Consistency of 2-step estimators
of Markov Switching Dynamic Factor Models

Catherine DOZ

Aarhus University Seminar. December 2, 2016

1 Introduction

- Knowledge about the current state of the economy is crucial for economic agents, particularly policy makers.
- Different institutions $\rightarrow$ different and sometimes not replicable methodologies.
  Turning points are identified with delays which can be important.
- Markov-Switching (MS) Models : replicable method + provides timely estimation of the probability to enter a recession (expansion)
- Initial introduction of these models was made using 1 series (Hamilton 1998,1999). ML estimation, numerical procedure.
- Introduction of MS Dynamic Factor Models (MS-DFM) : Kim (1994), Kim-Yoo (1995), Kim-Nelson (1998) : allows to use several series. ML estimation using KF (= 1-step method). Usable only with very small sets of series and parameters (4 or 5 series).

- Camacho, Perez-Quiroz, Poncela (2012) propose a 2-steps estimation method :

  i) estimation of the unobserved MS factor(s) by PCA

  ii) estimation of the MS parameters

  Comparison of 1-step and 2-step estimation methods on US data. Claim that 1-step method works far better, although the performance of both methods gets closer when the number of indicators grows.
- Doz, Petronevich (2016) : use a large set of series to analyze French business cycle. Show that the 2-steps procedure give good results.
- Aim of the present paper : prove that the 2-steps method indeed provides consistent estimators.

2 MS-DFM and two estimation methods

2.1 The model

- Factor Model :
  - x_t is a vector of N observable variables
  - x_t can be decomposed as a linear function of two unobservable latent variables f_t (common factors) and e_t (idiosyncratic components) as :

$$x_t = \Lambda f_t + e_t \tag{1}$$

    with (f_t) and (e_t) two orthogonal processes.
    In what follows, f_t has dimension 1 (i.e. one common factor).
    Results can be extended to r-dimensional factors.

- MS-DFM : (f_t) is an AR process with MS mean :

$$f_t = \beta_{S_t} + \phi_1 f_{t-1} + \ldots \phi_p f_{t-p} + \varepsilon_t \text{ where } \varepsilon_t \sim N(0, \sigma^2_{S_t}) \qquad (2)$$

In what follows $p = 1$ or $p = 2$. Results can be extended to any p.

- S_t is a Markov process which takes a finite number of values (states). In what follows S_t has two states.

$$S_t = \begin{cases} 1 & \text{if recession} \\ 0 & \text{if expansion} \end{cases}$$

S_t switches sates with transition probabilities defined as follows (Markov property) :

$$Pr(S_t = j | S_{t-1} = i,\ S_{t-2} = k, ...) = Pr(S_t = j | S_{t-1} = i) = p_{ij},\ i, j \in \{0, 1\}$$

2.2 Estimation methods

- **1-step method** :
  - Estimate parameters and factor simultaneously by ML through the Kalman filter and a numerical optimization procedure
  - The KF formulas are modified due to the MS behavior of the factor (Kim's (1994) filter).
  - The filtered probabilities $P(S_t = i|I_t)$ are also computed as an output of the filter.
  - Advantages of the procedure :
    . the MS behavior of the factor is directly taken into account
    . it can very easily accommodate individual AR process for the idiosyncratic component.
  - Drawbacks of the procedure :
    . Numerical optimization $\rightarrow$ can be done only for a very small N
    . Time consuming procedure + possible convergence problems.

- **2-steps method** :
  - 1st step : Estimate factor model (1) by PCA
  - 2nd step : If $\hat{f}_t$ is the estimated factor obtained at step 1, estimate model (2) with $\hat{f}_t$ playing the role of f_t.
    The estimation is made by ML using a numerical procedure (but here small set of parameters) and the filtered probabilities can be computed via Hamilton (1998,1999) formulas
  - Advantages of the method :
    . It can accommodate a (very) large number of series N
    . Under standard assumptions PCA gives consistent estimators and $p\lim(\hat{f}_t - f_t) = 0$
    . It can easily accommodate various specifications for the MS behaviour of f_t
    . It is very easy to implement.

– Drawbacks of the method :

. Does not take the AR behaviour of (f_t) and (e_t) into account at the 1st step (it's always the case when DFM are estimated by PCA).

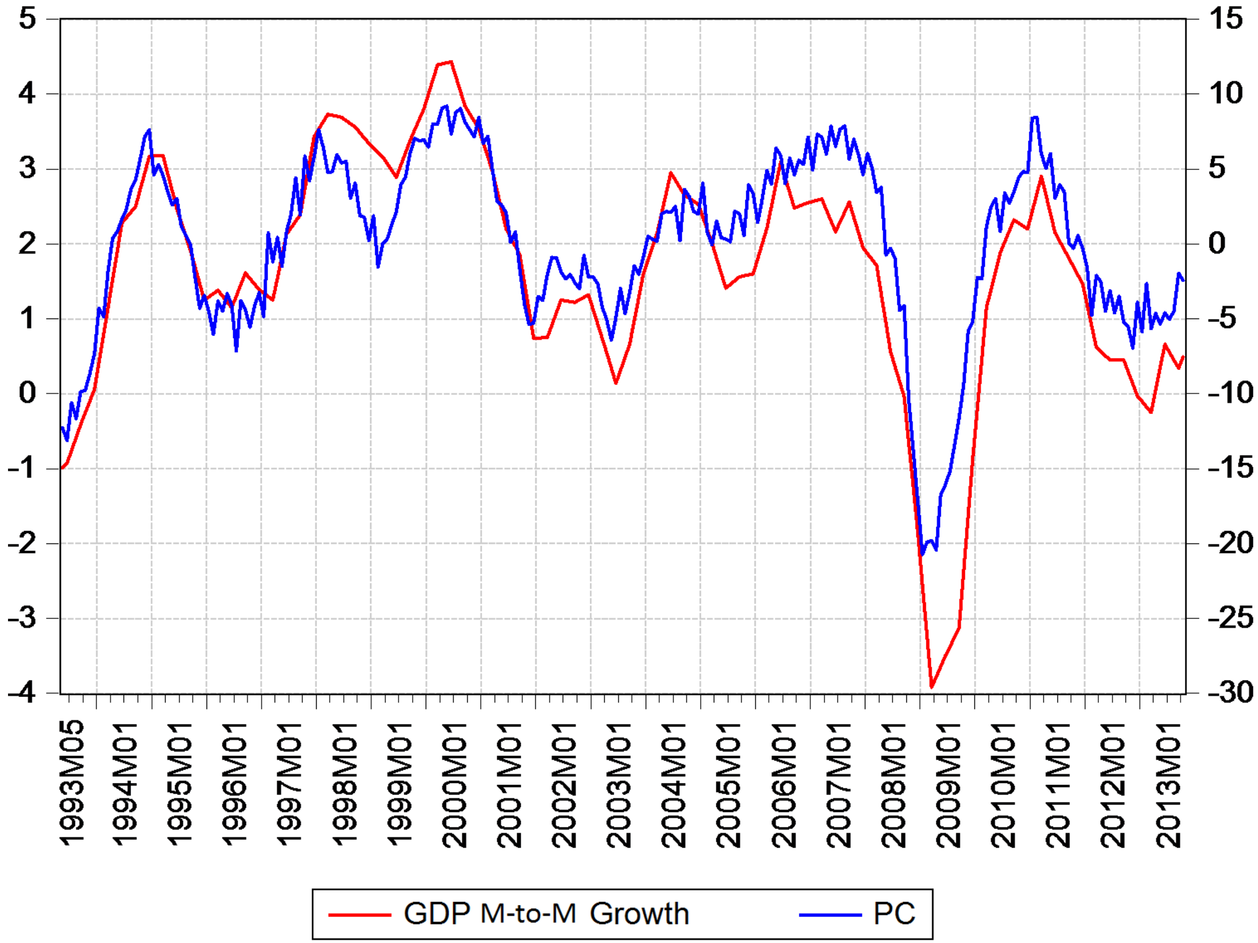
. Does not take the MS behaviour of (f_t) into account at the 1st step

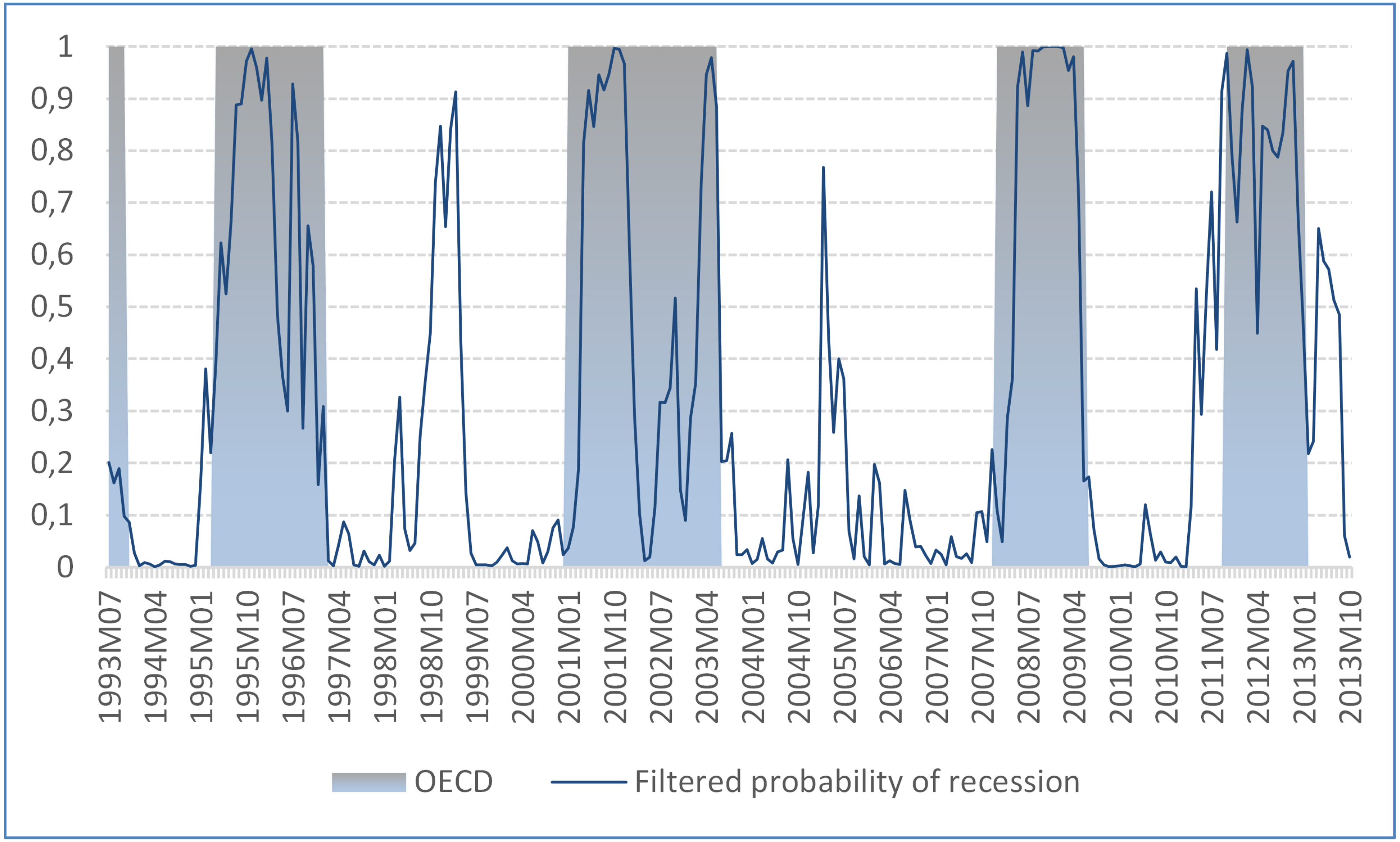
- Doz-Petronevich (2016) results :

. Data = 151 monthly series for French economy concerning : production sector, financial sector, employment, households, banking system, business surveys, international trade, monetary indicators, major world economic indicators and others.
The series are stationarized and then standardized.

. 2-step method, with several specifications for the MS-AR behaviour of (f_t).

. OECD business cycle dating as a reference.

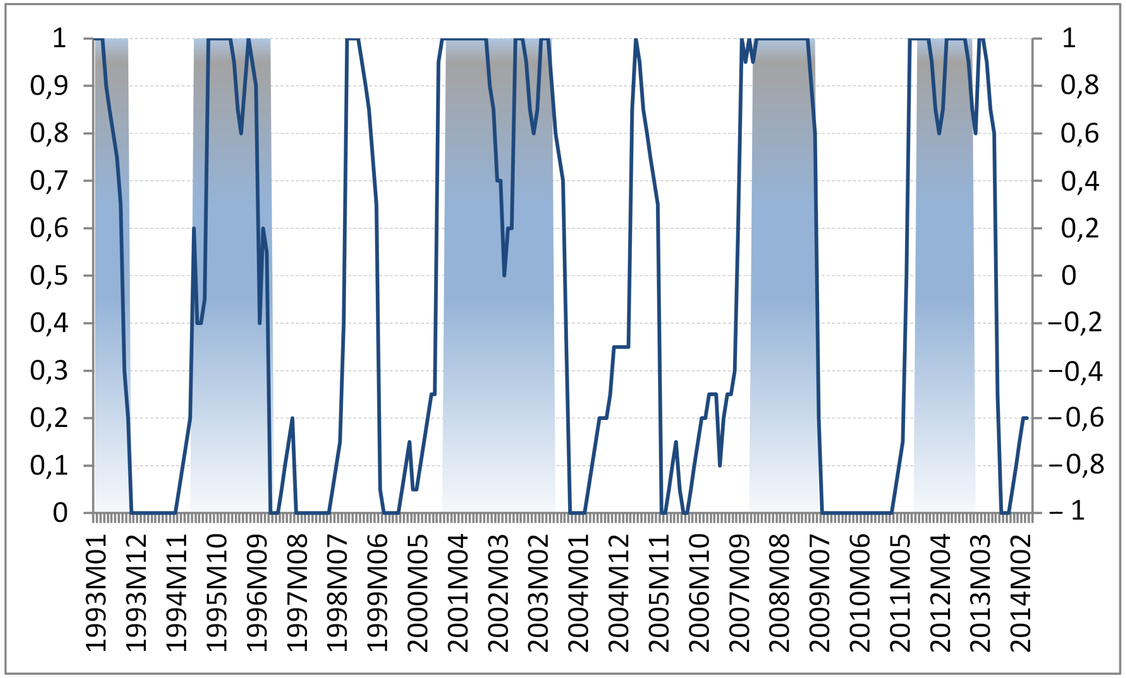

5
4
3
2
1
0
-1
-2
-3
-4
15
10
5
0
-5
-10
-15
-20
-25
-30
1993M05
1994M01
1995M01
1996M01
1997M01
1998M01
1999M01
2000M01
2001M01
2002M01
2003M01
2004M01
2005M01
2006M01
2007M01
2008M01
2009M01
2010M01
2011M01
2012M01
2013M01
GDP M-to-M Growth
PC



1
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
1993M07
1994M04
1995M01
1995M10
1996M07
1997M04
1998M01
1998M10
1999M07
2000M04
2001M01
2001M10
2002M07
2003M04
2004M01
2004M10
2005M07
2006M04
2007M01
2007M10
2008M07
2009M04
2010M01
2010M10
2011M07
2012M04
2013M01
2013M10
OECD
Filtered probability of recession

. Two slowdown periods (1998-1999 and 2005) are not detected by OECD due to the duration restrictions, but were indeed detected and analyzed by INSEE (see ”Note de Conjoncture”).

. Those periods were also detected by the reversal index (”indicateur de retournement”) published by INSEE.



3 Consistency of the 2-steps estimator

3.1 Consistency of ML estimators of HMM parameters when the variable is observable

- The model (AR(p) process with MS mean) is an Hidden Markov-switching Model (HMM) :

$$\begin{aligned} y_t &= \mu_{S_t} + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t \\ \varepsilon_t \quad &\text{i.i.d.} \quad \mathcal{N}(0, \sigma^2_{S_t}) \end{aligned}$$

  – y_t is observed
  – S_t is a two-state unobservable Markov variable such that :

$$P(S_t = 1 | S_{t-1} = 1) = p_1 \text{ and } P(S_t = 0 | S_{t-1} = 0) = p_0$$

- $p = 1$ below : identical proof for any p or any S_t with finite range.

- Conditional log-likelihood :

$$
\begin{aligned}
l(y_t|y_{t-1},\ldots,y_1,\theta) &= l(y_t|y_{t-1},\theta) \\
&= \frac{\pi_1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(y_t-\phi y_{t-1}-\mu_1)^2} + \frac{\pi_0}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(y_t-\phi y_{t-1}-\mu_0)^2}
\end{aligned}
$$

with : $\pi_1 = P(S_t = 1) = \frac{1-p_0}{2-(p_0+p_1)}$ and $\pi_0 = P(S_t = 0) = 1 - \pi_1 = \frac{1-p_1}{2-(p_0+p_1)}$
and $\theta = (\phi, \mu_0, \mu_1, \sigma_0^2, \sigma_1^2, p_0, p_1)'$

- Maximisation of the log-likelihood :

$$\hat{\theta} = \text{Argmax} Q_T(\theta) = \text{Argmax} \frac{1}{T} \sum_{t=1}^{T} \log l(y_t|y_{t-1},\ldots,y_1,\theta)$$

- No analytical solution : numerical optimization (e.g. EM-algorithm)
- General (complicated) proofs of MLE consistency for HMMs have been given in a very general framework (Leroux (1992), Cappe, Moulines, Rydden (2005) and others). Here : simpler proof for the particular case under study $\rightarrow$ extension to the 2-step estimator.

- In those mentioned papers, it is shown that $\hat{\theta}$ converges to θ^* or to a $\tilde{\theta}$ which is equivalent to θ^*, using the following definition :

$$\tilde{\theta} \sim \theta^* \Leftrightarrow P_{\tilde{\theta}} = P_{\theta^*}$$

It will be also the case here.
In practise, two parameters will be equivalent if they are equal up to an index permutation.

- Newey-Mac Fadden (1984) consistency conditions :
if $Q^*(\theta) = \mathrm{p}\lim Q_T(\theta)$ and if θ^* is the true value of θ, then the following conditions are sufficient to get : $\mathrm{p}\lim \hat{\theta} = \theta^*$
i) θ^* is the unique maximum of $Q^*(\theta)$
ii) $\theta \in \Theta$, with Θ a compact set containing θ^*
iii) Q^* is a continuous function
iv) $Q_T(\theta)$ converges in probability to $Q^*(\theta)$ uniformly on Θ

- In the present framework, (S_t) and (y_t) are stationary and

$$Q^*(\theta) = E_{\theta^*}\left(\log l(y_t|y_{t-1},\ldots,y_1,\theta)\right) = E_{\theta^*}\left(\log l(y_t|y_{t-1},\theta)\right)$$

- Sketch of the proof :

  ii) It can be easily shown that the maximization of $Q_T(\theta)$ can be done on a compact subset of the initial parameter space

  i) From the information inequality, it can be seen that θ^* is a maximum of Q^*.
  From the same inequality, it can be proved that any $\tilde{\theta}$ such that $Q^*(\tilde{\theta}) = Q^*(\theta^*)$ is equivalent to θ^*.
  Thus, any $\tilde{\theta}$ which is equivalent to θ is a potential limit for $\hat{\theta}$.

  iii) Can be obtained from Lebesgue continuity theorem.

  iv) It is sufficient to prove that (Q_T) is stochastically equicontinuous on Θ. This can be obtained by showing that :

$$\operatorname*{Sup}_{\theta\in\Theta} \|\nabla_\theta Q_T(\theta)\| = O_P(1)$$

3.2 Consistency of the 2-step estimator for an MS-DFM

- The model is now :

$$\begin{array}{rcl} x_t & = & \Lambda f_t + e_t \\ f_t & = & \mu_{S_t} + \phi_1 f_{t-1} + \cdots + \phi_p f_{t-p} + \varepsilon_t \\ \varepsilon_t & \text{i.i.d.} & \mathcal{N}(0, \sigma^2_{S_t}) \end{array}$$

  where
  - x_t is a $(n \times 1)$ vector of observed variables
  - f_t is a $(r \times 1)$ vector of unobserved factors
  - (f_t) and (e_t) are uncorrelated at all leads and lags
  - usual assumptions about Dynamic Factor Models (DFM) are satisfied

– S_t is a two-state unobservable Markov variable such that :

$$P(S_t = 1|S_{t-1} = 1) = p_1 \text{ and } P(S_t = 0|S_{t-1} = 0) = p_0$$

- Two-step estimator :

  1. Estimate the factor model (e.g. by PCA, but can be also done using QML) without taking the MS structure into account, and obtain an approximation $\hat{f}_t$ of f_t.
  2. Estimate an HMM for $\hat{f}_t$.

- Newey-Mac Fadden (1984) extension :

  using the same notations and conditions as before, if $\hat{\hat{\theta}}$ satisfies :

$$Q_T(\hat{\hat{\theta}}) \geq \underset{\theta \in \Theta}{\text{Max}} Q_T(\theta) + o_P(1)$$

  then :

$$\text{p}\lim \hat{\hat{\theta}} = \theta^*$$

- Application in our framework :
  - f_t plays the role which was played by y_t in the previous section
  - Q_T cannot be computed since f_t is not observable :

$$\begin{aligned} Q_T(\theta) &= \frac{1}{T}\sum_{t=1}^{T}\log l(f_t|f_{t-1},\ldots,f_1,\theta) \\ &= \frac{1}{T}\sum_{t=1}^{T}\log\left(\frac{\pi_1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{1}{2\sigma_1^2}(f_t-\phi f_{t-1}-\mu_1)^2}+\frac{\pi_0}{\sqrt{2\pi\sigma_0^2}}e^{-\frac{1}{2\sigma_0^2}(f_t-\phi f_{t-1}-\mu_0)^2}\right) \end{aligned}$$

  - Q_T is replaced by $\hat{Q}_T$:

$$\hat{Q}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\log\left(\frac{\pi_1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{1}{2\sigma_1^2}(\hat{f}_t-\phi \hat{f}_{t-1}-\mu_1)^2}+\frac{\pi_0}{\sqrt{2\pi\sigma_0^2}}e^{-\frac{1}{2\sigma_0^2}(\hat{f}_t-\phi \hat{f}_{t-1}-\mu_0)^2}\right)$$

  - $\hat{\hat{\theta}}$ is defined by $\hat{\hat{\theta}} = \text{Argmax}\hat{Q}_T(\theta)$.

– To get the consistency result, it is sufficient to prove that : $Q_T(\hat{\hat{\theta}}) \geq \underset{\theta\in\Theta}{\text{Max}} Q_T(\theta) + o_P(1)$.

– Remark : Here the consistency result will be analogous to the result obtained in section 1 :

$$\hat{\hat{\theta}} \text{ converges to a } \tilde{\theta} \text{ which is equivalent to } \theta^*.$$

- Sketch of the proof
  – $\forall\theta \in \Theta$, define

$$g(f_t, f_{t-1}, \theta) = \log\left(\sum_{i=0}^{1} \frac{\pi_i}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{1}{2\sigma_i^2}(f_t-\phi f_{t-1}-\mu_i)^2}\right)$$

so that

$$|Q_T(\theta) - \hat{Q}_T(\theta)| = |\frac{1}{T}\sum_{t=1}^{T}\Big(g(f_t, f_{t-1}, \theta) - g(\hat{f}_t, \hat{f}_{t-1}, \theta)\Big)|$$

– It can be shown that $\forall \theta \in \Theta$:

$$|\frac{\partial g}{\partial f_t}(f_t, f_{t-1}, \theta)| \leq M \text{ and } |\frac{\partial g}{\partial f_{t-1}}(f_t, f_{t-1}, \theta)| \leq M$$

– Thus :

$$\begin{aligned} |Q_T(\theta) - \hat{Q}_T(\theta)| &\leq \frac{1}{T}\sum_{t=1}^{T} |g(f_t, f_{t-1}, \theta) - g(\hat{f}_t, \hat{f}_{t-1}, \theta)| \\ &\leq M\frac{1}{T}\sum_{t=1}^{T} |f_t - \hat{f}_t|^2 = o_p(1) \text{ uniformly in } \theta \end{aligned}$$

– As $\hat{Q}_T(\hat{\hat{\theta}}) = \underset{\theta\in\Theta}{\text{Max}}\hat{Q}_T(\theta)$, we have :

$$\hat{Q}_T(\hat{\hat{\theta}}) \geq \hat{Q}_T(\hat{\theta})$$

where $\hat{\theta} = \text{Argmax } Q_T(\theta)$.

– This can be written as :

$$Q_T(\hat{\hat{\theta}}) + o_P(1) \geq Q_T(\hat{\theta}) + o_P(1)$$

which completes the proof.