A unified theory for ARMA models with varying coefficients: One solution fits all

M. Karanasos[†▷], A. Paraskevopoulos[‡], A. Magdalinos[*], A. Canepa[⋆]
[†]Brunel University London, [‡]University of Piraeus, [*]University of Southampton, [⋆] University of Turin

March 5, 2020

Abstract

For the large family of ARMA models with variable coefficients we provide an explicit and computationally tractable solution representation, which yields the fundamental properties of such processes, including the Wold-Crámer decomposition and the covariance structure. These results are founded on a banded Hessenbergian representation of the Green's function, built up solely of the autoregressive coefficients of the model. A generic condition, also expressed in terms of the Green's function, guarantees the convergence of the above mentioned properties. This condition is in line with the asymptotic stability and efficiency of such processes, while, their invertibility is guaranteed by an analogous condition, but now the Green's function is built up of the moving average coefficients. As a consequence, we illustrate mathematically a structural asymmetry between constant and 'time dependent' coefficient models, that is in the latter ones the backward and forward asymptotic efficiency differ structurally from one another. An alternative approach to the Hessenbergian solution representation is described by a simple procedure for manipulating polynomials with variable coefficients. The practical significance of the theoretical results in this work is illustrated with an application to U.S. inflation data. The main finding is that inflation persistence increased after 1976, whereas from 1986 onwards the persistence declines and stabilizes to even lower levels than the pre-1976 period.

Keywords: ARMA process, time variable coefficients, Green's function, Hessenbergians, asymptotic stability, asymptotic efficiency, Wold decomposition, invertibility, structural breaks, skew multiplication, time varying persistence.

JEL Classification: C13, C22, C32, E17, E31, E58.

[▷]M. Karanasos, A. Paraskevopoulos and A. Magdalinos are the lead authors, having defined the theoretical models, and having derived the theoretical results. A. Canepa is the fourth author, who carried out the empirical research.
We would like to thank S. Dafnos, who contributed to the interpretation and synthesis of the work, which led to an earlier version of the paper, and P. Koutroumpis, who provided the empirical section for this earlier version of the paper.

We gratefully acknowledge the helpful conversations we had with L. Giraitis, and G. Kapetanios in the preparation of the paper. We would also like to thank R. Baillie, M. Brennan, L. Bauwens, A. Demos, W. Distaso, D. van Dijk, C. Francq, P. Fryzlewicz, E. Guerre, C. Gourieroux, M. Guidolin, A. Harvey, C. Hommes, E. Kyriazidou, S. Leybourne, P. Marsh, P. Minford, C. Robotti, W. Semmler, R. Smith, T. Teräsvirta, E. Tzavalis, P. Zaffaroni, and J-M Zakoian for suggestions and comments which greatly improved many aspects of the paper. We are grateful to seminar participants at Aarhus University (Creates), CREST, London School of Economics (Department of Statistics), Erasmus University (Econometric Institute), Imperial College (Tanaka Business School), Queen Mary University of London (School of Economics and Finance), University of Essex (Business School), BI Norwegian Business School, University of Copenhagen (Department of Economics), Birkbeck College, University of London (Department of Economics, Mathematics and Statistics), University of Nottingham (Granger Centre for Time Series Econometrics), University of Manchester (Business School), Cardiff University (Business School), Lancaster University (Management School), University of Reading (ICMA Centre), King's College London (Business School), University of Turin (Department of Economics and Statistics), Athens University of Economics and Business (Department of Economics) and University of Piraeus (Department of Economics).

We have also benefited from the comments given by participants at the: SNDE 27th, 26th, 25th, 24th, 22nd and 21st Annual Symposiums (Federal Reserve Bank of Dallas, March 2019; Keio University, Tokyo, March 2018; ESSEC Business School, Paris, March 2017; University of Alabama, March 2016; The City University of New York, April 2014; University of Milano-Bicocca, March 2013, respectively), 5th and 4th AMEF Conferences (University of Macedonia, April 2019, 2018, respectively), 2nd DSSR Conference (University of Milano-Bicocca, February 2019), 5th, 3rd, 2nd and 1st IAAE Conferences (University of Montreal, June 2018; University of Milano-Bicocca, June 2016; University of Macedonia, Greece, June 2015; Queen Mary University of London, June 2014, respectively), 12th, 11th, 8th and 7th CFE Conferences (University of Pisa, December 2018, December 2014; Senate House, University of London, December 2013, December 2017), 17th, 13th and 12th CRETE Conferences (Tinos, Crete and Naxos, July 2018, 2014 and 2013, respectively), SEM Conference (MIT, Boston, July 2017), 69th and 67th ESEM European Meetings (University of Geneva, August 2016; Toulouse, France, August 2014, respectively), Final and 1st RastaNews Annual Conferences (Catholic University, Milan, January 2016; Palazzo Clerici, Milan, January 2014, respectively), 35th ISF Conference (Erasmus University, Rotterdam, July 2014), 2nd ISNPS Conference (Cadiz, Spain, June 2014), NSVSM Conference (University of Paderborn, July 2014), 9th and 8th BMRC-QASS conferences on Macro and Financial Economics (Brunel University London, May 2014 and 2013, respectively), and 3rd Humboldt-Copenhagen Conference on Financial Econometrics (Humboldt University, Berlin, March 2013).

We would also like to thank four anonymous referees. In particular, the comments of one of these referees led us to develop Sections 5.2.4, 5.3, 5 and Appendix D.

1 Introduction

Modelling time series processes with variable coefficients has received considerable attention in recent years in the wake of several financial crises and high volatility due to frequent changes in the market. Justification for the use of such structures can be found in Timmermann and van Dijk (2013); for example, for the dynamic econometric modelling and forecasting in the presence of instability see the papers in the corresponding Journal of Econometrics special issue, i.e., Pesaran et al. (2013), and Koop and Korobilis (2013). 'Time varying' systems are extensively applied by practitioners, and their importance is widely recognized (see, for example, Granger, 2007 and 2008).[1] Crucial advances in both the theory and the empirics for these structures are the works by Whittle (1965), Abdrabbo and Priestley (1967), Rao (1970), Hallin (1978, 1986), Kowalski and Szynal (1990, 1991) and Grillenzoni (1993, 2000). [2]

This research provides a general framework for the study of autoregressive moving average models with time varying coefficients and heteroscedastic errors (hereafter, TV-ARMA). There are two large classes of stochastic processes: the ones with deterministically and those with stochastically time varying coefficients. Both types have been widely applied in many fields of research, such as economics, finance and engineering, but traditionally they have been examined separately. The new framework unifies them by showing that one solution fits all. More specifically, we obtain explicit and computationally feasible solution representations that generate the fundamental properties of this family of models, whereas the useful tool which is traditionally used to obtain such representations, that is the characteristic polynomials, is not applicable when time variation is present (see, for details, Hallin, 1978, and Grillenzoni, 1990).

Miller (1968) established an explicit representation of a particular solution associated with the above mentioned type of models, which is expressed in terms of the (one-sided) Green's function, including the multivariate case. In a series of papers, Hallin (1978, 1984, 1986, 1987) employs the one sided Green's function (briefly Green's function) to describe the fundamental properties of the multivariate moving average models in connection with the invertibility and the asymptotic efficiency of these models. The Green's function itself was not explicitly described, however. In order to overcome this difficulty recursive methods were employed to compute it (see, for example, Grillenzoni, 2000, and Azrak and Melard, 2006). An explicit representation of the Green's function depends upon the availability of a fundamental set of solutions (as also marked by Hallin in the first of the above cited references) whose elements (known as fundamental or linearly independent solutions) are explicitly expressed and easily handled, which is an ongoing research issue. In this work we provide such a fundamental solution set yielding an explicit form of the Green's function as a banded Hessenbergian (determinant of a banded Hessenberg matrix), here also termed as principal determinant. The entries of the principal matrix determinant are the autoregressive (resp. moving average) coefficients evaluated at consecutive time instances. The first fundamental solution sequence is represented by the principal determinant. The remaining fundamental solutions are expressed in terms of the principal determinant and therefore as banded Hessenbergians too. As a consequence, an explicit form of the general solution is expressed in terms of the fundamental solutions and a particular solution, all represented by Hessenbergians. It leads to an efficient interpretation of the general solution, as a decomposition into four structurally different parts (see eq. (14)). In this interpretation the fundamental solutions could be portrayed as autoregressive coefficients of the observed random variables.

The Hessenbergian solution structure yields a simple (necessary and sufficient) condition, which guarantees the asymptotic backward and forward stability of the process. This solution structure also generates easily handled analytic representations (involving infinite sums) for the fundamental properties of the aforementioned models (such as forecasts, the unconditional first and second moments, the Wold-Crámer decomposition and impulse response functions). A generic condition ensuring the above mentioned properties (see Section 5.2) is grounded on the absolutely summability of terms generated by the principal determinant, whose nonzero entries are the autoregressive coefficients evaluated at consecutive time instants. The invertibility of the process is also guaranteed by the absolutely summability of terms generated by the principal determinant, while, in this case, its nonzero entries are moving average

[1]A growing empirical literature in macroeconomics is testimony to their importance. See, for example, Evans and Honkapohja (2001, 2009).

[2]See also Francq and Gautier (2004a, 2004b). We refer to the introduction of Azrak & Mélard (2006) and Alj et al. (2016) for further references.

coefficients of the model evaluated at corresponding time instants (see Section 5.4).

We illustrate mathematically one of the focal points in Hallin's (1986) analysis concerning the asymptotic efficiency of such models. Namely, that in a time varying setting two forecasts with identical forecasting horizons, but at different times, yield different mean squared errors. This implies that the backward asymptotic efficiency (when the vector of the observed random variables shifts into the remote past) is, in the general case, different from the forward (termed by Hallin Granger-Andersen) one, that is when the time at which a forecast is intended moves into the far future. In this case we provide sufficient conditions for the forward asymptotic boundedness and uniform boundedness of the mean square forecasting error. Equally important, we demonstrate how the linear algebra techniques, used to obtain the general solution, are equivalent to a simple procedure for manipulating polynomials with variable coefficients. In order to do so we employ the expression of the Green's function as a banded Hessenbergian in conjunction with the so called skew multiplication operator or symbolic operator (see, for example, Hallin, 1986, and Mrad and Farad, 2002).

Banded Hessenbergians are computationally tractable due to the linear running time for their calculation (see the Appendix A.1). Compact solution representations of banded Hessenbergians, established in Marrero and Tomeo (2012, 2017) and Paraskevopoulos and Karanasos (2019), can be applied to derive a compact representation for the principal determinant. These results modernize and enhance the explicit representations of 'time varying' models and their fundamental properties, by compact representations.

The definition of the principal determinant seems to be a somewhat arbitrary choice in our analysis. It naturally arises, however, from Paraskevopoulos' work (2014), who (by introducing the so called infinite Gauss-Jordan elimination algorithm) provides banded Hessenbergian representations for the p elements of a fundamental solution set associated with the homogeneous part of a TV-ARMA(p, q) model (see also Appendix A.1). We should highlight the fact that in the present paper a self-contained proving process is demonstrated without invoking the infinite elimination algorithm. To sum up, the unified theory enable us to study time series models of linear form with either stochastically or deterministically varying coefficients, thus allowing us to make theoretical comparisons between these two large classes of models.

This paper concludes with an empirical application on inflation persistence in the United States. Our main contribution is that we measure persistence by employing a 'time varying' model of inflation dynamics grounded on statistical theory. In particular, we estimate an autoregressive process with abrupt structural breaks and we compute an alternative measure of second-order time dependent persistence, which distinguishes between changes in the dynamics of inflation and its volatility and their persistence. Our main conclusion is that persistence increased after 1976, whereas from 1986 onwards it declines and stabilizes to even lower levels than the pre-1976 period. Our results are in line with those in Cogley and Sargent (2002), who find that the persistence of inflation in the United States rose in the 1970s and remained high during this decade, before starting a gradual decline from the 1980s until the early 2000s.

The outline of the paper is as follows. Section 2 introduces the notation used in the paper followed by the principal determinant. The next Section presents the general solution for an extensive family of 'time varying' ARMA models. In Section 4, we obtain a necessary and sufficient condition which guarantees the asymptotic stability of these processes. Section 5 derives explicit formulas for their fundamental properties including the problem of producing asymptotically efficient forecasts, and it deals with the invertibility issue. In Section 6 we examine an AR model where the drift and the autoregressive coefficients are stochastically varying. In Section 7 we introduce a simple procedure for manipulating polynomials with variable coefficients. The next Section gives an illustrative example with abrupt structural breaks and proposes a new measure of time varying persistence. Section 9 presents an empirical study on inflation persistence. Finally, Section 10 contains some concluding remarks. Note that throughout the paper all the proofs are delegated to the Appendix.

2 Time Varying ARMA

The aim of this section is to provide the definition of TV-ARMA models along with the main mathematical tool for their analysis. That is the banded Hessenbergian representation of the Green's function, associated with such models, which is referred to as principal determinant.

2.1 The Model

This Subsection introduces suitable notation and defines the basic process. Throughout the paper we adhere to the following conventions: ($\mathbb{Z}_{>0}$) $\mathbb{Z}$, and $\mathbb{Z}_{\geq 0}$ stand for the sets of (positive) integers, and non-negative integers respectively. Similarly, ($\mathbb{R}_{>0}$) $\mathbb{R}$ and $\mathbb{R}_{\geq 0}$ stands for the set of (positive) real numbers, and non-negative real numbers respectively. Let $(\Omega, \mathcal{F}, P)$ denote a complete probability space adapted to some filtration, $\{\mathcal{F}_s\}$, which is a non-decreasing sequence of σ-subfields of $\mathcal{F}$, that is $\mathcal{F}_{s-1} \subseteq \mathcal{F}_s$ for all $s \in \mathbb{Z}$, such that $\mathcal{F}_s \subseteq \mathcal{F}$. The space of P-equivalence classes of finite random variables with finite p-order moment is indicated by $L_p(\Omega, \mathcal{F}, P)$ (briefly L_p). In particular, L_2 stands for a Hilbert space of random variables with finite first and second moments.

A time varying ARMA(p, q) model ($p, q \in \mathbb{Z}_{\geq 0}$) with time dependent coefficients and heteroscedastic errors, hereafter termed TV-ARMA(p, q), is defined as

$$y_t = \varphi(t) + \sum_{m=1}^{p} \phi_m(t) y_{t-m} + u_t, \; t \in \mathbb{Z} \tag{1}$$

with moving average term u_t given by

$$u_t = \varepsilon_t + \sum_{l=1}^{q} \theta_l(t) \varepsilon_{t-l},$$

where $\varphi(t)$ is the time varying drift and $\{\varepsilon_t\}$ is a martingale difference defined on L_2 with time varying variance: $0 < \sigma^2(t) \leq M$, for each t and some $M \in \mathbb{R}_{>0}$. The forcing term in eq. (1) is assigned to be the time varying drift plus the moving average term: $\upsilon_t = \varphi(t) + u_t$.[3]

We have relaxed the assumption of homoscedasticity (see also, among others, Singh and Peiris, 1987, Kowalski and Szynal, 1990, 1991, and Azrak and Mélard, 2006), which is likely to be violated in practice and we allow ε_t to follow, for example, a stochastic volatility or a time varying GARCH type of process (see, for example, the earlier versions of the current paper (see Karanasos et al., 2014c [available upon request], and Karanasos et al., 2017) or we allow for abrupt structural breaks in the variance of ε_t (see the example in Section 8).

The TV-ARMA(p, q) model nests both the TV-AR(p) as a special case when $q = 0$ and the ARMA(p, q) specification when the drift, the autoregressive and moving average coefficients, and the variances are all constants, adopting for this purpose the conventional identifications: $\varphi(t) = \varphi, \; \phi_m(t) = \phi_m, \; \theta_l(t) = \theta_l, \sigma^2(t) = \sigma^2$ for all t.

The relation between the process under consideration and its innovations is essentially described by the Wold-Crámer decomposition (see Section 5.2), which is the main analytical tool for studying the backward asymptotic efficiency of the model. In this case, the latest time-index of the observed random variables, denoted here by s, moves to the remote past ($s \to -\infty$), while the forecast time-point, denoted here by t, is kept fixed. The forward asymptotic efficiency of the model (so-called by Hallin, 1986, Granger-Andersen) is strongly related to the forecasting problem. It directs attention to the asymptotic properties of the mean square forecasting error (MSE for short), as the time t moves to the far future, while s, is kept fixed (see Section 5.3).

One of the goals of this work is to obtain the unique inverse of the time varying autoregressive (AR) polynomial associated with eq. (1) being denoted by $\Phi_t(B)$, where B is the backshift or lag operator (see Section 7). In a time varying environment, the usual procedure employs the Green's function instead of the characteristic polynomials, which are used in the time invariant case. However, this is an implicit representation, due to the absence of an explicit and computationally feasible representation of the Green's function. To overcome this difficulty, we introduce the principal determinant (see Section 2.2) and make use of the so called multiplication skew operator (see the analysis in Section 7.1).

We should also mention that Kowalski and Szynal (1991) used the product of companion matrices to obtain the Green's function. Paraskevopoulos and Karanasos (2019) capitalized on the connection between the product of companion matrices and 'time varying' stochastic difference equations but in the

[3]Notice that in our setting the time varying coefficients can depend on the length of the series as well, as in Azeak and Melard (2006); see the examples in Section 5.2.

opposite direction. That is, they went the other way around and by finding an explicit and compact representation of the fundamental solutions for TV-ARMA models, they obtained an analogous representation for the elements of the associated companion matrix product.

2.2 The Principal Determinant

In this Subsection, we introduce the main mathematical tool of this paper, which is employed for the explicit representation of TV-ARMA models. This is the principal determinant (denoted as $\xi(t,s)$) associated with eq. (1). We show that amongst the various representations of the Green's function via corresponding fundamental sets of solutions (see Section 3.1 and Appendix A.2 for further details), the principal determinant provides an explicit and computationally tractable representation. As a consequence, the main properties of 'time varying' models acquire analogous representations, as presented in Sections 4 and 5.

To distinguish scalars from vectors we adopt lower and uppercase boldface symbols within square brackets for column vectors and matrices respectively: $\mathbf{x} = [x_i]$, $\mathbf{X} = [x_{ij}]$. Row vectors are indicated within round brackets and usually appear as transpositions of column vectors: $\mathbf{x}' = (x_i)$.

For every pair $(t,s) \in \mathbb{Z}^2$ such as $k = t - s \geq 1$ the principal matrix associated with the AR part of eq. (1), is defined by

$$\mathbf{\Phi}_{t,s} = \begin{bmatrix} \phi_1(s+1) & -1 & & & & & & \\ \phi_2(s+2) & \phi_1(s+2) & \ddots & & & & & \\ \vdots & \vdots & \ddots & \ddots & & & & \\ \phi_p(s+p) & \phi_{p-1}(s+p) & \ddots & \ddots & \ddots & & & \\ & \phi_p(s+p+1) & \ddots & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & \phi_p(t-1) & \phi_{p-1}(t-1) & \cdots & \phi_1(t-1) & -1 \\ & & & & \phi_p(t) & \cdots & \phi_2(t) & \phi_1(t) \end{bmatrix} \tag{2}$$

here and in what follows empty spaces in a matrix have to be replaced by zeros. The matrix $\mathbf{\Phi}_{t,s}$ is a lower Hessenberg matrix of order k. It is also a banded matrix with total bandwidth $p+1$ (the number of its non-zero diagonals, i.e., the diagonals whose elements are not all identically zero), upper bandwidth 1 (the number of non-zero super-diagonals) and lower bandwidth $p-1$ (the number of non-zero sub-diagonals). In particular, the elements of $\mathbf{\Phi}_{t,s}$ are: (-1) occupying the entries of the superdiagonal, the values of the first autoregressive coefficient $\phi_1(\cdot)$ (from time $s+1$ to time t), occupying the entries of the main diagonal, the values of the $(1+r)$-th autoregressive coefficient $\phi_{1+r}(\cdot)$ for $r = 1, 2, \ldots, p-1$ (from time $s+1+r$ to time t), occupying the entries of the r-th sub-diagonal, and zero entries elsewhere. It is clear that for $p \geq k$, $\mathbf{\Phi}_{t,s}$ is a full lower Hessenberg matrix.

For every pair $(t,s) \in \mathbb{Z}^2$ with $k \geq 1$ we also define the principal determinant associated with eq. (2):

$$\xi(t,s) = \det(\mathbf{\Phi}_{t,s}). \tag{3}$$

Formally $\xi(t,s)$ is a lower Hessenbergian (determinant of a lower Hessenberg matrix; for details on Hessenbergians see, for example, the book by Vein and Dale, 1999). We further extend the definition of $\xi(t,s)$ by assigning the initial values:

$$\xi(t,s) = \begin{cases} 1 & \text{for} \quad t = s \\ 0 & \text{for} \quad t < s. \end{cases} \tag{4}$$

Under the initial values defined in (4), for each fixed s the sequence $\{\xi(t,s)\}_{t \geq s+1-p}$ turns out to be a particular solution of the homogeneous equation

$$y_t = \sum_{m=1}^{p} \phi_m(t) y_{t-m} \tag{5}$$

called principal fundamental sequence (see Appendix A.2, Proposition A1 applied for $m = 1$). It is well known that the Green's function, often designated by $H(t, s)$ for $t > s$, solves the homogeneous equation (5), assuming the prescribed values $H(s, s) = 1$ and $H(t, s) = 0$ for $t = s + 1 - p, ..., s - 1$ (see Lakshmikantham and Trigiante, 2002, Theorem 3.4.1 p. 87). Since the principal determinant $\xi(t, s)$ solves eq. (5) under the same initial values, the uniqueness of the solution for an initial value problem entails that for every arbitrary but fixed $s \in \mathbb{Z}$ it follows that $H(t, s) = \xi(t, s)$ for all $t : t \geq s + 1 - p$ (a proof from first principles is presented in Paraskevopoulos and Karanasos, 2019). Therefore we will make use of two different terminologies: principal determinant or Green's function.

We conclude this section with an illustrative example, concerning the ARMA(p, q) model with constant coefficients.

Example 1 *The AR polynomial* $\Phi(B) = 1 - \sum_{m=1}^{p} \phi_m B^m$ *associated with eq. (5), whenever* $\phi_m(t) = \phi_m$, *is explicitly expressed in terms of the characteristic values as* $\Phi(B) = \prod_{m=1}^{p}(1 - \lambda_m B)$. *The definition in eq. (3) gives the determinant of a banded Hessenbergian Toeplitz matrix (for details on Toeplitz matrices see, for example, the book by Gray, 2006) and satisfies the identity (usually called Widom's formula; see Widom, 1958)*

$$\xi_k = \begin{vmatrix} \phi_1 & -1 & & & & & & & \\ \phi_2 & \phi_1 & -1 & & & & & & \\ \phi_3 & \phi_2 & \phi_1 & \ddots & & & & & \\ \vdots & \vdots & \vdots & \ddots & \ddots & & & & \\ \phi_p & \phi_{p-1} & \phi_{p-2} & \ddots & \ddots & \ddots & & & \\ & \phi_p & \phi_{p-1} & \ddots & \ddots & \ddots & \ddots & & \\ & & \phi_p & \phi_{p-1} & \phi_{p-2} & \cdots & \phi_1 & -1 & \\ & & & \phi_p & \phi_{p-1} & \cdots & \phi_2 & \phi_1 & -1 \\ & & & & \phi_p & \cdots & \phi_3 & \phi_2 & \phi_1 \end{vmatrix} = \sum_{m=1}^{p} \frac{\lambda_m^{k+p-1}}{\prod_{\substack{n=1 \\ n \neq m}}^{p} (\lambda_m - \lambda_n)}, \qquad (6)$$

where $\xi(t, s)$ *is now denoted as* ξ_k, *since it depends only on the forecasting horizon* k ($k = t - s$ *indicates the order of the matrix too). The second equality of eqs. (6) follows (only if* $\lambda_m \neq \lambda_n$) *from standard results in ARMA models (see, for example, Karanasos, 2001, and in particular, eq. (2.6) in his Corollary 2; see also Hamilton, 1994, pp. 12-13).*

3 Solution Representations

The main result of this section is an explicit formula for the general solution of TV-ARMA processes.[4] More specifically, we obtain explicit and computationally feasible solution representations that generate the fundamental properties of this family of models, i.e. their first two conditional and unconditional moments, the Wold-Crámer decomposition and their covariance structure.

3.1 Homogeneous Solution

The general solution of eq. (5) on the domain $\mathbb{Z}_{s+1-p} = \{s + 1 - p, ..., s, ..., t, ...\}$ is denoted by $y_{t,s}^{hom}$ and it is expressed below in terms of the principal determinant, the autoregressive coefficients and the prescribed values y_{s+1-m} for $m = 1, 2, ..., p$, as follows:

$$y_{t,s}^{hom} = \sum_{m=1}^{p} \sum_{r=1}^{p+1-m} \phi_{m-1+r}(s + r)\xi(t, s + r) y_{s+1-m}. \qquad (7)$$

[4]In linear algebra there have been some isolated attempts to deal with the problem, which have been criticized on a number of grounds. For example, Mallik (1998) provides an explicit solution for the aforementioned equations, but it appears not to be computationally tractable (see also Mallik, 1997 and 2000). Lim and Dai (2011) point out that "although explicit solutions for general linear difference equations are given by Mallik (1998), they appear to be unmotivated and no methods of solution are discussed".

A proof of this result is provided in Appendix A.4, Proposition A3. To the extent of our knowledge, the expression in (7) is first recorded in earlier versions of this paper (see also Paraskevopoulos and Karanasos, 2019). The relative consequences of eq. (7) (such as eq. (14)) gains a more efficient representation in this latter version of the paper.

Let us call

$$\xi^{(m)}(t,s) \overset{\text{def}}{=} \sum_{r=1}^{p+1-m} \phi_{m-1+r}(s+r)\xi(t,s+r) \tag{8}$$

(superscripts within parentheses or brackets [e.g., $(.)^{(m)}$] designate the index position of the corresponding term [e.g., m-th term] of a sequence, so as to distinguish position indices from power exponents). The principal determinant is identified with $\xi^{(1)}(t,s)$, that is $\xi(t,s) \overset{\text{def}}{=} \xi^{(1)}(t,s)$. Applying eq. (8) with $s=t-1$ on account of $\xi(t,t)=1$ and $\xi(t,t-1+r)=0$ for $r=2,...,p+1-m$ (see eq. (4)), we get: $\xi^{(m)}(t,t-1)=\phi_m(t)$.

It turns out (see eq. (A.5) in Proposition A1 of Appendix A.2) that $\{\xi^{(m)}(t,s)\}_{t\geq s+1-p}$ is the solution sequence of eq. (5) under the prescribed initial values:

$$y_{s+1-m}=1 \text{ and } y_{s+1-r}=0 \text{ for } 1\leq r\leq p \text{ and } r\neq m. \tag{9}$$

Applying the expression in eq. (8) to the right-hand side of eq. (7), the homogeneous solution takes a more condensed form:

$$y_{t,s}^{hom} = \sum_{m=1}^{p} \xi^{(m)}(t,s)y_{s+1-m}. \tag{10}$$

Eqs. in (9) applied with $m=1$ yield the initial condition vector: $[y_{s+1-p},...,y_{s-1},y_s]=[0,...,0,1]$. With these initial values, the right-hand side of eq. (10) turns into $\xi^{(1)}(t,s)=\xi(t,s)$ and thus the homogeneous solution in eq. (10) recovers the principal fundamental sequence $\{\xi(t,s)\}_{t\geq s+1-p}$.

In Appendix A.2 (as a result of Proposition A2), it is shown that the set

$$\Xi_s = \{\xi^{(1)}(t,s),\xi^{(2)}(t,s),...,\xi^{(p)}(t,s):\ t\geq s+1-p\}$$

is a fundamental (or linearly independent) set of solutions associated with eq. (5). Moreover the m-th fundamental solution can be expressed as a single banded Hessenbergian too. The difference between any two of these fundamental solutions lies only in the first column (see Proposition A1(i) in the Appendix).

3.2 Particular Solution

A particular solution of eq. (1) subject to the initial values $y_s=y_{s-1}=...=y_{s+1-p}=0$ is given by

$$y_{t,s}^{par} = \sum_{r=s+1}^{t} \xi(t,r)[\varphi(r)+u_r]. \tag{11}$$

A proof of the above formula is demonstrated in Appendix A.3 (see Proposition A4). Eq. (11) has to be compared with the equivalent result presented in Miller (1968, p. 40, eqs. (2.8) and (2.9)).

Next we state a Proposition that we will use in the next Section. But first we will introduce the following definition:

Definition 1 *Let $\xi_q(t,r)$ and $\xi_{s,q}(t,r)$ be defined as follows*

$$\xi_q(t,r) \overset{\text{def}}{=} \xi(t,r)+\sum_{l=1}^{q}\xi(t,r+l)\theta_l(r+l),\ \textit{for } r=s+1,\ldots,t, \tag{12}$$

$$\xi_{s,q}(t,r) \overset{\text{def}}{=} \sum_{l=s-r+1}^{q}\xi(t,r+l)\theta_l(r+l),\quad \textit{for } r=s+1-q,\ldots,s.$$

As $\xi_q(t,r)$ is equal to $\xi(t,r)$ plus a sum of terms consisting of the first q 'lead' evaluations of $\xi(t,r)$, multiplied by corresponding moving average coefficients, it can also be expressed as a banded Hessenbergian (the proof is deferred to the online Appendix F.1, which is available upon request). The same applies to $\xi_{s,q}(t,r)$. Therefore, we refer to $\xi_q(t,r)$ and $\xi_{s,q}(t,r)$ as banded Hessenbergian coefficients. Notice also that $\xi_q(t,t) = 1$ and $\xi_q(t,t+l) = 0$ for all l such that $l \in \mathbb{Z}_{\geq 1}$, that coincide with the corresponding values of $\xi(t,r)$ for $r = t$ and $r > t$ respectively (see eq. (4)). Finally, for a pure AR model, that is when $q = 0$, $\xi_0(t,r) = \xi(t,r)$ and $\xi_{s,0}(t,r) = 0$.

Proposition 1 *The innovation component of the particular solution (11) can be decomposed into two parts as follows:*

$$\sum_{r=s+1}^{t} \xi(t,r)u_r = \sum_{r=s+1}^{t} \xi_q(t,r)\varepsilon_r + \sum_{r=s+1-q}^{s} \xi_{s,q}(t,r)\varepsilon_r. \tag{13}$$

A formal proof of this result is provided in Appendix A.3. If $s = t-1$ (or $k = 1$), the first sum in the right hand side of the above equation reduces to 1 and the second sum reduces to $\sum_{l=1}^{r} \theta_l(t)$, a result which is in line with Remark 1 below (see also the online Appendix F.1). In the first summation in the right-hand side of the above equation the length of the time interval, extending from $s+1$ to t, coincides with the forecasting horizon. In the second one the time interval extends from $s+1-q$ to s (all its time points belong to the initial data information sequence). The above equation together with the general solution in eq. (14) will be used to obtain the results in Section 5.

3.3 General Solution

The general solution of eq. (1) is the sum of the homogeneous solution in eq. (10) plus the particular solution in (11). Using the result of Proposition 1, a representation of the general solution is provided in the following Theorem:

Theorem 1 *The solution of eq. (1) under the prescribed values y_{s+1-m}, $m = 1, 2, ..., p$, is*

$$y_{t,s} = \underbrace{\sum_{m=1}^{p} \xi^{(m)}(t,s)y_{s+1-m}}_{\textit{Homogeneous Solution}(y_{t,s}^{\text{hom}})} + \underbrace{\sum_{r=s+1}^{t} \xi(t,r)\varphi(r)}_{\substack{\textit{Particular Solution:}\\ \textit{Drift Part}}} + \underbrace{\sum_{r=s+1}^{t} \xi_q(t,r)\varepsilon_r + \sum_{r=s+1-q}^{s} \xi_{s,q}(t,r)\varepsilon_r}_{\textit{Particular Solution: Innovation Part}}. \tag{14}$$

A proof of this result is given in Appendix A.4. In eq. (14), the general solution comprises four (summation) parts. The first sum (the homogeneous solution, see eq. (10)) is a linear combination of m fundamental solutions times the prescribed values, taken from the data information sequence. The second sum (the drift part of the particular solution, see eq. (11)) is formed by products involving the principal determinant $\xi(t,r)$ multiplied by the drift $\varphi(r)$. The elements of the third sum (the first part of the 'MA decomposition', see eq. (13)) are the 'lead' values of the banded Hessenbergian coefficients $\xi_q(t,s)$ times the corresponding errors. Finally, the elements of the fourth sum (the second part of the 'MA decomposition') are the 'lead' values of the banded Hessenbergian coefficient $\xi_{s,q}(t,s-q)$ times the corresponding errors.

Remark 1 *When $s = t-1$ (or $k = 1$) the general solution in Theorem 1 coincides with eq. (1). This is a consequence of the following statements: i) $\xi^{(m)}(t,t-1) = \phi_m(t)$ (see the discussion next to eq. (8)) and ii) $\sum_{r=t}^{t} \xi(t,r)[\varphi(r) + u_t] = \varphi(t) + u_t$.*

Remark 2 *Replacing the homogeneous solution part in (14) by (7), and its innovation part by the left-hand side of eq. (13), the general solution can be solely expressed as a linear combination of the Green's function evaluated at corresponding time instances. The general solution in (14) has to be compared with the corresponding result in (Agarwal (2000) p.77, eq. (2.11.8)), in which the fundamental solutions, denoted there by $v_i(k)$, are not in general explicitly expressed, but only in specific cases (see the examples in the previously cited reference).*

The methodology presented in this Section can be used in the study of infinite order autoregression models as well as in the case of the fourth order moments for time varying GARCH models. In the interest of brevity the detailed examination of the aforementioned models will be the subject of future papers. We should also mention that another mathematical tool of constant use in difference equations is the generalized continuous fraction approach (see, Van de Cruyssen, 1979). The concept of matrix continued fraction was introduced in Hallin (1984), whereas Hallin (1986) showed the close connection between the convergence of matrix continued fractions and the existence of dominated solutions for multivariate difference equations of order two.

Apart from the unified explicit and easily handled representation in eq. (14) another advantage of our solution is its generality. That is, in deriving it we do not make any assumptions on the time dependent coefficients. Therefore, it does not require a case by case treatment. In other words, we suppose that the law of evolution of the coefficients is unknown, in particular they may be stochastic (either stationary or non stationary) or deterministic. Therefore, no restrictions are imposed on the functional form of the time varying autoregressive and moving average coefficients. In the non stochastic case the model allows for known abrupt changes, smooth changes and mixtures of them. If the changes are smooth the coefficients can depend on an exogenous variable x_t or t or both. In the case of stochastically varying coefficients the model includes the generalized random coefficient (GRC) AR specification (see, for example, Glasserman and Yao, 1995, and Hwang and Basawa, 1998) as a special case or allows for Markov switching behaviour (see, for example, Hamilton, 1989 and 1994, chapter 22). In both aforementioned cases it allows for periodicity. We should also mention that the solution includes the case where the variable coefficients depend on the length of the series (see the example in Section 5.2).

3.4 Gegenbauer Functions as Hessenbergians

We conclude this Section with an example. We show how the Gegengbauer functions can be expressed as Hessenbergians. For a discussion of the Gegenbauer processes and their applications to economics and finance see Baillie (1996) and Dissanayake et al. (2018); see also, Giraitis and Leipus (1995) and Caporale and Gil-Alana (2011).

Example 2 *The Gegenbauer (or ultraspherical) functions, denoted by $c_j^{(d)}(\phi)$ (hereafter, for notational simplicity we use c_j), are defined to be the coefficients in the power-series expansion of the following function:*

$$(1-2\phi z+z^2)^{-d}=\sum_{j=0}^{\infty}c_j z^j,$$

for $|z|\leq 1$, $|\phi|\leq 1$, and $0<d<\frac{1}{2}$. It is well known that c_j can be computed in several ways. The easiest way to compute c_j (for $j\geq 2$) using computers is based on the following time varying second order difference equation:

$$c_j=2\phi\left(\frac{d-1}{j}+1\right)c_{j-1}-\left(2\frac{d-1}{j}+1\right)c_{j-2}\ \text{ for } j\geq 2,$$

with initial values $c_0=1$ and $c_1=2\phi d$ (see, for example, Baillie, 1996, Chung, 1996, and the references therein). Notice that in this case the two variable coefficients are functions of the index j. Applying Theorem 1 the following Proposition holds.

Proposition 2 *The j-th (for $j\geq 2$) Gegenbauer coefficient is given by:*

$$\begin{aligned} c_j &= \xi(j,1)c_1-d\xi(j,2)c_0 \ \textit{or} \\ c_j &= \xi(j,1)2\phi d-d\xi(j,2), \end{aligned}$$

where $\xi(j,1)$ *is a* $(j-1)$*-th order banded Hessenbergian (or tridiagonal matrix):*

$$\xi(j,1)=\begin{vmatrix} \phi(d+1) & -1 & & & & \\ -d & 2\phi\left(\frac{d+2}{3}\right) & -1 & & & \\ & -\frac{2d+1}{3} & \phi\left(\frac{d+3}{2}\right) & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -\left(2\frac{d-1}{j-1}+1\right) & 2\phi\left(\frac{d-1}{j-1}+1\right) & -1 \\ & & & & -\left(2\frac{d-1}{j}+1\right) & 2\phi\left(\frac{d-1}{j}+1\right) \end{vmatrix}.$$

As a result of the multi-linearity property of determinants, the following Hessenbergian form of the Gegenbauer polynomials is obtained.

Corollary 1 *The* j*-th Gegenbauer polynomial can be expressed as a* $(j-1)$*-th order banded Hessenbergian:*

$$c_j=\begin{vmatrix} d[2\phi^2(d+1)-1] & -1 & & & & \\ -2\phi d^2 & 2\phi\left(\frac{d+2}{3}\right) & -1 & & & \\ & -\frac{2d+1}{3} & 2\phi\left(\frac{d+2}{3}\right) & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -\left(2\frac{d-1}{j-1}+1\right) & 2\phi\left(\frac{d-1}{j-1}+1\right) & -1 \\ & & & & -\left(2\frac{d-1}{j}+1\right) & 2\phi\left(\frac{d-1}{j}+1\right) \end{vmatrix}.$$

4 Asymptotic Stability

As pointed out by Grillenzoni (2000) stability is a useful feature of stochastic models because it is a sufficient (although non necessary) condition for optimal properties of parameter estimates and forecasts. Since model (1) can be expressed in Markovian form, the stability condition allows many other stability properties, such as irreducibility, recurrence, regularity, non evanescence and tightness (see Grillenzoni, 2000 for details).

4.1 Stability Condition

The asymptotic stability of a model ensures that all the solutions approach a solution independently of the initial values (the effect of the initial values is gradually dying out) as $s\to-\infty$ (backward stability) or $t\to\infty$ (forward stability). This is a standard requirement for a model to yield long-run predictions (see Section 5 below). In both cases, (backward and forward stability), the initial condition vector $\mathbf{c}'=[c_m]_{1\le m\le p}$ consists of p arbitrary but fixed components c_m and the forecasting horizon $k=t-s$ is getting longer.

The TV-ARMA process in eq. (1) is backward asymptotically stable if for every t, arbitrary but fixed, the sequence of homogeneous solutions $\{y^{hom}_{t,s}\}_{s\le t}$ of eq. (5), expressed explicitly in (7), tends to zero as the past time s decreases without bound (that is $\lim_{s\to-\infty} y^{hom}_{t,s}=0$), for any initial condition vector $\mathbf{c}'=[c_m]_{1\le m\le p}$ such that $y_{s+1-m}=c_m$ for all $s<t$ and each $m: 1\le m\le p$. The latter amounts to the same to say that the initial condition vector $\mathbf{c}$ moves further to the past ($s\to-\infty$), while the present time t is kept fixed. The TV-ARMA process in eq. (1) is forward asymptotically stable if for any initial condition vector $\mathbf{c}$ at present time s (arbitrary but fixed), then $y^{hom}_{t,s}$ tends to zero in some sense as the future time t increases without bound ($t\to\infty$). A sufficient and necessary stability condition for eq. (1) is presented in the following Theorem (the proof is given in Appendix B).

Theorem 2 *i) Let the autoregressive coefficients* $\phi_m(t)$ *be deterministic. If* $\sup_t|\phi_m(t)|<\infty$ *for all* m *such that* $1\le m\le p$*, then a necessary and sufficient condition for the TV-ARMA model in eq. (1) to be backward asymptotically stable is:* $\lim_{s\to-\infty}\xi(t,s)=0$ *for each* t*. Moreover, the condition* $\lim_{t\to\infty}\xi(t,s)=0$ *for each* s *is necessary and sufficient for the TV-ARMA model to be forward asymptotically stable.*

ii) Let the autoregressive coefficients $\phi_m(t)$ be stochastic. If $\sup_t \mathbb{E}(\phi_m^2(t)) < \infty$ for $1 \leq m \leq p$, then a necessary and sufficient condition for the TV-ARMA process to be backward asymptotically stable is $\xi(t,s) \overset{a.s.}{\to} 0$, as $s \to -\infty$ (almost surly convergence) for each t. Moreover, the condition $\lim_{t\to\infty} \xi(t,s) \overset{a.s.}{=} 0$ is necessary and sufficient for the TV-ARMA model with stochastic coefficients to be forward asymptotically stable.[5]

Notice that in the above Theorem the conditions $\sup_t |\phi_m(t)| < \infty$ and $\sup_t \mathbb{E}(\phi_m^2(t)) < \infty$, respectively, are not required for the forward asymptotic stability. Moreover, the conditions in Theorem 2(ii) include the 'bounded random walk' of Giraitis et al. (2014). Properties such as stability characterize the statistical properties ($\sqrt{T}$ convergence and asymptotic normality, where T is the sample size) of least squares (LS) and quasi-maximum likelihood (QML) estimators of the time varying coefficients.[6]

In the time invariant case since $\xi(t,s)$ depends neither on t nor on s but only on their difference, that is the forecasting horizon, k, the stability condition in Theorem 2(i) reduces to $\lim_{k\to\infty} \xi_k = 0$, which holds if and only if all the roots λ_m in eq. (6) lie inside the unit circle.

Since the condition in Theorem 2 is necessary not only for stability but for the existence of the moments as well (see Section 5), in a companion paper, we provide an explicit compact representation for $\xi(t,s)$ (see Paraskevopoulos and Karanasos, 2019; see also Marrero and Tomeo, 2012, 2017).

Kowalski and Szynal (1991) and Grillenzoni (2000) derived sufficient conditions for the model in eq. (1) with zero drift and non stochastic coefficients to be second-order, that is for every t $\sum_{r=-\infty}^{t} \xi^2(t,r) < \infty$ to hold (we provide not only sufficient but necessary conditions as well in Proposition 5 below), which, therefore, are sufficient conditions for $\lim_{s\to-\infty} \xi(t,s) = 0$ for all t. These are presented in the following Proposition.

Proposition 3 *Two sufficient conditions for the stability condition in Theorem 2(i) are:*
i) The deterministically varying polynomial $\Phi_t(z^{-1}) = 1 - \sum_{m=1}^{p} \phi_m(t) z^{-m}$ is regular. That is, $\phi_m(t)$ are such that there exist the limits $\lim_{t\to\infty} \phi_m(t) = \phi_m$ and $\sum_{r=1}^{\infty} \varrho^{2r} < \infty$, where $\varrho = \varrho(\Phi) + \epsilon$, $\epsilon > 0$, $\varrho(\Phi) = \max\{|z_m|, \Phi(z_m^{-1}) = 0\}$ with $\Phi(z^{-1}) = 1 - \sum_{m=1}^{p} \phi_m z^{-m}$ (see eq. (8) in Kowalski and Szynal, 1991). [7]
ii) The deterministically varying polynomial $\Phi_t(z^{-1})$ should have roots whose realizations entirely lie inside the unit circle, with the exception, at most, of a finite set of points (see Proposition 1 in Grillenzoni, 2000).

The sufficient conditions in Proposition 3 are not, however, necessary, whereas they do not cover the case of periodic coefficients, see Grillenzoni (1990) or Karanasos et al. (2014,a,b).

4.2 Two Illustrative Examples

The first of the following examples concerns the logistic smooth transition AR(1) model (see, for example, Teräsvirta, 1994).

[5] Goldie and Maller (2000) provided sufficient conditons for the backward stability of an AR(1) model with stochastically varying coefficients, that is the a.s. convergence of the solution, i.e. $\sum_{t=1}^{\infty} \phi_1(1)\phi_1(2)\cdots\phi_1(t-1)\varepsilon_t < \infty$, a.s. (see also Bougerol and Picard, 1992). Recently, in a multivariate setting, Erhardsson (2014) showed that the only sufficient condition is: $|\phi_1(1)\phi_1(2)\cdots\phi_1(t)| \overset{a.s.}{\to} 0$ as $t \to \infty$, a result which is in line with our Theorem 2 ii). See also Section 6 below.

[6] Azrak and Mélard (2006) have considered the asymptotic properties of quasi maximum likelihood estimators for a large class of ARMA models with time dependent coefficients and heteroscedastic innovations. The coefficients and the variance are assumed to be deterministic functions of time, and depend on a finite number of parameters which need to be estimated. Other researchers have also considered the statistical properties of maximum likelihood estimators for very general non stationary models. For example, Dahlhaus (1997) has obtained asymptotic results for a new class of locally stationary processes, which includes TV-ARMA processes (see Azrak and Mélard, 2006, and the references therein).

[7] Kowalski and Szynal (1991) showed that $\varrho(\Phi)$ is the spectral radius of the matrix $\mathbf{\Phi} = \begin{bmatrix} 0 & 0 & \cdots & 0 & -\phi_m \\ 1 & 0 & \cdots & 0 & -\phi_{m-1} \\ 0 & 1 & \cdots & 0 & -\phi_{m-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -\phi_1 \end{bmatrix}$

(see page 75 in their paper).

Example 3 *In the above mentioned model the autoregressive coefficient is given by (we drop the subscript* 1*):* $\phi(t) = \phi_1 F(t;\gamma,\tau) + [1 - F(t;\gamma,\tau)]\phi_2$ *and* $F(t;\gamma,\tau) = [1 + e^{\gamma(t-\tau)}]^{-1}$, $\gamma \in \dot{\mathbb{R}}_{\geq 0}$, $\tau \in \mathbb{Z}$, *is the first-order logistic function. Clearly, if* $t > \tau$, *then* $F(\cdot) < 0.5$ *and regime* 2 *prevails, whereas if* $t < \tau$, *then* $F(\cdot) > 0.5$ *and regime* 1 *prevails. Let also* t_2 *be the value of* t *for which* $F(t;\gamma,\tau) = 0$ *if* $t \geq t_2$ *and thus* $\phi(t) = \phi_2$. *Similarly, let* t_1 *be the value of* t *for which* $F(t;\gamma,\tau) = 1$ *if* $t \leq t_1$ *and thus* $\phi(t) = \phi_1$ *(clearly* $t_2 > t_1$, *since* $F(\cdot)$ *is a decreasing function of time). For this model, if* $t \geq t_2$ *and* $s \leq t_1$ *then* $\xi(t,s) = \phi_1^{t_1-s+1}\prod_{r=t_1+1}^{t_2-1}\phi(r)\phi_2^{t-t_2+1}$. *Clearly,* $\lim_{s\to-\infty}\xi(t,s) = 0$ *if and only if* $|\phi_1| < 1$, *whereas* $\lim_{t\to\infty}\xi(t,s) = 0$ *if and only if* $|\phi_2| < 1$.

In the second example we consider the periodic $\mathrm{AR}(1;\ell)$ model.

Example 4 *In this model* $\ell \in \mathbb{Z}_{\geq 1}$ *is the number of seasons (i.e., quarters) and let* ϕ_r, $r = 1,\ldots,\ell$ *denote the periodically varying autoregressive coefficients. Let* $T \in \mathbb{Z}_{\geq 0}$ *be the number of periods (i.e., years). Then* $t = T\ell + r$ *is time expressed in terms of seasons (i.e., if* $\ell = 4$, $r = 4$ *and* $T = 1$ *then* $t = 8$ *quarters). If we want to forecast* $k\ell$ *seasons ahead, that is* $t - s = k\ell$ *or* $s = t - k\ell$, *then:* $\xi(t,s) = [\prod_{r=1}^{\ell}(\phi_r)]^k$. *Clearly* $|\phi_r| < 1$ *for all* r *is a sufficient but not necessary condition for* $\lim_{s\to-\infty}\xi(t,s) = 0$ *(or equivalently* $\lim_{k\to\infty}\xi(t,s) = 0$*). The necessary and sufficient condition is* $\left|\prod_{r=1}^{\ell}(\phi_r)\right| < 1$.[8]

5 Second Order Properties

Having specified a general method for manipulating the TV-ARMA type of models we turn our attention to a consideration of their fundamental time series properties. In particular, we will provide their thorough description by deriving explicit formulas for i) their multistep ahead linear predictors, the associated forecast errors and the mean square errors (MSE), ii) the first two unconditional moments, iii) the Wold-Crámer decomposition, and, therefore, impulse response functions (IRFs), and iv) their covariance structure. In this section we shall restrict ourselves to a treatment of the models with ARMA structure and deterministic coefficients; we will term these processes deterministic time varying (DTV). In Section 6 we present a process with stochastically varying coefficients, which incorporates the GRC and double stochastic AR models.

5.1 Conditional Moments

In what follows we present an explicit formula for the first conditional moment of the DTV-ARMA(p,q) model.

Taking the conditional expectation of eq. (14) with respect to the σ field $\mathcal{F}_s$ generated by the data information sequence $\{y_s, y_{s-1},\ldots\}$, the following Proposition follows immediately.

Proposition 4 *The* k*-step-ahead optimal (in* L_2*-sense) linear predictor of the DTV-ARMA*(p,q) *process is*

$$\mathbb{E}(y_{t,s}\,|\mathcal{F}_s) = \sum_{m=1}^{p}\xi^{(m)}(t,s)y_{s+1-m} + \sum_{r=s+1}^{t}\xi(t,r)\varphi(r) + \sum_{r=s+1-q}^{s}\xi_{s,q}(t,r)\varepsilon_r.$$

In addition, the forecast error for the above k*-step-ahead predictor,* $\mathbb{FE}_{t,s} = y_{t,s} - \mathbb{E}(y_{t,s}\mid\mathcal{F}_s)$, *and the associated* $\mathbb{MSE}_{t,s} = \mathbb{V}ar(\mathbb{FE}_{t,s})$ *are given by*

$$\mathbb{FE}_{t,s} = \sum_{r=s+1}^{t}\xi_q(t,r)\varepsilon_r,\ \ \mathbb{MSE}_{t,s} = \sum_{r=s+1}^{t}\xi_q^2(t,r)\sigma^2(r)$$

(we recall that $\xi_{s,q}(t,r)$ *and* $\xi_q(t,r)$ *have been introduced in Definition (1)).*

[8] For a study of the periodic stationarity of a random coefficient periodic autoregression (RCPAR) see, for example, Aknouche and Guerbyenne (2009).

The explicit form of the variance for a Gaussian process is necessary for the determination of the confidence intervals of $\mathbb{E}(y_t \,|\mathcal{F}_s\,)$.

Singh and Peiris (1987), Kowalski and Szynal (1990, 1991), and Grillenzoni (1990, 2000) obtained the optimal forecasts using prediction algorithms and recursive computation.

As an alternative to the above computational approaches, in Proposition 4 we provide an explicit formula for the optimal linear predictors, which enables us to study the asymptotic forecasting efficiency of the DTV-ARMA(p, q) process. We should also mention that Whittle (1965) showed that in general the linear least-square predictor obeys a recursion (see his eq. (12)) and gave a recursive method for obtaining its coefficients. In Section 6 we describe how we can apply this technique in conjunction with our methodology to derive the optimal forecasts when the coefficients are varying stochastically. In particular, we examine the GRC-AR specification and a model with coefficients that follows AR processes.

Next, we formulate one of the main arguments made by Hallin (1986), which states that unlike the time invariant case, in a time varying setting two MSEs with the same forecasting horizon, but at different time points, are no longer equal. With this in mind, consider two pairs of time points, say (t_1, s_1) and (t_2, s_2), such that $t_1 - s_1 = t_2 - s_2 = k$. The MSEs associated with these two time points are:

$$\mathbb{MSE}_{t_1,s_1} = \sum_{r=s_1+1}^{t_1} \xi_q^2(t_1, r)\sigma^2(r), \quad \mathbb{MSE}_{t_2,s_2} = \sum_{r=s_2+1}^{t_2} \xi_q^2(t_2, r)\sigma^2(r).$$

Shifting the time intervals in the summation intervals ($[s_1 + 1, t_1]$ and $[s_2 + 1, t_2]$) both to $[0, k-1]$, we get:

$$\mathbb{MSE}_{t_1,s_1} = \sum_{r=0}^{k-1} \xi_q^2(t_1, t_1 - r)\sigma^2(t_1 - r), \quad \mathbb{MSE}_{t_2,s_2} = \sum_{r=0}^{k-1} \xi_q^2(t_2, t_2 - r)\sigma^2(t_2 - r).$$

In a time varying environment, a comparison between $\mathbb{MSE}_{t_1,s_1}$ and $\mathbb{MSE}_{t_2,s_2}$, whenever $t_1 - s_1 = t_2 - s_2$, entails, in the general case, that : $\mathbb{MSE}_{t_1,s_1} \neq \mathbb{MSE}_{t_2,s_2}$. On the other hand, in the time invariant case $\xi_q(t, t-r)$, which will now be denoted by $\xi_r^{(q)}$, becomes $\xi_r^{(q)} = \xi_r + \sum_{l=1}^{q} \xi_{r-l}\theta_l$ (see eqs. (6) and (12)). In this case the two MSEs coincide, as being identical to:

$$\mathbb{MSE}_k = \sigma^2 \sum_{r=0}^{k-1} (\xi_r^{(q)})^2.$$

Having derived an explicit formula for the first two conditional moments in the next Section, we turn our attention to the unconditional moments.

5.2 Unconditional Moments

In this Section we present analytic formulae for the first and second unconditional moments for the DTV-ARMA family of processes coupled with sufficient and necessary conditions for their existence (see Proposition 5). In particular, we provide a generic condition (the absolute summability of the principal determinant), which ensures the existence of the first two unconditional moments. This condition also ensures the convergence in L_2 of the Wold-Crámer Decomposition (Theorem 3) and therefore the existence of the autocovariance function (Proposition 6) for DTV-ARMA models. In addition, it has a strong effect on the forward asymptotic efficiency and totally guarantees both the backward asymptotic stability and efficiency.

In what follows we assume that $\sup_{s \le t} |\varphi(s)| < \infty$ and $\sup_t |\theta_l(t)| < \infty$ for all t and l such that $1 \le l \le q$.

Proposition 5 *A sufficient condition for the DTV-ARMA(p, q) model to be second-order is*

$$\sum_{r=-\infty}^{t} |\xi(t, r)| < \infty, \textit{ for all } t \textit{ (absolute summability).}$$

Under the absolute summability condition the unconditional mean of the process y_t in eq. (1), that is $\mathbb{E}(y_t) = \lim_{s\to-\infty} \mathbb{E}(y_t\,|\mathcal{F}_s\,)$, with non stochastic coefficients, exists in $\mathbb{R}$ and is given by

$$\mathbb{E}(y_t) = \sum_{r=-\infty}^{t} \xi(t,r)\varphi(r). \tag{15}$$

Under the absolute summability condition the unconditional variance of the process y_t in eq. (1), that is $\mathbb{V}ar(y_t) = \lim_{s\to-\infty} \mathbb{MSE}_{t,s}$, with non stochastic coefficients, exists in $\mathbb{R}$ and is given by

$$\mathbb{V}ar(y_t) = \sum_{r=-\infty}^{t} \xi_q^2(t,r)\sigma^2(r). \tag{16}$$

Necessary conditions for the DTV-ARMA(p,q) model to be first and second order respectively are:

$$\lim_{s\to-\infty} \xi(t,s)\varphi(s) = 0 \text{ and } \lim_{s\to-\infty} \xi_q^2(t,s)\sigma^2(s) = 0 \text{ for all } t.$$

Moreover, the stability condition, that is $\lim_{s\to-\infty} \xi(t,s) = 0$, is sufficient for the above two limits to exist, while it is necessary for the absolute summability to hold.

The main logical connections between the conditions, employed in the above Proposition, are illustrated in the following commutative diagrams:

$$\begin{array}{ccccccccc}
\sum_{r=-\infty}^{t} \xi(t,r)\varphi(r) \in \mathbb{R} & \Longleftarrow & & & \sum_{r=-\infty}^{t} |\xi(t,r)| < \infty & & & \Longrightarrow & \sum_{r=-\infty}^{t} \xi_q^2(t,r)\sigma^2(r) \in \mathbb{R}_{\geq 0} \\
\Downarrow & & \Swarrow & & \Downarrow & & \Searrow & & \Downarrow \\
\lim_{s\to-\infty} \xi(t,s)\varphi(s) = 0 & \Longleftarrow & & & \lim_{s\to-\infty} \xi(t,s) = 0 & & & \Longrightarrow & \lim_{s\to-\infty} \xi_q^2(t,s)\sigma^2(s) = 0
\end{array} \tag{17}$$

Commutative Diagrams

A proof of Proposition 5 along with the Diagrams in (17) is provided in Appendix C.1. Notice that the mean is the same for both the AR and the ARMA processes.

Wold-Crámer Decomposition

In view of the general solution in eq. (14) we obtain: $y_t \overset{L_2}{=} \lim_{s\to-\infty} y_{t,s}^{par}$. The Wold-Crámer decomposition[9] (see Crámer, 1961) of the DTV-ARMA(p,q) model is a fundamental result, ensuring the backward asymptotic efficiency of these models as described analytically in the following Theorem.

Theorem 3 *Let the absolute summability condition in Proposition 5 hold. The Wold-Crámer decomposition is a solution of (1) of the form:*

$$y_t = \sum_{r=-\infty}^{t} \xi(t,r)\varphi(r) + \sum_{r=-\infty}^{t} \xi_q(t,r)\varepsilon_r. \tag{18}$$

A formal proof of this result is given in Appendix C.2. In the above Theorem y_t is a solution of eq. (1) decomposed into a non random part and a zero mean random part. In particular, $\mathbb{E}(y_t)$ is the non random part of y_t while $\lim_{s\to-\infty} \mathbb{FE}_{t,s}$ is the zero mean random part. Hallin (1978), Singh and Peiris (1987), Kowalski and Szynal (1991), Grillenzoni (2000), and Azrak and Mélard (2006) obtained the Wold-Crámer decomposition through recursion. In sharp contrast, eq. (18) in Theorem 3 provides an analytic formula for the one-sided MA representation.

[9]Since a non-stationary generalization of Wold's result was given by Cramér, it is referred to as Wold-Cramér decomposition.

Autocovariance Function
Another consequence of Theorem 1 is the following Proposition (the proof is contained in Appendix C.3), where we state expressions for the second moment structure of the DTV-ARMA(p, q) process.

Proposition 6 *Let the absolute summability condition in Proposition 5 hold. Then the time varying ℓ order autocovariance function $\gamma_t(\ell) = \mathbb{C}ov(y_t, y_{t-\ell})$, $\ell \in \mathbb{Z}_{\geq 0}$, is given by*

$$\gamma_t(\ell) = \sum_{r=-\infty}^{t-\ell} \xi_q(t, r)\xi_q(t-\ell, r)\sigma^2(r). \tag{19}$$

The time varying variance of y_t, that is $\gamma_t(0) = \mathbb{V}ar(y_t)$, is given by eq. (16). Notice again that for the AR process $\xi_0(t, r) = \xi(t, r)$, and that the absolute summability condition implies absolute summable autocovariances: $\sum_{\ell=0}^{\infty} |\gamma_t(\ell)| < \infty$ for all t.

From a computational viewpoint, the covariance structure of $\{y_t\}_t$ can be numerically evaluated by computing the banded Hessenbergian coefficients, $\xi_q(t, r)$ in eq. (12) and substituting these in eq. (19).

The next remark highlights the importance of the existence of finite second moments.

Remark 3 *Azrak and Mélard (2006) considered the asymptotic properties of QML estimators for the DTV-ARMA family of models where the coefficients depend not only on t but on T as well (see Alj et al., 2017, for the multivariate case). In their Theorem and Lemma 1 the existence of finite second moments was required. They also show that the dependence of the model with respect to T has no substantial effect on their conclusions except that a.s. convergence is replaced by convergence in probability since convergence in L_2 norm implies convergence in probability (see Lemma 1′ in their paper).*

We conclude this Section with two more examples and a discussion of forward asymptotic stability.

Two More Examples Next, we consider two examples concerning AR(1) processes with variable autoregressive coefficients, taken from Azrad and Mélard (2006).

Example 5 *In the first example, the autoregressive coefficient is a periodic function of time defined by*

$$y_t = \phi(t)y_{t-1} + \varepsilon_t,$$

where ε_t is a martingale difference defined on L_2 with constant variance σ^2. Moreover, the autoregressive coefficient is given by $\phi(t) = \beta_{t-n[t/n]}$, where $n \in \mathbb{Z}_{\geq 2}$ and $[x]$ is the larger integer less or equal to x (see also Dahlhaus, 1996). By specializing the results of Proposition 5 and Theorem 3, the Wold-Crámer decomposition (if and only if $|\beta| < 1$, where $\beta = \beta_0 \cdot \beta_1 \cdot \dots \cdot \beta_{n-1}$) is given by

$$y_t = \sum_{r=-\infty}^{t} \xi(t, r)\varepsilon_r,$$

with unconditional variance

$$\mathbb{V}ar(y_t) = \sigma^2 \sum_{r=-\infty}^{t} \xi^2(t, r),$$

where

$$\xi(t, r) = \beta^{[\frac{t-r}{n}]} \left(\prod_{j=0}^{t-r-1-n[\frac{t-r}{n}]} \beta_{t-j} - n\left[\frac{t-j}{n}\right] \right),$$

and, therefore

$$\sum_{r=-\infty}^{t} \xi^2(t, r) = \frac{1}{1-\beta^2} \sum_{r=t-n+1}^{t} \left(\prod_{j=0}^{t-r-1-n[\frac{t-r}{n}]} \beta_{t-j} - n\left[\frac{t-j}{n}\right] \right)$$

(see also eq. (4.2) in Azrak and Mélard, 2006).

Example 6 *In the second example (see Example 2 in Azrak and Mélard, 2006), the autoregressive coefficient is an exponential function of time given by*

$$\phi(t) = \begin{cases} \phi & \text{for} \quad t \leq 0, \\ \phi\lambda^{t/T} & \text{for} \quad t = 1, \ldots, T-1, \\ \phi\lambda & \text{for} \quad t \geq T, \end{cases}$$

where $T \in \mathbb{Z}_{\geq 1}$ is the sample size. For this case (assuming that $t > T$)

$$\xi(t,r) = \begin{cases} (\phi\lambda)^{t-r} & \text{for} \quad r \in [T,t], \\ (\phi\lambda)^t \phi^{-r} \lambda^{-\left(\frac{T+1}{2} + \frac{r(r-1)}{2T}\right)} & \text{for} \quad r = 1, \ldots, T-1, \\ \phi^{1-r}\xi(t,1) & \text{for} \quad r \leq 0. \end{cases}$$

The condition $|\phi| < 1$ (necessary and sufficient) entails:

$$\sum_{r=-\infty}^{t} \xi^2(t,r) = \frac{1}{1-\phi^2}\xi^2(t,1) + (\phi\lambda)^{2t} \sum_{r=1}^{T-1} \phi^{-2r} \lambda^{-(T+1+\frac{r(r-1)}{T})} + \frac{1-(\phi\lambda)^{2(t-T+1)}}{1-(\phi\lambda)^2}.$$

As pointed out by Azrak and Mélard (2006) the use of variable coefficients, which depend on the length of the series, is compatible with the approach of Dahlhaus (1997).

5.3 Forward Asymptotic Efficiency

As pointed out by Hallin (1986), if a researcher wants to study the 'causal' properties of the observed process, then he/she should examine the Wold-Crámer decomposition (see Theorem 3 in Section 5.2). If forecasting is the main objective, then the forecast obtained by the model should be asymptotically efficient in some sense. Of course the asymptotic forecasting properties of the model rely on its behaviour in the far future, whereas its causal properties involve its remote past only. If the processes is of time constant coefficients, these two issues coincide. In the case of time dependent coefficients, however, they apparently differ strongly.

To reiterate one of the main purposes, in building models for stochastic processes, is to provide convenient forecast procedures. The researcher would like to minimize (asymptotically) the MSE or in other words to achieve asymptotic efficiency. The asymptotic efficiency of a forecasting procedure can be defined in two alternative ways (seemingly, analogous to each other, but indeed basically different). The first one (termed by Hallin backward efficiency, see Definition 5.1 in his paper) is obtained by considering the asymptotic forecasting performance of a model as s tends to $-\infty$ (see Proposition 5). A model produces backward efficient forecasts if and only if it is an invertible model.

A more realistic approach to efficiency consists of considering the asymptotic behaviour of the mean square forecasting error as $t \to \infty$ for s being arbitrary but fixed. This forward efficiency concept is also called the Granger-Andersen efficiency (see Definition 5.2 in Hallin, 1986, and the references therein).

In the following Proposition, we give a weak condition, which guarantees the forward asymptotic boundedness of the mean square error.

Proposition 7 *Let $F(t,s) \stackrel{\text{def}}{=} \sum_{r=s+1}^{t} |\xi(t,r)|$ for $t > s$. If $\{F(t,s)\}_t$ is bounded, as a function of t for each fixed s, then the mean square error is also bounded, as a function of t. Equivalently, the boundedness of $\{F(t,s)\}_t$ entails that for each s either $\lim_{t\to\infty} \mathbb{MSE}_{t,s}$ exists in $\mathbb{R}_{\geq 0}$ or $\{\mathbb{MSE}_{t,s}\}_t$ oscillates with oscillation: $\Omega(s) = \inf_t(\sup_{r\geq t} \mathbb{MSE}_{r,s} - \inf_{r\geq t} \mathbb{MSE}_{r,s})$.*

In the following Corollary, we provide a stronger condition, which guarantees the forward asymptotic uniform boundedness of the mean square error.

Corollary 2 *Let $\sum_{r=-\infty}^{t} |\xi(t,r)| < \infty$ for each $t \in \mathbb{Z}$ (absolutely summability condition). Let us further call $F_t = \sum_{r=-\infty}^{t} |\xi(t,r)|$. If $\{F_t\}_{t\in\mathbb{Z}_{\geq 0}}$ is a bounded function, then the sequence $\{\mathbb{MSE}_s(t)\}_s$ defined by $\mathbb{MSE}_s(t) \stackrel{\text{def}}{=} \mathbb{MSE}_{t,s}$, is uniformly bounded.*

Proofs of the above results are provided in Appendix C.4. As the absolute summability condition in Corollary 2 holds, all the results previously stated (asymptotic stability, Wold-Crámer Decomposition, e.t.c) also hold.

5.4 Invertibility

In Section 5.1 we obtained the k-step-ahead optimal (in L_2 sense) linear predictor along with the associated MSE.

For some given time t optimal forecasts are possible if and only if an invertible model is used. The model is invertible if and only if the current value of the input martingale difference can be recovered as a (converging) linear combination of the present and past values of the observed series. Our Theorem 4 (see below) establishes a sufficient invertibility condition (this has to be compared with Theorem 3.1 in Hallin, (1986), where he derives such conditions for the deterministically time varying multivariate MA).

The TV-ARMA model is backward efficient if and only if it is an invertible model. If it is not an invertible model, then the corresponding mean square forecasting error is unbounded.

The TV-ARMA model in eq. (1) can be rewritten as

$$\varepsilon_t = -\varphi(t) - \sum_{l=1}^{q} \theta_l(t)\varepsilon_{t-l} + y_t - \sum_{m=1}^{p} \phi_m(t) y_{t-m}. \tag{20}$$

The principal matrix associated with the moving average part is defined by

$$\boldsymbol{\Theta}(t,s) = \begin{bmatrix} -\theta_1(s+1) & -1 & & & & & & \\ -\theta_2(s+2) & -\theta_1(s+2) & \ddots & & & & & \\ \vdots & \vdots & \ddots & \ddots & & & & \\ -\theta_q(s+p) & -\theta_{q-1}(s+q) & \ddots & \ddots & \ddots & & & \\ & -\theta_q(s+q+1) & \ddots & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & -\theta_q(t-1) & \theta_{q-1}(t-1) & \cdots & -\theta_1(t-1) & -1 \\ & & & & -\theta_q(t) & \cdots & -\theta_2(t) & -\theta_1(t) \end{bmatrix} \tag{21}$$

(for $t-s>q$). The matrix $\boldsymbol{\Theta}(t,s)$ has a similar structure to the principal determinant associated with the AR operator, $\boldsymbol{\Phi}(t,s)$, that is both matrices are banded lower Hessenberg. It is clear that for $q \geq t-s$, $\boldsymbol{\Theta}(t,s)$ is a full lower Hessenberg matrix.

For every pair $(t,s) \in \mathbb{Z}^2$ with $t-s \geq 1$ we define the principal determinant associated with eq. (21):

$$\vartheta(t,s) = \det(\boldsymbol{\Theta}(t,s)).$$

Formally $\vartheta(t,s)$ (and similarly to $\xi(t,s)$), is a banded Hessenbergian. We further extend the definition of $\vartheta(t,s)$ by assigning the initial values: $\vartheta(s,s)=1$ and $\vartheta(t,s)=0$ for $t<s$. Accordingly, $\vartheta(t,s)$ is the Green's function associated with the MA(q) operator.

In analogy with the definition of $\xi_q(t,r)$ (see eq. (12)), we define:

$$\vartheta_p(t,r) = \vartheta(t,r) - \sum_{m=1}^{p} \vartheta(t,r+m)\phi_m(r+m) \text{ for } r = s+1,\ldots,t.$$

In the following Theorem we give a sufficient condition for the DTV-ARMA(p,q) model to be invertible. This extends the standard invertibility condition for time invariant ARMA(p,q) models (see Brockwell and Davis (2016) p.76).

Theorem 4 *Let the absolute summability condition* $\sum_{r\to-\infty}^{t}|\vartheta(t,r)|<\infty$ *hold for each* t. *Then the DTV-ARMA*(p,q) *model is invertible, that is*

$$\varepsilon_t=-\sum_{r=-\infty}^{t}\vartheta(t,r)\varphi(r)+\sum_{r=-\infty}^{t}\vartheta_p(t,r)y_r$$

solves eq. (20).

The proof of Theorem 4 essentially repeats the arguments of the first part in the proof of Theorem 3; switching the roles of y_r and ε_r, and replacing $\xi(t,r)$ with $\vartheta(t,r)$, $\xi_q(t,r)$ with $\vartheta_p(t,r)$ and $\varphi(r)$ with $-\varphi(r)$.

Following laborious research work, the literature contains a diversity of 'time varying' specifications of linear form whose main time series properties either remain unexplored or have not been fully examined. Making progress in interpreting seemingly different models requires us to provide a common platform for the investigation of their time series properties. In this Section we have developed a theoretical foundation on which work in synthesizing these models can be done. With the help of a few detailed examples, i.e., smooth transition AR processes, periodic and cyclical formulations, we have demonstrated how to encompass various time series processes within our unified theory. The main strength of our general solution and the way we have expressed it is that researchers can use it for a multiplicity of problems. The significance of our methodology is almost self-evident from the large number of problems that it can solve. Our proposed approach allows us to handle 'time varying' models of infinite order, by introducing unbounded order linear difference equation of index p. This type of equations also yield Hessenbergian solutions, but involving full lower Hessenberg matrices. The latter enables us to obtain easily handled analytic solutions for the infinite order case along with the fundamental properties of corresponding models. An advantage of our technique is that it can be applied with ease, that is without any major alterations, in a multivariate setting and provides a solution to the problem at hand without adding complexity.

6 Stochastic Coefficients

In this Section we will examine, for simplicity instead of the ARMA, the AR process with stochastically time varying coefficients (STV-AR). All the proofs are presented in Appendix D.

6.1 Companion Form

The STV-AR(p) process, can be expressed as

$$y_t=\phi_{0t}+\phi_t'\mathbf{y}_{t-1}+\varepsilon_t, \tag{22}$$

where $\mathbf{y}_{t-1}=(y_{t-1},\ y_{t-2},\dots,y_{t-p})'$ is a $p\times 1$ vector of past observations of y_t, and $\phi_t=(\phi_{1t},\ \phi_{2t},\ \dots,\phi_{pt})'$ is a $p\times 1$ vector of the autoregressive random coefficients. Notice that we denote the STV coefficients by ϕ_{mt}, $m=0,\dots,p$, instead of $\phi_m(t)$, which was the notation used for the DTV ones.

It is well known that the model (22) can be written in a companion form

$$\mathbf{y}_t=\phi_{0t}+\mathbf{\Phi}_t\mathbf{y}_{t-1}+\varepsilon_t, \tag{23}$$

where $\phi_{0t}=(\phi_{0t}\ 0\dots0)'$, $\varepsilon_t'=(\varepsilon_t\ 0\dots0)'$, and the companion (square) matrix $\mathbf{\Phi}_t$ of order p associated to the vector ϕ_t is given by

$$\mathbf{\Phi}_t=\begin{pmatrix}\phi_{1t} & \phi_{2t} & \dots & \phi_{p-1,t} & \phi_{pt}\\ 1 & 0 & \dots & 0 & 0\\ 0 & 1 & \dots & 0 & 0\\ \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots\\ 0 & 0 & \dots & 1 & 0\end{pmatrix}. \tag{24}$$

That is, the STV-AR(p) process is converted to a p-dimensional vector STV-AR(1) model. For any set of p prescribed values $\mathbf{y}_{t-k}$, iterating eq. (23) yields

$$\mathbf{y}_{t,k} = \mathbf{C}_{t,k}\mathbf{y}_{t-k} + \sum_{r=1}^{k-1} \mathbf{C}_{t,r}(\phi_{0,t-r} + \varepsilon_{t-r}), \tag{25}$$

where $\mathbf{y}_{t,k} = (y_{t,k}, y_{t-1,k-1}, \dots, y_{t-(p-1),k-(p-1)})'$ and $\mathbf{C}_{t,k} = \prod_{r=0}^{k-1}(\mathbf{\Phi}_{t-r})$ is the product of companion matrices with initial matrix value $\mathbf{C}_{t,0} = \mathbf{I}$. It follows directly from the above equation and Theorem 1 (see also, for more details, Paraskevopoulos and Karanasos, 2019) that the p-dimensional square matrix $\mathbf{C}_{t,k}$ is given by

$$\mathbf{C}_{t,k} = \begin{pmatrix} \xi^{(1)}_{t,k} & \xi^{(2)}_{t,k} & \cdots & \xi^{(p)}_{t,k} \\ \xi^{(1)}_{t-1,k-1} & \xi^{(2)}_{t-1,k-1} & \cdots & \xi^{(p)}_{t-1,k-1} \\ \vdots & \vdots & \vdots\vdots & \vdots \\ \xi^{(1)}_{t-p+1,k-p+1} & \xi^{(2)}_{t-p+1,k-p+1} & \cdots & \xi^{(p)}_{t-p+1,k-p+1} \end{pmatrix},$$

or in other words the element occuping the $(n+1, m)$-th entry of the matrix $\mathbf{C}_{t,k}$ $(n = 0, \dots, p-1)$ is the m-th fundamental solution $\xi^{(m)}_{t,k-n}$. We recall that $\xi^{(1)}_{t,k}$ is given in eqs. (2) and (3), where $\xi^{(1)}_{t,k}$ is given in eqs. (2) and (3) (where now, $\phi_m(t)$ is replaced by ϕ_{mt}) and, similarly, $\xi^{(m)}_{t,k}$ is given in eq. (8).

6.2 Random Coefficients AR Model

In this Section we will examine the random coefficient AR(p) model (with acronym RC-AR). It is given by eq. (22) where: $\phi^*_t = (\phi_{0t}\ \phi_t)'$, $t = s+1, s+2\dots$, is an *i.i.d.* $(p+1)$-dimensional random vector of the coefficients, and the *i.i.d.* errors, $\{\varepsilon_t;\ t = s+1, s+2, ...\}$ are independent of the random drift and autoregressive coefficients. Let also ε_s be a random variable independent of everything else (the initial state).

Next denote by $|\cdot|$ the Euclidean norm on the space $\mathbb{R}^p$. Let $\mathbb{R}^{p\times p}$ be the space of $p\times p$ matrices with elements in $\mathbb{R}$ and denote by $||\cdot||$ the matrix norm induced by $|\cdot|$ (this is known as the spectral norm, and is equal to the largest singular value of the matrix).

Condition 1 $\xi_{t,s} \overset{a.s}{\to} 0$ *as* $t \to \infty$.

Theorem 5 *Consider the RC-AR(p) model. Under Condition 1 the following are equivalent:*

i) $\mathbf{y}_t$ *convergences in distribution as* $t \to \infty$,

ii) $\sum_{r=s+1}^{\infty} |\mathbf{C}_{r-1,s}\varepsilon_r| < \infty$ *a.s.*,

iii) $\sum_{r=s+1}^{t} \mathbf{C}_{r-1,s}\varepsilon_r$ *converges a.s. as* $t \to \infty$,

iv) $\mathbf{C}_{t-1,s}\varepsilon_t \overset{a.s.}{\to} 0$ *as* $t \to \infty$,

v) $\sup_{t=s+1,s+2,\dots} |\mathbf{C}_{t-1,s}\varepsilon_t| < \infty$ *a.s.*

6.3 The Generalized RC-AR Model

The Generalized random coefficients (GRC) AR(p) model, generalizes the RC process by allowing contemporaneous dependence between the vector of the drift and the autoregressive coefficients, that is ϕ^*_t, and the vector of the errors ε_t. It integrates the following AR processes (for details, see Hwang and Basawa, 1998):

I. Random coefficient model: $\phi_{mt} = \phi_m + \eta_{mt}$, where $\{\eta_{mt}\}$ is a sequence of *i.i.d.* random variables with $\mathbb{E}(\eta_{mt}) = 0$, and $\{\eta_{mt}\}$ is independent of $\{\varepsilon_\tau\}$ for all m and $t \neq \tau$. Notice also that if we set $\eta_{mt} = 0$, for all m and t, we get the ordinary AR(p) process.

II. Markovian bilinear model: $\phi_{mt} = \phi_m + \vartheta_m \varepsilon_t$.

III. Generalized Markovian bilinear model: $\phi_{mt} = \varphi_m + \vartheta_m \varepsilon_t^{r_m}$, $r_m \in Z_{>0}$, and $\varepsilon_t^{r_m}$ has finite moments $\mathbb{E}(\varepsilon_t^{r_m})$ for all m. Note that if we set $r_m = 1$, for all m, we get the Markovian bilinear model coefficients.

IV. Random coefficients exponential model: $\phi_{mt} = \varphi_m + (\vartheta_{m1} + \vartheta_{m2} e^{-\vartheta_{m3}\varepsilon_t^2})\varepsilon_t$.

In models I-IV, ϕ_m, φ_m, ϑ_m, ϑ_{m1} and ϑ_{m2} are time invariant coefficients.

In the next Proposition and Theorem we will make use of the following notation:

Notation 1

$$\begin{aligned} \mathbb{E}(y_t) &= y, \mathbb{E}(\phi_{mt}) = \phi_m, \mathbb{E}(\phi_{mt}\varepsilon_t) = \sigma_{m\varepsilon}, \ \textit{for } m = 0, \dots, p, \\ \mathbb{C}ov(\phi_{mt}, \phi_{nt}) &= \sigma_{mn} = \phi_{mn} - \phi_m \phi_n, \ \textit{for } m, n = 0, \dots, p. \end{aligned}$$

Condition 2 $|\sum_{m=1}^{p} \phi_m| < 1$.

Proposition 8 *Under Condition 2 the expected value of the GRC-AR(p) model is given by*

$$y = \frac{\phi_0}{1 - \sum_{m=1}^{p} \phi_m}.$$

Next let $\mathbf{\Phi}_t^{\otimes 2} = \mathbf{\Phi}_t \otimes \mathbf{\Phi}_t$ where $\otimes$ is the Kronecker product, and let $\lambda_{\max}[\mathbb{E}(\mathbf{\Phi}_t^{\otimes 2})]$ refer to the modulus of the largest eigenvalue of $\mathbb{E}(\mathbf{\Phi}_t^{\otimes 2})$.

Condition 3 $\lambda_{\max}[\mathbb{E}(\mathbf{\Phi}_t^{\otimes 2})] < 1$.

In the next Theorem we will also make use of the following notation:

Notation 2 *Let i) φ_{ij} denote the element occuping the (i,j) -th entry of the matrix $[\mathbf{I}_{p^2} - \mathbb{E}(\mathbf{\Phi}_t^{\otimes 2})]^{-1}$,*

$$ii) \ \sigma^2 = (\sigma_\varepsilon + \sigma_0)^2 + 2\sum_{m=1}^{p}(\sigma_{m\varepsilon} + \sigma_{m0}) + \sum_{n=1}^{p}\sum_{m=1}^{p}\sigma_{mn}, \tag{26}$$

and iii) $\mathbb{E}(\xi_{t,k}^{(m)}) = \xi_k^{(m)}$, for example $\xi_k^{(1)}$ is the k-dimensional Toeplitz determinant:

$$\xi_\ell^{(1)} = \begin{vmatrix} \phi_1 & -1 & & & & & \\ \phi_2 & \phi_1 & \ddots & & & & \\ \vdots & \vdots & \ddots & \ddots & & & \\ \phi_p & \phi_{p-1} & \ddots & \ddots & \ddots & & \\ & \phi_p & \ddots & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & & \phi_p & \phi_{p-1} & \cdots & \phi_1 & -1 \\ & & & & \phi_p & \cdots & \phi_2 & \phi_1 \end{vmatrix} \tag{27}$$

(we recall that $\xi_{t,k}^{(1)}$ is given in eqs. (2) and (3), where now, $\phi_m(t)$ is replaced by ϕ_{mt} and, similarly, $\xi_{t,k}^{(m)}$ is given in eq. (8)).

Theorem 6 *Under Condition 3 the covariance structure of the GRC-AR(p) model is given by*

$$\gamma(\ell) = \begin{cases} \varphi_{\ell+1,1}\sigma^2 & \textit{for} \quad \ell = 0, \dots, p-1 \\ \sum_{m=1}^{2} \xi_\ell^{(m)} \gamma(m-1) & \textit{for} \quad \ell \geq p. \end{cases}$$

Notice that Condition 2 is equivalent to the condition: $\lim_{\ell\to\infty}\xi_\ell^{(1)}=0$, which implies that $\lim_{\ell\to\infty}\xi_\ell^{(m)}=0$ for all m, and, therefore, ensures that $\lim_{\ell\to\infty}\gamma(\ell)=0$.

Corollary 3 *Consider the GRC-AR(2) model and let the following condition hold:*

$$1-\frac{2\phi_{12}\phi_1}{1-\phi_2}>\phi_{11}+\phi_{22}>0.$$

Under the above Condition the covariance structure of the GRC-AR(2) model is given by

$$\gamma(\ell)=\begin{cases}\varphi_{\ell+1,1}\sigma^2 & \text{for} \quad \ell=0,1\\ \sum_{m=1}^{2}\xi_\ell^{(m)}\gamma(m-1) & \text{for} \quad \ell\geq 2,\end{cases}$$

where

$$\begin{aligned}\varphi_{1,1} &= \frac{1-\phi_2^2}{(1-\phi_2^2)(1-\phi_{11}-\phi_{22})-2\phi_{12}\phi_1(1+\phi_2)},\\ \varphi_{2,1} &= \varphi_{2,1}=\frac{\phi_1(1+\phi_2)}{(1-\phi_2^2)(1-\phi_{11}-\phi_{22})-2\phi_{12}\phi_1(1+\phi_2)},\end{aligned}$$

and the tridiagonal matrices of order ℓ, that is $\xi_\ell^{(m)}$, $m=1,2$, are given by

$$\xi_\ell^{(1)}=\begin{bmatrix}\phi_1 & -1 & & & \\ \phi_2 & \phi_1 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \phi_1 & -1\\ & & & \phi_2 & \phi_1\end{bmatrix}, \quad \xi_\ell^{(2)}=\phi_1\xi_{\ell-1}^{(1)}.$$

6.4 Double Stochastic AR Models

In this Section we will investigate the double stochastic AR model, hereafter termed DS-AR (for double stochastic processes, and in particular ARMA processes with ARMA coefficients, see, for example, Grillenzoni, 1993, and the references therein). In this case the autoregressive coefficients in eq. (22), ϕ_{mt}, for $m=1,\ldots,p$, follow AR processes:

$$\phi_{mt}=\beta_{m0}+\sum_{l=1}^{p_m}\beta_{ml}\phi_{m,t-l}+e_{mt}, \tag{28}$$

where β_{m0} and β_{ml} are constant coefficients and $p_m\in\mathbb{Z}_{\geq 0}$ for all m. $\{e_{mt}\}$ are martingale differences defined on L_2, where e_{mt} and $\varepsilon_{t\pm b}$, $b\in\mathbb{Z}$, *are independent of each other for all* m, and t. For simplicity, we will assume that the drift in eq. (22) is time invariant. That is, $\phi_{0t}=\phi_0$ for all t.

For this class of models the results in Sections 5.1 and 5.2 (in the paper) can be easily modified by replacing the fundamental solutions with their respective (conditional and unconditional) expectations. More specifically, we have the following three Propositions followed by a Theorem.

Proposition 9 *The k-step-ahead optimal (in L_2-sense) linear predictor of the DS-AR(p) process is*

$$\mathbb{E}(y_{t,s}\,|\mathcal{F}_s)=\sum_{m=1}^{p}\mathbb{E}[\xi^{(m)}(t,s)\,|\mathcal{F}_s]y_{s+1-m}+\phi_0\sum_{r=s+1}^{t}\mathbb{E}[\xi(t,r)\,|\mathcal{F}_s].$$

In addition, the forecast error for the above k-step-ahead predictor, $\mathbb{FE}_{t,s}$, is given by

$$\mathbb{FE}_{t,s}=\phi_0\sum_{r=s+1}^{t}[\xi(t,r)-\mathbb{E}[\xi(t,r)\,|\mathcal{F}_s]]+\sum_{r=s+1}^{t}\xi(t,r)\varepsilon_r+\sum_{m=1}^{p}\{\xi^{(m)}(t,s)-\mathbb{E}[\xi^{(m)}(t,s)\,|\mathcal{F}_s]\}y_{s+1-m}.$$

The associated $\mathbb{MSE}_{t,s} = \mathbb{V}ar(\mathbb{FE}_{t,s}\,|\mathcal{F}_s\,)$ *is given by*

$$\mathbb{MSE}_{t,s} = \phi_0^2 \sum_{r=s+1}^{t} Var[\xi(t,r)\,|\mathcal{F}_s] + \sigma_\varepsilon^2 \sum_{r=s+1}^{t} [\;\mathbb{E}[\xi(t,r)^2\,|\mathcal{F}_s] + \sum_{m=1}^{p} Var[\xi^{(m)}(t,s)\,|\mathcal{F}_s] y_{s+1-m}^2.$$

(we recall that $\xi(t,r)$ *have been introduced in eqs. (2) and (3), where now* $\phi_m(t)$ *are replaced by* ϕ_{mt}*).*

Proposition 10 *A sufficient condition for the DS-AR(p) model to be first-order stationary is*

$$\sum_{r=-\infty}^{t} |\mathbb{E}[\xi(t,r)]| < \infty, \text{ for all } t \text{ (first-order absolute summability).}$$

Under the absolute summability condition the unconditional mean of the DS-AR(p) process, y_t*, that is* $\mathbb{E}(y_t) = \lim_{s\to-\infty} \mathbb{E}(y_t\,|\mathcal{F}_s\,)$*, exists in* $\mathbb{R}$ *and is given by*

$$\mathbb{E}(y_t) = \phi_0 \sum_{r=-\infty}^{t} \mathbb{E}[\xi(t,r)]. \tag{29}$$

A necessary condition for the absolute summability to hold is

$$\lim_{s\to-\infty} \mathbb{E}[\xi(t,s)] = 0.$$

Proposition 11 *Sufficient conditions for the DS-AR(p) model to be second-order stationary are the absolute summability and*

$$\sum_{r=-\infty}^{t} \mathbb{E}[\xi(t,r)^2] < \infty, \text{ for all } t \quad \text{(second-order summability).}$$

Under the two summability conditions the unconditional variance of the DS-AR(p) process, that is $\mathbb{V}ar(y_t) = \lim_{s\to-\infty} \mathbb{MSE}_{t,s}$*, exists in* $\mathbb{R}$ *and is given by*

$$\mathbb{V}ar(y_t) = \phi_0^2 \sum_{r=-\infty}^{t} \mathbb{V}ar[\xi(t,r)] + \sigma_\varepsilon^2 \sum_{r=s+1}^{t} \mathbb{E}[\xi(t,r)^2]. \tag{30}$$

A necessary condition for the second-order summability to hold is

$$\lim_{s\to-\infty} \mathbb{E}[\xi(t,s)^2] = 0.$$

Remark 4 *Notice that* $\lim_{s\to-\infty} \mathbb{E}[\xi(t,s)^2] = 0$ *is a sufficient condition for* $\lim_{s\to-\infty} \mathbb{E}[\xi(t,s)] = 0$*, and, hence, for* $\xi(t,s) \overset{a.s.}{\to} 0$ *as* $s \to -\infty$*.*

Theorem 7 *Let the two summability conditions in Propositions 10 and 11 hold. The Wold-Crámer decomposition is a solution of the DS-AR(p) model of the form:*

$$y_t = \sum_{r=-\infty}^{t} \xi(t,r)(\phi_0 + \varepsilon_r). \tag{31}$$

Remark 5 *Generally, it is very difficult to verify if the two summability conditions are fulfilled. Only some special cases allow to write explicit solutions (see, Andel, 1991, and the references therein). A sufficient condition for the absolute summability to hold is that* $\{\sum_{m=1}^{p} \phi_{mt}\}$ *belongs with probability one to the interval* $(-1,1)$*, nearly everywhere, that is, with the exception, at most, of a finite number of t (see, for example, Grillenzoni, 1993). Similarly, a sufficient condition for the square summability to hold is that* $\lambda_t^{(\max)}[\mathbf{\Phi}_t^{\otimes 2}] < 1$*, nearly everywhere, where* $\lambda_t^{(\max)}[\mathbf{\Phi}_t^{\otimes 2}]$ *refers to the modulus of the largest eigenvalue of* $\mathbf{\Phi}_t^{\otimes 2}$*.*

Theorem 8 *If all the autoregressive coefficients,* ϕ_{mt}*,* $m = 1,\ldots,p$*, are strictly stationary, then eq. (22) has a stationary solution of the type (31) if and only if*

$$\sum_{r=1}^{\infty} |\mathbb{E}[\xi(r,1)]| < \infty, \text{ and } \sum_{r=1}^{\infty} \mathbb{E}[\xi(r,1)^2] < \infty.$$

7 Time Varying Polynomials

In Section 3 we employed techniques of linear algebra in order to obtain the general solution of the TV-ARMA(p, q) model and its first two (conditional and unconditional) moments. The main mathematical tool used was the Hessenbergian determinant. Now that we have expressed the Green's function as a Hessenbergian we will see how the summation terms in the various equations in Sections 3 and 5 can be expressed as time varying polynomials.

Recall that B denotes the backshift (or lag operator), defined such that $By_t = y_{t-1}$. The time varying AR and moving average (MA) polynomial (backshift) operators associated with the TV-ARMA(p, q) model are denoted as:

$$\Phi_t(B) = 1 - \sum_{m=1}^{p} \phi_m(t) B^m, \;\; \Theta_t(B) = 1 + \sum_{l=1}^{q} \theta_l(t) B^l. \tag{32}$$

Under this notation eq. (1) can be written in a more condensed form

$$\Phi_t(B) y_t = \varphi(t) + \Theta_t(B)\varepsilon_t. \tag{33}$$

In the time invariant case one can employ the roots of the time invariant polynomial $\Phi(z^{-1})$ to obtain its general time series properties such as the Wold decomposition and the second moment structure. In a time varying environment, according to Grillenzoni (1990), the so called here principal fundamental sequence $\{\xi(t, s)\}_{t \geq s+1-p}$ cannot be obtained as in stationarity, that is by expanding in Taylor series the rational polynomial $\Phi_t^{-1}(B)$. As an alternative, Hallin (1986) used some results on difference operators involving the symbolic product of operators, which has also been termed by researchers, in the field of engineering, as the skew multiplication operator (see, for example, Mrad and Farag, 2002). Hence, now that we have at our disposal an explicit and computationally tractable representation of the Green's function as a banded Hessenbergian, coupled with the use of the time-domain noncommutative multiplication operation -which, as pointed out by Mrad and Farag (2002), is based on the manipulation of polynomial operators with time varying coefficients using operations restricted to the time domain- we are able to state some important Theorems in relation to the results in Sections 3 and 5.

7.1 The Skew Multiplication Operator

In a time varying environment, the time varying polynomial operators in eq. (32) can be manipulated by using the 'skew' multiplication operator "$\circ$" defined by

$$B^i \circ B^j = B^{i+j} \;\; \text{and} \;\; B^i \circ f(t) = f(t-i) B^i, \tag{34}$$

where $f(t)$ is a function of time. This time-domain multiplication operation is associative but noncummutative (see Karmen, 1988, Bouthellier and Ghosh, 1988, and Mrad and Farag, 2002). Using the properties of "$\circ$", from eq. (33), under the necessary and sufficient conditions in Proposition 5, we can obtain the unique inverse of $\Phi_t(B)$, that is $\Phi_t(B)^{-1} \circ \Phi_t(B) = 1$, where $\Phi_t(B)^{-1}$ is provided in eq. (40).

7.2 Polynomial Operators

Next, and equally important, we will provide a critical and essential further link between the linear algebra techniques used in Section 3 (to obtain the general solution of the TV-ARMA model) and the time varying polynomial approach, in which we make use of the 'skew' multiplication operator. Certainly, from an operational point of view, both are equally satisfying and recommendable. First, let us define the two time varying polynomial (backshift) operators associated with the general solution.

Definition 2 *I) Let* $\Xi_{t,p}^{(k)}(B)$ *be defined as follows*

$$\Xi_{t,p}^{(k)}(B) = 1 - \sum_{m=1}^{p} \xi^{(m)}(t, s) B^{k-1+m} \tag{35}$$

IIa) Let $\Xi_t^{(k)}(B)$ be defined as follows

$$\Xi_t^{(k)}(B) = \sum_{r=s+1}^{t} \xi(t,r)B^{t-r} \overset{\text{or}}{=} \sum_{r=0}^{k-1} \xi(t,t-r)B^r \tag{36}$$

IIb) The limit of $\Xi_t^{(k)}(B)$ as $k \to \infty$ is denoted by $\Xi_t(B)$.

Remark 6 *$\Xi_{t,p}^{(k)}(B)$ in Definition 2(I) is a polynomial of order $p+k-1$ associated with the homogeneous solution (10), and it is expressed in terms of the m fundamental solutions, defined in eq. (8).*
Notice that: i) $\Xi_{t,p}^{(1)}(B) = \Phi_t(B)$, since $\xi^{(m)}(t,t-1) = \phi_m(t)$ (see the discussion next to eq. (8)), and ii) under the stability condition in Theorem 2, $\lim_{k\to\infty} \Xi_{t,p}^{(k)}(B) = \Xi_{t,p}(B) = 1$, since $\lim_{s\to-\infty} \xi^{(m)}(t,s) = 0$ (see in Appendix B, Lemma B1).

Remark 7 *$\Xi_t^{(k)}(B)$ is a polynomial of order $k-1$ associated with the particular solution (11), and it is expressed in terms of the 'lead' values of the principal determinant, $\xi(t,t-k)$. Notice also that $\Xi_t^{(1)}(B) = 1$.*

Next we define two additional time varying polynomial operators associated with the innovation part of the particular solution (see eq. 13) and the Wold-Crámer decomposition (see eq. 18) respectively.

Definition 3 *I) Let $\Xi_{t,q}^{(k)}(B)$ be defined as follows*

$$\Xi_{t,q}^{(k)}(B) = \sum_{r=0}^{k-1} \xi_q(t,t-r)B^r + \sum_{r=k}^{k-1+q} \xi_{s,q}(t,t-r)B^r. \tag{37}$$

II) Let $\Xi_{t,q}(B)$ be defined as follows

$$\Xi_{t,q}(B) = \sum_{r=0}^{\infty} \xi_q(t,t-r)B^r.$$

Remark 8 *$\Xi_{t,q}^{(k)}(B)$ is a polynomial of order $q+k-1$ associated with the innovation part of the particular solution, and is expressed in terms of the 'lead' values of the two Hessenbergian coefficients, $\xi_q(t,t-k)$ and $\xi_{s,q}(t,t-k-q)$, which have been defined in Definition 1.*
Notice that, i) for the pure AR model $\Xi_{t,0}^{(k)}(B) = \Xi_t^{(k)}(B)$, since $\xi_0(t,t-r) = \xi(t,t-r)$ and the second summation in eq. (37) (adopting the convention $\sum_{r=k}^{k-1} a_r = 0$) vanishes. ii) $\Xi_{t,q}^{(1)}(B) = \Theta_t(B)$, since the first summation is equal to $\xi(t,t) = 1$, and the second summation is equal to $1-\Theta_t(B)$ (see the discussion next to Proposition 1).

7.3 General Solution

The next Proposition is an alternative to Proposition 1, expressed in terms of polynomial operators. Proposition 12 and Theorems 9, 10 can be deduced by applying the properties of the skew multiplication operator "$\circ$"(see eq. (34)). By analogy to the decomposition in eq. (13), we deduce the following two part-decomposition:

Proposition 12 *$\Xi_t^{(k)}(B) \circ u_t$ takes the following alternative expressions:*

$$\Xi_t^{(k)}(B) \circ u_t = \Xi_{t,q}^{(k)}(B)\varepsilon_t \overset{\text{or}}{=} \sum_{r=0}^{k-1} \xi_q(t,t-r)\varepsilon_{t-r} + \sum_{r=k}^{k-1+q} \xi_{s,q}(t,t-r)\varepsilon_{t-r}. \tag{38}$$

A proof of the above result is provided in Appendix E.
The first of the following Theorems (Theorem 9) is equivalent to Theorem 1.

Theorem 9 *The general solution of eq. (1) given in eq. (14), can be equivalently expressed, in terms of the polynomial operators given in Definitions 2 and 3, as:*

$$\Xi_{t,p}^{(k)}(B)y_t = \Xi_t^{(k)}(B) \circ [\varphi(t) + u_t] = \Xi_t^{(k)}(B) \circ \varphi(t) + \Xi_{t,q}^{(k)}(B)\varepsilon_t. \tag{39}$$

Clearly, the results for the pure AR model are obtained by setting $u_t = \varepsilon_t$ in the first equality of eq. (39) or by noticing that in the second equality of eq. (39) $\Xi_{t,0}^{(k)}(B) = \Xi_t^{(k)}(B)$ (see Definition 3).

Notice also that Proposition 4 can be expressed in terms of the time varying polynomial operators as well (results not reported).

In what follows we will make use of the infinite order polynomials, $\Xi_t(B)$ and $\Xi_{t,q}(B)$, which have been introduced in Definitions 2(IIb) and 3(II), respectively.

The next Theorem shows that $\Xi_{t,q}(B)$ is the time varying Wold-Crámer polynomial operator associated with the DTV-ARMA(p,q) model.

Theorem 10 *Under the absolute summability condition in Proposition 5, the Wold-Crámer operators are given by*

$$\Xi_{t,q}(B) = \Phi_t^{-1}(B) \circ \Theta_t(B) \quad \textit{and} \quad \Xi_t(B) = \Phi_t^{-1}(B) \quad \text{w.r.t} \quad \text{“}\circ\text{”}. \tag{40}$$

The Wold-Crámer decomposition in Theorem 3 can be written in terms of the Wold-Crámer operator as

$$y_t = \Xi_t(B) \circ \varphi(t) + \Xi_{t,q}(B)\varepsilon_t, \tag{41}$$

which of course implies that

$$\mathbb{E}(y_t) = \Xi_t(B) \circ \varphi(t).$$

Appendix E contains also the proofs of Theorems 9 and 10.

In the online Appendix Section G, we show how the time varying techniques, introduced here, that is, time varying polynomials with the time dependent coefficients expressed as Hessenbergians (coupled with the usage of the 'skew' multiplication operator), incorporate as a special case the standard approach to time series analysis which is based on characteristic polynomials. In the online Appendix Section H we present a summary of the 'time varying polynomials' results.

8 An Example

In this Section we will consider the AR(2) process with 2 deterministic abrupt breaks [DAB-AR$(2;2)$] at fixed points of time t_1 and t_2, where $t_1 > t_2$.

Example 7 *The process is defined by*

$$y_t = \begin{cases} \varphi_1 + \phi_{1,1}y_{t-1} + \phi_{2,1}y_{t-2} + \sigma_1 e_t & \textit{for} \quad t > t_1, \\ \varphi_2 + \phi_{1,2}y_{t-1} + \phi_{2,2}y_{t-2} + \sigma_2 e_t & \textit{for} \quad t_2 < t \leq t_1, \\ \varphi_3 + \phi_{1,3}y_{t-1} + \phi_{2,3}y_{t-2} + \sigma_3 e_t & \textit{for} \quad t \leq t_2, \end{cases} \tag{42}$$

where $e_t \sim$ i.i.d. $(0,1)$ for all t and $0 < \sigma_i^2 \leq M$, $i = 1,2,3$. Applying the results of Theorem 3 the following Corollary provides the general solution (at time $t_1 + l$, $l \in \mathbb{Z}_{\geq 0}$) of eq. (42). But first, we define the following matrices.

Definition 4 *Let the two tridiagonal matrices of order $r \in \mathbb{Z}_{\geq 1}$, denoted by $\Phi_r^{(j)}$, $j = 1,2$, be defined as*

$$\Phi_r^{(j)} = \begin{bmatrix} \phi_{1,j} & -1 & & & & \\ \phi_{2,j} & \phi_{1,j} & -1 & & & \\ & \phi_{2,j} & \phi_{1,j} & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \phi_{2,j} & \phi_{1,j} & -1 \\ & & & & \phi_{2,j} & \phi_{1,j} \end{bmatrix},$$

Definition 5 *i) The tridiagonal matrix (of order $l-r$) Φ_{t_1+l,t_1+r}, for $r=1,\ldots,l-1$ and $l\geq 1$, is defined as:*

$$\Phi_{t_1+l,t_1+r}=\Phi^{(1)}_{l-r},$$

where its determinant is $\xi(t_1+l,t_1+r)=|\Phi_{t_1+l,t_1+r}|$ with initial values $\xi(t_1+l,t_1+l)=1$ and $\xi(t_1,t_1+r)=0$
ii) The Hessenberg matrix Φ_{t_1+l,t_1-r}, for $r=0,\ldots,t_1-t_2$ and $r+l>0$, is defined as:

$$\Phi_{t_1+l,t_1-r}=\begin{bmatrix}\Phi^{(2)}_r & \overline{0}\\ \widetilde{0} & \Phi^{(1)}_l\end{bmatrix},$$

where (for $r,l\neq 0$) $\overline{0}$ is an $r\times l$ matrix of zeros except for -1 in its $r\times 1$ entry, and $\widetilde{0}$ is an $l\times r$ matrix of zeros except for $\phi_{2,1}$ in its $1\times r$ entry. Notice that Φ_{t_1+l,t_1-r} is a block square matrix of order $l+r$. Its determinant is $\xi(t_1+l,t_1-r)=|\Phi_{t_1+l,t_1-r}|$ with initial value $\xi(t_1,t_1)=1$.

Applying Theorem 1 to the DAB-AR(2; 2) model we obtain the following Corollary.

Corollary 4 *The general solution of y_{t_1+l} in eq. (42), taking on the initial values y_{t_2}, y_{t_2-1}, is given by*

$$y_{t_1+l,t_2}=\sum_{r=t_2+1}^{t_1+l}\xi(t_1+l,r)(\varphi_r+\varepsilon_r)+\xi(t_1+l,t_2)y_{t_2}+\phi_{2,1}\xi(t_1+l,t_2+1)y_{t_2-1},\tag{43}$$

8.1 Second Moment Structure

In this section we will examine the second moment structure of the DAB-AR (2; 2) model. To obtain the time varying variance of y_{t_1+l}, we will directly apply Corollary 4.

First, let $1-\phi_{1,i}B-\phi_{2,i}B^2=(1-\lambda_{1,i}B)(1-\lambda_{2,i}B)$, for $i=1,2,3$.
Assumption 2 (Second-Order). $|\lambda_{m,i}|<1$, $m=1,2$, for $i=1,3$.
Assumption 2 implies that the DAB-AR(2; 2)-process is second-order.

The following Proposition states expressions for the time varying variance of y_{t_1+l} in eq. (43).

Proposition 13 *Consider the model in eq. (42). Under Assumption 2, the $\mathbb{V}ar(y_{t_1+l})$ is given by*

$$\mathbb{V}ar(y_{t_1+l})=A_{t_1+l}\sigma_1^2+B_{t_1+l}\sigma_2^2+C_{t_1+l}\sigma_3^2,\tag{44}$$

where

$$A_{t_1+l}=\sum_{r=1}^{l}\xi^2(t_1+l,t_1+r),\ B_{t_1+l}=\sum_{r=0}^{t_1-t_2-1}\xi^2(t_1+l,t_1-r),$$
$$C_{t_1+l}=\frac{\left[(1-\phi_{2,3})\left(\xi^2(t_1+l,t_2)+\phi_{2,1}^2\xi^2(t_1+l,t_2+1)\right)+2\phi_{1,3}\xi(t_1+l,t_2)\phi_{2,1}\xi(t_1+l,t_2+1)\right]}{(1+\phi_{2,3})[(1-\phi_{2,3})^2-\phi_{1,3}^2]}.$$

Further, if in the above expression we set: $t_1=t_2$, and therefore $\phi_{m,1}=\phi_{m,2}$ for $m=1,2$, and $\sigma_1=\sigma_2$, then we obtain the $\mathbb{V}ar(y_{t_2+l})$, which is equivalent to the case of one break (notice that in this case $B_{t_2+l}=0$):

$$\mathbb{V}ar(y_{t_2+l})=A_{t_2+l}\sigma_2^2+C_{t_2+l}\sigma_3^2.$$

Finally, if in addition we set $l=0$ then we obtain the $\mathbb{V}ar(y_{t_2})$, which (since $A_{t_2}=0$, $\xi_{t_2,t_2}=1$, $\xi_{t_2,t_2+1}=0$) is the well known formula for the time invariant AR(2) model:

$$\mathbb{V}ar(y_{t_2})=\frac{(1-\phi_{2,3})\sigma_3^2}{(1+\phi_{2,3})[(1-\phi_{2,3})^2-\phi_{1,3}^2]}.$$

In the next Section we will show how the above results can be used to derive a time varying second-order measure of persistence.

8.2 Time Varying Persistence

The most often applied time invariant measures of first-order (or mean) persistence are the largest autoregressive root (LAR), and the sum of the autoregressive coefficients (SUM); see, e.g., Pivetta and Reis (2007). As pointed out by Pivetta and Reis in relation to the issue of recidivism by monetary policy its occurrence depends very much on the model used to test the natural rate hypothesis, i.e., the hypothesis that the SUM or the LAR for inflation data is equal to one. Obviously, if both measures ignore the presence of breaks then they will potentially under or over estimate the persistence in the levels. The LAR has been used to measure persistence in the context of testing for the presence of unit roots (see, for details, Pivetta and Reis, 2007).

In the following, we suggest a time varying second-order (or variance) persistence measure that is able to take into account the presence of breaks not only in the mean but in the variance as well. Fiorentini and Sentana (1998) argue that any reasonable measure of shock persistence should be based on the IRFs. For a univariate process x_t with *i.i.d.* errors, e_t, they define the persistence of a shock e_t on x_t as $P(x_t\,|e_t\,) \overset{\text{def}}{=} \dfrac{\mathbb{V}ar(\pi_t)}{\mathbb{V}ar(\varepsilon_t)}$. Clearly $P(x_t\,|e_t\,)$ will take its minimum value of one if x_t is white noise and it will not exist (will be infinite) for an I(1) process. It follows directly from eq. (44) that $P(y_{t_1+l}\,|\varepsilon_{t_1+l}\,) = \dfrac{\mathbb{V}ar(y_{t_1+l})}{\sigma_1^2}$, is given by

$$P(y_{t_1+l}\,|\varepsilon_{t_1+l}\,) = A_{t_1+l} + B_{t_1+l}\frac{\sigma_2^2}{\sigma_1^2} + C_{t_1+l}\frac{\sigma_3^2}{\sigma_1^2}. \tag{45}$$

If Assumption 1 is violated then conditional measures of second-order persistence can be constructed using the variance of the forecast error instead of the unconditional variance (results not reported but are available upon request).

Having derived explicit formulas for time varying second-order (or variance) persistence measures, in the next Section we show the empirical relevance of these results using U.S. inflation data.[10]

9 Inflation Data

In this Section we directly link econometric theory with empirical evidence. In our empirical application we consider the possible presence of structural breaks in inflation for the United States. We use quarterly data on the GDP deflator as the measure of price level. The data set consists of observations from 1963Q4 to 2018Q1. Inflation is calculated as the quarterly change of price level at an annualized rate calculated as $\pi_t = 400(\ln(P_t/P_{t-1})$.

In terms of inflation modelling, the period under consideration is of particular interest as it covers the boom-time inflation of the late 1960s, the stagflation in the 1970s, and the double-digit inflation of the early 1980s. During this period substantial shifts in monetary policy occurred, most notably the Fed's radical step of switching policy from targeting interest rates to targeting the money supply in the early 1980s. Therefore when modelling inflation it is important to allow for time varying parameters.

9.1 Unit Root Tests

Although we allow for regime shifts, we are particularly interested in modelling changes in inflation persistence. In the related literature inflation persistence is defined as the tendency of inflation to converge at a slow pace to the long run equilibrium level after a shock. Monetary policy authorities are particularly interested in knowing the speed at which the inflation rate converges to the central bank's inflation target following macroeconomics shocks. However, as shown in Levin and Piger (2004) not accounting for structural breaks may lead to overestimating inflation persistence. In this regards, a growing body of research has found evidence that the monetary policy target has an impact on the persistence properties

[10] Cogley and Sargent (2002) measured persistence by the spectrum at frequency zero, S_0. As an example, for the time invariant AR(2) model this will be given by: $S_0 = \dfrac{\sigma_\varepsilon^2}{2\pi(1-\phi_1-\phi_2)^2}$.

of inflation as well as on its volatility (see for example Brainard and Perry, 2000 or Taylor, 2000). That monetary policy actions affect persistence of inflation is of interest as it has important implications for inflation modelling as changes in regimes of monetary policy may leave econometric models open to the Lucas critique.

In the empirical literature a common approach for modelling inflation persistence is to estimate a univariate AR(p) model where the sum of the estimated autoregressive coefficients is used to approximate the slugginess with which the inflation process responds to macroeconomic shocks (see for example Pivetta and Reis, 2007) and/or apply unit root tests.

Accordingly, in Table 2 a number of common unit roots tests are reported. Namely: ADF (Augmented Dickey–Fuller), ERS, by Elliott *et al.* (1996), and MZ GLS, suggested by Perron and Ng (1996) and Ng and Perron (2001). As recommended by Ng and Perron (2001), the choice of the number of lags is based on the modified Akaike information criterion (AIC). The results in Table 2 show that, in general, we can reject the null hypothesis of a unit root in inflation series.

Table 2. Unit root Tests.

Test Statistic	
ADF	-3.229^{**}
ERS	-3.154^{**}
MZ_a	-19.642^{*}
MZ_t	-3.1331^{*}

The notation: *), **), ***) indicates statistical significance at 1%, 5% and 10%, respectively.

9.2 Structural Breaks

Next we estimate AR(p) models with abrupt structural breaks. The optimal model is the DAB-AR(2; 2) model in eq. (42). The choice of the number of lags was based on the modified AIC and the Bayesian information criteria. The break points are treated as unknown. Note that breaks in the variance are permitted provided that they occur at the same dates as the break in the autoregressive parameters. Benati (2008) also used an AR model allowing for time varying volatility. Cogley and Sargent (2005) also estimated a model in which the variance of innovations can vary over time. For each l partition $(T_1, ..., T_l)$ the DAB-HAR($2; l$) model can be estimated using the least-squared principle by minimizing the sum of the squared residuals where the minimization is taken over all partitions. Since the break points are discrete parameters and can only take a finite number of values they can be estimated by grid search using dynamic programming (see Bai and Perron, 2003, for more details).

Coming to the estimation procedure, the first step is to identify possible points of parameter changes. In order to do so the Bai and Perron (2003) sequential tests on inflation rates is used to identify possible breaks during the sample period. They propose an F-type test for l versus $l+1$ breaks, which we refer to as $\sup F_t(l+1|l)$. The testing procedure allows for a specific to general modelling strategy for the determination of the number of breaks in each series. The test is applied to each segment containing the T_{i-1} to T_i $(i = 1, ..., l+1)$. In particular, the procedure involves using a sequence of $(l+1)$ tests, where the conclusion of a rejection in favour of a model with $(l+1)$ breaks if the overall minimal value of the sum of squared residuals is sufficiently smaller than the sum of the squared residuals from the l break model.

Note that the sum of the squared residuals is calculated over all segments where an additional break is included and compared with the residuals from the l model. Therefore, the break date selected is the one associated with the overall minimum.

The results of the structural break test are reported in Panel A of Table 3. The first column reports the null hypothesis of l breaks versus the alternative hypothesis of $l+1$ breaks, the second column reports the calculated value of the statistics and the third column the critical value of the test. Looking at the calculated values of the test it appears that the null hypothesis zero versus one break is rejected in favour of the alternative hypothesis. Similarly, the hypothesis of one break versus two breaks is rejected. However, the null hypothesis of two versus three breaks in not rejected, therefore we conclude that there are two structural breaks.

The first break occurred in the mid-1970's, when the Fed tightened monetary policy to fight the high inflation rate after the end of the Bretton Woods period. The second break occurred in 1986 when the

Fed embarked on an aggressive policy to reduce inflation, which reached unusually high levels starting from the 70s. As a result, inflation fell from 10.5% at the end of 1980 to 1.1% in 1986Q2, which is also the date of the estimated break. [11]

Table 3. Structural break test and estimation results.

Panel A: Bai-Perron tests of $L+1$ vs. L sequentially determined breaks				
Null hypotheses	F-Statistic	Critical Value		
$H_0 : 0\ vs\ 1$	57.96^{**}	13.98		
$H_0 : 1\ vs\ 2$	18.13^{**}	15.72		
$H_0 : 2\ vs\ 3$	13.57	16.83		
Panel B: Model Estimation and Misspecification Tests				
Period	φ_i	$\phi_{1,i}$	$\phi_{2,i}$	σ_i
1964Q2-1976Q3	0.496^{**} (0.224)	0.470^{*} (0.108)	0.376^{*} (0.102)	1.077
1976Q4-1986Q2	3.637^{*} (0.954)	0.710^{*} (0.119)	0.127 (0.112)	2.300
1986Q3-2018Q1	2.859^{*} (0.396)	0.247^{*} (0.082)	-0.314^{*} (0.077)	2.160
R^2		0.614		
Breusch-Godfrey Test		2.055 (0.561)		
White Test		3.103 (0.376)		

Panel A reports the calculated Bai-Perron test for structural breaks along with the critical value of the test taken from Bai-Perron (2003). Panel B reports the estimated parameters and associated standard errors. The notation: *), **), ***) indicates statistical significance at 1%, 5% and 10%, respectively. The p-values for the misspecification tests are given in parenthesis.

9.3 Estimation Results

As far as the estimation results are concerned Panel B of Table 3 reports the estimated model and the relative misspecification tests. Looking now at the estimated parameters, according to the estimates in Panel B the inflation process is well approximated by a second-order autoregression. With respect to the estimated parameters, the drift parameters φ_i, $i = 1, 2, 3$ increased from $\varphi_3 = 0.496$ before 1976Q3 to $\varphi_2 = 3.637$ during the period 1976-1986. The increase in the drift reflects the fact that toward the second half of the 70s until the middle of the 80s the inflation level was stubbornly high. After 1986 the smaller magnitude of the estimated drift reflects the lower average inflation rates that the US enjoyed over the last three decades. This is in line with the finding in Levin and Piger (2004), who provide statistical evidence for a fall in the intercept after the early 1990s. Kozicki and Tinsley (2002) interpreted this shift as change in the long-run inflation target of the Federal Reserve.

Considering now the estimated autoregressive parameters, $\phi_{1,i}$ and $\phi_{2,i}$, according to the estimates until 1986 the inflation process had a high *intrinsic* persistence ($\phi_{1,3}+\phi_{2,3} = 0.846 \simeq \phi_{1,2}+\phi_{2,2} = 0.837$), but it has fallen ever since. These results are consistent with the findings in Cogley and Sargent (2002) (see also Brainard and Perry, 2000, and Taylor, 2000). With respect to the variance parameter σ_i, we see that the volatility of the innovation was relatively high during the decade 1976-1986 ($\sigma_2 = 2.30$) and it has fallen slightly in the last thirty years ($\sigma_3 = 2.160$). However, it did not go back to the relatively low level before 1976 ($\sigma_1 = 1.077$). This is probably due to the fact that the last period included the turmoil of the financial crisis that started in 2005 (see, for example, Stock and Watson, 2009).

Our estimated model confirms that changes in inflation dynamics can be explained by changes in the drift, the *intrinsic* persistence and the variance parameter. To summarize our results, we find evidence that the parameters in the models capturing persistence change over time. Therefore, not allowing for time varying coefficients in the estimation procedure would result in a less accurate modelling of the inflation process. This, in light of the simulation results in Section 8.4, may lead to poor forecasting. Finally, the misspecification tests are reported at the bottom of Panel B. It appears that the Breusch-Godfrey for

[11] McConnell and Perez-Quiros (2000) have detected a fall in the volatiity of output after 1984 as well.

autocorrelation does not reject the null hypothesis of no serial correlation. Similarly, the White test for heteroscedasticity does not reject the null hypothesis of homoscedasticity, therefore indicating that the model does not suffer from misspecification.

9.4 Forecasting

We now consider the out-of-sample forecasting performance of the model estimated in Table 3. In order to investigate the effect of model misspecification on the forecasted inflation level we compare three models. The first model, which we label as Model 1, is the estimated DAB-AR(2;2). The second model, which we refer to as Model 2, is the true model which we obtained by simulating the inflation process using the estimated parameters in Table 3 as a data generating process. Finally, the third model is the misspecified AR(2) model with no time varying parameters, which we label as Model 3.

The evaluation of the out-of-sample forecast exercise does not rely on a single criterion; for robustness we compare the results of three different forecasting measures, namely, the root mean square forecast error (RMSE), the mean absolute error (MAE) and the Theil Inequality Coefficient (U Coefficient). Table 4 reports the results of the forecasting exercise. [12] In columns 1 and 2 the forecasting horizon and the performance measure are reported, respectively; whereas in columns 3-5 the forecasting results are reported.

Table 4. Forecasting inflation in the United States: point predictive performances.

Forecast Horizon	Forecast Error Measure	Model 1	Model 2	Model 3
1	RMSE	0.0134	0.0110	0.0194
4		0.0141	0.0121	0.0167
8		0.0132	0.0149	0.0242
1	MAE	0.0166	0.0111	0.0944
4		0.0112	0.0101	0.0144
8		0.0112	0.0104	0.0208
1	U Coefficient	0.323	0.251	0.293
4		0.258	0.241	0.243
8		0.327	0.287	0.407

Note: The table compares the out-of-sample point forecasts of three models. Model 1 is the model DAB-AR (2;2) model estimated in Table 3, Model 2 is obtained using simulated data, and Model 3 is an AR(2) process with no time varying parameters. The forecast measures are i) the RMSE, ii) MAE, and iii) the U Coeff.. The forecast horizon is $1, 4$, and 8 quarters.

From Table 4 it is clear that according to the RMSE and MAE criteria the DAB-AR (2;2) model performs better than its misspecified counterpart. According to these two performance measures Model 1 has forecasting properties in line with those obtained using the true model, Model 2. However, looking at the U coefficient measure the results are more mixed with Model 3 outperforming Model 1 in the short horizon and Model 1 having superior performance in the longer horizon.

Having investigated the out-of-sample forecasting performance of the DAB-AR-(2;2) model we next investigate whether inflation and its volatility are highly persistent.

9.5 Inflation Persistence

Pivetta and Reis (2007) employ different estimation methods and measures of persistence. Estimating the persistence of inflation over time using different measures and procedures is beyond the scope of this paper.[13] In this Section we depart from their study in an important way, that is we contribute to

[12] For forecasting under structural breaks, see for example, Pesaran and Timmermann (2005).

[13] Pivetta and Reis (2007) applied a Bayesian approach, which explicitly treats the autoregressive parameters as being stochastically varying and it provides their posterior densities at all points in time. From these, they obtained posterior

the measurement over time of inflation persistence by taking a different approach to the problem and estimate a DAB-HAR model of inflation dynamics grounded on econometric theory, and we compute an alternative measure of persistence, that is, the second-order persistence (using the methodology in Sections 8.1 and 8.2), which distinguishes between changes in the dynamics of inflation and its volatility (and their persistence).

As pointed out by Pivetta and Reis (2007) estimates of the inflation persistence affect the tests of the natural hypothesis neutrality. Therefore detecting whether persistence has recently fallen is key in assessing the likelihood of recidivism by the central bank. In addition, if the central bank feels encouraged to exploit an illusory inflation-output trade off, the result could be high inflation without any accompanying output gains. Furthermore, research on dynamic price adjustment has emphasized the need for theories that generate inflation persistence.

Table 5 presents the within each period time invariant first and second-order measures of persistence for all three periods. The first three columns report the three first-order measures of persistence (LAR, $1/(1-SUM)$ and $\mathbb{E}(\pi_t)$). For the first two measures Period 1 yields the highest persistence. In particular, the persistence (measured by $1/(1-SUM)$) decreases by 5.5% in the post-1976 period and it decreases further by 85% in the post-1986 period. The mean of inflation, $\mathbb{E}(\pi_t)$, increases by 59.3% in the second period and it decreases by 88% in the third period. The last three columns of Table 5 report the three second-order measures of persistence, i.e. S_0, $P_2(\pi_t\,|\varepsilon_t\,) \stackrel{\text{def}}{=} \dfrac{\mathbb{V}ar(\pi_t)}{\mathbb{V}ar(\varepsilon_t)}$ and $\mathbb{V}ar(\pi_t)$. For two out of the three measures the post-1986 period exhibits the lowest persistence whereas in the second period the persistence is the highest. The variance of inflation, $Var(\pi_t)$, from 1976 to 1986 is almost five times the variance of inflation of the pre-1976 period and it is almost three times the variance of the post-1986 period.

Table 5. Persistence for each of the three periods/models.

First and Second-order Measures of Persistence						
	First-Order			Second-Order		
	LAR	1/(1-SUM)	$\mathbb{E}(\pi_t)$	S_0	$P(\pi_t\,\|\varepsilon_t\,)$	$\mathbb{V}ar(\pi_t)$
$1964Q_2 - 1976Q_3$	0.892	6.493	3.221	7.784	2.692	3.122
$1976Q_4 - 1986Q_2$	0.858	6.135	22.313	31.688	3.002	15.881
$1986Q_3 - 2018Q_1$	0.560	0.937	2.679	0.652	1.150	5.365

Note: For each period, $n = 1, 2, 3$ we use the six alternative measures to calculate the (within each period time invariant) first and second-order persistence.

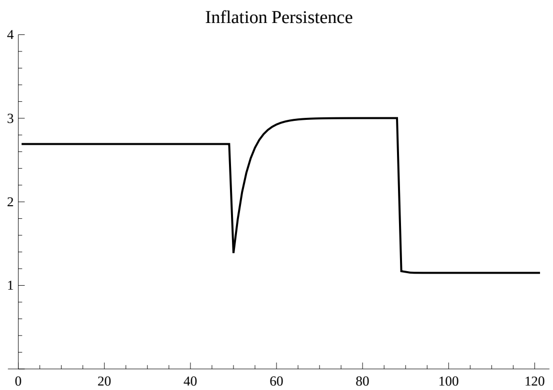
The following graphs[14] of the measures $P_2(\pi_t\,|\varepsilon_t\,)$ and $\mathbb{V}ar(\pi_t)$ reflect the dynamics of the second-order time varying inflation persistence. Details of how we construct the graphs are presented in the online Appendix I. In the x-axis each unit represents a year-quarter starting with $1964Q_2$, chosen as the first. In particular, $1976Q_3 = 49$ (49-th year-quarter) and $1986Q_2 = 88$ (88-th year-quarter).

The key features for the graph of the inflation variance $Var(\pi_t)$ are discussed below: i) In the pre-76 period the graph is constant: $Var(\pi_t) = 3.122$. ii) Within the post-76 and pre-86 period, the graph increases abruptly next to the quarter $1976Q_4$, but at a decreasing rate, reaching in the end the highest value $Var(\pi_t) = 15.881$. iii) In the post-86 period the graph stabilizes to $Var(\pi_t) = 5.365$, after an abrupt drop next to the quarter $1986Q_3$. Analogous statements can be addressed for the inflation persistence graph $P(\pi_t\,|\varepsilon_t\,)$. As illustrated above, the main difference between the shapes of the two graphs is due to the abrupt drop next to the quarter $1976Q_4$ followed shortly afterwards by an abrupt increase at a decreasing rate.

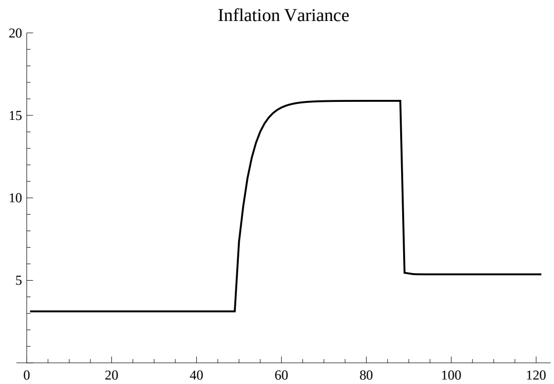
densities for the measures of inflation persistence. Such estimates of persistence are forward-looking, since they are meant to capture the perspective of a policy maker who at a point in time is trying to forsee what the persistence of inflation will be. They also estimated backward-looking measures of persistence that the applied economist forms at a point in time, given all the sample until then.

Pivetta and Reis (2007) also used an alternative set of estimation techniques for persistence. They assumed time invariant autoregressive parameters and re-estimated their AR model on different sub-samples of the data, obtaining median unbiased estimates of persistence for each regression. Finally, Pivetta and Reis also employed rolling and recursive unit root tests.

[14]The graphs have been designed and plotted with the Mathematica drawing tools.



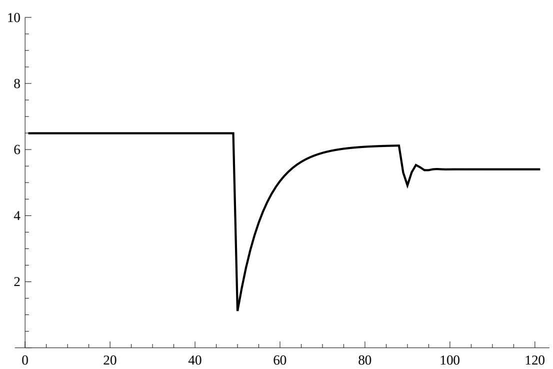
(a) $P_2(\pi_t\,|\varepsilon_t)$



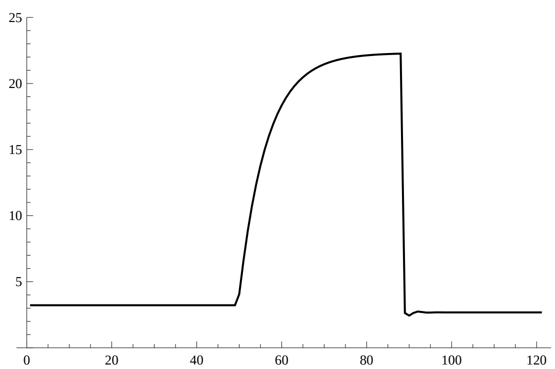
(b) $Var(\pi_t)$

Second-Order Time Varying Persistence

The graphs of the two measures $P_1(\pi_t\,|\varepsilon_t) \overset{\text{def}}{=} \dfrac{\mathbb{E}(\pi_t)}{\varphi(t)}$ and $\mathbb{E}(\pi_t)$ for the first-order persistence are shown below.



(a) $P_1(\pi_t\,|\varepsilon_t)$



(b) $\mathbb{E}(\pi_t)$

First-Order Time Varying Persistence

In sum our main conclusion is that for our chosen specification (DAB-HAR model) the preferred measure of persistence, that is the second-order persistence, as measured by the conditional variance of inflation, increased considerably from 1976 onwards, whereas in the post-1986 period the persistence falls to even lower levels than the pre-1976 period. Our results are in line with those in Cogley and Sargent (2002), who find that the persistence of inflation in the United States rose in the 1970s and remained high during this decade, before starting a gradual decline from the 1980s until the early 2000s (similar to the results of Brainard and Perry, 2000, and Taylor, 2000). Stock and Watson (2002) found no evidence of a change in persistence in U.S. inflation. However, they found strong evidence of a fall in volatility. Therefore their results are in agreement with ours.

10 Conclusions and Future Work

It is important to understand the fundamental properties of ‘linear form’ time series models with variable coefficients in order to handle these more complicated structures efficiently. We have put forward a methodology for solving linear stochastic time varying difference equations. The theory presented makes no claim to being applicable in all 'linear form' processes with variable coefficients. However, the cases covered are those which belong to the large family of 'time varying' models with ARMA representations. Our methodology is a practical tool that can be applied to many dynamic problems. As an illustration we studied an AR specification with abrupt breaks, which is grounded on econometric theory. The second moment structure of this construction was employed to obtain a new time varying measure of second-order persistence.

To summarize, we identified a lack of a universally applicable approach yielding an explicit solution to TV-ARMA models. Our response was to try and fill the gap by developing a coherent body of theory, which implicitly contains the invertibility of a time varying polynomial, and, therefore, can replace the convenient tool of characteristic polynomials. In particular, the general theory does three things: first, it provides a new technique that gives the general solution of such schemes; second, it derives the necessary and sufficient conditions for their stability; and third it generates the second moments of these schemes as well as necessary and sufficient conditions for their existence, which (in the case of the deterministically varying coefficients) are required for the quasi maximum likelihood and central least squares estimation.

We developed this new technique, which can be applied virtually unchanged in every 'ARMA' environment, that is to the even larger family of 'time varying' models, with ARMA representations (i.e., GARCH type of [or stochastic] volatility and Markov switching processes; for the abundant literature on weak ARMA representations see, for example, Francq and Zakoïan, 2005, and the references therein). Thus our results are applied to TV-GARCH models as well without any significant difficulties. This generic framework that forms a base for such a general approach releases us from the need to work with characteristic polynomials and, by enabling us to examine a variety of specifications and solve a number of problems, helps us to deepen our familiarity with their distinctive features.

The empirical relevance of the theory has been illustrated through an application to inflation rates. Our estimation results led to the conclusion that U.S. inflation persistence has been high since 1976, whereas after 1986 the persistence falls to even lower levels than the pre-1976 period, a finding which agrees with those of Brainard and Perry (2000), Taylor (2000) and Cogley and Sargent (2002).

The usefulness of our unified theory is apparent from the fact that it enables us to analyze an abundance of models and solve a plethora of problems. In addition, an extension of the methodology developed in this paper enables us to (just to mention a few consequences): i) examine in depth infinite order autoregressions with either constant or variable coefficients, since it releases us from the need to work with characteristic polynomials and ii) obtain the fourth moments of TV-GARCH models, which themselves follow linear time varying difference equations of infinite order, taking advantage of the fact that the various GARCH formulations have weak ARMA representations (see, for example, Karanasos, 1999, and Francq and Zakoïan, 2005), iii) work out the fundamental time series properties of time varying 'linear form' VAR systems (since it can be easily modified and applied to a multivariate setting; see, for example, Karanasos et al., 2014), iv) derive explicit formulas for the nonnegativity constraints and the second moment structure of both constant and time varying multivariate GARCH processes (thus extending the results in He and Teräsvirta, 2004, Conrad and Karanasos, 2010, and Karanasos and Hu, 2017).

Hallin (1986) applied recurrences in a multivariate context to obtain the Green's matrices. Work is at present continuing on the multivariate case. When this has been completed one should be able to apply, without any major alterations, the methods of this paper to multivariate TV ARMA and GARCH models. Spectral factorization is another important problem that can be solved by our new representations.

Some of these research issues are already work in progress and the rest will be addressed in future work.

References

[1] ABDRABBO, N. A. AND M. B. PRIESTLEY (1967): "On the Prediction of Non-stationary Processes," Journal of the Royal Statistical Society, Series B, 29, 570-585.

[2] AGARWAL, R. P. (2000): "Difference Equations and Inequalities: Theory, Methods, and Applications," Pure and Applied Mathematics, A Series of Monographs and Textbooks.

[3] AKNOUCHE, A. AND H. GUERBYENNE (2009): "Periodic Stationarity of Random Coefficient Periodic Autoregressions," Statistics and Probability Letters, 79, 990-996.

[4] ALJ, A., R. AZRAK, C. LEY, AND G. MÉLARD (2017): "Asymptotic Properties of QML Estimators for Time-Dependent VARMA Models," Scandinavian Journal of Statistics, 44, 617-635.

[5] ANDEL, J. (1991): "ON Stationarity of a Multiple Doubly Stochastic Model," Kybernetica, 2, 114-119.

[6] AZRAK, R. AND G. MÉLARD (2006): "Asymptotic Properties of Quasi-Maximum Likelihood Estimators for ARMA Models with Time-Dependent Coefficients," Statistical Inference for Stochastic Processes, 9, 279-330.

[7] BAI, J. AND P. PERRON (2003): "Computation and Analysis of Multiple Structural Change Models," Journal of Applied Econometrics, 18, 1-22.

[8] BAILLIE, R. (1996): "Long Memory Processes and Fractional Integration in Econometrics," Journal of Econometrics, 73, 5-59.

[9] BENATI, L. (2008): "Investigating Inflation Persistence Across Monetary Regimes," The Quarterly Journal of Economics, 123, 1005-1060.

[10] BOUTHELLIER, P. AND GHOSH. P. K. (1988): "Robust Stabilization of Discrete Time, Single Input Single Output, Linear Time Varying Systems," Proceedings of the 27th IEEE Conference on Decision and Control, Austin, Texas.

[11] BRAINARD, W, AND G. PERRY (2000): "Making Policy in a Changing World," IN: Perry, G., Tobin, J. (Eds.), "Economic Events, Ideas, and Policies: The 1960s and After", Brookings Institution Press.

[12] MARIA CAPORALE, G. AND L. A. GIL-ALANA (2011): "Multi-factor Gegenbauer Processes and European Inflation Rates," Journal of Economic Integration, 26, 386-409.

[13] CHUNG, C-F. (1996): "Estimating a Generalized Long Memory Process," Journal of Econometrics, 73, 237-259.

[14] COGLEY, T. AND T. J. SARGENT (2002): "Evolving Post-World War II U.S. Inflation Dynamics," In Getler, M., Rogoff, K. (Eds.), NBER Macroeconomics Annual, pp. 331-373. MIT Press, Cambridge.

[15] COGLEY, T. AND T. J. SARGENT (2005): "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US," Review of Economic Dynamics, 8, 262-302.

[16] CONRAD, C. AND M. KARANASOS (2010): "Negative Volatility Spillovers in the Unrestricted ECCC-GARCH Model," Econometric Theory, 26, 838-862.

[17] CRAMER, H. (1961): "On Some Classes of Nonstationary Stochastic Processes," Proc. 4th Berkeley Symp. Math. Statist. Prob. 2, 57-78.

[18] DAHLHAUS, R. (1996): "Maximum Likelihood Estimation and Model Selection for Locally Stationary Processes," Journal of Nonparametric Statistics 6, 171–191.

[19] DAHLHAUS, R. (1997): "Fitting Time Series Models to Nonstationary Processes," The Annals of Statistics, 25, 1-37.

[20] DISSANAYKE, G. S, M. S. PEIRIS, AND T. PROIETTI (2018): "Fractionally Differenced Gegenbauer Processes with Long Memory: A Review," Statistical Science, 33, 413-426.

[21] ELAYDI, S. (2005): "An Introduction to Difference Equations," Springer.

[22] ELLIOTT, G., T. J. ROTHENBERG, AND J. H. STOCK (1996): “Efficient Tests for an Autoregressive Unit Root,” Econometrica, 75, 813-836.

[23] ERHARDSSON, T. (2014): ”Conditions for Convergence of Random Coefficient AR(1) Processes and Perpetuities in Higher Dimensions,” Bernoulli, 20, 990-1005.

[24] EVANS, G. W. AND S. HONKAPOHJA (2001): Learning and Expectations in Macroeconomics. Princeton University Press.

[25] EVANS, G. W. AND S. HONKAPOHJA (2009): “Learning and Macroeconomics,” Annual Review of Economics, 1, 421-449.

[26] FIORENTINI, G. AND E. SENTANA (1998): “Conditional Means of Time Series Processes and Time Series Processes for Conditional Means,” International Economic Review, 39, 1101-1018.

[27] FRANCQ, C. AND A. GAUTIER (2004a): “Large Sample Properties of Parameter Least Squares Estimates for Time-Varying ARMA Models,” Journal of Time Series Analysis, 25, 765-783.

[28] FRANCQ, C. AND A. GAUTIER (2004b): “Estimation of Time-Varying ARMA models with Markovian Changes in Regime,” Statistics & Probability Letters, 70, 243-251.

[29] FRANCQ, C. and J.-M. ZAKOÏAN. (2005): “Recent Results for Linear Time Series Models with Non Independent Innovations,” In: DUCHESNE, P., RÉMILARD, B., Eds., “Statistical Modeling and Analysis for Complex Data Problems”, Springer.

[30] FULKERSON, D. R. (1951): “Quasi-Hermite Forms of Row-finite Matrices,” Ph.D. Thesis, University of Wisconsin, URL: http://www.researchgate.net/publication/36218809.

[31] GIRAITIS, L., G. KAPETANIOS, AND T. YATES (2014): “Inference on Stochastic Time-Varying Coefficient Models,” Journal of Econometrics, 179, 46-65.

[32] GIRAITIS, L. AND R. LEIPUS (1995): “ A Generalized Fractionally Differencing Approach in Long-memory modeling,” Lithuanian Mathematical Journal, 35, 53-65.

[33] GLASSERMAN, P. AND D. D. YAO (1995): “Stochastic Vector Difference Equations with Stationary Coefficients,” Journal of Applied Probability, 32, 851-866.

[34] GOLDIE, C. AND R. MALLER (2000): “Stability of Perpetuities,” The Annals of Probability, 28, 1195-1218.

[35] GRANGER, C. W. J. (2007): “Forecasting - Looking Back and Forward: Paper to Celebrate the 50th Anniversary of the Econometrics Institute at the Erasmus University, Rotterdam,” Journal of Econometrics, 138, 3-13.

[36] GRANGER, C. W. J. (2008): “Non-Linear Models: Where Do We Go Next - Time Varying Parameter Models?,” Studies in Nonlinear Dynamics & Econometrics, 12, 1-9.

[37] GRAY, R. M. (2006): Toeplitz and Circulant Matrices: A Review. Now Publishers.

[38] GRILLENZONI, C. (1990): “Modelling Time-Varying Dynamical Systems,” Journal of the American Statistical Association, 85, 499-507.

[39] GRILLENZONI, C. (1993): “ARIMA Processes with ARIMA Parameters,” Journal of Business & Economic Statistics, 11, 235-250.

[40] GRILLENZONI, C. (2000): “Time-Varying Parameters Prediction,” Annals of the Institute of Statistical Mathematics, 52, 108-122.

[41] HALLIN, M. (1984): “Spectral Factorization of Nonstationary Moving Average Processes,” Annals of Statistics, 12, 172-192.

[42] HALLIN, M. (1986): "Non-stationary q-Dependent Processes and Time-Varying Moving-Average Models: Invertibility Properties and the Forecasting Problem," Advances in Applied Probability, 18, 170-210.

[43] HALLIN, M. (1987). Fractions continuées matricielles et matrices-bandes définies positives infinies. In *Statistique et Calcul Scientifique, Proceedings of the Seventh Franco-Belgian Meeting of Statisticians, Rouen, November 1986*, J.-P. Florens and J.-P. Raoult Eds., *Cahiers du Centre d'Etudes de Recherche Opérationnelle* **32**, 37-54.

[44] HAMILTON, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," Econometrica, 57, 357-384.

[45] HAMILTON, J. D. (1994): "Time Series Analysis," Princeton University Press.

[46] HE, C. AND T. TERÄSVIRTA (2004): " An extended Constant Conditional Correlation GARCH Model and its Fourth-Moment Structure," Econometric Theory, 20, 904-926.

[47] HWANG, S. Y. AND I. V. BASAWA (1998): "Parameter Estimation for Generalized Random Coefficient Autoregressive Processes," Journal of Statistical Planning and Inference, 68, 323-337.

[48] KARANASOS, M. (1999): "The Second Moment and the Autocovariance Function of the Squared Errors of the GARCH Model," Journal of Econometrics, 90, 63-76.

[49] KARANASOS, M. (2001): "Prediction in ARMA Models with GARCH in Mean Effects," Journal of Time Series Analysis, 22, 555-576.

[50] KARANASOS, M., A. G. PARASKEVOPOULOS, AND S. DAFNOS (2014a): "The Fundamental Properties of Time Varying AR Models with Non Stochastic Coefficients," arXiv.1403.3359.

[51] KARANASOS, M., A. G. PARASKEVOPOULOS, AND S. DAFNOS (2014b): "A Univariate Time Varying Analysis of Periodic ARMA Processes," arXiv.1403.4803.

[52] KARANASOS, M., A. G. PARASKEVOPOULOS, AND S. DAFNOS (2014c): "A Unified Theory for Time Varying Models: Foundations with Applications in the Presence of Breaks and Heteroscedasticity," Unpublished Paper.

[53] KARANASOS, M., A. G. PARASKEVOPOULOS, S. DAFNOS, AND P. KOUTROUMPIS (2017): "A Unified Theory for Time Varying Models: Foundations with Applications in the Presence of Breaks and Heteroscedasticity," URL: https://research.mbs.ac.uk/accounting-finance/Portals/0/docs/A%20unified%20theory%20for%20time%20varying%20models.pdf

[54] KARANASOS, M., A. G. PARASKEVOPOULOS, M. KAROGLOU, F. MENLA ALI, AND S. YFANTI (2014): "Modelling Returns and Volatilities During Financial Crises: A Time Varying Coefficient Approach," Journal of Empirical Finance, 29, 113-128.

[55] KARANASOS, M. AND Y. HU (2017): "Matrix Inequality Constraints for Vector (Asymmetric Power) GARCH Models and MEM with Spillovers: some New (Mixture) Formulations," Unpublished Paper.

[56] KARMEN, E. W. (1988): "The Poles and Zeros of a Linear Time-Varying System," Linear Algebra and Applications, 98, 263-289.

[57] KITTAPPA, R. K. (1993): "A Representation of the Solution of the nth Order Linear Difference Equation with Variable Coefficients," Linear Algebra and its Applications, 193, 211-222.

[58] KOOP, G. AND D. KOROBILIS (2013): "Large Time-Varying Parameter VARs," Journal of Econometrics, 177, 185-198.

[59] KOWALSKI, A. AND D. SZYNAL (1990): "An Optimal Prediction in General ARMA Models," Journal of Multivariate Analysis, 34, 14-36.

[60] KOWALSKI, A. AND D. SZYNAL (1991): "On a Characterization of Optimal Predictors for Nonstationary ARMA Processes," Stochastic Processes and their Applications, 37, 71-80.

[61] KOZICKI, S. AND P. TINSLEY (2000): "Alternative Sources of the Lag Dynamics of Inflation," FRB Kansas City Research Working Paper No. 02-12

[62] LAKSHMIKANTHAM, V. AND D. TRIGIANTE (2002): "Linear Difference Equations: Numerical Methods and Applications," Marcel Dekker.

[63] LEVIN, A. AND J. PIGGER (2004): "Is Inflation Persistence Intrinsic in Industrial Economics?" Federal Reserve Bank of St. Louis, Working Paper, No. 202-023A.

[64] LIM, A. AND J. DAI (2011): "On Product of Companion Matrices," Linear Algebra and its Applications, 435, 2921-2935.

[65] MALLIK, R. K. (1997): "On the Solution of a Second Order Linear Homogeneous Difference Equation with Variable Coefficients," Journal of Mathematical Analysis and Applications, 215, 32-47.

[66] MALLIK, R. K. (1998): "Solutions of Linear Difference Equations with Variable Coefficients," Journal of Mathematical Analysis and Applications, 222, 79-91.

[67] MALLIK, R. K. (2000): "On the Solution of a Linear Homogeneous Difference Equation with Varying Coefficients," SIAM Journal on Mathematical Analysis, 31, 375-385.

[68] MARRERO, A. AND V. TOMEO (2012): "On the Closed Representation for the Inverse of Hessenberg Matrices," Journal of Computational and Applied Mathematics, 236, 2962-2970.

[69] MARRERO, A. AND V. TOMEO (2017): "On Compact Representations for the Solutions of Linear Difference Equations With Variable Coefficients," J. Comput. Appl. Math., 318, 411-421.

[70] McCONNELL, M. AND G. PEREZ-QUIROS, 2000: "Output Fluctuations in the United States: What has Changed Since the Early 1980's?", American Economic Review, 90, 1464-1476.

[71] MILLER, K. S. (1968): "Linear Difference Equations," Benjamin.

[72] MRAD, R. B. AND FARAGE. (2002): "Identification of Armax Models With Time Dependent Coefficients," Journal of Dynamic Systems, Measurement, and Control, 124, 464-467.

[73] NG, S. AND P. PERRON (2001): "Lag Length Selection and the Construction of Unit Root Tests With Good Size and Power," Econometrica, 69, 1519-1554.

[74] PARASKEVOPOULOS, A. G. (2012): "The Infinite Gauss-Jordan Elimination on Row-Finite $\omega \times \omega$ Matrices," arXiv:1201.2950.

[75] PARASKEVOPOULOS, A. G. (2014): "The Solution of Row-Finite Linear Systems with the Infinite Gauss-Jordan Elimination," arXiv 1403.2624.

[76] PARASKEVOPOULOS, A. G. AND M. KARANASOS (2019): "Explicit and Compact Representations for the Green's Function and the Solution of Linear Difference Equations with Variable Coefficients," arXiv 1906.07130.

[77] PERRON, P. AND S. NG (1996): "Useful Modifications to Some Unit Root Tests With Dependent Errors and Their Local Asymptotic Properties," The Review of Economic Studies, 63, 435–463.

[78] PESARAN, M. H., A. PICK, AND M. PRANOVICH (2013): "Optimal Forecasts in the Presence of Structural Breaks," Journal of Econometrics, 177, 134-152.

[79] PESARAN, M. H. AND A. TIMMERMANN (2005): "Small Sample Properties of Forecasts from Autoregressive Models under Structural Breaks," Journal of Econometrics, 129, 183-217.

[80] PIVETTA, F. AND R. REIS (2007): " The Persistence of Inflation in the United States," Journal of Economic Dynamics and Control, 31, 1326-1358.

[81] RAO, S. T. (1970): "The Fitting of Non-stationary Time-Series Models with Time-Dependent Parameters," Journal of the Royal Statistical Society, Series B, 32, 312-322.

[82] SINGH, N. AND M. S. PEIRIS (1987): "A Note on the Properties of Some Nonstationary ARMA Processes," Stochastic Processes and their Applications, 24, 151-155.

[83] STOCK, J. H. AND M. W. WATSON (2002): "Has the Business Cycle Changed and Why?," In: Gertler, M., Rogoff, K. (Eds.), NBER Macroeconomics Annual, 159-218. MIT Press, Cambridge.

[84] STOCK, J.H., and M. W. WATSON (2009): "Phillips Curve Inflation Forecasts," In: Fuhrer, J., Kodrzycki, Y., Little, J., Olivei, J. (Eds.), "Understanding Inflation and the Implications for Monetary Policy", Cambridge, MIT Press.

[85] TAYLOR, J. (2000): "Low Inflation, Pass Through, and the Pricing Power of Firms," European Economic Review, 44, 1389-1408.

[86] TERÄSVIRTA, T. (1994): "Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models," Journal of the American Statistical Association, 89, 208-218.

[87] THORSON, J. (2000): "Gaussian Elimination on a Banded Matrix", Stanford Exploration Project, Stanford University, 12pp, Web Service.
URL:http://sepwww.stanford.edu/oldreports/sep20/20_11.pdf

[88] TIMMERMANN, A. AND H. K. VAN DIJK (2013): "Dynamic Econometric Modeling and Forecasting in the Presence of Instability", Journal of Econometrics, 177, 131-133.

[89] TOEPLITZ, O. (1909): "Über die Auflösung Unendich Vieler Linearer Gleichungen mit Unendlich Vielen Unbekannten," Rend. Circ. Mat. Palermo., 28, 88-96.

[90] VAN DER CRUYSSEN, P. (1979): "Linear Difference Equations and Generalized Continued Fractions," Computing, 22, 269-278.

[91] VEIN, R. AND P. DALE (1999): "Determinants and their Application in Mathematical Physics," Springer.

[92] WHITTLE, P. (1965): "Recursive Relations for Predictors of Non-stationary Processes," Journal of the Royal Statistical Society, Series B, 27, 523-532.

Appendices

In the appendices we provide proofs for the statements and formulas presented in the paper. The standard notation used in the main body of the paper is adopted throughout the appendices.

A Time Varying ARMA

In this Appendix Section, we present an autonomous procedure for the proofs of the statements of Section 3. The mathematical origins of the solution sequences including the main tool of our analysis (the principal determinant $\xi(t,s)$; see eq. (3)) introduced in this paper along with some computational issues, are discussed in Subsection A.1. We show in Subsection A.2 that the functions $\xi^{(m)}(t,s)$ for $t \geq s+1-p$ (s fixed), defined in eq. (8), form a fundamental solution set associated with eq. (1).

A.1 The Principal determinant

Linear difference equations with variable coefficients of order p (shortly TV-LDEs(p)), therefore TV-ARMA(p,q) models as well, can be represented by infinite linear systems whose coefficient matrix is row-finite[15] the elements of which are the autoregressive coefficients at consecutive instances, starting at instance $s+1$:

$$\begin{bmatrix} \phi_p(s+1) & \phi_{p-1}(s+1) & \phi_{p-2}(s+1)\dots & \phi_1(s+1) & -1 & 0 & 0 & \dots \\ 0 & \phi_p(s+2) & \phi_{p-1}(s+2)\dots & \phi_2(s+2) & \phi_1(s+2) & -1 & 0 & \dots \\ 0 & 0 & \phi_p(s+3)\dots & \phi_3(s+3) & \phi_2(s+3) & \phi_1(s+3) & -1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix} \begin{bmatrix} y_{s-p+1} \\ y_{s-p+2} \\ y_{s-p+3} \\ \vdots \\ y_s \\ y_{s+1,s} \\ y_{s+2,s} \\ \vdots \end{bmatrix} = \begin{bmatrix} v_{s+1} \\ v_{s+2} \\ \vdots \end{bmatrix}. \quad \text{(A.1)}$$

Row-finite systems, in a general form, were first studied by Toeplitz (1909), in which some results on finite linear systems were extended to cover infinite row-finite ones. The representation of their solution was further developed in Fulkerson (1951). He devised and proved the existence of a reduced form (defined by three postulates) for any arbitrary row-finite matrix with the aid of which the general solution of the system is derived. The lack of a method transforming row-finite matrices into a Fulkerson's reduced form has recently been highlighted in Paraskevopoulos (2012), who in responding to this challenge has introduced a modified version of the Gauss-Jordan elimination algorithm equipped with a rightmost pivot strategy (called Infinite Gauss-Jordan elimination algorithm). In a companion paper, Paraskevopoulos (2014) has further developed the infinite Gauss-Jordan elimination algorithm, yielding an analytic form to the general solution of row-finite linear systems. The algorithm is effectively applied to infinite system representations of TV-LDEs(p) constructing a fundamental solution set. The elements of this set, occupying the first p columns in the Fulkerson's reduced matrix, are linearly independent solution sequences of the homogeneous system associated with eq. (A.1), that is, whenever $v_{s+i}=0$ for all i. The first p terms of each fundamental solution are occupied by the components of corresponding unit vectors, which coincide with the initial conditions under which eq. (A.1) yields a unique solution. Their remaining terms come out as expansions of banded Hessenbergians, denoted here as $\xi^{(m)}(t,s)$. The principal determinant, $\xi(t,s) \stackrel{\text{def}}{=} \xi^{(1)}(t,s)$, represents the first of these solution sequences, that is the principal fundamental sequence. A few first terms of $\{\xi(t,s)\}_{t\geq s-p+1}$ are:

$$y_{s-p+1}=y_{s-p+2}=\dots=y_{s-1}=0, y_s=1, y_{s+1,s}=\phi_1(s+1), y_{s+2,s}=\begin{vmatrix} \phi_1(s+1) & -1 \\ \phi_2(s+2) & \phi_1(s+2) \end{vmatrix}, \dots$$

Applying the above values of y's to eq. (A.1), provided that $v_{s+i}=0$ for all $i\in\mathbb{Z}_{\geq 1}$, the first equations of the homogeneous system are easily verified. Applying the same sequence of elementary operations to the sequence of forcing terms $\{v_{s+i}\}_{i\geq 1-p}$, a particular solution sequence is constructed. It is also represented by a Hessenbergian which coincides with eq. (11) (see Proposition A4). The general solution turns out to be a linear combination of fundamental solutions (that is the general homogeneous solution) with coefficients the observed values plus the particular solution mentioned above.

The linear time complexity for the calculation of banded-matrix determinants of order k, that is $O(k)$, entails that the principal determinant (or the Green's function) is computationally tractable. This is due to the Gaussian elimination algorithm, which uses approximately $\dfrac{k(p+1)^2}{4}$ multiplications, where $(p+1)$ is the bandwidth of the matrix (see Thorson, 2000). $O(k)$ time complexity is comparable with the computational time complexity of algorithms that calculate the Hessenbergians by recursion.

A.2 A Fundamental Set of Solutions

In the literature about 'time varying' models fundamental solution sets play a decisive role in the explicit representation of their solution. Their existence is theoretically guaranteed by the Fundamental Theorem of LDEs (see the Fundamental Theorem of LDEs in Elaydi, 2005 p.74). As a consequence of the superposition principle (see the previously cited reference) the general homogeneous solution of eq.

[15] A row-finite matrix is an $\mathbb{N}\times\mathbb{N}$ infinite matrix, each row of which comprises a finite number of non-zero entries.

(5) is explicitly expressed as a linear combination of the fundamental solutions with coefficients arbitrary constants determined by the initial conditions. Moreover, the determinant form of the Green's function (see Agarwal, 2000 p.77, eq. (2.11.7)) is expressed in terms of the p fundamental solutions occupying corresponding matrix entries. Having a fundamental solution set at our disposal, the Green's function is also explicitly expressed, yielding a particular solution, which involves this function and the forcing term (see Miller, 1968 pp 40, eqs. (2.8), (2.9)).

In this Appendix Section, the principal matrix in eq. (2) is slightly modified to provide an explicit representation of the remaining $p-1$ elements of a fundamental solution set associated with any arbitrary TV-LDE of p order. Moreover, we derive the formula (A.4) below, which shows that every fundamental solution can also be expressed in terms of the principal determinant $\xi(t,s)$ exclusively.

In order to save space, we shall interchangeably use the notation $t-s$ and k. We define the matrix:

$$\mathbf{\Phi}_{t,s}^{(m)} = \tag{A.2}$$

$$\begin{bmatrix}
\phi_m(s+1) & -1 & & & & & & & & \\
\phi_{m+1}(s+2) & \phi_1(s+2) & \ddots & & & & & & & \\
\vdots & \vdots & \ddots & \ddots & & & & & & \\
\phi_p(s+p+1-m) & \phi_{p-m}(s+p+1-m) & \ddots & \ddots & \ddots & & & & & \\
 & \vdots & \ddots & \ddots & \ddots & \ddots & & & & \\
 & \phi_{p-1}(s+p) & \ddots & \ddots & \ddots & \ddots & \ddots & & & \\
 & \phi_p(s+p+1) & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & & \\
 & & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\
 & & & \phi_p(t-1) & \phi_{p-1}(t-1) & \cdots & \phi_{p-m}(t-1) & \cdots & \phi_1(t-1) & -1 \\
 & & & & \phi_p(t) & \cdots & \phi_{p+1-m}(t) & \cdots & \phi_2(t) & \phi_1(t)
\end{bmatrix}.$$

We show next that the sequences $\{\xi^{(m)}(t,s)\}_{t\geq s-p+1}$ for $1\leq m\leq p$ determined by

$$\xi^{(m)}(t,s) = \begin{cases} \det(\Phi_{t,s}^{(m)}) & \text{If } t>s \\ 1 & \text{If } t=s+1-m \\ 0 & \text{If } s+1-p\leq t\leq s \text{ and } t\neq s+1-m, \end{cases} \tag{A.3}$$

form a fundamental solution set associated with the eq. (5).

The matrix $\mathbf{\Phi}_{t,s}^{(m)}$, $m\geq 2$, is derived by replacing the first column of $\mathbf{\Phi}_{t,s}$ (see eq. (2) in the main body of the paper) with the column vector: $[\phi_{t,s}^{(m)}]$, given by

$$[\phi_{t,s}^{(m)}]' = \begin{pmatrix} \phi_m(s+1), & \phi_{m+1}(s+2), & \ldots & ,\phi_p(s+p+1-m), & 0, & \ldots & ,0 \end{pmatrix}.$$

In other words, each matrix in the sequence $\{\mathbf{\Phi}_{t,s}^{(m)}\}_{1\leq m\leq p}$, differs from any other matrix in this sequence only in the first column $[\phi_{t,s}^{(m)}]$. Formally $\mathbf{\Phi}_{t,s}^{(m)}$ is a banded Hessenberg matrix of order $k=t-s$. The principal matrix $\mathbf{\Phi}_{t,s}$ (resp. principal determinant $\xi(t,s)$) is identified with $\mathbf{\Phi}_{t,s}^{(1)}$ (resp. $\xi^{(1)}(t,s)$), that is $\mathbf{\Phi}_{t,s}^{(1)} \stackrel{\text{def}}{=} \mathbf{\Phi}_{t,s}$ (resp. $\xi^{(1)}(t,s) \stackrel{\text{def}}{=} \xi(t,s)$). For notational convenience we will interchangeably use $\mathbf{\Phi}_{t,s}^{(1)}$ (resp. $\xi^{(1)}(t,s)$) in place of $\mathbf{\Phi}_{t,s}$ (resp. $\xi(t,s)$).

Next, we show that $\xi^{(m)}(t,s)$, defined by the sum expansion in eq. (8), coincides with the definition in eq. (A.3). In the proof we use the following Lemma (see Paraskevopoulos and Karanasos, 2019; Lemma 1 for a formal proof):

Lemma A1
i) The cofactor of the coefficient $\phi_{m+i}(s+i+1)$ *in the first column of* $\Phi_{t,s}^{(m)}$ *coincides with* $\xi(t,s+1+i)$ *for* $i=0,...,p-m$.
ii) The cofactor of the coefficient $\phi_i(t)$, *in the last row of* $\Phi_{t,s}^{(m)}$ *coincides with* $\xi^{(m)}(t-i,s)$.

As a consequence of the above Lemma, we have:

Proposition A1 *i) The cofactor expansion of* $\xi^{(m)}(t,s)$ *along the first column of* $\Phi_{t,s}^{(m)}$ *is given by*

$$\xi^{(m)}(t,s) = \sum_{r=1}^{p+1-m} \phi_{m-1+r}(s+r)\xi(t,s+r), \tag{A.4}$$

which coincides with the definition in eq. (8) in Subsection 3.1.
ii) The cofactor expansion of $\xi^{(m)}(t,s)$ *along the last row of* $\Phi_{t,s}^{(m)}$ *gives*

$$\xi^{(m)}(t,s) = \phi_1(t)\xi^{(m)}(t-1,s)+\phi_2(t)\xi^{(m)}(t-2,s)+...+\phi_p(t)\xi^{(m)}(t-p,s) = \sum_{j=1}^{p}\phi_j(t)\xi^{(m)}(t-j,s). \tag{A.5}$$

Eq. (A.5) entails that $\{\xi^{(m)}(t,s)\}_{t\geq s+1-p}$ is the solution sequence of eq. (5) under the initial values given in eq. (9), that is $\xi^{(m)}(s+1-m,s) = 1$ and $\xi^{(m)}(s+1-r,s) = 0$, whenever $r = 1,2,...,p$ and $r \neq m$.

The linear independence of the solutions $\xi^{(m)}(t,s)$ for $1 \leq m \leq p$, is verified in the following Proposition:

Proposition A2 *For any arbitrary but fixed* $s \in \mathbb{Z}$ *the set of the solutions*

$$\Xi_s = \{\xi^{(1)}(t,s), \xi^{(2)}(t,s), ..., \xi^{(p)}(t,s): \; t \geq s+1-p\}$$

is a fundamental solution set associated with eq. (5).

Proof. Let us consider the sequence of Casorati matrices associated with the set Ξ_s:

$$\mathbf{\Xi}_{t,s} = \begin{bmatrix} \xi^{(1)}(t,s) & \xi^{(2)}(t,s) & ... & \xi^{(p)}(t,s) \\ \xi^{(1)}(t-1,s) & \xi^{(2)}(t-1,s) & ... & \xi^{(p)}(t-1,s) \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ \xi^{(1)}(t+1-p,s) & \xi^{(2)}(t+1-p,s) & ... & \xi^{(p)}(t+1-p,s) \end{bmatrix}.$$

The Definition in (A.3) entails that the matrix $\mathbf{\Xi}_{s,s}$ is the identity matrix of order p. Therefore the first Casoratian $|\mathbf{\Xi}_{s,s}|$, of the set Ξ_s is $|\mathbf{\Xi}_{s,s}| = 1 \neq 0$. It turns out that $|\mathbf{\Xi}_{t,s}| \neq 0$ for all $t \geq s$ and the set Ξ_s is linearly independent (see Elaydi, 2005, Corollary 2.14. pp 69). Moreover, as the dimension of the homogeneous solution space of eq. (5) is p, the set Ξ_s is a fundamental solution set associated with eq. (5). ∎

A.3 Decomposition

In this Appendix Subsection we prove Proposition 1, that is the decomposition of the innovation part of the particular solution.

Proof of Proposition 1. Let us write u_t in eq. (1) as $u_t = \sum_{l=0}^{q}\theta_l(r)\varepsilon_{r-l}$, provided that $\theta_0(t) \overset{\text{def}}{=} 1$ for all t. The left side of eq. (13) can be expressed as:

$$\begin{aligned}\sum_{r=s+1}^{t}\xi(t,r)u_r &= \sum_{r=s+1}^{t}\xi(t,r)\sum_{l=0}^{q}\theta_l(r)\varepsilon_{r-l} = \sum_{l=0}^{q}\sum_{r=s+1}^{t}\xi(t,r)\theta_l(r)\varepsilon_{r-l} \\ &= \sum_{r=s+1}^{t}\xi(t,r)\theta_0(r)\varepsilon_r + \sum_{l=1}^{q}\sum_{r=s+1}^{t}\xi(t,r)\theta_l(r)\varepsilon_{r-l}.\end{aligned}$$

By splitting the second double sum in the right-hand side of the above equation into two parts, it takes the form:

$$\sum_{r=s+1}^{t} \xi(t,r)u_r = \sum_{r=s+1}^{t} \xi(t,r)\theta_0(r)\varepsilon_r + \sum_{l=1}^{q}\sum_{r=s+1}^{s+l} \xi(t,r)\theta_l(r)\varepsilon_{r-l} + \sum_{l=1}^{q}\sum_{r=s+1+l}^{t} \xi(t,r)\theta_l(r)\varepsilon_{r-l}. \tag{A.6}$$

As the extended definition of $\xi(t,s)$ in eq. (4) entails that $\xi(t,r+l)=0$, whenever $r+l>t$ (or $r>t-l$), the second sum in the last double sum of eq. (A.6) can be rewritten as:

$$\sum_{r=s+1+l}^{t} \xi(t,r)\theta_l(r)\varepsilon_{r-l} = \sum_{r=s+1}^{t-l} \xi(t,r+l)\theta_l(r+l)\varepsilon_r = \sum_{r=s+1}^{t} \xi(t,r+l)\theta_l(r+l)\varepsilon_r.$$

Substituting the above sum in eq. (A.6) we get:

$$\sum_{r=s+1}^{t} \xi(t,r)u_r = \sum_{r=s+1}^{t} \xi(t,r)\theta_0(r)\varepsilon_r + \sum_{l=1}^{q}\sum_{r=s+1}^{t} \xi(t,r+l)\theta_l(r+l)\varepsilon_r + \sum_{l=1}^{q}\sum_{r=s+1}^{s+l} \xi(t,r)\theta_l(r)\varepsilon_{r-l},$$

or equivalently

$$\sum_{r=s+1}^{t} \xi(t,r)u_r = \sum_{l=0}^{q}\sum_{r=s+1}^{t} \xi(t,r+l)\theta_l(r+l)\varepsilon_r + \sum_{l=1}^{q}\sum_{r=s+1}^{s+l} \xi(t,r)\theta_l(r)\varepsilon_{r-l}. \tag{A.7}$$

Using the definition of $\xi_q(t,r)$ in eq. (12), eq. (A.7) can be rewritten as:

$$\sum_{r=s+1}^{t} \xi(t,r)u_r = \sum_{r=s+1}^{t} \xi_q(t,r)\varepsilon_r + \sum_{l=1}^{q}\sum_{r=s+1}^{s+l} \xi(t,r)\theta_l(r)\varepsilon_{r-l}. \tag{A.8}$$

By expanding the double sum in eq. (A.8), we have:

$$\begin{aligned}\sum_{l=1}^{q}\sum_{r=s+1}^{s+l} \xi(t,r)\theta_l(r)\varepsilon_{r-l} \;=\; & \underbrace{\xi(t,s+1)\theta_1(s+1)\varepsilon_s}_{l=1,\ r=s+1} + \underbrace{\xi(t,s+2)\theta_2(s+2)\varepsilon_s + \xi(t,s+1)\theta_2(s+1)\varepsilon_{s-1}}_{l=2,\ r=s+1,s+2} \\ & + \underbrace{\xi(t,s+3)\theta_3(s+3)\varepsilon_s + \xi(t,s+2)\theta_3(s+2)\varepsilon_{s-1} + \xi(t,s+1)\theta_3(s+1)\varepsilon_{s-2}}_{l=3,\ r=s+1,s+2,s+3} + \cdots \\ & + \underbrace{\xi(t,s+q)\theta_q(s+q)\varepsilon_s + \cdots + \xi(t,s+1)\theta_q(s+1)\varepsilon_{s+1-q}}_{l=q,\ r=s+1,s+2,\ldots,s+q}.\end{aligned}$$

By rearranging terms, we can rewrite the latter double sum as:

$$\begin{aligned}\sum_{l=1}^{q}\sum_{r=s+1}^{s+l} \xi(t,r)\theta_l(r)\varepsilon_{r-l} \;=\; & \underbrace{[\xi(t,s+1)\theta_1(s+1) + \xi(t,s+2)\theta_2(s+2) + \cdots + \xi(t,s+q)\theta_q(s+q)]\varepsilon_s}_{\sum_{l=1}^{q} \xi(t,s+l)\theta_l(s+l)\varepsilon_s} \\ & + \underbrace{[\xi(t,s+1)\theta_2(s+1) + \cdots + \xi(t,s+q-1)\theta_q(s+q-1)]\varepsilon_{s-1}}_{\sum_{l=2}^{q} \xi(t,s-1+l)\theta_l(s-1+l)\varepsilon_{s-1}} + \cdots + \underbrace{\xi(t,s+1)\theta_q(s+1)\varepsilon_{s+1-q}}_{\sum_{l=q}^{q} \xi(t,s+1-q+l)\theta_l(s+1-q+l)\varepsilon_{s+1-q}} = \\ & \sum_{r=s+1-q}^{s}\;\sum_{l=s-r+1}^{q} \xi(t,r+l)\theta_l(r+l)\varepsilon_r.\end{aligned} \tag{A.9}$$

Therefore, substituting the result of eq. (A.9) back into eq. (A.8), we obtain the expression:

$$\sum_{r=s+1}^{t} \xi(t,r)u_r = \sum_{r=s+1}^{t} \xi_q(t,r)\varepsilon_r + \sum_{r=s+1-q}^{s}\;\sum_{l=s-r+1}^{q} \xi(t,r+l)\theta_l(r+l)\varepsilon_r. \tag{A.10}$$

Substituting the defining formula of $\xi_{s,q}(t,r)$ (see eq. (12)) into eq. (A.10) the latter takes the form:

$$\sum_{r=s+1}^{t} \xi(t,r)u_r = \sum_{r=s+1}^{t} \xi_q(t,r)\varepsilon_r + \sum_{r=s+1-q}^{s} \xi_{s,q}(t,r)\varepsilon_r,$$

that is eq. (13), as required. ■

A.4 The General Solution

In this Section we show the explicit representation of the solution in eq. (14).

We start with the following Proposition, which provides an expression of the homogeneous solution as a linear combination of the fundamental solutions, as shown below.

Proposition A3 *The solution of eq. (5) assuming the prescribed initial values* $\{y_{s+1-m}\}_{m=1,2,...,p}$ *is given by*

$$y_{t,s}^{hom} = \sum_{m=1}^{p} \xi^{(m)}(t,s)y_{s+1-m}. \tag{A.11}$$

Proof. As Ξ_s, defined in Proposition A2, is a fundamental solution set, every solution can be expressed as $y_{t,s}^{hom} = \sum_{m=1}^{p} a_m\xi^{(m)}(t,s)$. Fixing the initial conditions at $y_{s+1-m} = c_m$ for $m = 1, 2, ..., p$, it remains to show that $c_m = a_m$ for all $m :\ 1 \leq m \leq p$. Taking into account that for $1 \leq m \leq p$

$$\xi^{(m)}(s+1-m,s) = 1 \ \text{ and } \xi^{(m)}(s+1-r,s) = 0, \text{ whenever } 1 \leq r \leq p \text{ and } r \neq m,$$

we have: $c_m = y_{s+1-m} = y_{s+1-m,s}^{hom} = \sum_{r=1}^{p} a_r\xi^{(r)}(s+1-r,s) = a_m\xi^{(m)}(s+1-m,s) = a_m$ and the proof is complete. ■

Applying eq. (A.5) to eq. (A.11), we obtain the general homogeneous solution in (7), which is exclusively expressed in terms of the Green's function.

In the following Proposition we provide a particular solution for eq. (1) and we show that this solution is a Hessenbergian representation of eq. (11).

Proposition A4 *i) The solution of eq. (1) subject to zero initial values* $y_{s-r} = 0$ *for* $0 \leq r \leq p-1$, *can be expressed as a single Hessenbergian:*

$$y^*_{t,s} = \tag{A.12}$$

$$\begin{vmatrix}
v_{s+1} & -1 & & & & & & & & \\
v_{s+2} & \phi_1(s+2) & \ddots & & & & & & & \\
\vdots & \vdots & \ddots & \ddots & & & & & & \\
v_{s+p+1-m} & \phi_{p-m}(s+p+1-m) & & \ddots & \ddots & & & & & \\
\vdots & \vdots & \ddots & & \ddots & \ddots & & & & \\
v_{s+p} & \phi_{p-1}(s+p) & & \ddots & & \ddots & \ddots & & & \\
v_{s+p+1} & \phi_p(s+p+1) & \ddots & & \ddots & & \ddots & \ddots & & \\
\vdots & & \ddots & \ddots & & \ddots & & \ddots & \ddots & \\
v_{t-1} & & & \phi_p(t-1) & \phi_{p-1}(t-1) & \cdots & \phi_{p-m}(t-1) & \cdots & \phi_1(t-1) & -1 \\
v_t & & & & \phi_p(t) & \cdots & \phi_{p+1-m}(t) & \cdots & \phi_2(t) & \phi_1(t)
\end{vmatrix}.$$

ii) The expressions in eqs. (11) and (A.12) coincide, that is $y^{par}_{t,s} = y^*_{t,s}$.

Proof. i) Working with elementary properties of determinants, it turns out that the cofactor of the coefficients $\phi_m(t)$ for $1 \leq m \leq p$, in the last row of the determinant in eq. (A.12) is

$$y^*_{t-m,s} =$$

$$\begin{vmatrix}
v_{s+1} & -1 & & & & & & \\
v_{s+2} & \phi_1(s+2) & \ddots & & & & & \\
\vdots & \vdots & \ddots & \ddots & & & & \\
v_{s+p} & \phi_{p-1}(s+p) & & \ddots & \ddots & & & \\
v_{s+p+1} & \phi_p(s+p+1) & \ddots & & \ddots & \ddots & & \\
\vdots & & \ddots & \ddots & & \ddots & \ddots & \\
\vdots & & & \ddots & \ddots & & \ddots & \ddots \\
v_{t-m-1} & & & \phi_p(t-m-1) & \phi_{p-1}(t-m-1) & \cdots & \phi_1(t-m-1) & -1 \\
v_{t-m} & & & & \phi_p(t-m) & \cdots & \phi_2(t-m) & \phi_1(t-m)
\end{vmatrix}$$

(see Paraskevopoulos and Karanasos, 2019, for a formal proof). Moreover the cofactor of v_t in eq. (A.12) is $(-1)^{k+1}(-1)^{k-1} = 1$. As a consequence, the cofactor expansion of (A.12) along the last row gives:

$$y^*_{t,s} = \sum_{m=1}^{p} \phi_m(t) y^*_{t-m,s} + v_t.$$

Adding to the right-hand side of this expression the homogeneous solution in (A.11) with $y_{s+1-m} = 0$, we conclude that $y^*_{t,s}$ satisfies eq. (1) subject to zero initial values, as required.
ii) Working similarly, the cofactor of v_{s+i} in the first column of eq. (A.12) is $\xi(t, s+i+1)$ for $1 \leq i \leq k$. Thus, the cofactor expansion of the determinant in (A.12) along the first column yields the expression in eq. (11). ■

Using the result in Proposition 1 we can rewrite the particular solution in (11) as

$$y^{par}_{t,s} = \sum_{r=s+1}^{t} \xi(t,r)\varphi(r) + \sum_{r=s+1}^{t} \xi_q(t,r)\varepsilon_r + \sum_{r=s+1-q}^{s} \xi_{s,q}(t,r)\varepsilon_r. \tag{A.13}$$

Proof of Theorem 1. The solution of eq. (1), represented by eq. (14), can be obtained by adding the homogeneous and particular solutions in eqs. (A.11) and (11) respectively, yielding:

$$y_{t,s} = \sum_{m=1}^{p} \xi^{(m)}(t,s) y_{s+1-m} + \sum_{r=s+1}^{t} \xi(t,r)\varphi(r) + \sum_{r=s+1}^{t} \xi(t,r) u_r.$$

Applying eq. (A.13) to the above expression of $y_{t,s}$, eq. (14) follows. ■

An explicit representation of the general solution of eq. (14) and its equivalence with the single determinant representation (see Kittappa, 1993) are shown in (Paraskevopoulos and Karanasos, 2019).

B Asymptotic Stability

In this Appendix Section we provide a proof for the stability of DTV-ARMA processes associated with non-stochastic coefficients, presented in Theorem 2. In the case of stochastic coefficients, the proof is analogous.

In the following Lemma we adhere to the notation:

$$\tilde{\phi}_{t,m,r} = \sup_{s \le t} |\phi_{m+r-1}(s+r)| \ \text{ and } \tilde{\phi}_{t,m} = \max_{1 \le r \le p-m+1} \tilde{\phi}_{t,m,r}.$$

The hypothesis $\sup_t |\phi_m(t)| < \infty$ for each m such that $1 \le m \le p$, in the stability Theorem 2, entails that $\tilde{\phi}_{t,m,r} < \infty$. It follows from the finiteness of the range values of r that $\tilde{\phi}_{t,m} < \infty$ too.

Lemma B2 *i) If the backward stability condition holds, that is if* $\lim_{s \to -\infty} \xi(t,s) = 0$ *for each* $t \in \mathbb{Z}$, *then*

$$\lim_{s \to -\infty} \xi^{(m)}(t,s) = 0 \text{ for all } t \in \mathbb{Z} \text{ and } m \in \mathbb{Z} : \ 1 \le m \le p,$$

provided that $\sup_t |\phi_m(t)| < \infty$ *for each* m.
ii) If the forward stability condition holds, that is $\lim_{t \to \infty} \xi(t,s) = 0$ *for each* $s \in \mathbb{Z}$, *then*

$$\lim_{t \to \infty} \xi^{(m)}(t,s) = 0 \text{ for each } s \in \mathbb{Z} \text{ and } m \in \mathbb{Z} : \ 1 \le m \le p.$$

Proof. i) In view of eq. (A.4) we have:

$$\begin{aligned} \left|\xi^{(m)}(t,s)\right| = \left|\sum_{r=1}^{p+1-m} \phi_{m-1+r}(s+r)\xi(t,s+r)\right| &\le \sum_{r=1}^{p+1-m} |\phi_{m-1+r}(s+r)||\xi(t,s+r)| \\ &\le \sum_{r=1}^{p+1-m} \tilde{\phi}_{t,m}|\xi(t,s+r)| \\ &= \tilde{\phi}_{t,m} \sum_{r=1}^{p+1-m} |\xi(t,s+r)|. \end{aligned}$$

Letting $s \to -\infty$ in the above inequalities we get:

$$\lim_{s \to -\infty} |\xi^{(m)}(t,s)| \le \lim_{s \to -\infty} \tilde{\phi}_{t,m} \sum_{r=1}^{p+1-m} |\xi(t,s+r)| = \tilde{\phi}_{t,m} \sum_{r=1}^{p+1-m} \lim_{s \to -\infty} |\xi(t,s+r)| = 0.$$

Hence $\lim_{s \to -\infty} \xi^{(m)}(t,s) = 0$ for each t, m.

ii) The assertion follows from:

$$\lim_{t \to \infty} \xi^{(m)}(t,s) = \lim_{t \to \infty} \sum_{r=1}^{p+1-m} \phi_{m-1+r}(s+r)\xi(t,s+r) = \sum_{r=1}^{p+1-m} \phi_{m-1+r}(s+r) \lim_{t \to \infty} |\xi(t,s+r)| = 0.$$

This completes the proof of the Lemma. ■

The above Lemma states that if the principal fundamental sequence $\{\xi(t,s)\}_{s\le t}$ converges to zero, as $s \to -\infty$ (or $t \to \infty$), then all of the fundamental sequences $\{\xi^{(m)}(t,s)\}_{s\le t}$ converge to zero too.

In the proof of the backward stability in Theorem 2, in addition to the notation of Subsection 4.1, we further call $\mathfrak{z}(s) = \mathbf{c}$ for all s such that $s \le t$, where $\mathbf{c}' = [c_m]_{1\le m\le p}$ is the initial condition vector. Accordingly, $\mathfrak{z}$ is a constant vector-valued function yielding the sequence of the same repeated term, the vector $\mathbf{c}$. In view of the solution expression in eq. (7), the TV-ARMA process associated with eq. (1) is backwards asymptotically stable if and only if $y^{hom}_{t,s} \to 0$, as $s \to -\infty$ for the initial values being the components of $\mathbf{c}$. In other words, as the initial condition vector $\mathfrak{z}(s) = \mathbf{c}$ moves further to the past, i.e. $s \to -\infty$, its effects on the solution $y_{t,s}$ are gradually dying out.

Proof of Theorem 2. (Sufficient) In view of eq. (A.11), Lemma B2 implies that

$$\lim_{s\to-\infty} y^{hom}_{t,s} = \lim_{s\to-\infty} \sum_{m=1}^{p} c_m \, \xi^{(m)}(t,s) = \sum_{m=1}^{p} c_m \lim_{s\to-\infty} \xi^{(m)}(t,s) = \sum_{m=1}^{p} c_m \cdot 0 = 0,$$

which shows the backward stability of the process, as required.

(Necessary) The formula $y^{hom}_{t,s} = \sum_{m=1}^{p} \xi^{(m)}(t,s)c_m$ applied with $\{c_1 = 1, c_2 = 0, c_3 = 0, ..., c_p = 0\}$ yields $\xi^{(1)}(t,s)$ ($\xi(t,s)$ for short), that is $y^{hom}_{t,s} = \xi(t,s)$. This amounts to the same as saying that $\xi(t,s)$ is the solution of eq. (5) subject to the initial values: $\{y_s = 1, y_{s-1} = 0, ..., y_{s-p} = 0\}$ for all s with $s \le t$. The assumption $\lim_{s\to-\infty} y^{hom}_{t,s} = 0$ for any initial condition vector $\mathfrak{z}(s) = \mathbf{c}$ and all t implies that $\lim_{s\to-\infty} \xi(t,s) = 0$ for all t, as required.

Replacing $(s \to -\infty)$ by $(t \to \infty)$ in the above statements we deduce the forward asymptotic stability of the model, without invoking the condition $\sup_t |\phi_m(t)| < \infty$ for each m. ■

C Second Order Properties

We show in this Section that the first and the second unconditional moments exist, provided that the absolute summability condition holds. Under this condition the Wold-Cramér decomposition of the DTV-ARMA(p,q) processes is obtained along with the second order properties of such processes.

C.1 Unconditional Moments

In this Subsection we give a proof for the existence of the first and second unconditional moments, described in Proposition 5, completed by the logical implications that render the Diagrams in eq. (17) commutative.

Proof of Proposition 5. First, we verify that the first unconditional moment in eq. (15) exists in $\mathbb{R}$, that is $\sum_{r=-\infty}^{t} \xi(t,r)\varphi(r)$ converges, provided that $\sum_{r=-\infty}^{t} |\xi(t,r)| < \infty$ for each t (absolute summability condition). Employing the notation $\tilde{\varphi}_t = \sup_{s\le t} |\varphi(s)| \in \mathbb{R}_{\ge 0}$ for each t, we have:

$$\sum_{r=s}^{t} |\xi(t,r)\varphi(r)| = \sum_{r=s}^{t} |\xi(t,r)||\varphi(r)| \le \sum_{r=s}^{t} |\xi(t,r)|\tilde{\varphi}_t = \tilde{\varphi}_t \sum_{r=s}^{t} |\xi(t,r)|, \text{ for all } s : s \le t. \tag{C.1}$$

Letting $s \to -\infty$ in the inequality (C.1) and taking into account that absolute summability implies summability, the result follows from:

$$\sum_{r=-\infty}^{t} |\xi(t,r)\varphi(r)| \le \tilde{\varphi}_t \sum_{r=-\infty}^{t} |\xi(t,r)| < \infty \text{ for all t.} \tag{C.2}$$

Second, we show that the second unconditional moment in eq. (16) exists in $\mathbb{R}_{\ge 0}$, provided that $\sum_{r=-\infty}^{t} |\xi(t,r)| < \infty$ for all t. In view of eq. (12), we show first that $\sum_{r=-\infty}^{t} |\xi_q(t,r)| < \infty$ for all t. Let us call $\tilde{\theta}_l = \sup_r |\theta_l(l+r)| \in \mathbb{R}_{\geq 0}$ for each $l = 1, ..., q$ and $\Theta = \max_{0 \leq l \leq q} \tilde{\theta}_l$, where $\theta_0(t) \stackrel{\text{def}}{=} 1$ for all t. Then $\xi_q(t,r)$ can be rewritten as $\xi_q(t,r) = \sum_{l=0}^{q} \xi(t, r+l)\theta_l(r+l)$ and

$$|\xi_q(t,r)| \leq \left| \sum_{l=0}^{q} \xi(t, r+l)\tilde{\theta}_l \right| \leq \Theta \left| \sum_{l=0}^{q} \xi(t, r+l) \right|. \tag{C.3}$$

It follows from Tonelli's Theorem for series (either convergent or divergent) that we can switch the summation order, that is $\sum_{r=-\infty}^{t} \sum_{l=0}^{q} |\xi(t, r+l)| = \sum_{l=0}^{q} \sum_{r=-\infty}^{t} |\xi(t, r+l)|$, whence

$$\sum_{r=-\infty}^{t} |\xi_q(t,r)| \leq \sum_{r=-\infty}^{t} \Theta \left| \sum_{l=0}^{q} \xi(t, r+l) \right| \leq \Theta \sum_{r=-\infty}^{t} \sum_{l=0}^{q} |\xi(t, r+l)| = \Theta \sum_{l=0}^{q} \sum_{r=-\infty}^{t} |\xi(t, r+l)|. \tag{C.4}$$

The hypothesis $\sum_{r=-\infty}^{t} |\xi(t,r)| < \infty$ along with the fact that $\sum_{r=t+1}^{t+l} |\xi(t,r)| = 0$ implies that

$$\sum_{r=-\infty}^{t} |\xi(t, r+l)| = \sum_{r=-\infty}^{t+l} |\xi(t,r)| = \sum_{r=-\infty}^{t} |\xi(t,r)| + \sum_{r=t+1}^{t+l} |\xi(t,r)| = \sum_{r=-\infty}^{t} |\xi(t,r)| < \infty \tag{C.5}$$

for all t and any $l : 0 \leq l \leq q$. Let us call $g(t,l) = \sum_{r=-\infty}^{t} |\xi(t, r+l)|$. It follows from eq. (C.5) that $g(t,l) \in \mathbb{R}_{\geq 0}$ for all t and any $l : 0 \leq l \leq q$. Accordingly $\Theta \sum_{l=0}^{q} g(t,l) \in \mathbb{R}_{\geq 0}$ for all t and any l such that $0 \leq l \leq q$ (as being a multiple of a finite sum of real numbers). It follows from inequality (C.4) that

$$\sum_{r=-\infty}^{t} |\xi_q(t,r)| \leq \Theta \sum_{l=0}^{q} \sum_{r=-\infty}^{t} |\xi(t, r+l)| = \Theta \sum_{l=0}^{q} g(t,l) < \infty,$$

for all t, as claimed. We recall that $0 < \sigma^2(r) \leq M$ for all r. Taking into account that absolute summability implies square summability, that is: $\sum_{r=-\infty}^{t} |\xi_q(t,r)| < \infty \implies \sum_{r=-\infty}^{t} \xi_q^2(t,r) < \infty$ for all t, the existence of variance follows from:

$$\mathbb{V}ar(y_t) = \sum_{r=-\infty}^{t} \xi_q^2(t,r)\sigma^2(r) \leq \sum_{r=-\infty}^{t} \xi_q^2(t,r)M = M \sum_{r=-\infty}^{t} \xi_q^2(t,r) < \infty \quad \text{for all } t. \tag{C.6}$$

It follows from eqs. (C.2) and (C.6) that $\lim_{s \to -\infty} \xi(t,s)\varphi(s) = 0$ and $\lim_{s \to -\infty} \xi_q^2(t,s)\sigma^2(s) = 0$, respectively, are necessary conditions for the existence of the first and second unconditional moments respectively.

Finally, in view of the diagrams in eq. (17), it remains to show the following implications:

$$\begin{array}{ccc} & & \lim_{s \to -\infty} \xi(t,s)\varphi(s) = 0 \\ & \nearrow\!\!\!\!\nearrow & \\ \lim_{s \to -\infty} \xi(t,s) = 0 & & \qquad \text{for all } t \in \mathbb{Z}. \\ & \searrow\!\!\!\!\searrow & \\ & & \lim_{s \to -\infty} \xi_q^2(t,s)\sigma^2(s) = 0 \end{array} \tag{C.7}$$

Since

$$|\xi(t,s)\varphi(s)| = |\xi(t,s)||\varphi(s)| \leq |\xi(t,s)|\tilde{\varphi}_t = \tilde{\varphi}_t|\xi(t,s)| \quad \text{for all } t, s \in \mathbb{Z},$$

and $\tilde{\varphi}_t \lim_{s \to -\infty} |\xi(t,s)| = 0$ it follows from the squeeze Theorem that $\lim_{s \to -\infty} |\xi(t,s)\varphi(s)| = 0$ for all $t \in \mathbb{Z}$. As $\left(\lim_{s \to -\infty} |\xi(t,s)\varphi(s)| = 0 \iff \lim_{s \to -\infty} \xi(t,s)\varphi(s) = 0 \right)$, the first implication in diagram (C.7) follows. Taking into account the implications

$$\sum_{r=-\infty}^{t} |\xi(t,r)| < \infty \implies \sum_{r=-\infty}^{t} \xi_q^2(t,r) < \infty \implies \lim_{s\to-\infty} \xi_q^2(t,s) = 0 \implies \lim_{s\to-\infty} M\xi_q^2(t,s) = 0, \text{ for all } t \in \mathbb{Z}$$

along with the fact that $\xi_q^2(t,s)\sigma^2(s) \leq M\xi_q^2(t,s)$ for all $t, s \in \mathbb{Z}$, it follows from the squeeze Theorem that $\lim_{s\to-\infty} \xi_q^2(t,s)\sigma^2(s) = 0$. This shows the second implication in diagram (C.7) and completes the proof of the Proposition. ■

C.2 Wold-Cramér Decomposition

In the next Theorem we prove the Wold-Cramér decomposition of a DTV-ARMA(p, q) process.

Proof of Theorem 3. The homogeneous solution associated with the data information sequence $\{y_{s+1-m}\}_{1\leq m\leq p}$ is given by eq. (10). For each $t \in \mathbb{Z}$ arbitrary but fixed, we define the random variables in the extended real line $(\mathbb{R} \cup \{\pm\infty\})$: $\tilde{y}_{t,m} = \sup_{s\leq t} |y_{s+1-m}|$, for each m and $\tilde{y}_t = \max_{1\leq m\leq p} \tilde{y}_{t,m}$. The following inequality trivially holds:

$$\left|y_{t,s}^{hom}\right| = \left|\sum_{m=1}^{p} \xi^{(m)}(t,s) y_{s+1-m}\right| \leq \sum_{m=1}^{p} \left|\xi^{(m)}(t,s)\right| \tilde{y}_t = \tilde{y}_t \sum_{m=1}^{p} |\xi^{(m)}(t,s)|.$$

Taking the limits to both sides of the above inequality and using the measure theory convention that $+\infty \cdot 0 = 0$ we have:

$$\lim_{s\to-\infty} \left|y_{t,s}^{hom}\right| \leq \lim_{s\to-\infty} \tilde{y}_t \sum_{m=1}^{p} |\xi^{(m)}(t,s)| = \tilde{y}_t \sum_{m=1}^{p} \lim_{s\to-\infty} |\xi^{(m)}(t,s)| = \tilde{y}_t \sum_{m=1}^{p} 0 = 0$$

(notice that we allow $\tilde{y}_t = +\infty$). As a consequence $\lim_{s\to-\infty} y_{t,s}^{hom} = 0$. In view of the definition of u_r in eq. (1) we have:

$$\sum_{r=s+1}^{t} \xi(t,r) u_r = \sum_{r=s+1}^{t} \sum_{l=0}^{q} \xi(t,r)\theta_l(r)\varepsilon_{r-l}. \tag{C.8}$$

As we have shown in the proof of Proposition 5 the infinite sum $\sum_{r=-\infty}^{t} \xi(t,r)[\varphi(r) + u_r]$ converges in L_2, provided that the condition of the absolute summability holds. Taking the limits in eq. (C.8) as $s \to -\infty$ and recalling that $\xi(t,r) = 0$ for $r > t$ we have:

$$\begin{aligned}
\sum_{r=-\infty}^{t} \xi(t,r) u_r \quad &= \sum_{r=-\infty}^{t} \sum_{l=0}^{q} \xi(t,r)\theta_l(r)\varepsilon_{r-l} \\
\text{(switching summation)} \quad &= \sum_{l=0}^{q} \sum_{r=-\infty}^{t} \xi(t,r)\theta_l(r)\varepsilon_{r-l} \\
\text{(changing the summation limits)} \quad &= \sum_{l=0}^{q} \sum_{r=-\infty}^{t-l} \xi(t,r+l)\theta_l(r+l)\varepsilon_r \\
\text{(adding some zero terms)} \quad &= \sum_{l=0}^{q} \sum_{r=-\infty}^{t-l} \xi(t,r+l)\theta_l(r+l)\varepsilon_r + \sum_{l=0}^{q} \sum_{r=t-l+1}^{t} \xi(t,r+l)\theta_l(r+l)\varepsilon_r \\
\text{(condensed sum)} \quad &= \sum_{l=0}^{q} \sum_{r=-\infty}^{t} \xi(t,r+l)\theta_l(r+l)\varepsilon_r \\
\text{(switching summation)} \quad &= \sum_{r=-\infty}^{t} \sum_{l=0}^{q} \xi(t,r+l)\theta_l(r+l)\varepsilon_r \\
\text{(definition in eq. (12))} \quad &= \sum_{r=-\infty}^{t} \xi_q(t,r)\varepsilon_r.
\end{aligned} \tag{C.9}$$

As $\lim_{s\to-\infty} y^{hom}_{t,s} = 0$, it follows from eqs. (14), (11) and (C.9) that:

$$\begin{aligned}\lim_{s\to-\infty} y_{t,s} &= \lim_{s\to-\infty} y^{hom}_{t,s} + \lim_{s\to-\infty} y^{par}_{t,s} = \lim_{s\to-\infty} y^{par}_{t,s} \\ &= \sum_{r=-\infty}^{t} \xi(t,r)[\varphi(r)+u_r] = \sum_{r=-\infty}^{t} \xi(t,r)\varphi(r) + \sum_{r=-\infty}^{t} \xi_q(t,r)\varepsilon_r.\end{aligned}$$

Let us recall the notation $v_r = \phi(r) + u_r$. It remains to show that $y_t = \sum_{r=-\infty}^{t} \xi(t,r)v_r$ (or $y_t = \lim_{s\to-\infty} y_{t,s}$) solves eq. (1). Applying the expression $y_{t-m} = \sum_{r=-\infty}^{t-m} \xi(t-m,r)v_r$ for $m = 0, 1, ..., p$ to eq. (1), we show below that its right-hand side, i.e. $\sum_{m=1}^{p} \phi_m(t) \sum_{r=-\infty}^{t-m} \xi(t-m,r)v_r$, equals its left side, i.e. $\sum_{r=-\infty}^{t} \xi(t,r)v_r$:

$$\begin{aligned}\sum_{j=1}^{p} \phi_j(t) \sum_{r=-\infty}^{t-j} \xi(t-j,r)v_r &= \sum_{j=1}^{p}\sum_{r=-\infty}^{t-j} \phi_j(t)\xi(t-j,r)v_r \\ \text{(adding some zero terms)} &= \sum_{j=1}^{p}\sum_{r=-\infty}^{t-j} \phi_j(t)\xi(t-j,r)v_r + \sum_{j=1}^{p}\sum_{r=t-j+1}^{t} \phi_j(t)\xi(t-j,r)v_r \\ \text{(condensed sum)} &= \sum_{j=1}^{p}\sum_{r=-\infty}^{t} \phi_j(t)\xi(t-j,r)v_r \\ \text{(switching summation)} &= \sum_{r=-\infty}^{t}\sum_{j=1}^{p} \phi_j(t)\xi(t-j,r)v_r \\ \text{(factoring)} &= \sum_{r=-\infty}^{t} v_r \sum_{j=1}^{p} \phi_j(t)\xi(t-j,r) \\ \text{(applying eq. (A.5) for m = 1)} &= \sum_{r=-\infty}^{t} v_r\xi(t,r).\end{aligned}$$

This completes the proof of the Theorem. ■

C.3 Autocovariance Function

By virtue of Theorem 3, the stochastic part of the one sided MA representation of a stochastic process $\{y_t\}_t$ associated with a DTV-ARMA(p,q) is given by $y_t = \sum_{r=-\infty}^{t} \xi_q(t,r)\varepsilon_r$. A proof of Proposition 6 is given below.

Proof of Proposition 6.

In what follows we use the statements:

i) As $\{\varepsilon_t\}$ is a martingale difference, it follows that: $\mathbb{E}(\varepsilon_{r_1} \cdot \varepsilon_{r_2}) = 0$, whenever $r_1 \neq r_2$.

ii) As $\phi_m(t)$ are deterministic, it follows that $\xi_q(t,r)$ is deterministic too, whence: $\mathbb{E}(\xi_q(t,r) \cdot \xi_q(t-\ell,r)) = \xi_q(t,r) \cdot \xi_q(t-\ell,r)$.

iii) As shown in the proof of Proposition 5, it follows from our general condition $\sum_{s=-\infty}^{t} |\xi(t,s)| < \infty$ that $\sum_{s=-\infty}^{t} |\xi_q(t,s)| < \infty$. We remark that the absolute summability is a sufficient condition for switching expectation with infinite summation.

iv) It follows from statements (i)-(iii) that

$$\mathbb{E}\left(\sum_{r=t-\ell+1}^{t} \xi_q(t,r)\varepsilon_r \cdot \sum_{r=-\infty}^{t-\ell} \xi_q(t-\ell,r)\varepsilon_r \right) = 0. \quad \text{(C.10)}$$

Employing the autocovariance function notation $\gamma_t(\ell) \overset{\text{def}}{=} \mathbb{C}ov(y_t, y_{t-\ell})$, in view of Theorem 3, the following equalities hold:

$$\begin{aligned}
\gamma_t(\ell) =& \ \mathbb{E}\left(\sum_{r=-\infty}^{t} \xi_q(t,r)\varepsilon_r \cdot \sum_{r=-\infty}^{t-\ell} \xi_q(t-\ell,r)\varepsilon_r\right) \\
\text{(breaking first sum)} \quad =& \ \mathbb{E}\left(\left(\sum_{r=-\infty}^{t-\ell} \xi_q(t,r)\varepsilon_r + \sum_{r=t-\ell+1}^{t} \xi_q(t,r)\varepsilon_r\right)\sum_{r=-\infty}^{t-\ell} \xi_q(t-\ell,r)\varepsilon_r\right) \\
\text{(associative law)} \quad =& \ \mathbb{E}\left(\sum_{r=-\infty}^{t-\ell} \xi_q(t,r)\varepsilon_r \sum_{r=-\infty}^{t-\ell} \xi_q(t-\ell,r)\varepsilon_r + \sum_{r=t-\ell+1}^{t} \xi_q(t,r)\varepsilon_r \sum_{r=-\infty}^{t-\ell} \xi_q(t-\ell,r)\varepsilon_r\right) \\
\text{(by eq. (C.10))} \quad =& \ \mathbb{E}\left(\sum_{r=-\infty}^{t-\ell} \xi_q(t,r)\varepsilon_r \sum_{r=-\infty}^{t-\ell} \xi_q(t-\ell,r)\varepsilon_r\right) \\
\text{(by statement(i))} \quad =& \ \mathbb{E}\left(\sum_{r=-\infty}^{t-\ell} \xi_q(t,r)\xi_q(t-\ell,r)\varepsilon_r^2\right) \\
\text{(by statements (ii) \& (iii))} \quad =& \sum_{r=-\infty}^{t-\ell} \xi_q(t,r)\xi_q(t-\ell,r)\sigma^2(r),
\end{aligned}$$

as required. ∎

C.4 Forward asymptotic efficiency

Before proving Proposition 7 and Corollary 2, we provide some elementary definitions for oscillation.

Let $\{x_t\}_{t\in\mathbb{Z}_{\geq 0}}$ be a bounded sequence, that is $|x_t| \leq M$ for some $M \in \mathbb{R}_{\geq 0}$ and for all $t \in \mathbb{Z}_{\geq 0}$. Every bounded sequence is either convergent or oscillating and divergent with oscillation given by

$$\mathit{\Omega} \overset{\text{def}}{=} \limsup_{t\to\infty} x_t - \liminf_{t\to\infty} x_t \overset{\text{or}}{=} \lim_{t\to\infty}(\sup_{r\geq t} x_r - \inf_{r\geq t} x_r) \overset{\text{or}}{=} \inf_t(\sup_{r\geq t} x_r - \inf_{r\geq t} x_r)$$

If the oscillation is zero, then the sequence converges, otherwise it diverges.

Proof of Proposition 7. As $\{F(t,s)\}_t$ is bounded, there exists some $N_s \in \mathbb{R}_{\geq 0}$ such that $F(t,s) \leq N_s$ for all t. First we show that the sequence $\{F_q(t,s)\}_t$ defined by $F_q(t,s) = \sum_{r=s+1}^{t} |\xi_q(t,r)|$ is also bounded, as a function of t, for each fixed s. Using the notation of the proof of Proposition 5 and taking into account inequality (C.3) we deduce that:

$$\begin{aligned}
\sum_{r=s+1}^{t} |\xi_q(t,r)| \leq \sum_{r=s+1}^{t} \Theta\left|\sum_{l=0}^{q} \xi(t,r+l)\right| \quad &\leq \quad \Theta \sum_{r=s+1}^{t}\sum_{l=0}^{q} |\xi(t,r+l)| \\
\text{(switching summation)} \quad &= \quad \Theta \sum_{l=0}^{q}\sum_{r=s+1}^{t} |\xi(t,r+l)| \\
\text{(breaking summation)} \quad &= \quad \Theta \sum_{l=0}^{q}\left(\sum_{r=s+1}^{t-l} |\xi(t,r+l) + \sum_{r=t-l+1}^{t} |\xi(t,r+l)|\right) \\
\text{(subtracting zero terms)} \quad &= \quad \Theta \sum_{l=0}^{q}\sum_{r=s+1}^{t-l} |\xi(t,r+l)| \\
\text{(shifting summation index)} \quad &= \quad \Theta \sum_{l=0}^{q}\sum_{r=s+l+1}^{t} |\xi(t,r)| \\
\text{(by definitions)} \quad &= \quad \Theta \sum_{l=0}^{q} F(t,s+l) \leq \Theta \sum_{l=0}^{q} N_{s+l}
\end{aligned}$$

The first requirement follows from:

$$F_q(t,s) \le \Theta \sum_{l=0}^{q} N_{s+l} \quad \text{for all } t. \tag{C.11}$$

Let us call $S(t,s) = \sum_{r=s+1}^{t} \xi_q^2(t,r)$ for $t > s$. Employing the notation of Proposition 2, it follows from the well known identity

$$(\sum_{r=s+1}^{t} |\xi_q(t,r)|)^2 = \sum_{r=s+1}^{t} \xi_q^2(t,r) + 2 \sum_{i=s+1}^{t-1} \sum_{j=i+1}^{t} |\xi_q(t,i)\xi_q(t,j)|$$

that

$$S(t,s) = \sum_{r=s+1}^{t} \xi_q^2(t,r) \le \sum_{r=s+1}^{t} \xi_q^2(t,r) + 2 \sum_{i=s+1}^{t-1} \sum_{j=i+1}^{t} |\xi_q(t,i)\xi_q(t,j)| = (\sum_{r=s+1}^{t} |\xi_q(t,r)|)^2 = F_q^2(t,s) \tag{C.12}$$

for all $t > s$. As $F_q(t,s) \ge 0$, inequality (C.11) implies: for every $s \in \mathbb{Z}$, $F_q^2(t,s) \le (\Theta \sum_{l=0}^{q} N_{s+l})^2 \in \mathbb{R}_{\ge 0}$ for all $t > s$. Thus $\{F_q^2(t,s)\}_t$ is bounded in t for each s. Let us call $S_s \overset{\text{def}}{=} \sup_t F_q^2(t,s)$. Now inequality (C.12) implies that for every s: $0 \le S(t,s) \le F_q^2(t,s) \le S_s$ for all $t > s$. Thus $\{S(t,s)\}_t$ is bounded in t for each $s \in \mathbb{Z}$. As $0 < \sigma^2(r) \le M$, we have:

$$\mathbb{MSE}_{t,s} = \sum_{r=s+1}^{t} \xi_q^2(t,r)\sigma^2(r) \le M \sum_{r=s+1}^{t} \xi_q^2(t,r) = M \cdot S(t,s) \le M \cdot S_s.$$

Accordingly $\{\mathbb{MSE}_{t,s}\}_t$ is bounded in t for each s, as claimed. ■

Proof of Corollary 2. Let us call $U = \sup_t F_t$. Then

$$F(t,s) = \sum_{r=s+1}^{t} |\xi(t,r)| \le \sum_{r=-\infty}^{t} |\xi(t,r)| = F_t \le U.$$

Thus the condition of Proposition 7 is fulfilled and therefore $\{\mathbb{MSE}_{t,s}\}_t$ is bounded in t for each s. The uniformly boundedness can be derived as follows. If we replace N_s by U, in the proof of Proposition 7, then inequality (C.11) entails that: $F_q(t,s) \le \Theta(q+1)U$. Also, it follows from inequality (C.12) that: $S(t,s) \le (\Theta(q+1)U)^2$. Thus, $\mathbb{MSE}_{t,s} \le M \cdot (\Theta(q+1)U)^2$. As the latter bound is in $\mathbb{R}_{\ge 0}$ and independent of t, s the result follows. ■

D Stochastic Coefficients

Proof of Theorem 5. The proof follows from the fact that $\xi_{t,s} \overset{a.s}{\to} 0$ as $t \to \infty$ implies that $||\mathbf{C}_{t,s}|| \overset{a.s.}{\to} 0$ as $t \to \infty$ and, therefore, Theorem 2.1 in Erhardsson (2014) applies. ■

Notation D1 *In the Lemma and proof below we will make use of the following notation in relation to the companion form in eq. (23):*

$$\begin{aligned}
i)\mathbb{E}(\mathbf{\Phi}_t) &= \mathbf{\Phi}, \\
ii)\ \widetilde{\mathbf{y}}_t &= \mathbf{y}_t - \mathbf{y},\ \widetilde{\phi}_{0t} = \phi_{0t} - \phi_0,\ \widetilde{\mathbf{\Phi}}_t = \mathbf{\Phi}_t - \mathbf{\Phi}\ \textit{(deviations from the mean)}, \\
iii)\ \widetilde{\varepsilon}_t &= \widetilde{\phi}_{0t} + \varepsilon_t + \widetilde{\mathbf{\Phi}}_t \mathbf{y}, \\
iv)\ \mathbf{\Phi}^{(m)} &= \begin{pmatrix} \sigma_{m1} + \phi_m\phi_1 & \sigma_{m2} + \phi_m\phi_2 & \cdots & \sigma_{m,p-1} + \phi_m\phi_{p-1} & \sigma_{mp} + \phi_m\phi_p \\ \phi_m & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & \phi_m & 0 \end{pmatrix}, \\
\textit{for } m &= 1, \ldots, p.
\end{aligned}$$

(we recall that ϕ_m and σ_{mn} are given in Notation 1)
and v) $\mathbf{\Gamma}(\ell) = \mathbb{E}(\widetilde{\mathbf{y}}_t \widetilde{\mathbf{y}}'_{t-\ell})$ *with* $\gamma(\ell) = vec(\mathbf{\Gamma}(\ell))$.

The proof of the following Lemma is straightforward.

Lemma D3 *The expected value of the vec form of* $\widetilde{\varepsilon}_t \widetilde{\varepsilon}'_t$ *is given by*

$$\mathbb{E}[vec(\widetilde{\varepsilon}_t\ \widetilde{\varepsilon}'_t)] = \mathbb{E}[vec(\widetilde{\phi}_{0t}\widetilde{\phi}'_{0t} + \varepsilon_t\ \varepsilon'_t + 2\widetilde{\phi}_{0t}\ \varepsilon'_t)] + 2\mathbb{E}[\widetilde{\mathbf{\Phi}}_t \otimes (\widetilde{\phi}_{0t} + \varepsilon_t)]vec(\mathbf{y}') + \mathbb{E}(\widetilde{\mathbf{\Phi}}_t^{\otimes 2})vec(\mathbf{y}\mathbf{y}') = \sigma^2 \mathbf{i}_0,$$

where $\mathbf{i}_0$ *is a vector of order* p^2 *with first element unity and zero otherwise (recall that* σ^2 *has been defined in eq. (26)).*

Proof of Proposition 8 and Theorem 6. Let us rewrite the Companion form of the GRC-AR(p) model, that is eq. (22):

$$\mathbf{y}_t = \phi_{0t} + \mathbf{\Phi}_t \mathbf{y}_{t-1} + \varepsilon_t. \tag{D.1}$$

Under Condition 2 taking expectations on both sides of the above equation gives

$$\mathbf{y} = \phi_0 + \mathbf{\Phi}\mathbf{y}, \tag{D.2}$$

or

$$y = \frac{\phi_0}{1 - \sum_{m-1}^{p} \phi_m}.$$

Subtracting eq. (D.2) from eq. (D.1), the latter is expressed in terms of deviations from the mean as follows:

$$\begin{aligned} \widetilde{\mathbf{y}}_t &= \widetilde{\phi}_{0t} + \mathbf{\Phi}_t \mathbf{y}_{t-1} - \mathbf{\Phi}\mathbf{y} + \varepsilon_t \\ &= \widetilde{\phi}_{0t} + \mathbf{\Phi}_t \widetilde{\mathbf{y}}_{t-1} + \widetilde{\mathbf{\Phi}}_t \mathbf{y} + \varepsilon_t \\ \text{(by Notation D1(iii))} \quad &= \mathbf{\Phi}_t \widetilde{\mathbf{y}}_{t-1} + \widetilde{\varepsilon}_t. \end{aligned} \tag{D.3}$$

Taking the transpose on both sides of eq. (D.3) yields

$$\widetilde{\mathbf{y}}'_t = \widetilde{\mathbf{y}}'_{t-1}\mathbf{\Phi}'_t + \widetilde{\varepsilon}'_t. \tag{D.4}$$

Right-multiplying eq. (D.3) by eq. (D.4) yields

$$\widetilde{\mathbf{y}}_t \widetilde{\mathbf{y}}'_t = \mathbf{\Phi}_t \widetilde{\mathbf{y}}_{t-1} \widetilde{\mathbf{y}}'_{t-1} \mathbf{\Phi}'_t + \widetilde{\varepsilon}_t \widetilde{\varepsilon}'_t + \mathbf{\Phi}_t \widetilde{\mathbf{y}}_{t-1} \widetilde{\varepsilon}'_t + \widetilde{\mathbf{y}}'_{t-1} \mathbf{\Phi}'_t \widetilde{\varepsilon}_t.$$

Under Conditions 2 and 3, taking expectations on both sides of the above equation, ignoring zero terms, and applying the vec operator, gives:

$$\begin{aligned} \gamma(0) &= \mathbb{E}(\mathbf{\Phi}_t^{\otimes 2})\ \gamma(0) + \mathbb{E}[vec(\widetilde{\varepsilon}_t \widetilde{\varepsilon}'_t)] \\ \text{(by Lemma D3)} \quad &= \mathbb{E}(\ \mathbf{\Phi}_t^{\otimes 2})\gamma(0) + \sigma^2 \mathbf{i}_0. \end{aligned} \tag{D.5}$$

Solving eq. (D.5) for $\gamma(0)$ gives

$$\gamma(0) = [\mathbf{I}_{p^2} - \mathbb{E}(\mathbf{\Phi}_t^{\otimes 2})]^{-1} \sigma^2 \mathbf{i}_0. \tag{D.6}$$

Next, note that $\mathbf{I}_{p^2} - \mathbb{E}(\mathbf{\Phi}_t^{\otimes 2})$ can be written as a block matrix

$$\mathbf{I}_{p^2} - \mathbb{E}(\mathbf{\Phi}_t^{\otimes 2}) = \begin{pmatrix} \mathbf{I}_p - \mathbf{\Phi}^{(1)} & -\mathbf{\Phi}^{(2)} & \cdots & -\mathbf{\Phi}^{(p-1)} & -\mathbf{\Phi}^{(p)} \\ -\mathbf{\Phi} & \mathbf{I}_p & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & -\mathbf{\Phi} & \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & -\mathbf{\Phi} & \mathbf{I}_p \end{pmatrix},$$

We recall that the element occuping the (i,j)-th entry of $[\mathbf{I}_{p^2} - \mathbb{E}(\mathbf{\Phi}_t^{\otimes 2})]^{-1}$ is denoted by φ_{ij} (see Notation 2 (i)). Thus eq. (D.6) implies that

$$\gamma(\ell) = \varphi_{\ell+1,1}\sigma^2, \text{ for } \ell = 0, \ldots, p-1,$$

Finally it is not difficult to show that, for $\ell \geq p$:

$$\gamma(\ell) = (\mathbf{I}_{p^2} \otimes \mathbf{C}_\ell)\gamma(0), \tag{D.7}$$

with

$$\mathbf{C}_\ell = \mathbb{E}(\mathbf{C}_{t,\ell}) = \begin{pmatrix} \xi_\ell^{(1)} & \xi_\ell^{(2)} & \cdots & \xi_\ell^{(p)} \\ \xi_{\ell-1}^{(1)} & \xi_{\ell-1}^{(2)} & \cdots & \xi_{\ell-1}^{(p)} \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ \xi_{\ell-p+1}^{(1)} & \xi_{\ell-p+1}^{(2)} & \cdots & \xi_{\ell-p+1}^{(p)} \end{pmatrix}, \tag{D.8}$$

where $\xi_\ell^{(m)}$ has been defined in Notation 2 (iii). Eqs. (D.7) and (D.8) imply that

$$\gamma(\ell) = \sum_{m=1}^{p} \xi_\ell^{(m)} \gamma(m-1).$$

■

The proofs of Propositions 9, 10, 11 and Theorem 7 are similar to those presented for the deterministic coefficient case.

Proof of Theorem 8. The two summability conditions must hold for all integers t. Since all the autoregressive coefficients are strictly stationary, the Green's function is strictly stationary, and, therefore, the left hand-sides of the two inequality conditions, does not change when we subtract $t-r$ from each index (see also Anděl, 1991). ■

E Time Varying Polynomials

In this Appendix we will prove Proposition 12 and Theorems 9, 10. Let us first rewrite the two equations in Definition 1 by replacing r with $t-r$ as follows:

$$\xi_q(t, t-r) = \xi(t, t-r) + \sum_{l=1}^{q} \xi(t, t-r+l)\theta_l(t-r+l), \text{ for } r = 0, \ldots, k-1, \tag{E.1}$$

$$\xi_{s,q}(t, t-r) = \sum_{l=r-k+1}^{q} \xi(t, t-r+l)\theta_l(t-r+l), \text{ for } r = k-1+q, \ldots, k. \tag{E.2}$$

Proof of Proposition 12.

On account of $u_t = \Theta_t(B)\varepsilon_t$, it suffices to show that

$$\Xi_t^{(k)}(B) \circ \Theta_t(B) = \Xi_{t,q}^{(k)}(B).$$

In view of eqs. (36) and (32), respectively, the left-hand side of the above equation can be rewritten as:

$$\Xi_t^{(k)}(B) \circ \Theta_t(B) = \sum_{r=0}^{k-1} \xi(t, t-r)B^r \circ \left(1 + \sum_{l=1}^{q} \theta_l(t)B^l\right).$$

Multiplying the two polynomials (using the properties of '$\circ$'), we get:

$$\Xi_t^{(k)}(B) \circ \Theta_t(B) = \sum_{r=0}^{k-1} \xi(t, t-r)B^r + \sum_{r=0}^{k-1}\sum_{l=1}^{q} \xi(t, t-r)\theta_l(t-r)B^{r+l}.$$

Collecting the terms of the right-hand side polynomials in the above equation whose backshift operator is raised to the same powers, we obtain:
i) For powers from zero up to $k-1$:

$$\begin{aligned} S_1 = 1 + [\xi(t,t-1) + \xi(t,t)\theta_1(t)]B + \cdots + \\ [\xi(t,s+2) + \xi(t,s+3)\theta_1(s+3) + \cdots + \xi(t,s+2+q)\theta_q(s+2+q)]B^{k-2} \\ + [\xi(t,s+1) + \xi(t,s+2)\theta_1(s+2) + \cdots + \xi(t,s+1+q)\theta_q(s+1+q)]B^{k-1}. \end{aligned}$$

ii) For powers from k up to $k+q-1$:

$$\begin{aligned} S_2 = [\xi(t,s+1)\theta_1(s+1) + \cdots + \xi(t,s+q)\theta_q(s+q)]B^k \\ + [\xi(t,s+1)\theta_2(s+1) + \cdots + \xi(t,s-1+q)\theta_q(s-1+q)]B^{k+1} + \cdots + \\ [\xi(t,s+1)\theta_{q-1}(s+1) + \xi(t,s+2)\theta_q(s+2)]B^{k-2+q} + \xi(t,s+1)\theta_q(s+1)B^{k-1+q}. \end{aligned}$$

In view of eqs. (E.1) and (E.2) the above sums can be condensed as $S_1 = \sum_{r=0}^{k-1} \xi_q(t,t-r)B^r$ and $S_2 = \sum_{r=k}^{k-1+q} \xi_{s,q}(t,t-r)B^r$. Taking into account that $\Xi_t^{(k)}(B) \circ \Theta_t(B) = S_1 + S_2$, the result follows from the subsequent equalities

$$\Xi_t^{(k)}(B) \circ \Theta_t(B) = S_1 + S_2 = \sum_{r=0}^{k-1} \xi_q(t,t-r)B^r + \sum_{r=k}^{k-1+q} \xi_{s,q}(t,t-r)B^r = \Xi_{t,q}^{(k)}(B),$$

where the last equality arises from the definition in eq. (37). ■

Next we show Theorem 9, but first we apply eq. (8) in the main body of the paper

$$\xi^{(m)}(t,s) = \sum_{r=1}^{p-m+1} \phi_{m-1+r}(s+r)\xi(t,s+r). \tag{E.3}$$

for various values of m and s. For $m=1$, eq. (E.3) gives:

$$\xi(t,s) = \sum_{r=1}^{p} \phi_r(s+r)\xi(t,s+r). \tag{E.4}$$

Next, we evaluate $\xi(t,s)$, by applying eq. (E.4) for specific values of s starting from $t-1$ and moving backwards, that is for $s = t-1, t-2, ...$, as illustrated below:

$$\begin{array}{rcl} \xi(t,t-1) & = & \phi_1(t) \\ \xi(t,t-2) & = & \phi_1(t-1)\xi(t,t-1) + \phi_2(t) \\ \xi(t,t-3) & = & \phi_1(t-2)\xi(t,t-2) + \phi_2(t-1)\xi(t,t-1) + \phi_3(t) \\ \vdots & \vdots & \vdots \\ \xi(t,t-i) & = & \sum_{r=1}^{p} \phi_r(t-i+r)\xi(t,t-i+r). \end{array} \tag{E.5}$$

Applying eq. (E.3) for specific values of $2 \le m \le p$, we get:

$$\begin{array}{rcl} \xi^{(2)}(t,s) & = & \sum_{r=1}^{p-1} \phi_{1+r}(s+r)\xi(t,s+r) \\ \xi^{(3)}(t,s) & = & \sum_{r=1}^{p-2} \phi_{2+r}(s+r)\xi(t,s+r) \\ \vdots & \vdots & \vdots \\ \xi^{(p-1)}(t,s) & = & \phi_{p-1}(s+1)\xi(t,s+1) + \phi_p(s+2)\xi(t,s+2) \\ \xi^{(p)}(t,s) & = & \phi_p(s+1)\xi(t,s+1). \end{array} \tag{E.6}$$

Proof of Theorem 9. It suffices to show that

$$\Xi_t^{(k)}(B) \circ \Phi_t(B) = \Xi_{t,p}^{(k)}(B).$$

The left-hand side of the above equation, using eqs. (36) and (32), is equal to

$$\Xi_t^{(k)}(B) \circ \Phi_t(B) = \left(1 + \sum_{r=1}^{k-1} \xi(t, t-r) B^r\right) \circ \left(1 - \sum_{m=1}^{p} \phi_m(t) B^m\right). \tag{E.7}$$

Multiplying the two polynomials (using the properties of '$\circ$') and collecting terms with the same powers of the backshift operator, the right-hand side of eq. (E.7) (separated into two parts: S_1, S_2) gives:
i) For powers from zero up to $k-1$:

$$\begin{aligned} S_1 = 1 &+ [\xi(t,t-1) - \phi_1(t)]B + [\xi(t,t-2) - \phi_1(t-1)\xi(t,t-1) - \phi_2(t)]B^2 \\ &+ [\xi(t,t-3) - \phi_1(t-2)\xi(t,t-2) - \phi_2(t-1)\xi(t,t-1) - \phi_3(t)]B^3 \\ &\cdots \\ &+ \left[\xi(t,t-k+2) - \sum_{r=1}^{p} \phi_r(t-k+2+r)\xi(t,t-k+2+r)\right] B^{k-2} \\ &+ \left[\xi(t,t-k+1) - \sum_{r=1}^{p} \phi_r(t-k+1+r)\xi(t,t-k+1+r)\right] B^{k-1}. \end{aligned} \tag{E.8}$$

ii) For powers from k up to $k-1+p$:

$$\begin{aligned} S_2 = &- \left[\sum_{r=1}^{p} \phi_r(s+r)\xi(t,s+r)\right] B^{k-1+1} - \left[\sum_{r=1}^{p-1} \phi_{1+r}(s+r)\xi(t,s+r)\right] B^{k-1+2} \\ &- \left[\sum_{r=1}^{p-2} \phi_{2+r}(s+r)\xi(t,s+r)\right] B^{k-1+3} \\ &\cdots \\ &- [\phi_{p-1}(s+1)\xi(t,s+1) + \phi_p(s+2)\xi(t,s+2)]B^{k-1+p-1} \\ &- \phi_p(s+1)\xi(t,s+1)B^{k-1+p}. \end{aligned} \tag{E.9}$$

The final step of the proof is to notice that
i) on account of eq. (E.5), the ℓth ($\ell = 1, \ldots, k-1$) coefficient of the $(k-1)$th order polynomial in eq. (E.8) is equal to zero since $\xi(t,t-\ell) - \sum_{r=1}^{p} \phi_r(t-\ell+r)\xi(t,t-\ell+r) = 0$ and thus $S_1 = 1$, and
ii) in view of eqs. (E.4) and (E.6), the time varying coefficient of B^{k-1+m} ($m = 1, \ldots, p$) in the polynomial in eq. (E.9) is equal to minus the mth fundamental solution, $\xi^{(m)}(t,s)$, and, therefore: $S_2 = -\sum_{m=1}^{p} \xi^{(m)}(t,s)B^{k-1+m}$.
Thus eq. (E.7) gives

$$\Xi_t^{(k)}(B) \circ \Phi_t(B) = 1 - \sum_{m=1}^{p} \xi^{(m)}(t,s) B^{k-1+m} = \Xi_{t,p}^{(k)}(B),$$

as required. ■

Theorems 9 and 1 are equivalent. Therefore, the former implies the latter and vice versa. The proof of Theorem 10 (under the absolute summability condition in Proposition 5) follows along the same lines as the proof of Theorem 9 except that this time we let $k \to \infty$, thus only the argument in part i) applies (see eq. E.8).