WORKING PAPER N° 2017 – 42

On the consistency of the two-step estimates of the MS-DFM: a Monte Carlo study

Catherine Doz
Anna Petronevich

JEL Codes:
Keywords: Markov-switching, Dynamic Factor models, two-step estimation, small-sample performance, consistency, Monte Carlo simulations

On the consistency of the two-step estimates of the MS-DFM: a Monte Carlo study[☆]

Catherine Doz[a], Anna Petronevich[a,b]

[a]*Paris School of Economics, Université Paris 1 Panthéon-Sorbonne*

[b]*CREST, Università Ca'Foscari Venezia, NRU HSE*

Abstract

The Markov-Switching Dynamic Factor Model (MS-DFM) has been used in different applications, notably in the business cycle analysis. When the cross-sectional dimension of data is high, the Maximum Likelihood estimation becomes unfeasible due to the excessive number of parameters. In this case, the MS-DFM can be estimated in two steps, which means that in the first step the common factor is extracted from a database of indicators, and in the second step the Markov-Switching autoregressive model is fit to this extracted factor. The validity of the two-step method is conventionally accepted, although the asymptotic properties of the two-step estimates have not been studied yet. In this paper we examine their consistency as well as the small-sample behavior with the help of Monte Carlo simulations. Our results indicate that the two-step estimates are consistent when the number of cross-section series and time observations is large, however, as expected, the estimates and their standard errors tend to be biased in small samples.

Keywords: Markov-switching, Dynamic Factor models, two-step estimation, small-sample performance, consistency, Monte Carlo simulations

1. Introduction

Markov-Switching Dynamic Factor model (MS-DFM) has proved to be a useful instrument in a number of applications. Among them are tracking of labor productivity (Dolega (2007)), modeling the joint dynamics of the yield curve and the GDP (Chauvet and Senyuz (2012)), examination of fluctuations in the employment rates (Juhn et al. (2002)) and many others. However, the major application of the MS-DFM is the analysis of the business cycle turning points (see, for example, Kim and Yoo (1995), Darné and Ferrara (2011), Camacho et al. (2012), Chauvet and Yu (2006), Wang et al. (2009)). The initially suggested univariate Markov-switching model in the seminal paper by Hamilton (1989) was extended to the multivariate case, the MS-DFM, by Kim (1994). The model allows to obtain the turning points in a transparent and replicable way, and, importantly,

[☆]The authors are thankful to Monica Billio, Jean-Michel Zakoian, Christian Francq and the participants of the internal Macro Seminar of the University Ca'Foscari Venezia for their fruitful suggestions. We also acknowledge financial support by the European Commission in the framework of the European Doctorate in Economics Erasmus Mundus (EDEEM).

in a more timely manner than the official institutions (OECD and NBER, for example).

The MS-DFM formalizes the idea of Diebold and Rudebusch (1996) that the economic variables comove and follow a pattern with alternating periods of growth and decline, this comovement essentially representing the business cycle. More precisely, the model assumes that the economic indicators have a factor structure, i.e. are driven by some common factor, which itself follows a Markov-switching dynamics with two regimes.

Depending on the number of economic series under consideration, the model can be estimated using different techniques. The original paper by Kim (1994) as well as some of the following applications is based on just a few economic indicators so the parameters and the factor can be estimated simultaneously with Kalman filter and Maximum Likelihood. However, when the number of series increases, convergence problems may arise and, besides, the estimation may become time-consuming since the number of parameters expands proportionally to the number of series. In the same time, the use of many economic indicators is desirable in order to consider as much information on the business cycle as possible. A natural solution to this trade-off between the size of the information set and the computational time is the the estimation of the Markov-Switching and Dynamic Factor parts of the MS-DFM separately, i.e. in two steps. Attractive in terms of applicability and information treatment, the two-step estimation method has been used in several studies (see, for example, Chauvet and Senyuz (2012), Darné and Ferrara (2011), Bessec and Bouabdallah (2015), Brave and Butters (2010), Davig (2008), Paap et al. (2009)). This method implies that the factor is extracted[1] on the first step, and then the classical univariate Markov-switching model à la Hamilton (1989) is fit to the estimated factor on the second step. The two-step procedure is much easier to implement[2] and does not impose any restrictions on the number of series by default.

Previous studies show the importance of the number of cross-sections N and the number of observations T for the accuracy of estimates on each of the steps. Connor and Korajczyk (1986) proved the consistency of the principal components estimator $\hat{f}_t$ (commonly used in the first step) for fixed T and $N \to \infty$ under general assumptions used by Chamberlain (1983) and Chamberlain and Rothschild (1983) for the definition of the approximate factor structure. Stock and Watson (2002) examined conditions for the rates of N and T under which $\hat{f}_t$ can be treated as data for the OLS regression. More precisely, they show that the performance of the PCA is very good even when the sample-size and the number of series are relatively small, $N = 100$ and $T = 100$. Further on, Bai and Ng (2006) and Bai and Ng (2013) extended this result showing that, under a standard set of assumptions usually used in factor analysis, the $\hat{f}_t$ can be treated as data in subsequent regressions when $N \to \infty$, $T \to \infty$ and $N^2/T \to \infty$. As for the second step, Kiefer (1978), Francq and Roussignol (1997), Francq and Roussignol (1998), Krishnamurthy and Ryden (1998), Douc et al. (2004) and Douc et al. (2011) show that under particular conditions the maximum likelihood estimators of an autoregressive model with Markov regimes are consistent. Francq and Roussignol (1997) also propose a gaussian maximum pseudo-likelihood estimator, Krishnamurthy and Ryden

[1]Different methods of factor extraction can be applied, Kalman filter (for a small number of series) and PCA are the most common ones.

[2]The procedures for the estimation of the Markov-Switching models are installed in some econometric software such as Eviews, Stata. The corresponding packages exist for Matlab and R.

(1998) derived the conditions under which the MLE are asymptotically Gaussian.

Even though the consistency of the estimates of the factor on the first step and the consistency of the ML estimates of the Markov-switching model on the second step has been already shown, it is not straightforward that the ML estimates of the Markov-switching model which is fit to the estimated (and not observed directly) factors are consistent. To the best of our knowledge, the consistency of the two-step estimates has not been shown yet. It is not very surprising since the asymptotic properties of Markov-switching autoregressive models are rather complicated and are difficult to derive analytically.

Another concern of the two-step approach are the small-sample properties of the estimates. Indeed, Hosmer (1973), Hamilton (1991) and Hamilton (1996) find that asymptotic approximations of the sampling distribution of the MLE may be inadequate in small-sample cases. In their Monte Carlo study, Psaradakis and Sola (1998) have shown that "the performance of the MLE was often unsatisfactory even for sample sizes as large as 400", pointing out the non-normality of the empirical distribution of the estimates and the bias that that takes place.

The purpose of this paper is thus twofold. First, we study the consistency of the two-step estimates, where the factor is estimated with PCA. Secondly, we would like to examine the behavior of the estimates in small samples and identify the minimum N and T required to obtain estimates with a reasonably small bias. In addition, we check whether the distribution of the two-step estimates approaches normal (as is the case for the Maximum Likelihood estimates of an MS-AR) given the amount of data usually available in macroeconomic applications. In this paper, we study the aforementioned questions with the help of Monte Carlo simulations. The analytic proof of consistency is being prepared in a companion paper to this work.

This paper thus contributes to the literature on the analysis of the two-step estimates of the MS-DFM. Previously, Camacho et al. (2015) studied the performance of the two-step estimates where the factor is extracted with a linear DFM on the first step. Their study focused on the quality of identification of states and was based on the use of a few series (N not higher than 8). Having compared the two-step results to regular (one-step) Maximum Likelihood estimates, the authors showed that the two-step results diverge from the one-step ones when the common factor is extracted with the help of a linear DFM while the data-generating process is a nonlinear MS-DFM, although the difference decreases when N rises or when the data are less noisy.

The paper is organized as follows. Section 2 describes the MS-DFM model. Section 3 presents the two-step estimation technique. In section 4 we describe the design of the Monte Carlo experiment and discuss the simulation results. Section 5 concludes.

2. Markov-Switching Dynamic Factor Model

In the present paper, we take the basic specification of the MS-DFM for the business cycle as in the seminal paper by Kim and Yoo (1995), and we assume that the growth rate cycle of the economic activity has only two regimes (or states), associated with its low and high levels. The economic activity itself is represented by an unobservable factor, which summarizes the common dynamics of several observable variables. It is assumed that the switch between regimes happens instantaneously, without any transition period (as is considered, for example, by STAR family models). This assumption can be motivated by the fact that the transition period before deep crises is normally short enough to be omitted. For example, the growth rate of French GDP fell from 0.5% in the first quarter of 2008 to -0.51% in the second quarter of the same year, and further down to -1.59% in the first quarter of 2009[3].

The model is thus decomposed into two equations, the first one defining the factor model, and the second one describing the Markov switching autoregressive model which is assumed for the common factor. More precisely, in the first equation, each series of the information set is decomposed into the sum of a common component (the common factor loads each of the observable series with a specific weight) and an idiosyncratic component:

$$y_t = \lambda f_t + z_t, \tag{1}$$

where $t = 1, ..., T$, y_t is a $N \times 1$ vector of economic indicators, f_t is a univariate common factor, z_t is a $N \times 1$ vector of idiosyncratic components uncorrelated with f_t at all leads and lags, λ is a $N \times 1$ vector. In this equation all series are supposed to be stationary, so that some of the components of y_t may be the first differences of the initially non stationary economic indicator.

The idiosyncratic components z_{it}'s, $i = 1, ..., N$, are mutually uncorrelated at all leads and lags, and each of them follows an autoregressive process

$$\psi_i(L) z_{it} = \varepsilon_{it}, \tag{2}$$

where $\psi_i(L)$ is a lag polynomial such that $\psi_i(0) = 1$, $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_i^2)$ and $cov(\varepsilon_{it}, \varepsilon_{jt}) = 0$ for all $i \neq j$.

The second equation describes the behavior of the factor f_t, which is supposed to follow an autoregressive Markov Switching process with constant transition probabilities[4]. In what follows, we consider that the change in regime affects only the level of the constant with the high level corresponding to the expansion state and the low level to the recession state:

$$\varphi(L) f_t = \beta_{S_t} + \eta_t, \tag{3}$$

where $\eta_t \sim i.i.d.\ \mathcal{N}(0, 1)$, $\varphi(L)$ is an autoregressive polynomial such that $\varphi(0) = 1$.

The switching mean is defined as:

[3] INSEE, France, Gross Domestic Product, Total, Contribution to Growth, Calendar Adjusted, Constant Prices, SA, Chained, Change P/P

[4] Kim and Yoo (1995) showed that, in the business cycle applications, although the assumption of the time dependent probabilities improves the quality of the model, the gain in terms of loglikelihood is not very large.

$$\beta_{S_t} = \beta_0(1 - S_t) + \beta_1 S_t, \tag{4}$$

where S_t takes a value 0 when the economy is in expansion and 1 otherwise, so $\beta_0 > \beta_1$. S_t follows an ergodic Markov chain, i.e.

$$P(S_t = j|S_{t-1} = i, S_{t-2} = k, ...) = P(S_t = j|S_{t-1} = i) = p_{ij}. \tag{5}$$

As it is assumed that there are two states only, S_t switches states according to a 2×2 transition probabilities matrix defined as $\begin{bmatrix} p_0 & 1-p_0 \\ 1-p_1 & p_1 \end{bmatrix}$, where

$$\begin{aligned} P(S_t = 0|S_{t-1} = 0) &= p_0, \\ P(S_t = 1|S_{t-1} = 1) &= p_1. \end{aligned} \tag{6}$$

There is no restriction on the duration of each state, and the states are defined point-wise, i.e. a recession period may last one period only.

The present framework can be generalized to the case of a higher number of states and/or to regime dependence in the other parameters of the model (the variance of the error term, the coefficients of the autoregressive polynomial). In our study, we consider the simplest case with two regimes and a switch in constant as in this specification it is easier to control the data generating process. It is also often selected by information criteria in the empirical applications.

3. Two-step estimation method

The model presented above can be cast into the state-space form and estimated with Maximum Likelihood. However, the estimation is complicated as the likelihood function has to take into account all possible paths of S_t, which is 2^T, and the number of parameters to estimate grows proportionally to the number of series in y_t. While the first issue can be solved with collapsing procedure suggested by Kim (1994), the second problem makes the solution unfeasible for large N. This is computationally challenging, and there are several ways to solve this issue. The one that we consider in this paper is estimating the model in two steps, where on the first step the factor is extracted from the data y_t, while on the second step the estimated factor is used as if it were the true factor in order to obtain the estimates of equation (3). More precisely, the procedure is the following.

Step 1

The factor f_t is extracted from a large database of economic indicators according to equation (1) without taking its Markov-Switching dynamics into account. In the present paper, we use principal component analysis to compute an approximation $\hat{f}_t$ of the true factor. Indeed, since (f_t) is a stationary process, as we have discussed in the introduction, under a mild set of assumptions, it is consistently estimated by $\hat{f}_t$. The factor can be extracted with a different method, for example, using the two-step estimator suggested by Doz et al. (2011) or Quasi-Maximum Likelihood estimator by Doz et al. (2012).

If we denote by $\hat{\Sigma} = \frac{1}{T}\tilde{y}'\tilde{y}$ the empirical correlation matrix of y (where $\tilde{y}$ is the standardized y), by $\hat{D}$ the $N \times N$ diagonal matrix of with the eigenvalues of $\hat{\Sigma}$ in decreasing order, by $\hat{V}$ the $N \times N$ matrix of the unitary eigenvectors corresponding to $\hat{D}$, then the matrix of the principal components $\hat{F}$ is defined as:

$$\hat{F} = y\hat{V},$$

and the corresponding matrix of loadings $\hat{\Lambda}$ is:

$$\hat{\Lambda} = \hat{V}',$$

The first column of the matrix $\hat{F}$ is then the estimate $\hat{f}_t$ of the true factor f_t, whereas the first column of $\hat{\Lambda}$ is the estimate $\hat{\lambda}$.

Step 2

The parameters of the autoregressive Markov-Switching model described by equations (3) and (5) are estimated by maximum likelihood, with f_t replaced by $\hat{f}_t$. This amounts to fit the univariate model of Hamilton (1989) to the estimated factor $\hat{f}_t$, which is taken as if it were an observed variable:

$$\varphi(L)\hat{f}_t = \beta_{S_t} + u_t, \tag{7}$$

where $u_t \sim N(0, \sigma^2)$. Suppose that $\theta = (\varphi(L), \psi_1(L), ..., \psi_N(L), \lambda_1, ..., \lambda_N, \sigma_1^2, ..., \sigma_N^2, \beta_0, \beta_1, p_0, p_1, \sigma^2)'$ is the vector of unknown parameters, $g(\cdot)$ is the Gaussian density function. The log-likelihood function takes the following form:

$$\mathcal{L}_T(\hat{f}, \theta) = ln\, l(\hat{f}_1, \hat{f}_2, ..., \hat{f}_T, \theta) = \sum_{t=1}^{T} ln\, g(\hat{f}_t|I_{t-1}, \theta), \tag{8}$$

where we denote $\hat{f} = (\hat{f}_1, ..., \hat{f}_T)$ and $I_{t-1} = \{\hat{f}_1, ..., \hat{f}_{t-1}\}$. The density

$$g(\hat{f}_t|I_{t-1}, \theta) = \sum_{j=0}^{1}\sum_{i=0}^{1} g(\hat{f}_t, S_t = j, S_{t-1} = i|I_{t-1}, \theta), \tag{9}$$

is computed using filtered probability $P(S_t = j|I_t, \theta)$ on the basis of Bayes' theorem:

$$g(\hat{f}_t, S_t = j, S_{t-1} = i|I_{t-1}, \theta) = g(\hat{f}_t|S_t = j, S_{t-1} = i, \theta) \times P(S_t = j, S_{t-1} = i|I_{t-1}, \theta), \tag{10}$$

where

$$g(\hat{f}_t|S_t = j, S_{t-1} = i, I_{t-1}, \theta) = (2\pi\sigma^2)^{-1/2} exp\left\{-\frac{1}{2}\frac{(\varphi(L)\hat{f}_t - \beta_{S_t})^2}{\sigma^2}\right\} \tag{11}$$

$$P(S = j, S_{t-1} = i|I_{t-1}, \theta) = P(S_t = j|S_{t-1} = i, \theta) \times P(S_{t-1} = i|I_{t-1}, \theta), \tag{12}$$

and

$$P(S_t = j|I_t, \theta) = \sum_{i=0}^{1} P(S = j, S_{t-1} = i|I_t, \theta) \tag{13}$$

$$\begin{aligned} P(S_t = j, S_{t-1} = i|I_t, \theta) &= \frac{g(\hat{f}_t, S_t = j, S_{t-1} = i|I_{t-1}, \theta)}{g(\hat{f}_t|I_{t-1}, \theta)} \\ &= \frac{g(\hat{f}_t|S_t = j, S_{t-1} = i, I_{t-1}, \theta) \times P(S_t = j, S_{t-1} = i|I_{t-1}, \theta)}{g(\hat{f}_t|I_{t-1}, \theta)} \end{aligned} \tag{14}$$

The recursion is initialized with the steady state probability π of being in state $j \in \{0; 1\}$ at time $t = 0$:

$$\pi = P(S_0 = 1|I_0, \theta) = \frac{1 - p_0}{2 - p_0 - p_1}, \tag{15}$$

$$P(S_0 = 0|I_0, \theta) = 1 - P(S_1 = 1|I_0, \theta) = 1 - \pi. \tag{16}$$

The two-step estimates $\hat{\theta}(\hat{f})$ are obtained as the maximum of the likelihood function $\mathcal{L}_T(\hat{f}, \theta)$ using numerical optimization algorithms. Then, for $\hat{\theta}(\hat{f})$ given, we can infer the associated filtered probability $P(S_t = j|I_t, \hat{\theta})$[5] with formulas (9)-(14). Also, it is possible to compute the smoothed probabilities of each state $P(S_t = 1|I_T, \hat{\theta})$ using backward filtering (see Hamilton (1989)).

In the majority of studies on business cycle fluctuations analyzed with MS-DFM, the filtered probability of recession is the main focus. Since it allows to have an estimate of the state at time t on the basis of the information available up to the moment t, it is used for the purposes of nowcasting. The smoothed probabilities are often used to establish business cycle dating retrospectively on the basis of the full information set, i.e. a posteriori. It is therefore important to verify if the two-step estimates provide good quality of identification of states in terms of both filtered and smoothed probability.

[5] To simplify notations, we denote $\hat{\theta} = \hat{\theta}(\hat{f})$.

4. Monte Carlo simulations

We use Monte Carlo simulations to examine the consistency of the two-step estimates as well as their small-sample properties. We first discuss the experimental design. The numerical results follow.

4.1. Experimental setup

4.1.1. The DGP

The data generating process (DGP) used in simulations is described in equations (1)-(5). We assume for simplicity that the order of $\varphi(L)$ is one and the order of $\psi_i(L)$ is zero, $i = 1, ..., N$. The autoregressive polynomials of higher order would complicate the control over the variance of the factor and the idiosyncratic components without changing the essence of the dynamics of the underlying processes (unless it renders the dynamics nonstationary). The DGP is therefore:

$$y_{it} = \lambda_i f_t + \varepsilon_{it}, \tag{17}$$

$$(1 - \varphi L) f_t = \beta_{S_t} + \eta_t, \tag{18}$$

$$\beta_{S_t} = \beta_1 S_t + \beta_0 (1 - S_t) = \beta_0 + (\beta_1 - \beta_0) S_t, \tag{19}$$

$$\begin{aligned} P(S_t = 0 | S_{t-1} = 0) &= p_0, \\ P(S_t = 1 | S_{t-1} = 1) &= p_1, \end{aligned} \tag{20}$$

where $i = 1, ..., N$, $t = 1, ..., T$, $\eta_t \sim N(0, 1)$. The factor loadings λ_i are generated from a normal distribution and are normalized to have a unit sum of squares, i.e. $\lambda_i = \frac{\gamma_i}{\sqrt{\gamma'\gamma}}$, $\gamma_i \sim N(0, 1)$.[6] The idiosyncratic disturbance terms ε_{it} are cross-sectionally independent and have a Gaussian distribution $\varepsilon_{it} \sim N(0, \sigma_i^2)$. The state variable S_t is a Markov switching process with two states $S_t \in \{0; 1\}$ (0 corresponds to expansion, and 1 to recession) and transition probability matrix $\begin{pmatrix} p_0 & 1 - p_0 \\ 1 - p_1 & p_1 \end{pmatrix}$.

We put the unconditional mean of the factor to zero, a classical assumption for the factor models. Since $ES_t = \pi$, this imposes a fixed relation between β_0 and β_1:

$$\beta_1 = \beta_0 (1 - \frac{1}{\pi}). \tag{21}$$

4.1.2. The parameters of control

Intuitively, besides the size of N and T, there are four aspects of the dynamics of the DGP that might affect the quality of the estimates of the MLE. These are:

1. The persistence of each regime. For a given sample size, a more persistent regime is better identified since it is activated during longer periods of time.

[6] The vector λ is unitary in order to provide the same scale to the generated factor and the estimated factor obtained by PCA with normalized loadings.

2. The noise-to-signal ratios $s_i = \frac{\sigma_i^2}{V(y_{it})}$. When the data are less noisy, the estimate of the factor is more precise.
3. The persistence of the autoregressive dynamics of the factor. Presumably, the closer the root of the autoregressive polynomial is to one in absolute value, the more difficult it is to distinguish between the change in regime and the long-lasting effect of a shock in the error term.
4. The share of variance of the factor due to the switch. If most of the variance of the factor is generated by the error term η_t, the states are more difficult to identify.

   Under the assumption that the unconditional mean of the factor is zero, it is possible to show that (see Appendix for the details) that the unconditional variance of the true factor f_t is:

$$V(f_t) = \frac{1}{1-\varphi^2}\left(\sigma^2 + \beta_0^2\left(\frac{1-p_1}{1-p_0}\right)\left(\frac{1+\varphi(p_0+p_1-1)}{1-\varphi(p_0+p_1-1)}\right)\right). \tag{22}$$

We control the first item directly by changing p_0, p_1. We suppose that the s_i's are uniformly distributed: $s_i \sim \mathcal{U}[u; 1-u]$, and we control the noisiness by choosing u. We control the persistence of the factor by changing the autoregressive term φ. The share of variance of the factor due to the error term is varied by changing the ratio $c = \frac{V(f_t)}{\sigma^2}$. Therefore, the free parameters of the simulation are c, p_0, p_1, u, φ. The parameters to be estimated are $\theta = (\beta_0, \beta_1, \varphi, \sigma^2, p_0, p_1)$.

In order to examine behavior of the two-step method under different conditions, we run Monte Carlo simulations for the following scenarios:

1. baseline scenario: $c = 5$, $p_0 = 0.9$, $p_1 = 0.8$, $u = 0.1$, $\varphi = 0.3$;
2. noisy factor: $c = 2$, $p_0 = 0.9$, $p_1 = 0.8$, $u = 0.1$, $\varphi = 0.3$;
3. persistence of the factor dynamics:
   3.1. high autocorrelation: $c = 5$, $p_0 = 0.9$, $p_1 = 0.8$, $u = 0.1$, $\varphi = 0.9$;
   3.2. medium autocorrelation: $c = 5$, $p_0 = 0.9$, $p_1 = 0.8$, $u = 0.1$, $\varphi = 0.6$;
4. persistence of states:
   4.1. impersistent states: $c = 5$, $p_0 = 0.5$, $p_1 = 0.5$, $u = 0.1$, $\varphi = 0.3$;
   4.2. very persistent states: $c = 5$, $p_0 = 0.95$, $p_1 = 0.95$, $u = 0.1$, $\varphi = 0.3$;
5. homogeneous data: $c = 5$, $p_0 = 0.9$, $p_1 = 0.8$, $u = 0.5$, $\varphi = 0.3$;

The values of the parameters in the baseline scenario have been taken from existing empirical studies using MS-DFM. Thus, φ close to 0.3 has been obtained by Chauvet (1998), Kim and Yoo (1995), Kim and Nelson (1998), whereas the ratio of factor variance and variance of the error term varies between 2 in Kim and Yoo (1995) and 7.5 in Kim and Nelson (1998). The transition probabilities p_0 and p_1 are often estimated to be around 0.9 and 0.8, respectively, i.e. the two states are very persistent and the probability of staying in expansion is always higher than the probability to stay in recession. We generate $K = 2000$ replications of each scenario and estimate the MS-DFM with the help of the two-step method for all combinations of $N \in \{25, 50, 100, 150, 300\}$ and $T \in \{25, 50, 100, 150, 300\}$.

As the consistency properties of the PCA estimates have already been shown in previous literature, we take as given that, with high values of N and T, we obtain a good estimate $\hat{f}_t$ for the factor f_t in the first step. Interestingly, our simulations[7] show that the factor can be estimated well even for $N = 50$ and $T = \{25, 50\}$. In their Monte-Carlo study, Stock and Watson (2002) confirm this result.For this reason, we also report the behavior of the two-step estimates samples as small as $N = 25$ and $T = 25$.[8]

4.1.3. Estimation

It is known that the PCA estimates of the factors are identified up to a sign change. We manually control the sign of the estimated factor by multiplying the $\hat{f}$ by -1 if its correlation with the true factor is negative. In practice, is it also usually possible to recover the sign of the true factor.

For each replication, the likelihood function $\mathcal{L}_T(\hat{f}, \theta)$ is maximized under constraints on transition probabilities (to insure that they lie in the open unit interval) and variance (to insure that it is positive). The maximum likelihood estimate is obtained with SQP[9] (sequential quadratic programming) method, which is essentially a version of Newton's method for constrained optimization. At each iteration, the Hessian of the function is updated using the Broyden-Fletcher-Goldfarb-Shranno (BFGS) method. The variance-covariance matrix $\hat{\Omega}_T^{-1}$ of the estimates $\hat{\theta}(\hat{f})$ is then estimated as

$$\hat{\Omega}_T = -\frac{1}{T}\left(\frac{\partial^2 \mathcal{L}_T(\theta, \hat{f})}{\partial\theta\partial\theta'}\right)\Bigg|_{\theta=\hat{\theta}}.$$

4.1.4. Measures of quality

In order to quantify the impact of the use of $\hat{f}$ instead of f on the ML estimates of the second step, we compare the empirical distributions of the two-step estimates to the MLE obtained on the observed factor. We denote by $\hat{\theta}(f)$ (resp. $\hat{\theta}(\hat{f})$) the vector estimates obtained with equation (3) (resp. equation (7)), and by $\hat{\theta}_i(f)$ (resp. $\hat{\theta}_i(\hat{f})$) the i-th component of $\hat{\theta}(f)$ (resp. $\hat{\theta}(\hat{f})$). For each pair of elements $\hat{\theta}_i(f)$ and $\hat{\theta}_i(\hat{f})$, we compute two following measures:

1. the Kullback-Leibler divergence

$$D_{KL}(F_{\hat{f}}||F_f) = \sum_j F^i_{\hat{f}}(j) \ln \frac{F^i_{\hat{f}}(j)}{F^i_f(j)},$$

where $F^i_{\hat{f}}$ and F^i_f are the empirical cumulative distribution functions of $\hat{\theta}_i(f)$ and $\hat{\theta}_i(\hat{f})$, and $F_f(j)$ and $F_{\hat{f}}(j)$ are probability measures on a bin j.[10] This is a measure of the information

[7]To save space, we do not report these results here, but they are surely available on request.

[8]This setting may be interesting for the empirical studies of the business cycle in countries with limited availability of data.

[9]which corresponds to the *sqp* optimization algorithm of the *fmincon* optimizer in Matlab R2015a.

[10]In order to render the KL distances of the parameters comparable between each other, we set the width of a bin to 0.25, the minimum value which guarantees non-emptiness of $F^i_f(j)$ for the distributions of each element of $\hat{\theta}_i(\hat{f})$.

loss when $F^i_{\hat{f}}$ is used to approximate F^i_f and corresponds to a proxy of the expectation of the logarithmic difference.

2. the Kolmogorov-Smirnov statistic

$$KS_i = \sup_{\hat{\theta}_i} |F_f(\hat{\theta}_i) - F_{\hat{f}}(\hat{\theta}_i)|,$$

where $\sup_{\hat{\theta}_i}$ is the supremum of the set of distances between $F^i_{\hat{f}}$ and F^i_f at different values of $\hat{\theta}_i$. KS_i shows the maximum deviation of $F_f(\hat{\theta}_i)$ from $F_{\hat{f}}(\hat{\theta}_i)$. The statistic is used for the Kolmogorov-Smirnov test with the null hypothesis that $\hat{\theta}_i(f)$ and $\hat{\theta}_i(\hat{f})$ come from the same distribution. The null is rejected at 5% when $KS_i > 0.043$.[11] The KS statistic and the corresponding test thus show whether the two empirical distributions are statistically different. It's important to underline that the test requires that the two empirical distributions correspond to independent samples. For this reason, we compute $F_f(\hat{\theta}_i)$ and $F_{\hat{f}}(\hat{\theta}_i)$ using two disjoint subsets of 1000 replications, i.e. using the first 1000 replications for $F_f(\hat{\theta}_i)$ and the last 1000 replications for $F_{\hat{f}}(\hat{\theta}_i)$.

To study the small-sample behavior and consistency of the two-step estimates, we report the ratio $\frac{\hat{\theta}_i(\hat{f})}{\theta_{0i}}$, where θ_{0i} is the genuine value of the parameter, and the ratio between the mean estimated standard error and sampling standard error of $\hat{\theta}_i(\hat{f})$.

To measure the ability of the model to identify states we compare the obtained estimates of filtered probability $P(S_t = 1|I_t)$ to the true sequence of states. To simplify notations, we use $FP_t(\hat{f}) = P(S_t = 1|\hat{f}_t, \hat{f}_{t-1}, ..., \hat{f}_1)$ and $FP_t(f) = P(S_t = 1|f_t, f_{t-1}, ..., f_1)$ for the filtered probability corresponding to equations (7) and (3), respectively. We use the following quality indicators:

1. the quadratic probability score by Brier (1950), which measures the average quadratic deviation of the filtered probability from the true state and is defined as

$$QPS = \frac{1}{T}\sum_{t=1}^{T}(S_t - FP_t(\hat{f}))^2;$$

2. false positives score, which measures the average number of wrongly identified states in the sample, under the assumption that $S_t = 1$ when $FP_t(\hat{f}) > 0.5$[12] and $S_t = 0$ otherwise; FPS is defined as

$$FPS = \frac{1}{T}\sum_{t=1}^{T}(S_t - I_{FP_t(\hat{f})>0.5})^2;$$

[11] In the general case, the null is rejected when $KS_{n,n'} > c(\alpha)\sqrt{\frac{n+n'}{nn'}}$ where $c(\alpha)$ is the quantile of the Kolmogorov distribution ($c(\alpha) = 1.36$ for $\alpha = 0.05\%$), n and n' are the sizes of first and second sample respectively.

[12] The cut-off threshold of 0.5 for the filtered probability of recession is chosen arbitrary, however, it is quite common in the literature.

3. the correlation between the true state and $FP_t(\hat{f})$

$$r_1 = corr(FP_t(\hat{f}),\, S_t);$$

4. the correlation between the true state and the filtered probability of recession inferred from the dynamics of the true factor

$$r_2 = corr(FP_t(f),\, S_t)$$

   which allows to evaluate the performance of the Markov-Switching model for the identification of the state S_t in finite samples. By comparing r_1 and r_2, we can assess the impact of the use of the proxy of the factor $\hat{f}_t$ instead of the factor f_t itself.

While the correlations measure how well the filtered probability follows the business cycle, the QPS and FPS show how reliable it is about the estimate of the state and how often it fails. The same indicators QPS, FPS, r_1 and r_2 are computed for the smoothed probability of recession.

Finally, it is interesting to study whether the distribution of the two-step estimates has the same properties as the MLE of the MS-AR model. Indeed, in case of a regular Markov-Switching autoregressive model, is often assumed that under sufficient regularity conditions,[13] the MLE $\hat{\theta}$ is Gaussian and so $\sqrt{T}(\hat{\theta} - \theta_0)$ converges in distribution to $N(0, \Omega_0^{-1})$ as $T \to \infty$, where

$$\Omega_0 = \lim_{T\to\infty} \frac{1}{T} \left(\frac{\partial^2 \mathcal{L}_T(\hat{f}, \theta)}{\partial\theta\partial\theta'} \right)\Bigg|_{\theta=\theta_0},$$

and Ω_0 is the information matrix.

In order to verify whether the two-step estimates have normal distribution (or tend to it asymptotically), we study the conventional t-statistics corresponding to the elements of $\hat{\theta}_i$:

$$t_i = \frac{\hat{\theta}_i(\hat{f}) - \theta_{0i}}{\hat{\sigma}_{\hat{\theta}_i(\hat{f})}}.$$

If the two-step estimates are asymptotically normal, the t-statistics should also have asymptotically Gaussian distribution. In this case, the use of Wald-type tests (including significance tests) when interpreting the results of the MS-DFM is justified. We examine this hypothesis by analyzing the mean, the skewness and the excess kurtosis of the distribution of t_i, as well as run the Kolmogorov-Smirnov normality test. As an additional indicator of gaussianity, we also compute the empirical rejection rates of the test with the null $H_0 : E(\hat{\theta}_i(\hat{f})) = \theta_{0i}$. If the empirical rejection rate coincides with the theoretical one, this is regarded as an additional sign of normality of the distribution.

[13]see, for example, Kiefer (1978).

4.2. Simulation results

In this section we provide simulation results for the baseline scenario. The experiments were performed on SCSCF, a multiprocessor cluster system owned by Università Ca'Foscari Venezia.

4.2.1. The impact of the first step

Figure 1 and Table 1 provide the information on how the first step - the use of estimated factor instead of the true one - modifies the ML estimates of the Markov-Switching autoregressive model. For each matrix in Figure 1, the change in the color columnwise corresponds to the effect of the increase of the number of series N, while the change rowwise to the increase of the number of observations T. As expected, the Kullback-Leibler distance between the empirical distributions of $\hat{\theta}_i(\hat{f})$ and $\hat{\theta}_i(f)$ decreases when N rises. However, we observe the distance increase when T rises for a given N. This intuitively contradictory finding is connected to the presence of replications with aberrant results, i.e. replications with estimated transition probabilities very close to 0 or 1 (which is an implausible result since it implies that the underlying Markov chain is not irreducible), or with $\hat{\beta}_0$ is very close to $\hat{\beta}_1$ (in this case, the states are not identified, neither are the parameters). In most cases, these estimates correspond to the convergence of the maximum likelihood optimizer to a wrong local maximum. Present in both $\hat{\theta}(\hat{f})$ and $\hat{\theta}(f)$, this kind of estimates form an additional mode which distorts the distributions. Since these obviously abnormal values of the estimators can be easily identified as such while working with the real data, we discard them from our analysis. Typically, their amount is not large (5%-10% of replications, depending on the parameter), so the remaining number of replications is still large enough to analyze the properties of the two-step estimator.[14]

Figure 2 reports the Kullback-Leibler divergence between $\hat{\theta}_i(\hat{f})$ and $\hat{\theta}_i(f)$ when the implausible replications are discarded. In this case we observe convergence both in N and T. Importantly, for all parameters the distance is high for $N < 100$. When $N > 150$, little improvement can be achieved by increasing the number of series even more, so the major factor of proximity between the two distributions is the number of observations. This observation is validated by the Kolmogorov-Smirnov tests reported in Table 1, where the null that $\hat{\theta}_i(\hat{f})$ and $\hat{\theta}_i(f)$ come from the same distribution is not rejected in the majority of cases.

[14]For the purpose of comparison, we compute several tables reported in this section for all replications (see Appendix C). Other tables are available on request.

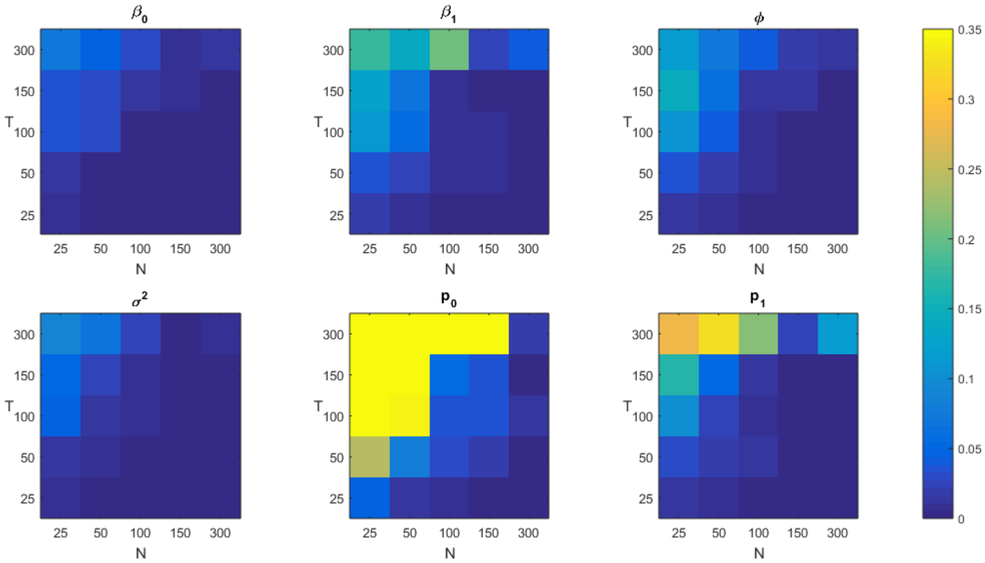


Figure 1: Kullback-Leibler distance between $\hat{\theta}_i(\hat{f})$ and $\hat{\theta}_i(f)$, with $\theta = (\beta_0, \beta_1, \varphi, \sigma^2, p_0, p_1)$

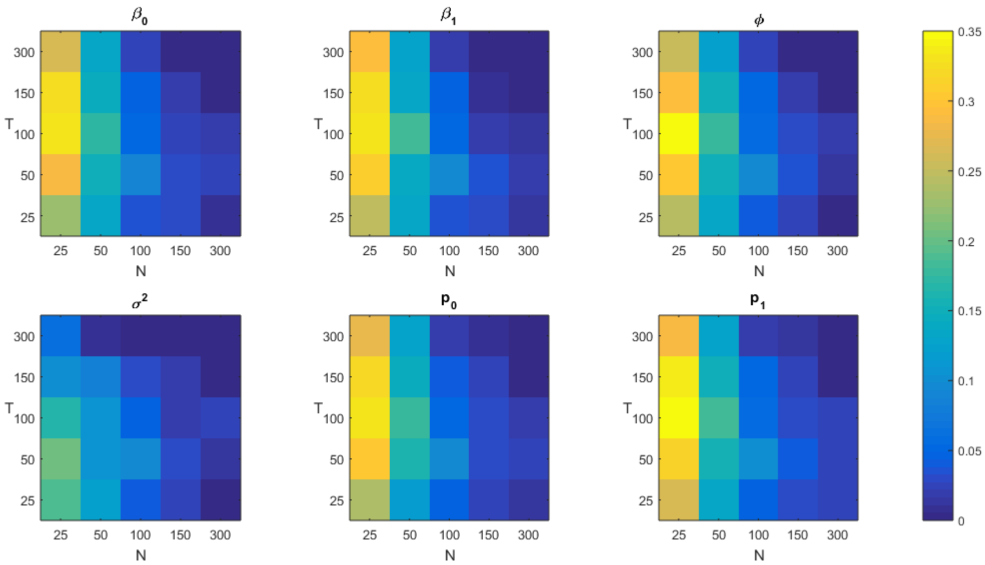


Figure 2: Kullback-Leibler distance between $\hat{\theta}_i(\hat{f})$ and $\hat{\theta}_i(f)$, with $\theta = (\beta_0, \beta_1, \varphi, \sigma^2, p_0, p_1)$, aberrant replications excluded

Table 1: Test statistic of the Kolmogorov-Smirnov test

N	T	β_0	β_1	φ	σ^2	p_0	p_1
25	25	0.05	0.07	0.08	0.10	**0.04**	**0.04**
25	50	0.05	0.05	**0.03**	0.11	**0.04**	**0.04**
25	100	0.08	0.07	0.06	0.14	0.05	**0.03**
25	150	0.09	**0.03**	**0.04**	0.15	**0.03**	**0.04**
25	300	0.09	**0.03**	0.07	0.15	**0.04**	**0.04**
50	25	**0.04**	0.05	**0.04**	**0.04**	0.05	0.06
50	50	**0.03**	0.06	**0.03**	0.10	**0.04**	**0.04**
50	100	**0.03**	**0.04**	**0.03**	0.10	**0.04**	**0.04**
50	150	0.07	0.06	0.06	0.09	0.06	**0.03**
50	300	**0.04**	0.05	**0.03**	0.12	0.06	**0.03**
100	25	0.06	**0.03**	**0.04**	0.05	**0.04**	0.05
100	50	**0.03**	**0.03**	**0.03**	0.05	0.05	0.05
100	100	**0.03**	0.06	**0.03**	**0.04**	**0.04**	**0.04**
100	150	**0.03**	**0.03**	**0.03**	0.07	**0.04**	0.06
100	300	**0.03**	**0.03**	**0.03**	0.08	**0.04**	**0.04**
150	25	**0.04**	0.06	0.05	**0.03**	0.06	0.06
150	50	0.07	**0.02**	0.03	**0.03**	0.05	0.06
150	100	**0.02**	0.05	0.04	**0.04**	**0.04**	0.05
150	150	**0.03**	**0.03**	**0.03**	**0.04**	0.06	**0.03**
150	300	**0.02**	0.05	**0.04**	0.07	**0.02**	0.06
300	25	**0.03**	0.05	0.05	**0.02**	0.06	0.05
300	50	**0.03**	**0.04**	**0.03**	**0.03**	**0.04**	**0.04**
300	100	**0.03**	**0.03**	**0.02**	0.06	**0.04**	**0.04**
300	150	**0.04**	**0.03**	**0.02**	**0.04**	**0.03**	**0.03**
300	300	**0.04**	**0.03**	**0.03**	**0.04**	**0.03**	**0.04**

The null hypothesis of the test is that $\hat{\theta}_i(\hat{f})$ and $\hat{\theta}_i(f)$, with $\theta = (\beta_0, \beta_1, \varphi, \sigma^2, p_0, p_1)$, are from the same continuous distribution. The null is rejected when $KS > 0.043$. The cases when the null is not rejected are marked with bold font.

4.3. Consistency and small-sample performance of the two-step estimates $\hat{\theta}(\hat{f})$

4.3.1. Mean bias

Figure 3 provides the ratios of the two-step estimates $\hat{\theta}_i(\hat{f})$ to the true values of the parameters θ_{0i} averaged over replications (the exact values of the ratios are given in Table B.1 in Appendix), whereas B.1 sheds light on their distributions.

We observe that, as T and N rise, the ratio for all elements of $\hat{\theta}(\hat{f})$ approaches one indicating consistency of the estimates. Interestingly, the convergence is achieved faster for the estimates of the parameters corresponding to the switch, i.e. β_0, β_1, p_0 and p_1, the deviation of the estimates of transition probabilities being no more than 3% even for very small N and T. The estimates of φ and σ^2 require greater N and T to approach their true values (the deviation is around 2%-4% for and φ and is much higher for σ^2). As expected, the rate of convergence is lower for the two-step estimates in comparison to the estimates computed with the observed factor (see Table D.1 in the Appendix), the convergence of $\hat{\theta}(f)$ is achieved at $T = 150$ already.

In case of small T and N, the estimated values of the parameters deviate from their true values. The bias is generally a decreasing function of the sample size and the number of series, and for most design points is substantially different from zero. Consistent with Psaradakis and Sola (1998), the estimates of φ, p_0, p_1 and β_1[15] are always downward biased, whereas the estimates of β_0 and σ^2

[15]In our setting, its true value is $\beta_1 = -2$; the values above 1 in Table B.1 indicate that the average value of $\hat{\beta_1}$ is

are upward biased.

4.3.2. Standard error bias

The accuracy of the estimated small-sample standard errors as approximations of the sampling standard deviation of the two-step estimates also provides additional information on the convergence. Figure 4 reports the ratio between the estimated standard errors averaged over replications and sampling standard deviation of $\hat{\theta}_i(\hat{f})$.

The ratio approaches 1 as T and N rise, which advocates for consistency of the two-step estimates and for the accuracy of the numerical estimates of the standard errors. The only exception is the standard error of $\hat{\sigma}^2$, which tends to have an overestimated standard error when N is small, however, we observe convergence to one for $N > 300$. For smaller samples and number of observations, the standard errors appear to be overestimated for β_0, β_1, φ and p_0 and underestimated for p_1.

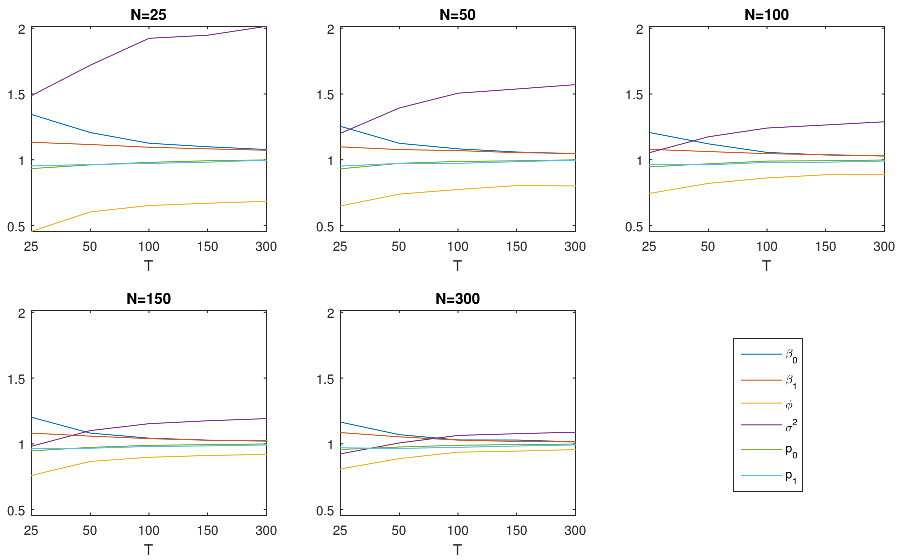


Figure 3: Mean ratio between the two-step estimate of the parameter $\hat{\theta}_i(\hat{f})$ and its true value θ_{0i}

more negative, i.e. downward biased, in small samples.

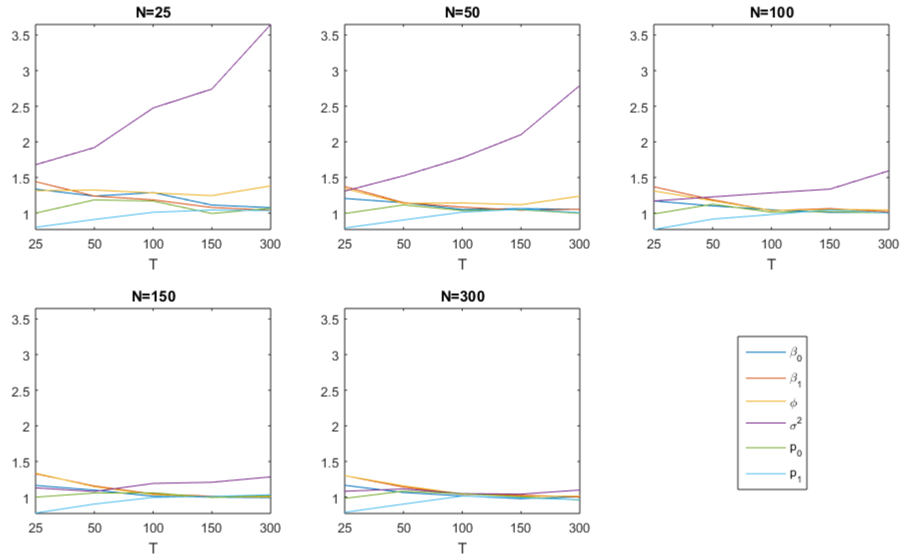


Figure 4: Ratio between mean estimated standard deviation $\hat{\sigma}_{\hat{\theta}_i(\hat{f})}$ and sample standard deviation of $\hat{\theta}_i(\hat{f})$

4.4. Identification of states

Figures 5 and 6 demonstrate the ability of the model to identify recession states.

As expected, the quality of state identification increases with the precision with which the factor is estimated, and thus with the number of series N. In the same time, the factor reveals more information about existing states when T is higher.

By comparing the sequence of filtered probabilities with the sequence of realized states we assess the quality of nowcasts of the current state of the cycle. Figure 5 shows that with $T > 50$ and $N > 100$, the model erroneously assigns a high probability of recession in at most 10% of cases ($FPS < 0.10$). In real empirical applications, the number of series that would produce the same quality should be lower, as the data are usually much less noisy (the results of scenario 4.1 confirm this hypothesis). The correlation of the filtered probability of recession with the true states r_1 is getting closer to r_2, its counterpart obtained with the observed factor, as N rises (up to almost coinciding when $N = 300$) and is very high under values of N and T close to those usually used in practice (above 0.8 for all T and $N > 150$).

The recession identification performance of the model is even higher for the retrospective analysis of the cycles, i.e. for the smoothed probabilities. With $T > 100$ and $N > 50$, the FPS is below 0.10 and attains 0.03 with 300 observations on 300 series.

To conclude, it is worth noting that notwithstanding the bias in the two-step estimates when the number of series and observations are small, the two-step estimates seem to be precise enough to insure high performance of the MS-DFM in terms of identification of states, especially a posteriori.

This finding seems particularly encouraging to us, since dating of the moments of changes of states is the major output of the MS-DFM.

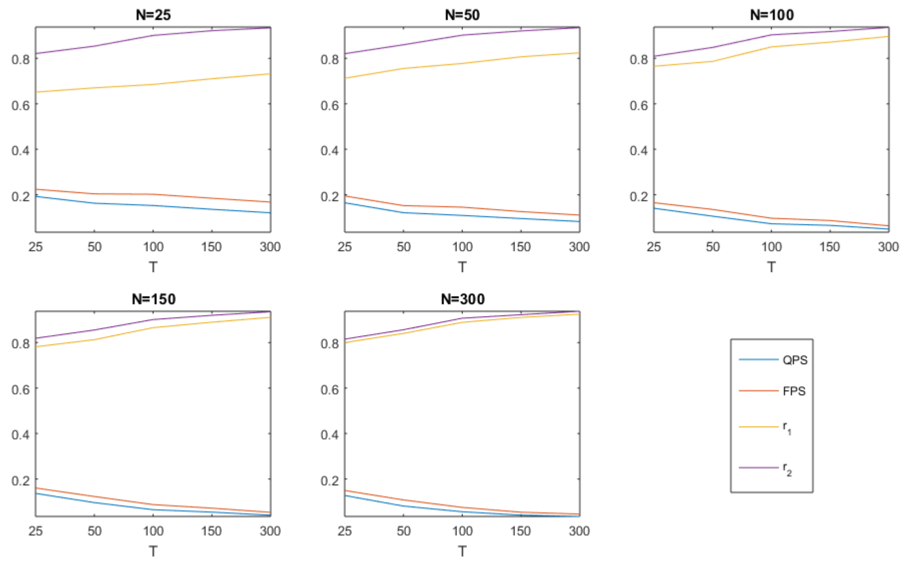


Figure 5: Quality of state identification: filtered probability of recession

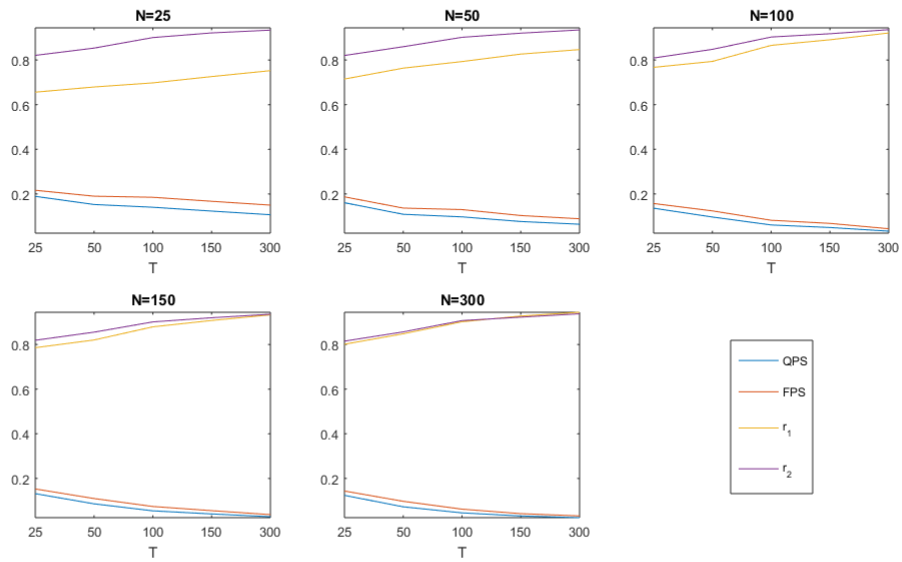


Figure 6: Quality of state identification: smoothed probability of recession

4.5. Analysis of t-statistics

Table E.1, Figure E.1 and Table E.2 in the Appendix show the mean, skewness and excess kurtosis of the t-statistics. As Table B.1, Table E.1 reports the bias of the estimates, but measured in standard deviations of the estimates. Once the standard deviations are accounted for, we observe that the bias of the estimates of p_0 and p_1 is very close to zero. The direction of bias of the t-statistics of the other estimates is in accordance with the results of Table B.1. One may notice that the values of the t-statistics in most design points is below 1.96 in absolute value, indicating that if the two-step estimates were asymptotically normal, the null $H_0 : E(\hat{\theta}_i) = \theta_{0i}$ would not be rejected.

For all values of N, the distribution of the t-statistics of the estimates are skewed, and skewness often changes sign when passing from small T to higher T and diminishes as N rises. Following the sign of the sample mean of the t-statistics, the skewness is negative for $t_{\hat{\varphi}}$ and $t_{\hat{\beta}_1}$, and positive for the other parameters. In particular, the empirical distribution of the two-step estimates of β_0, β_1 and σ^2 have thick right tails, whereas those of φ, p_0 and p_1 have thick left tails. The skewness of the two-step estimates of σ^2, p_0 and p_1 is probably not very surprising since the estimation was implemented under the constraint that the variance should be positive and the transition probabilities should lie in the open unit interval. These restrictions are likely to lead to finite-sample distributions resembling truncated ones, since more probability mass would lie in the required interval than would in a model with no restriction on the parameters. This effect is clearly visible on the boxplots of the ratio of the two-step estimates to their true values (see Figure B.1 in the Appendix).

Like the estimates obtained on the observed factors in small samples (see Psaradakis and Sola (1998)), the two-step estimates and their t-statistics are skewed and leptokurtic, at least in the samples of size considered in this paper. We find nevertheless that in some cases the t-statistics are Gaussian (see Figure F.1): the Kolmogorov-Smirnov test shows that normality of the distribution of $t_{\hat{p}_1}$ is not rejected at 5% for all N and high T (and for small N and high T in case of $t_{\hat{p}_0}$), however it is rejected for all other parameters. To the contrary, Jarque-Bera test does not reject normality for other parameters (see Figure F.2), so no unambiguous conclusion on the normality of the two-step estimates can be derived.[16]

To get additional insight on the potential normality of the distribution of the two-step estimates, we analyze the empirical size of the tests of the null $H_0 : E(\hat{\theta}_i) = \theta_{0i}$ (see Tables E.3 and E.4). We observe that due to the distortions in the distributions, the empirical size of the tests is always above its the nominal size. As distortions attenuate, the empirical size gets closer to its nominal counterpart, which we observe in case of β_0, β_1 (when the number of series and observations is high) and, in particular, p_0 and p_1 (for $N > 100$). Consistent with Psaradakis and Sola (1998), the difficulties are the most serious for the t-statistics of $\hat{\sigma}^2$.

All the observations listed above lead us to the conclusion that the two-step estimates $\hat{\theta}(\hat{f})$[17] are usually not normal for $N < 300$ and $T < 300$. This implies that the tests using this assumption (such as t-tests on significance of the coefficients, Wald-type tests, and other) are likely to be invalid and should be used with caution.

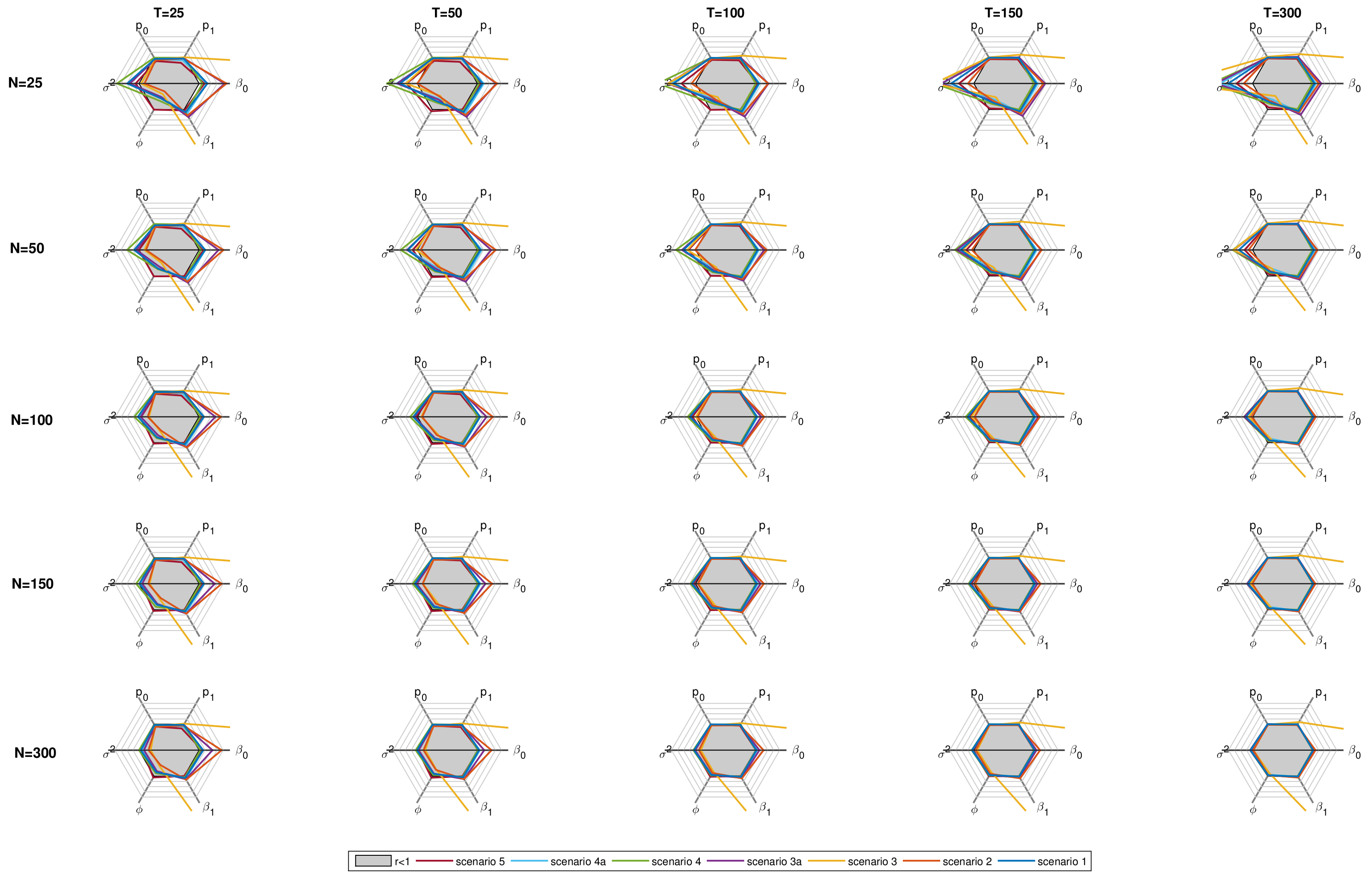
[16]Different normality tests are known to have different power depending on the shape of the distribution. Jarque-Bera is considered to be the most powerful when the distribution is symmetrical (i.e. in case of $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\varphi}$), but it is overcome by Kolmogorov-Smirnov test in other cases (see Thadewald and Bning (2007) for details).

[17]as well as their counterpart $\hat{\theta}(f)$, as reported by Psaradakis and Sola (1998).

4.6. Other scenarios

To analyze the behavior of two-step estimates under different parameter sets, we discuss the results obtained with the other scenarios in terms of their deviation from the baseline scenario. Figures 7 and 8 show the ratio $\frac{\hat{\theta}_i(\hat{f})}{\theta_{0i}}$ and the indicators of quality of state identification, allowing us to compare the small-sample bias and the reliability of estimates of the current and past states.

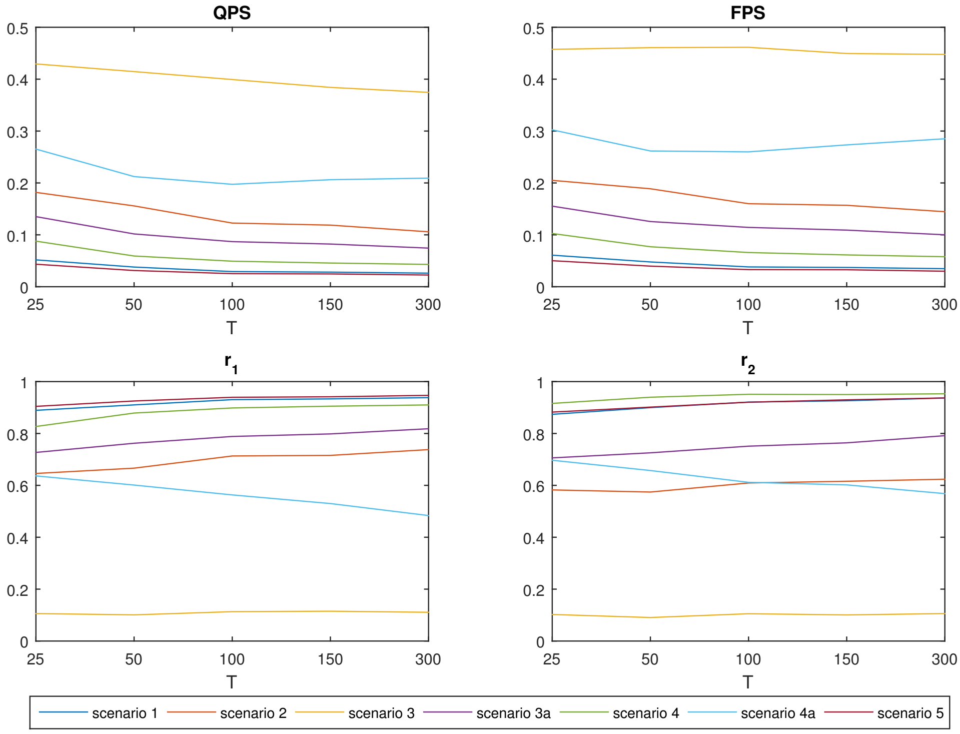
Figure 7: Mean values of two-step estimates for different scenarios



^{a}

[a]Note: The values on each axis varies from 0 to 2, the shaded area marking the values below or equal 1. The scenarios under consideration are: scenario 1 (baseline scenario): $c = 5$, $p_0 = 0.9$, $p_1 = 0.8$, $u = 0.1$, $\varphi = 0.3$; scenario 2 (noisy factor): $c = 2$, $p_0 = 0.9$, $p_1 = 0.8$, $u = 0.1$, $\varphi = 0.3$; scenario 3.1 (high autocorrelation): $c = 5$, $p_0 = 0.9$, $p_1 = 0.8$, $u = 0.1$, $\varphi = 0.9$; scenario 3.2 (medium autocorrelation): $c = 5$, $p_0 = 0.9$, $p_1 = 0.8$, $u = 0.1$, $\varphi = 0.6$; scenario 4.1 (impersistent states): $c = 5$, $p_0 = 0.5$, $p_1 = 0.5$, $u = 0.1$, $\varphi = 0.3$; scenario 4.2 (very persistent states): $c = 5$, $p_0 = 0.95$, $p_1 = 0.95$, $u = 0.1$, $\varphi = 0.3$; scenario 5 (homogeneous data): $c = 5$, $p_0 = 0.9$, $p_1 = 0.8$, $u = 0.5$, $\varphi = 0.3$

Figure 8: Indicators of quality of state identification for various scenarios, smoothed probability, $N = 100$



18

In spite of the fact that scenarios relate to very different conditions, we can track several commonalities:

- for small N and T, φ tends to be underestimated, whereas β_0 and σ^2 are overestimated;
- σ^2 is estimated better when N rises; $\hat{\beta}_0$ and $\hat{\beta}_1$ are more accurate with higher T;
- for all scenarios except scenario 3.1, the two-step estimates are very close to their true values for $N > 150$ and $T > 150$;
- the two-step estimates of transition probabilities are the most accurate.

The differences between scenarios are clearly visible under small N and T. When the factor is noisy (scenario 2), the bias of the two-step estimates amplifies greatly for the autoregressive coefficient and β_0 and β_1. When the factor dynamics has high persistence (φ is high, scenario 3.1), the two-step method tends to confuse it with a large distance in mean, overestimating both constants to a large extent, and underestimating φ. The problem, however, disappears once the characteristic root is far enough from unity (scenario 3.2). The two-step method appears to be resistant to different degrees of persistence in states. For both low persistence and high persistence cases (scenario 4.1 and 4a, respectively), the distortions are comparable to the baseline scenario, being slightly higher in case of frequently changing regimes when N is low. The last scenario (homogeneous data) is, not surprisingly, the most favourable of all, bringing two-step estimates close to their true values at $N > 100$ and $T > 150$.

In terms of quality of state identification, with $N \geqslant 100$ the two-step estimates lead to performance usually considered as acceptable in all scenarios except the case of high autoregressive coefficient (scenario 3).[19] Not surprisingly, more favourable scenarios (homogeneous data, low persistence in states) lead to more accurate estimates of state, whereas high φ and very persistent states deteriorate the ability of the model to identify states.

Finally, Figures F.1 and F.2 show the results of the Kolmogorov-Smirnov and Jarque-Berra normality tests. The results appear to be contradictory, leaving the asymptotic distribution of the two-step estimates an open question.

5. Conclusion

In this paper we analyze the consistency and small-sample performance of a two-step estimator of the Markov-Switching Dynamic Factor model with the help of Monte Carlo simulations.

We observe that the empirical average of the estimates approaches the true value of the parameters when the number of observations and number of series rise. Together with convergence of the mean estimated standard errors to the sampling standard errors of the estimates, these two facts indicate consistency of the two-step method.

We find that, under values of parameters of the data generating process close to the ones usually observed in empirical applications, the estimates of the switching constants and transition probabilities are close to their true values when the the dataset contains more than 150 series and with at least 150 observations. The convergence of the estimates of the autoregressive coefficient and the variance of the error term requires higher N and T. The results of the baseline scenario can be improved by the use of informative and homogeneous data (in terms of signal-to-noise), which brings the two-step estimates close to the true values at $N > 100$ and $T > 150$.

Consistent with previous results concerning the estimates of a simple autoregressive Markov Switching model (in the context of this paper this is equivalent to the hypothesis that the factor is observed), our findings indicate that the estimates are biased when T and N are small. In fact, the autoregressive coefficient tends to be underestimated, whereas the variance of the error terms and the constants (in absolute value) are overestimated. The precision of the variance of the error term increases when N rises, while the estimates of constants are more accurate with higher T. Importantly, the estimates of transition probabilities have almost no bias with N and T as small as $N = 50$ and $T > 50$. These observations are also valid for various deviations from the baseline DGP.

[19] Interestingly, the estimates of the model on the observed factor also lead to poor estimates of current and past states. This is somehow in contradiction with Psaradakis and Sola (1998), where the consistency is achieved for all parameters of the model (including switching constants). This difference might be explained by the use of a different specification of the model, i.e. with a switch in mean instead of constant; this feature softens the effect of the switch).

When the baseline DGP is modified, we observe that the bias of the two-step estimates increases a lot for the autoregressive coefficient and the constants when the factor is noisier. When the autoregressive coefficient is close to unity, the two-step method tends to confuse the effect of high persistence in the dynamics with a large distance in mean, overestimating both constants to a large extent, and underestimating the autoregressive component. The problem, however, disappears once the characteristic root is far enough from unity. The two-step method appears to be resistant to different degrees of persistence in states. For both low persistence and high persistence cases the distortions are comparable to the baseline scenario, being slightly higher in case of frequently changing regimes when N is low.

In spite of the bias in small samples, the two-step estimates still lead to plausible state-detection performance of the MS-DFM with a dataset of dimensions commonly used in the business cycle analysis ($T > 100$, $N > 100$), producing the filtered and smoothed probability of recession which are highly correlated with the true underlying sequence of states and giving a reasonable amount of false recession signals.

The empirical distributions of the t-statistics associated to the two-step estimates in finite samples tend to be skewed and leptokurtic and non-normal according to the Kolmogorov-Smirnov test. Therefore, some of the traditional tests using the normality assumption (such as t-student significance test or Wald-type test) are likely to be invalid. Similarly, the standard errors of the estimates should better be bootstrapped for small N and T. A positive exception are the parameters of transition probabilities, which were found to be normally distributed when $T > 300$.

This paper shows the general validity of the two-step estimation method for small-samples. It seems however likely that its performance can be improved by using more efficient estimates to estimate the factor on the first step. For example, the use of two-step method or Quasi-Maximum likelihood estimates proposed by Doz et al. (2011) and Doz et al. (2012) for the first step might probably lead to more precise estimates in the second step.

References

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

Bai, J., Ng, S.. Principal components estimation and identification of static factors. Journal of Econometrics 2013;176(1):18–29.

Bessec, M., Bouabdallah, O.. Forecasting GDP over the Business Cycle in a Multi-Frequency and Data-Rich Environment. Oxford Bulletin of Economics and Statistics 2015;77(3):360–384.

Brave, S., Butters, R.A.. Gathering insights on the forest from the trees: a new metric for financial conditions. Technical Report; 2010.

Brier, G.W.. Verification of forecasts expressed in terms of probability. Monthly Weather Review 1950;78(1):1–3.

Camacho, M., PerezQuiros, G., Poncela, P.. Markov-switching dynamic factor models in real time. Working Papers 1205; Banco de Espaa;Working Papers Homepage; 2012.

Camacho, M., PerezQuiros, G., Poncela, P.. Extracting Nonlinear Signals from Several Economic Indicators. Journal of Applied Econometrics 2015;30(7):1073–1089.

Chamberlain, G.. Funds, factors, and diversification in arbitrage pricing models. Econometrica 1983;51(5):1305–1323.

Chamberlain, G., Rothschild, M.. Arbitrage, factor structure, and mean-variance analysis on large asset markets. Econometrica 1983;51(5):1281–1304.

Chauvet, M.. An econometric characterization of business cycle dynamics with factor structure and regime switching. International Economic Review 1998;39(4):969–96.

Chauvet, M., Senyuz, Z.. A Dynamic Factor Model of the Yield Curve as a Predictor of the Economy. Finance and Economics Discussion Series 2012-32; Board of Governors of the Federal Reserve System (U.S.); 2012.

Chauvet, M., Yu, C.. International business cycles: G7 and oecd countries. Economic Review 2006;(Q 1):43–54.

Connor, G., Korajczyk, R.A.. Performance measurement with the arbitrage pricing theory: A new framework for analysis. Journal of Financial Economics 1986;15(3):373 – 394.

Darné, O., Ferrara, L.. Identification of Slowdowns and Accelerations for the Euro Area Economy. Oxford Bulletin of Economics and Statistics 2011;73(3):335–364.

Davig, T.A.. Detecting recessions in the Great Moderation: a real-time analysis. Economic Review 2008;(Q IV):5–33.

Diebold, F.X., Rudebusch, G.D.. Measuring Business Cycles: A Modern Perspective. The Review of Economics and Statistics 1996;78(1):67–77.

Dolega, M.. Tracking Canadian Trend Productivity: A Dynamic Factor Model with Markov Switching. Discussion Papers 07-12; Bank of Canada; 2007.

Douc, R., Moulines, E., Olsson, J., van Handel, R.. Consistency of the maximum likelihood estimator for general hidden markov models. Ann Statist 2011;39(1):474–513.

Douc, R., Moulines, E., Rydén, T.. Asymptotic properties of the maximum likelihood estimator in autoregressive models with markov regime. The Annals of Statistics 2004;32(5):2254–2304.

Doz, C., Giannone, D., Reichlin, L.. A two-step estimator for large approximate dynamic factor models based on Kalman filtering. Journal of Econometrics 2011;164(1):188–205.

Doz, C., Giannone, D., Reichlin, L.. A QuasiMaximum Likelihood Approach for Large, Approximate Dynamic Factor Models. The Review of Economics and Statistics 2012;94(4):1014–1024.

Francq, C., Roussignol, M.. On white noises driven by hidden markov chains. Journal of Time Series Analysis 1997;18(6):553–578.

Francq, C., Roussignol, M.. Ergodicity of autoregressive processes with markov-switching and consistency of the maximum-likelihood estimator. Statistics 1998;32(2):151–173.

Hamilton, J.. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 1989;57(2):357–84.

Hamilton, J.D.. A Quasi-Bayesian Approach to Estimating Parameters for Mixtures of Normal Distributions. Journal of Business & Economic Statistics 1991;9(1):27–39.

Hamilton, J.D.. Specification testing in Markov-switching time-series models. Journal of Econometrics 1996;70(1):127–157.

Hosmer, D.W.. On mle of the parameters of a mixture of two normal distributions when the sample size is small. Communications in Statistics 1973;1(3):217–227.

Juhn, C., Potter, S., Chauvet, M.. Markov switching in disaggregate unemployment rates. Empirical Economics 2002;27(2):205–232.

Kiefer, N.M.. Discrete parameter variation: Efficient estimation of a switching regression model. Econometrica 1978;46(2):427–434.

Kim, C.J.. Dynamic linear models with Markov-switching. Journal of Econometrics 1994;60(1-2):1–22.

Kim, C.J., Nelson, C.R.. Business Cycle Turning Points, A New Coincident Index, And Tests Of Duration Dependence Based On A Dynamic Factor Model With Regime Switching. The Review of Economics and Statistics 1998;80(2):188–201.

Kim, M.J., Yoo, J.S.. New index of coincident indicators: A multivariate markov switching factor model approach. Journal of Monetary Economics 1995;36(3):607 – 630.

Krishnamurthy, V., Ryden, T.. Consistent estimation of linear and non-linear autoregressive models with markov regime. Journal of Time Series Analysis 1998;19(3):291–307.

Paap, R., Segers, R., van Dijk, D.. Do leading indicators lead peaks more than troughs? Journal of Business & Economic Statistics 2009;27(4):528–543.

Psaradakis, Z., Sola, M.. Finite-sample properties of the maximum likelihood estimator in autoregressive models with Markov switching. Journal of Econometrics 1998;86(2):369–386.

Stock, J., Watson, M.. Forecasting Using Principal Components From a Large Number of Predictors. Journal of the American Statistical Association 2002;97:1167–1179.

Thadewald, T., Bning, H.. Jarquebera test and its competitors for testing normality a power comparison. Journal of Applied Statistics 2007;34(1):87–105.

Wang, J.m., Gao, T.m., McNown, R.. Measuring Chinese business cycles with dynamic factor models. Journal of Asian Economics 2009;20(2):89–97.

Appendix A. Variance of the factor

The formula for the variance of the factor is obtained in the following way.

$$V(f_t) = V(\beta_{S_t}) + \varphi^2 V(f_{t-1}) + V(\eta_t) + 2\varphi Cov(\beta_{S_t}, f_{t-1}). \tag{A.1}$$

The process (f_t) is stationary, so

$$V(f_t) = \frac{1}{1-\varphi^2}\left[V(\beta_{S_t}) + V(\eta_t) + 2\varphi Cov(\beta_{S_t}, f_{t-1})\right]. \tag{A.2}$$

Let us consider each part of $V(f_t)$ separately. The variance of the switching constant is:

$$\begin{aligned} V(\beta_{S_t}) &= V(\beta_0 + (\beta_1 - \beta_0)S_t) \\ &= (\beta_1 - \beta_0)^2 V(S_t) \\ &= (\beta_1 - \beta_0)^2 (E(S_t^2) - E^2(S_t)) \\ &= (\beta_1 - \beta_0)^2 (\pi - \pi^2). \end{aligned} \tag{A.3}$$

where $E(S_t) = E(S_t^2) = \pi$.

The covariance term $Cov(\beta_{S_t}, f_{t-1})$ is

$$\begin{aligned} Cov(\beta_{S_t}, f_{t-1}) &= Cov(\beta_{S_t}, \sum_{i=0}^{\infty} \varphi^i \beta_{S_{t-1-i}} + \sum_{i=0}^{\infty} \varphi^i \eta_{S_{t-1-i}}) \\ &= Cov(\beta_{S_t}, \sum_{i=0}^{\infty} \varphi^i \beta_{S_{t-1-i}}) \\ &= \sum_{i=0}^{\infty} \varphi^i Cov(\beta_{S_t}, \beta_{S_{t-1-i}}) \\ &= (\beta_1 - \beta_0)^2 \sum_{i=0}^{\infty} \varphi^i Cov(S_t, S_{t-i-1}), \end{aligned} \tag{A.4}$$

where

$$\begin{aligned} Cov(S_t, S_{t-i-1}) &= E(S_t S_{t-i-1}) - E(S_t)E(S_{t-i-1}) \\ &= E(S_t S_{t-i-1}) - \pi^2 \\ &= \sum_{j=0}^{1}\sum_{k=0}^{1} jkP(S_t = j|S_{t-i-1} = k)P(S_{t-i-1} = k) - \pi^2 \\ &= P(S_t = 1|S_{t-i-1} = 1)\pi - \pi^2. \end{aligned} \tag{A.5}$$

If $P(S_t|S_{t-1})$ is the transition probability matrix for one time step:

$$P(S_t|S_{t-1}) = \begin{pmatrix} p_0 & 1-p_0 \\ 1-p_1 & p_1 \end{pmatrix}, \tag{A.6}$$

and $P(S_t|S_{t-i-1})$ is the transition probability matrix for i time steps, then, according to Chapman-Kolmogorov theorem,

$$P(S_t|S_{t-i-1}) = (P(S_t|S_{t-1}))^{i+i}. \tag{A.7}$$

Using Cayley-Hamilton theorem or a diagonalization of this matrix, it can be shown that:

$$\begin{pmatrix} p_0 & 1-p_0 \\ 1-p_1 & p_1 \end{pmatrix}^n = \frac{\lambda_2\lambda_1^n - \lambda_1\lambda_2^n}{\lambda_2 - \lambda_1} I_2 + \frac{\lambda_2^n - \lambda_1^n}{\lambda_2 - \lambda_1} \begin{pmatrix} p_0 & 1-p_0 \\ 1-p_1 & p_1 \end{pmatrix}, \tag{A.8}$$

where λ_1 and λ_2 are the eigenvalues of the matrix $\begin{pmatrix} p_0 & 1-p_0 \\ 1-p_1 & p_1 \end{pmatrix}$, such that $\lambda_1 = 1$, $\lambda_2 = p_0 + p_1 - 1$. So,

$$\begin{pmatrix} p_0 & 1-p_0 \\ 1-p_1 & p_1 \end{pmatrix}^n = \frac{1}{p_0+p_1-2} \begin{bmatrix} (p_0-1)(p_0+p_1-1)^n + p_1 - 1 & (1-p_0)(p_0+p_1-1)^n + p_0 - 1 \\ (1-p_1)(p_0+p_1-1)^n + p_1 - 1 & (p_1-1)(p_0+p_1-1)^n + p_0 - 1 \end{bmatrix}. \tag{A.9}$$

Therefore,

$$\begin{aligned} P(S_t = 1|S_{t-i-1} = 1) &= (P(S_t|S_{t-1}))^{i+i}_{(2,2)} \\ &= \frac{1}{p_0+p_1-2}\left((p_1-1)(p_0+p_1-1)^{i+1} + p_0 - 1\right) \\ &= (1-\pi)(p_0+p_1-1)^{i+1} + \pi. \end{aligned} \tag{A.10}$$

Coming back to $Cov(S_t, S_{t-i-1})$:

$$\begin{aligned} Cov(S_t, S_{t-i-1}) &= (1-\pi)\pi(p_0+p_1-1)^{i+1} + \pi^2 - \pi^2 \\ &= (1-\pi)\pi(p_0+p_1-1)^{i+1}. \end{aligned} \tag{A.11}$$

Putting all terms together,

$$\begin{aligned} Cov(\beta_{S_t}, f_{t-1}) &= \sum_{i=0}^{\infty} \varphi^i (\beta_1 - \beta_0)^2 (1-\pi)\pi(p_0+p_1-1)^{i+1} \\ &= \frac{(\beta_1-\beta_0)^2(1-\pi)\pi(p_0+p_1-1)}{1-\varphi(p_0+p_1-1)}. \end{aligned} \tag{A.12}$$

Finally,

$$
\begin{aligned}
V(f_t) &= \frac{1}{1-\varphi^2}\left[(\beta_1-\beta_0)^2(\pi-\pi^2)+\sigma^2+\frac{2\varphi(\beta_1-\beta_0)^2(\pi-\pi^2)(p_0+p_1-1)}{1-\varphi(p_0+p_1-1)}\right] \\
&= \frac{1}{1-\varphi^2}\left[\sigma^2+(\beta_1-\beta_0)^2(\pi-\pi^2)\frac{1+\varphi(p_0+p_1-1)}{1-\varphi(p_0+p_1-1)}\right].
\end{aligned} \tag{A.13}
$$

The nullity of the $E(f_t)$ implies that $\beta_1 = \beta_0(1-\frac{1}{\pi})$, so the final expression for $V(f_t)$ becomes:

$$
V(f_t) = \frac{1}{1-\varphi^2}\left[\sigma^2+\beta_0^2\left(\frac{1-p_1}{1-p_0}\right)\left(\frac{1+\varphi(p_0+p_1-1)}{1-\varphi(p_0+p_1-1)}\right)\right]. \tag{A.14}
$$

Appendix B. Two-step estimates distribution

Table B.1: Mean ratio between the two-step estimate of the parameter $\hat{\theta}_i(\hat{f})$ and its true value θ_0

N	T	$\hat{\beta_0}/\beta_0$	$\hat{\beta_1}/\beta_1$	$\hat{\varphi}/\varphi$	$\hat{\sigma^2}/\sigma^2$	$\hat{p_0}/p_0$	$\hat{p_1}/p_1$
25	25	1.34	1.13	0.46	1.49	0.93	0.95
25	50	1.21	1.12	0.60	1.72	0.96	0.96
25	100	1.13	1.10	0.65	1.92	0.98	0.97
25	150	1.10	1.08	0.67	1.95	0.99	0.98
25	300	1.08	1.07	0.68	2.01	1.00	1.00
50	25	1.25	1.10	0.65	1.20	0.93	0.95
50	50	1.13	1.08	0.74	1.39	0.97	0.97
50	100	1.08	1.07	0.77	1.51	0.99	0.97
50	150	1.06	1.05	0.80	1.54	0.99	0.99
50	300	1.05	1.05	0.80	1.57	1.00	1.00
100	25	1.21	1.08	0.74	1.05	0.95	0.97
100	50	1.12	1.06	0.82	1.17	0.97	0.96
100	100	1.06	1.05	0.86	1.24	0.99	0.98
100	150	1.04	1.04	0.89	1.26	0.99	0.98
100	300	1.03	1.03	0.89	1.29	1.00	0.99
150	25	1.20	1.08	0.76	0.98	0.95	0.96
150	50	1.08	1.06	0.87	1.10	0.97	0.97
150	100	1.04	1.04	0.90	1.15	0.99	0.98
150	150	1.03	1.03	0.91	1.18	0.99	0.98
150	300	1.02	1.02	0.92	1.19	1.00	0.99
300	25	1.17	1.09	0.81	0.92	0.96	0.97
300	50	1.07	1.05	0.89	1.01	0.98	0.97
300	100	1.03	1.03	0.94	1.06	0.99	0.97
300	150	1.03	1.02	0.94	1.08	0.99	0.98
300	300	1.02	1.02	0.96	1.09	1.00	0.99

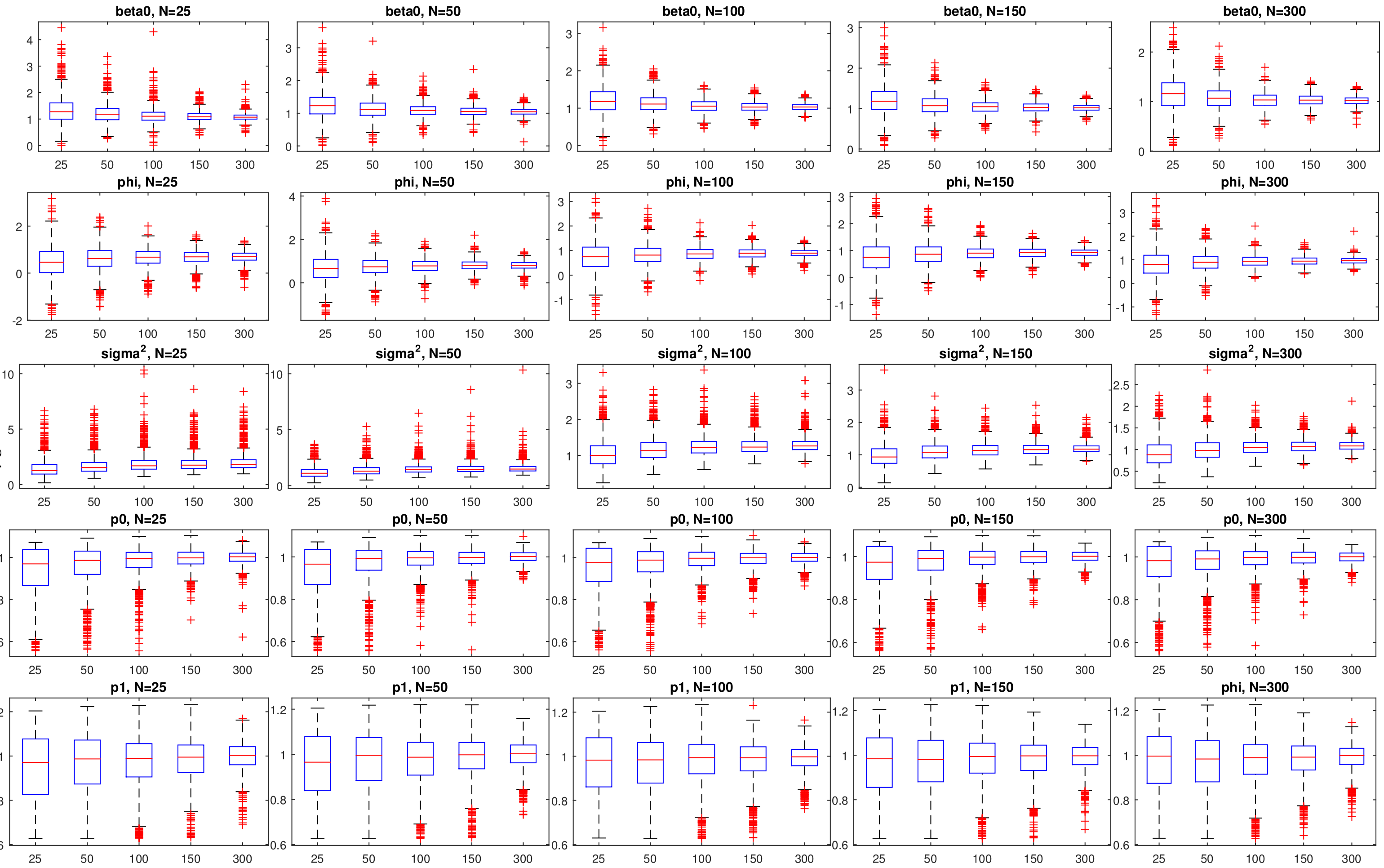


Figure B.1: Sample characteristics of the ratio of the two-step estimates to their true values

Appendix C. Results on unfiltered data

Table C.1: Test statistic of the Kolmogorov-Smirnov test

N	T	β_0	β_1	φ	σ^2	p_0	p_1
25	25	**0.04**	0.11	0.08	0.07	0.20	0.08
25	50	0.08	0.11	0.12	0.08	0.27	0.09
25	100	0.10	0.16	0.18	0.12	0.37	0.12
25	150	0.07	0.14	0.19	0.12	0.42	0.18
25	300	0.12	0.22	0.16	0.15	0.44	0.30
50	25	**0.04**	0.08	**0.03**	**0.03**	0.15	0.10
50	50	**0.04**	0.08	0.07	0.08	0.14	0.06
50	100	0.06	0.10	0.10	0.10	0.29	0.11
50	150	0.12	0.15	0.11	0.12	0.32	0.16
50	300	0.07	0.20	0.15	0.15	0.40	0.26
100	25	0.06	0.06	**0.03**	0.06	0.11	0.08
100	50	**0.04**	0.06	**0.04**	**0.04**	0.11	0.05
100	100	**0.03**	**0.04**	0.07	**0.03**	0.10	0.09
100	150	0.06	0.09	0.08	0.08	0.13	0.05
100	300	0.06	0.11	0.10	0.10	0.20	0.16
150	25	0.05	**0.03**	0.05	**0.02**	0.13	0.09
150	50	0.06	**0.04**	**0.04**	**0.02**	0.08	0.05
150	100	**0.03**	**0.03**	0.05	**0.04**	0.11	0.06
150	150	0.05	0.06	0.07	0.05	0.14	**0.04**
150	300	**0.04**	0.09	0.09	0.08	0.12	0.06
300	25	0.05	**0.03**	**0.04**	**0.02**	0.05	0.07
300	50	**0.04**	**0.03**	**0.04**	**0.04**	0.06	0.05
300	100	**0.03**	**0.04**	**0.04**	0.05	0.06	0.05
300	150	**0.04**	**0.02**	**0.02**	**0.04**	**0.02**	**0.02**
300	300	0.08	0.06	0.07	0.06	0.10	0.06

The null hypothesis of the test is that $\hat{\theta}_i(\hat{f})$ and $\hat{\theta}_i(f)$, with $\theta = (\beta_0, \beta_1, \varphi, \sigma^2, p_0, p_1)$, are from the same continuous distribution. The null is rejected when $KS > 0.043$. The cases when the null is not rejected are marked with bold font.

Table C.2: Mean ratio between the two-step estimate of the parameter $\hat{\theta}_i(\hat{f})$ and its true value

N	T	$\hat{\beta_0}/\beta_0$	$\hat{\beta_1}/\beta_1$	$\hat{\varphi}/\varphi$	$\hat{\sigma^2}/\sigma^2$	$\hat{p_0}/p_0$	$\hat{p_1}/p_1$
25	25	1.40	0.90	0.87	1.45	0.74	0.72
25	50	1.20	0.93	1.05	1.94	0.77	0.74
25	100	1.11	0.91	1.07	2.28	0.78	0.79
25	150	1.06	0.92	1.04	2.37	0.81	0.81
25	300	1.03	0.93	0.97	2.50	0.85	0.84
50	25	1.27	0.90	1.00	1.18	0.78	0.73
50	50	1.10	0.94	1.07	1.52	0.84	0.81
50	100	1.05	0.93	1.06	1.72	0.86	0.86
50	150	1.03	0.95	1.02	1.74	0.90	0.89
50	300	1.01	0.99	0.98	1.77	0.93	0.91
100	25	1.23	0.88	1.01	1.04	0.82	0.76
100	50	1.08	0.91	1.12	1.26	0.85	0.82
100	100	1.02	0.97	1.03	1.33	0.92	0.90
100	150	1.01	0.98	1.01	1.36	0.95	0.94
100	300	1.01	1.01	0.94	1.34	0.98	0.97
150	25	1.22	0.88	0.98	0.97	0.83	0.78
150	50	1.06	0.92	1.11	1.19	0.87	0.84
150	100	1.02	0.96	1.04	1.23	0.92	0.92
150	150	1.01	0.98	1.01	1.23	0.96	0.95
150	300	1.01	1.01	0.96	1.22	0.99	0.98
300	25	1.19	0.89	1.02	0.93	0.84	0.79
300	50	1.04	0.93	1.08	1.09	0.90	0.86
300	100	1.00	0.97	1.05	1.13	0.94	0.93
300	150	1.01	0.99	1.00	1.11	0.97	0.96
300	300	1.01	1.01	0.98	1.11	0.99	0.98

Appendix D. ML estimates obtained with the observable factor f_t

Table D.1: Mean ratio between $\hat{\theta}_i(f)$ and its true value θ_{0i}

T	$\hat{\beta}_0/\beta_0$	$\hat{\beta}_1/\beta_1$	$\hat{\varphi}/\varphi$	$\hat{\sigma^2}/\sigma^2$	$\hat{p_0}/p_0$	$\hat{p_1}/p_1$
25	1.13	0.84	1.04	0.82	0.84	0.80
50	1.03	0.92	1.08	0.96	0.91	0.87
100	1.00	0.97	1.04	1.00	0.96	0.94
150	1.00	0.98	1.03	1.00	0.98	0.96
300	1.00	1.00	1.00	1.00	1.00	0.99

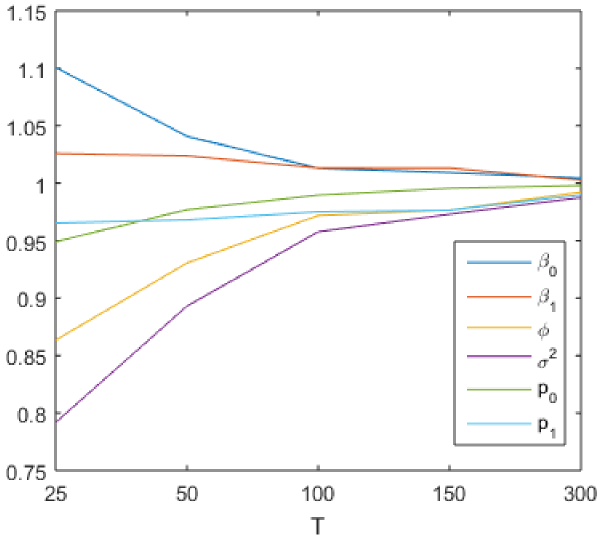


Figure D.1: Mean ratio between $\hat{\theta}_i(f)$ and its true value θ_{0i}

Appendix E. Properties of empirical distributions of t-statistics corresponding to the two-step estimates

Table E.1: Sample mean of the t-statistics corresponding to $\hat{\theta}(\hat{f})$

N	T	$t_{\hat{\beta}_0}$	$t_{\hat{\beta}_1}$	$t_{\hat{\varphi}}$	$t_{\hat{\sigma}^2}$	$t_{\hat{p}_0}$	$t_{\hat{p}_1}$
25	25	0.82	-0.35	-1.03	0.37	-0.24	0.06
25	50	0.66	-0.50	-0.96	1.27	-0.05	0.08
25	100	0.54	-0.57	-1.16	2.26	0.01	0.05
25	150	0.56	-0.63	-1.34	2.90	0.09	0.06
25	300	0.63	-0.75	-1.80	4.35	0.17	0.14
50	25	0.70	-0.29	-0.66	-0.01	-0.25	0.07
50	50	0.46	-0.37	-0.68	0.82	0.02	0.13
50	100	0.45	-0.48	-0.82	1.64	0.04	0.03
50	150	0.38	-0.47	-0.86	2.16	0.06	0.07
50	300	0.43	-0.59	-1.22	3.30	0.16	0.12
100	25	0.60	-0.25	-0.50	-0.35	-0.15	0.11
100	50	0.50	-0.32	-0.48	0.32	-0.06	0.04
100	100	0.34	-0.36	-0.54	0.93	0.07	0.06
100	150	0.28	-0.37	-0.55	1.32	0.04	0.01
100	300	0.31	-0.40	-0.75	2.12	0.10	0.00
150	25	0.63	-0.24	-0.47	-0.56	-0.15	0.10
150	50	0.34	-0.32	-0.37	0.11	-0.03	0.07
150	100	0.28	-0.31	-0.40	0.58	0.04	0.06
150	150	0.22	-0.28	-0.43	0.92	0.07	0.03
150	300	0.24	-0.35	-0.56	1.54	0.13	0.03
300	25	0.52	-0.31	-0.35	-0.75	-0.08	0.15
300	50	0.32	-0.29	-0.31	-0.27	0.03	0.06
300	100	0.19	-0.24	-0.25	0.18	0.02	-0.01
300	150	0.24	-0.19	-0.28	0.38	0.06	0.01
300	300	0.19	-0.23	-0.31	0.74	0.10	0.01

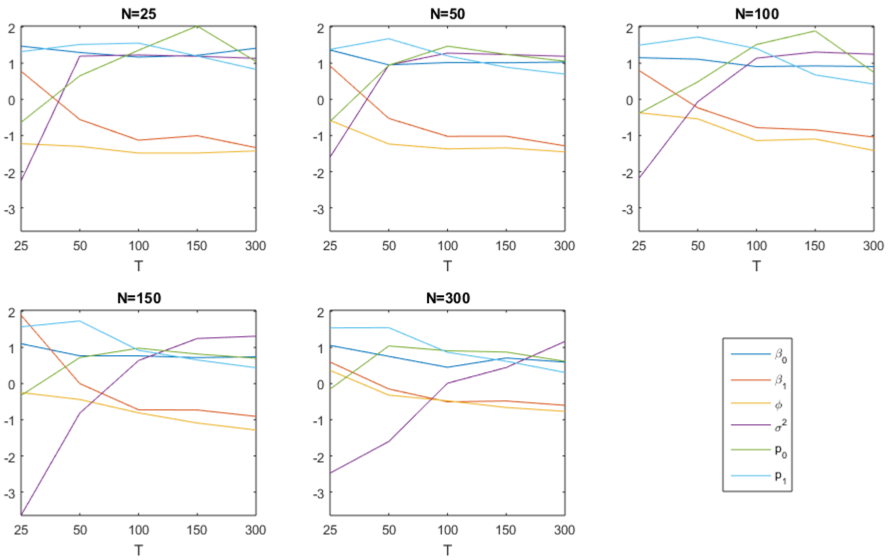


Figure E.1: Skewness of the t-statistics corresponding to the estimates $\hat{\theta}_i(\hat{f})$ at different N and T

Table E.2: Excess kurtosis of t-statistics of the estimates $\hat{\theta}_i(\hat{f})$

N	T	$t_{\hat{\beta}_0}$	$t_{\hat{\beta}_1}$	$t_{\hat{\varphi}}$	$t_{\hat{\sigma^2}}$	$t_{\hat{p_0}}$	$t_{\hat{p_1}}$
25	25	0.50	5.38	0.27	21.22	-0.64	1.79
25	50	0.29	0.51	0.37	-1.07	1.11	3.79
25	100	0.11	0.17	-0.11	-1.35	4.87	5.34
25	150	0.00	-0.11	-0.31	-1.45	13.38	3.62
25	300	0.13	-0.31	-0.62	-1.63	1.53*	0.82*
50	25	0.71	4.39	1.57	4.01	-0.75	1.79
50	50	0.20	0.63	0.05	-0.66	1.73	4.13
50	100	-0.09	0.15	-0.14	-1.14	6.55	3.49
50	150	-0.02	-0.01	-0.31	-1.31	4.57	2.46
50	300	0.17	0.13	-0.36	-1.44	2.36*	0.51*
100	25	-0.04	5.32	1.82	5.29	-0.75	2.30
100	50	0.13	1.06	1.04	0.35	1.01	6.28
100	100	0.20	0.18	0.16	-0.77	5.64	7.32
100	150	0.32	-0.12	-0.05	-0.97	18.36	1.77
100	300	0.16	0.02	-0.07	-1.27	0.94	0.53*
150	25	0.35	12.71	0.79	24.17	-0.84	2.50
150	50	0.23	2.57	1.07	1.86	1.29	5.65
150	100	0.29	0.10	0.52	-0.21	2.21	1.73
150	150	0.21	0.10	0.06	-0.70	1.05	0.70*
150	300	0.15	-0.04	0.03	-1.02	0.22	0.42*
300	25	0.18	4.12	3.41	6.04	-0.83	2.26
300	50	0.24	1.67	1.39	3.08	1.69	4.75
300	100	-0.02	0.19	0.45	0.25	2.92	2.82
300	150	0.01	0.05	-0.12	-0.08	1.67	0.67
300	300	0.11	0.01	0.62	-0.69	0.02	0.46*

Note: the symbol '*' marks the cases for which the null of normality of the Kolmogorov-Smirnov test is not rejected at 5% level of confidence probability.

Table E.3: Empirical size of two-tailed tests based on t-statistics, nominal size=5%

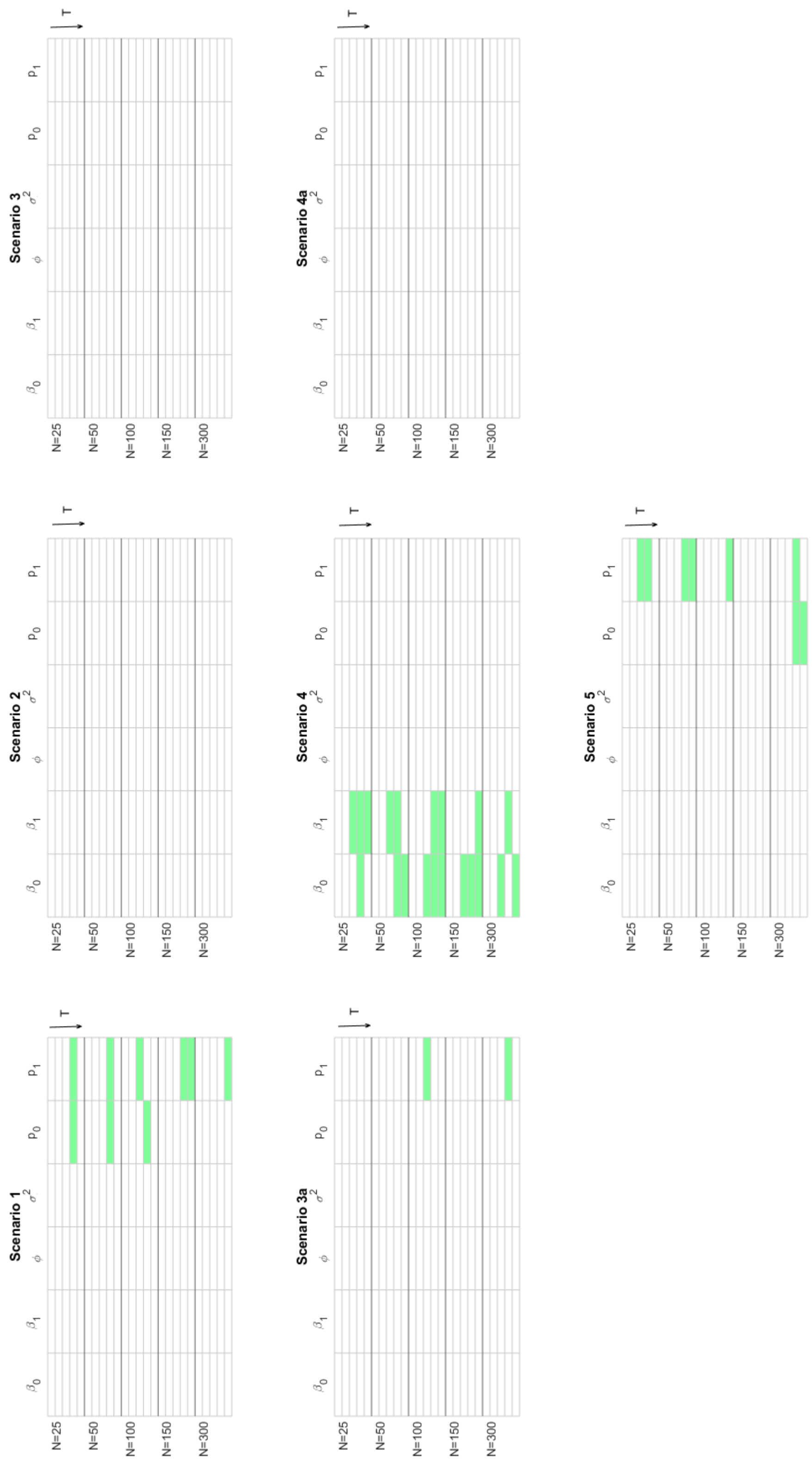
N	T	$t_{\hat{\beta}_0}$	$t_{\hat{\beta}_1}$	$t_{\hat{\varphi}}$	$t_{\hat{\sigma^2}}$	$t_{\hat{p_0}}$	$t_{\hat{p_1}}$
25	25	0.17	0.17	0.23	0.09	0.01	0.06
25	50	0.13	0.13	0.21	0.26	0.08	0.07
25	100	0.11	0.11	0.25	0.61	0.08	0.07
25	150	0.10	0.10	0.28	0.79	0.06	0.07
25	300	0.10	0.12	0.42	0.96	0.06	0.07
50	25	0.14	0.15	0.18	0.09	0.01	0.06
50	50	0.11	0.10	0.14	0.13	0.07	0.06
50	100	0.08	0.09	0.17	0.38	0.05	0.06
50	150	0.08	0.08	0.15	0.59	0.05	0.07
50	300	0.08	0.09	0.26	0.86	0.06	0.06
100	25	0.11	0.15	0.16	0.11	0.00	0.05
100	50	0.10	0.10	0.11	0.08	0.06	0.06
100	100	0.07	0.07	0.09	0.14	0.05	0.06
100	150	0.06	0.07	0.09	0.25	0.05	0.06
100	300	0.06	0.07	0.12	0.56	0.05	0.05
150	25	0.12	0.13	0.15	0.13	0.00	0.05
150	50	0.08	0.10	0.10	0.06	0.05	0.06
150	100	0.06	0.07	0.08	0.09	0.06	0.06
150	150	0.05	0.06	0.08	0.15	0.05	0.06
150	300	0.05	0.07	0.09	0.34	0.06	0.05
300	25	0.10	0.13	0.14	0.15	0.00	0.06
300	50	0.07	0.08	0.10	0.09	0.05	0.05
300	100	0.06	0.06	0.07	0.05	0.05	0.06
300	150	0.05	0.06	0.06	0.05	0.05	0.05
300	300	0.06	0.06	0.06	0.12	0.05	0.05

Table E.4: Empirical size of two-tailed tests based on t-statistics, nominal size=10%

N	T	$t_{\hat{\beta}_0}$	$t_{\hat{\beta}_1}$	$t_{\hat{\varphi}}$	$t_{\hat{\sigma^2}}$	$t_{\hat{p_0}}$	$t_{\hat{p_1}}$
25	25	0.24	0.24	0.33	0.19	0.04	0.08
25	50	0.20	0.21	0.31	0.38	0.13	0.11
25	100	0.18	0.18	0.35	0.72	0.12	0.12
25	150	0.17	0.18	0.38	0.86	0.11	0.12
25	300	0.16	0.18	0.52	0.97	0.10	0.11
50	25	0.20	0.24	0.26	0.15	0.03	0.09
50	50	0.18	0.16	0.21	0.24	0.10	0.10
50	100	0.13	0.15	0.24	0.52	0.10	0.10
50	150	0.14	0.13	0.24	0.70	0.10	0.11
50	300	0.13	0.17	0.36	0.91	0.11	0.11
100	25	0.20	0.23	0.23	0.16	0.03	0.07
100	50	0.17	0.17	0.18	0.14	0.09	0.09
100	100	0.13	0.13	0.15	0.26	0.09	0.10
100	150	0.11	0.14	0.17	0.38	0.10	0.11
100	300	0.11	0.13	0.19	0.69	0.10	0.09
150	25	0.20	0.21	0.23	0.17	0.03	0.08
150	50	0.14	0.16	0.16	0.11	0.08	0.09
150	100	0.11	0.13	0.13	0.17	0.11	0.11
150	150	0.10	0.12	0.14	0.25	0.10	0.10
150	300	0.10	0.13	0.16	0.47	0.11	0.10
300	25	0.18	0.19	0.22	0.20	0.02	0.08
300	50	0.13	0.15	0.16	0.14	0.09	0.09
300	100	0.11	0.12	0.12	0.10	0.09	0.10
300	150	0.10	0.11	0.13	0.11	0.09	0.09
300	300	0.11	0.12	0.11	0.20	0.10	0.09

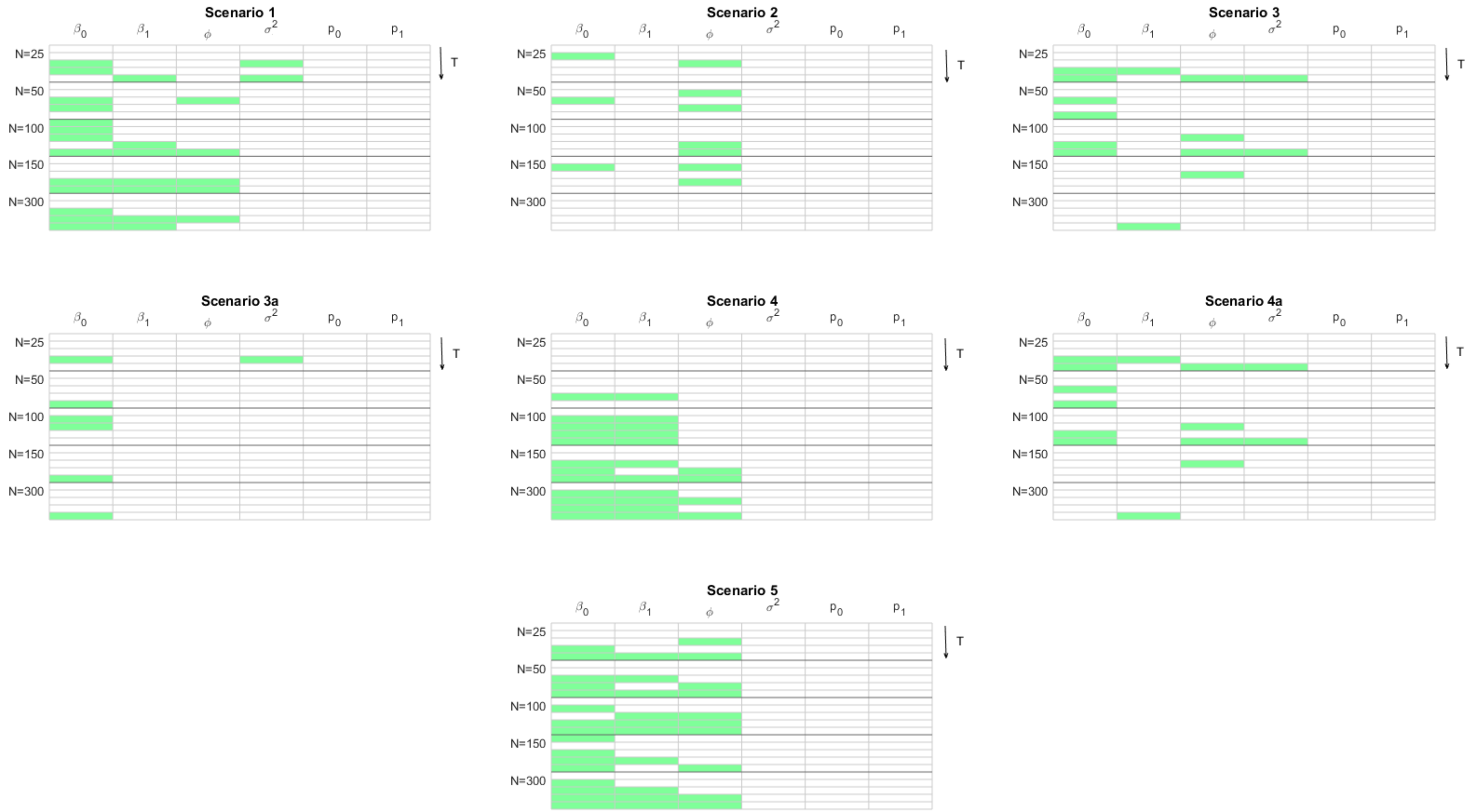
Appendix F. Simulations with various scenarios

Figure F.1: Results of the Kolmogorov-Smirnov normality test for the t-statistics of the two-step estimators, $\alpha = 5\%$ [a]



[a] Note: cases when the hypothesis is not rejected are marked in green.

Figure F.2: Results of the Jarque-Berra normality test for the t-statistics of the two-step estimators, $\alpha = 5\%$



a

[a]Note: cases when the hypothesis is not rejected are marked in green.