



**HAL**  
open science

## Estimation of linear autoregressive models with Markov-switching, the E.M. algorithm revisited.

Joseph Rynkiewicz

► **To cite this version:**

Joseph Rynkiewicz. Estimation of linear autoregressive models with Markov-switching, the E.M. algorithm revisited.. INVESTIGACIÓN OPERACIONAL, 2004, 25 (2), pp.166-173. hal-00258137

**HAL Id: hal-00258137**

**<https://hal.science/hal-00258137>**

Submitted on 21 Feb 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Estimation of linear autoregressive models with Markov-switching, the E.M. algorithm revisited.

Joseph Rynkiewicz\*

February 21, 2008

## Abstract

This work concerns estimation of linear autoregressive models with Markov-switching using expectation maximisation (E.M.) algorithm. Our method generalise the method introduced by Elliot for general hidden Markov models and avoid to use backward recursion.

Keywords : Maximum likelihood estimation, Expectation-Maximisation algorithm, Hidden Markov models, Switching models.

## 1 Introduction

In the present paper we consider an extension of basic (HMM). Let  $(X_t, Y_t)_{t \in \mathbb{Z}}$  be the process such that

1.  $(X_t)_{t \in \mathbb{Z}}$  is a Markov chain in a finite state space  $\mathbb{E} = \{e_1, \dots, e_N\}$ , which can be identified without loss of generality with the simplex of  $\mathbb{R}^N$ , where  $e_i$  are unit vector in  $\mathbb{R}^N$ , with unity as the  $i$ th element and zeros elsewhere.
2. Given  $(X_t)_{t \in \mathbb{Z}}$ , the process  $(Y_t)_{t \in \mathbb{Z}}$  is a sequence of linear autoregressive model in  $\mathbb{R}$  and the distribution of  $Y_n$  depends only of  $X_n$  and  $Y_{n-1}, \dots, Y_{n-p}$ .

Hence, for a fixed  $t$ , the dynamic of the model is :

$Y_{t+1} = F_{X_{t+1}}(Y_{t-p+1}^t) + \sigma_{X_{t+1}} \varepsilon_{t+1}$  with  $F_{X_{t+1}} \in \{F_{e_1}, \dots, F_{e_N}\}$  linear functions,  $\sigma_{X_{t+1}} \in \{\sigma_{e_1}, \dots, \sigma_{e_N}\}$  strictly positive numbers and  $(\varepsilon_t)_{t \in \mathbb{N}^*}$  a i.i.d sequence of Gaussian random variable  $\mathcal{N}(0, 1)$ .

---

\*SAMOS/MATISSE, University of ParisI, 90 rue de Tolbiac, Paris, France, rynkiewi@univ-paris1.fr

**Definition 1** Write  $\mathcal{F}_t = \sigma \{X_0, \dots, X_t\}$ , for the  $\sigma$ -field generated by  $X_0, \dots, X_t$ ,  $\mathcal{Y}_t = \sigma \{Y_0, \dots, Y_t\}$ , for the  $\sigma$ -field generated by  $Y_0, \dots, Y_t$  and  $\mathcal{G}_t = \sigma \{(X_0, Y_0), \dots, (X_t, Y_t)\}$ , for the  $\sigma$ -field generated by  $X_0, \dots, X_t$  and  $Y_0, \dots, Y_t$ .

The Markov property implies here that  $P(X_{t+1} = e_i | \mathcal{F}_t) = P(X_{t+1} = e_i | X_t)$ . Write  $a_{ij} = P(X_{t+1} = e_i | X_t = e_j)$  and  $A = (a_{ij}) \in \mathbb{R}^{N \times N}$  and define :  $V_{t+1} := X_{t+1} - E[X_{t+1} | \mathcal{F}_t] = X_{t+1} - AX_t$ . With the previous notations, we obtain the general equation of the model, for  $t \in \mathbb{N}$  :

$$\begin{cases} X_{t+1} = AX_t + V_{t+1} \\ Y_{t+1} = F_{X_{t+1}}(Y_{t-p+1}^t) + \sigma_{X_{t+1}} \varepsilon_{t+1} \end{cases} \quad (1)$$

The parameters of the model are the transition probabilities of the matrix  $A$ , the coefficients of the linear functions  $F_{e_i}$  and the variances  $\sigma_{e_i}$ . A successful method for estimating such model is to compute the maximum likelihood estimator<sup>1</sup> with the E.M. algorithm introduced by Demster, Lair and Rubin (1977). Generally, this algorithm demands the calculus of the conditional expectation of the hidden states knowing the observations (the E.-step), this can be done with the Baum and Welch forward-backward algorithm (see Baum et al. (1970)). The derivation of the M-step of the E.M. algorithm is then immediate since we can compute the optimal parameters of the regression functions thanks weighted linear regression.

However we show here that we can also embed these two steps in only one. Namely we can compute, for each step of the E.M. algorithm, directly the optimal coefficients of the regression functions as the variances and the transition matrix thanks a generalisation of the method introduced by Elliott (1994).

## 2 Change of measure

The fundamental technique employed throughout this paper is the discrete time change of measure. Write  $\sigma$  the vector  $(\sigma_{e_1}, \dots, \sigma_{e_N})$ ,  $\phi(\cdot)$  for the density of  $\mathcal{N}(0, 1)$  and  $\langle \cdot, \cdot \rangle$  the inner product in  $\mathbb{R}^N$ .

We wish to introduce a new probability measure  $\bar{P}$ , using a density  $\Lambda$ , so that  $\frac{d\bar{P}}{dP} = \Lambda$  and under  $\bar{P}$  the random variables  $y_t$  are  $\mathcal{N}(0, 1)$  i.i.d. random variables.

---

<sup>1</sup>This likelihood is computed conditionally to the first “p” observations.

Define

$$\lambda_l = \frac{\langle \sigma, X_{l-1} \rangle \phi(y_l)}{\phi(\varepsilon_l)}, l \in \mathbb{N}^*, \text{ with } \Lambda_0 = 1 \text{ and } \Lambda_t = \prod_{l=1}^t \lambda_l$$

and construct a new probability measure  $\bar{P}$  by setting the restriction of the Radon-Nikodym derivative to  $\mathcal{G}_t$  equal to  $\Lambda_t$ . Then the following lemma is a straightforward adaptation of lemma 4.1 of Elliot (1994) (see annexe).

**Lemma 1** *Under  $\bar{P}$  the  $Y_t$  are  $\mathcal{N}(0, 1)$  i.i.d. random variables.*

Conversely, suppose we start with a probability measure  $\bar{P}$  such that under  $\bar{P}$

1.  $(X_t)_{t \in \mathbb{N}}$  is a Markov chain with transition matrix  $A$ .
2.  $(Y_t)_{t \in \mathbb{N}}$  is a sequence of  $\mathcal{N}(0, 1)$  i.i.d. random variable.

We construct a new probability measure  $P$  such that under  $P$  we have  $Y_{t+1} = F_{X_t}(Y_{t-p}^t) + \sigma_{X_t} \varepsilon_{t+1}$ . To construct  $P$  from  $\bar{P}$ , we introduce  $\bar{\lambda}_l := (\lambda_l)^{-1}$  and  $\bar{\Lambda}_t := (\Lambda_t)^{-1}$  and we define  $P$  by putting  $(\frac{dP}{d\bar{P}}) |_{\mathcal{G}_t} = \bar{\Lambda}_t$ ,

**Definition 2** *let  $(H_t)$ ,  $t \in \mathbb{N}$  be a sequence adapted to  $(\mathcal{G}_t)$ , We shall write :*

$$\gamma_t(H_t) = \bar{E} [\bar{\Lambda}_t H_t | \mathcal{Y}_t] \text{ and } \Gamma^i(Y_{t+1}) = \frac{\phi\left(\frac{Y_{t+1} - F_{X_t}(Y_{t-p+1}^t)}{\langle \sigma, e_i \rangle}\right)}{\langle \sigma, e_i \rangle \phi(Y_{t+1})}.$$

The proof of the following theorem is a detailed adaption of the proof of theorem 5.3 of Elliott (1994) (see annexe).

**Theorem 1** *Suppose  $H_t$  is a scalar  $\mathcal{G}$ -adapted process of the form :  $H_0$  is  $\mathcal{F}_0$  measurable,  $H_{t+1} = H_t + \alpha_{t+1} + \langle \beta_{t+1}, V_{t+1} \rangle + \delta_{t+1} f(Y_{t+1})$ ,  $k \geq 0$ , where  $V_{t+1} = X_{t+1} - AX_t$ ,  $f$  is a scalar valued function and  $\alpha, \beta, \delta$  are  $\mathcal{G}$  predictable process ( $\beta$  will be  $N$ -dimensional vector process). Then :*

$$\begin{aligned} \gamma_{t+1}(H_{t+1} X_{t+1}) &:= \gamma_{t+1, t+1}(H_{t+1}) \\ &= \sum_{i=1}^N \{ \langle \gamma_t(H_t X_t), \Gamma^i(y_{t+1}) \rangle a_i \\ &\quad + \gamma_t(\alpha_{t+1} \langle X_t, \Gamma^i(y_{t+1}) \rangle) a_i \\ &\quad + \gamma_t(\delta_{t+1} \langle X_t, \Gamma^i(y_{t+1}) \rangle) f(y_{t+1}) a_i \\ &\quad + (\text{diag}(a_i) - a_i a_i^T) \gamma_t(\beta_{t+1} \langle X_t, \Gamma^i(y_{t+1}) \rangle) \end{aligned} \quad (2)$$

where  $a_i := Ae_i$ ,  $a_i^T$  is the transpose of  $a_i$  and  $\text{diag}(a_i)$  is the matrix with vector  $a_i$  for diagonal and zeros elsewhere.

We will now consider special cases of processes H. In all cases, we will calculate the quantity  $\gamma_{t,t}(H_t)$  and deduce  $\gamma_t(H_t)$  by summing the components of  $\gamma_{t,t}(H_t)$ . Then, we deduce from the conditional Bayes' theorem the conditional expectation of  $H_t$  :

$$\hat{H}_t := E[H_t | \mathcal{Y}_t] = \frac{\gamma_t(H_t)}{\gamma_t(1)}.$$

### 3 Application to the Expectation (E.-step) of the E.M. algorithm

We will use the previous theorem in order to compute conditional quantities needed by the E.M. algorithm.

Let  $\mathcal{J}_t^{rs} = \sum_{l=1}^t \langle X_{l-1}, e_r \rangle \langle X_l, e_s \rangle$  be the number of jump from state  $e_r$  to state  $e_s$  at time  $t$ , we obtain :

$$\gamma_{t+1,t+1}(\mathcal{J}_{t+1}^{rs}) = \sum_{i=1}^N \langle \gamma_{t,t}(\mathcal{J}_t^{rs}), \Gamma^i(Y_{t+1}) \rangle a_i + \langle \gamma_t(X_t), \Gamma^r(Y_{t+1}) \rangle a_{sr} e_s. \quad (3)$$

Write now  $\mathcal{O}_t^r = \sum_{n=1}^{t+1} \langle X_n, e_r \rangle$  for the number of times, up to  $t$ , that  $X$  occupies the state  $e_r$ . We obtain

$$\gamma_{t+1,t+1}(\mathcal{O}_{t+1}^r) = \sum_{i=1}^N \langle \gamma_{t,t}(\mathcal{O}_t^r), \Gamma^i(Y_{t+1}) \rangle a_i + \langle \gamma_t(X_t), \Gamma^r(Y_{t+1}) \rangle a_r. \quad (4)$$

For the regression functions, the M-Step of the E.M. algorithm is achieved by finding the parameters minimising the weighted sum of squares :

$$\sum_{t=1}^n \gamma_i(t) \left( y_t - (a_0^i + a_1 y_{t-1} + \dots + a_p y_{t-p})^2 \right)$$

where  $\gamma_i(t)$  is the conditional expectation of the hidden  $e_i$  at time  $t$  knowing the observations  $y_{-p+1}, \dots, y_n$ .

Write  $\psi^T(t) = (1, y_{t-1}, \dots, y_{t-p})$  and  $\theta_i = (a_0^i, \dots, a_p^i)$ , suppose that the matrix  $[\sum_{t=1}^n \gamma_i(t) \psi(t) \psi^T(t)]$  is invertible. The estimator  $\hat{\theta}_i(n)$  of  $\theta_i$  is given by :

$$\hat{\theta}_i(n) = \left[ \sum_{t=1}^n \gamma_i(t) \psi(t) \psi^T(t) \right]^{-1} \sum_{t=1}^n \gamma_i(t) \psi(t) Y_t.$$

Hence, in order to compute  $\hat{\theta}_i(n)$ , we need to estimate the conditional expectation of the following processes :

1.

$$\mathcal{TA}_{t+1}^r(j) = \sum_{l=1}^{t+1} \langle X_l, e_r \rangle Y_{l-j} Y_{l+1}$$

for  $-1 \leq j \leq p$  and  $1 \leq r \leq N$ .

2.

$$\mathcal{TB}_{t+1}^r(i, j) = \sum_{l=1}^{t+1} \langle X_l, e_r \rangle Y_{l-j} Y_{l-i}$$

for  $0 \leq j, i \leq p$  and  $1 \leq r \leq N$ .

3.

$$\mathcal{TC}_{t+1}^r = \sum_{l=1}^{t+1} \langle X_l, e_r \rangle Y_{l+1}.$$

4.

$$\mathcal{TD}_{t+1}^r(j) = \sum_{l=1}^{t+1} \langle X_l, e_r \rangle Y_{l-j}$$

for  $0 \leq j \leq p$  and  $1 \leq r \leq N$ .

Applying theorem (2) with  $H_{t+1}(j) = \mathcal{TA}_{t+1}^r(j)$ ,  $H_0 = 0$ ,  $\alpha_{t+1} = 0$ ,  $\beta_{t+1} = 0$ ,  $\delta_{t+1} = \langle X_t, e_r \rangle Y_{t-j}$  and  $f(Y_{t+1}) = Y_{t+1}$ , if  $j \neq -1$  or  $\delta_{t+1} = \langle X_t, e_r \rangle$  and  $f(Y_{t+1}) = Y_{t+1}^2$  if  $j = -1$ , gives us

$$\gamma_{t+1, t+1}(\mathcal{TA}_{t+1}^r(j)) = \sum_{i=1}^N \frac{\langle \gamma_{t, t}(\mathcal{TA}_t^r(j)), \Gamma^i(Y_{t+1}) \rangle a_i}{+ \langle \gamma_t(X_t), \Gamma^r(Y_{t+1}) \rangle Y_{t-j} Y_{t+1} a_r}, \quad (5)$$

where  $a_r$  is the  $r$ -th column of  $A$ .

Then, applying theorem (2) with

$H_{t+1}(j) = \mathcal{TB}_{t+1}^r(i, j)$ ,  $H_0 = 0$ ,  $\alpha_{t+1} = 0$ ,  $\beta_{t+1} = 0$ ,  $\delta_{t+1} = \langle X_t, e_r \rangle Y_{t-j} Y_{t-i}$  and  $f(Y_{t+1}) = 1$  gives :

$$\gamma_{t+1, t+1}(\mathcal{TB}_{t+1}^r(i, j)) = \sum_{i=1}^N \frac{\langle \gamma_{t, t}(\mathcal{TB}_t^r(j)), \Gamma^i(Y_{t+1}) \rangle a_i}{+ \langle \gamma_t(X_t), \Gamma^r(Y_{t+1}) \rangle Y_{t-j} Y_{t-i} a_r}. \quad (6)$$

Next, applying theorem (2) with

$H_{t+1} = \mathcal{TC}_{t+1}^r$ ,  $H_0 = 0$ ,  $\alpha_{t+1} = 0$ ,  $\beta_{t+1} = 0$ ,  $\delta_{t+1} = \langle X_t, e_r \rangle$  and  $f(Y_{t+1}) = Y_{t+1}$  gives :

$$\gamma_{t+1,t+1}(\mathcal{TC}_{t+1}^r) = \sum_{i=1}^N \frac{\langle \gamma_{t,t}(\mathcal{TC}_t^r(j)), \Gamma^i(Y_{t+1}) \rangle a_i}{\langle \gamma_t(X_t), \Gamma^r(Y_{t+1}) \rangle Y_{t+1} a_r}. \quad (7)$$

Finally, applying theorem (2) with

$H_{t+1}(j) = \mathcal{TD}_{t+1}^r(j)$ ,  $H_0 = 0$ ,  $\alpha_{t+1} = 0$ ,  $\beta_{t+1} = 0$ ,  $\delta_{t+1} = \langle X_t, e_r \rangle Y_{t-j}$  and  $f(Y_{t+1}) = 1$  gives :

$$\gamma_{t+1,t+1}(\mathcal{TD}_{t+1}^r(j)) = \sum_{i=1}^N \frac{\langle \gamma_{t,t}(\mathcal{TD}_t^r(j)), \Gamma^i(Y_{t+1}) \rangle a_i}{\langle \gamma_t(X_t), \Gamma^r(Y_{t+1}) \rangle Y_{t-j} a_r}. \quad (8)$$

The ‘‘Maximisation’’ pass of the E.M. algorithm is now achieved by updating the parameters in the following way.

**Parameters of the transition matrix** The parameter of the transition matrix will be updates with the formula :

$$\hat{a}_{sr} = \frac{\gamma_T(\mathcal{J}_T^{sr})}{\gamma_T(\mathcal{O}_T^r)}. \quad (9)$$

**Parameters of the regression functions** For  $1 \leq r \leq N$ , let

$R^r := (R_{ij}^r)_{1 \leq i,j \leq p+1}$  be the symmetric with

$R_{11}^r = 1, R_{1j}^r = R_{j1}^r = \hat{\mathcal{T}}\mathcal{D}^r(j)$ ,  $R_{ij}^r = \hat{\mathcal{T}}\mathcal{B}^r(i-1, j-1)$  and

$C^r = (\hat{\mathcal{T}}\mathcal{C}^r, (\hat{\mathcal{T}}\mathcal{A}^r(i))_{0 \leq i \leq p})$  we can then compute the updated parameter  $\hat{\theta}_r$  of the regression function  $F_{e_r}$  with the formula :

$$\hat{\theta}_r = (R^r)^{-1} C^r \quad (10)$$

**Parameters of the variances** Finally, thanks the previous conditional expectations, we can directly calculate the parameters  $\hat{\sigma}_1, \dots, \hat{\sigma}_N$ , since for  $1 \leq r \leq N$  the conditional expectation of the mean square error of the rth model is

$$\hat{\sigma}_r^2 = \frac{1}{\mathcal{O}_r} \left( \hat{\mathcal{T}}\mathcal{A}^r(-1) + \hat{\theta}_r^T R^r \hat{\theta}_r - 2\hat{\theta}_r^T C^r \right). \quad (11)$$

This complete the M-step of the E.M. algorithm.

## 4 conclusion

Using the discrete Girsanov measure transform, we propose an new way to apply the E.M. algorithm in the case of Markov-switching linear autoregressions.

Note that, contrary to the Baum and Welch algorithm, we don't use backward recurrence, although the cost of calculus slightly increase since the number of operations is multiplied by  $\frac{N}{2}$ , where  $N$  is the number of hidden state of the Markov chain.

## References

Baum, L.E., Petrie, T., Soules, G. and Weiss N. A maximisation technique occurring in the statistical estimation of probabilistic functions of Markov processes. *Annals of Mathematical statistics*, 41:1:164-171, 1970

Demster, A.P., Lair N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the E.M. algorithm. *Journal of the Royal statistical society of London*, Series B:39:1-38, 1966.

Elliott,R.J. (1994) Exact Adaptative Filters for Markov chains observed in Gaussian Noise *Automatica* 30:9:1399-1408, 1994.

## Annexe

### Proof of lemma 1

**Lemma 2** Under  $\bar{P}$  the  $Y_t$  are  $\mathcal{N}(0, 1)$  i.i.d. random variables.

**Proof** The proof is based on the conditionnal Bayes' Theorem, it is a simple rewriting of the Proof of Elliot , hence we have

$$\bar{P}(Y_{t+1} \leq \tau | \mathcal{G}_t) = \bar{E} [1_{\{Y_{t+1} \leq \tau\}} | \mathcal{G}_t]$$

Thanks the conditionnal Bayes' Theorem we have :

$$\begin{aligned} & \bar{E} [1_{\{Y_{t+1} \leq \tau\}} | \mathcal{G}_t] \\ &= \frac{E [\Lambda_{t+1} 1_{\{Y_{t+1} \leq \tau\}} | \mathcal{G}_t]}{E [\Lambda_{t+1} | \mathcal{G}_t]} \end{aligned}$$



$$= \frac{\Lambda_t}{\Lambda_t} \times \frac{E[\lambda_{t+1} 1_{\{Y_{t+1} \leq \tau\}} | \mathcal{G}_t]}{E[\lambda_{t+1} | \mathcal{G}_t]}.$$

Now

$$\begin{aligned} E[\lambda_{t+1} | \mathcal{G}_t] &= \int_{-\infty}^{\infty} \frac{\langle \sigma, X_t \rangle \phi(Y_{t+1})}{\phi(\varepsilon_{t+1})} \times \phi(\varepsilon_{t+1}) d\varepsilon_{t+1} \\ &= \int_{-\infty}^{\infty} \langle \sigma, X_t \rangle \phi(F_{X_t}(Y_{t-p+1}^t) + \langle \sigma, X_t \rangle \times \varepsilon_{t+1}) d\varepsilon_{t+1} = 1 \end{aligned}$$

and since  $\varepsilon_{t+1} = \frac{Y_{t+1} - F_{X_t}(Y_{t-p+1}^t)}{\langle \sigma, X_t \rangle}$  :

$$\begin{aligned} \bar{P}(Y_{t+1} \leq \tau | \mathcal{G}_t) &= E[\lambda_{t+1} 1_{\{Y_{t+1} \leq \tau\}} | \mathcal{G}_t] \\ &= \int_{-\infty}^{\infty} \frac{\langle \sigma, X_t \rangle \phi(Y_{t+1})}{\phi(\varepsilon_{t+1})} \times 1_{\{Y_{t+1} \leq \tau\}} \times \phi(\varepsilon_{t+1}) d\varepsilon_{t+1} \\ &= \int_{-\infty}^{\tau} \phi(Y_{t+1}) dy_{t+1} = \bar{P}(Y_{t+1} \leq \tau) \end{aligned}$$

■

## Proof of Theorem 2

**Theorem 2** Suppose  $H_t$  is a scalar  $\mathcal{G}$ -adapted process of the form :  $H_0$  is  $\mathcal{F}_t$  measurable,  $H_{t+1} = H_t + \alpha_{t+1} + \langle \beta_{t+1}, V_{t+1} \rangle + \delta_{t+1} f(Y_{t+1})$ ,  $k \geq 0$ , where  $V_{t+1} = X_{t+1} - AX_t$ ,  $f$  is a scalar valued function and  $\alpha, \beta, \delta$  are  $\mathcal{G}$  predictable process ( $\beta$  will be  $N$ -dimensional vector process). Then :

$$\begin{aligned} \gamma_{t+1}(H_{t+1} X_{t+1}) &:= \gamma_{t+1,t+1}(H_{t+1}) \\ &= \sum_{i=1}^N \{ \langle \gamma_t(H_t X_t), \Gamma^i(y_{t+1}) \rangle a_i \\ &\quad + \gamma_t(\alpha_{t+1} \langle X_t, \Gamma^i(y_{t+1}) \rangle) a_i \\ &\quad + \gamma_t(\delta_{t+1} \langle X_t, \Gamma^i(y_{t+1}) \rangle) f(y_{t+1}) a_i \\ &\quad + (\text{diag}(a_i) - a_i a_i^T) \gamma_t(\beta_{t+1} \langle X_t, \Gamma^i(y_{t+1}) \rangle) \} \end{aligned} \quad (12)$$

where  $a_i := Ae_i$ ,  $a_i^T$  is the transpose of  $a_i$  and  $\text{diag}(a_i)$  is the matrix with vector  $a_i$  for diagonal and zeros elsewhere.

**Proof** Here again it is only a rewriting of the proof of Elliot.

We begin with the two following results :

### Result 1

$$\begin{aligned} \bar{E}[V_{t+1} | \mathcal{Y}_{t+1}] &= \bar{E}[\bar{E}[V_{t+1} | \mathcal{G}_t, \mathcal{Y}_{t+1}] | \mathcal{Y}_{t+1}] \\ &= \bar{E}[\bar{E}[V_{t+1} | \mathcal{G}_t] | \mathcal{Y}_{t+1}] = 0. \end{aligned} \quad (13)$$

**Result 2**

$$X_{t+1}X_{t+1}^T = AX_t(AX_t)^T + AX_tV_{t+1}^T + V_{t+1}(AX_t)^T + V_{t+1}V_{t+1}^T.$$

Since  $X_t$  is of the form  $(0, \dots, 0, 1, 0, \dots, 0)$  we have

$$X_{t+1}X_{t+1}^T = \text{diag}(X_{t+1}) = \text{diag}(AX_t) + \text{diag}(V_{t+1})$$

so

$$V_{t+1}V_{t+1}^T = \text{diag}(AX_t) + \text{diag}(V_{t+1}) - A \text{diag}(X_t) A^T - AX_tV_{t+1}^T - V_{t+1}(AX_t)^T.$$

Finally we obtain the result

$$\begin{aligned} \langle V_{t+1} \rangle &:= E[V_{t+1}V_{t+1}^T | \mathcal{F}_t] \\ &= E[V_{t+1}V_{t+1}^T | X_t] \\ &= \text{diag}(AX_t) - A \text{diag}(X_t) A^T. \end{aligned} \tag{14}$$

**Main proff** We have

$$\begin{aligned} \gamma_{t+1,t+1}(H_{t+1}) &= \bar{E} [\bar{\Lambda}_{t+1}H_{t+1}X_{t+1} | \mathcal{Y}_{t+1}] \\ &= \bar{E} [(AX_t + V_{t+1})(H_t + \alpha_{t+1} + \langle \beta_{t+1}, V_{t+1} \rangle + \delta_{t+1}f(y_{t+1})) \times \bar{\Lambda}_{t+1} | \mathcal{Y}_{t+1}] \end{aligned}$$

Thanks equation (13),

$$\gamma_{t+1,t+1}(H_{t+1}) = \bar{E} [((H_t + \alpha_{t+1} + \delta_{t+1}f(y_{t+1})) AX_t + \langle \beta_{t+1}, V_{t+1} \rangle) \times \bar{\Lambda}_{t+1} | \mathcal{Y}_{t+1}].$$

so :

$$\begin{aligned} \gamma_{t+1,t+1}(H_{t+1}) &= \sum_{j=1}^N \{ \bar{E} [((H_t + \alpha_{t+1} + \delta_{t+1}f(y_{t+1})) \langle AX_t, e_j \rangle) \bar{\Lambda}_{t+1} | \mathcal{Y}_{t+1}] \} \\ &+ \bar{E} [\langle \beta_{t+1}, V_{t+1} \rangle \times \bar{\Lambda}_{t+1} | \mathcal{Y}_{t+1}], \end{aligned}$$

hence

$$\begin{aligned} \gamma_{t+1,t+1}(H_{t+1}) &= \sum_{j=1}^N \sum_{i=1}^N \{ \bar{E} [((H_t + \alpha_{t+1} + \delta_{t+1}f(y_{t+1})) \langle X_t, e_i \rangle) \bar{\Lambda}_{t+1} a_{ji} e_j | \mathcal{Y}_{t+1}] \} \\ &+ \bar{E} [\langle \beta_{t+1}, V_{t+1} \rangle \times \bar{\Lambda}_{t+1} | \mathcal{Y}_{t+1}]. \end{aligned}$$

we have noted  $a_i = Ae_i$ , so

$$\begin{aligned} \gamma_{t+1,t+1}(H_{t+1}) &= \sum_{i=1}^N \{ \bar{E} [((H_t + \alpha_{t+1} + \delta_{t+1}f(y_{t+1})) \langle X_t, e_i \rangle) \bar{\Lambda}_{t+1} a_i | \mathcal{Y}_{t+1}] \} \\ &+ \bar{E} [\langle \beta_{t+1}, V_{t+1} \rangle \times \bar{\Lambda}_{t+1} | \mathcal{Y}_{t+1}]. \end{aligned}$$

Since for an adapted process  $H_t$  to the sigma-algebra  $\mathcal{G}_t$

$$\bar{E} [\bar{\Lambda}_{t+1} H_t | \mathcal{Y}_{t+1}] = \sum_{i=1}^N \langle \gamma_t(H_t X_t), \Gamma^i(y_{t+1}) \rangle$$

So, for all  $e_r \in \mathbb{E}$

$$\begin{aligned} \bar{E} [\bar{\Lambda}_{t+1} H_t \langle X_t, e_r \rangle | \mathcal{Y}_{t+1}] &= \sum_{i=1}^N \langle \gamma_t(H_t X_t \langle X_t, e_r \rangle), \Gamma^i(y_{t+1}) \rangle \\ &= \sum_{i=1}^N \langle \gamma_t(H_t X_t X_t^T e_r), \Gamma^i(y_{t+1}) \rangle \end{aligned}$$

But we have also :

$$\gamma_t(H_t X_t X_t^T) = \sum_{i=1}^N \langle \gamma_t(H_t X_t), e_i \rangle e_i e_i^T,$$

So we have :

$$\bar{E} [\bar{\Lambda}_{t+1} H_t \langle X_t, e_r \rangle | \mathcal{Y}_{t+1}] = \sum_{i=1}^N \langle \gamma_t(H_t X_t X_t^T e_r), \Gamma^i(y_{t+1}) \rangle = \langle \gamma_t(H_t X_t), \Gamma^r(y_{t+1}) \rangle.$$

Since  $\alpha, \beta, \delta$  are  $\mathcal{G}$  predictable and  $f(y_{t+1})$  measurable with respect to  $\mathcal{Y}_{t+1}$ , the result (14) yield us the conclusion ■