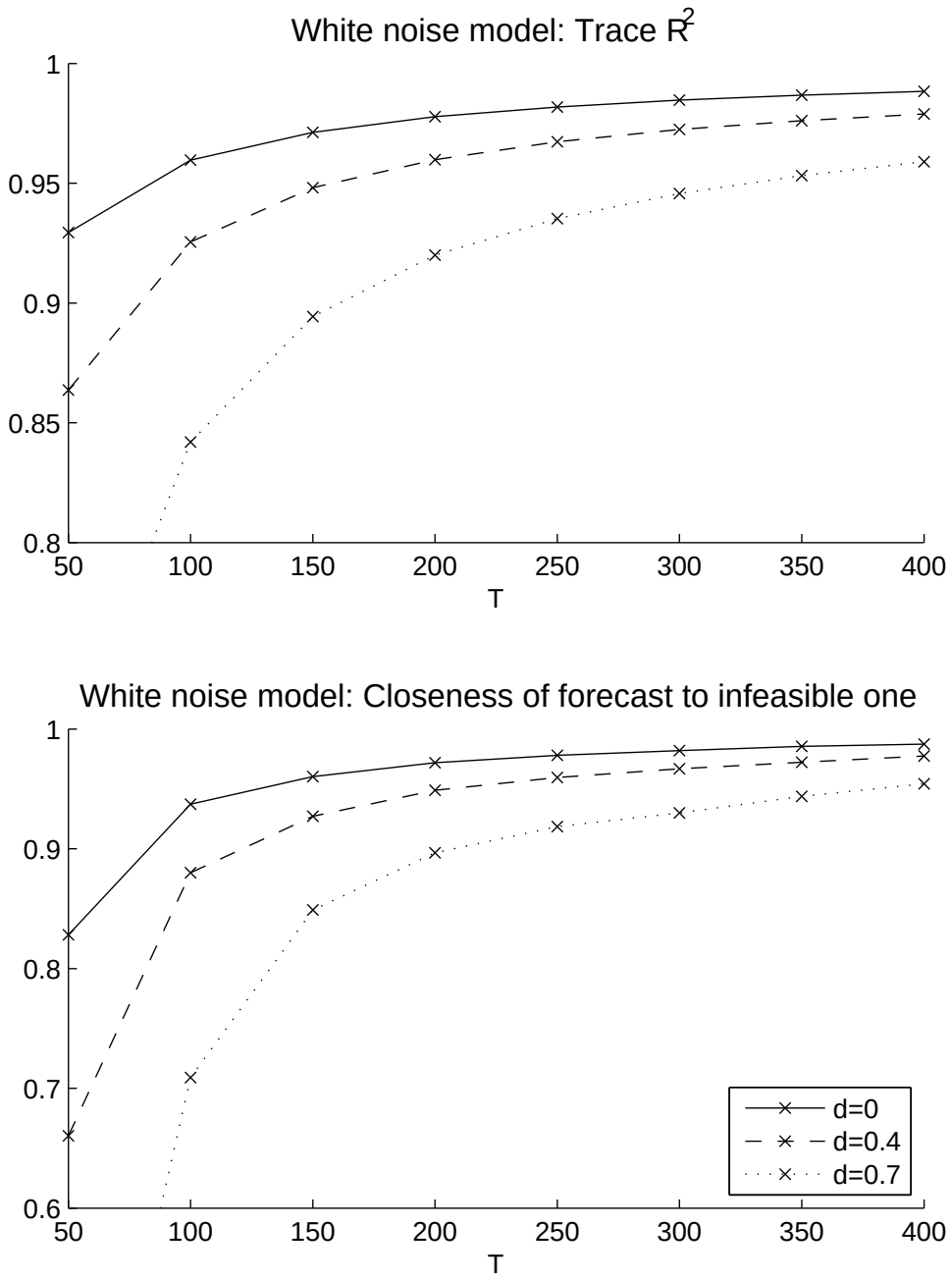

Figure 1: Simulation results for the white noise model, benchmark parameter and rate choices. Actual observations are marked with "x." Each is based on 5,000 Monte Carlo repetitions. The lines are piecewise linear interpolations.

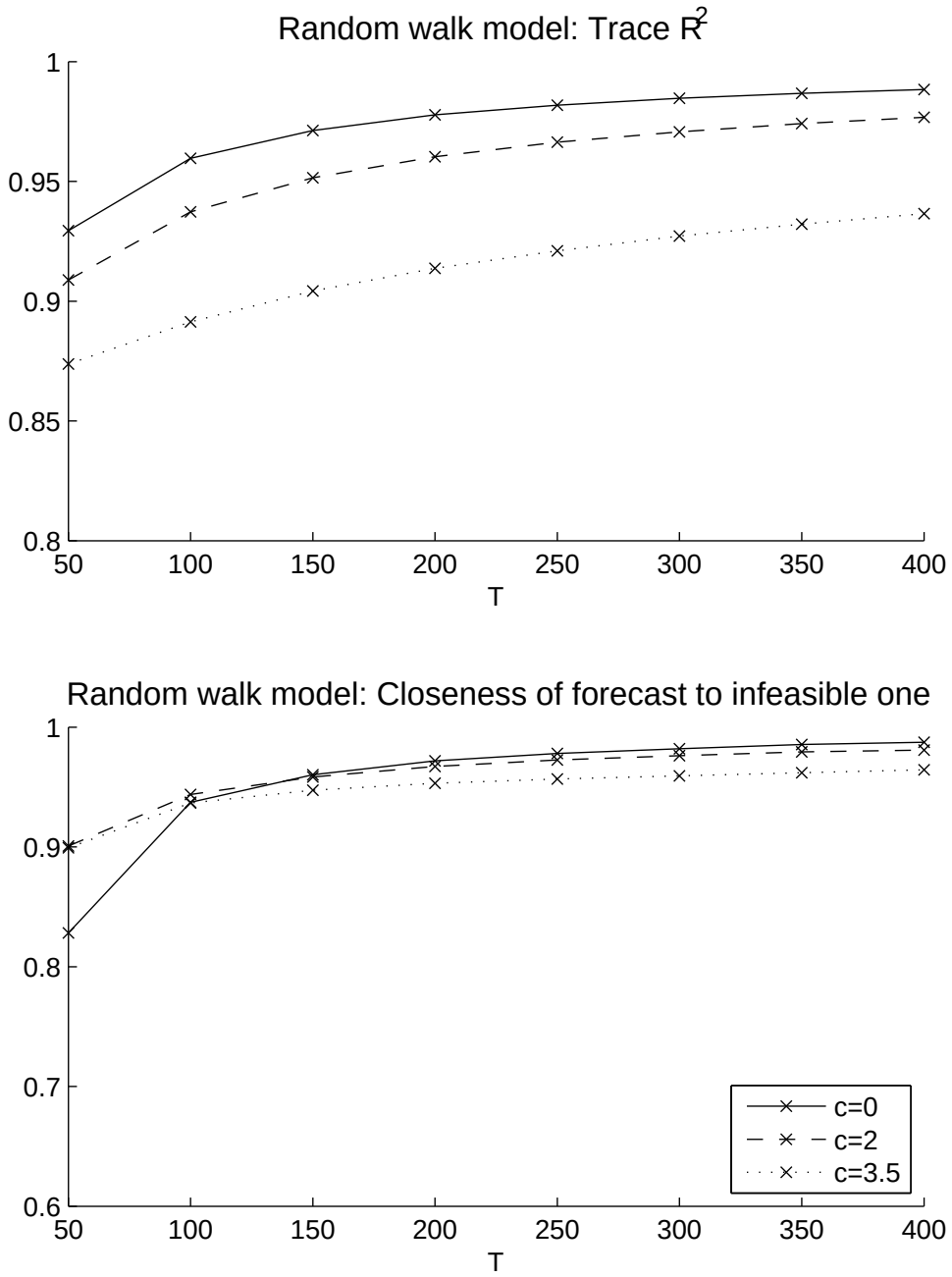

Figure 2: Simulation results for the random walk model, benchmark parameter and rate choices.

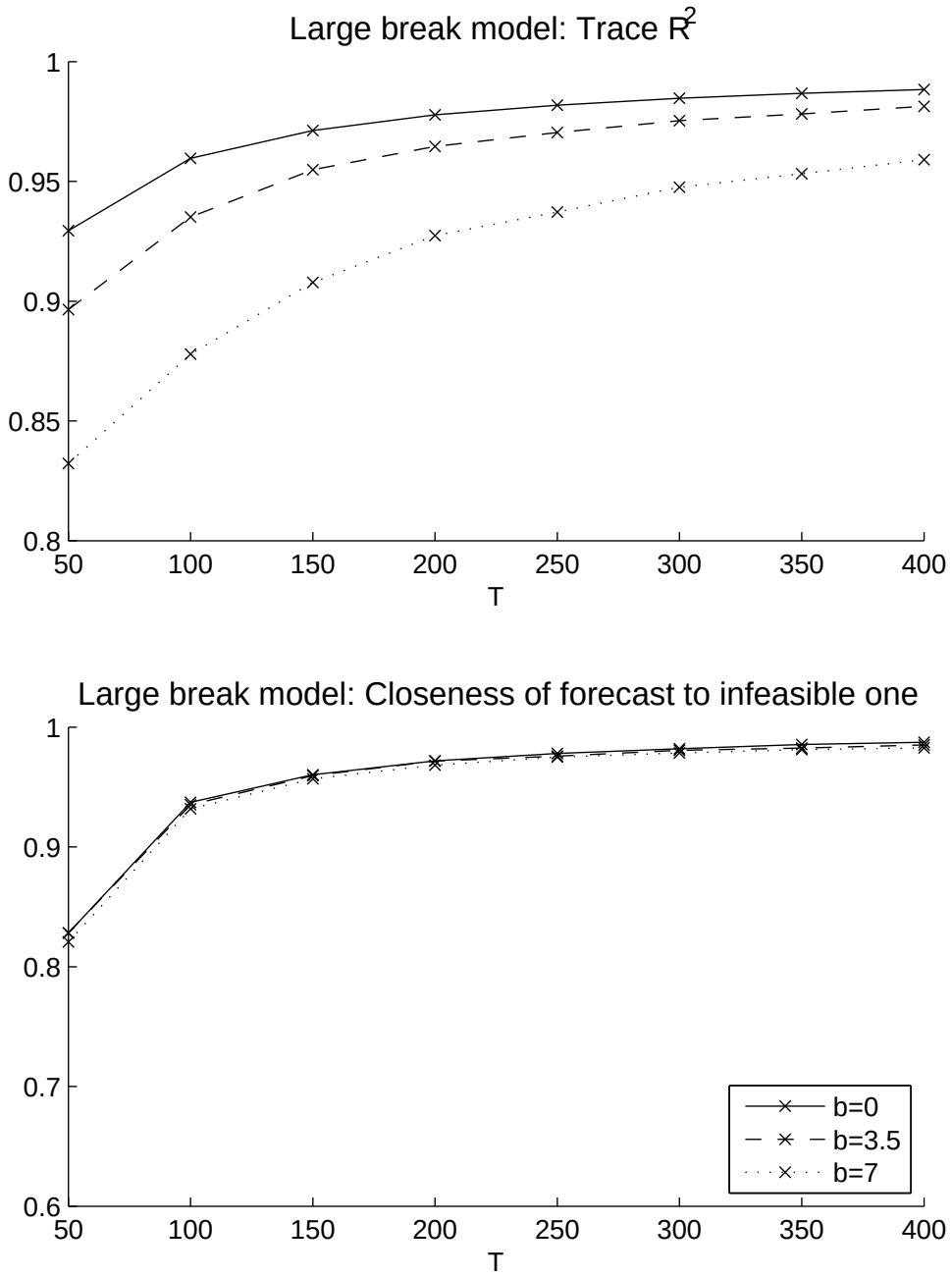

Figure 3: Simulation results for the large break model, benchmark parameter and rate choices.

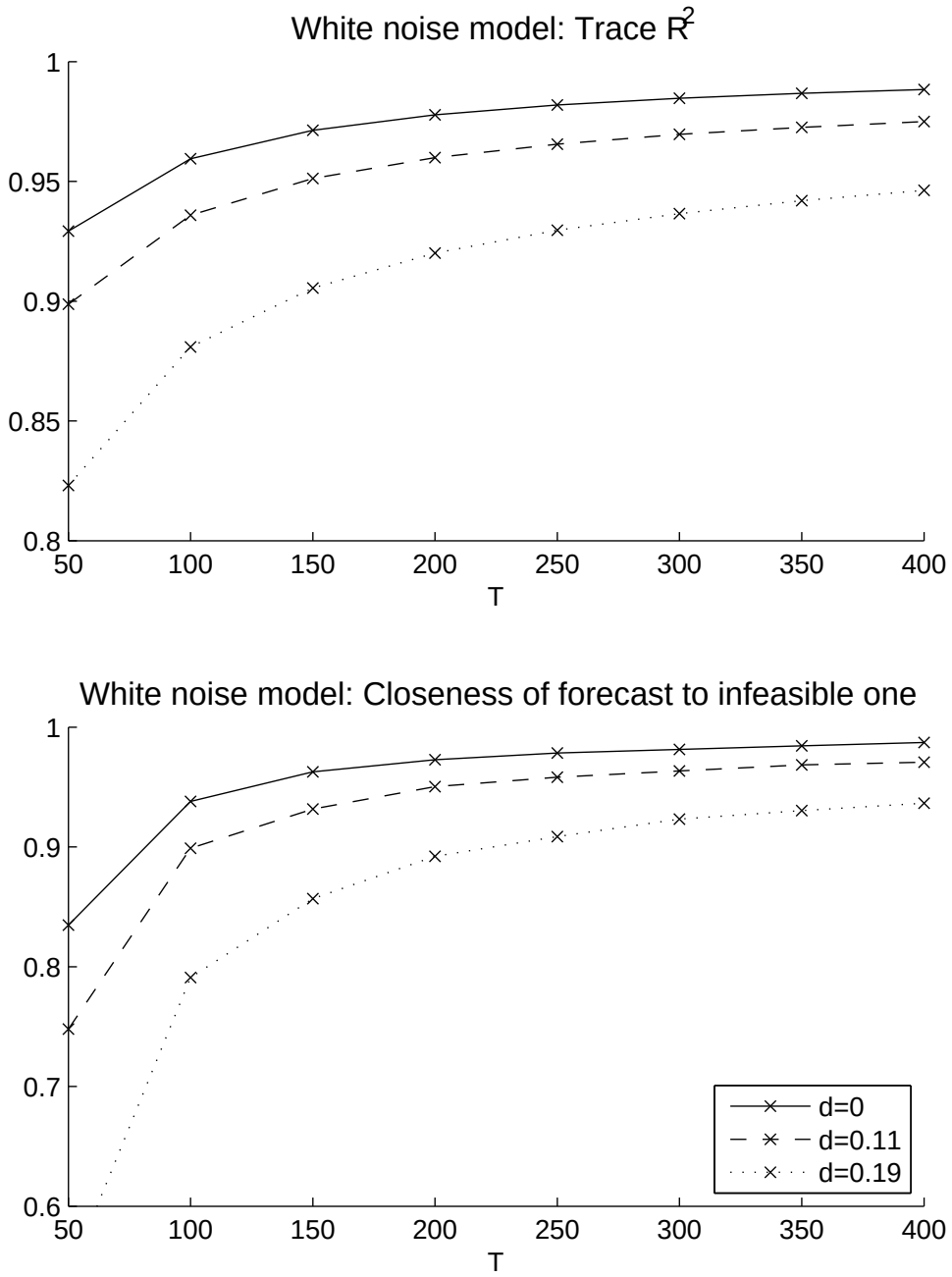

Figure 4: Simulation results for the white noise model, alternative rates.

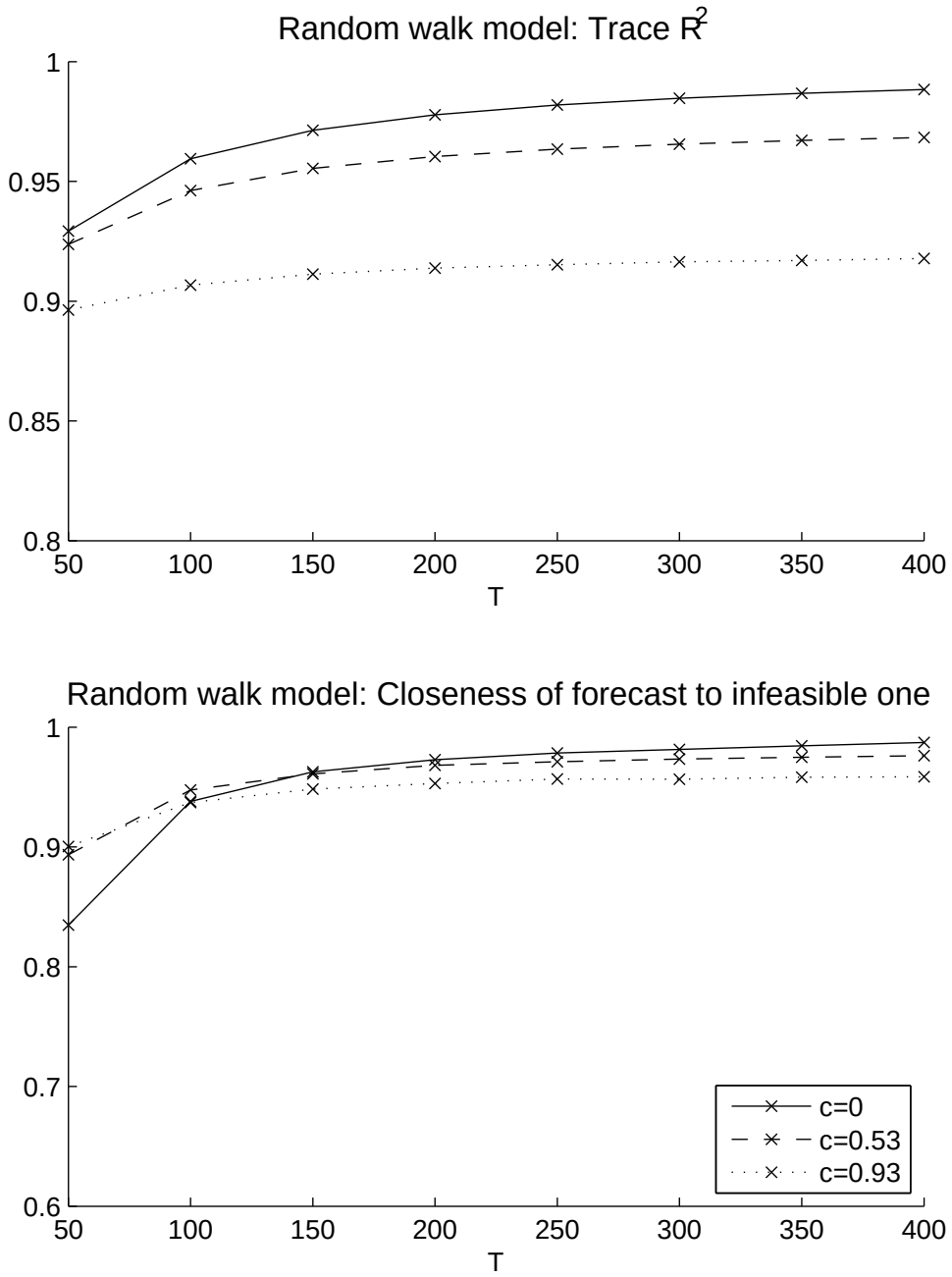

Figure 5: Simulation results for the random walk model, alternative rates.

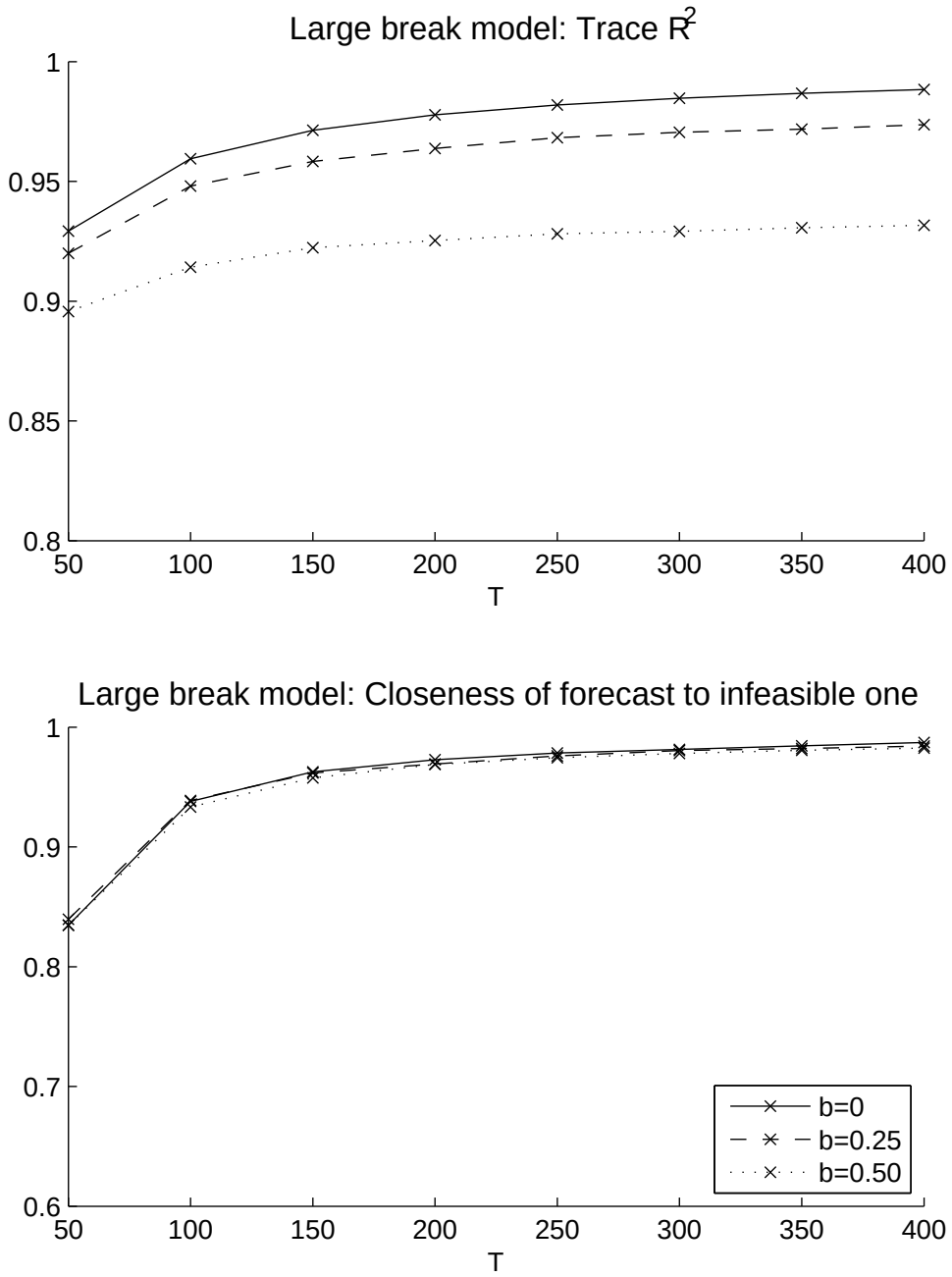

Figure 6: Simulation results for the large break model, alternative rates.

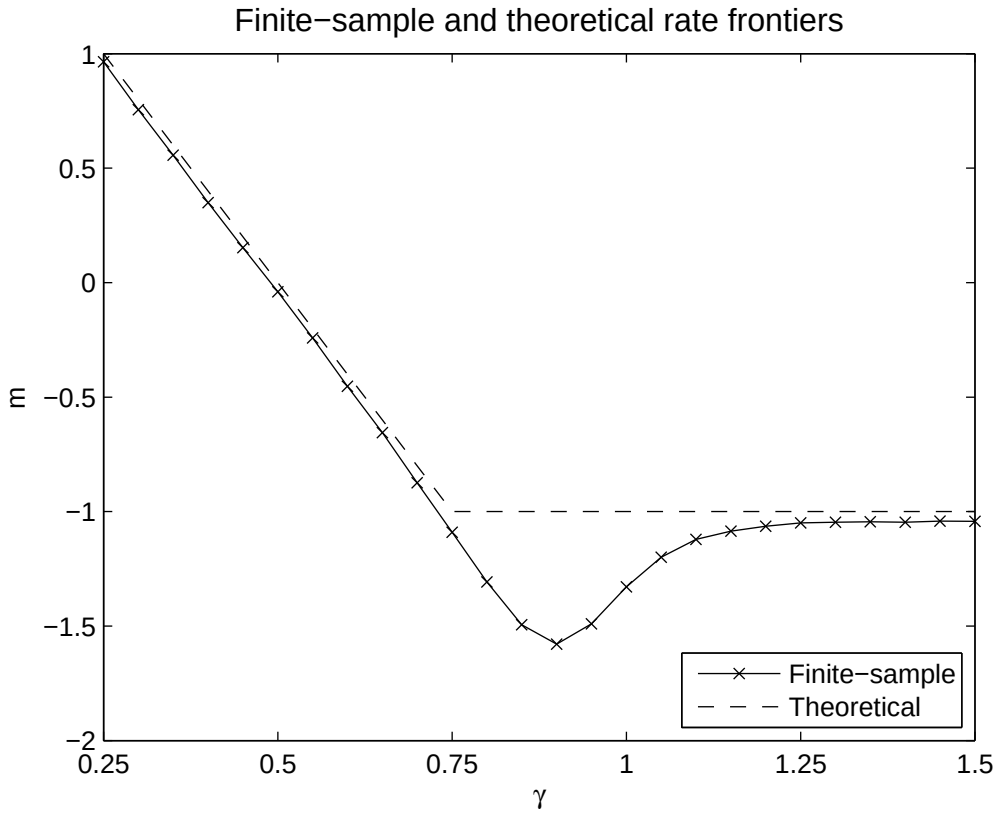

Figure 7: Rate frontiers for the random walk model with $c = 5$, $N = T$ and $h_{NT} = cT^{-\gamma}$. The solid line interpolates between the finite-sample rate exponent estimates $\hat{m}_\gamma$ (observations are marked with "x"), while the dotted line represents the theoretical rate exponent $m(\gamma)$.

Monte Carlo simulations: White noise model

		$d=0$					$d=0.4$					$d=0.7$				
		$k=r$		IC			$k=r$		IC			$k=r$		IC		
T	N	$R^2_{\hat{F},F}$	$S^2_{\hat{y},\tilde{y}}$	$R^2_{\hat{F},F}$	$S^2_{\hat{y},\tilde{y}}$	$\hat{E}(\hat{k})$	$R^2_{\hat{F},F}$	$S^2_{\hat{y},\tilde{y}}$	$R^2_{\hat{F},F}$	$S^2_{\hat{y},\tilde{y}}$	$\hat{E}(\hat{k})$	$R^2_{\hat{F},F}$	$S^2_{\hat{y},\tilde{y}}$	$R^2_{\hat{F},F}$	$S^2_{\hat{y},\tilde{y}}$	$\hat{E}(\hat{k})$
$a=0,\ \beta=0,\ \rho=0$																
50	50	0.93	0.83	0.94	0.50	3.4	0.86	0.64	0.91	−0.83	1.6	0.71	0.25	0.84	−2.77	1.0
50	100	0.96	0.93	0.96	0.88	4.7	0.92	0.84	0.94	0.35	2.9	0.81	0.59	0.89	−1.72	1.1
100	100	0.96	0.94	0.96	0.93	5.0	0.93	0.88	0.93	0.66	3.9	0.84	0.70	0.90	−1.88	1.3
100	200	0.98	0.97	0.98	0.97	5.0	0.96	0.94	0.96	0.93	5.0	0.91	0.86	0.93	0.13	2.7
200	100	0.96	0.94	0.96	0.94	5.0	0.93	0.89	0.93	0.86	4.8	0.86	0.77	0.88	−0.20	2.4
200	200	0.98	0.97	0.98	0.97	5.0	0.96	0.95	0.96	0.95	5.0	0.92	0.90	0.93	0.70	4.1
200	400	0.99	0.99	0.99	0.99	5.0	0.98	0.97	0.98	0.97	5.0	0.96	0.95	0.96	0.94	5.0
$a=0.5,\ \beta=0,\ \rho=0$																
50	50	0.91	0.77	0.93	0.53	3.7	0.86	0.64	0.91	−0.42	2.0	0.75	0.38	0.86	−2.39	1.1
50	100	0.95	0.90	0.95	0.88	4.8	0.92	0.84	0.93	0.57	3.6	0.84	0.68	0.91	−1.01	1.4
100	100	0.96	0.93	0.96	0.92	5.0	0.93	0.89	0.93	0.78	4.4	0.87	0.77	0.91	−0.70	1.9
100	200	0.98	0.97	0.98	0.97	5.0	0.96	0.95	0.96	0.94	5.0	0.93	0.89	0.93	0.65	3.8
200	100	0.96	0.94	0.96	0.94	5.0	0.93	0.90	0.93	0.90	4.9	0.88	0.82	0.89	0.41	3.4
200	200	0.98	0.97	0.98	0.97	5.0	0.96	0.95	0.96	0.95	5.0	0.93	0.92	0.94	0.87	4.8
200	400	0.99	0.99	0.99	0.99	5.0	0.98	0.98	0.98	0.98	5.0	0.96	0.96	0.96	0.96	5.0
$a=0,\ \beta=0.5,\ \rho=0$																
50	50	0.91	0.76	0.93	0.53	3.7	0.85	0.58	0.90	−0.87	1.7	0.70	0.22	0.83	−3.09	1.0
50	100	0.95	0.91	0.96	0.87	4.7	0.92	0.83	0.93	0.39	3.0	0.80	0.57	0.89	−1.58	1.1
100	100	0.96	0.92	0.96	0.92	5.0	0.92	0.87	0.93	0.66	4.0	0.84	0.70	0.89	−1.90	1.3
100	200	0.98	0.97	0.98	0.97	5.0	0.96	0.94	0.96	0.93	5.0	0.91	0.86	0.93	0.15	2.7
200	100	0.96	0.94	0.95	0.94	5.0	0.92	0.88	0.92	0.86	4.8	0.85	0.76	0.88	−0.19	2.4
200	200	0.98	0.97	0.98	0.97	5.0	0.96	0.95	0.96	0.95	5.0	0.92	0.90	0.92	0.70	4.1
200	400	0.99	0.99	0.99	0.99	5.0	0.98	0.97	0.98	0.97	5.0	0.96	0.95	0.96	0.94	5.0
$a=0,\ \beta=0,\ \rho=0.9$																
50	50	0.95	0.81	0.97	0.43	2.3	0.61	0.03	0.84	−1.18	1.0	0.32	−0.96	0.52	−2.94	1.0
50	100	0.97	0.91	0.98	0.69	2.9	0.70	0.37	0.91	−0.75	1.0	0.38	−0.38	0.67	−1.74	1.0
100	100	0.97	0.94	0.97	0.81	3.9	0.75	0.43	0.89	−1.46	1.0	0.39	−0.43	0.67	−2.48	1.0
100	200	0.98	0.97	0.98	0.94	4.6	0.84	0.65	0.94	−0.75	1.3	0.48	−0.00	0.80	−1.60	1.0
200	100	0.97	0.94	0.97	0.93	4.9	0.79	0.55	0.88	−1.68	1.2	0.43	−0.50	0.68	−3.19	1.0
200	200	0.98	0.97	0.98	0.97	5.0	0.88	0.80	0.93	−0.52	1.7	0.57	0.11	0.81	−2.02	1.0
200	400	0.99	0.99	0.99	0.99	5.0	0.94	0.90	0.95	0.40	2.7	0.69	0.45	0.88	−1.99	1.0
$a=0.5,\ \beta=0.5,\ \rho=0.9$																
50	50	0.94	0.74	0.95	0.65	3.7	0.71	0.24	0.88	−1.05	1.0	0.41	−0.55	0.66	−1.91	1.0
50	100	0.97	0.86	0.97	0.83	4.5	0.80	0.51	0.93	−0.52	1.2	0.49	−0.09	0.79	−1.40	1.0
100	100	0.96	0.90	0.97	0.88	4.6	0.82	0.56	0.92	−1.05	1.3	0.50	−0.07	0.77	−2.04	1.0
100	200	0.98	0.96	0.98	0.95	4.9	0.90	0.76	0.95	−0.07	1.9	0.61	0.25	0.87	−1.51	1.0
200	100	0.96	0.93	0.96	0.92	5.0	0.84	0.66	0.90	−0.66	1.8	0.55	−0.07	0.76	−2.91	1.0
200	200	0.98	0.97	0.98	0.96	5.0	0.91	0.85	0.94	0.19	2.5	0.70	0.39	0.86	−1.93	1.0
200	400	0.99	0.98	0.99	0.98	5.0	0.95	0.93	0.96	0.71	3.6	0.80	0.64	0.92	−1.79	1.0

Table 1: Simulation results for DGPs with white noise disturbances in the factor loadings.

Monte Carlo simulations: Random walk model

		$c=0$					$c=2$					$c=3.5$				
		$k=r$		IC			$k=r$		IC			$k=r$		IC		
T	N	$R^2_{\hat{F},F}$	$S^2_{\hat{y},\tilde{y}}$	$R^2_{\hat{F},F}$	$S^2_{\hat{y},\tilde{y}}$	$\hat{E}(\hat{k})$	$R^2_{\hat{F},F}$	$S^2_{\hat{y},\tilde{y}}$	$R^2_{\hat{F},F}$	$S^2_{\hat{y},\tilde{y}}$	$\hat{E}(\hat{k})$	$R^2_{\hat{F},F}$	$S^2_{\hat{y},\tilde{y}}$	$R^2_{\hat{F},F}$	$S^2_{\hat{y},\tilde{y}}$	$\hat{E}(\hat{k})$
$a=0,\ \beta=0,\ \rho=0$																
50	50	0.93	0.83	0.94	0.50	3.4	0.91	0.90	0.91	0.89	4.8	0.87	0.89	0.87	0.90	5.2
50	100	0.96	0.93	0.96	0.88	4.7	0.93	0.93	0.93	0.93	5.0	0.89	0.92	0.88	0.91	5.7
100	100	0.96	0.94	0.96	0.93	5.0	0.94	0.94	0.94	0.95	5.0	0.89	0.94	0.89	0.94	5.4
100	200	0.98	0.97	0.98	0.97	5.0	0.95	0.96	0.95	0.96	5.0	0.90	0.94	0.89	0.94	6.8
200	100	0.96	0.94	0.96	0.94	5.0	0.94	0.95	0.94	0.95	5.0	0.90	0.94	0.90	0.94	5.6
200	200	0.98	0.97	0.98	0.97	5.0	0.96	0.97	0.96	0.97	5.0	0.91	0.95	0.91	0.95	7.0
200	400	0.99	0.99	0.99	0.99	5.0	0.97	0.98	0.97	0.97	5.0	0.92	0.96	0.91	0.96	9.5
$a=0.5,\ \beta=0,\ \rho=0$																
50	50	0.91	0.77	0.93	0.53	3.7	0.91	0.90	0.91	0.87	4.7	0.88	0.89	0.88	0.90	5.1
50	100	0.95	0.90	0.95	0.88	4.8	0.94	0.94	0.94	0.94	5.0	0.90	0.92	0.89	0.92	5.5
100	100	0.96	0.93	0.96	0.92	5.0	0.94	0.95	0.94	0.95	5.0	0.90	0.94	0.90	0.94	5.2
100	200	0.98	0.97	0.98	0.97	5.0	0.96	0.96	0.96	0.96	5.0	0.91	0.95	0.91	0.94	6.1
200	100	0.96	0.94	0.96	0.94	5.0	0.95	0.95	0.95	0.95	5.0	0.92	0.95	0.91	0.95	5.2
200	200	0.98	0.97	0.98	0.97	5.0	0.97	0.97	0.97	0.97	5.0	0.93	0.96	0.92	0.96	5.8
200	400	0.99	0.99	0.99	0.99	5.0	0.98	0.98	0.98	0.98	5.0	0.93	0.96	0.93	0.96	8.4
$a=0,\ \beta=0.5,\ \rho=0$																
50	50	0.91	0.76	0.93	0.53	3.7	0.91	0.89	0.91	0.89	4.8	0.87	0.89	0.87	0.90	5.3
50	100	0.95	0.91	0.96	0.87	4.7	0.93	0.93	0.93	0.93	5.0	0.89	0.92	0.88	0.91	5.8
100	100	0.96	0.92	0.96	0.92	5.0	0.94	0.94	0.94	0.95	5.0	0.89	0.94	0.89	0.94	5.5
100	200	0.98	0.97	0.98	0.97	5.0	0.95	0.96	0.95	0.96	5.0	0.90	0.94	0.89	0.94	6.9
200	100	0.96	0.94	0.95	0.94	5.0	0.94	0.95	0.94	0.95	5.0	0.90	0.95	0.90	0.94	5.8
200	200	0.98	0.97	0.98	0.97	5.0	0.96	0.97	0.96	0.97	5.0	0.91	0.95	0.91	0.95	7.1
200	400	0.99	0.99	0.99	0.99	5.0	0.97	0.98	0.97	0.98	5.0	0.92	0.96	0.91	0.96	9.5
$a=0,\ \beta=0,\ \rho=0.9$																
50	50	0.95	0.81	0.97	0.43	2.3	0.94	0.90	0.94	0.91	5.5	0.94	0.90	0.94	0.92	7.3
50	100	0.97	0.91	0.98	0.69	2.9	0.95	0.92	0.95	0.94	6.4	0.94	0.92	0.94	0.94	8.6
100	100	0.97	0.94	0.97	0.81	3.9	0.93	0.93	0.92	0.95	6.7	0.91	0.93	0.91	0.95	9.2
100	200	0.98	0.97	0.98	0.94	4.6	0.93	0.94	0.93	0.97	7.8	0.92	0.94	0.91	0.96	9.9
200	100	0.97	0.94	0.97	0.93	4.9	0.91	0.94	0.90	0.95	7.4	0.88	0.94	0.88	0.96	9.9
200	200	0.98	0.97	0.98	0.97	5.0	0.92	0.95	0.91	0.97	8.3	0.89	0.95	0.88	0.97	10.0
200	400	0.99	0.99	0.99	0.99	5.0	0.92	0.95	0.91	0.98	9.6	0.89	0.95	0.88	0.97	10.0
$a=0.5,\ \beta=0.5,\ \rho=0.9$																
50	50	0.94	0.74	0.95	0.65	3.7	0.94	0.90	0.94	0.91	6.8	0.94	0.90	0.94	0.92	8.3
50	100	0.97	0.86	0.97	0.83	4.5	0.95	0.92	0.95	0.95	8.4	0.94	0.92	0.94	0.94	9.7
100	100	0.96	0.90	0.97	0.88	4.6	0.93	0.93	0.92	0.95	7.7	0.92	0.93	0.91	0.95	9.7
100	200	0.98	0.96	0.98	0.95	4.9	0.94	0.95	0.93	0.97	9.2	0.92	0.94	0.91	0.97	10.0
200	100	0.96	0.93	0.96	0.92	5.0	0.92	0.94	0.91	0.95	8.0	0.89	0.94	0.88	0.95	10.0
200	200	0.98	0.97	0.98	0.96	5.0	0.93	0.95	0.92	0.97	8.9	0.89	0.95	0.89	0.97	10.0
200	400	0.99	0.98	0.99	0.98	5.0	0.93	0.96	0.92	0.98	9.9	0.90	0.95	0.89	0.98	10.0

Table 2: Simulation results for DGPs with random walk factor loadings.

Monte Carlo simulations: Large break model

		$b=0$					$b=3.5$					$b=7$				
		$k=r$		IC			$k=r$		IC			$k=r$		IC		
T	N	$R^2_{\hat{F},F}$	$S^2_{\hat{y},\tilde{y}}$	$R^2_{\hat{F},F}$	$S^2_{\hat{y},\tilde{y}}$	$\hat{E}(\hat{k})$	$R^2_{\hat{F},F}$	$S^2_{\hat{y},\tilde{y}}$	$R^2_{\hat{F},F}$	$S^2_{\hat{y},\tilde{y}}$	$\hat{E}(\hat{k})$	$R^2_{\hat{F},F}$	$S^2_{\hat{y},\tilde{y}}$	$R^2_{\hat{F},F}$	$S^2_{\hat{y},\tilde{y}}$	$\hat{E}(\hat{k})$
$a=0,\ \beta=0,\ \rho=0$																
50	50	0.93	0.83	0.94	0.50	3.4	0.90	0.83	0.90	0.59	3.6	0.83	0.82	0.84	0.58	3.7
50	100	0.96	0.93	0.96	0.88	4.7	0.94	0.92	0.94	0.89	4.7	0.89	0.91	0.89	0.89	4.8
100	100	0.96	0.94	0.96	0.93	5.0	0.94	0.94	0.94	0.93	5.0	0.88	0.93	0.88	0.93	5.1
100	200	0.98	0.97	0.98	0.97	5.0	0.97	0.97	0.96	0.97	5.1	0.93	0.96	0.93	0.97	5.3
200	100	0.96	0.94	0.96	0.94	5.0	0.93	0.94	0.93	0.94	5.2	0.87	0.94	0.87	0.94	5.5
200	200	0.98	0.97	0.98	0.97	5.0	0.96	0.97	0.96	0.97	5.2	0.93	0.97	0.92	0.97	5.6
200	400	0.99	0.99	0.99	0.99	5.0	0.98	0.98	0.98	0.99	5.2	0.96	0.98	0.95	0.99	5.6
$a=0.5,\ \beta=0,\ \rho=0$																
50	50	0.91	0.77	0.93	0.53	3.7	0.88	0.79	0.89	0.61	3.8	0.82	0.79	0.83	0.62	3.9
50	100	0.95	0.90	0.95	0.88	4.8	0.93	0.90	0.93	0.89	4.9	0.88	0.89	0.88	0.88	5.0
100	100	0.96	0.93	0.96	0.92	5.0	0.93	0.93	0.93	0.93	5.0	0.88	0.93	0.88	0.93	5.2
100	200	0.98	0.97	0.98	0.97	5.0	0.96	0.97	0.96	0.97	5.1	0.93	0.96	0.92	0.97	5.4
200	100	0.96	0.94	0.96	0.94	5.0	0.93	0.94	0.93	0.94	5.2	0.87	0.94	0.87	0.94	5.5
200	200	0.98	0.97	0.98	0.97	5.0	0.96	0.97	0.96	0.97	5.2	0.93	0.97	0.92	0.97	5.6
200	400	0.99	0.99	0.99	0.99	5.0	0.98	0.98	0.98	0.98	5.2	0.96	0.98	0.95	0.98	5.6
$a=0,\ \beta=0.5,\ \rho=0$																
50	50	0.91	0.76	0.93	0.53	3.7	0.87	0.77	0.88	0.61	3.9	0.80	0.77	0.81	0.61	4.1
50	100	0.95	0.91	0.96	0.87	4.7	0.93	0.90	0.93	0.89	4.8	0.88	0.90	0.88	0.88	5.0
100	100	0.96	0.92	0.96	0.92	5.0	0.93	0.93	0.93	0.93	5.1	0.87	0.92	0.87	0.93	5.3
100	200	0.98	0.97	0.98	0.97	5.0	0.96	0.96	0.96	0.96	5.1	0.93	0.96	0.92	0.97	5.4
200	100	0.96	0.94	0.95	0.94	5.0	0.93	0.94	0.92	0.94	5.3	0.86	0.93	0.86	0.94	5.7
200	200	0.98	0.97	0.98	0.97	5.0	0.96	0.97	0.96	0.97	5.3	0.92	0.97	0.91	0.97	5.7
200	400	0.99	0.99	0.99	0.99	5.0	0.98	0.98	0.98	0.98	5.3	0.96	0.98	0.95	0.99	5.7
$a=0,\ \beta=0,\ \rho=0.9$																
50	50	0.95	0.81	0.97	0.43	2.3	0.94	0.81	0.95	0.52	2.4	0.92	0.81	0.93	0.56	2.6
50	100	0.97	0.91	0.98	0.69	2.9	0.97	0.90	0.97	0.72	3.0	0.95	0.90	0.95	0.73	3.1
100	100	0.97	0.94	0.97	0.81	3.9	0.96	0.93	0.96	0.83	4.0	0.93	0.91	0.93	0.83	4.2
100	200	0.98	0.97	0.98	0.94	4.6	0.98	0.97	0.98	0.94	4.7	0.96	0.96	0.96	0.94	4.8
200	100	0.97	0.94	0.97	0.93	4.9	0.95	0.94	0.95	0.93	5.0	0.91	0.93	0.91	0.93	5.1
200	200	0.98	0.97	0.98	0.97	5.0	0.97	0.97	0.97	0.97	5.1	0.95	0.96	0.94	0.97	5.3
200	400	0.99	0.99	0.99	0.99	5.0	0.99	0.98	0.98	0.98	5.1	0.97	0.98	0.97	0.98	5.4
$a=0.5,\ \beta=0.5,\ \rho=0.9$																
50	50	0.94	0.74	0.95	0.65	3.7	0.93	0.75	0.94	0.70	4.1	0.90	0.75	0.91	0.71	4.3
50	100	0.97	0.86	0.97	0.83	4.5	0.96	0.85	0.96	0.84	4.8	0.94	0.86	0.94	0.85	5.0
100	100	0.96	0.90	0.97	0.88	4.6	0.95	0.89	0.95	0.88	4.8	0.92	0.88	0.92	0.88	5.0
100	200	0.98	0.96	0.98	0.95	4.9	0.97	0.95	0.97	0.95	5.2	0.95	0.94	0.95	0.95	5.4
200	100	0.96	0.93	0.96	0.92	5.0	0.94	0.92	0.94	0.93	5.3	0.90	0.91	0.90	0.93	5.6
200	200	0.98	0.97	0.98	0.96	5.0	0.97	0.96	0.97	0.97	5.3	0.94	0.95	0.94	0.97	5.6
200	400	0.99	0.98	0.99	0.98	5.0	0.98	0.98	0.98	0.98	5.3	0.97	0.98	0.96	0.98	5.7

Table 3: Simulation results for DGPs with a single large break in the factor loadings.