
Working Paper Series

08/20

A UNIFIED THEORY FOR THE LARGE FAMILY OF TIME VARYING MODELS WITH ARMA REPRESENTATIONS: ONE SOLUTION FITS ALL

**MENELAOS KARANASOS, ALEXANDROS PARASKEVOPOULOS
and ALESSANDRA CANEPA**

A unified theory for the large family of ARMA models with varying coefficients: One solution fits all

M. Karanasos^{†▷}, A. Paraskevopoulos[‡], A. Canepa^{*}

[†]Brunel University London, [‡]University of Piraeus, ^{*} University of Turin

Abstract

For the large family of ARMA models with variable coefficients we obtain an explicit and computationally tractable solution that generates all their fundamental properties, including the Wold-Cr amer decomposition and their covariance structure, thus unifying the invertibility conditions which guarantee both their asymptotic stability and main properties. The one sided Green's function, associated with the homogeneous solution, is expressed as a banded Hessenbergian formulated exclusively in terms of the autoregressive parameters of the model. The proposed methodology allows for a unified treatment of these 'time varying' systems.

We also illustrate mathematically one of the focal points in Hallin's (1986) analysis. Namely, that in a time varying setting the backward asymptotic efficiency is different from the forward one. Equally important it is shown how the linear algebra techniques, used to obtain the general solution, are equivalent to a simple procedure for manipulating polynomials with variable coefficients.

The practical significance of the suggested approach 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 reduces and stabilizes to even lower levels than the pre-1976 period.

Keywords: asymptotic forecasting efficiency, ARMA process, General solution, Green's function, Hessenbergians, invertibility, structural breaks, symbolic product of operators, time dependent persistence, time varying models, Wold-Cr amer decomposition.

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

[▷]M. Karanasos and A. Paraskevopoulos are the lead authors having defined the theoretical models, and having derived the theoretical results. A. Canepa is the third 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, G. Kapetanios and A. Magdalinos 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), 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), University of Reading (ICMA Centre), King's College London (Business School), 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: SEM Conference (MIT, Boston, July 2017); 2nd and 1st IAAE Conferences (University of Macedonia, Greece, June 2015; Queen Mary University of London, June 2014, respectively), 8th and 7th CFE Conferences (University of Pisa, December 2014; Senate House, University of London, December 2013, respectively), 68th ESEM European Meeting (Toulouse, France, August 2014), 35th ISF Conference (Erasmus University, Rotterdam, July 2014), 2nd ISNPS Conference (Cadiz, Spain, June 2014), 13th and 12th CRETE Conferences (AUEB, Crete and Naxos, July 2014 and 2013, respectively), 9th and 8th BMRC-QASS conferences on Macro and Financial Economics (Brunel University, London, May 2014 and 2013, respectively), the SNDE 22nd and 21st Annual Symposiums (The City University of New York, April 2014; University of Milan-Bicocca, March 2013, respectively), 1st RastaNews Annual Conference (Palazzo Clerici, Milan, January 2014), 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 lead us to develop Sections 5.2.3, 5.3 and 6.

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 modeling 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-HARMA). There are two large classes of stochastic processes. The ones with deterministically and those with stochastically 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 (see, for details, Hallin, 1978, and Grillenzoni, 1990), is not applicable when time variation is present.

Miller (1968) established an explicit (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. The Green’s function itself was not explicitly described, however, in order to overcome this difficulty Hallin and Melárd (1977) proposed a recursive method for its computation. Moreover, Hallin (1978, for the multivariate case) presented conditions in terms of the Green’s function, which make the above mentioned models invertible. An efficient explicit representation of the Green’s function depends upon the availability of a fundamental set of solutions whose elements (known as fundamental or linearly independent solutions) are explicitly formulated and computationally 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), called here principal determinant. The entries of the principal matrix determinant are defined in terms of the autoregressive (resp. moving average) coefficients. The first fundamental solution sequence is represented by the principal determinant. The remaining $p - 1$ fundamental solutions are also expressed as Hessenbergians. 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.

The Hessenbergian solution structure yields a simple (necessary and sufficient) condition, which guarantees the asymptotic backward and forward stability of the process. It also generates easily handled analytic representations (as infinite sums are involved) for the fundamental properties of the aforementioned models (such as forecasts, the unconditional first and second moments, the Wold-Cr amer decomposition and impulse response functions) and for the conditions that ensure their existence.

We also illustrate mathematically one of the focal points in Hallin’s (1986) analysis. Namely, that in a time varying setting two forecasts with identical forecasting horizons, but at different times, have different mean squared errors. This implies that backward asymptotic efficiency (when the initial observation vector shifts into the remote past) is 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. Equally important we show 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 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

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

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

in Marrero and Tomeo (2012, 2016) and Paraskevopoulos and Karanasos (2019), can be applied to derive a compact representation for the principal determinant. This result modernizes and enhances the explicit representations of the ‘time varying’ models and their fundamental properties.

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 employing the so called infinite elimination algorithm) for the homogenous part of a HARMA(p, q) model with variable coefficients constructs the p elements of a fundamental solution set. We should highlight the fact that in the present paper a self-contained proving process is demonstrated without invoking the infinite elimination algorithm.

For various lengths of the forecasting horizon, $k = 1, 2, \dots$, it is described as a sequence of TV-HARMA representations of order $(p+k-1, q+k-1)$. The first term of the sequence, that is when $k = 1$, coincides with the initial model, whereas when $k \rightarrow \infty$ we obtain its unique Wold-Cr amer decomposition. In this interpretation the fundamental solutions are portrayed as autoregressive coefficients.

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 reduces 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 and the principal determinant. The next Section presents the general solution for the large 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 Section 6 deals with the invertibility issue. 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 HARMA

The aim of this section is to provide explicit solution expressions for a fairly large family of ARMA models with time dependent coefficients.

2.1 Preliminaries

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 (Ω, \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 . Finally, $H = L_2(\Omega, \mathcal{F}, P)$ 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-HARMA(p, q), is defined as

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

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 $L, M \in \mathbb{R}_{>0}$. The forcing term of eq. (1) is assigned to be the time varying drift plus the moving average term: $v_t = \varphi(t) + u_t$.

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: Karanasos et al., 2014 [which is 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 7).

The above process nests the TV-HAR(p) model as a special case when $q = 0$ and the ARMA(p, q) specification when the drift, the autoregressive coupled with the moving average coefficients, and the variances are all constants using the conventional identifications: $\varphi(t) = \varphi$, $\phi_m(t) = \phi_m$, $\theta_l(t) = \theta_l$, $\sigma^2(t) = \sigma^2$ for all t .

If the researcher wants to obtain a sensible insight into the ‘causal’ structure of the observed process then the integer t , which is an arbitrary point in time, should be considered as the present and $s = t - k$, $k \in \mathbb{Z}_{\geq 1}$ as the past. In this case the Wold-Cr amer decomposition, which is obtained when $s \rightarrow -\infty$ (or, equivalently, when $k \rightarrow \infty$) and relates the process under study to its innovations, should be examined (see Section 5.2.1). If forecasting is the main objective then t should be considered as the future, indicating an arbitrary point of time at which a forecast is intended. In this case k is the forecasting horizon and s represents the ‘present’ time (that is the right end-point of the information time interval $\mathbf{I}_s = (-\infty, s] \cap \mathbb{Z}$) such that at time s and back the information sequence $\{y_s, y_{s-1}, \dots\}$ is accessible and generates the σ -field \mathcal{F}_s . Moreover, p consecutive terms of the information sequence are the components of the initial value vector ($y_{s+1-p} = c_p, \dots, y_s = c_1$). Backward forecasting refers to the case where we keep the future time t fixed and we move the present, that is the initial observation on time s , backwards. Backward asymptotic efficiency directs attention to $s \rightarrow -\infty$ (see Section 5.1). Forward forecasting refers to the case where we keep the present time s fixed and we move the future time t forwards. Therefore, asymptotic forward (called by Hallin, 1986, Granger-Andersen) efficiency directs attention to $t \rightarrow \infty$.

One of the goals of this work is to obtain the unique inverse of the time varying autoregressive (AR) polynomial associated with eq. (1) and denoted by $\Phi_t(B)$, where B is the backshift or lag operator (see Section 6). In a time varying environment, the usual procedure employs the Green’s function instead of the characteristic polynomials, that 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 (therefore in the literature recursive methods are used to compute them, see, for example, Grillenzoni, 2000, and Azrak and Mélard, 2006). To overcome this difficulty, we introduce the principal determinant below and make use of the so called multiplication skew operator (see the analysis in Section 6.1).

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 opposite direction. That is, they went the other way around and by finding an explicit and compact representation of the fundamental solutions for TV-HARMA 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-HARMA 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

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

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

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

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

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 $y_s = y_{s-1} = \dots = y_{s+1-p} = 0$ is given by

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

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

$$\begin{aligned} \xi_q(t,r) &= \xi(t,r) + \sum_{l=1}^q \xi(t,r+l) \theta_l(r+l), \text{ for } r = s+1, \dots, t, \\ \xi_{s,q}(t,r) &= \sum_{l=s-r+1}^q \xi(t,r+l) \theta_l(r+l), \text{ for } r = s+1-q, \dots, s. \end{aligned} \quad (12)$$

As $\xi_q(t,r)$ is equal to $\xi(t,r)$ plus a sum of terms which involves the first q ‘lead’ values of $\xi(t,r)$ (a banded Hessenbergian), it can also be expressed as a banded Hessenbergian (the proof is deferred to the online Appendix F.1). 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 $l \in \mathbb{Z}_{\geq 1}$, which coincides with the corresponding values of $\xi(t,r)$ (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 stochastic component of the particular solution (11) can be decomposed into two parts as follows:

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

A formal proof of this result is given in Appendix A.4. 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 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 information sequence). The above equation together with the general solution (14) will be used to obtain all 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 initial values y_{s+1-m} , $m = 1, 2, \dots, p$, is*

$$y_{t,s} = \underbrace{\sum_{m=1}^p \xi^{(m)}(t,s)y_{s+1-m}}_{\text{Homogeneous Solution}(y_{t,s}^{\text{hom}})} + \underbrace{\sum_{r=s+1}^t \xi(t,r)\varphi(r)}_{\text{Particular Solution: Deterministic Part}} + \underbrace{\left(\sum_{r=s+1}^t \xi_q(t,r)\varepsilon_r + \sum_{r=s+1-q}^s \xi_{s,q}(t,r)\varepsilon_r \right)}_{\text{Particular Solution: Stochastic Part}} \quad (14)$$

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 initial values, taken from the information sequence. The second sum (the deterministic part of the particular solution, see eq. (11)) is formed by products involving the principal determinant $\xi(t,r)$ times 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) $\sum_{r=t}^t \xi(t,r)[\varphi(r) + u_t] = \varphi(t) + u_t$.*

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 introduced in Hallin (1984), whereas Hallin (1986) show the close connection that exists between the convergence of matrix continued fractions and the existence of dominated solutions of multivariate difference equations of order 2.

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 periodicity, unknown 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) HAR 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). 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).

3.4 Gegenbauer Functions as Hessenbergians

We conclude this Section with an example. In particular, we show how the Gegenbauer functions can be expressed as Hessenbergians.

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:

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

In view of eq. (14), the TV-HARMA process in eq. (1) is backward asymptotically stable if for every t , arbitrary but fixed, the sequence of homogeneous solutions $\{y_{t,s}^{hom}\}_{s \leq t}$ of eq. (5), as a function of s , whose terms are given in (7), tends to zero as $s \rightarrow -\infty$ (that is $\lim_{s \rightarrow -\infty} y_{t,s}^{hom} = 0$), for any initial condition vector $\mathbf{c}' = [c_m]_{1 \leq m \leq p}$ such that $y_{s+1-m} = c_m$ for all $s < t$ and each $m : 1 \leq m \leq p$. That is the initial condition vector \mathbf{c} moves further to the past ($s \rightarrow -\infty$), while the future time t is fixed. The TV-HARMA process in eq. (1) is forward asymptotically stable if for every s , arbitrary but fixed, $\lim_{t \rightarrow \infty} y_{t,s}^{hom} = 0$ for any initial condition vector \mathbf{c} . 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 $1 \leq m \leq p$, then a necessary and sufficient condition for the TV-HARMA model in eq. (1) to be backward asymptotically stable is: $\lim_{s \rightarrow -\infty} \xi(t, s) = 0$ for each t . Moreover, the condition $\lim_{t \rightarrow \infty} \xi(t, s) = 0$ is necessary and sufficient for the TV-HARMA 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-HARMA process to be backward asymptotically stable is $\xi(t, s) \xrightarrow{P} 0$, as $s \rightarrow -\infty$ (probability convergence) for each t . Moreover, the condition $\lim_{t \rightarrow \infty} \xi(t, s) \xrightarrow{P} 0$ is necessary and sufficient for the TV-HARMA model with stochastic coefficients to be forward asymptotically stable.

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.⁴

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 $\lim_{k \rightarrow \infty} \xi_k = 0$, 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, 2016).

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 $\lim_{s \rightarrow -\infty} \xi(t, s) = 0$ 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 $\lim_{t \rightarrow \infty} \phi_m(t) = \phi_m$ and $\sum_{r=1}^{\infty} \rho^{2r} < \infty$, where $\rho = \rho(\Phi) + \epsilon$, $\epsilon > 0$, $\rho(\Phi) = \max\{|z_m|, \Phi(z_m^{-1}) = 0\}$ with $\Phi(z^{-1}) = 1 - \sum_{m=1}^p \phi_m z^{-m}$ (see eq. (8) in Kowalski and Szynal, 1991).⁵*
ii) The deterministically varying polynomial $\Phi_t(z^{-1})$ should have roots whose realizations entirely lie

⁴Azrak 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, which 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-HARMA processes (see Azrak and Mélard, 2006, and the references therein).

⁵Kowalski and Szynal (1991) showed that $\rho(\Phi)$ is the spectral radius of the matrix $\Phi = \begin{bmatrix} 0 & 0 & \cdots & 0 & -\phi_m \\ 1 & 0 & \cdots & 0 & -\phi_{m-1} \\ 0 & 1 & \cdots & 0 & -\phi_{m-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -\phi_1 \end{bmatrix}$

(see page 75 in their paper).

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

As an example, consider the logistic smooth transition AR(1) model (see, for example, Teräsvirta, 1994) where 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 $F(t; \gamma, \tau) = [1 + e^{\gamma(t-\tau)}]^{-1}$, $\gamma \in \mathbb{R}_{\geq 0}$, $\tau \in \mathbb{Z}$, 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, $\lim_{s \rightarrow -\infty} \xi(t, s) = 0$ if and only if $|\phi_1| < 1$, whereas $\lim_{t \rightarrow \infty} \xi(t, s) = 0$ if and only if $|\phi_2| < 1$.

We conclude this section with another example. Consider the periodic AR(1; ℓ) model where $\ell \in \mathbb{Z}_{\geq 1}$ is the number of seasons (i.e., quarters) and let ϕ_r , $r = 1, \dots, \ell$ 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 $\lim_{s \rightarrow -\infty} \xi(t, s) = 0$ (or equivalently $\lim_{k \rightarrow \infty} \xi(t, s) = 0$). The necessary and sufficient condition is $|\prod_{r=1}^{\ell} (\phi_r)| < 1$.

5 Second Order Properties

Having specified a general method for manipulating the TV-HARMA 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-Cr amer 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 HARMA structure and deterministic coefficients; we will term these processes deterministic time varying (DTV). To save space we will discuss the case of the stochastic coefficients in a future paper. However, in Appendix D 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-HARMA(p, q) model.

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

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

$$\mathbb{E}(y_t | \mathcal{F}_s) = \sum_{m=1}^p \xi^{(m)}(t, s) y_{s+1-m} + \sum_{r=s+1}^t \xi(t, r) \varphi(r) + \sum_{r=s+1-q}^s \xi_{s,q}(t, r) \varepsilon_r. \quad (15)$$

In addition, the forecast error for the above k -step-ahead predictor, $\mathbb{FE}_{t,s} = y_t - \mathbb{E}(y_t, s | \mathcal{F}_s)$, and the associated $\text{MSE}_{t,s} = \text{Var}(\mathbb{FE}_{t,s})$ are given by

$$\mathbb{FE}_{t,s} = \sum_{r=s+1}^t \xi_q(t, r) \varepsilon_r, \quad \text{MSE}_{t,s} = \sum_{r=s+1}^t \xi_q^2(t, r) \sigma^2(r) \quad (16)$$

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

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 obtain an explicit formula for the optimal linear predictors, which enables us to study the asymptotic forecasting efficiency of the DTV-HARMA(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 Appendix D 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-HAR specification and a model with coefficients that follow 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 not any more equal. At this aim, 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:

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

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

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

A comparison between MSE_{t_1, s_1} and MSE_{t_2, s_2} associated with a time varying environment, in the general case, entails that : $\text{MSE}_{t_1, s_1} \neq \text{MSE}_{t_2, s_2}$. 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:

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

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 coupled with sufficient and necessary conditions ensuring their existence. In what follows we assume that $\sup_{s \leq t} |\varphi(s)| < \infty$ and $\sup_t |\theta_l(t)| < \infty$ for all t and l such that $1 \leq l \leq q$.

Sufficient and necessary conditions ensuring that the first and second unconditional moments of the DTV-HARMA family of processes exist are given below. In particular, in the following Proposition we provide sufficient and necessary conditions ensuring that i) $\sum_{r=-\infty}^t \xi(t, r)$ convergence for all t , and thus for $\sum_{r=s+1}^t \xi(t, r) \varphi(r)$ to converge as $s \rightarrow -\infty$ for all t , and ii) $\sum_{r=-\infty}^t \xi_q^2(t, r) \sigma^2(r) < \infty$ for all t , and thus for $\sum_{r=s+1}^t \xi_q(t, r) \varepsilon_r$ to converge in L_2 as $s \rightarrow -\infty$ for all t (backward asymptotic efficiency). We recall that $\xi_q(t, r)$ has been defined in Definition 1.

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

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

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

$$\mathbb{E}(y_t) = \sum_{r=-\infty}^t \xi(t, r) \varphi(r). \quad (17)$$

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

$$\text{Var}(y_t) = \sum_{r=-\infty}^t \xi_q^2(t, r) \sigma^2(r). \quad (18)$$

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

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

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

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

$$\begin{array}{ccccc} \sum_{r=-\infty}^t \xi(t, r) \varphi(r) \in \mathbb{R} & \Leftarrow & \sum_{r=-\infty}^t |\xi(t, r)| < \infty & \Rightarrow & \sum_{r=-\infty}^t \xi_q^2(t, r) \sigma^2(r) \in \mathbb{R}_{\geq 0} \\ \Downarrow & \swarrow & \Downarrow & \searrow & \Downarrow \\ \lim_{s \rightarrow -\infty} \xi(t, s) \varphi(s) = 0 & \Leftarrow & \lim_{s \rightarrow -\infty} \xi(t, s) = 0 & \Rightarrow & \lim_{s \rightarrow -\infty} \xi_q^2(t, s) \sigma^2(s) = 0 \end{array} \quad (19)$$

Commutative Diagrams

Notice that the absolute summability condition is the unique condition guaranteeing the first and second unconditional moments coupled with the backward asymptotic stability. A proof of Proposition 5 along with the Diagrams in (19) is provided in Appendix C.1. Notice that the mean is the same for both the AR and the ARMA processes.

5.2.1 Wold-Cr amer Decomposition

In view of the general solution in eq. (14) we obtain: $y_t \stackrel{L_2}{=} \lim_{s \rightarrow -\infty} y_{t,s}^{par}$. The Wold-Cr amer decomposition⁶ (see Cr amer, 1961) of the DTV-HARMA(p, q) model is described in the following Theorem.

Theorem 3 *Let the absolute summability condition in Proposition 5 holds. The Wold-Cr amer decomposition takes the form*

$$y_t = \sum_{r=-\infty}^t \xi(t, r) \varphi(r) + \sum_{r=-\infty}^t \xi_q(t, r) \varepsilon_r. \quad (20)$$

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 $\lim_{s \rightarrow -\infty} \mathbb{F}\mathbb{E}_{t,s}$ is the zero mean random part. Hallin (1978), Singh and Peiris (1987), Kowalski and Szynal (1991), Grillenzoni (2000), and Azrak and M elard (2006) obtained the Wold-Cr amer decomposition through recursion. In sharp contrast, eq. (20) in Theorem 3 provides an analytic formula for the one-sided MA representation.

⁶As pointed out by Hallin (1986) since a non-stationary generalization of Wold's result was given by Cram er it is referred to as Wold-Cram er decomposition.

5.2.2 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-HARMA(p, q) process.

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

$$\gamma_t(\ell) = \sum_{r=-\infty}^{t-\ell} \xi_q(t, r) \xi_q(t-\ell, r) \sigma^2(r). \quad (21)$$

The time varying variance of y_t , that is $\gamma_t(0) = \text{Var}(y_t)$, is given by eq. (18). 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. (21).

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

Remark 2 *Azrak and M elard (2006) considered the asymptotic properties of QML estimators for the TV HARMA 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.

5.2.3 Two More Examples

In this Section we consider two examples concerning AR(1) processes with variable autoregressive coefficients, taken from Azrad and M elard (2006). In the first example, the autoregressive coefficient is a periodic function of time. In the second example, the coefficient is an exponential function of time. In particular, we consider the process 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 . In the first example, the autoregressive coefficient is given by $\phi(t) = \beta_{t-n[t/n]}$, where $n \in \mathbb{Z}_{\geq 1}$ 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-Cr amer decomposition (if and only if $|\beta| < 1$, where $\beta = \beta_0 \cdot \beta_1 \cdots \beta_{s-1}$) is given by

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

with unconditional variance

$$\text{Var}(y_t) = \sigma^2 \sum_{r=-\infty}^t \xi^2(t, r),$$

where

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

and, therefore

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

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

In the second example (see example 2 in Azrak and Mélard, 2006), the autoregressive coefficient is given by

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

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

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

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).

5.2.4 Forward Asymptotic Efficiency

As a consequence of the work of Hallin (1986), if a researcher wants to obtain a sensible insight into the ‘causal structure’ of the observed process then he should examine the Wold-Cr amer decomposition relating the process under study to its innovations (see Theorem 3 in Section 5.2.1 of the above cited reference). If forecasting is the main objective then the forecast produced 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 stationary processes take place, these two issues coincide. In the non-stationary case, 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 the initial observation on time 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 $t \rightarrow \infty$ 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).

Sufficient conditions for the forward asymptotic efficiency of a DTV-HARMA(p, q) process are presented below.

Proposition 7 *Let $\sum_{r=-\infty}^t |\xi(t, r)| < \infty$ for each $t \in \mathbb{Z}$ (absolutely summability condition). Let us call*

$F_t = \sum_{r=-\infty}^t |\xi(t, r)|$. *If $\{F_t\}_{t \in \mathbb{Z}_{\geq 0}}$ is bounded, then the sequence $\{\text{MSE}_s(t)\}_s$ defined by $\text{MSE}_s(t) = \text{MSE}_{t,s}$, is uniformly bounded.*

A proof of the above Proposition is provided in Appendix C.4. In Lemma C3 a weaker condition is assumed, which guarantees the boundedness of $\text{MSE}_{t,s}$ as a function of t for each s . In particular, if $F(t, s) \stackrel{\text{def}}{=} \sum_{r=s+1}^t |\xi(t, r)|$ is bounded, as a function of t for each s , we show that $\{\text{MSE}_{t,s}\}_t$ is also bounded. As a direct consequence, for every arbitrary but fixed s either $\{\text{MSE}_{t,s}\}_t$ is convergent in t (namely $\lim_{t \rightarrow \infty} \text{MSE}_{t,s}$ exists in $\mathbb{R}_{\geq 0}$) or oscillating with oscillation: $\Omega(s) = \lim_{t \rightarrow \infty} (\sup_{r \geq t} \text{MSE}_{r,s} - \inf_{r \geq t} \text{MSE}_{r,s})$.

As the absolutely summability condition in Proposition 7 holds, all the results previously stated (asymptotic stability, Wold-Cr amer Decomposition, e.t.c) also hold.

The proof of Theorem 4 essentially repeats the arguments of 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 linear ‘time varying’ specifications 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. An advantage of our technique is that it can be applied with ease in a multivariate setting and provides a solution to the problem at hand without adding complexity.

6 Time Varying Polynomials

In Section 3 we employed techniques of linear algebra in order to obtain the general solution of the TV-HARMA(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 denote 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-HARMA(p, q) model are denoted as:

$$\Phi_t(B) = 1 - \sum_{m=1}^p \phi_m(t)B^m, \quad \Theta_t(B) = 1 + \sum_{l=1}^q \theta_l(t)B^l. \quad (24)$$

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

$$\Phi_t(B)y_t = \varphi(t) + \Theta_t(B)\varepsilon_t. \quad (25)$$

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 generating sequence $\{\xi(t, s)\}_{t \geq s+1-p}$ cannot be obtained as in stationarity, by expanding in Taylor series the rational polynomial $\Phi_t(B)^{-1}$. As an alternative Hallin (1986) introduced some results on difference operators such as the symbolic product of operators, which has also been termed by researchers in the field of engineering 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 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.

6.1 The Skew Multiplication Operator

In a time varying environment, the time varying polynomial operators in eq. (24) can be manipulated by using the ‘skew’ multiplication operator “ \circ ” defined by

$$B^i \circ B^j = B^{i+j} \quad \text{and} \quad B^i \circ f(t) = f(t-i)B^i, \quad (26)$$

where $f(t)$ is a function of time. This time-domain multiplication operation is associative but noncommutative (see Karmen, 1988, Bouthellier and Ghosh, 1988, and Mrad and Farag, 2002). Using the properties of “ \circ ”, from eq. (25), under the necessary and sufficient conditions in Proposition 5, we can obtain the unique inverse of $\Phi_t(B)$, that is $\Phi_t(B)^{-1} \circ \Phi_t(B) = 1$.

6.2 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-HARMA 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

$$\Xi_{t,p}^{(k)}(B) = 1 - \sum_{m=1}^p \xi^{(m)}(t, s) B^{k-1+m} \quad (27)$$

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

$$\Xi_t^{(k)}(B) = \sum_{r=s+1}^t \xi(t, r) B^{t-r} = \sum_{r=0}^{k-1} \xi(t, t-r) B^r \quad (28)$$

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

Remark 3 $\Xi_{t,p}^{(k)}(B)$ 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, $\lim_{k \rightarrow \infty} \Xi_{t,p}^{(k)}(B) = \Xi_{t,p}(B) = 1$, since $\lim_{s \rightarrow -\infty} \xi^{(m)}(t, s) = 0$ (see in Appendix B, Lemma B1).

Remark 4 $\Xi_t^{(k)}(B)$ 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, $\xi(t, t-k)$. It can also be expressed as a single banded Hessenbergian (see the online Appendix F.2). Notice also that $\Xi_t^{(1)}(B) = 1$.

Next we define two additional time varying polynomial operators associated with the MA part of the particular solution and the Wold-Cr amer decomposition respectively.

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

$$\Xi_{t,q}^{(k)}(B) = \sum_{r=0}^{k-1} \xi_q(t, t-r) B^r + \sum_{r=k}^{k-1+q} \xi_{s,q}(t, t-r) B^r. \quad (29)$$

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 5 $\Xi_{t,q}^{(k)}(B)$ is a polynomial of order $q+k-1$ associated with the stochastic 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. (29) (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).

6.3 General Solution

The next Proposition is analogous to Proposition 1.

Proposition 8 *Similarly to the decomposition in eq. (13) we can also decompose $\Xi_t^{(k)}(B) \circ u_t$ into two parts as follows*

$$\Xi_t^{(k)}(B) \circ u_t = \Xi_{t,q}^{(k)}(B)\varepsilon_t, \quad (30)$$

where the two parts of the right side of (30) are indicated in (29).

The following Theorems can be deduced by applying the properties of the skew multiplication operator “ \circ ” (see eq. (26)). The first of these Theorems (Theorem 5) is equivalent to Theorem 1.

Theorem 5 *The general solution in eq. (14), can be equivalently expressed, in terms of the polynomial operators given in Definitions 2 and 3, as:*

$$\Xi_{t,p}^{(k)}(B)y_t = \Xi_t^{(k)}(B) \circ [\varphi(t) + u_t] = \Xi_t^{(k)}(B) \circ \varphi(t) + \Xi_{t,q}^{(k)}(B)\varepsilon_t. \quad (31)$$

Clearly, the results for the pure AR model are obtained by setting $u_t = \varepsilon_t$ in the first equality of eq. (31) or by noticing that in the second equality of eq. (31) $\Xi_{t,0}^{(k)}(B) = \Xi_t^{(k)}(B)$ (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 defined in Definitions 2(IIb) and 3(II), respectively.

The next Theorem shows that $\Xi_{t,q}(B)$ is the time varying Wold-Cr amer polynomial operator associated with the TV-HARMA(p, q) model.

Theorem 6 *Under the absolute summability condition in Proposition 5, the Wold-Cr amer operators are given by*

$$\Xi_{t,q}(B) = \Phi_t(B)^{-1} \circ \Theta_t(B) \quad \text{and} \quad \Xi_t(B) = \Phi_t(B)^{-1}. \quad (32)$$

The Wold-Cr amer decomposition in Theorem 3 can be written in terms of the Wold-Cr amer operator as

$$y_t = \Xi_t(B) \circ \varphi(t) + \Xi_{t,q}(B)\varepsilon_t, \quad (33)$$

which of course implies that

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

Appendix E contains the proofs of Theorems 5 and 6.

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 I we present a summary of the ‘time varying polynomials’ results.

6.4 Interpretation

In Section 3.1 we show that each fundamental solution $\xi^{(m)}(t, s)$, $m = 1, \dots, p$, is the solution of eq. (5) under the initial conditions $y_{s+1-m} = 1$ and $y_{s+1-r} = 0$ for $1 \leq r \leq p$ and $r \neq m$. In this Section we make use of the general solution, expressed in terms of the three time varying polynomial operators (see Theorem 5), in order to provide a further interpretations of the p fundamental solutions.

Infinite TV-HARMA($p + k - 1, q + k - 1$) Models

First, we will set side by side $\left(1 - \Xi_{t,p}^{(k)}(B)\right) y_t$ (see eq. (27)) and $(1 - \Phi_t(B))y_t$ (see eq. (25)). What does $\left(1 - \Xi_{t,p}^{(k)}(B)\right)$ in comparison to $(1 - \Phi_t(B))$? It simply shifts, $k - 1 (= t - s - 1)$ periods backward in

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

Corollary 2 *The general solution of y_{t_1+l} in eq. (34), subject to the initial conditions y_{t_2} , y_{t_2-1} , is given by*

$$y_{t_1+l,t_2} = \sum_{r=t_2+1}^{t_1+l} \xi(t_1+l,r)(\varphi_r + \varepsilon_r) + \xi(t_1+l,t_2)y_{t_2} + \phi_{2,1}\xi(t_1+l,t_2+1)y_{t_2-1}, \quad (35)$$

7.1 Second Moment Structure

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

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 1 (Second-Order). $|\lambda_{m,i}| < 1$, $m = 1, 2$, for $i = 1, 3$.

Assumption 1 implies that the DAB-HAR(2;2)-process is second-order.

The following Proposition states expressions for the time varying variance of y_{t_1+l} in eq. (35).

Proposition 9 *Consider the general model in eq. (34). Then under Assumption 1, the $\text{Var}(y_{t_1+l})$ is given by*

$$\text{Var}(y_{t_1+l}) = A_{t_1+l}\sigma_1^2 + B_{t_1+l}\sigma_2^2 + C_{t_1+l}\sigma_3^2, \quad (36)$$

where

$$A_{t_1+l} = \sum_{r=1}^l \xi^2(t_1+l, t_1+r), \quad B_{t_1+l} = \sum_{r=0}^{t_1-t_2-1} \xi^2(t_1+l, t_1-r),$$

$$C_{t_1+l} = \frac{[(1 - \phi_{2,3})(\xi^2(t_1+l, t_2) + \phi_{2,1}^2\xi^2(t_1+l, t_2+1)) + 2\phi_{1,3}\xi(t_1+l, t_2)\phi_{2,1}\xi(t_1+l, t_2+1)]}{(1 + \phi_{2,3})[(1 - \phi_{2,3})^2 - \phi_{1,3}^2]}.$$

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 $\text{Var}(y_{t_2+l})$, which is equivalent to the case of one break (notice that in this case $B_{t_2+l} = 0$):

$$\text{Var}(y_{t_2+l}) = A_{t_2+l}\sigma_2^2 + C_{t_2+l}\sigma_3^2.$$

Finally, if in addition we set $l = 0$ then we obtain the $\text{Var}(y_{t_2})$, 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:

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

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

7.2 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 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 $P(x_t|e_t) = \text{Var}(x_t)/\text{Var}(e_t)$. 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. (36) that $P(y_{t_1+l}|\varepsilon_{t_1+l}) = \frac{\text{Var}(y_{t_1+l})}{\sigma_1^2}$, is given by

$$P(y_{t_1+l}|\varepsilon_{t_1+l}) = A_{t_1+l} + B_{t_1+l} \frac{\sigma_2^2}{\sigma_1^2} + C_{t_1+l} \frac{\sigma_3^2}{\sigma_1^2}. \quad (37)$$

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.⁷

8 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 United States. We use quarterly data on the GDP deflator as 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 annualized rate calculated as $\pi_t = 400(\ln(P_t/P_{t-1}))$.

In term 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 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.

8.1 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 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 overestimate inflation persistence. In this regards, a growing body of research has found evidence that monetary policy target has an impact on the persistence properties 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 sluggishness 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.

⁷Cogley 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: $S_0 = \frac{\sigma_\varepsilon^2}{2\pi(1-\phi_1-\phi_2)^2}$.

Table 2. Unit root Tests.

Test Statistic	
ADF	-3.229**
ERS	-3.154**
MZ _{α}	-19.642*
MZ _{t}	-3.1331*

8.2 Structural Breaks

Next we estimate AR(p) models with abrupt structural breaks. The optimal model is the DAB-AR(2; 2) model in eq. (34). 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, \dots, 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. Bai and Perron (2003) 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 ($i = 1, \dots, l + 1$). 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 is 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 high inflation rate after the end of the Bretton Woods period. The second break occurred in 1986 when the Fed embarked in 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. ⁸

⁸McConnell and Perez-Quiros (2000) have detected a fall in the volatility of output after 1984 as well.

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		
$H_0 : 0 \text{ vs } 1$	57.96**	13.98		
$H_0 : 1 \text{ vs } 2$	18.13**	15.72		
$H_0 : 2 \text{ vs } 3$	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 parenthesis.

8.3 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 inflation level was stubbornly high. After 1986 the smaller magnitude of the estimated drift reflects the lower average inflation rates that 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 slightly reduced in the last thirty years ($\sigma_3 = 2.160$). However, it did not go back to the relatively low level before the 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 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.

8.4 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 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.⁹ 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 model. 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 root mean square forecast error (RMSE), *ii*) the mean absolute error (MAE), *iii*) the Theil Inequality Coefficient (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 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.

8.5 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.¹⁰ In this Section we depart from their study in an important way, that is we contribute to

⁹For forecasting under structural breaks, see for example, Pesaran and Timmermann (2005).

¹⁰Pivetta 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 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 foresee 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.

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 7.1 and 7.2), 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 persistence of inflation 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, $E(\pi_t)$, increases by 59.3% in the second period and it decreases by 88% in the third period.

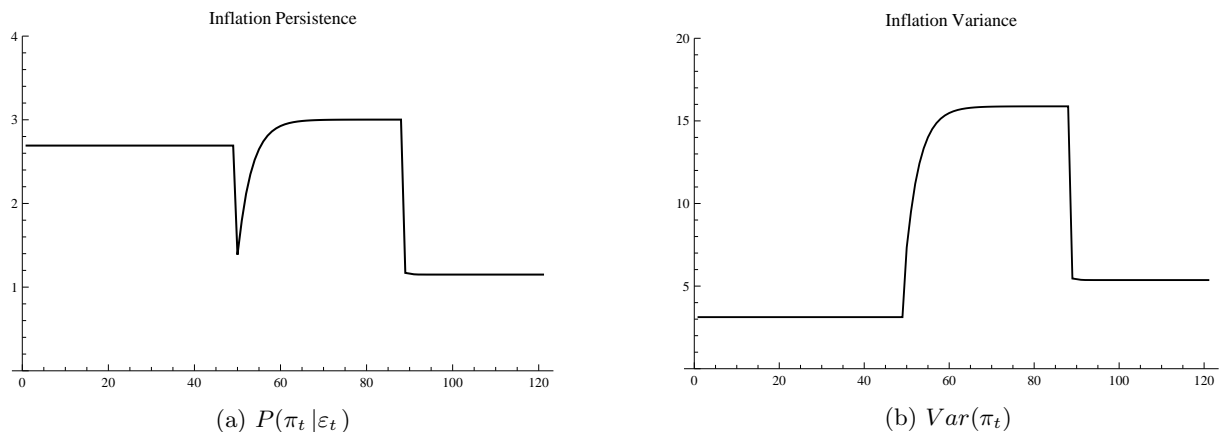
The last three columns report the three second-order measures of persistence (S_0 , $P(\pi_t|\varepsilon_t)$, and $\text{Var}(\pi_t)$). For two out of the three measures the post-1986 period exhibits the lowest persistence whereas in the second period the persistence is the highest. The variance of inflation, $\text{Var}(\pi_t)$, 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	$P(\pi_t \varepsilon_t)$	$\text{Var}(\pi_t)$
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 of the measures $P(\pi_t|\varepsilon_t)$ and $\text{Var}(\pi_t)$ reflect the dynamics of the second-order time varying inflation persistence. Details of how we construct the graphs are presented in the online Appendix H. In the x-axis each unit represents a year-quarter starting with 1964Q₂, chosen as first. In particular, 1976Q₃ = 48 (48-th year-quarter) and 1986Q₂ = 87 (87-th year-quarter).

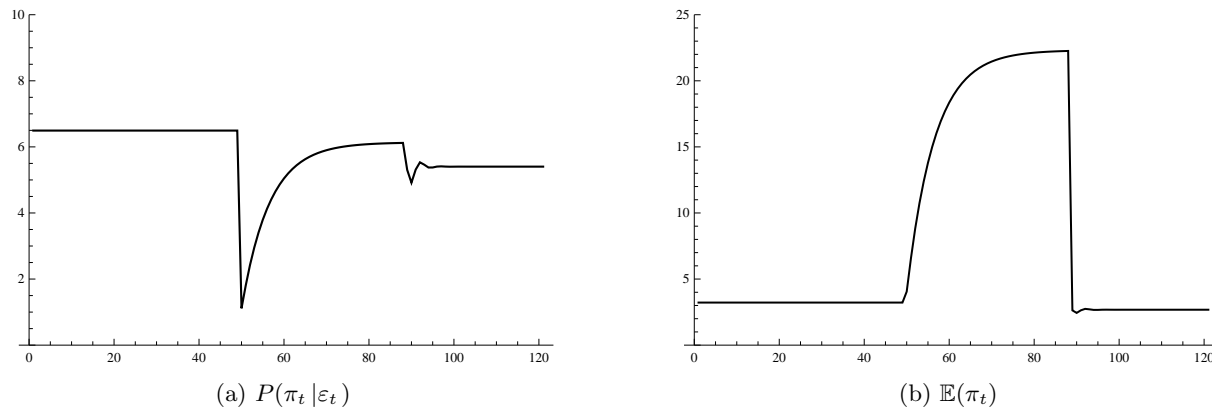


Second-Order Time Varying Persistence

The main futures for the graph of the inflation variance $\text{Var}(\pi_t)$ 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 1976 Q_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 1986 Q_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 1976 Q_4 followed shortly afterwards by an abrupt increase at a decreasing rate.

The graphs of the two measures $P(\pi_t | \varepsilon_t)$ and $\mathbb{E}(\pi_t)$ for the first-order persistence are shown below.



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 reduces 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.

9 Conclusions and Future Work

It is important to understand the fundamental properties of ‘linear’ time series models with **variable** coefficients in order to efficiently handle these more complicated structures. 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’ 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-HARMA 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 reduces 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 particular, just to mention a few examples, it allows us to: i) tackle infinity and examine in depth infinite order autoregressions with either constant or variable coefficients, since it releases us from the need to work with characteristic polynomials, 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) and, in view of being easily applied to a multivariate setting (see, for example, Karanasos et al., 2014) to: iii) work out the fundamental time series properties of time varying linear VAR systems, 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 the methods of this paper to multivariate TV HARMA and GARCH models.

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

References

- [1] ABDRAHMO, 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] 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.
- [4] 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.
- [5] BAI, J. AND P. PERRON (2003): “Computation and Analysis of Multiple Structural Change Models,” *Journal of Applied Econometrics*, 18, 1-22.
- [6] BAILLIE, R. (1996): “Long Memory Processes and Fractional Integration in Econometrics,” *Journal of Econometrics*, 73, 5-59.
- [7] Chung, C-F. (1996): “Estimating a Generalized Long Memory Process,” *Journal of Econometrics*, 73, 237-259.

- [8] BENATI, L. (2008): “Investigating Inflation Persistence Across Monetary Regimes,” *The Quarterly Journal of Economics*, 123, 1005-1060.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] CONRAD, C. AND M. KARANASOS (2010): “Negative Volatility Spillovers in the Unrestricted ECCG-GARCH Model,” *Econometric Theory*, 26, 838-862.
- [14] CRAMER, H. (1961): “On Some Classes of Nonstationary Stochastic Processes,” *Proc. 4th Berkeley Symp. Math. Statist. Prob.* 2, 57-78.
- [15] DAHLHAUS, R. (1996): “Maximum Likelihood Estimation and Model Selection for Locally Stationary Processes,” *Journal of Nonparametric Statistics* 6, 171–191.
- [16] DAHLHAUS, R. (1997): “Fitting Time Series Models to Nonstationary Processes,” *The Annals of Statistics*, 25, 1-37.
- [17] ELAYDI, S. (2005): “An Introduction to Difference Equations,” Springer.
- [18] ELLIOTT, G., T. J. ROTHENBERG, AND J. H. STOCK (1996): “Efficient Tests for an Autoregressive Unit Root,” *Econometrica*, 75, 813-836.
- [19] EVANS, G. W. AND S. HONKAPOHJA (2001): *Learning and Expectations in Macroeconomics*. Princeton University Press.
- [20] EVANS, G. W. AND S. HONKAPOHJA (2009): “Learning and Macroeconomics,” *Annual Review of Economics*, 1, 421-449.
- [21] 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.
- [22] FRANCO, 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.
- [23] FRANCO, C. AND A. GAUTIER (2004b): “Estimation of Time-Varying ARMA models with Markovian Changes in Regime,” *Statistics & Probability Letters*, 70, 243-251.
- [24] FRANCO, 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.
- [25] FULKERSON, D. R. (1951): “Quasi-Hermite Forms of Row-finite Matrices,” Ph.D. Thesis, University of Wisconsin, available at <http://www.researchgate.net/publication/36218809>.
- [26] GIRAITIS, L., G. KAPETANIOS, AND T. YATES (2014): “Inference on Stochastic Time-Varying Coefficient Models,” *Journal of Econometrics*, 179, 46-65.

- [27] GLASSERMAN, P. AND D. D. YAO (1995): “Stochastic Vector Difference Equations with Stationary Coefficients,” *Journal of Applied Probability*, 32, 851-866.
- [28] 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.
- [29] 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.
- [30] GRAY, R. M. (2006): *Toeplitz and Circulant Matrices: A Review*. Now Publishers.
- [31] GRILLENZONI, C. (1990): “Modelling Time-Varying Dynamical Systems,” *Journal of the American Statistical Association*, 85, 499-507.
- [32] GRILLENZONI, C. (1993): “ARIMA Processes with ARIMA Parameters,” *Journal of Business & Economic Statistics*, 11, 235-250.
- [33] GRILLENZONI, C. (2000): “Time-Varying Parameters Prediction,” *Annals of the Institute of Statistical Mathematics*, 52, 108-122.
- [34] HAMILTON, J. D. (1989): “A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle,” *Econometrica*, 57, 357-384.
- [35] HAMILTON, J. D. (1994): “Time Series Analysis,” Princeton University Press.
- [36] HALLIN, M. ET MÉLARD (1977): “Indéterminabilité pure et inversibilité des processus autorégressifs-moyenne mobile à coefficients dépendant du temps.” *Cahiers Center Études Recherche Opér.* 19 385-392.
- [37] HALLIN, M. (1978): “Mixed Autoregressive-Moving Average Multivariate Processes with Time-Dependent Coefficients,” *Journal of Multivariate Analysis*, 8, 567-572.
- [38] HALLIN, M. (1984): “Spectral Factorization of Nonstationary Moving Average Processes,” *Annals of Statistics*, 12, 172-192.
- [39] 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.
- [40] HE, C. AND T. TERÄSVIRTA (2004): “An extended Constant Conditional Correlation GARCH Model and its Fourth-Moment Structure,” *Econometric Theory*, 20, 904-926.
- [41] 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.
- [42] KARANASOS, M. (1999): “The Second Moment and the Autocovariance Function of the Squared Errors of the GARCH Model,” *Journal of Econometrics*, 90, 63-76.
- [43] KARANASOS, M. (2001): “Prediction in ARMA Models with GARCH in Mean Effects,” *Journal of Time Series Analysis*, 22, 555-576.
- [44] KARANASOS, M., A. G. PARASKEVOPOULOS, AND S. DAFNOS (2014a): “The Fundamental Properties of Time Varying AR Models with Non Stochastic Coefficients,” arXiv.1403.3359.
- [45] KARANASOS, M., A. G. PARASKEVOPOULOS, AND S. DAFNOS (2014b): “A Univariate Time Varying Analysis of Periodic ARMA Processes,” arXiv.1403.4803.
- [46] 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.

- [47] **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,” available online:
- [48] **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.
- [49] **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.
- [50] **KARMEN, E. W.** (1988): “The Poles and Zeros of a Linear Time-Varying System,” *Linear Algebra and Applications*, 98, 263-289.
- [51] **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.
- [52] **KOOP, G. AND D. KOROBILIS** (2013): “Large Time-Varying Parameter VARs,” *Journal of Econometrics*, 177, 185-198.
- [53] **KOWALSKI, A. AND D. SZYNAL** (1990): “An Optimal Prediction in General ARMA Models,” *Journal of Multivariate Analysis*, 34, 14-36.
- [54] **KOWALSKI, A. AND D. SZYNAL** (1991): “On a Characterization of Optimal Predictors for Non-stationary ARMA Processes,” *Stochastic Processes and their Applications*, 37, 71-80.
- [55] **KOZICKI, S. AND P. TINSLEY** (2000): “Alternative Sources of the Lag Dynamics of Inflation,” FRB Kansas City Research Working Paper No. 02-12
- [56] **LAKSHMIKANTHAM, V. AND D. TRIGIANTE** (2002): “Linear Difference Equations: Numerical Methods and Applications,” Marcel Dekker.
- [57] **LEVIN, A. AND J. PIGGER** (2004): “Is inflation Persistence Intrinsic in Industrial Economics?” Federal Reserve Bank of St. Louis, Working Paper, No. 202-023A.
- [58] **LIM, A. AND J. DAI** (2011): “On Product of Companion Matrices,” *Linear Algebra and its Applications*, 435, 2921-2935.
- [59] **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.
- [60] **MALLIK, R. K.** (1998): “Solutions of Linear Difference Equations with Variable Coefficients,” *Journal of Mathematical Analysis and Applications*, 222, 79-91.
- [61] **MALLIK, R. K.** (2000): “On the Solution of a Linear Homogeneous Difference Equation with Varying Coefficients,” *SIAM Journal on Mathematical Analysis*, 31, 375-385.
- [62] **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.
- [63] **MARRERO, A. AND V. TOMEO** (2016): “On Compact Representations for the Solutions of Linear Difference Equations With Variable Coefficients,” *J. Comput. Appl. Math.*, 318, 411-421.
- [64] **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.
- [65] **MILLER, K. S.** (1968): “Linear Difference Equations,” Benjamin.
- [66] **MRAD, R. B. AND FARAG E.** (2002): “Identification of Armax Models With Time Dependent Coefficients,” *Journal of Dynamic Systems, Measurement, and Control*, 124, 464-467.

- [67] 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.
- [68] PARASKEVOPOULOS, A. G. (2012): “The Infinite Gauss-Jordan Elimination on Row-Finite $\omega \times \omega$ Matrices,” arXiv:1201.2950.
- [69] PARASKEVOPOULOS, A. G. (2014): “The Solution of Row-Finite Linear Systems with the Infinite Gauss-Jordan Elimination,” arXiv 1403.2624.
- [70] PARASKEVOPOULOS, A. G. AND M. KARANASOS (2019): “Explicit and compact representations for the one sided Green’s function and the Solution of Linear Difference Equations with variable coefficients,” arXiv .
- [71] 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.
- [72] PESARAN, M. H., A. PICK, AND M. PRANOVICH (2013): “Optimal Forecasts in the Presence of Structural Breaks,” *Journal of Econometrics*, 177, 134-152.
- [73] PESARAN, M. H. AND A. TIMMERMANN (2005): “Small Sample Properties of Forecasts from Autoregressive Models under Structural Breaks,” *Journal of Econometrics*, 129, 183-217.
- [74] PIVETTA, F. AND R. REIS (2007): “The Persistence of Inflation in the United States,” *Journal of Economic Dynamics and Control*, 31, 1326-1358.
- [75] 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.
- [76] 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.
- [77] 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.
- [78] STOCK, J. H. AND M. W. WATSON (2007): “Why has U.S. inflation Become Harder to Forecast?,” *Journal of Money, Credit and Banking*, 39, 3–33.
- [79] STOCK, J.H., and M. W. WATSON (2009): “Phillips Curve Inflation Forecasts,” In: Fuhrer, J., Kordzycki, Y., Little, J., Olivei, J. (Eds.), “Understanding Inflation and the Implications for Monetary Policy”, Cambridge, MIT Press.
- [80] TAYLOR, J. (2000): “Low Inflation, Pass Through, and the Pricing Power of Firms,” *European Economic Review*, 44, 1389-1408.
- [81] TERÄSVIRTA, T. (1994): “Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models,” *Journal of the American Statistical Association*, 89, 208-218.
- [82] 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
- [83] TIMMERMANN, A. AND H. K. VAN DIJK (2013): “Dynamic Econometric Modeling and Forecasting in the Presence of Instability“, *Journal of Econometrics*, 177, 131-133.
- [84] TOEPLITZ, O. (1909): “Über die Auflösung Unendlich Vieler Linearer Gleichungen mit Unendlich Vielen Unbekannten,” *Rend. Circ. Mat. Palermo.*, 28, 88-96.
- [85] VAN DER CRUYSSSEN, P. (1979): “Linear Difference Equations and Generalized Continued Fractions,” *Computing*, 22, 269-278.

- [86] VEIN, R. AND P. DALE (1999): “Determinants and their application in Mathematical Physics,” Springer.
- [87] 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 Subsection 2.2 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 (TV-LDEs(p)), thus TV-HARMA(p, q) models as well, can be represented as infinite linear systems whose coefficient matrix is row-finite¹¹ of dimension $\mathbb{N} \times \mathbb{N}$ consisting of the autoregressive coefficients:

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

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 a reduced form (defined by three postulates) of 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 been recently highlighted in Paraskevopoulos (2012), who has introduced a modified version of the Gauss-Jordan elimination algorithm in responding to this challenge. 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 there to infinite system representations of TV-LDEs of p order constructing a fundamental solution set. The fundamental solution sequences occupy the first p columns in the Fulkerson’s reduced matrix and the principal determinant, $\xi(t, s)$, represents the first of these solution sequences of the homogeneous system associated with (A.1), that is

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

This is easily verified by applying the above values of $y_{s+r,s}$ for $r = s - p + 1, \dots, s, s + 1, s + 2, \dots$ to eq. (A.1), assuming the right side values being zero, that is $v_{s+i} = 0$ for all $i \in \mathbb{Z}_{\geq 1}$. Moreover, fundamental solutions come out as expansions of banded Hessenbergians in terms of the time varying coefficients.

¹¹A row-finite matrix is an infinite matrix, each row of which comprises a finite number of non-zero entries.

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

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

Formally $\Phi_{t,s}^{(m)}$ is a banded Hessenberg matrix of order $k = t - s$. Each matrix in the sequence $\{\Phi_{t,s}^{(m)}\}_{1 \leq m \leq p}$, differs from any other matrix in this sequence only in the first column $[\phi_{t,s}^{(m)}]$. The principal matrix $\Phi_{t,s}$ (resp. principal determinant $\xi(t, s)$) is identified with $\Phi_{t,s}^{(1)}$ (resp. $\xi^{(1)}(t, s)$), that is $\Phi_{t,s}^{(1)} = \Phi_{t,s}$ (resp. $\xi^{(1)}(t, s) = \xi(t, s)$) (for notational convenience we will interchangeably use $\Phi_{t,s}^{(1)}$ (resp. $\xi^{(1)}(t, s)$) in place of $\Phi_{t,s}$ (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 of the next Lemma):

Lemma A1

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

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

that 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

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

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 $r = 1, 2, \dots, p$ 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

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

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

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

$$\Xi_{t,s} = \begin{bmatrix} \xi^{(1)}(t, s) & \xi^{(2)}(t, s) & \dots & \xi^{(p)}(t, s) \\ \xi^{(1)}(t-1, s) & \xi^{(2)}(t-1, s) & \dots & \xi^{(p)}(t-1, s) \\ \vdots & \vdots & \ddots & \vdots \\ \xi^{(1)}(t+1-p, s) & \xi^{(2)}(t+1-p, s) & \dots & \xi^{(p)}(t+1-p, s) \end{bmatrix}.$$

The Definition in (A.3) entails that the matrix $\Xi_{s,s}$ is the identity matrix of order p . Therefore the first Casoratian $|\Xi_{s,s}|$, of the set Ξ_s is $|\Xi_{s,s}| = 1 \neq 0$. It turns out that $|\Xi_{t,s}| \neq 0$ 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 stochastic 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 $\theta_0(t) = 1$ 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 side of the above equation into two parts, it takes the form:

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

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

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

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

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

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

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

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

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

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).

An expression of the homogeneous solution as a linear combination of the fundamental solutions is given below.

Proposition A3 *The solution of eq. (5) assuming the prescribed initial values $\{y_{s-m+1}\}_{m=1,2,\dots,p}$ 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 $y_{s-m+1} = c_m$ for $m = 1, 2, \dots, p$, it remains to show that $c_m = a_m$ for all $m : 1 \leq m \leq p$. Taking into account that for $1 \leq m \leq p$

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

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. ■

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

$$\begin{array}{c}
 y_{t,s}^* = \left| \begin{array}{cccccccc}
 v_{s+1} & & -1 & & & & & \\
 v_{s+2} & & \phi_1(s+2) & & \ddots & & & \\
 \vdots & & \vdots & & \ddots & & \ddots & \\
 v_{s+p+1-m} & & \phi_{p-m}(s+p+1-m) & & \ddots & & \ddots & \\
 \vdots & & \vdots & & \ddots & & \ddots & \\
 v_{s+p} & & \phi_{p-1}(s+p) & & \ddots & & \ddots & \\
 v_{s+p+1} & & \phi_p(s+p+1) & & \ddots & & \ddots & \\
 \vdots & & \vdots & & \ddots & & \ddots & \\
 v_{t-1} & & & \phi_p(t-1) & \phi_{p-1}(t-1) & \cdots & \phi_{p-m}(t-1) & \cdots & \phi_1(t-1) & -1 \\
 v_t & & & \phi_p(t) & \phi_{p-1}(t) & \cdots & \phi_{p+1-m}(t) & \cdots & \phi_2(t) & \phi_1(t)
 \end{array} \right|.
 \end{array} \tag{A.12}$$

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

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

$$\begin{array}{c}
 y_{t-m,s}^* = \left| \begin{array}{cccccccc}
 v_{s+1} & & -1 & & & & & \\
 v_{s+2} & & \phi_1(s+2) & & \ddots & & & \\
 \vdots & & \vdots & & \ddots & & \ddots & \\
 v_{s+p} & & \phi_{p-1}(s+p) & & \ddots & & \ddots & \\
 v_{s+p+1} & & \phi_p(s+p+1) & & \ddots & & \ddots & \\
 \vdots & & \vdots & & \ddots & & \ddots & \\
 \vdots & & \vdots & & \ddots & & \ddots & \\
 v_{t-m-1} & & & \phi_p(t-m-1) & \phi_{p-1}(t-m-1) & \cdots & \phi_1(t-m-1) & -1 \\
 v_{t-m} & & & \phi_p(t-m) & \phi_{p-1}(t-m) & \cdots & \phi_2(t-m) & \phi_1(t-m)
 \end{array} \right|
 \end{array}$$

(see Paraskevopoulos and Karanasos (2019) for an alternative 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 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 is identical to the expression (11). ■

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

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

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 to 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-HARMA 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:

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

The hypothesis $\sup_t |\phi_m(t)| < \infty$ for each m such that $1 \leq m \leq p$, 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 $\lim_{s \rightarrow -\infty} \xi(t,s) = 0$ for each $t \in \mathbb{Z}$, then*

$$\lim_{s \rightarrow -\infty} \xi^{(m)}(t,s) = 0 \quad \text{for all } t \in \mathbb{Z} \text{ and } m \in \mathbb{Z} : 1 \leq m \leq p, \quad (\text{B.1})$$

provided that $\sup_t |\phi_m(t)| < \infty$ for each m .

ii) If the forward stability condition holds, that is $\lim_{t \rightarrow \infty} \xi(t,s) = 0$ for each $s \in \mathbb{Z}$, then

$$\lim_{t \rightarrow \infty} \xi^{(m)}(t,s) = 0 \quad \text{for each } s \in \mathbb{Z} \text{ and } m \in \mathbb{Z} : 1 \leq m \leq p. \quad (\text{B.2})$$

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

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

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

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

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

ii) The assertion follows from:

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

This completes the proof of Lemma. ■

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

In the proof of the backward stability in Theorem 2, we shall use the notation of Subsection 4.1. Let us further call $\mathfrak{z}(s) = \mathbf{c}$ for all s such that $s \leq t$, where $\mathbf{c}' = [c_m]_{1 \leq m \leq p}$ 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-HARMA process associated with eq. (1) is backwards asymptotically stable if and only if $y_{t,s}^{hom} \rightarrow 0$, as $s \rightarrow -\infty$ 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. $s \rightarrow -\infty$, 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

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

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

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

Replacing $(s \rightarrow -\infty)$ by $(t \rightarrow \infty)$ in the above statements we deduce the forward asymptotic stability of the model. ■

C Second Order Properties

In this section we show that the first and the second unconditional moments exist, provided that the absolute summability condition holds. Under the above mentioned condition the Wold-Cramér decomposition of the DTV-HARMA(p, q) processes is derived along with the second order structure of these 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 Diagrams I and II commutative.

Proof of Proposition 5. First, we verify that the first unconditional moment in eq. (17) 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 $\tilde{\varphi}_t = \sup_{s \leq t} |\varphi(s)| \in \mathbb{R}_{\geq 0}$ for each t , we have:

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

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

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

Second, we show that the second unconditional moment in eq. (18) exists in $\mathbb{R}_{\geq 0}$, 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 $l = 1, \dots, q$ 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

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

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

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

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

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

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 (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:

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

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

Finally, in view of diagrams in eq. (19), it remains to show the following implications:

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

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 $\tilde{\varphi}_t \lim_{s \rightarrow -\infty} |\xi(t, s)| = 0$ it follows from the squeeze Theorem that $\lim_{s \rightarrow -\infty} |\xi(t, s)\varphi(s)| = 0$ for all $t \in \mathbb{Z}$.

As $\left(\lim_{s \rightarrow -\infty} |\xi(t, s)\varphi(s)| = 0 \iff \lim_{s \rightarrow -\infty} \xi(t, s)\varphi(s) = 0 \right)$, the first implication in diagram (C.7) follows.

Taking into account the implications

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

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 $\lim_{s \rightarrow -\infty} \xi_q^2(t, s)\sigma^2(s) = 0$. This shows the second implication in (C.7) and the proof is complete. ■

C.2 Wold-Cramér Decomposition

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

Proof of Theorem 3. The homogeneous solution subject to the 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 holds:

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

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

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

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

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

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 $s \rightarrow -\infty$ and recalling that $\xi(t, r) = 0$ for $r > t$ we have:

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

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

$$\begin{aligned} \lim_{s \rightarrow -\infty} y_{t,s} &= \lim_{s \rightarrow -\infty} y_{t,s}^{hom} + \lim_{s \rightarrow -\infty} y_{t,s}^{par} = \lim_{s \rightarrow -\infty} y_{t,s}^{par} \\ &= \sum_{r=-\infty}^t \xi(t,r)[\varphi(r) + u_r] = \sum_{r=-\infty}^t \xi(t,r)\varphi(r) + \sum_{r=-\infty}^t \xi_q(t,r)\varepsilon_r. \end{aligned} \quad (\text{C.10})$$

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 $y_t = \lim_{s \rightarrow -\infty} y_{t,s}$)

solves eq. (1). Applying the expression $y_{t-m} = \sum_{r=-\infty}^{t-m} \xi(t-m,r)v_r$ for $m = 0, 1, \dots, p$ to eq. (1), we show

below that its right 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. $\sum_{r=-\infty}^t \xi(t,r)v_r$:

$$\begin{aligned} \sum_{j=1}^p \phi_j(t) \sum_{r=-\infty}^{t-j} \xi(t-j,r)v_r &= \sum_{j=1}^p \sum_{r=-\infty}^{t-j} \phi_j(t)\xi(t-j,r)v_r \\ (\text{adding some zero terms}) &= \sum_{j=1}^p \sum_{r=-\infty}^{t-j} \phi_j(t)\xi(t-j,r)v_r + \sum_{j=1}^p \sum_{r=t-j+1}^t \phi_j(t)\xi(t-j,r)v_r \\ (\text{condensed sum}) &= \sum_{j=1}^p \sum_{r=-\infty}^t \phi_j(t)\xi(t-j,r)v_r \\ (\text{switching summation}) &= \sum_{r=-\infty}^t \sum_{j=1}^p \phi_j(t)\xi(t-j,r)v_r \\ (\text{factoring}) &= \sum_{r=-\infty}^t v_r \sum_{j=1}^p \phi_j(t)\xi(t-j,r) \\ (\text{applying eq. (A.5) for } m=1) &= \sum_{r=-\infty}^t v_r \xi(t,r). \end{aligned} \quad (\text{C.11})$$

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 the stochastic process $\{y_t\}_t$ associated with a DTV-HARMA(p, q), is given by

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

Next, we give a proof to Proposition 6.

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

$$\mathbb{E} \left(\sum_{r=t-\ell+1}^t \xi_q(t, r) \varepsilon_r \cdot \sum_{r=-\infty}^{t-\ell} \xi_q(t-\ell, r) \varepsilon_r \right) = 0. \quad (\text{C.12})$$

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

$$\begin{aligned} \gamma_t(\ell) &= \mathbb{E} \left(\sum_{r=-\infty}^t \xi_q(t, r) \varepsilon_r \cdot \sum_{r=-\infty}^{t-\ell} \xi_q(t-\ell, r) \varepsilon_r \right) \\ (\text{breaking first sum}) &= \mathbb{E} \left(\left(\sum_{r=-\infty}^{t-\ell} \xi_q(t, r) \varepsilon_r + \sum_{r=t-\ell+1}^t \xi_q(t, r) \varepsilon_r \right) \sum_{r=-\infty}^{t-\ell} \xi_q(t-\ell, r) \varepsilon_r \right) \\ (\text{associative law}) &= \mathbb{E} \left(\sum_{r=-\infty}^{t-\ell} \xi_q(t, r) \varepsilon_r \sum_{r=-\infty}^{t-\ell} \xi_q(t-\ell, r) \varepsilon_r + \sum_{r=t-\ell+1}^t \xi_q(t, r) \varepsilon_r \sum_{r=-\infty}^{t-\ell} \xi_q(t-\ell, r) \varepsilon_r \right) \\ (\text{by eq.(C.12)}) &= \mathbb{E} \left(\sum_{r=-\infty}^{t-\ell} \xi_q(t, r) \varepsilon_r \sum_{r=-\infty}^{t-\ell} \xi_q(t-\ell, r) \varepsilon_r \right) \\ (\text{by statement(i)}) &= \mathbb{E} \left(\sum_{r=-\infty}^{t-\ell} \xi_q(t, r) \xi_q(t-\ell, r) \varepsilon_r^2 \right) \\ (\text{by statements (ii) \& (iii)}) &= \sum_{r=-\infty}^{t-\ell} \xi_q(t, r) \xi_q(t-\ell, r) \sigma^2(r), \end{aligned}$$

as required. ■

C.4 Forward asymptotic efficiency

Before proving Lemma C3 and Proposition 7, we provide some useful notes on 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

$$\Omega \stackrel{\text{def}}{=} \limsup_{t \rightarrow \infty} x_t - \liminf_{t \rightarrow \infty} x_t \stackrel{\text{or}}{=} \lim_{t \rightarrow \infty} (\sup_{r \geq t} x_r - \inf_{r \geq t} x_r) \stackrel{\text{or}}{=} \inf_t (\sup_{r \geq t} x_r - \inf_{r \geq t} x_r) \quad (\text{C.13})$$

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

Lemma C3 *Let $F(t, s) = \sum_{r=s+1}^t |\xi(t, r)|$ 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, under the boundedness condition of $\{F(t, s)\}_t$, for every s either $\lim_{t \rightarrow \infty} \text{MSE}_{t,s}$ exists in $\mathbb{R}_{\geq 0}$ or $\{\text{MSE}_{t,s}\}_t$ oscillates with oscillation: $\Omega(s) = \inf_t (\sup_{r \geq t} \text{MSE}_{r,s} - \inf_{r \geq t} \text{MSE}_{r,s})$.*

Proof of Lemma C3. 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:

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

The first requirement follows from:

$$F_q(t, s) \leq \Theta \sum_{l=0}^q N_{s+l} \quad \text{for all } t. \quad (\text{C.14})$$

Let us call $S(t, s) = \sum_{r=s+1}^t \xi_q^2(t, r)$ for $t > s$. Employing the notation of Proposition 7, it follows from the well known identity

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

that

$$S(t, s) = \sum_{r=s+1}^t \xi_q^2(t, r) \leq \sum_{r=s+1}^t \xi_q^2(t, r) + 2 \sum_{i=s+1}^{t-1} \sum_{j=i+1}^t |\xi_q(t, i)\xi_q(t, j)| = \left(\sum_{r=s+1}^t |\xi_q(t, r)| \right)^2 = F_q^2(t, s) \quad (\text{C.15})$$

for all $t > s$. As $F_q(t, s) \geq 0$, inequality (C.14) implies: for every $s \in \mathbb{Z}$, $F_q^2(t, s) \leq \left(\Theta \sum_{l=0}^q N_{s+l} \right)^2 \in \mathbb{R}_{\geq 0}$

for all $t > s$. Thus $\{F_q^2(t, s)\}_t$ is bounded in t for each s . Let us call $S_s \stackrel{\text{def}}{=} \sup_t F_q^2(t, s)$. Now inequality (C.15) implies that for every s : $0 \leq S(t, s) \leq F_q^2(t, s) \leq S_s$ for all $t > s$. Thus $\{S(t, s)\}_t$ is bounded in t for each $s \in \mathbb{Z}$. As $0 < \sigma^2(r) \leq M$, we have:

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

Accordingly $\{\text{MSE}_{t,s}\}_t$ is bounded in t for each s and the result follows. ■

Proof of Proposition 7. Let us call $U = \sup_t F_t$. Then

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

Thus the condition of Lemma C3 is fulfilled and therefore $\{\text{MSE}_{t,s}\}_t$ is bounded in t for each s . The uniformly boundedness is derived as follows. If we replace N_s by U , in the proof of Lemma C3, then inequality (C.14) entails that: $F_q(t, s) \leq \Theta(q+1)U$. Also, it follows from inequality (C.15) that: $S(t, s) \leq \left(\Theta(q+1)U \right)^2$. Thus, $\text{MSE}_{t,s} \leq M \cdot \left(\Theta(q+1)U \right)^2$. As the latter bound is in $\mathbb{R}_{\geq 0}$ and independent of t, s the result follows. ■

D Stochastic Coefficients

In this Appendix we will examine a HAR model where the drift and the autoregressive coefficients are stochastically varying (STV). That is, y_t satisfies the following process:

$$y_t = \phi_{0t} + \sum_{m=1}^p \phi_{mt} y_{t-m} + \varepsilon_t, \quad (\text{D.1})$$

where $\{\varepsilon_t\}$ is a martingale difference defined on L_2 , with $\sup_t \sigma_\varepsilon^2(t) < \infty$.

The conditional expectations of y_t and ϕ_{mt} , for $m = 0, \dots, p$, with respect to the sigma field \mathcal{F}_s are denoted as

$$\bar{y}_{t,s} \stackrel{\text{def}}{=} \mathbb{E}(y_t | \mathcal{F}_s) \text{ and } \phi_{t,s}^{(m)} \stackrel{\text{def}}{=} \mathbb{E}(\phi_{mt} | \mathcal{F}_s). \quad (\text{D.2})$$

It is assumed that the conditional expectation of the product $\phi_{mt} y_{t-m}$ satisfies the following hypothesis:

Assumption 1 $\mathbb{E}(\phi_{mt} y_{t-m} | \mathcal{F}_s) = \phi_{t,s}^{(m)} \bar{y}_{t-m,s}$ for all m such that $1 \leq m \leq p$.

The STV-HAR process incorporates two classes of models:

Model Class 1. The GRC (see section 3.3) model, where ϕ_{mt} is independent of ε_τ and $\phi_{m\tau}$, $\tau \in \mathbb{Z}$, for all m and $t \neq \tau$. It integrates the following AR processes (for details, see Hwang and Basawa, 1998):

I. Random coefficient model: $\phi_{mt} = \phi_m + \eta_{mt}$, where ϕ_m are constant coefficients, $\{\eta_{mt}\}$ 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, t and τ . 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: $\phi_{mt} = \phi_m + \vartheta_m \varepsilon_t$, with ϕ_m as defined in I, and ϑ_m are constant coefficients.

III. Generalized Markovian bilinear model: $\phi_{mt} = \phi_m + \vartheta_m \varepsilon_t^{r_m}$, $r_m \in \mathbb{Z}_{>0}$, (ϕ_m and ϑ_m are as in I and II, respectively), 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 Markovian bilinear model.

IV. Random coefficients exponential model: $\phi_{mt} = \phi_m + (\vartheta_{m1} + \vartheta_{m2} e^{-\vartheta_{m3} \varepsilon_t^2}) \varepsilon_t$ (with ϕ_m as defined in I), where ϑ_{mi} , $i = 1, 2, 3$, are constant coefficients.

To sum up, $\phi_{t,s}^{(m)}$ (which in all Class 1 models is time invariant) is given by:

$$\begin{aligned} \text{Models I and II:} \quad \phi_{t,s}^{(m)} &= \phi_m, & (\text{D.3}) \\ \text{Model III:} \quad \phi_{t,s}^{(m)} &= \phi_m + \vartheta_m \mathbb{E}(\varepsilon_t^{r_m}), \\ \text{Model IV:} \quad \phi_{t,s}^{(m)} &= \phi_m + \vartheta_{m2} \mathbb{E}(e^{-\vartheta_{m3} \varepsilon_t^2} \varepsilon_t). \end{aligned}$$

Model Class 2. The double stochastic HAR model, hereafter termed DSHAR (for double stochastic processes, and in particular ARMA processes with ARMA coefficients, see Grillenzoni, 1993, and the references therein). In this case the random drift and autoregressive coefficients, ϕ_{mt} , for $m = 0, \dots, p$, follow AR processes:

$$\phi_{mt} = \varphi_{m0} + \sum_{l=1}^{p_m} \varphi_{ml} \phi_{m,t-l} + e_{mt}, \quad (\text{D.4})$$

where φ_{m0} and φ_{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 .

Proposition D2 Under Assumption 1, the k -step-ahead optimal (in L_2 -sense) linear predictor of the STV-HAR(p) process in eq. (D.1) satisfies the following ‘time varying’ difference equation¹²:

$$\bar{y}_{t,s} = \phi_{t,s}^{(0)} + \sum_{m=1}^p \phi_{t,s}^{(m)} \bar{y}_{t-m,s}, \quad (\text{D.6})$$

with solution given by

$$\bar{y}_{t,s} = \sum_{m=1}^p \xi^{(m)}(t,s) y_{s+1-m} + \sum_{r=s+1}^t \xi(t,r) \varphi_{r,s}^{(0)}. \quad (\text{D.7})$$

Proof of Proposition D2. Under Assumption 1, taking conditional expectations as of time s on both sides of eq. (D.1), we obtain eq. (D.6). Applying to eq. (D.6) the methodology in the proof of Theorem 1, eq. (D2) follows. ■

E Time Varying Polynomials

In this Appendix we will prove Theorems 5, 6 and Proposition 8. Let us recall eq. (8) in the main body of the paper:

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

Applying the above equation for $m = 1$, it gives:

$$\xi(t,s) = \sum_{r=1}^p \phi_r(s+r) \xi(t,s+r). \quad (\text{E.2})$$

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

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

Applying eq. (E.1) for specific values of $2 \leq m \leq p$, we get:

$$\begin{aligned} \xi^{(2)}(t,s) &= \sum_{r=1}^{p-1} \phi_{1+r}(s+r) \xi(t,s+r) \\ \xi^{(3)}(t,s) &= \sum_{r=1}^{p-2} \phi_{2+r}(s+r) \xi(t,s+r) \\ &\vdots \\ \xi^{(p-1)}(t,s) &= \phi_{p-1}(s+1)\xi(t,s+1) + \phi_p(s+2)\xi(t,s+2) \\ \xi^{(p)}(t,s) &= \phi_p(s+1)\xi(t,s+1). \end{aligned} \quad (\text{E.4})$$

Proof of Theorem 5. 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. (28) and (24), is equal to

$$\Xi_t^{(k)}(B) \circ \Phi_t(B) = \left(1 + \sum_{r=1}^{k-1} \xi(t,t-r) B^r \right) \circ \left(1 - \sum_{m=1}^p \phi_m(t) B^m \right). \quad (\text{E.5})$$

¹²Actually, for the GRC specification, since $\phi_{t,s}^{(m)}$, $m = 0, 1, \dots, p$, are constants, the difference equation is time invariant.

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.5) (separated into two the parts S_1, S_2) gives:

i) for powers from zero up to $k - 1$:

$$\begin{aligned}
S_1 &= 1 + [\xi(t, t - 1) - \phi_1(t)]B + [\xi(t, t - 2) - \phi_1(t - 1)\xi(t, t - 1) - \phi_2(t)]B^2 \\
&\quad + [\xi(t, t - 3) - \phi_1(t - 2)\xi(t, t - 2) - \phi_2(t - 1)\xi(t, t - 1) - \phi_3(t)]B^3 \\
&\quad \dots \\
&\quad + \left[\xi(t, t - k + 2) - \sum_{r=1}^p \phi_r(t - k + 2 + r)\xi(t, t - k + 2 + r) \right] B^{k-2} \\
&\quad + \left[\xi(t, t - k + 1) - \sum_{r=1}^p \phi_r(t - k + 1 + r)\xi(t, t - k + 1 + r) \right] B^{k-1}
\end{aligned} \tag{E.6}$$

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

$$\begin{aligned}
S_2 &= - \left[\sum_{r=1}^p \phi_r(s + r)\xi(t, s + r) \right] B^{k-1+1} - \left[\sum_{r=1}^{p-1} \phi_{1+r}(s + r)\xi(t, s + r) \right] B^{k-1+2} \\
&\quad - \left[\sum_{r=1}^{p-2} \phi_{2+r}(s + r)\xi(t, s + r) \right] B^{k-1+3} \\
&\quad \dots \\
&\quad - [\phi_{p-1}(s + 1)\xi(t, s + 1) + \phi_p(s + 2)\xi(t, s + 2)]B^{k-1+p-1} \\
&\quad - \phi_p(s + 1)\xi(t, s + 1)B^{k-1+p}.
\end{aligned} \tag{E.7}$$

The final step of the proof is to notice that

i) on account of eq. (E.3), the ℓ th ($\ell = 1, \dots, k - 1$) coefficient of the $(k - 1)$ th order polynomial in eq. (E.6) is equal to zero and $\xi(t, t - \ell) - \sum_{r=1}^p \phi_r(t - \ell + r)\xi(t, t - \ell + r) = 0$ thus $S_1 = 1$, and

ii) in view of eqs. (E.2) and (E.4), the time varying coefficient of B^{k-1+m} ($m = 1, \dots, p$) in the polynomial in eq. (E.7) is equal to minus the m th 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.5) 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 5 and 1 are equivalent, therefore, the former implies the latter and vice versa. The proof of Theorem 6 (under the absolute summability condition in Proposition 5) follows along the same lines as the proof of Theorem 5 except that this time we let $k \rightarrow \infty$, thus only the argument in part i) (see eq. E.6) applies.

Next we will proof Proposition 8. But first, we will rewrite (by replacing r with $t - r$) the two equations in Definition 1 as follows:

$$\xi_q(t, t - r) = \xi(t, t - r) + \sum_{l=1}^q \xi(t, t - r + l)\theta_l(t - r + l), \text{ for } r = 0, \dots, k - 1, \tag{E.8}$$

$$\xi_{s,q}(t, t - r) = \sum_{l=r-k+1}^q \xi(t, t - r + l)\theta_l(t - r + l), \text{ for } r = k - 1 + q, \dots, k. \tag{E.9}$$

Proof of Proposition 8. It suffices to show that

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

In view of eqs. (28) and (24), 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 ’), and collecting terms with the same powers of the backshift operator, the right-hand side of the above equation (separated into two parts) gives:

i) for powers from zero up to $k-1$:

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

On account of eq. (E.8) the latter summation is condensed to $\sum_{r=0}^{k-1} \xi_q(t, t-r) B^r$.

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

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

In view of eq. (12) the latter summation is condensed to $\sum_{r=k}^{k-1+q} \xi_{s,q}(t, t-r) B^r$. Adding the two parts/summations we obtain

$$\Xi_t^{(k)}(B) \circ \Theta_t(B) = \sum_{r=0}^{k-1} \xi_q(t, t-r) B^r + \sum_{r=k}^{k-1+q} \xi_{s,q}(t, t-r) B^r \stackrel{\text{eq. (29)}}{=} \Xi_{t,q}^{(k)}(B),$$

as claimed. ■