BUSINESS CYCLE AND MARKOV SWITCHING MODELS WITH DISTRIBUTED LAGS: A COMPARISON BETWEEN US AND EURO AREA

by Monica Billio * and Maddalena Cavicchioli †

FOR PhD SESSION

Abstract. *This paper shows that well-known State Space systems used to analyse business cycle phases in several empirical works can be comprised into a broad class of non linear models, the MSI-VARMA. These processes are M-state Markov switching VARMA(p, q) models for which the intercept term depends not only on the actual regime but also on the last r regimes. We give stable finite order VARMA(p^*,q^*) representations for these processes, where upper bounds for p^* and q^* are elementary functions of the dimension K of the process, the number M of regimes, the number of regimes r on the intercept and the orders p and q. If there is no cancellation, the bounds become equalities, and this solves the identification problem. This result allows us to study US and European business cycles and to determine the number of regimes most appropriate for the description of the economic systems. Two regimes are confirmed for the US economy; the European business cycle exhibits, instead, strong non-linearities and more regimes are necessary. This is taken into account when performing estimation and regime identification.*

Keywords: Time series, Varma models, Markov chains, changes in regime, regime number, business cycle models.

JEL Classification: C01, C50, C32, E32.

---

*Department of Economics, "Cá Foscari" University of Venice, Cannaregio 873, 30121 Venice, Italy. E-mail: billio@unive.it

†Advanced School of Economics, Department of Economics, "Cá Foscari" University of Venice, Cannaregio 873, 30121 Venice, Italy. E-mail: maddalena.cavicchioli@unive.it (corresponding author)

1. Introduction

In this paper we consider a generalization of dynamic models denoted as Markov Switching Vector Autoregressive Moving Average (in short, MS(M)-VARMA(p,q)), which are models for time series where we allow the parameters to change as a result of the outcome of an observed Markov chain. In particular, one may also assume that the intercept term depends not only on the actual regime, as in the baseline case (see Cavicchioli (2013)), but also on the last r regimes and we we denote it as MSI(M,r)-VARMA(p,q). It turns out that this is a useful representation of a dynamic economic system since some business cycle models can be comprised in this framework. As a consequence, a key problem is the determination of the number of states which best describes the observed data. Our method for the determination of the number of regimes relies on the computation of the autocovariance function and on finite order (or stable) VARMA representation of the initial switching model (see Krolzig (1997) and Cavicchioli (2013)). More precisely, the parameters of the VARMA representation can be determined by evaluating the autocovariance function of the Markov-switching model. It turns out that the orders p^* and q^* of the finite order VARMA representations are elementary functions of the dimension K of the process, the number M of regimes, the number of lags r on the intercept and the orders p and q of the initial swiching model. Moreover, if there is no cancellation among the roots of the autoregressive and moving average polynomials, the bounds become equalities, and this solves the identification problem.

We note that well-known State-Space models of business cycle can be rewritten into the MSI-VARMA framework. This is convenient since it allows simpler inference and gives a way to determine the number of regimes which better describes the economic system. Business cycle analysis takes its first steps from the main empirical facts establishing that during a postwar period, contraction has typically been followed by a high-growth recovery that quickly boosts output to its prerecession level. This two phase pattern was initially proposed by Hamilton (1989) and Lam (1990) and we call it the Lam-Hamilton-Kim model. An alternative description was initiated by Friedman (1964; 1993) who observes that postwar fluctuations in real output should be thought of as having three phases rather than two - contractions, high-growth recoveries and moderate-growth subsequent recoveries (see also DeLong and Summers (1988) and Kim and Nelson (1999)). In the sequel we will refer to this alternative model as Friedman-Kim-Nelson. In recent years, interest has increased in the ability of business cycle models to forecast economic growth rates and turning points or structural breaks in economic activity (see Krolizig (2001,2004), Billio and Casarin (2010) or Billio, Ferrara, Guégan and Mazzi (2013)). However, in empirical applications where such non-linear specifications are employed, the number of regimes is sometimes dictated by the particular application or is determined in an informal manner by visual inspection of plots. Since Hamilton (1989) and his application, for the study of US cycles, two regimes have been considered in many studies. On the contrary, in some recent papers which analyze the Euro area dynamics, more regimes have been suggested. For example, Billio, Casarin, Ravazzolo and Van Dijk (2012) considered Markov Switching models and in their application to US and EU industrial production data, for a period of time including the last recession, they find that four regimes (strong-recession, contraction, normal-growth, and high-growth) are necessary to identify some important features of the cycle.

The main contributions of our paper are twofold. Firstly, we show that the most used models for business cycle analysis can be comprised into a broad class of non linear model, the MSI-VARMA, and we obtain new results related to it. Specifically, we give its stable VARMA(p^*, q^*) representation and the orders can be determined by evaluating the autocovariance function of the initial switching model. Secondly, we propose a new and more rigorous way for the determination of the number of regimes and apply it to analyse the US and Euro business cycles. In particular, we are able to assess that two regimes are sufficient when modeling the US business cycle but more regimes are necessary when we consider the Euro area. This is the preliminary stage to obtain correct estimation and to identify regimes.

The rest of the paper is organized as follows. In Section 2 we review some facts linking Markov switching models to business cycle analysis and thus we introduce the MS-VAR model. In Section 3 we study a generalization denoted as MSI-VARMA starting from the baseline case of an hidden Markov chain process with distributed lags in the regime (in short, MSI(M, r)-VAR(0)); here we give upper bounds for the stable VARMA order both via autocovariance function and via explicit determination of the stable VARMA. Then we extend these results in Section 4 where theorems are stated for every MSI(M, r)-VARMA(p, q) model in which the regime variable is uncorrelated with the observable. Section 5 introduces the Lam-Hamilton-Kim and the Friedman-Kim-Nelson models of business cycle fluctuations and shows that these State Space models can be expressed as MSI-VARMA. Finally, an application on the determination of the number of regimes for the US and Euro real GDP is conducted in Section 6, followed by estimation and regime identification. Section 7 concludes. Proofs are given in the Appendix.

2. Markov Switching Models and Business Cycle

Many economic time series occasionally exhibit dramatic breaks in their behavior, associated with events such as financial crises, abrupt changes in government policy or in the price of production factors. Of particular interest for economists is the statistical measurement and forecasting of business cycles. Since the early work of Bruns and Mitchell (1946), many attempts have been made to identify cycles and to provide a turning-point chronology that dates the cycle for a given country or economic area. The modern tools to deal with business cycle analysis refer to nonlinear parametric modelling, which are flexible enough to take into account certain stylized facts, such as asymmetries in the phase of the cycle. In fact, there is a large literature that uses Markov switching (MS) models to recognize business cycle phases. The starting point of this strand of literature is the recognition that there is a relationship between the concepts of changes in cyclical phases and changes in regime. The most representative works are the univariate regime-switching model proposed by Hamilton (1989) and its multivariate extension allowing both for co-movement of macroeconomic variables and switching regime as in Kim and Nelson (1999). The relationship between turning point and change in regime has been confirmed by a number of empirical studies, as Krolzig (2001, 2004) and Billio, Ferrara, Guégan and Mazzi (2013) among the others. When using parametric models to analyze the cycle, the most used models are certainly MS-VAR. This approach do not assume any a priori definition of the business cycle: by means of the switching approach, different regimes are identified. Indeed, these regimes differ in terms of average growth rates and/or growth volatilities. Let us now introduce the MS-VAR model. Consider the K-dimensional second-order stationary dynamic process $\mathbf{y} = (\mathbf{y}_t)$ satisfying the following Markov switching autoregressive model

$$\boldsymbol{\phi}_{s_t}(L)\mathbf{y}_t = \boldsymbol{\nu}_{s_t} + \boldsymbol{\Sigma}_{s_t}\mathbf{u}_t \tag{2.1}$$

where $\mathbf{u}_t \sim IID(\mathbf{0}, \mathbf{I}_K)$ and $\boldsymbol{\phi}_{s_t}(L) = \sum_{i=0}^{p} \boldsymbol{\phi}_{s_t,i} L^i$ with $\boldsymbol{\phi}_{s_t,0} = \mathbf{I}_k$ and $\boldsymbol{\phi}_{s_t,p} \neq \mathbf{0}$. As usual, we assume that the polynomials $|\boldsymbol{\phi}_{s_t}(z)|$ have all their roots strictly outside the unit circle. Sufficient conditions ensuring second-order stationarity for Markov-switching VAR models and Markov-switching VARMA models can be found, for example, in Francq and Zakoïan (2001), respectively. The regime (s_t) follows an M-state ergodic irreducible Markov chain with $\mathbf{P} = (p_{ij})$ being the $(M \times M)$ matrix of transition probabilities $p_{ij} = Pr(s_t = j|s_{t-1} = i)$, for $i, j = 1, \ldots, M$. In general, from a statistical point of view, the order p, the number of states M, the parameters and the transition matrix $\mathbf{P}$ are unknown. However, it is established in the literature that such models admit finite-order VARMA$(p^*; q^*)$ representations. Several authors (Krolzig (1997), Zhang and Stine (2001), Francq and Zakoan (2001,2002), Cavicchioli (2013) have looked at the problem of finding upper bounds for p^* and q^*, expressed as functions of various parameters of the initial switching model. One possible application of such bounds are corresponding lower bounds for M which in principle could be useful in real data situation. Another method for the determination of regimes' number refers to complexity-penalized likelihood criteria, such as AIC, BIC, HQC (see Psaradakis and Spagnolo 2003, Olteanu and Rynkiewicz 2007, Rios and Rodriguez 2008). However, these criteria are not widely used in empirical literature, possibly for the computational burder required. The sample autocovariances are instead more easily calculated than maximum (penalized) likelihood estimates of the model parameters and the bounds arising from the above-mentioned elementary functions are very useful for selecting the number of regimes and the orders of the switching autoregression. In the sequel we will consider a generalization of these models where the intercept depends not only on the actual regime but also on the some previous regimes and we propose a method for the determination of regimes number. This is interesting since well-known models used for business cycle analysis can be recomprised in this framework, thus having a rigorous way for detection of regimes and estimation.

3. The MSI(M, r) - VAR(0) Model

Let us consider a generalization of the MS(M)-VAR(p) model (see Cavicchioli (2013)), for which we assume that the intercept term depends not only on the actual regime but also on the last r ($r \geq 0$) regimes

$$\boldsymbol{\nu}_t = \boldsymbol{\nu}_{s_t, s_{t-1}, \dots, s_{t-r}} = \sum_{j=0}^{r} \boldsymbol{\nu}_{j, s_{t-j}} = \sum_{j=0}^{r} \sum_{m=1}^{M} \boldsymbol{\nu}_{jm} I(s_{t-j} = m)$$

where the indicator function $I(s_t = \ \cdot)$ takes on the value 1 if $s_t = m$ and zero otherwise. This specification is called MSI(M, r) - VAR(p) model. Here we treat the case $r > 0$ (for $r = 0$ see Cavicchioli (2013)). (The basic reference for our arguments and techniques is the Krolzig book (1997)). First we consider the K-dimensional MSI(M, r)-VAR(0) process:

$$\mathbf{y}_t = \sum_{j=0}^{r} \boldsymbol{\nu}_{j, s_{t-j}} + \boldsymbol{\Sigma}_{s_t} \mathbf{u}_t \tag{3.1}$$

where $\mathbf{u}_t \sim IID(\mathbf{0}, \mathbf{I}_K)$ and (s_t) follows an M-state ergodic irreducible Markov chain. The Markov chain follows an AR(1) process

$$\boldsymbol{\xi}_t = \mathbf{P}^{'} \boldsymbol{\xi}_{t-1} + \mathbf{v}_t.$$

where $\mathbf{P} = (p_{ij})$ is the $(M \times M)$ matrix of transition probabilities $p_{ij} = Pr(s_t = j | s_{t-1} = i)$, for $i, j = 1, \dots, M$, and $\boldsymbol{\xi}_t$ denotes the random $(M \times 1)$ vector whose mth element is equal to 1 if $s_t = m$ and zero otherwise. Here the innovation process $(\mathbf{v}_t)$ is a martingale difference sequence defined by $\mathbf{v}_t = \boldsymbol{\xi}_t - E(\boldsymbol{\xi}_t | \boldsymbol{\xi}_{t-1})$; it is uncorrelated with $\mathbf{u}_t$ and past values of $\mathbf{u}$, $\boldsymbol{\xi}$ or $\mathbf{y}$. The $(M \times 1)$ vector of the ergodic probabilities is denoted by $\boldsymbol{\pi} = E(\boldsymbol{\xi}_t) = (\pi_1, \dots, \pi_M)^{'}$. It turns out to be the eigenvector of $\mathbf{P}$ associated with the unit eigenvalue, that is, the vector $\boldsymbol{\pi}$ satisfies $\mathbf{P}^{'} \boldsymbol{\pi} = \boldsymbol{\pi}$. The eigenvector $\boldsymbol{\pi}$ is normalized so that its elements sum to unity. Irreducibility implies that $\pi_m > 0$, for $m = 1, \dots, M$, meaning that all unobservable states are possible.

The process in (3.1) has a first State-Space representation as follows

$$\begin{cases} \mathbf{y}_t = \sum_{j=0}^{r} \boldsymbol{\Lambda}_j \boldsymbol{\xi}_{t-j} + \boldsymbol{\Sigma}(\boldsymbol{\xi}_t \otimes \mathbf{I}_K)\mathbf{u}_t \\ \boldsymbol{\xi}_t = \mathbf{P}^{'} \boldsymbol{\xi}_{t-1} + \mathbf{v}_t \end{cases} \tag{3.2}$$

where $\boldsymbol{\Lambda}_j = (\boldsymbol{\nu}_{j1} \, \ldots \, \boldsymbol{\nu}_{jM})$, for $j = 0, \ldots, r$, and $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1 \, \ldots \, \boldsymbol{\Sigma}_M)$.

Theorem 3.1 *For every $h \geq r > 0$, the autocovariance function of the process $(\mathbf{y}_t)$ in (3.1) satisfies*

$$\boldsymbol{\Gamma}_{\mathbf{y}}(h) = \mathbf{A}^{'}(\mathbf{Q}^{'})^h \mathbf{B}$$

where

$$\mathbf{A}^{'} = \sum_{i=0}^{r} \boldsymbol{\Lambda}_i (\mathbf{P}^{'})^{-i} \qquad \mathbf{B} = \sum_{j=0}^{r} (\mathbf{P}^{'})^j \mathbf{D} \boldsymbol{\Lambda}_j^{'} \qquad \mathbf{Q} = \mathbf{P} - \mathbf{P}_\infty \qquad \mathbf{P}_\infty = \lim_n \mathbf{P}^n.$$

We can always obtain a second State-Space representation in the following way:

$$\mathbf{y}_t = \sum_{j=0}^{r} \boldsymbol{\Lambda}_j (\boldsymbol{\xi}_{t-j} - \boldsymbol{\pi}) + (\sum_{j=0}^{r} \boldsymbol{\Lambda}_j)\boldsymbol{\pi} + \boldsymbol{\Sigma}((\boldsymbol{\xi}_t - \boldsymbol{\pi}) \otimes \mathbf{I}_K)\mathbf{u}_t + \boldsymbol{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{u}_t$$

or equivalently

$$\mathbf{y}_t = (\sum_{j=0}^{r} \boldsymbol{\Lambda}_j)\boldsymbol{\pi} + \sum_{j=0}^{r} \widetilde{\boldsymbol{\Lambda}}_j \boldsymbol{\delta}_{t-j} + \widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\delta}_t \otimes \mathbf{I}_K)\mathbf{u}_t + \boldsymbol{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{u}_t$$

where

$$\widetilde{\boldsymbol{\Lambda}}_j = (\boldsymbol{\nu}_{j1} - \boldsymbol{\nu}_{jM} \, \ldots \, \boldsymbol{\nu}_{jM-1} - \boldsymbol{\nu}_{jM}) \qquad \widetilde{\boldsymbol{\Sigma}} = (\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma}_M \, \ldots \, \boldsymbol{\Sigma}_{M-1} - \boldsymbol{\Sigma}_M).$$

Here $\boldsymbol{\delta}_t$ is the $(M-1) \times 1$ vector formed by the columns, but the last one, of $\boldsymbol{\xi}_t - \boldsymbol{\pi}$. Of course, the last formula is obtained by using the restrictions $\mathbf{i}_M^{'} \boldsymbol{\xi}_t = 1$ and $\mathbf{i}_M^{'} \boldsymbol{\pi} = 1$, where $\mathbf{i}_M^{'}$ denotes the $(M \times 1)$ vector of ones. So we get

$$\begin{cases} \mathbf{y}_t = (\sum_{j=0}^{r} \boldsymbol{\Lambda}_j)\boldsymbol{\pi} + \sum_{j=0}^{r} \widetilde{\boldsymbol{\Lambda}}_j \boldsymbol{\delta}_{t-j} + \widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\delta}_t \otimes \mathbf{I}_K)\mathbf{u}_t + \boldsymbol{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{u}_t \\ \boldsymbol{\delta}_t = \mathbf{F}\boldsymbol{\delta}_{t-1} + \mathbf{w}_t \end{cases} \tag{3.3}$$

where $E(\mathbf{w}_t \mathbf{w}_t^{'}) = \widetilde{\mathbf{D}} - \mathbf{F}\widetilde{\mathbf{D}}\mathbf{F}^{'}$, $\quad E(\mathbf{w}_t \mathbf{w}_\tau^{'}) = \mathbf{0}$ for $t \neq \tau$,

$$\widetilde{\mathbf{D}} = \begin{pmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & \cdots & -\pi_1\pi_{M-1} \\ -\pi_1\pi_2 & \pi_2(1-\pi_2) & \cdots & -\pi_2\pi_{M-1} \\ \vdots & \vdots & & \vdots \\ -\pi_{M-1}\pi_1 & -\pi_{M-1}\pi_2 & \cdots & \pi_{M-1}(1-\pi_{M-1}) \end{pmatrix}$$

and

$$\mathbf{F} = \begin{pmatrix} p_{11} - p_{M1} & \cdots & p_{M-1,1} - p_{M1} \\ \vdots & & \vdots \\ p_{1,M-1} - p_{M,M-1} & \cdots & p_{M-1,M-1} - p_{M,M-1} \end{pmatrix}$$

which is an $(M-1) \times (M-1)$ matrix with all eigenvalues inside the unit circle. Since $\boldsymbol{\delta}_t$ has zero mean, the unconditional expectation of the initial process is given by $E(\mathbf{y}_t) = \boldsymbol{\mu}_{\mathbf{y}} = (\sum_{j=0}^{r} \boldsymbol{\Lambda}_j)\boldsymbol{\pi}$, as before. By iteration of the transition equation in (3.3), we also obtain $E(\boldsymbol{\delta}_t \boldsymbol{\delta}_{t+h}^{'}) = \widetilde{\mathbf{D}}(\mathbf{F}^{'})^h$ for every $h \geq 0$.

For Model (3.1), we obtain the following main result:

Theorem 3.2. *The second-order stationary dynamic process defined in (3.1), with* $r > 0$*, has a stable VARMA(*p^*, q^**) representation, where* $p^* \leq M-1$ *and* $q^* \leq M+r-2$*. If the lag polynomials of the AR and MA parts of the VARMA(*p^*, q^**) have no roots in common, equalities hold in the previous relations, and the identification problem is completely solved, that is,* $M = p^* + 1$ *and* $r = q^* - p^* + 1$ *(in this case, we have* $q^* \geq p^*$*).*

Now we also determine explicitly a stable VARMA(p^*, q^*) representation for the process $(\mathbf{y}_t)$ in (3.1), with $r > 0$. In our computation, the autoregressive lag polynomial of such a stable VARMA is shown to be scalar.

Theorem 3.3. *The second-order stationary dynamic process defined in (3.1), with* $r > 0$*, has a stable VARMA(*p^*, q^**) representation, with* $p^* \leq M-1$ *and* $q^* \leq M+r-2$*,*

$$\gamma(L)(\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{y}}) = \mathbf{C}(L)\boldsymbol{\epsilon}_t$$

where $\gamma(L) = |F(L)|$ *is a scalar polynomial of degree* $M-1$ *in* L *(that is, the determinant of the matrix* $F(L) = \mathbf{I}_{M-1} - \mathbf{F}L$*, where* L *is the lag operator) and* $\mathbf{C}(L)$ *is the* $[K \times (KM+M-1)]$*-dimensional lag polynomial matrix of degree* $M+r-2$ *in* L *given by*

$$\mathbf{C}(L) = [\sum_{j=0}^{r} \widetilde{\boldsymbol{\Lambda}}_j F(L)^* L^j \qquad \widetilde{\boldsymbol{\Sigma}}(F(L)^* \otimes \mathbf{I}_K) \qquad |F(L)|\boldsymbol{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_K)]$$

where $F(L)^*$ *is the adjoint matrix of* $F(L)$*. Furthermore,* $\boldsymbol{\epsilon}_t = (\mathbf{w}_t^{'} \quad \mathbf{u}_t^{'}(\mathbf{w}_t^{'} \otimes \mathbf{I}_K) \quad \mathbf{u}_t^{'})^{'}$ *is a zero mean white noise process with* $\mathrm{var}(\boldsymbol{\epsilon}_t) = \mathrm{diag}(\widetilde{\mathbf{D}} - \mathbf{F}\widetilde{\mathbf{D}}\mathbf{F}^{'}, (\widetilde{\mathbf{D}} - \mathbf{F}\widetilde{\mathbf{D}}\mathbf{F}^{'}) \otimes \mathbf{I}_K, \mathbf{I}_K)$*. If* $\gamma(L)$ *and* $\mathbf{C}(L)$ *are coprime, then* $p^* = M-1$ *and* $q^* = M+r-2$*, and the identification problem is completely solved.*

4. The MSI(M, r)-VARMA(p, q) Model

Firstly, we consider the following K-dimensional second-order stationary MSI(M, r)-VAR(p) process, with $r > 0$ and $p > 0$,

$$\boldsymbol{\phi}_{s_t}(L)\mathbf{y}_t = \sum_{j=0}^{r} \boldsymbol{\nu}_{j,s_{t-j}} + \boldsymbol{\Sigma}_{s_t}\mathbf{u}_t \tag{4.1}$$

where $\mathbf{u}_t \sim IID(\mathbf{0}, \mathbf{I}_K)$ and $\boldsymbol{\phi}_{s_t}(L) = \boldsymbol{\phi}_{0,s_t} + \boldsymbol{\phi}_{1,s_t} L + \cdots + \boldsymbol{\phi}_{p,s_t} L^p$ with $\boldsymbol{\phi}_{0,s_t} = \mathbf{I}_K$ and $\boldsymbol{\phi}_{p,s_t} \neq \mathbf{0}$. As usual, we assume that the polynomials $|\boldsymbol{\phi}_{s_t}(z)|$ have all their roots strictly outside the unit circle. Sufficient conditions ensuring second-order stationarity for Markov-switching VARMA models can be found, for example, in Francq and Zakoïan (2001) .

For every $j = 0, \dots, r$, define $\boldsymbol{\Lambda}_j = (\boldsymbol{\nu}_{j1} \dots \boldsymbol{\nu}_{jM})$. Then define $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1 \dots \boldsymbol{\Sigma}_M)$ and

$$\boldsymbol{\phi}(L) = [\mathbf{I}_K + \boldsymbol{\phi}_{1,1} L + \cdots + \boldsymbol{\phi}_{p,1} L^p \dots \mathbf{I}_K + \boldsymbol{\phi}_{1,M} L + \cdots + \boldsymbol{\phi}_{p,M} L^p].$$

The process in (4.1) has a first State-Space representation as follows

$$\begin{cases} \boldsymbol{\phi}(L)(\boldsymbol{\xi}_t \otimes \mathbf{I}_K)\mathbf{y}_t = \sum_{j=0}^{r} \boldsymbol{\Lambda}_j \boldsymbol{\xi}_{t-j} + \boldsymbol{\Sigma}(\boldsymbol{\xi}_t \otimes \mathbf{I}_K)\mathbf{u}_t \\ \boldsymbol{\xi}_t = \mathbf{P}' \boldsymbol{\xi}_{t-1} + \mathbf{v}_t \end{cases} \tag{4.2}$$

Moreover, we can always obtain a second State-Space representation:

$$\begin{cases} \widetilde{\boldsymbol{\phi}}(L)(\boldsymbol{\delta}_t \otimes \mathbf{I}_K)\mathbf{y}_t + \boldsymbol{\phi}(L)(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{y}_t \\ \qquad = (\sum_{j=0}^{r} \boldsymbol{\Lambda}_j)\boldsymbol{\pi} + \sum_{j=0}^{r} \widetilde{\boldsymbol{\Lambda}}_j \boldsymbol{\delta}_{t-j} + \widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\delta}_t \otimes \mathbf{I}_K)\mathbf{u}_t + \boldsymbol{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{u}_t \\ \boldsymbol{\delta}_t = \mathbf{F}\boldsymbol{\delta}_{t-1} + \mathbf{w}_t \end{cases} \tag{4.3}$$

where $\boldsymbol{\delta}_t$, $\widetilde{\boldsymbol{\Lambda}}_j$ and $\widetilde{\boldsymbol{\Sigma}}$ are defined as in the previous section, and

$$\widetilde{\boldsymbol{\phi}}(L) = [(\boldsymbol{\phi}_{1,1} - \boldsymbol{\phi}_{1,M})L + \cdots + (\boldsymbol{\phi}_{p,1} - \boldsymbol{\phi}_{p,M})L^p \dots (\boldsymbol{\phi}_{1,M-1} - \boldsymbol{\phi}_{1,M})L + \cdots + (\boldsymbol{\phi}_{p,M-1} - \boldsymbol{\phi}_{p,M})L^p].$$

Theorem 4.1. *For every $h \geq r > 0$, assume that the regime variable $\boldsymbol{\xi}_{t+h}$ is uncorrelated with $\mathbf{y}_t$. Then the autocovariance function of the second-order stationary process in (4.1) satisfies*

$$B(L)\boldsymbol{\Gamma}_{\mathbf{y}}(h) = \mathbf{A}' \mathbf{F}^h \mathbf{B}$$

where $\mathbf{A}' = \sum_{j=0}^{r} \widetilde{\boldsymbol{\Lambda}}_j \mathbf{F}^{-j}$ and $\mathbf{B} = E(\boldsymbol{\delta}_t \boldsymbol{y}_t')$, which are assumed to be nonzero matrices.

Now, applying Theorem 2.2 from Cavicchioli (2013) for $q = r$ and taking in mind that $\mathbf{F}$ is $(M-1) \times (M-1)$, we have the following main result for model (4.1):

Theorem 4.2. *Under the hypothesis that the regime variable is uncorrelated with the observable, the K-dimensional second-order stationary process in (4.1), with $r > 0$ and $p > 0$, admits a stable VARMA(p^*, q^*) representation with $p^* \leq M + p - 1$ and $q* \leq M + r - 2$. If we require that autoregressive lag polynomial of such a stable representation is scalar, then the bounds become $p^* \leq M + Kp - 1$ and $q* \leq M + (K-1)p + r - 2$. If the lag polynomials of the AR and MA parts of the former VARMA(p^*, q^*) representation* *have no roots in common, equalities hold in the previous relations, that is,* $p^* = M + p - 1$ *and* $q^* = M + r - 2$.

To end the section we compute explicitly a VARMA representation for a general MSI-VARMA. This gives a new proof of Theorem 3.3 (case $q = 0$) and extends Proposition 5 from Krolzig (1997), Section 10.2.2. We start with the more simple case in which the autoregressive lag polynomial of the initial process is state independent.

Theorem 4.3. *Let* $\mathbf{y} = (\mathbf{y}_t)$ *be an* K*–dimensional second-order stationary MSI(*M, r*)-VARMA(*p, q*) process, with* $r > 0$,

$$\mathbf{A}(L)\mathbf{y}_t = \sum_{j=0}^{r} \boldsymbol{\nu}_{j,s_{t-j}} + \sum_{i=0}^{q} \boldsymbol{\Theta}_{s_t}(L)\mathbf{u}_t$$

where $\mathbf{A}(L) = \sum_{\ell=0}^{p} \mathbf{A}_\ell L^\ell$ *with* $\boldsymbol{A}_0 = \mathbf{I}_K$, $|\mathbf{A}_p| \neq \mathbf{0}$*, and* $\boldsymbol{\Theta}_{s_t}(L) = \sum_{i=0}^{q} \boldsymbol{\Theta}_{s_t,i} L^i$*, with* $\boldsymbol{\Theta}_{s_t,0} = \boldsymbol{\Sigma}_{s_t}$ *(nonsingular symmetric* $K \times K$ *matrix) and* $|\boldsymbol{\Theta}_{s_t,q}| \neq \mathbf{0}$ *are full rank matrix lag polynomials. Under quite general regularity conditions, the dynamic process* $\mathbf{y} = (\mathbf{y}_t)$ *admits a stable VARMA(*p^*, q^**) representation, with* $p^* \leq M + Kp - 1$ *and* $q* \leq M + (K-1)p + \max\{r, q+1\} - 2$,

$$\gamma(L)(\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{y}}) = \mathbf{C}(L)\boldsymbol{\epsilon}_t$$

where $\gamma(L) = |F(L)||\mathbf{A}(L)|$ *is the scalar AR operator of degree* $M + Kp - 1$*,* $\mathbf{C}(L)$ *is a matrix lag polynomial of degree* $M + (K-1)p + \max\{r, q+1\} - 2$*, and* $\boldsymbol{\epsilon}_t$ *is a zero mean vector white noise process.*

In the general case in which the autoregression part of the process in Theorem 4.3 is state dependent but the regime variable is uncorrelated with the observable, we can proceed as follows. By Theorem 4.1 the autocovariances of the process satisfy a finite difference equation of order $p^* = M + p - 1$ and rank $q^* + 1 = M + \max\{r, q+1\} - 1$. Then the process can be represented by a stable VARMA(p^*, q^*). Given the process $(\mathbf{y}_t)$, we can estimate the coefficients of the stable VARMA(p^*, q^*) with usual procedures. If there is no cancellation between the AR and MA part of the estimated VARMA(p^*, q^*), then we get the representation as in Theorem 4.3 with equalities.

5. Business Cycle Models

In this Section we show that business cycle models widely used in empirical works can be rewritten as MSI-VARMA. Therefore, we can formally test the number of regimes as well as lags in the intercept or autoregressive lags, given the above results. This avoids informal determination of regimes' number by the researcher and leads to correct estimation results and possibly to reliable forecasting conclusions.

(5.1) *The Lam-Hamilton-Kim model.* In modeling the time series behaviour of real Gross National Product (GNP) of United States, Hamilton (1989) considered the case in which real GNP is generated by the sum of two independent unobserved components, one following an autoregressive process with a unit root, and the other following a random walk with a Markov switching error term. Lam (1990) generalized the Hamilton model to the case in which the autoregressive component need not contain a unit root. Finally, Kim (1994) showed that the Lam-Hamilton model can be written in the following State-Space form, which we shall call *the Lam-Hamilton-Kim model* :

$$\begin{cases} y_t = \mathbf{H}\mathbf{x}_t + \beta_{s_t} \\ \mathbf{x}_t = \mathbf{\Phi}\mathbf{x}_{t-1} + \mathbf{e}_t \end{cases} \tag{5.1}$$

where y_t is scalar ($K = 1$), $\mathbf{x}_t = (x_t \ x_{t-1} \cdots x_{t-r+1})'$ and $\mathbf{e}_t = (u_t \ 0 \cdots 0)'$ are $r \times 1$, with $u_t \sim IID(0, \sigma^2)$, $\mathbf{H} = (1 \ -1 \ 0 \cdots 0)$ is $1 \times r$ and

$$\mathbf{\Phi} = \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \cdots & \phi_r \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

with $\phi_i \in \mathbb{R}$, $i = 1, \ldots, r$, $(r \geq 2)$. Here $\beta_{s_t} = \delta_0 + \delta_1 s_t$, where $\delta_j \in \mathbb{R}$, $j = 0, 1$, and (s_t) is an M-state Markov chain. Kim (1994) estimated this model by using suitable (filtering and smoothing) algorithms that he also constructed for more general State-Space representations with Markov switching. Now we are going to show that Model (5.1) can be interpreted as a model with distributed lags in the regime. Then we give a stable VARMA representation of it. This allows us to estimate the model via MLE, i.e., a very simple procedure which is alternative to that employed by Kim (1994). From the transition equation in (5.1) we get

$$\Phi(L)\mathbf{x}_t = \mathbf{e}_t$$

hence

$$|\Phi(L)|\mathbf{x}_t = \Phi(L)^*\mathbf{e}_t \tag{5.2}$$

where $\Phi(L) = \mathbf{I}_r - \mathbf{\Phi}L$. Substituting (5.2) into the measurement equation in (5.1) after premultiplying by the determinant of $\Phi(L)$ gives

$$|\Phi(L)|y_t = |\Phi(L)|\beta_{s_t} + \mathbf{H}\Phi(L)^*\mathbf{e}_t \tag{5.3}$$

which is an MSI(M, r)-VARMA(p, q) model in the sense of Section 4, where $p = r$ and $q = r-1$ (recall $r \geq 2$). So Theorem 4.3 directly implies that the Lam-Hamilton-Kim model

admits a stable VARMA(p^*, q^*) representation (whose autoregressive lag polynomial is scalar) with $p^* \leq M + r - 1$ and $q^* \leq M + \max\{p, p\} - 2 = M + p - 2$.
We now determine explicitly the final form of this stable VARMA and we need some notation. Define $\boldsymbol{\beta} = (\beta_1 \cdots \beta_M)$, where $\beta_m = \delta_0 + \delta_1 m$ for every $m = 1, \ldots, M$. Then we get $\beta_{s_t} = \boldsymbol{\beta}\boldsymbol{\xi}_t$. Substituting this relation into (5.3) yields

$$\begin{aligned}
|\Phi(L)|y_t &= |\Phi(L)|\boldsymbol{\beta}\boldsymbol{\xi}_t + \mathbf{H}\Phi(L)^*\mathbf{e}_t \\
&= |\Phi(L)|\boldsymbol{\beta}(\boldsymbol{\xi}_t - \boldsymbol{\pi}) + |\Phi(1)|\boldsymbol{\beta}\boldsymbol{\pi} + \mathbf{H}\Phi(L)^*\mathbf{e}_t \\
&= |\Phi(L)|\widetilde{\boldsymbol{\beta}}\boldsymbol{\delta}_t + |\Phi(1)|\boldsymbol{\beta}\boldsymbol{\pi} + \mathbf{H}\Phi(L)^*\mathbf{e}_t
\end{aligned}$$

where $\widetilde{\boldsymbol{\beta}} = (\beta_1 - \beta_M \cdots \beta_{M-1} - \beta_M)$. Then we get the State-Space representation

$$\left\{\begin{aligned}
|\Phi(L)|(y_t - \mu_y) &= |\Phi(L)|\widetilde{\boldsymbol{\beta}}\boldsymbol{\delta}_t + \mathbf{H}\Phi(L)^*\mathbf{e}_t \\
\boldsymbol{\delta}_t &= \mathbf{F}\boldsymbol{\delta}_{t-1} + \mathbf{w}_t
\end{aligned}\right. \tag{5.4}$$

where $|\Phi(L)|\mu_y = |\Phi(1)|\boldsymbol{\beta}\boldsymbol{\pi}$ (in fact, $\mu_y = \boldsymbol{\beta}\boldsymbol{\pi}$). Substituting $\boldsymbol{\delta}_t = F(L)^{-1}\mathbf{w}_t$ into the measurement equation in (5.4) and premultiplying by $|F(L)|$ yield

$$|F(L)||\Phi(L)|(y_t - \mu_y) = \widetilde{\boldsymbol{\beta}}|\Phi(L)|F(L)^*\mathbf{w}_t + \mathbf{H}|F(L)|\Phi(L)^*\mathbf{e}_t$$

which is a stable VARMA(p^*, q^*) with $p^* \leq M + r - 1$ and $q^* \leq M + r - 2$, and whose autoregressive lag polynomial is scalar. Summarizing we have the following result:

Theorem 5.1. *The Lam-Hamilton-Kim model for US real GNP admits a stable VARMA(p^*, q^*) representation*

$$\gamma(L)(y_t - \mu_y) = \mathbf{C}(L)\boldsymbol{\epsilon}_t$$

with $p^* \leq M + r - 1$ and $q* \leq M + r - 2$. Under quite general regularity conditions, $\gamma(L) = |F(L)||\Phi(L)|$ is the scalar AR operator of degree $M + r - 1$ in the lag operator L and $\mathbf{C}(L)$ is a $[1 \times (M + r - 1)]$-dimensional matrix lag polynomial of degree $M + r - 2$ in L given by

$$\mathbf{C}(L) = [\mathbf{H}|F(L)|\Phi(L)^* \qquad \widetilde{\boldsymbol{\beta}}|\Phi(L)|F(L)^*]$$

and $\boldsymbol{\epsilon}_t = (\mathbf{e}_t' \quad \mathbf{w}_t')'$ is a zero mean $(M + r - 1) \times 1$ vector white noise process. If $\gamma(L)$ and $\mathbf{C}(L)$ are coprime, then equalities hold in the previous relations, hence $p^* = M + r - 1$ and $q^* = p^* - 1$.

(5.2) *The Friedman-Kim-Nelson model.* Friedman's plucking model (1964) of business fluctuations suggests that output cannot exceed a ceiling level, and it is occasionally plucked downward by recession. The model implies that business fluctuations are asymmetric, that recessions have only a temporary effect on output, and that recessions are duration dependent while expansions are not. Subsequent literature has provided copious empirical support for these statements. See, for example, De Simone and Clarke (2007) and its references. Kim and Nelson (1998) showed that the Friedman model can be written in the following State-Space form, which we shall call *the Friedman-Kim-Nelson model* :

$$\begin{cases} y_t = \mathbf{H}\mathbf{x}_t \\ \mathbf{x}_t = \boldsymbol{\mu}_{s_t} + \boldsymbol{\Phi}\mathbf{x}_{t-1} + \boldsymbol{\Sigma}_{s_t}\mathbf{e}_t \end{cases} \tag{5.5}$$

where y_t is scalar ($K = 1$), $\mathbf{x}_t$, $\boldsymbol{\mu}_{s_t}$ and $\mathbf{e}_t$ are 4×1 with $\mathbf{e}_t \sim IID(\mathbf{0}, \mathbf{I}_4)$, $\boldsymbol{\Sigma}_{s_t}$ is a 4×4 diagonal matrix, $\mathbf{H} = (1\ 1\ 0\ 0)$, and

$$\boldsymbol{\Phi} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & \phi_1 & \phi_2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

with $\phi_i \in \mathbb{R}$, $i = 1, 2$; here (s_t) is an M-state Markov chain ($M = 2$ in the quoted papers). Kim and Nelson (1998) estimated the model by using Kim's approximate MLE. See also Kim (1994). We show that Model (5.5) can be viewed as a model with distributed lags in the regime and then we give a stable VARMA representation of it. From the transition equation in (5.5) we get

$$\Phi(L)\mathbf{x}_t = \boldsymbol{\mu}_{s_t} + \boldsymbol{\Sigma}_{s_t}\mathbf{e}_t$$

where

$$\Phi(L) = \mathbf{I}_4 - \boldsymbol{\Phi}L = \begin{pmatrix} 1-L & 0 & 0 & -L \\ 0 & 1-\phi_1 L & -\phi_2 L & 0 \\ 0 & -L & 1 & 0 \\ 0 & 0 & 0 & 1-L \end{pmatrix}$$

hence $|\Phi(L)| = (1-L)^2(1-\phi_1 L - \phi_2 L^2)$. Premultiplying by the adjoint matrix $\Phi(L)^*$ we obtain

$$|\Phi(L)|\mathbf{x}_t = \Phi(L)^*\boldsymbol{\mu}_{s_t} + \Phi(L)^*\boldsymbol{\Sigma}_{s_t}\mathbf{e}_t \tag{5.6}$$

where

$$\Phi(L)^* = \begin{pmatrix} (1-L)(1-\phi_1 L-\phi_2 L^2) & 0 & 0 & L(1-\phi_1 L-\phi_2 L^2) \\ 0 & (1-L)^2 & \phi_2 L(1-L)^2 & 0 \\ 0 & L(1-L)^2 & (1-\phi_1 L)(1-L)^2 & 0 \\ 0 & 0 & 0 & (1-L)(1-\phi_1 L-\phi_2 L^2) \end{pmatrix}.$$

Substituting (5.6) into the measurement equation in (5.5) gives

$$|\Phi(L)|y_t = \mathbf{H}\Phi(L)^*\boldsymbol{\mu}_{s_t} + \mathbf{H}\Phi(L)^*\boldsymbol{\Sigma}_{s_t}\mathbf{e}_t$$

which is an MSI(M,r)-VARMA(p,q) with $p=4$ and $r=q=3$. So Theorem 4.3 implies that such a model has a stable VARMA(p^*,q^*) representation, with $p^* \le M+p-1 = M+3$ and $q^* \le M + \max\{r, q+1\} - 2 = M+2$. We now determine explicitly the final form of this stable VARMA. Model (5.5) has the following State-Space representation

$$\begin{cases} y_t = \mathbf{H}\mathbf{x}_t \\ \mathbf{x}_t = \widetilde{\boldsymbol{\mu}}\boldsymbol{\delta}_t + \boldsymbol{\mu}\boldsymbol{\pi} + \boldsymbol{\Phi}\mathbf{x}_{t-1} + \widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\delta}_t \otimes \mathbf{I}_4)\mathbf{e}_t + \boldsymbol{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_4)\mathbf{e}_t \\ \boldsymbol{\delta}_t = \mathbf{F}\boldsymbol{\delta}_{t-1} + \mathbf{w}_t \end{cases} \tag{5.7}$$

where $\boldsymbol{\mu} = (\boldsymbol{\mu}_1 \cdots \boldsymbol{\mu}_M)$ is $4 \times M$, $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1 \cdots \boldsymbol{\Sigma}_M)$ is $4 \times (4M)$, and $\widetilde{\boldsymbol{\mu}}$ and $\widetilde{\boldsymbol{\Sigma}}$ are defined in the usual way. Furthermore, we have $E(y_t) = \mathbf{H}E(\mathbf{x}_t)$ and $\Phi(L)E(\mathbf{x}_t) = \boldsymbol{\mu}\boldsymbol{\pi}$, hence $y_t - E(y_t) = \mathbf{H}(\mathbf{x}_t - E(\mathbf{x}_t))$. From (5.7) we get

$$\begin{aligned} \Phi(L)(\mathbf{x}_t - E(\mathbf{x}_t)) &= \widetilde{\boldsymbol{\mu}}\boldsymbol{\delta}_t + \widetilde{\boldsymbol{\Sigma}}(\boldsymbol{\delta}_t \otimes \mathbf{I}_4)\mathbf{e}_t + \boldsymbol{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_4)\mathbf{e}_t \\ &= \widetilde{\boldsymbol{\mu}}F(L)^{-1}\mathbf{w}_t + \widetilde{\boldsymbol{\Sigma}}(F(L)^{-1}\mathbf{w}_t \otimes \mathbf{I}_4)\mathbf{e}_t + \boldsymbol{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_4)\mathbf{e}_t \end{aligned}$$

hence

$$|F(L)|\Phi(L)(\mathbf{x}_t - E(\mathbf{x}_t)) = \widetilde{\boldsymbol{\mu}}F(L)^*\mathbf{w}_t + \widetilde{\boldsymbol{\Sigma}}(F(L)^*\mathbf{w}_t \otimes \mathbf{I}_4)\mathbf{e}_t + \boldsymbol{\Sigma}|F(L)|(\boldsymbol{\pi} \otimes \mathbf{I}_4)\mathbf{e}_t.$$

Premultiplying by the determinant of $\Phi(L)$ yields

$$|F(L)||\Phi(L)|(\mathbf{x}_t - E(\mathbf{x}_t)) = \Phi(L)^*\widetilde{\boldsymbol{\mu}}F(L)^*\mathbf{w}_t + \Phi(L)^*\widetilde{\boldsymbol{\Sigma}}(F(L)^*\mathbf{w}_t \otimes \mathbf{I}_4)\mathbf{e}_t + \Phi(L)^*\boldsymbol{\Sigma}|F(L)|(\boldsymbol{\pi} \otimes \mathbf{I}_4)\mathbf{e}_t.$$

Assuming $\mu_y = E(y_t)$ time invariant, we obtain

$$\begin{aligned} &|F(L)||\Phi(L)|(y_t - \mu_y) = \mathbf{H}|F(L)||\Phi(L)|(\mathbf{x}_t - E(\mathbf{x}_t)) \\ &= \mathbf{H}\Phi(L)^*\widetilde{\boldsymbol{\mu}}F(L)^*\mathbf{w}_t + \mathbf{H}\Phi(L)^*\widetilde{\boldsymbol{\Sigma}}(F(L)^* \otimes \mathbf{I}_4)(\mathbf{w}_t \otimes \mathbf{I}_4)\mathbf{e}_t + \mathbf{H}\Phi(L)^*\boldsymbol{\Sigma}|F(L)|(\boldsymbol{\pi} \otimes \mathbf{I}_4)\mathbf{e}_t \end{aligned}$$

which is a stable VARMA(p^*,q^*) with $p^* \le M+3$ and $q^* \le M+2$, and whose autoregressive lag polynomial is scalar. Summarizing we have the following result:

Theorem 5.2. *The Friedman-Kim-Nelson model of business fluctuations admits a stable VARMA(p^*, q^*) representation*

$$\gamma(L)(y_t - \mu_y) = \mathbf{C}(L)\boldsymbol{\epsilon}_t$$

with $p^* \le M+3$ and $q* \le M+2$. Under quite general regularity conditions, $\gamma(L) = |F(L)||\Phi(L)|$ is the scalar AR operator of degree $M+3$ in the lag operator L and $\mathbf{C}(L)$ is a $[1 \times (5M-1)]$-dimensional matrix lag polynomial of degree $M+2$ in L given by

$$\mathbf{C}(L) = [\mathbf{H}\Phi(L)^*\widetilde{\boldsymbol{\mu}}F(L)^* \qquad \mathbf{H}\Phi(L)^*\widetilde{\boldsymbol{\Sigma}}(F(L)^* \otimes \mathbf{I}_4) \qquad \mathbf{H}\Phi(L)^*\boldsymbol{\Sigma}|F(L)|(\boldsymbol{\pi} \otimes \mathbf{I}_4)]$$

and $\boldsymbol{\epsilon}_t = (\mathbf{w}_t' \quad \mathbf{e}_t'(\mathbf{w}_t' \otimes \mathbf{I}_4) \quad \mathbf{e}_t')'$ *is a zero mean* $(5M-1)\times 1$ *vector white noise process. If* $\gamma(L)$ *and* $\mathbf{C}(L)$ *are coprime, then equalities hold in the previous relations, hence* $p^* = M+3$ *and* $q^* = p^* - 1$.

6. Empirical Application

In our study we consider the Gross Domestic Product (GDP) from FRED at a quarterly frequency for the United States (US), from 1951:1 to 2012:4, and from EUROSTAT for the European Union (EU 12), from 1973:1 to 2012:4 [1]. The presence of unit root in the data has been checked by augmented Dickey-Fuller (ADF) test which points out the non-stationarity of both series. For the null hypothesis of unit roots, the test statistic gives 1.724 (with $p = 12$) for US GDP and -0.4037 (with p=5) for EU GDP. In both cases the null hypothesis of a unit root cannot be rejected. For differenced time series, the ADF test rejects the unit root hypothesis on the 1% significance level (with test statistics of -4.4853 for US and -5.3384 for EU). In the following analysis we consider the growth rate of quarterly real GDP data for US and EU and the series are plotted in Figure 1 and Figure 2.

We model both series as a MSI(M,r)-AR(0), with state-dependent mean and variance and no autoregressive part. This follows from several empirical studies which find that most part of the forecast errors is due to time changes in some parameters of the prediction models. In particular, we follow Krolzig (2000) and Anas et al. (2008) where only intercept and volatility are assumed to be driven by a regime-switching variable. In fact, with regards to the Euro area, Anas et al. (2008) and Billio et al. (2013) find that allowing regime switching-autoregressive coefficients deteriorates the detection of the business cycle turning points. Note that we can now apply the bounds proposed in Section 3 which simultaneously define number of regimes and lags in the intercept. Those can be obtained having estimates of the stable VARMA orders $\hat{p}^*$ and $\hat{q}^*$ as follows

$$\begin{cases} \hat{M} = \hat{p}^* + 1 \\ \hat{r} = 1 - \hat{p}^* + \hat{q}^*. \end{cases}$$

For the computation of the orders of the stable VARMA we use the 3-pattern method (TPM) proposed by Choi (1992). This gives the results reported in Table 1.

[1] Data are taken from Fred website: $< research.stlouisfed.org >$ and from Eurostat website: $< http : //epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/ >$. Data are seasonally adjusted.

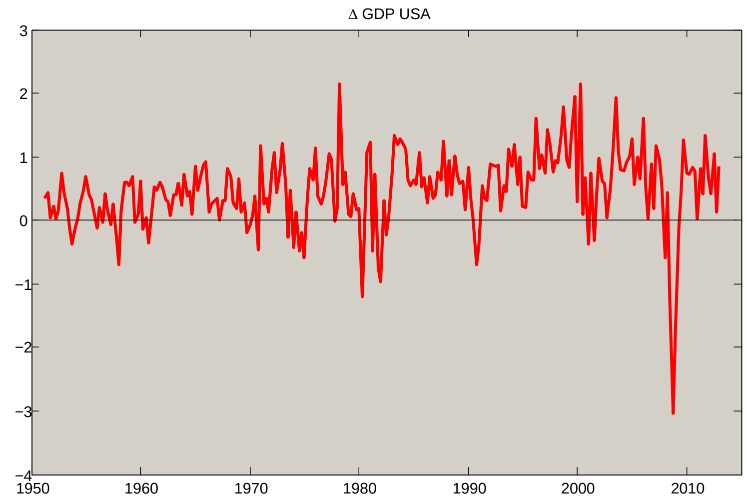

Figure 1: US quarterly growth rate of real GDP for the period 1951:1 - 2012:4. Data are taken from Fred website.

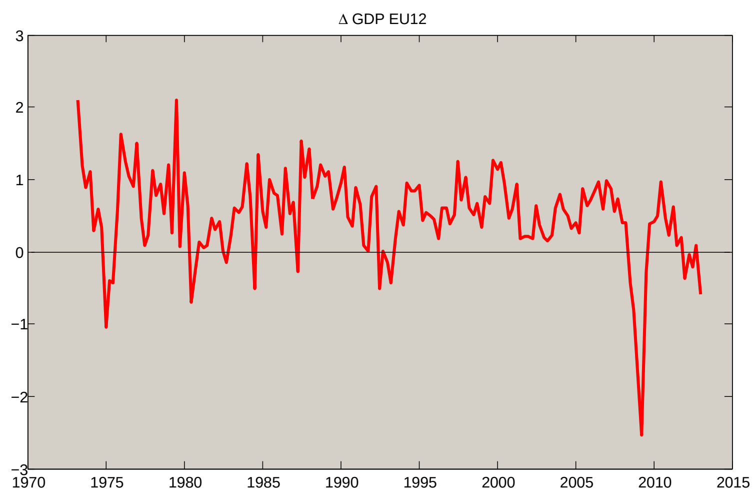

Figure 2: European (EU12) quarterly growth rate of real GDP for the period 1973:1 - 2012:4. Data are taken from Eurostat website.

	$ARMA(\hat{p}^*, \hat{q}^*)$	$\hat{M}$	$\hat{r}$
US	ARMA(1,1)	2	1
EU	ARMA(3,2)	4	0

Table 1: Estimated number of regimes and distributed lags in the intercept for US and EU GDP. The time lags are from 1951:1 to 2012:4 for US GDP and from 1973:1 to 2012:4 for EU GDP. The procedure uses the bounds obtained in Section 3.

Mean		St.Deviation	
μ_1	-0.437	σ_1	1.274
	(0.604)		(0.322)
μ_2	-0.180	σ_2	0.260
	(0.132)		(0.048)
μ_3	0.394	σ_3	0.184
	(0.058)		(0.042)
μ_4	0.893	σ_4	0.289
	(0.058)		(0.042)

Table 2: Estimated coefficients of the MSI(4,0)-AR(0) model for EU quarterly growth rate of real GDP for the period 1973:1 - 2012:4. Standard errors are in parenthesis. The log-likelihood value is 268.4987.

Our results suggest that US real GDP is sufficiently good described with two regimes and one lag in the intercept, which is in line with the estimated Markov-switching State Space models of the previous section and with several empirical works studying US business cycle. Then we are not going to indagate further. On the contrary, when modeling GDP of the Euro Area four regimes are more appropriate. In order to identify regimes for the European economy, we proceed with the estimation of a MSI(4,0)-AR(0) model, as suggested from the above step. Tables 2 and 3 report estimated parameters and their standard errors, the transition matrix and the expected duration of the regimes. Moreover, Figure 2 plots the smoothed probabilities of the four regimes.

All regimes can be matched with a plausible economic interpretation. In fact, recession phases are identified with Regime 1 and 2, being Regime 1 a stronger recession than Regime 2 (contraction), while Regime 3 is moderate/normal growth and Regime 4 is high growth.

	Regime 1	Regime 2	Regime 3	Regime 4
Regime 1	0.49	0.00	0.09	0.00
	(0.45)	(0.01)	(0.06)	(0.01)
Regime 2	0.51	0.58	0.02	0.05
	(0.33)	(0.26)	(0.02)	(0.03)
Regime 3	0.00	0.20	0.68	0.18
	(0.01)	(0.12)	(0.16)	(0.07)
Regime 4	0.00	0.22	0.22	0.76
	(0.01)	(0.12)	(0.09)	(0.13)

Table 3: Transition probability matrix of the estimated MSI(4,0)-AR(0) model for EU quarterly growth rate of real GDP for the period 1973:1 - 2012:4. Standard errors are in parenthesis.

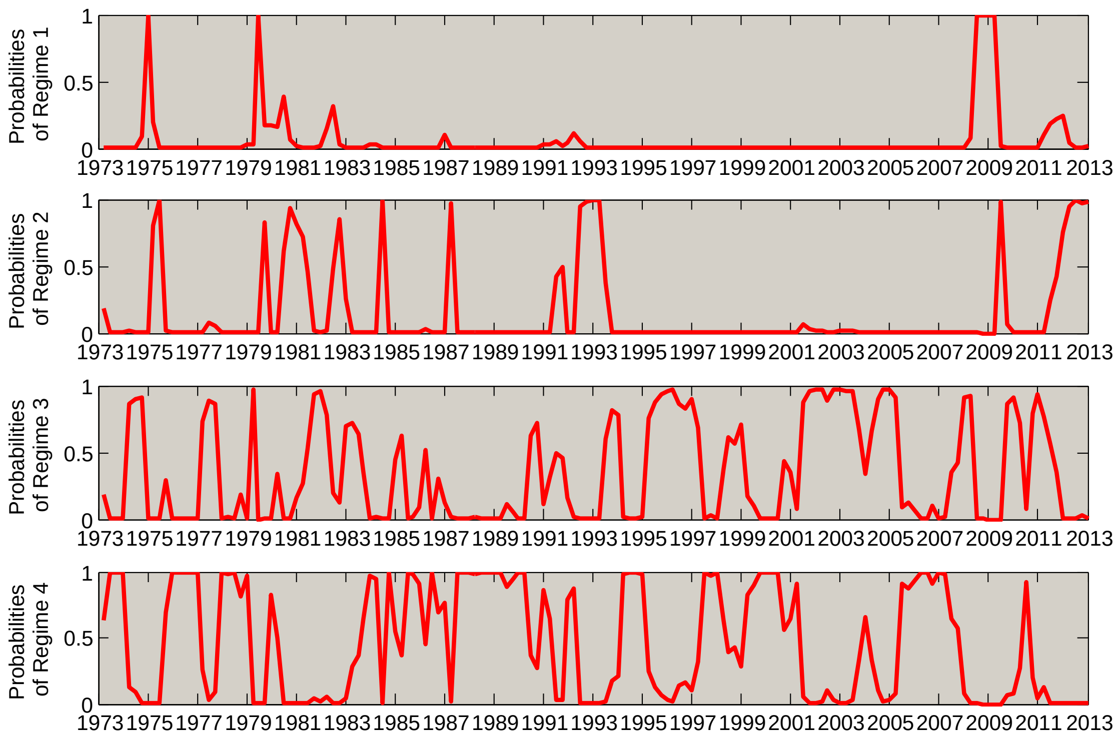

Figure 3: Smoothed probabilities from the estimated MSI(4,0)-AR(0) model for EU quarterly growth rate of real GDP for the period 1973:1 - 2012:4. Data are taken from Eurostat. Regime 1 corresponds to strong recession, Regime 2 to contraction, Regime 3 to moderate/normal growth and Regime 4 to high growth.

The expansionary regimes are more persistent than the others, in fact probabilities of staying in those regimes are 0.68 and 0.76. Note also that all volatility coefficients are significant at 1% level. The expected duration of recession phases is 4/5 quarters, while expansionary regimes cover about 7 quarters. Here we detect turning points as the last quarter of each regime phase and the following recession periods can be inferred from the estimation: 1974:1 - 1975:3, 1979:1 - 1981:3, 1982:1 - 1983:1, 1991:1 - 1993:3, 2008:1 - 2010:1 and 2011:1 - 2012:4. These conclusions are in line with well-recognised recession phases, see, for instance, Anas et al.(2007).

7. Conclusion

In this paper, for a general class of Markov-switching VARMA models with distributed lags in the regime, in symbols MSI(M, r)- VARMA(p, q), we give finite order VARMA(p^*, q^*) representations where the parameters can be determined by evaluating the autocovariance function of the Markov-switching models. It turns out that upper bounds for p^* and q^* are elementary functions of the dimension K of the process, the number M of regimes, the number of regimes r on the intercept and the orders p and q. If there is no cancellation, the bounds become equalities, and this solves the identification problem. This result produces an easy method for setting a lower bound on the number of regimes from the estimated autocovariance function. Of particular interest is how some well-known State Space systems, introduced in the literature for business cycle analysis, are shown to be comprised in this general MSI-VARMA model, such as the Lam-Hamilton-Kim and the Friedman-Kim-Nelson models of business fluctuations. In the application we determine the number of regimes which turns out to be more appropriate for the description of US and EU economic systems by using the bounds obtained in this work. In particular, US real GDP is better described with two regimes, as is usually assumed in the estimation of such State Space systems. However, EU business cycle exhibits strong non-linearities and more regimes are necessary. This is taken into account when performing estimation and regime identification.

References

[1] Anas, J., Billio, M., Ferrara, L., LoDuca, M. (2007) *Business Cycle Analysis with Multivariate Markov Switching Models*, in Growth and Cycle in the Euro Zone, Mazzi GL and Savio G (eds), Palgrave Macmillan: New York, 249–260.

[2] Anas, J., Billio, M., Ferrara, L., LoDuca, M. (2008) *A system for dating and detecting turning points in the Euro area*, The Manchester School, **76**, 549–577.

[3] Billio, M., and Casarin, R. (2010) *Identifying business cycle turning points with sequential monte carlo methods: an online and real-time application to Euro area*, Journal of Forecasting, **29**, 145–167.

[4] Billio, M., Casarin, R., Ravazzolo, F., and Van Dijk, H.K. (2012) *Combination scheme for turning point predictions*, The Quartely Review of Economics and Finance, **52**, 402–412.

[5] Billio, M., Ferrara, L., Gugan, D., and Mazzi, G.L. (2013) *Evaluation of regime switching models for real-time business cycle analysis of the Euro area*, Journal of Forecasting, DOI: 10.1002/for.2260, 2013.

[6] Burns, A.F. and Mitchell, W.C. (1946), *Measuring Business Cycles*, New York: National Bureau of Economic Research.

[7] Cavicchioli, M. (2012) *Determining the number of regimes in Markov-Switching Varma models*, Working Paper Series Unive N.03/2013.

[8] Choi, B. S. (1992) ARMA Model Identification, New York, Springer-Verlag.

[9] DeLong B., and Summers, L. (1988) *How does macroeconomic policy affect output?*, Brookings paper on economic activity **2**, 433–494.

[10] Francq, C. and Zakoïan, J.M. (2001) *Stationarity of Multivariate Markov-Switching ARMA Models*, Journal of Econometrics **102**, 339–364.

[11] Francq, C. and Zakoïan, J.M. (2002) *Autocovariance Structure of Powers of Switching-Regime ARMA Processes*, ESAIM: Probability and Statistics **6**, 259–270.

[12] Friedman, M. (1964) *Monetary Studies of the National Bureau* , The National Bureau Enters Its 45th Year, 44th Annual Report, 7-25; Reprinted in Milton Friedman (1969). The Optimum Quantity of Money and Other Essays. Chicago: Aldine, Chp. **12**, 261–284.

[13] Friedman, M. (1993) *The "plucking model" of business fluctuations Revisited* , Economic Inquiry **31**, 171–177.

[14] Hamilton, J. (1989) *A new approach to the economic analysis of nonstationary time series and the business cycle*, Econometrica **57**, 357–384.

[15] Kim, C.J. (1994) *Dynamic linear models with Markov-switching*, Journal of Econometrics **60**, 1–22.

[16] Kim, C.J. and Nelson, C.R. (1999) *Friedman's plucking model of business fluctuations: tests and estimates of permanent and transitory components*, Journal of Money, Credit, and Banking **31** (3), 317–334.

[17] Kim, C.J. and Nelson, C.R. (1999) *State-Space models with regime switching: classical and Gibbs-sampling approaches with applications*, The MIT Press.

[18] Krolzig, H.M. (1997) *Markov-Switching Vector Autoregressions: Modelling, Statistical Inference and Application to Business Cycle Analysis*, Springer Verlag, Berlin-Heidelberg-New York.

[19] Krolzig, H.M. (2000) *Predicting Markov Switching vector autoregressive processes*, Nuffield College Economics Working Papers, 2000-WP31.

[20] Krolzig, H.M. (2001) *Markov Switching procedures for dating the Euro-zone business cycle*, Quarterly Journal of Economic Research, **3**, 339–351.

[21] Krolzig, H.M. (2004) *Constructing turning point chronologies with Markov-switching vector autoregressive models: the euro-zone business cycle*, in Mazzi GL and Savio G (eds), Monographs of Official Statistics, Luxembourg: European Commission, 147–190.

[22] Lam, P.S. (1990) *The Hamilton model with a general autoregressive component: estimation and comparison with other models of economic time series*, Journal of Monetary Economics **26**, 409–432.

[23] Sichel, D.E. (1994) *Inventories and the three phases of business cycle*, Journal of Business and Economic Statistics, **12** (3), 269–277.

Appendix

Proof of Theorem 3.1. Using the measurement equation in (3.2) , we can easily compute

$$E(\mathbf{y}_t) = \sum_{j=0}^{r} \mathbf{\Lambda}_j E(\boldsymbol{\xi}_{t-j}) = (\sum_{j=0}^{r} \mathbf{\Lambda}_j)\boldsymbol{\pi}$$

and

$$E(\mathbf{y}_t)E(\mathbf{y}_{t+h}^{'}) = (\sum_{j=0}^{r} \mathbf{\Lambda}_j)\boldsymbol{\pi}\boldsymbol{\pi}^{'}(\sum_{j=0}^{r} \mathbf{\Lambda}_j^{'}) = (\sum_{j=0}^{r} \mathbf{\Lambda}_j)\mathbf{D}\mathbf{P}_\infty(\sum_{j=0}^{r} \mathbf{\Lambda}_j^{'})$$

where $\mathbf{D} = diag(\pi_1, \ldots, \pi_M)$. For every $h \geq r > 0$, we get

$$\begin{aligned}
E(\mathbf{y}_t\mathbf{y}_{t+h}^{'}) &= E[\sum_{j=0}^{r} \mathbf{\Lambda}_j\boldsymbol{\xi}_{t-j} + \mathbf{\Sigma}(\boldsymbol{\xi}_t \otimes \mathbf{I}_K)\mathbf{u}_t, \sum_{i=0}^{r} \boldsymbol{\xi}_{t+h-i}^{'}\mathbf{\Lambda}_i^{'} + \mathbf{u}_{t+h}^{'}(\boldsymbol{\xi}_{t+h}^{'} \otimes \mathbf{I}_K)\mathbf{\Sigma}^{'}] \\
&= \sum_{i=0}^{r}\sum_{j=0}^{r} \mathbf{\Lambda}_j E(\boldsymbol{\xi}_{t-j}\boldsymbol{\xi}_{t+h-i}^{'})\mathbf{\Lambda}_i^{'} \\
&= \sum_{i=0}^{r}\sum_{j=0}^{r} \mathbf{\Lambda}_j E(\boldsymbol{\xi}_{t-j}\boldsymbol{\xi}_{t-j+h-i+j}^{'})\mathbf{\Lambda}_i^{'} \\
&= \sum_{i=0}^{r}\sum_{j=0}^{r} \mathbf{\Lambda}_j \mathbf{D}\mathbf{P}^{h-i+j}\mathbf{\Lambda}_i^{'} \\
&= \sum_{i=0}^{r}\sum_{j=0}^{r} \mathbf{\Lambda}_j \mathbf{D}\mathbf{Q}^{h-i+j}\mathbf{\Lambda}_i^{'} + \sum_{i=0}^{r}\sum_{j=0}^{r} \mathbf{\Lambda}_j \mathbf{D}\mathbf{P}_\infty\mathbf{\Lambda}_i^{'}
\end{aligned}$$

because $h + r \geq h - i + j \geq h - r \geq 0$ for every $i, j = 0, \ldots, r$. Here we have used the well-known property $E(\boldsymbol{\xi}_t\boldsymbol{\xi}_{t+h}^{'}) = \mathbf{D}\mathbf{P}^h$ for every $h \geq 0$ (where we set $\mathbf{P}^h = \mathbf{I}_M$ for $h = 0$). Thus

$$\mathbf{\Gamma}_{\mathbf{y}}(-h) = cov(\mathbf{y}_t, \mathbf{y}_{t+h}) = E(\mathbf{y}_t\mathbf{y}_{t+h}^{'}) - E(\mathbf{y}_t)E(\mathbf{y}_{t+h}^{'}) = (\sum_{j=0}^{r} \mathbf{\Lambda}_j\mathbf{D}\mathbf{P}^j)\mathbf{Q}^h(\sum_{i=0}^{r} \mathbf{P}^{-i}\mathbf{\Lambda}_i^{'})$$

for every $h \geq r > 0$, and taking the transpose gives the result. Here we have used the relations $\mathbf{Q}^{h-i+j} = \mathbf{P}^{h-i+j} - \mathbf{P}_\infty = \mathbf{P}^j(\mathbf{P}^h - \mathbf{P}_\infty)\mathbf{P}^{-i} = \mathbf{P}^j\mathbf{Q}^h\mathbf{P}^{-i}$ as $\mathbf{P}^n\mathbf{P}_\infty = \mathbf{P}_\infty\mathbf{P}^n = \mathbf{P}_\infty$ and $\mathbf{Q}^n = \mathbf{P}^n - \mathbf{P}_\infty$ for every $n \geq 1$. $\square$

Proof of Theorem 3.3. For every $h \geq r > 0$, we get

$$\mathbf{\Gamma}_{\mathbf{y}}(-h) = E(\boldsymbol{y}_t\boldsymbol{y}_{t+h}^{'}) - (\sum_{j=0}^{r} \mathbf{\Lambda}_j)\mathbf{D}\mathbf{P}_\infty(\sum_{i=0}^{r} \mathbf{\Lambda}_i^{'})$$

and

$$E(\boldsymbol{y}_t\boldsymbol{y}_{t+h}^{'}) = (\sum_{j=0}^{r} \mathbf{\Lambda}_j)\mathbf{D}\mathbf{P}_\infty(\sum_{j=0}^{r} \mathbf{\Lambda}_j^{'}) + \sum_{j=0}^{r}\sum_{i=0}^{r} \widetilde{\mathbf{\Lambda}}_j E(\boldsymbol{\delta}_{t-j}\boldsymbol{\delta}_{t-j+h-i+j}^{'})\widetilde{\mathbf{\Lambda}}_i^{'},$$

by using the State-Space representation in (3.3). Then it follows that

$$\mathbf{\Gamma_y}(-h) = \sum_{j=0}^{r}\sum_{i=0}^{r} \widetilde{\mathbf{\Lambda}}_j \widetilde{\mathbf{D}} (\mathbf{F}^{'})^{h-i+j} \widetilde{\mathbf{\Lambda}}_i^{'} = (\sum_{j=0}^{r} \widetilde{\mathbf{\Lambda}}_j \widetilde{\mathbf{D}} (\mathbf{F}^{'})^{j})(\mathbf{F}^{'})^{h} (\sum_{i=0}^{r} (\mathbf{F}^{'})^{-i} \widetilde{\mathbf{\Lambda}}_i^{'})$$

hence

$$\mathbf{\Gamma_y}(h) = \mathbf{A}^{'} \mathbf{F}^h \mathbf{B}$$

where $\mathbf{A}^{'} = \sum_{i=0}^{r} \widetilde{\mathbf{\Lambda}}_i \mathbf{F}^{-i}$ and $\mathbf{B} = \sum_{j=0}^{r} \mathbf{F}^j \widetilde{\mathbf{D}} \widetilde{\mathbf{\Lambda}}_j^{'}$, which we assume to be nonzero matrices. Now apply Theorem 3.2 with $p = 0$, $q = r > 0$ and $M - 1$ instead of M as $\mathbf{F}$ is $(M-1) \times (M-1)$. $\square$

Proof of Theorem 3.4. The stable VAR(1) process $(\boldsymbol{\delta}_t)$ possesses the vector MA(∞) representation $\boldsymbol{\delta}_t = F(L)^{-1}\mathbf{w}_t$. Since the inverse matrix polynomial can be reduced to the inverse of the determinant, that is, $|F(L)|^{-1}$, and the adjoint matrix $F(L)^*$, we have $\boldsymbol{\delta}_t = |F(L)|^{-1}F(L)^*\mathbf{w}_t$. Inserting this transformed state equation into the measurement equation in (3.3) and multiplying by the determinant of $F(L)$ yield

$$|F(L)|(\mathbf{y}_t - \boldsymbol{\mu}_\mathbf{y}) = \sum_{j=0}^{r} \widetilde{\mathbf{\Lambda}}_j F(L)^* L^j \mathbf{w}_t + \widetilde{\mathbf{\Sigma}}(F(L)^* \otimes \mathbf{I}_K)(\mathbf{w}_t \otimes \mathbf{I}_K)\mathbf{u}_t + \ |F(L)|\mathbf{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{u}_t$$

which is a stable VARMA whose autoregressive lag polynomial is scalar, and where the orders of the stable VARMA are as in the statement. $\square$

Proof of Theorem 4.1. Set $\mathbf{x}_t = \sum_{j=0}^{r} \mathbf{\Lambda}_j \boldsymbol{\xi}_{t-j} + \mathbf{\Sigma}(\boldsymbol{\xi}_t \otimes \mathbf{I}_K)\mathbf{u}_t$. For every $h \geq r > 0$, we have

$$\begin{aligned} cov(\mathbf{x}_{t+h}, \mathbf{y}_t) &= cov(\boldsymbol{\phi}(L)(\boldsymbol{\xi}_{t+h} \otimes \mathbf{I}_K)\mathbf{y}_{t+h}, \mathbf{y}_t) \\ &= \boldsymbol{\phi}(L)[E(\boldsymbol{\xi}_{t+h}) \otimes cov(\mathbf{y}_{t+h}, \mathbf{y}_t)] \\ &= \boldsymbol{\phi}(L)(\boldsymbol{\pi} \otimes \mathbf{I}_K)[1 \otimes cov(\mathbf{y}_{t+h}, \mathbf{y}_t)] \\ &= B(L)\mathbf{\Gamma_y}(h) \end{aligned} \tag{A.1}$$

where $B(L) = \boldsymbol{\phi}(L)(\boldsymbol{\pi} \otimes \mathbf{I}_K)$ is a $K \times K$ matrix lag polynomial of degree p. Since the process is second-order stationary, the above formula implies that $cov(\mathbf{x}_{t+h}, \mathbf{y}_t)$ is time invariant. Using the unrestricted State-Space representation (4.3) and the relation $\boldsymbol{\delta}_{t+h} = \mathbf{F}^h \boldsymbol{\delta}_t + \sum_{j=0}^{h-1} \mathbf{F}^j \mathbf{w}_{t+h-j}$, we have

(A.2)

$$\begin{aligned} \mathbf{x}_{t+h} =& (\sum_{j=0}^{r} \mathbf{\Lambda}_j)\boldsymbol{\pi} + \sum_{j=0}^{r} \widetilde{\mathbf{\Lambda}}_j \mathbf{F}^h \boldsymbol{\delta}_{t-j} + \sum_{i=0}^{h-1}\sum_{j=0}^{r} \widetilde{\mathbf{\Lambda}}_j \mathbf{F}^i \mathbf{w}_{t+h-i-j} + \widetilde{\mathbf{\Sigma}}[(\mathbf{F}^h \boldsymbol{\delta}_t) \otimes \mathbf{I}_K]\mathbf{u}_{t+h} \\ &+ \sum_{i=0}^{h-1} \widetilde{\mathbf{\Sigma}}[(\mathbf{F}^i \mathbf{w}_{t+h-i}) \otimes \mathbf{I}_K]\mathbf{u}_{t+h} + \mathbf{\Sigma}(\boldsymbol{\pi} \otimes \mathbf{I}_K)\mathbf{u}_{t+h} \end{aligned}$$

By (A.2), for every $h \geq r > 0$, we obtain

$$
\begin{aligned}
(3.6)\qquad cov(\mathbf{x}_{t+h}, \mathbf{y}_t) &= cov((\sum_{j=0}^{r} \mathbf{\Lambda}_j)\boldsymbol{\pi} + \sum_{j=0}^{r} \widetilde{\mathbf{\Lambda}}_j \mathbf{F}^h \boldsymbol{\delta}_{t-j}, \mathbf{y}_t) = \sum_{j=0}^{r} \widetilde{\mathbf{\Lambda}}_j \mathbf{F}^h cov(\boldsymbol{\delta}_{t-j}, \mathbf{y}_t) \\
&= \sum_{j=0}^{r} \widetilde{\mathbf{\Lambda}}_j \mathbf{F}^{h-j} E(\boldsymbol{\delta}_t \mathbf{y}_t^{'}) = \mathbf{A}' \mathbf{F}^h \mathbf{B}
\end{aligned}
$$

where $\mathbf{A}^{'} = \sum_{j=0}^{r} \widetilde{\mathbf{\Lambda}}_j \mathbf{F}^{-j}$ and $\mathbf{B} = E(\boldsymbol{\delta}_t \boldsymbol{y}_t^{'})$, as requested. Now we see that $E(\boldsymbol{\delta}_t \mathbf{y}_t^{'})$ is time invariant as $cov(\mathbf{x}_{t+h}, \mathbf{y}_t)$ is. Collecting formulae (A.1) and (A.3) gives the result. □

Proof of Theorem 4.3. We have $\boldsymbol{\delta}_t = F(L)^{-1}\mathbf{w}_t$ as usual. Substituting the last formula in the State-Space representation of the initial process obtained in the same manner as in (4.3), we get

$$
\begin{aligned}
(A.4)\qquad \mathbf{A}(L)(\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{y}}) &= \sum_{j=0}^{r} \widetilde{\mathbf{\Lambda}}_j F(L)^{-1} L^j \mathbf{w}_t + \sum_{i=0}^{q} \widetilde{\mathbf{\Theta}}_i (F(L)^{-1} \otimes \mathbf{I}_K)(\mathbf{w}_t \otimes \mathbf{I}_K) L^i \mathbf{u}_t \\
&+ \sum_{i=0}^{q} \mathbf{\Theta}_i (\boldsymbol{\pi} \otimes \mathbf{I}_K) L^i \mathbf{u}_t
\end{aligned}
$$

where $\widetilde{\mathbf{\Theta}}_i$ is given by the usual construction applied to $\mathbf{\Theta}_i = (\mathbf{\Theta}_{1i} \ldots \mathbf{\Theta}_{Mi})$ for $i = 0, \ldots, q$. Premultiplying (A.4) by $|F(L)|$ yields

$$
\begin{aligned}
(A.5)\qquad |F(L)|\mathbf{A}(L)(\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{y}}) &= \sum_{j=0}^{r} \widetilde{\mathbf{\Lambda}}_j F(L)^{*} L^j \mathbf{w}_t + \sum_{i=0}^{q} \widetilde{\mathbf{\Theta}}_i (F(L)^{*} \otimes \mathbf{I}_K)(\mathbf{w}_t \otimes \mathbf{I}_K) L^i \mathbf{u}_t \\
&+ |F(L)| \sum_{i=0}^{q} \mathbf{\Theta}_i (\boldsymbol{\pi} \otimes \mathbf{I}_K) L^i \mathbf{u}_t.
\end{aligned}
$$

Now the regularity conditions of the statement mean that $\mathbf{A}(L)$ is invertible, that is, $\mathbf{A}(L)^{*}\mathbf{A}(L) = |\mathbf{A}(L)|\mathbf{I}_K$. Premultiplying (A.5) by $\mathbf{A}(L)^{*}$, we get the VARMA(p^{*}, q^{*}) representation, with $p^{*} \leq M + Kp - 1$ and $q^{*} \leq M + (K-1)p + \max\{r, q+1\} - 2$ (use the fact that the degree of $|\mathbf{A}(L)|$ is Kp):

$$
\begin{aligned}
&|F(L)||\mathbf{A}(L)|(\mathbf{y}_t - \boldsymbol{\mu}_{\mathbf{y}}) = \sum_{j=0}^{r} \mathbf{A}(L)^{*} \widetilde{\mathbf{\Lambda}}_j F(L)^{*} L^j \mathbf{w}_t \\
&+ \sum_{i=0}^{q} \mathbf{A}(L)^{*} \widetilde{\mathbf{\Theta}}_i (F(L)^{*} \otimes \mathbf{I}_K)(\mathbf{w}_t \otimes \mathbf{I}_K) L^i \mathbf{u}_t + |F(L)|\mathbf{A}(L)^{*} \sum_{i=0}^{q} \mathbf{\Theta}_i (\boldsymbol{\pi} \otimes \mathbf{I}_K) L^i \mathbf{u}_t
\end{aligned}
$$

which is a stable model as required in the statement. □