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

December 13, 2017

Abstract

The paper develops a widely applicable approach to examine the dynamics of the large family of 'time varying' models with ARMA representations. We provide the general solution, which gives rise to explicit or analytic formulas for all the main time series properties of these processes, such as the optimal forecasts and impulse response functions. The practical significance of the proposed methodology is illustrated with an application to U.S. inflation data. The main finding is that inflation persistence has been high since 1967, whereas in the post-crisis period the persistence reduces but it remains higher than the pre-1967 levels.

Keywords: General solution, models with ARMA representations, Monte Carlo simulations, optimal forecasts, structural breaks, time varying persistence.

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

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).

1 Introduction

Modelling time series processes with time dependent 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 models 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).[1]

Despite their widely recognized importance (see, for example, Granger, 2007 and 2008) and the fact that 'time varying' models are extensively applied by practitioners, there is a lack of a unified theory that supports the use of such a large class of processes. In this respect, a few exceptions are the pioneering works by Whittle (1965), Abdrabbo and Priestley (1967), Rao (1970), Hallin (1978, 1986), Kowalski and Szynal (1990,1991) and Grillenzoni (1993, 2000).[2] However, the stumbling block to the development of such a theory is the lack of a suitable explicit general solution for this large family of models. In particular, the problem arises because the convenient tool which is traditionally used to obtain a general solution, that is the characteristic polynomials (see, for details, Hallin, 1978, and Grillenzoni, 1990), is not applicable when time variation is present.

Against this background, in this paper a new technique for analyzing linear time series processes without the need to work with lag polynomials is presented. The proposed methodology can be applied to a large class of 'time varying' models that admit an ARMA representation (i.e., smooth transition AR specifications and Markov switching models), and ultimately enhance our understanding of their distinctive features.

In particular, we provide an autonomous method which yields explicit solutions expressed in terms of Hessenbergians, that is determinants of Hessenberg matrices. Our approach is rooted in the work of Toeplitz (1909) on the solution of infinite row-finite linear systems. The existence and representation of such solutions was established in Fulkerson (1951) by introducing a heuristic reduced form of row-finite matrices leading to systems that preserve the solutions of the original one. The lack of a constructive method transforming these 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 solving the problem. In a companion paper, Paraskevopoulos (2014) has further developed the above

[1] A growing empirical literature in macroeconomics is testimony to their importance. See, for example, Cooley and Prescott (1976) Evans and Honkapohja (2001, 2009), Cogley and Sargent (2005), Primiceri (2005), Stock and Watson (2012), and Koop and Korobilis (2012).

[2] Some other important steps in the development of theories for deterministically 'time varying' models include Francq and Gautier (2004a, 2004b), and Azrak and Mélard (2006).

noted algorithm yielding an analytic form to the general solution of row-finite linear systems. The algorithm directly applies to infinite system representations of ‘time varying’ linear difference equations and leads to Hessenbergian solutions for the large family of ARMA type of models with time dependent coefficients.

An important result is that the general solution makes it possible to derive explicit formulas for the fundamental time series properties of these complicated structures, such as optimal forecasts, the Wold-Crámer representation, impulse response functions and the unconditional first and second moments. Moreover, for the case of deterministically varying coefficients, we present necessary and sufficient conditions for the existence of the second moments, which are needed in order to obtain the asymptotic properties of the quasi maximum likelihood estimators (for related literature see, for example, Kowalski and Szynal, 1991, Grillenzoni, 2000, and Azrak and Mélard, 2006).

The proposed methodology enables us to reconcile a basic dichotomy in the research on these types of linear ‘time varying’ schemes. They are often categorized into models with either deterministically (i.e., periodic AR formulations, and processes with abrupt breaks) or stochastically varying coefficients (i.e., generalized random coefficients AR specifications). Although the two strands of the literature have considerably increased over time the overlap between them has not. Judging from the missing cross references, these strands are quite independent of each other. The unified theory that we propose in this paper provides the common ground between them. Trying to fill this gap, we hope that our research will be a decisive step towards their integration.

Another contribution of this work is to investigate to what extent commonly used unit root tests are robust to structural breaks in the times series process. It is well known that the performance of such tests depends on a number of factors that are not easily observed by applied economists trying to discriminate between stationarity and non stationarity. In addition, empirical research has often found evidence of GARCH effects with highly persistent volatility in situations where the conditional second moment affects the level of the series. However, the performance of the unit root tests for these types of stochastic processes has not been widely investigated. Thus, we consider a ‘time varying’ GARCH-in-mean specification and we carry out an extensive Monte Carlo experiment in order to examine the size and power of these tests in the presence of abrupt breaks in the in-mean parameter. The results indicate that the performance of the test statistics under consideration is severely affected by these breaks. The above considerations reinforce the argument (and extend it to a dynamic environment) made by Conrad and Karanasos (2015a) that conventional time invariant measures of persistence, such as unit roots, might result in misleading conclusions regarding the persistence in the level. Similarly, it is well known that unexpected shifts in a time series can lead to huge forecasting errors and unreliability of the model in general. Therefore, in a companion exercise we use simulated data to evaluate the reliability of the out-of-sample forecasts in the context of the GARCH-in-mean specification with abrupt breaks and find that the location and the magnitude of the breaks severely affects the forecasting performance of the models.

A well known example of stochastic processes where the persistence of the conditional variance is transmitted to the conditional mean are inflation time series (see for example Fountas and Karanasos, 2007). Accordingly, 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 model of inflation dynamics grounded on economic (instead of a statistical) theory. In particular, we estimate a GARCH-in-mean specification with variable coefficients and we compute an alternative measure of second-order time varying persistence, which not only distinguishes between changes in the dynamics of inflation and its volatility (and their persistence) but also allows for feedback from nominal uncertainty to inflation. Our results concur with the findings of Pivetta and Reis (2007), who by computing alternative statistical measures of persistence came to the conclusion that inflation persistence in the United States has been high since 1967. We also find that in the post-crisis period the persistence reduces but it remains higher than the pre-1967 levels.

Before concluding this section it may be useful to summarize a few more advantages of our technique. In a nutshell, it allows us to study stochastic linear difference equations (with variable coefficients) of ascending order and handle 'time varying' infinite order autoregressions, which include long memory processes as a special case (see the twin paper by Karanasos et al., 2017). Equally important, capitalizing on the connection between linear difference equations and the product of companion matrices, our general approach can also be employed to obtain an explicit formula for the latter (to save space these results are included in a companion paper by Paraskevopoulos and Karanasos, 2014). Interestingly, applying our technique to investigate ARMA processes with multiple abrupt breaks we show that working out these models by repeated substitution- a challenging problem that, until now, remained unsolved- is equivalent to calculating the determinant of a Block banded lower Hessenberg matrix (see, for example, the twin paper by Karanasos et al., 2017). In addition, because of its simplicity and generality our proposed methodology is applicable to a wide range of 'time varying' GARCH models (see, for example, Karanasos et al., 2014, and Karanasos et al., 2017).

The outline of the paper is as follows. Section 2 introduces the notation used in the paper and presents the general solution for the large family of 'time varying' models with ARMA representations. In Section 3, we derive explicit formulas for the main properties of these models. To show how our results can be easily extended to a VAR structure, in Section 4.1 we express a GARCH-in-mean specification with abrupt breaks as a bivariate system. We derive the optimal forecasts (in Section 4.3) and the second moments of this construction (see Section 4.5), which we utilize in order to obtain a new time varying measure of second-order persistence (in Section 4.6), thus extending the results in Conrad and Karanasos (2015a) to the case of variable coefficients. In Section 4.2 we consider the performance of commonly used unit root tests when the data generating process is a GARCH-in mean process with deterministic abrupt breaks. Section 5 presents an empirical study on inflation persistence. Finally, Section 6 contains some concluding remarks. Note that throughout the paper all the proofs are delegated to the Appendix.

2 Time Varying ARMA family of Models

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 model. Throughout the paper we adhere to the following conventions: ($\mathbb{Z}^+$) $\mathbb{Z}$, and $\mathbb{Z}^*$ stand for the sets of (positive) integers, and non-negative integers respectively. Similarly, ($\mathbb{R}^+$) $\mathbb{R}$ stands for the set of (positive) real numbers. Let the triple $(\Omega, \{\mathcal{F}_t, t \in \mathbb{Z}\}, P)$ denote a complete probability space endowed with a filtration, $\{\mathcal{F}_t\}$, which is a non-decreasing sequence of σ-fields $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t \subseteq \mathcal{F}_{t+1}$. The space of P-equivalence classes of finite random variables with finite p-order moment is indicated by L_p. Finally, $H = L_2(\Omega, \mathcal{F}_t, P)$ stands for a Hilbert space of random variables with finite first and second moments.

A time varying ARMA model of order $(p, q) \in \mathbb{Z}^+ \times \mathbb{Z}^*$, with time dependent coefficients and heteroscedastic errors (hereafter, termed TV-HARMA(p, q)) is defined as

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

with

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

where $\varphi(t)$ is the time varying drift and $\{\varepsilon_t\}$ is a martingale difference defined on L_2 with variance: $L < \sigma_t^2 < M$, for each t and some $L, M \in \mathbb{R}^+$. In the sequel we use $\phi_m(t) = 0$, $m > p$, and $\theta_l(t) = 0$, $l > q$. Notice that in our analysis this assumption will be relaxed when we consider the asceding order (see Appendix A).

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, Karanasos et al., 2014, and the twin paper by Karanasos et al., 2017).

The above process nests the TV-HAR(p) model as a special case when $q = 0$ (hereafter adopting the convention $\sum_{r=l}^{q}(.) = 0$, whenever $l > q$) 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: $\varphi(t) = \varphi,\ \phi_m(t) = \phi_m,\ \theta_l(t) = \theta_l,\ \sigma_t^2 = \sigma^2$ for all t.

To allow for statistical treatment of the TV-HARMA model (1) we follow Grillenzoni (2000) and assume that the parameters $\{\varphi(t),\ \phi_m(t),\ \theta_l(t)\}$ are smooth functions. The smoothness assumption is a necessary condition to have a good statistical performance.

The integer t indicates an arbitrary point of time at which a forecast y_t is intended and $k \in \mathbb{Z}^*$ the number of forecasting periods (or forecasting horizon) such that at time $t-k$ and back to time $t-k+1-p$ information is accessible. In particular, the integer $\tau = t - k$ indicates the 'present' time, that is the right end-point of the information time interval $\mathbb{I}_\tau = [\tau + 1 - p, \tau] \cap \mathbb{Z}$. The time interval $\mathbb{I}_\tau$ also coincides with the indexing set of the initial condition values $y_{t-k} = c_1, ..., y_{t-k+1-p} = c_p$. The forcing term of eq. (1) is assigned to be the time varying drift plus the moving average term: $\upsilon_t = \varphi(t) + u_t$.

A solution of eq. (1) is considered to be an explicit form of y_t expressed in terms of the initial conditions, the autoregressive coefficients $\phi_m(t)$ and the forcing term υ_t, such that all the other terms of the solution sequence, apart from the initial condition values, are generated by y_t for distinct values of t.

In this subsection we have defined the general TV-HARMA model. In the time invariant case we can express the simple ARMA(p, q) model in terms of lag polynomials as: $\Phi(B)y_t = \Theta(B)\varepsilon_t$, where B is the backshift or lag operator, $\Phi(B) = 1 - \sum_{m=1}^{p} \phi_m B^m$, $\Theta(B) = 1 + \sum_{l=1}^{q} \theta_l B^l$, and employ the roots of $\Phi(z^{-1})$ to obtain its general time series properties such as the second moment structure (see, for example, Karanasos, 1998). However, in a time varying environment the aforementioned procedure is not feasible since the convenient tool of characteristic polynomials disappears (see Hallin, 1978). To overcome this difficulty, we introduce the generating sequence, denoted below as $\{\xi_{t,r}\}_{r \geq 1-p}$. This allows us to derive the general solution and its stability, and the properties of the model without the need to work with characteristic polynomials.

2.2 The Generating Sequence

In this subsection, we present the mathematical tools used for analyzing TV-HARMA models. The starting point is to define the principal determinant associated with a TV-HARMA model (identified as $\xi_{t,k}$) being the determinant of a band and lower Hessenberg matrix built out of autoregressive coefficients $\phi_m(t-k+r)$ exclusively. This is the building element of the formulas of this paper including the generating sequence, the general solution and the analytic expressions for the first two unconditional moments. The generating sequence associated with TV-HARMA processes is the analytical tool for the study of the asymptotic stability and the second order properties of these models. It makes it possible to set up a generic condition ensuring both the asymptotic stability of the model together with the existence of its first two unconditional moments.

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

We start with the definition of the principal matrix:

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

(for $k \geq 1$), here and in what follows empty spaces in a matrix have to be replaced by zeros.

The matrix $\mathbf{\Phi}_{t,k}$ is a band $k \times k$ matrix with total bandwidth $p+1$ (the number of its nonzero diagonals, i.e., the diagonals whose elements are not all identically zero), upper bandwidth 1 (the number of nonzero superdiagonals) and lower bandwidth $p-1$ (the number of nonzero subdiagonals). That is, it is a band lower Hessenberg matrix.

Each non zero diagonal of the $\mathbf{\Phi}_{t,k}$ matrix, below the superdiagonal, consists of the time varying coefficients $\phi_r(\cdot)$, $r = 1, \ldots, \min(k, p)$, from $t-k+r$ to t. For example, the main diagonal contains the k values of the first time varying autoregressive coefficient, $\phi_1(\cdot)$, from time t to time $t-k+1$. We may also characterize $\mathbf{\Phi}_{t,k}$ as a 'time varying' Toeplitz matrix, because its time invariant version is a Toeplitz matrix of total bandwidth $p+1$ (see eq. (5) below; for details on Toeplitz matrices see, for example, the book by Gray, 2006 and Zemke, 2006).

For every pair $(t,k) \in \mathbb{Z} \times \mathbb{Z}^*$, we define the principal determinant associated with eq. (1) by

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

We further extend the definition of $\xi_{t,k}$ by assigning the values: $\xi_{t,0} = 1$ and $\xi_{t,-r} = 0$ for $r = 1, 2, ..., p-1$.

If $r \geq 0$, we can use the forecasting horizon k in place of r. The principal determinant $\xi_{t,r}$ for $r \geq 1$ in eq. (3) is an $r \times r$ band lower Hessenbergian. It gives rise to two sequences: First, for each arbitrary but fixed forecasting time point t, the generating sequence is defined by $\{\xi_{t,r}\}_{r \geq 1-p}$, i.e.,

$$\{\xi_{t,1-p}, \xi_{t,2-p}, ..., \xi_{t,0}, ..., \xi_{t,k}, \xi_{t,k+1}, \dots\}.$$

In this definition the variable r varies independently of t. Any term of the generating sequence at time t is obtained by assigning a certain value to r. For example, when $r = 0$ the outcome is $\xi_{t,0} = 1$, occurred at 'present' time $t = \tau$.

Second, the number of forecasting periods k coupled with the forecasting time t are kept fixed. We define the sequence $\{\xi_{t-k+r,r}\}_{r \geq 1-p}$ with independent variable r, that is:

$$\{\xi_{t-k+1-p,1-p}, \xi_{t-k+2-p,2-p}, ..., \xi_{t-k,0}, \dots, \xi_{t-1,k-1}, \xi_{t,k}, \xi_{t+1,k+1}, ...\}.$$

Unlike the generating sequence, the first variable $(t - k + r)$ of the above sequence varies in a linear fashion with r, being a dependent variable. Setting $r = k$ in $\{\xi_{t-k+r,r}\}_{r \geq 1-p}$, we obtain the term $\xi_{t,k}$ of the sequence, that is the y_t term of the sequence derived by solving the initial value problem

$$y_{\mathrm{t}} = \sum_{m=1}^{p} \phi_m(\mathrm{t}) y_{\mathrm{t}-m}, \tag{4}$$

subject to the initial conditions: $y_{t-k} = 1$ and $y_{t-k+1-m} = 0$, $m = 2, \dots, p$. As a consequence, $\{\xi_{t-k+r,r}\}_{r \geq 1-p}$ is one out of p solutions of eq. (4), forming a fundamental set of solutions (see the next subsection). We shall refer to this solution as the primary fundamental solution.

Having at our disposal the principal determinant $\xi_{t,k}$ all the terms for both types of sequences are fully determined. As opposed to ARMA models with constant coefficients, in the time varying case, for each distinct value of k (or r), while keeping t fixed, a new primary fundamental solution sequence is generated, which does not include, in general, the terms of a primary fundamental solution sequence obtained for lower values of k. This is due to the fact that the coefficients $\phi_m(t - k + r)$ vary along with k despite the fact that the initial condition values remain the same. It follows that any two terms of a generating sequence belong to different primary fundamental solution sequences (see Appendix A.4.2).

Before presenting the fundamental solution set it may be useful to discuss two illustrative examples

For the ARMA(p, q) model with constant coefficients, that is $\Phi_t(B) = \Phi(B) = 1 - \sum_{m=1}^{p} \phi_m B^m = \prod_{m=1}^{p}(1 - \lambda_m B)$, eq. (3) gives the determinant of a Toeplitz matrix of bandwith $p + 1$:

$$\xi_k = |\mathbf{\Phi}_k| = \begin{vmatrix} \phi_1 & -1 & & & & & & \\ \phi_2 & \phi_1 & -1 & & & & & \\ \vdots & \ddots & \ddots & \ddots & & & & \\ \vdots & \vdots & \ddots & \ddots & \ddots & & & \\ \phi_p & \phi_{p-1} & \cdots & \ddots & \ddots & \ddots & & \\ & \phi_p & \ddots & \cdots & \ddots & \ddots & -1 & \\ & & \ddots & \phi_{p-1} & \cdots & \phi_2 & \phi_1 & -1 \\ & & & \phi_p & \cdots & \cdots & \phi_2 & \phi_1 \end{vmatrix} = \sum_{m=1}^{p} \frac{\lambda_m^{k+p-1}}{\prod_{\substack{n=1 \\ n \neq m}}^{p} (\lambda_m - \lambda_n)}, \tag{5}$$

where the second equality follows (only if $\lambda_m \neq \lambda_n$) from standard results in ARMA models (see, for example, Karanasos, 2001, and in particular, eq. (2.6) in his Corollary 2). Note that the stability condition $\lim_{k \to \infty} \xi_k = 0$ (see Theorem 2(i) below) holds if and only if all the roots λ_m lie inside the unit circle.

As another example consider the TV-AR(1) model, where there is time variation only in the segment $(t - k_1,\, t - k_2)$, where $k > k_2 > k_1$. In particular,

$$\phi_1(s) = \begin{cases} \phi_1 & \text{if} \quad s > t - k_1 \\ \phi(s) & \text{if} \quad t - k_2 < s \leq t - k_1 \\ \phi_2 & \text{if} \quad s \leq t - k_2, \end{cases}$$

where $s \in \mathbb{Z}$. In this case, $\xi_{t,k} = \phi_1^{k_1} \prod_{\ell=k_1}^{k_2-1} \phi(t - \ell) \phi_2^{k-k_2}$, since for an AR(1) process all the elements in the subdiagonals of $\xi_{t,k}$ are zero. In this case, the stability condition $\lim_{k \to \infty} \xi_{t,k} = 0$ for all t holds if and only if $|\phi_2| < 1$. Similarly, the stability condition $\lim_{k \to \infty} \xi_{t+k,k} = 0$ for all t holds if and only if $|\phi_1| < 1$ (see again Theorem 2(i) below).

The current subsection defined the generating sequence and the primary fundamental solution. The next one presents the fundamental set of solutions.

2.3 Fundamental Set of Solutions

In what follows, we derive a linearly independent set of p sequences spanning the homogeneous solution space of TV-HARMA models. This set is known as the fundamental solution set (see also Appendix A). Accordingly, every homogeneous solution can be expressed uniquely as a linear combination of fundamental solutions (elements of the fundamental solution set) whose coefficients are the initial condition values. We will show that every fundamental solution can also be expressed in terms of the generating sequence, $\{\xi_{t,r}\}_r$, and, hence, the autoregressive coefficients.

The primary fundamental solution sequence together with the remaining $p-1$ fundamental solution sequences make up the fundamental solution set associated with eq. (1). The principal matrix $\mathbf{\Phi}_{t,k}$ is identified with $\mathbf{\Phi}_{t,k}^{(1)}$, that is $\mathbf{\Phi}_{t,k}^{(1)} = \mathbf{\Phi}_{t,k}$ (for notational convenience we will interchangeably use $\mathbf{\Phi}_{t,k}^{(1)}$ or $\mathbf{\Phi}_{t,k}$; from here onwards, we use superscripts within parentheses [e.g., $(.)^{(m)}$] to designate the index position of the corresponding term [e.g., m-th term] of a sequence, so as to distinguish position indices from power exponents). The matrix $\mathbf{\Phi}_{t,k}^{(m)}$, $m \geq 2$, is derived by replacing the first column of $\mathbf{\Phi}_{t,k}$ with the column vector: $[\phi_{t,k}^{(m)}]$, given by

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

(see also eq. (B.3) in Appendix B). Formally for every m and every $k > 1$ the matrix $\mathbf{\Phi}_{t,k}^{(m)}$ is $k \times k$ band lower Hessenberg.

Clearly, each matrix in the sequence $\{\mathbf{\Phi}_{t,k}^{(m)}\}_{1 \leq m \leq p}$, differs from any other matrix in the sequence only in the first column $[\phi_{t,k}^{(m)}]$.

For $k \leq p$, $\mathbf{\Phi}_{t,k}^{(m)}$ is a full lower Hessenberg matrix with last non zero row

$$(\phi_{m+k-1}(t), \phi_{k-1}(t), ..., \phi_1(t)).$$

In analogy with the definition of formula (3), we define

$$\xi_{t,k}^{(m)} = \det(\mathbf{\Phi}_{t,k}^{(m)}), \tag{6}$$

assigning $\xi_{t,1-m}^{(m)} = 1$ and $\xi_{t,-r}^{(m)} = 0$, for $r = 0, 1, ..., p-1$, and $r \neq 1-m$. Applying the above equation for $m = 1$, we infer the identification $\xi_{t,r}^{(1)} = \xi_{t,r}$. The expression (6) yields the analogue of the generating sequence, that is $\{\xi_{t,r}^{(m)}\}_{r \geq 1-p}$ for each t. Moreover, it generates the m-th fundamental solution sequence

$\{\xi^{(m)}_{t+r,r}\}_{r\geq 1-p}$, that is the solution sequence of the homogeneous equation (4) under the initial conditions $y_{t-k+1-m} = 1$ and $y_{t-k+1-s} = 0$ for $s = 1, \ldots, p$, $s \neq m$.

By expanding the determinant representation of $\xi^{(m)}_{t,k}$ (for $k \geq 1$) along the first column, we deduce an expression in terms of the autoregressive coefficients and the generating sequence:

$$\xi^{(m)}_{t,k} = \sum_{r=1}^{p-m+1} \phi_{m-1+r}(t-k+r)\xi_{t,k-r}. \tag{7}$$

Notice that $\xi^{(m)}_{t,1} = \phi_m(t)$ since $\xi_{t,0} = 1$ and $\xi_{t,-k} = 0$. Moreover, $\xi_{t,k-r} = \det(\mathbf{\Phi}_{t,k-r})$ for $r < k$, where $\mathbf{\Phi}_{t,k-r}$ equals the matrix $\mathbf{\Phi}_{t,k}$ without its first r rows and columns.

The homogeneous solution sequence of eq. (4) subject to the initial conditions $y_{t-k-p+m} = c_m$, $m = 1, 2, ..., p$ is denoted by $\{y^{hom}_{t-k+r,r}\}_{r\geq 1-p}$. As r increases along with $t-k+r$, the homogeneous solution sequence is directed forward to the future. Fixing t and letting r vary, $\{y^{hom}_{t,r}\}_{r\geq 1-p}$ matches the behavior of the generating sequence.

The set

$$\Xi_{t,k} = \{\{\xi^{(1)}_{t-k+,r}\}_{r\geq 1-p}, \{\xi^{(2)}_{t-k+,r}\}_{r\geq 1-p}, ..., \{\xi^{(p)}_{t-k+,r}\}_{r\geq 1-p}\},$$

is a fundamental solution set.

Therefore, the general homogeneous solution can be uniquely expressed as a linear combination of fundamental solutions as

$$y^{hom}_{t,k} = \sum_{m=1}^{p} c_m \xi^{(m)}_{t,k}, \tag{8}$$

for all t, k. A combination of eqs. (7) and (8) entails that $y^{hom}_{t,k}$ can be exclusively expressed in terms of autoregressive coefficients, the generating sequence, and the initial conditions.

In the following subsection we shall use the previously obtained results to provide an efficient form for the general solution of TV-HARMA processes. In particular, we will develop a framework for solving ascending order time varying linear difference equations (ATV-LDEs) and we will derive their two solution parts, that is the homogeneous and particular one (for details on TV-LDEs, see Miller, 1968).

2.4 The General Solution

The main contribution of this section is the development of a new technique that yields an applicable explicit formula for the general (or generating) solution of TV-HARMA processes. As pointed out by Hallin (1978) (see also Grillenzoni, 1990) the classical time series properties of ARMA models with constant coefficients follow from the properties of the underlying LDEs, the solution of which is based on the corresponding characteristic polynomials. However, this convenient tool disappears in the case where the coefficients are time varying. In what follows we present a methodology, which allows us easily to deal with the problem of time variation. [3]

The solution method, leading to Theorem 1 below, is initiated by a decomposition of eq. (1) into two parts, a weighted sum of recursively generated solution terms occupying the left hand side of the equation and a weighted sum comprising all the initial conditions plus the forcing term, occupying the right hand side of the equation. It enables us to represent eq. (1) as an infinite non-singular system of linear equations whose right hand side is made up exclusively of quantities in term of which the solution is intended to be expressed. On account of the non-singular nature of the system, Cramer's rule is applicable to a finite subsystem of the above mentioned system, obtaining the general term of the solution sequence as a Hessenbergian. By expanding the Hessenbergian solution along the first column we explicitly obtain the general solution in terms of only the generating sequence $\{\xi_{t,k}\}_k$ (for details see Appendix A and in particular eq. (A. 15)).

The formula in (A.15) is the main (or generating) solution expression of this paper. The combination of eq. (A.15) with the expression of $\xi_{t,k}^{(m)}$ in eq. (7) leads to an equivalent solution expression presented in the following theorem. The theorem provides an explicit form to the general solution of (1), which is the building block for the results in Section 3.

Theorem 1 *The general solution of eq. (1) subject to the initial conditions $y_{t-k+1-m}$, $m = 1, 2, ..., p$, can be expressed as*

$$y_{t,k} = \underbrace{\sum_{m=1}^{p} \xi_{t,k}^{(m)} y_{t-k+1-m}}_{Homogeneous\ Solution\ (y_{t,k}^{\text{hom}})} + \underbrace{\sum_{r=0}^{k-1} \xi_{t,r}[\varphi(t-r) + u_{t-r}]}_{Particular\ Solution\ (y_{t,k}^{par})}, \tag{9}$$

In the above theorem the general solution is described by the general term $y_{t,k}$ of the solution sequence, which is explicitly written as a sum of homogeneous and particular solutions.

The solution expressions are established (see Appendix A) in a wider framework concerning ATV-LDEs. This class of models operates as a bridge, which enables us to transfer results from TV-LDEs of finite order to the ones of infinite order. The solution expression when the order is infinite turns out to be the limit of a sequence of solutions of ATV-LDEs. This is an analytic solution expression that 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 (see, for example, Karanasos et al., 2014, and the twin paper by Karanasos

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

et al., 2017). Note that in the interest of brevity the detailed examination of the aforementioned models will be the subject of future companion papers.

In what follows, we apply formula (9) to recover two special instances of the TV-HARMA process. The first concerns the extreme case when $t = \tau$ (or, since $\tau = t - k$, $k = 0$). The second recovers the value $y_{\tau+1}$ explicitly obtained by formula (9), which was initially derived recursively from eq. (1).

1) Let $t = \tau$. Since $k = 0$, it follows from (9) that $y^{par}_{t,0} = 0$ since $\sum_{r=0}^{-1}(\cdot) = 0$. Moreover as $\xi_{t,0} = 1$ and $\xi^{(m)}_{t,0} = 0$ for $m \geq 2$, we infer that

$$y_{t,0} = y^{hom}_{t,0} = \sum_{m=1}^{p} \xi^{(m)}_{t,0} y_{t+1-m} = \xi_{t,0} y_t = y_t = y_\tau.$$

This means that $y_{t,0}$ coincides with the right end-point initial condition value ($c_1 = y_t$), as expected.

2) Let $t = \tau + 1$. Since $k = 1$, it follows from (9) that

$$y^{par}_{t,1} = \sum_{r=0}^{0} \xi_{t,r} \upsilon_{t-r} = \xi_{t,0}[\varphi(t) + u_t] = \varphi(t) + u_t.$$

Taking into account that $\xi^{(m)}_{t,1} = \phi_m(t)$ (see eqs. (6) or (7)), it follows from (9) and (1) that

$$y_{t,1} = y^{hom}_{t,1} + y^{par}_{t,1} = \sum_{m=1}^{p} \xi^{(m)}_{t,1} y_{t-m} + \upsilon_t = \sum_{m=1}^{p} \phi_m(t) y_{t-m} + \varphi(t) + u_t = y_t = y_{\tau+1}.$$

The main 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 unknown abrupt changes, smooth changes and mixtures of them. If the changes are smooth the coefficients can depend on an exogenous variable s_t or t or both. In the case of stochastically varying coefficients the model allows for Markov switching behavior (see, for example, Hamilton, 1989) or includes the generalized random coefficient (GRC) AR specification (see, for example, Glasserman and Yao, 1995, and Hwang and Basawa, 1998) as a special case.

In the next section we provide a necessary and sufficient condition for the stability of eq. (4).

2.5 Stability Condition

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

The stability conditions ensure that all the solutions approach a solution independent of the initial conditions (the initial conditions are gradually dying out), this being a standard requirement for a model to yield long-run predictions (see Section 3 below).

In particular, the TV-HARMA process in eq. (1) is asymptotically stable if $y_{t,k}^{hom} \to 0$, as $k \to \infty$ for an initial condition sequence of p arbitrary but fixed values $\{y_{t-k+1-m} = c_m\}_{1\leq m\leq p}$ for all t, k. A sufficient and necessary stability condition for eq. (1) is presented in the following theorem. In what follows we assume that $\sup_t |\phi_m(t)| < \infty$ (for the deterministic case) and $\sup_t \mathbb{E}(\phi_m^2(t)) < \infty$ (for the stochastic case) for all m.

Theorem 2 *i) Let the autoregressive coefficients $\phi_m(t)$ be deterministic. A necessary and sufficient condition for the TV-HARMA model in eq. (1) to be asymptotically stable is that $\lim_{k\to\infty} \xi_{t,k} = 0$ for all t. ii) Let the autoregressive coefficients $\phi_m(t)$ be stochastic. A necessary and sufficient condition for the TV-HARMA process to be asymptotically stable is that $\xi_{t,k} \xrightarrow{p} 0$, as $k \to \infty$ (probability convergence) for all t.*

Notice that the condition in Theorem 2(ii) includes 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]

Within our general framework we gave a necessary and sufficient condition which guarantees the stability of model (1). Since the condition in Theorem 2 is necessary not only for stability but for the existence of the moments as well (see Section 3) in a companion paper we provide an explicit formula for $\xi_{t,k}$ (see Paraskevopoulos and Karanasos, 2014).

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 $\sum_{r=0}^{\infty} \xi_{t,k}^2 < \infty$

[4] Azrak and Mélard (2006) have considered the asymptotic properties of QML 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).

to hold (see Proposition 4 below), which, therefore, are sufficient conditions for $\lim_{k\to\infty}\xi_{t,k}=0$ for all t. These are presented in the following proposition.

Proposition 1 *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* $\lim_{r\to\infty}\phi_m(t-r)=\phi_m$ *and* $\sum_{r=1}^{\infty}\varrho^{2r}<\infty$*, where* $\varrho=\varrho(\Phi)+\epsilon$*,* $\epsilon>0$*,* $\varrho(\Phi)=\max\{|z_m|\,,\ \Phi(z_m^{-1})=0\}$ *with* $\Phi(z^{-1})=1-\sum_{m=1}^{p}\phi_m z^{-m}$ *(see eq. 8 in Kowalski and Szynal, 1991).* [5]

ii) The deterministically varying polynomial $\Phi_t(z^{-1})$ *should have roots whose realizations entirely lie inside the unit circle, with the exception, at most, of a finite set of points (see Proposition 1 in Grillenzoni, 2000).*

With respect to Proposition 1(ii) trivially, in the case of periodic coefficients this is not a necessary condition, see Grillenzoni (1990) or Karanasos et al. (2014,a,b).

As an example, consider the logistic smooth transition AR(1) model, where the autoregressive coefficient is given by (we drop the subscript 1): $\phi(t)=\phi_1F(t;\gamma,s)+[1-F(t;\gamma,s)]\phi_2$ and $F(t;\gamma,s)=[1+e^{\gamma(t-s)}]^{-1}$, $\gamma\in\mathbb{R}^*$, $s\in\mathbb{Z}$, is the first-order logistic function. Clearly, if $t>s$, $F(\cdot)<0.5$ and regime 2 prevails, whereas if $t<s$ then $F(\cdot)<0.5$ and regime 1 prevails. Next recall that $\tau=t-k$ and let $\tau_+=t+k$. Then as $k\to\infty$, $F(\tau_+;\gamma,s)\to 0$ and $\phi(\tau_+)=\phi_2$, whereas $F(\tau;\gamma,s)\to 1$ and $\phi(\tau)=\phi_1$. Let also k_+ be the value of k for which $F(\tau_+;\gamma,s)=0$, and similarly k_- be the value of k for which $F(\tau;\gamma,s)=1$. For this model, if $k\geq k_-$ then $\xi_{t,k}=\prod_{\ell=0}^{k_--1}\phi(t-\ell)\phi_1^{k-k_-}$. Clearly, $\lim_{k\to\infty}\xi_{t,k}=0$ if and only if $|\phi_1|<1$. In addition, if $k\geq k_+$ then $\xi_{t+k,k}=\prod_{\ell=0}^{k_+-1}\phi(t+\ell)\phi_2^{k-k_+}$. Clearly, $\lim_{k\to\infty}\xi_{t+k,k}=0$ if and only if $|\phi_2|<1$.

We conclude this section with another example. Consider the periodic AR$(1;\ell)$ model where ℓ is the number of seasons (i.e., quarters) and let ϕ_s, $s=1,\ldots,\ell$ denote the periodically varying autoregressive coefficients. Then $\xi_{t_{T,s},k\ell}=[\prod_{s=1}^{\ell}(\phi_s)]^k$, where $t_{T,s}=T\ell+s$ and $T\in Z^*$ is the number of periods (i.e., years). Clearly $|\phi_s|<1$ for all s is a sufficient but not necessary condition for $\lim_{k\to\infty}\xi_{t_{T,s},k\ell}=0$. The necessary condition is $\left|\prod_{s=1}^{\ell}(\phi_s)\right|<1$.

[5] Kowalski and Snyzal (1991) showed that $\varrho(\Phi)$ is the spectral radius of the matrix $\mathbf{\Phi}=\begin{bmatrix}0&0&\cdots&0&-\phi_m\\1&0&\cdots&0&-\phi_{m-1}\\0&1&\cdots&0&-\phi_{m-2}\\\vdots&\vdots&\vdots&\vdots&\vdots\\0&0&\cdots&1&-\phi_1\end{bmatrix}$ (see page 75 in their paper).

3 Second Moments (Non Stochastic Coefficients)

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 predictors, the associated forecast errors and the mean square errors (MSE), ii) the first two unconditional moments, iii) the Wold-Crámer decomposition, and, therefore, impulse response functions (IRFs), and iv) their covariance structure. In this section we shall restrict ourselves to a treatment of the models with HAR structure and deterministic coefficients; we will term these models DTV-HAR. To save space we will discuss the case of the stochastic coefficients in a companion paper. The MA part complicates the notation but not the mathematics. However, the underlying concepts for the ARMA model are similar to the AR one. Therefore these details are delegated to Appendix C.

In what follows we present explicit formulas for the first conditional moment of the DTV-HAR(p) model.

3.1 Conditional Moments

Taking the conditional expectation of eq. (9) (when the coefficients are non stochastic and $u_t = \varepsilon_t$ or $\Theta_t(B) = 0$ for all t) with respect to the σ field $\mathcal{F}_{t-k}$ generated by $\{y_{t-k}, y_{t-k-1}, \ldots, y_{t-k+1-p}\}$ the following proposition holds.

Proposition 2 *The k-step-ahead optimal (in L_2-sense) linear predictor of the DTV-HAR(p) process is*

$$\mathbb{E}(y_t \,|\mathcal{F}_{t-k}) = \sum_{r=0}^{k-1} \xi_{t,r}\varphi(t-r) + \sum_{m=1}^{p} \xi_{t,k}^{(m)} y_{t-k+1-m}. \tag{10}$$

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

$$\mathbb{FE}(y_t \,|\mathcal{F}_{t-k}) = \sum_{r=0}^{k-1} \xi_{t,r}\varepsilon_{t-r}, \;\; \mathbb{V}ar[\mathbb{FE}(y_t \,|\mathcal{F}_{t-k})] = \sum_{r=0}^{k-1} \xi_{t,r}^2 \sigma_{t-r}^2.$$

The explicit form of the variance is necessary for the determination of the confidence intervals of $\mathbb{E}(y_t \,|\mathcal{F}_{t-k})$. Notice that the forecast error associated with eq. (10) is a finite sum of k errors from time $t-k+1$ to time t. Each error is multiplied by the corresponding term of the generating sequence $\{\xi_{t,r}\}_r$ (recall that $\xi_{t,r}$ is a band lower Hessenbergian). Similarly, the respective MSE is a finite sum of k terms.

Each term is the product of the variance, σ^2_{t-r}, multiplied by the corresponding squared value of $\xi_{t,r}$.

Singh and Peiris (1987), Kowalski and Szynal (1990, 1991), and Grillenzoni (1990, 2000), obtained the optimal forecasts using prediction algorithms and recursive computation. In sharp contrast, in the above proposition we provided an explicit formula for the optimal linear predictors. In the following remark we describe how we can use our widely applicable methodology to derive the optimal forecasts when the time varying coefficients are stochastic.

Remark 1 *In the case of stochastically varying coefficients we can proceed as follows:*
i) Take the conditional expectation of eq. (1) and express $\mathbb{E}(y_t\,|\mathcal{F}_{t-k})$ *as a time varying (with respect to* k*) LDE.*
ii) Solve the resulting difference equation using Theorem 1.

To save space the conditional (and unconditional moments) for the large family of ARMA processes with stochastically varying coefficients will be the subject of a future companion paper.

Having derived an explicit formula for the first conditional moment in the next section we turn our attention to the unconditional moments.

3.2 Unconditional Moments

In this section we present an explicit formula for the first unconditional moment and sufficient and necessary conditions for its existence. In what follows we assume that $\sup_t |\varphi(t)| < \infty$ for all m.

FIRST MOMENTS

Sufficient and necessary conditions ensuring that $\sum_{r=0}^{k-1} \xi_{t,r}\varphi(t-r)$ converge as $k \to \infty$ for all t, and thus that the unconditional means of the DTV HARMA family of processes exist are given below.

Proposition 3 *A sufficient condition for the DTV-HAR(p) model to be first-order is:*

$$\sum_{r=0}^{\infty} |\xi_{t,r}| < \infty, \text{ for all } t \text{ (absolute summability).}$$

A necessary condition is:

$$\lim_{r\to\infty} \xi_{t,r}\varphi(t-r) = 0 \text{ for all } t.$$

Under the absolute summability condition the unconditional mean of the process y_t *in eq. (1), that is*

$\mathbb{E}(y_t) = \lim_{k\to\infty} \mathbb{E}(y_t \,|\mathcal{F}_{t-k})$, *with non stochastic coefficients, is given by*

$$\mathbb{E}(y_t) = \sum_{r=0}^{\infty} \xi_{t,r}\varphi(t-r)$$

(clearly the mean is the same for both the AR and the ARMA processes). Notice that the absolute summability condition implies the stability condition.

The mean is an infinite sum of the time varying drifts multiplied by the corresponding term of the generating sequence $\{\xi_{t,r}\}_r$.

Another immediate consequence of Theorem 1 is the following result where we state expressions for the Wold-Crámer decomposition of the DTV-HAR(p) process. [6]

WOLD-CRÁMER DECOMPOSITION

Next we derive the one-sided MA representation of our model and sufficient and necessary conditions for its existence in the L_2 sense.

In particular, in the following proposition we provide sufficient and necessary conditions ensuring that $\sum_{r=0}^{\infty}\xi_{t,r}^2\sigma_{t-r}^2$ is bounded for all t and thus for $\sum_{r=0}^{k-1}\xi_{t,r}\varepsilon_{t-r}$ to converge in L_2 as $k\to\infty$ for all t (we recall that $L < \sigma_t^2 < M$ for all t).

Proposition 4 *A sufficient condition for* $\sum_{r=0}^{k-1}\xi_{t,r}^2\sigma_{t-r}^2$ *to converge as* $k\to\infty$ *for all* t *is*[7]*:*

$$\sum_{r=0}^{\infty}\xi_{t,r}^2 < \infty \text{ for all } t \text{ (Square Summability).}$$

A necessary condition is the stability condition in Theorem 2(i).

It is well known that absolute summability implies square summability. In our case $\sum_{r=0}^{\infty}|\xi_{t,r}| < \infty$ implies $\sum_{r=0}^{\infty}\xi_{t,r}^2 < \infty$. As a consequence, absolute summability is a sufficient condition for the DTV-HAR model to admit a second-order Wold-Crámer decomposition. Under the absolute summability condition we have $\lim_{k\to\infty} y_{t,k}^{hom} = 0$ (see Theorem 2). In view of eq. (9) we obtain: $y_t \overset{L_2}{=} \lim_{k\to\infty} y_{t,k} = \lim_{k\to\infty} y_{t,k}^{par}$. The following theorem is deduced directly from the above results.

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

[7] Kowalski and Szynal (1991) show that another sufficient condition for $\sum_{r=0}^{k-1}\xi_{t,r}\varepsilon_{t-r}$ to converge in L_2 as $k\to\infty$ is the AR-regularity of the DTV-HARMA process, that is $\sum_{r=1}^{\infty}\varrho^{2r}\sigma_{t-r}^2 < \infty$ for a sufficiently small t, where ϱ has been defined in Proposition 1. Kowalski and Szynal (1990) derived a sufficient condition for $\sum_{r=0}^{k-1}\xi_{t,r}\varepsilon_{t-r}$ to converge both a.s. and in L_2 as $k\to\infty$ for all t (see Lemma 5.1 in their paper).

Grillenzoni (2000) also showed that under the conditions of his Proposition 1 the DTV-ARMA process with zero mean is fourth order as well. The existence of fourth order moments is required for the central least squares (CLS) estimation.

Theorem 3 *Let the absolute summability condition in Proposition 3 hold. Then y_t is given by*

$$y_t = \sum_{r=0}^{\infty} \xi_{t,r}\varphi(t-r) + \sum_{r=0}^{\infty} \xi_{t,r}\varepsilon_{t-r},$$

which is the Wold-Crámer representation of the DTV-HAR(p) model.

In the above theorem y_t is 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 $\lim_{k\to\infty} \mathbb{FE}(y_t \,|\mathcal{F}_{t-k})$ is the zero mean random part.

Hallin (1978), Singh and Peiris (1987), Kowalski and Szynal (1991), Grillenzoni (2000), and Azrak and Mélard (2006) obtained the Wold-Crámer decomposition using recurrences. In sharp contrast, in Theorem 3 we have provided an explicit formula for the one-sided MA representation.[8]

Next we state an important corollary in relation to the Wold-Crámer decomposition.

Corollary 1 *Theorem 3 implicitly contains the invertibility of the time varying polynomial $\Phi_t(B)$ as follows: $[\Phi_t(B)]^{-1} = \Xi_t(B) = \sum_{r=0}^{\infty}\xi_{t,r}B^r$.*[9]

We should also mention that sufficient conditions for the process in eq. (1) to be invertible (i.e. $\{\varepsilon_t\}$ can be generated from $\{y_t\}$) are similar to those in Proposition 1 but with respect to the time varying MA polynomial $\Theta_t(z^{-1}) = 1 + \sum_{l=1}^{q}\theta_l(t)z^{-l}$. [10]

SECOND MOMENTS

Another immediate consequence of Theorem 1 is the following proposition, where we state expressions for the second moments of the DTV-HAR(p) process.

Proposition 5 *Let the absolute summability condition in Proposition 3 hold. Then the time varying ℓ order autocovariance function $\gamma_t(\ell) = \mathbb{C}ov(y_t, y_{t-\ell})$, $\ell \in \mathbb{Z}^*$, is given by*

$$\gamma_t(\ell) = \sum_{r=0}^{\infty} \xi_{t,\ell+r}\xi_{t-\ell,r}\sigma^2_{t-\ell-r} \tag{11}$$

The time varying variance of y_t, that is $\gamma_t(0) = \sum_{r=0}^{\infty} \xi^2_{t,r}\sigma^2_{t-r}$, is an infinite sum of the time varying variances of the errors. Each of these variances is multiplied by the corresponding squared ξ. Notice that the absolute summability condition implies absolute summable autocovariances: $\sum_{r=0}^{\infty} |\gamma_t(\ell)| < \infty$ for all t.

[8] Kowalski and Szynal (1991) extended the Wold decomposition to a class of stochastic processes without moment conditions.

[9] As pointed out by Grillenzoni (1990) the generating sequence $\{\xi_{t,r}\}_r$ cannot be obtained as in stationarity, by expanding in Taylor series the rational polynomial $\frac{1}{\Phi_t(B)}$. As an alternative Hallin (1978) introduced some results on difference operators such as adjoint operators and symbolic products of operators.

[10] Hallin (1986, see Theorem 4.2 ad 5.4) shows that whenever a non-stationary process admits a Granger-Andersen invertible model, then its Wold-Cramér decomposition is also Granger-Andersen invertible, an important result which reconciles causal 'explanation' and forecasting for the non-stationary case.

Although it may be difficult to explicitly evaluate the covariance structure of $\{y_t\}$, for numerical work, it can be calculated by computing the band lower Hessenbergians, $\xi_{t,r}$ in eq. (3) and substituting these in eq. (11).

We conclude this section with a remark.

Remark 2 *Azrak and Mélard (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. 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).*

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. Our proposed approach allows us to study stochastic linear difference equations of ascending order and handle 'time varying' models of infinite order. 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. 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.

To show how our results can be easily extended to a VAR structure, in the next section we examine a GARCH-in-mean model with abrupt breaks by expressing it as a bivariate system.

4 GARCH-in-mean Model

In this section, we consider an AR(1) process with GARCH-in mean (M) effects -that is, a model in which the conditional variance affects the conditional mean- and deterministic abrupt breaks (hereafter, DAB-AR$(1;2)$-M model). In particular, we will examine the case of two breaks ($N = 2$) which occur at times $t - k_1$ and $t - k_2$ (with $k_2 > k_1$, $k_2 \in \mathbb{Z}^+$; of course when $k_2 = k_1$ we have the case of one break), where the switch from one set of parameters to another is abrupt. The time invariant version of the model has been introduced by Engle et al. (1987) and applied by, for example, Glosten et al. (1993), Christensen and Nielsen (2007) and Conrad and Karanasos (2015a).

The DAB-AR(1; 2)-M model is given by

$$y_t = \varphi(t) + \phi(t)y_{t-1} + \varsigma(t)\sigma_t^\delta + \varepsilon_t, \tag{12}$$

where $\varepsilon_t = e_t\sigma_t$, and the vector of the three deterministically varying coefficients, $\mathbf{m}(\tau)' = (\varphi(\tau), \phi(\tau), \varsigma(\tau))$ is given by

$$\mathbf{m}(\tau)' = \begin{cases} (\varphi_1, \phi_1, \varsigma_1) & \text{if} \quad \tau > t - k_1 \\ (\varphi_2, \phi_2, \varsigma_2) & \text{if} \quad t - k_2 < \tau \leq t - k_1 \\ (\varphi_3, \phi_3, \varsigma_3) & \text{if} \quad \tau \leq t - k_2, \end{cases}$$

with φ_n, ϕ_n, $\varsigma_n \in \mathbb{R}$, $n = 1, 2, 3$, $\delta \in \mathbb{R}^+$, $\{e_t\}$ is a sequence of independent and indentically distributed (*i.i.d*) random variables with zero mean and unit variance, and σ_t^2 is the conditional variance of y_t. [11] The time dependent autoregressive coefficient $\phi(t)$ naturally measures the intrinsic persistence in the level of y_t. By including σ_t^δ in the conditional mean we allow for feedback from the power transformed conditional variance of y_t to its level, captured by the deterministically varying in-mean coefficient $\varsigma(t)$. We denote the size of the breaks by $\Delta\phi_n = \phi_n - \phi_{n-1}$ and $\Delta\varsigma_n = \varsigma_n - \varsigma_{n-1}$, for $n = 2, 3$. For example, $\phi_2 = \phi_3 - \Delta\phi_3$ and $\phi_1 = \phi_3 - \Delta\phi_3 - \Delta\phi_2$.

The power transformed conditional variance, σ_t^δ, is positive with probability one and is a measurable function of $\mathcal{F}_{t-1}$, which in turn is the sigma-algebra generated by $\{y_{t-1}, y_{t-2}, \ldots\}$. We assume that σ_t^δ is specified as a time invariant asymmetric power (AP) GARCH(1, 1) process:

$$(1 - \beta B)\sigma_t^\delta = \omega + \alpha f(\varepsilon_{t-1}), \tag{13}$$

with

$$f(\varepsilon_{t-1}) = (|\varepsilon_{t-1}| - \gamma\varepsilon_{t-1})^\delta,$$

where $|\gamma| < 1$ (for the APGARCH model with time invariant parameters see, for example, Ding et al., 1993, and Karanasos and Kim, 2006). The following conditions are necessary and sufficient for $\sigma_t^\delta > 0$, for all t: $\omega > 0$, $\alpha, \beta \geq 0$.

Next we will introduce some important notation.

Notation 1 *i) We denote the time invariant* r*th moment* ($r \in \mathbb{Z}^+$) *of the power transformed variance by*

[11] Within the class of ARMA processes this specification is quite general and allows for intercept and slope shifts (see also Pesaran and Timmermann, 2005, Pesaran et al., 2006, and Koop and Potter, 2007).

$\mu_r = \mathbb{E}(\sigma_t^{\delta r})$.

ii) Similarly, κ_r *denotes the* r*th moment of* $f(e_t)$*:* $\kappa_r = \mathbb{E}[[f(e_t)]^r]$.

Clearly for $\delta \geq 1$, $\mu_{2/\delta} = \mathbb{E}(\sigma_t^2)$ is not a fractional moment only if δ is equal to 1 or 2. In all other cases $\mu_{2/\delta}$ have to be calculated numerically. However, if $\delta > 2$, the existence of the first moment, μ_1 guarantees that of $\mu_{2/\delta}$. Similarly, $\mu_{1+1/\delta} = \mathbb{E}(\sigma_t^{\delta+1})$ is not a fractional moment only if $\delta = 1/\lambda$ where $\lambda \in \mathbb{Z}^+$. In all other cases $\mu_{1+1/\delta}$ have to be calculated numerically.

The APGARCH$(1,1)$ formulation in eq. (13) can readily be interpreted as having an ARMA$(1,1)$ representation for the conditional variance:

$$(1 - cB)\sigma_t^\delta = \omega + \alpha v_{t-1}, \tag{14}$$

where

$$c = \alpha\kappa_1 + \beta, \quad \text{and} \quad v_t = f(\varepsilon_t) - \mathbb{E}[f(\varepsilon_t)\,|\mathcal{F}_{t-1}] = f(\varepsilon_t) - \kappa_1\sigma_t^\delta,$$

and v_t is, by construction, an uncorrelated term with expected value 0. While the ε_t are the innovations to the level of y_t, the v_t can be considered the 'innovations' to the power transformed conditional variance of y_t. Note that the parameter c measures the *intrinsic* memory or persistence in the conditional variance.

Next we will define the covariance matrix of the two 'shocks' ε_t and v_t, $\mathbf{\Sigma} = \mathbf{E}(\varepsilon_t\varepsilon_t')$, where $\mathbf{E}(\cdot)$ denotes the elementwise expectation operator. First, we will denote the variances of the two 'shocks' and their covariance by

$$\sigma_\varepsilon = \mathbb{E}(\varepsilon_t^2), \sigma_v = \mathbb{E}(v_t^2), \sigma_{\varepsilon v} = \mathbb{E}(\varepsilon_t v_t).$$

The covariance matrix $\mathbf{\Sigma}$ is given by

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_\varepsilon & \sigma_{\varepsilon v} \\ \sigma_{\varepsilon v} & \sigma_v \end{bmatrix} = \begin{bmatrix} \mu_{2/\delta} & \mu_{1+1/\delta}\widetilde{\kappa} \\ \mu_{1+1/\delta}\widetilde{\kappa} & \mu_2\kappa \end{bmatrix}, \tag{15}$$

where

$$\kappa = (\kappa_2 - \kappa_1^2), \ \widetilde{\kappa} = \mathbb{E}[e_t f(e_t)].$$

In the following corollary we present expressions for κ_r and $\widetilde{\kappa}$ under the Assumption of Normality (see also Karanasos and Kim, 2006 and Conrad and Karanasos, 2015a).

Corollary 2 *Consider the case where the term e_t is standard normal. Then κ_r and $\widetilde{\kappa}$ are given by*

$$\kappa_r = \frac{1}{\sqrt{\pi}}\left[(1-\gamma)^{r\delta} + (1+\gamma)^{r\delta}\right] 2^{(\frac{r\delta}{2}-1)}\Gamma\left(\frac{r\delta+1}{2}\right),$$
$$\widetilde{\kappa} = \frac{1}{\sqrt{2\pi}}\left[[1-\gamma]^{\delta} - [1+\gamma]^{\delta}\right]2^{(\delta/2)}\Gamma\left(\frac{\delta}{2}+1\right),$$

where $\Gamma(\cdot)$ is the Gamma function.
When $\delta = 1$ the above expressions reduce to $\widetilde{\kappa} = -\gamma$, $\kappa_1 = \sqrt{\frac{2}{\pi}}$, $\kappa_2 = 1+\gamma^2$ and therefore $\kappa = (\kappa_2 - \kappa_1^2) = 1+\gamma^2 - \frac{2}{\pi}$, which implies that $\mathbf{\Sigma}$ becomes

$$\mathbf{\Sigma} = \mu_2 \begin{bmatrix} 1 & -\gamma \\ -\gamma & 1+\gamma^2-\frac{2}{\pi} \end{bmatrix}. \tag{16}$$

Having defined the deterministically varying extension of the AR-(APGARCH) M model, which was examined in Conrad and Karanasos (2015a), hereafter we will start to deviate from their analysis and methodology.

4.1 VAR Formulation

To obtain the optimal predictors and the variance of y_t for the DAB-AR-M model in eqs. (12) and (13) we could express it as a DAB-ARMA$(2,1;2)$ representation and then directly apply the results of Section 3. However, in what follows we will adopt an alternative methodology. More specifically, in the next lemma we will express eqs. (12) and (14) in a matrix form.

Lemma 1 *Eqs. (12) and (14) can be expressed in a matrix form as*

$$\mathbf{y}_\tau = \varphi(\tau) + \mathbf{\Phi}(\tau)\mathbf{y}_{\tau-1} + \mathbf{J}\varepsilon_\tau + \mathbf{Z}(\tau)\varepsilon_{\tau-1}, \tag{17}$$

with $\mathbf{y}_\tau = (y_\tau,\ \sigma_\tau^\delta)'$, $\varepsilon_\tau = (\varepsilon_\tau,\ v_\tau)'$, $\mathbf{J} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, where the three time varying coefficient matrices, $\varphi(\tau)$, $\mathbf{\Phi}(\tau)$, and $\mathbf{Z}(\tau)$ are time invariant in each of the three segments:

$$\varphi_n = \begin{bmatrix} \varphi_n + \varsigma_n\omega \\ \omega \end{bmatrix}, \mathbf{\Phi}_n = \begin{bmatrix} \phi_n & \varsigma_n c \\ 0 & c \end{bmatrix}, \mathbf{Z}_n = \begin{bmatrix} 0 & \varsigma_n\alpha \\ 0 & \alpha \end{bmatrix}, \begin{cases} n=1 & \text{if} \quad \tau > t-k_1, \\ n=2 & \text{if} \quad t-k_2 < \tau \le t-k_1, \\ n=3 & \text{if} \quad \tau \le t-k_2. \end{cases}$$

For notational convenience we will interchangeably use $\mathbf{\Phi}_3$ or $\mathbf{\Phi}$ and $\mathbf{Z}_3$ or $\mathbf{Z}$. We will term the deterministically varying bivariate expression in eq. (17) the DAB-BVARMA$(1, 1; 2)$ representation.[12]

In what follows we will employ the above representation to derive explicit formulas for the optimal predictors and the variance of y_t and σ_t^δ in eqs. (12) and (13), respectively.[13] These are needed in order to obtain time varying first and second-order measures of persistence. Such a measure for the time invariant case was developed in Conrad and Karanasos (2015a). But first, since the LAR has been commonly used as a measure of persistence in the context of testing for the presence of unit roots, we will use Monte Carlo simulations to examine the performance of unit root tests when the data are generated from an AR-(APGARCH) M process with unknown structural breaks in the in-mean coefficient.

4.2 Monte Carlo Experiment

A decision whether a series is treated as integrated of order zero, $I(0)$, or $I(1)$ has important implications for the subsequent modelling, hypothesis testing and forecasting. A frequent criticism of unit root tests concerns the poor power and size properties that many such tests exhibit (see Conrad and Karanasos, 2015a, and the references therein). Since standard unit root tests are based on the assumption that some type of heteroscedasticity is present but ignore the possibility that the volatility has a direct impact on the level, we investigate the size and power properties of common unit root tests in the presence of GARCH-M effects and unknown structural breaks in the in-mean parameter.

The two unit root tests considered are the Dickey-Fuller test (DF) proposed by Dickey and Fuller (1981) and the M test proposed by Perron and Ng (1996). As far as the estimation of the autoregressive parameter ϕ is concerned both ordinary least squared method (OLS) and the generalized least squared method (GLS) suggested by Elliott et al. (1996) are considered. This gives us two DF statistics, which we define as DF_{OLS} or DF_{GLS} depending on the estimation method used for ϕ. Likewise, the M tests are defined as M_{OLS} and M_{GLS} respectively.

To examine the properties of these tests we consider the DAB-AR$(1;2)$-M model (data generating process, DGP) in eqs. (12) and (13) for the Monte Carlo simulation experiment where

$$\varphi(t) = \phi(t) = 1 \text{ for all } t,\ \delta = 1,\ \omega = 1 - \alpha - \beta,\ \alpha = 0.1,\ \beta = 0.70,\ \gamma = 0, \tag{18}$$

and there are two abrupt breaks in the time varying in-mean coefficient, $\varsigma(t)$, at times $t - k_1$ and $t - k_2$.

[12] As pointed out by Conrad and Karanasos (2015a) the AR(1)-[APGARCH$(1, 1)$]-M model is observationally equivalent to an ARMA$(2, 1)$ process, with the largest autoregressive root (LAR) being close to one. Clearly, if $\phi = 0$, $c = 1$ and there are no breaks the reduced form representation of the AR-M specification coincides with the IMA$(1, 1)$ model proposed by Stock and Watson (2007).

[13] Notice that, as pointed ouy by Pivetta and Reis (2007), including other variables would lead to an assessment of predictability. Since here we focus on persistence, not predictability, we work with a univariate GARCH-M model.

In particular, $\varsigma(\tau) = \varsigma_1$ for $\tau < t - k_2$ and $\tau > t - k_1$, whereas $\varsigma(\tau) = \varsigma_2 = \varsigma_1 + \Delta_\varsigma$ for $t - k_2 \leq \tau \leq t - k_1$. The magnitude of the break is denoted by Δ_ς and the length of the break by $\Delta_k = k_2 - k_1$. Therefore, time variation is caused only by the in-mean coefficient. We also set the sample size k equal to $1,000$. Finally, $\{e_t\}$ are $i.i.d \sim N(0,1)$ random variables.

4.2.1 Empirical Sizes

The Monte Carlo simulation experiment design is targeted at investigating the effect of the in-mean breaks on the empirical sizes of the test statistics under consideration. However, as the magnitude of the in-mean parameter itself is likely to affect the performance of the test statistics (see Conrad and Karanasos, 2015a) we investigate this latter issue before considering the former. Accordingly, the Monte Carlo experiment is aimed at investigating the effects on the empirical sizes of *i*) the magnitude of the in-mean parameter, *ii*) the magnitude of the break, Δ_ς, and *iii*) the timing (k_1, k_2) and the length or duration (Δ_k) of the breaks as a fraction of the sample size, k. To address point *i)* a set of simulation experiments was undertaken with the *DGP* in eqs. (12), (13) and (18) with increasing magnitude of the in-mean parameter, namely $\varsigma_1 \in \{0.1, 0.3, 0.9\}$. Similarly, to investigate point *ii*) simulation experiments were undertaken with $\Delta_\varsigma \in \{0.07, 0.25, 0.50\}$ with the case of $\Delta_\varsigma = 0.00$ set as benchmark. Finally, to tackle point *iii*) in the experiment design we considered the above *DGP* with $k_1/k = (k - k_2)/k \in \{0.100, 0.333, 0.450\}$, that is $k_1 = (k - k_2) \in \{100, 333, 450\}$. In other words, we consider three values for the length of the in-mean break: $\Delta_k/k = (k_2 - k_1)/k \in \{0.80, 0.333, 0.10\}$ or $\Delta_k \in \{800, 333, 100\}$.

Note that all experiments were performed over $10,000$ Monte Carlo replications using, as mentioned earlier, a sample size $k = 1,000$, with a further 50 observations created and discarded in order to avoid the influence of the initial values. The sequence $\{e_t\}$ was generated using pseudo $i.i.d \sim N(0,1)$ random numbers from the RNDNS procedure in GAUSS with the value of y_0 set as a $N(0,1)$ random number.

Table 1 reports the results for the empirical sizes of the inference procedures under consideration for the 5% nominal significant level. The top panel reports the empirical sizes resulting from the simulation experiment with the aforementioned *DPG* with $\Delta_k = 800$, whereas the results for $\Delta_k = 333$ and $\Delta_k = 100$ are given in the middle and bottom panel, respectively.

Table 1. Empirical sizes of unit root tests: the case of two unknown breaks in the in-mean parameter.

		DF_{OLS}	DF_{GLS}	M_{OLS}	M_{GLS}
		$\Delta_k = 800$ or $\Delta_k/k = 0.80$			
$\varsigma_1 = 0.1$	$\Delta_\varsigma = 0.00$	0.049	0.054	0.054	0.054
	$\Delta_\varsigma = 0.07$ (ς_2=0.17)	0.048	0.042	0.046	0.040
	$\Delta_\varsigma = 0.25$	0.048	0.037	0.037	0.037
	$\Delta_\varsigma = 0.50$	0.017	0.011	0.012	0.011
$\varsigma_1 = 0.3$	$\Delta_\varsigma = 0.00$	0.049	0.040	0.042	0.040
	$\Delta_\varsigma = 0.07$	0.045	0.028	0.030	0.028
	$\Delta_\varsigma = 0.25$	0.029	0.014	0.014	0.013
	$\Delta_\varsigma = 0.50$	0.012	0.007	0.005	0.006
$\varsigma_1 = 0.9$	$\Delta_\varsigma = 0.00$	0.015	0.001	0.005	0.001
	$\Delta_\varsigma = 0.07$	0.013	0.001	0.001	0.001
	$\Delta_\varsigma = 0.25$	0.016	0.000	0.000	0.000
	$\Delta_\varsigma = 0.50$	0.008	0.000	0.000	0.000
		$\Delta_k = 333$ or $\Delta_k/k = 0.333$			
$\varsigma_1 = 0.1$	$\Delta_\varsigma = 0.07$	0.047	0.045	0.046	0.045
	$\Delta_\varsigma = 0.25$	0.049	0.042	0.043	0.041
	$\Delta_\varsigma = 0.50$	0.020	0.027	0.022	0.025
$\varsigma_1 = 0.3$	$\Delta_\varsigma = 0.07$	0.042	0.030	0.032	0.030
	$\Delta_\varsigma = 0.25$	0.037	0.025	0.024	0.025
	$\Delta_\varsigma = 0.50$	0.013	0.010	0.009	0.010
$\varsigma_1 = 0.9$	$\Delta_\varsigma = 0.07$	0.013	0.001	0.002	0.001
	$\Delta_\varsigma = 0.25$	0.009	0.000	0.000	0.000
	$\Delta_\varsigma = 0.50$	0.003	0.000	0.000	0.000
		$\Delta_k = 100$ or $\Delta_k/k = 0.10$			
$\varsigma_1 = 0.1$	$\Delta_\varsigma = 0.07$	0.049	0.053	0.043	0.053
	$\Delta_\varsigma = 0.25$	0.052	0.038	0.044	0.037
	$\Delta_\varsigma = 0.50$	0.037	0.045	0.046	0.045
$\varsigma_1 = 0.3$	$\Delta_\varsigma = 0.07$	0.048	0.038	0.031	0.038
	$\Delta_\varsigma = 0.25$	0.050	0.036	0.031	0.035
	$\Delta_\varsigma = 0.50$	0.022	0.023	0.019	0.023
$\varsigma_1 = 0.9$	$\Delta_\varsigma = 0.07$	0.013	0.001	0.002	0.001
	$\Delta_\varsigma = 0.25$	0.013	0.000	0.000	0.000
	$\Delta_\varsigma = 0.50$	0.008	0.001	0.001	0.001

Note: The *DGP* is $y_t = 1 + y_{t-1} + \varsigma(t)\sigma_t + e_t\sigma_t$ and $\sigma_t = 0.2 + 0.1\,|e_{t-1}\sigma_{t-1}| + 0.7\sigma_{t-1}$, where $\varsigma(\tau) = \varsigma_1$ if $\tau > t - k_1$ or $\tau < t - k_2$, and $\varsigma(\tau) = \varsigma_2 = \varsigma_1 + \Delta_\varsigma$ otherwise with $\varsigma_1 \in \{0.1, 0.3, 0.9\}$, $\Delta_\varsigma \in \{0.07, 0.25, 0.50\}$, $k = 1{,}000$, $k_1 = (k - k_2) \in \{100, 333, 450\}$ or $\Delta_k \in \{800, 333, 100\}$.

Looking at the results in Table 1 we first notice that all inference procedures appear to be robust to small values of the in-mean parameter ($\varsigma_1 = 0.1$) and of the breaks ($\Delta_\varsigma = 0.07$). However, the magnitude of the in-mean parameter appears to have a significant effect on the size distortion of all test statistics as, even when $\Delta_{\varsigma} = 0.00$, for $\varsigma_1 = 0.9$ all the test statistics are severely undersized. Similarly, both the magnitude and the location of the breaks affect the size properties of the inference procedures under consideration as from the top panel of Table 1 it is clear that the worst case scenario appears to be when $\Delta_k = 800$ and $\Delta_{\varsigma} \geq 0.25$. In this case the break occurs very early and the stochastic process stays in the second regime for 80% of the time period, only to go back to the first regime for the last 100 observations.[14]

Looking now at the performance of the individual tests, it appears that the *OLS* based test are more robust to regime shifts in the in-mean parameter than the *GLS* based tests, as both DF_{OLS} and M_{OLS} enjoy smaller size distortion than their *GLS* based counterparts.

4.2.2 Empirical Power

The empirical sizes of the unit root tests presented in Table 1 are constructed to generate a test with asymptotic size of 5% under the null hypothesis of a unit root. We now focus on examining the power of the inference procedures to reject the null hypothesis of $\phi(t) = 1$ for all t when in fact the process is second-order, that is $\phi(t) = \phi$ with $|\phi| < 1$ for all t.

As for the size, the Monte Carlo experiment design is meant to investigate the effects for points *i*) - *iii*) above. With this target in mind, the asymptotic local power functions for the 5% nominal level test have been calculated. To model the sequence of stationary alternatives near the null hypothesis of unit root, we consider the aforementioned *DGP* but now with $\phi(t) = 1 - \frac{l}{k}$ for all t (instead of $\phi(t) = 1$) in eq. (12) where $l = 30, 29, ..., 1, 0$ controlling the size of the departure from a unit root.

To investigate the issue in point *i*) simulation experiments were undertaken setting different values of the in-mean parameter under the alternative hypothesis. The simulation results are summarized in Figure 1, where the asymptotic local power curves are plotted for the *DGP* when the magnitude of the parameter is increased from the modest value of $\varsigma_1 = 0.1$ to a relatively large value $\varsigma_1 = 0.9$, with the break parameter fixed at $\Delta_{\varsigma} = 0$. In the x-axis the value taken by l is reported, whereas in the y-axis the empirical rejection frequencies are reported. Looking at the plot of the asymptotic power curves for the tests under consideration from Figure 1 it appears that all test statistics are sensitive to the magnitude of the in-mean parameter. However, it is clear that DF_{OLS} and M_{OLS} are less sensitive to the magnitude of ς than the *GLS* based counterparts.

Coming to target point *ii*), in Figure 2 we report the results of simulation experiments obtained by

[14] We also find (results not reported) that in the presence of asymmetries the size distortion of the unit root tests is stronger, reinforcing the argument made by Conrad and Karanasos (2015a).

fixing the in-mean parameter at 0.9 and $\Delta_k = 800$, then comparing the resulting power curves of the test statistics when $\Delta_\varsigma = 0$ and $\Delta_\varsigma = 0.5$. Interestingly enough, the DF_{OLS} procedure appears to be the most robust to the regime shift of the in-mean parameter. By contrast both GLS based statistics are severely affected by the magnitude of the break.

Finally, we consider the issue of the timing and duration of the in-mean regime shift as stated in target point *iii*). In this case the simulation experiment was undertaken with $\Delta_k \in \{800, 100\}$ and Δ_ς fixed at the smallest value 0.07. Figure 3 plots the asymptotic local power function for DF_{OLS}, DF_{GLS}, M_{OLS} and M_{GLS} respectively. From the results in Figure 3 it appears that the empirical power of all inference procedure is less affected by the timing and the duration of the regime shift than the size reported in Table 1. Note that in the interest of brevity not all the values of the parameter space considered in Table 1 have been reported, but results are available upon request.

Having investigated the size and power properties of unit root tests in the presence of GARCH-M effects and unknown structural breaks in the in-mean parameter, next we will derive an explicit formula for the general solution of the DAB-BVARMA$(1, 1; 2)$ representation.

4.3 VAR General Solution

In this section we provide the generating solution of the DAB-BVARMA$(1, 1; 2)$ representation, which generates explicit formulas for the optimal predictors and the bidimensional time varying covariance matrix of $\{\mathbf{y}_\tau\}$, $\tau = t + r$, $r \in \mathbb{Z}^*$.

First, let $\lambda_{\max}(\mathbf{X})$ denote the modulus of the largest eigenvalue of $\mathbf{X}$. The following theorem holds.

Theorem 4 *The general solution of the bivariate system in eq. (17), subject to the initial condition* $\mathbf{y}_{\tau-k}$*, for* $k \geq k_2 + r$*, is given by*

$$\mathbf{y}_{\tau,k} = \mathbf{E}(\mathbf{y}_\tau \,|\mathcal{F}_{\tau-k}\,) + \mathbf{FE}(\mathbf{y}_\tau \,|\mathcal{F}_{\tau-k}\,), \tag{19}$$

where

$$\begin{aligned}
\mathbf{E}(\mathbf{y}_\tau \,|\mathcal{F}_{\tau-k}\,) &= \varphi_k(\tau) + \mathbf{\Phi}_1^{k_1+r}\mathbf{\Phi}_2^{k_2-k_1}\mathbf{\Phi}^{k-k_2-1}(\mathbf{\Phi}\mathbf{y}_{\tau-k} + \mathbf{Z}\varepsilon_{\tau-k}),\\
\mathbf{FE}(\mathbf{y}_\tau \,|\mathcal{F}_{\tau-k}\,) &= \mathbf{J}\varepsilon_\tau + \sum_{\ell=1}^{k_1+r}\mathbf{\Phi}_1^{\ell-1}(\mathbf{\Phi}_1\mathbf{J} + \mathbf{Z}_1)\varepsilon_{\tau-\ell} + \mathbf{\Phi}_1^{k_1+r}\{\sum_{\ell=1}^{k_2-k_1}\mathbf{\Phi}_2^{\ell-1}(\mathbf{\Phi}_2\mathbf{J} + \mathbf{Z}_2)\varepsilon_{t-k_1-\ell}\\
&\quad + \mathbf{\Phi}_2^{k_2-k_1}[\sum_{\ell=1}^{k-k_2-1}\mathbf{\Phi}^{\ell-1}(\mathbf{\Phi}\mathbf{J} + \mathbf{Z})\varepsilon_{t-k_2-\ell}]\},
\end{aligned}$$

and if $\lambda_{\max}(\mathbf{\Phi}_n) \neq 1$, $n = 1, 2, 3$, *then*

$$\varphi_k(\tau) = (\mathbf{I} - \mathbf{\Phi}_1^{k_1+r})(\mathbf{I} - \mathbf{\Phi}_1)^{-1}\varphi_1 + \mathbf{\Phi}_1^{k_1+r}[(\mathbf{I} - \mathbf{\Phi}_2^{k_2-k_1})(\mathbf{I} - \mathbf{\Phi}_2)^{-1}\varphi_2 + \mathbf{\Phi}_2^{k_2-k_1}(\mathbf{I} - \mathbf{\Phi}^{(k-k_2)})(\mathbf{I} - \mathbf{\Phi})^{-1}\varphi].$$

In the above expression if $\lambda_{\max}(\mathbf{\Phi}_n) = 1$, then $(\mathbf{I} - \mathbf{\Phi}_n^{k_n - k_{n-1}})(\mathbf{I} - \mathbf{\Phi}_n)^{-1}$, with $k_0 = -r$ and $k_3 = k$, should be replaced by $\sum_{\ell=0}^{k_n-k_{n-1}-1} \mathbf{\Phi}_n^{\ell}$ (a similar argument holds for any of the analogous cases that follow).

The above theorem expresses the general solution, $\mathbf{y}_{\tau,k}$, in terms of the $(k + r)$-step ahead optimal in (L_2 sense) linear predictor, $\mathbf{E}(\mathbf{y}_\tau \,|\mathcal{F}_{\tau-k})$, and the associated forecast error, $\mathbf{FE}(\mathbf{y}_\tau \,|\mathcal{F}_{\tau-k})$. Clearly, if $k_2 = k_1$ eq. (19) gives the solution in the case of one break, whereas if $k_2 = k_1 = k$, it gives the general solution when there is no time variation. For example, for the time invariant case, since $\mathbf{\Phi}_1 = \mathbf{\Phi}_2 = \mathbf{\Phi}$ and $\mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{Z}$, the forecast error in eq. (19) reduces to

$$\mathbf{FE}(\mathbf{y}_\tau \,|\mathcal{F}_{\tau-k}) = \mathbf{J}\varepsilon_\tau + \sum_{\ell=1}^{k+r} \mathbf{\Phi}^{\ell-1}(\mathbf{\Phi}\mathbf{J} + \mathbf{Z})\varepsilon_{\tau-\ell}. \tag{20}$$

The general solutions when $k \leq k_1 + r$ and $k_1 + r < k < k_2 + r$ can be obtained along the lines of Theorem 4 and are equivalent to the time invariant case and the case when there is one break, respectively. In this section, in the context of the DAB-AR-M model, we show the importance of taking into account abrupt breaks for the in-sample forecasting. In the next section we use Monte Carlo simulations to examine the out-of-sample performance of the model under consideration.

4.4 Forecasting

In this section we investigate the out-of-sample forecasting performance of the model in eqs. (12)-(13). The *DGP* was generated by Monte Carlo simulation as explained in Section 4.2 with $\phi(t) = \phi = 0.8$, $\varsigma_1 = 0.3$ and the other parameters as specified in eq. (18). [15]

In order to investigate the effects of the time varying in-mean parameter the model with $\Delta_\varsigma = 0.00$ was considered as a benchmark and then the magnitude of the break increased as in Table 1. Similarly, the duration of the regime shift was decreased from $\Delta_k = 800$ to $\Delta_k = 100$.

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 MSE, the mean absolute error (MAE) and the root mean square forecast error (RMSE). Table 2 reports the results of the forecasting exercise. In columns 1 and 2 the forecasting horizon and the break magnitude under consideration are

[15] See Elliott and Timmermann (2008) for an excellent review on economic forecasting.

reported, respectively, whereas in columns 3-8 the forecasting results for the conditional mean and the conditional variance are reported.

Table 2. Forecasting with a DAB-AR(1; 2)-M model. Point predictive performances.

Forecast Horizon	Break size	Conditional Mean			Conditional Variance		
		MSE	*MAE*	*RMSE*	*MSE*	*MAE*	*RMSE*
		$\Delta_k = 800$					
1	$\Delta_\varsigma = 0.00$	0.002	0.040	0.040	0.016	0.127	0.126
	$\Delta_\varsigma = 0.07$	0.002	0.045	0.044	0.018	0.134	0.139
	$\Delta_\varsigma = 0.25$	0.002	0.048	0.048	0.018	0.137	0.137
	$\Delta_\varsigma = 0.50$	0.002	0.050	0.050	0.019	0.137	0.137
5	$\Delta_\varsigma = 0.00$	0.009	0.082	0.098	0.022	0.142	0.148
	$\Delta_\varsigma = 0.07$	0.011	0.084	0.101	0.023	0.145	0.150
	$\Delta_\varsigma = 0.25$	0.012	0.103	0.107	0.024	0.150	0.153
	$\Delta_\varsigma = 0.50$	0.013	0.104	0.118	0.026	0.153	0.155
10	$\Delta_\varsigma = 0.00$	0.117	0.300	0.343	0.026	0.158	0.164
	$\Delta_\varsigma = 0.07$	0.131	0.319	0.361	0.028	0.162	0.162
	$\Delta_\varsigma = 0.25$	0.141	0.333	0.376	0.029	0.169	0.172
	$\Delta_\varsigma = 0.50$	0.174	0.381	0.417	0.032	0.175	0.177
		$\Delta_k = 333$					
1	$\Delta_\varsigma = 0.07$	0.002	0.046	0.046	0.018	0.135	0.135
	$\Delta_\varsigma = 0.25$	0.002	0.048	0.048	0.018	0.136	0.138
	$\Delta_\varsigma = 0.50$	0.002	0.049	0.049	0.019	0.137	0.137
5	$\Delta_\varsigma = 0.07$	0.009	0.083	0.100	0.022	0.144	0.149
	$\Delta_\varsigma = 0.25$	0.010	0.084	0.103	0.023	0.147	0.151
	$\Delta_\varsigma = 0.50$	0.011	0.087	0.107	0.025	0.150	0.153
10	$\Delta_\varsigma = 0.07$	0.135	0.325	0.368	0.027	0.160	0.165
	$\Delta_\varsigma = 0.25$	0.140	0.333	0.375	0.030	0.164	0.168
	$\Delta_\varsigma = 0.50$	0.144	0.336	0.379	0.029	0.168	0.171
		$\Delta_k = 100$					
1	$\Delta_\varsigma = 0.07$	0.002	0.048	0.049	0.018	0.137	0.137
	$\Delta_\varsigma = 0.25$	0.002	0.049	0.049	0.019	0.137	0.137
	$\Delta_\varsigma = 0.50$	0.002	0.050	0.050	0.019	0.137	0.137
5	$\Delta_\varsigma = 0.07$	0.009	0.080	0.092	0.019	0.140	0.143
	$\Delta_\varsigma = 0.25$	0.010	0.082	0.099	0.022	0.144	0.149
	$\Delta_\varsigma = 0.50$	0.011	0.083	0.100	0.027	0.145	0.150
10	$\Delta_\varsigma = 0.07$	0.140	0.332	0.374	0.027	0.159	0.164
	$\Delta_\varsigma = 0.25$	0.143	0.335	0.378	0.027	0.160	0.165
	$\Delta_\varsigma = 0.50$	0.145	0.337	0.380	0.027	0.161	0.166

Note: The DGP is $y_t = 1 + 0.8y_{t-1} + \varsigma(t)\sigma_t + \varepsilon_t$ and $\sigma_t = 0.2 + 0.1\,|\varepsilon_{t-1}\sigma_{t-1}| + 0.7\sigma_{t-1}$, where $\varsigma(\tau) = \varsigma_1$ if $\tau > t - \kappa_1$ or $\tau < t - \kappa_2$, and $\varsigma(\tau) = \varsigma_2 = \varsigma_1 + \Delta_\varsigma$ otherwise with $\varsigma_1 = 0.3$, $k = 1{,}000$, $k_1 = (k - k_2) \in \{100, 333, 450\}$ or $\Delta_k \in \{800, 333, 100\}$.

Looking now at the results, from the top panel of Table 2 it appears that the forecasting accuracy deteriorates when the forecasting horizon under consideration increases, as all three performance criteria considered are considerably larger for the 10-steps ahead period. However, comparing the top and bottom part of Table 2 it is clear that the location of the breaks does affect the forecasting performance of the model. Similarly, comparing the benchmark case of $\Delta_\varsigma = 0.00$ in the top panel of Table 2 with $\Delta_\varsigma = 0.50$ it appears that, when the forecasting horizon increases, the greater the magnitude of the break the worse the forecasting accuracy.

Having investigated the out-of-sample forecasting performance of the DAB-AR-M model in the next section we will derive an explicit formula for the bidimensional time varying covariance matrix of $\{\mathbf{y}_t\}$, which, as noted above, is needed in order to obtain a time varying measure of second-order persistence.

4.5 Second Moment Structure

In this section we will examine the second moment structure of the DAB-BVARMA representation in eq. (17). First we will introduce some further notation.

Let $\mathbf{X}^{\otimes 2} = \mathbf{X} \otimes \mathbf{X}$ where $\otimes$ is the Kronecker product. In addition, let $vec(\mathbf{X})$ be a vector in which the columns of matrix $\mathbf{X}$ are stacked one underneath the other, and $\mathbf{s} = vec(\mathbf{\Sigma})$. Finally, let $\mathbf{\Gamma}_\tau$ denote the zero order bidimensional time varying covariance matrix of $\{\mathbf{y}_\tau\}$ and $\gamma_\tau = vec(\mathbf{\Gamma}_\tau)$, that is $\gamma_\tau = (\mathbb{V}ar(y_\tau)$, $\mathbb{C}ov(y_\tau, \sigma_\tau^\delta)$, $\mathbb{C}ov(y_\tau, \sigma_\tau^\delta)$, $\mathbb{V}ar(\sigma_\tau^\delta))'$.

Assumption 1 (Second-Order). $\lambda_{\max}(\mathbf{\Phi}_n)^{\otimes 2} < 1$, $n = 1, 3$.

Assumption 1 implies that the DAB-BVARMA$(1, 1; 2)$ representation is second-order. The equivalent Assumption for this representation to be first-order is: $\lambda_{\max}(\mathbf{\Phi}_n) < 1$, $n = 1, 3$. Clearly, this condition is sufficient for the condition in Assumption 1 to hold. Due to space considerations the first moment structure of the above process and its Wold-Crámer decomposition are not reported but are available upon request.

The following theorem states expressions for the $\mathbf{\Gamma}_\tau$ (in the interest of brevity the results for higher order time varying covariances are not reported but are available upon request).

Theorem 5 *Consider the general model in eq. (17). Then under Assumption 1 γ_τ is given by*

$$\gamma_\tau = \mathbf{G}(\tau)\mathbf{s}, \tag{21}$$

where

$$\mathbf{G}(\tau)=[g_{ij}(\tau)] = \mathbf{J}^{\otimes 2} + [\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_1^{k_1+r})^{\otimes 2}][\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_1)^{\otimes 2}]^{-1}(\mathbf{\Phi}_1\mathbf{J} + \mathbf{Z}_1)^{\otimes 2}$$
$$+ (\mathbf{\Phi}_1^{k_1+r})^{\otimes 2}\{[\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_2^{k_2-k_1})^{\otimes 2}][\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_2)^{\otimes 2}]^{-1}(\mathbf{\Phi}_2\mathbf{J} + \mathbf{Z}_2)^{\otimes 2}$$
$$+ (\mathbf{\Phi}_2^{k_2-k_1})^{\otimes 2}(\mathbf{I}^{\otimes 2} - \mathbf{\Phi}^{\otimes 2})^{-1}(\mathbf{\Phi}\mathbf{J} + \mathbf{Z})^{\otimes 2}\},$$

and thus $\mathbf{G}_1 = [g_{ij,1}] = \lim_{r\to\infty} \mathbf{G}(\tau)$ *is given by*

$$\mathbf{G}_1 = \mathbf{J}^{\otimes 2} + [\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_1)^{\otimes 2}]^{-1}(\mathbf{\Phi}_1\mathbf{J} + \mathbf{Z}_1)^{\otimes 2}. \tag{22}$$

Clearly, if we set $k_1 = k_2$, and therefore $\mathbf{\Phi}_1 = \mathbf{\Phi}_2$ and $\mathbf{Z}_1 = \mathbf{Z}_2$ (the case of one break), then we obtain the results for the simpler case, where we have only one abrupt break, at time $t - k_2$. In this case the form for $\mathbf{G}(\tau)$ simplifies to

$$\mathbf{G}(\tau) = \mathbf{J}^{\otimes 2} + [\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_2^{k_2+r})^{\otimes 2}][\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_2)^{\otimes 2}]^{-1}(\mathbf{\Phi}_2\mathbf{J} + \mathbf{Z}_2)^{\otimes 2} + (\mathbf{\Phi}_2^{k_2+r})^{\otimes 2}(\mathbf{I}^{\otimes 2} - \mathbf{\Phi}^{\otimes 2})^{-1}(\mathbf{\Phi}\mathbf{J} + \mathbf{Z})^{\otimes 2},$$

and thus $\mathbf{G}_2 = [g_{ij,2}] = \lim_{r\to\infty} \mathbf{G}(\tau)$ is given by

$$\mathbf{G}_2 = \mathbf{J}^{\otimes 2} + [\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_2)^{\otimes 2}]^{-1}(\mathbf{\Phi}_2\mathbf{J} + \mathbf{Z}_2)^{\otimes 2}. \tag{23}$$

Further, if in addition $k_2 = k$, then $\mathbf{G}(\tau)$, since $\mathbf{\Phi}_2 = \mathbf{\Phi}$ and $\mathbf{Z}_2 = \mathbf{Z}$, reduces to the well known formula for the time invariant model

$$\mathbf{G} = \mathbf{J}^{\otimes 2} + [\mathbf{I}^{\otimes 2} - \mathbf{\Phi}^{\otimes 2}]^{-1}(\mathbf{\Phi}\mathbf{J} + \mathbf{Z})^{\otimes 2}, \tag{24}$$

which is the result obtained in Conrad and Karanasos (2015a), but it is expressed in a more compact way. It follows directly from the above theorem that the first element of the time varying covariance vector, γ_τ, which is the time dependent variance of y_τ, is given by

$$\mathbb{V}ar(y_\tau) = \sigma_\varepsilon g_{11}(\tau) + \sigma_{\varepsilon v} 2 g_{12}(\tau) + \sigma_v g_{14}(\tau). \tag{25}$$

Notice that the three time invariant variances for each of the three periods, denoted by $\mathbb{V}ar_n(y_\tau)$, $n = 1, 2, 3$, are obtained from the above expression by replacing $g_{1j}(\tau)$ with $g_{1j,n}$ (see eqs. (22)-(24)). In addition, since the matrices $\mathbf{\Phi}_n$ are upper triangular, the $\mathbf{G}(\tau)$ matrix is also upper triangular and its

$(4,4)$ time invariant element is $g_{44} = \frac{\alpha^2}{1-c^2}$. Thus, the fourth element of γ_τ, which is the time invariant unconditional variance of σ_τ^δ, is given by

$$\mathbb{V}ar(\sigma_\tau^\delta) = \mu_2 - \mu_1^2 = \frac{\alpha^2}{1-c^2}\sigma_v.$$

Since $\sigma_v = \mu_2\kappa$ (see eq. 15) and using $\mu_1 = \frac{\omega}{1-c}$ we obtain (if and only if $c^2 - \alpha^2\kappa < 1$) by straightforward manipulation the standard result (see, e.g., Karanasos, 1999, He and Teräsvirta, 1999, and Karanasos and Kim, 2006):

$$\mu_2 = \frac{(1+c)\omega^2}{(1-c)(1-c^2-\alpha^2\kappa)}. \tag{26}$$

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

4.6 Time Varying Persistence

The most often applied time invariant measures of first-order (or mean) persistence are the 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, both measures would ignore the presence of breaks and in-mean effects and, hence, potentially under or over estimate the persistence in the levels, which is partly induced by the persistence in the conditional variance (see, e.g., Conrad and Karanasos, 2015a).

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). The authors find no evidence pointing to a rejection of a unit root in inflation. However, as we show in Section 4.2 if the in-mean mechanism together with the possible presence of breaks in the in-mean parameter are ignored, then conventional procedures (such as unit root tests) for estimating the persistence in the mean may lead to biased estimates. In particular, they might falsely indicate a unit root, and, hence, suggest the modelling of the differenced series rather than their levels.

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 and to distinguish between the effects of a *mean shock* and a *volatility shock* on the level and conditional variance respectively. 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 $P(x_t\,|e_t) = \mathbb{V}ar(x_t)/\mathbb{V}ar(e_t)$.

As pointed out by Conrad and Karanasos (2015a), 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.

Regarding the DAB-AR-(APGARCH) M model, if $\sigma_{\epsilon v} = 0$, that is there are no asymmetries (see eq. (15)), then ε_t and v_t can be viewed as 'structural' shocks. Thus it follows directly from eq. (25) that:

$$\mathbb{V}ar(y_\tau) = \sigma_\varepsilon g_{11}(\tau) + \sigma_v g_{14}(\tau). \tag{27}$$

Correlated Shocks

In general, the two shocks will be correlated with covariance matrix $\mathbf{\Sigma}$. In this case we will define two uncorrelated shocks with variances equal to one.

The new orthogonal shocks, $\widetilde{\varepsilon}_t$ and $\widetilde{v}_t$, can be obtained from the original shocks via the transformation:

$$\widetilde{\varepsilon}_\tau = \frac{\varepsilon_\tau}{\sqrt{\sigma_\epsilon}}, \;\; \widetilde{v}_\tau = \frac{1}{\sqrt{1-\rho_{\varepsilon v}^2}}(-\rho_{\varepsilon v}\widetilde{\varepsilon}_\tau + \frac{v_\tau}{\sqrt{\sigma_v}})$$

(see Conrad and Karanasos, 2015a [p. 714] for details).

Now, the persistence of the two shocks, $\widetilde{\varepsilon}_t$ and $\widetilde{v}_t$, for the variance of y_t, can be decomposed as follows (see also eqs. (15-17) in Conrad and Karanasos, 2015a):

$$\mathbb{V}ar(y_\tau) = P(y_\tau \, |\widetilde{\varepsilon}\,) + \; P(y_\tau \, |\widetilde{v}\,), \tag{28}$$

where

$$P(y_\tau \, |\widetilde{\varepsilon}\,) = \sigma_\epsilon g_{11}(\tau) + \rho_{\varepsilon v}^2 \sigma_v g_{14}(\tau) + 2\sigma_{\epsilon v} g_{12}(\tau), \;\; P(y_\tau \, |\widetilde{v}\,) = \sigma_v(1-\rho_{\varepsilon v}^2) g_{14}(\tau).$$

Clearly, if we have the symmetric case, that is $\gamma = 0$ and, therefore, $\sigma_{\epsilon v} = 0$ the above expression reduces to the one in eq. (27). To save space the equivalent persistence measures for the power transformed conditional variance and also for the product $y_t \sigma_t^\delta$ are not reported but are available upon request. Notice that when $\delta = 1$ the above expressions (since $\sigma_\epsilon = \mu_2$, $\sigma_v = \mu_2(1+\gamma^2 - \frac{2}{\pi})$ and $\sigma_{\epsilon v} = -\mu_2 \gamma$ [see Corollary 2]) reduce to

$$P(y_\tau \, |\widetilde{\varepsilon}\,) = \mu_2[g_{11}(\tau) + \gamma^2 g_{14}(\tau) - 2\gamma g_{12}(\tau)], \;\; P(y_\tau \, |\widetilde{v}\,) = \mu_2(1-\frac{2}{\pi}) g_{14}(\tau), \tag{29}$$

where μ_2 is given in eq. (26) with $c = \alpha\sqrt{\frac{2}{\pi}} + \beta$ (see eq. (14) and Corollary 2).

If Assumption 1 is violated then conditional measures of second-order persistence can be constructed using the variance of the forecast error (see eq. D.3 in the Appendix D) 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.[16]

5 Inflation Data

As noted earlier the two strands of literature on 'time varying'models (the one with stochastic coefficients and the other with deterministically varying coefficients) have evolved rather separately so far. However, there is an even greater dichotomy between theoretical and empirical work. So we have a dichotomy within the dichotomy. By bringing the two strands closer to each other in Section 2, and also by a direct confrontation of economic theory with empirical evidence in this section, we hope if not to eliminate these two dichotomies at least to reduce them.

In our empirical application we consider log-differences of quarterly data of Personal Consumption Expenditure (CPE) in the United States from 1947Q1 to 2016Q3. The CPE index is used by the Federal Reserve as inflation proxy when reviewing economic conditions and charting a course of action designed to impact on the real economy. It is therefore crucially important to be able to make appropriate modelling, inference and forecasting in order to avoid unwanted effects of monetary policy actions.

In general, economists have placed considerable emphasis on the impact of inflation uncertainty on both inflation and output growth. Friedman (1977) states that nominal uncertainty causes an adverse output effect. This argument is based on the viewpoint that uncertainty about future inflation distorts the allocative efficiency aspect of the price mechanism (for details, see for example, Fountas et al., 2006, and the references therein). Following the influential work of Friedman a rich literature highlights the importance of nominal uncertainty for macroeconomic modelling and policy making. In particular, according to Cukierman and Meltzer (1986) in the presence of uncertainty about the rate of monetary growth and, therefore, inflation, the policymaker applies an expansionary monetary policy in order to surprise the agents and enjoy output gains. The argument that Central Banks tend to create inflation surprises in the presence of more inflation uncertainty (hereafter, termed the Cukierman and Meltzer hypothesis) implies a positive causal effect from inflation uncertainty to inflation (for details, see for example, Fountas and Karanasos, 2007, and the references therein).

One of the first papers to test for the Cukierman and Meltzer hypothesis in a context of a GARCH-M

[16] Cogley and Sargent (2001) 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: $S_0 = \frac{\sigma_\varepsilon^2}{2\pi(1-\phi_1-\phi_2)^2}$.

model was Baillie et al. (1996); see also Brunner and Hess (1993). However, the econometric specifications which are employed in most of these studies do not take into consideration the time dependent characteristics of the models. Time variation may explain why inflation in the United States has become harder to be modelled and forecasted in recent years (see, for example, Stock and Watson, 2007). In this respect, allowing for time dependent coefficients should be able to accommodate for structural changes in the economy and the resulting shifts in the private sector behavior. Recently, Conrad and Karanasos (2015b), using monthly data, have found strong evidence that higher U.S. nominal uncertainty increases the average inflation rate for the period 1960-2010. To control for possible changes in the conduct of monetary policy, they re-estimate their favoured specification by interacting the main variables of interest with dummy variables for the period 1980–2010.

The first step in the estimation procedure is to identify possible points of parameter changes. In order to do so the Bai and Perron (2003) breakpoint estimation technique on inflation rates is used to identify possible breaks during the sample period.[17] Using the Bai and Perron procedure two significant breaks were identified. The first break took place in the 1960's, when an expansion of social programs was undertaken by the U.S. administration in the aftermath of a contraction period when unemployment and inflation reached high levels. The second break occurred in 2008 at the height of the financial crisis and during a spike of the oil price. In particular, the breaks were identified in 1966Q4 and 2008Q3. [18]

Accordingly, below we estimate the DAB-AR$(1;2)$-M model in eqs. (12)-(13) (with $\delta = 1$)[19] allowing for both the *intrinsic* persistence of inflation (as captured by the autoregressive coefficient, $\phi(t)$) and the in-mean coefficient, $\varsigma(t)$, to switch across breakpoints. [20] This should allow us to determine whether changes in the structure of the conditional mean of inflation recently observed in the U.S. derive from changes in either $\phi(t)$ or $\varsigma(t)$, or possibly both. To capture these changes we use dummy variables that take the value zero in the period before each break and the value one after the break. [21]

[17] Since the seminal paper by Perron (1989) a great deal of research has been directed to the detection and estimation of breaks, and forecasting in the presence of breaks (see, e.g., Andrews, 1993; Andrews and Ploberger, 1994; Bai and Perron, 1998).

[18] Kim et al. (2004) found evidence for a structural break in inflation in late 1979, resulting in lower persistence. The Bai and Perron methodology also identified a third break in 1977Q4. However, in the estimation of the DAB-AR-M model the corresponding dummy variable was insignificant. Pivetta and Reis (2007) point out that extra data points in their sample (1965-2001) might show a break in 1991. The Bai and Perron methodology also identified a fourth break in 1991Q3. However, in the estimation of our model the corresponding dummy variable was insignificant.

Another line of work has identified changes in the way monetary policy is conducted in the United States. Therefore we also add a dummy variable for the period 1981–1983, which was an anomalous period in the data for inflation, commonly referred to as the Volcker disinflation. The dummy was insignificant.

[19] Karanasos and Schurer (2008) show that it is optimal to model the conditional standard deviation of inflation instead of the conditional variance. So far the relevant empirical literature has ignored this important characteristic of the inflation data.

[20] The asymptotic theory for the QML estimator of the parametric GARCH-M model has recently been developed by Conrad and Mammen (2016). However, this theory does not yet treat all standard specifications.

[21] An alternative approach to account for structural breaks would be to estimate similar models for subperiods. Due to the limited number of quarterly observations we do not consider this option.

As far as the QML estimation results are concerned Table 3 reports the estimated parameters for each of the three models and the relative misspecification tests. In particular, the top part of Table 3 reports the estimated parameters for the conditional mean, whereas the ones for the conditional variance are given in Panel B.

Table 3. Estimated DAB-AR$(1;2)$-M model using U.S. inflation data

	Model 1	Model 2	Model 3
Panel A: Conditional Mean			
φ	0.0009 (0.0006)	0.0004 (0.0005)	0.0004 (0.0005)
ϕ_3	0.405^{*} (0.093)	0.545^{*} (0.059)	0.377^{*} (0.088)
$\Delta\phi_3$ ($\phi_2{=}\phi_3{-}\Delta\phi_3$)	-0.318^{*} (0.085)	–	-0.28^{*} (0.153)
$\Delta\phi_2$ ($\phi_1{=}\phi_3{-}\Delta\phi_3{-}\Delta\phi_2$)	0.263^{**} (0.142)	–	–
ς_3	0.481^{**} (0.217)	0.432^{**} (0.188)	0.643^{**} (0.191)
$\Delta\varsigma_3$ ($\varsigma_2{=}\varsigma_3{-}\Delta\varsigma_3$)	–	-0.613^{**} (0.153)	–
$\Delta\varsigma_2$ ($\varsigma_1{=}\varsigma_3{-}\Delta\varsigma_3{-}\Delta\varsigma_2$)	–	0.376^{***} (0.227)	0.181 (0.230)
Panel B: Conditional Variance			
ω	0.0001^{**} (0.000)	0.0002^{**} (0.000)	0.0001 (0.000)
α	0.092^{**} (0.044)	0.103^{**} (0.052)	0.103^{**} (0.048)
γ	-0.803^{**} (0.370)	-0.791^{***} (0.358)	-0.671^{**} (0.305)
β	0.871^{*} (0.044)	0.857^{*} (0.055)	0.868^{*} (0.048)
$c = \alpha\kappa_1 - \beta$	0.944	0.939	0.950
R^2	0.625	0.626	0.616
Panel C: Q-Statistics and Information Criteria			
Q-Statistics (4)	1.702 [0.199]	0.852 [0.310]	0.335 [0.562]
Akaike	−8.403	−8.004	−8.432
Schwarz	−8.269	−7.789	−8.376
Panel D: Forecasting			
Conditional Mean			
MSE	0.000	0.000	0.000
MAE	0.005	0.005	0.004
RMSE	0.004	0.006	0.006
Conditional Variance			
MSE	0.000	0.000	0.000
MAE	0.006	0.005	0.008
RMSE	0.007	0.002	0.005

Note: *, **, *** indicate statistical significance at 1%, 5% and 10%, respectively.
The numbers in parentheses are standard errors. The numbers in brackets are p-values.
$-\Delta\phi_3$, $-\Delta\varsigma_3$ and $-\Delta\phi_2$, $-\Delta\varsigma_2$ are the estimated parameters for the first and second dummy, respectively.
Accordingly, in Model 1: $\phi_2 = 0.723$ and $\phi_1 = 0.460$; In Model 2, $\varsigma_2 = 1.045$ and $\varsigma_1 = 0.669$; In Model 3: $\phi_2 = 0.657$ and $\varsigma_1 = 0.462$.

In order to investigate whether changes in inflation were due to breaks in the *intrinsic* persistence coefficient or the in-mean coefficient the model in eqs. (12)-(13) was estimated with no dummy variables for the latter (i.e., with $\varsigma_2 = \varsigma_1 = 0$ in Table 3) and the resulting specification is labelled as Model 1. Similarly, to investigate if breaks in the in-mean coefficient did affect inflation, the model was estimated with no dummies capturing breaks in the *intrinsic* persistence (i.e., with the parameters $\phi_2 = \phi_1 = 0$) and the resulting specification is referred to as Model 2. Finally, to investigate the joint effects of the two types of breaks various attempts were made to estimate a model with breaks in both coefficients. The best (based on information criteria and likelihood ratio tests) resulting specification is labelled as Model 3. [22]

Looking now at the estimated parameters, according to the estimates in Model 1 until 1966 the inflation process was well approximated by a first-order autoregression with low *intrinsic* persistence ($\phi_3 = 0.40$), but from 1967 onwards the autoregression coefficient increased considerably ($\phi_2 = 0.72$). It was only after 2008 that the *intrinsic* persistence level went back to roughly its previous regime ($\phi_3 = 0.46$).

According to Model 2 it appears that inflation variability imposed upward pressure on inflation ($\varsigma = 0.43$). In general, higher inflation uncertainty increases long term risk premia, inducing extra hedging costs due to inflation risks therefore shifting upward inflation levels, as predicted by Cukierman and Meltzer. Interestingly, the in-mean effect becomes stronger (it more than doubles in size) after 1966 ($\varsigma_2 = 1.04$). However, in the post-(global financial) crisis it decreases ($\varsigma_1 = 0.67$) although it does not return to the pre-1967 level. Finally, Model 3 confirms that changes in inflation dynamics can be explained by both changes in the *intrinsic* persistence and the in-mean coefficient. In all models the estimated intrinsic variance persistence is high (c is either 0.94 or 0.95) and the asymmetry coefficient is negative, indicating that negative shocks have a higher impact on the volatility than positive shocks.

Looking now at the specification tests in Panel C the Q-Statistics do not reject the null hypothesis of no serial correlation, therefore indicating that the models do not suffer from misspecification. Also, from the information criteria it appears that Model 2 offers the best specification for the inflation process. Finally, the bottom part reports the 5-step ahead forecasting criteria for the models under consideration. It appears that all three models have relatively good forecasting properties.

To summarize our results, we find evidence that the parameters in the models capturing *intrinsic* persistence and in-mean effects 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 the light of the simulation results in Section 4.4, may lead to poor forecasting.

[22] There is also the possibility of a change in the drift of inflation. Such a shift can be interpreted as a change in the long-run inflation target of the Federal Reserve (see, for details, Pivetta and Reis, 2007 and the references therein). Our DAB-AR-M model also allows for changes in the intercept. However, and in spite of allowing for a time varying drift, we find that the second-order persistence is unchanged (see the analysis below), a result which is in agreement with the one in Pivetta and Reis.

Next we will investigate whether inflation and its variability are highly persistent.

5.1 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. [23] In this section we depart from their study in an important way, that is we contribute to the measurement over time of inflation persistence by taking a different approach to the problem and estimating a model of inflation dynamics grounded on economic (instead of a statistical) theory. In particular, we employ the DAB AR-(APGARCH) M model and we compute an alternative measure of persistence, that is, the second-order persistence (using the methodology in Sections 4.5 and 4.6), which not only distinguishes between changes in the dynamics of inflation and its volatility (and their persistence) but also allows for feedback from volatility (inflation uncertainty) to the level of the process (inflation).

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 4 presents the within each period time invariant second-order measures of persistence for all three models and periods. The first three columns report the persistence for the mean shock ($\widetilde{\varepsilon}$), the next three columns the persistence for the volatility shock ($\widetilde{v}$), and the last three columns the sum of the two shocks (see eq. (29)). Model 2, which is the preferred one, generates the highest persistence. In particular, the persistence increases by 32% in the post-1966 period and decreases by 16% in the post-crisis period. Model 1 is the one with the lowest persistence. For this model the persistence doubles in the second period but after the global financial crisis it almost returns to the pre-1967 levels. In model 3 the persistence increases by 75% in the period 1967-2008 and decreases by 8% in the post-crisis period. Figure 4 presents the time varying inflation persistence for the three models.

[23] Pivetta and Reis (2007) applied a Bayesian approach, which explicitly treats the autoregressive parameters as being stochastically varying and it provides their posterior densities at all points in time. From these, they obtained posterior densities for the measures of inflation persistence. Such estimates of persistence are forward-looking, since they are meant to capture the perspective of a policy maker who at a point in time is trying to forsee what the persistence of inflation will be. They also estimated backward-looking measures of persistence that the applied economist forms at a point in time, given all the sample until then.

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

Table 4. 'Second-order persistence' for each of the three periods and models.

	$P(y_t \,\|\widetilde{\varepsilon}) \times 10^6$			$P(y_t \,\|\widetilde{v}) \times 10^6$			$[P(y_t \,\|\widetilde{\varepsilon}) + P(y_t \,\|\widetilde{v})] \times 10^6$		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
1947Q1-1966Q4	4.51	18.81	5.64	0.062	0.32	0.18	4.57	19.13	5.82
1967Q1-2008Q2	9.01	23.31	9.67	0.26	1.87	0.55	9.26	25.18	10.22
2008Q3-2016Q3	4.84	20.31	9.12	0.07	0.79	0.28	4.92	21.10	9.40

Note: We use eq. (29) to calculate the (within each period time invariant) second-order persistence for the three models. For each period, $n = 1, 2, 3$, we obtained the $g_{1j,n}$, using eqs. (22)-(24).

For comparison Table A.1 in the Appendix presents the (time invariant within each period) 'first-order persistence' for all three models. The pre-1967 period exhibits the lowest persistence whereas in the second period the persistence is the highest. In particular, for the second model from 1967 onwards it increases by 60%. For all models in the post global financial crisis period the persistence reduces but it remains higher than the pre-1967 levels except perhaps for model 1, where it more than doubles in the second period and then it almost halves in the post-crisis period.

Therefore our findings regarding the first-order persistence are in line with the findings of i) Barsky (1987), who found very low inflation persistence for the pre-war period but that inflation is very persistent since the 1960s, and ii) Pivetta and Reis (2007), who by computing alternative statistical measures of persistence came to the conclusion that inflation persistence in the United States is best described as unchanged over their sample, which was 1965-2001.

In sum our main conclusion is that for our chosen specification (model 2) the preferred measure of persistence, that is the second-order persistence, increased considerably from 1967 onwards, whereas in the post-crisis period the persistence reduces but it remains higher than the pre-1967 levels. Pivetta and Reis (2007) and Stock (2001), who applied a subset of the classical methods used by Pivetta and Reis, also found no evidence of a change in the US inflation persistence. [24]

[24] Stock and Watson (2002) also found no evidence of a change in persistence in U.S. inflation. Therefore their results are in agreement with ours. However, they found strong evidence of a fall in volatility. We also checked for possible changes in the unconditional volatility by adding dummy variables in the drift of the conditional variance. However, they are all insignificant.

6 Conclusions

It is important to understand the fundamental properties of 'linear' time series models with time dependent 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 a GARCH-in-mean specification with abrupt breaks, which is grounded on economic theory. In the context of this in-mean formulation we show how our results can be easily extended to a VAR system. 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 stochastic linear time varying difference equations. 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, Markov switching processes, and state-space formulations). 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 1967, whereas in the post-crisis period the persistence reduces but it remains higher than the pre-1967 levels, a finding which agrees with those of Stock and Watson (2002) and Pivetta and Reis (2007). The important implication of this outcome is that theories should predict very persistent inflation rates, as argued by Pivetta and Reis.

The importance 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 changing coefficients, since it releases us from the need to work with characteristic polynomials,

ii) obtain the fourth moments of 'time varying' GARCH models, which themselves follow linear time varying difference equations of infinite order, taking advantage of the fact that the various GARCH formulations have ARMA representations,

and, in view of being easily applied to a multivariate setting, to:

iv) work out the fundamental time series properties of 'time varying' linear VAR systems,

v) 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).

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

References

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

[2] ANDREWS, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821-856.

[3] ANDREWS, D. W. K. AND W. PLOBERGER (1994): "Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative," *Econometrica*, 62, 1383-1414.

[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 (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47-78.

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

[7] BAILLIE, R., C. CHUNG, AND M. TIESLAU (1996): "Analyzing Inflation by the Fractionally Integrated ARFIMA-GARCH Model," *Journal of Applied Econometrics*, 11, 23-40.

[8] BARSKY, R. (1987): "The Fisher Hypothesis and the Forecastability and Persistence of Inflation," *Journal of Monetary Economics*, 19, 3–24.

[9] BRUNNER, A. D. AND G. D. HESS (1993): "Are Higher Levels of Inflation Less Predictable? a State-Dependent Conditional Heteroscedasticity Approach," *Journal of Business & Economic Statistics*, 11, 187-197.

[10] CHRISTENSEN, B. J. AND M. Ø. NIELSEN (2007): "The Effect of Long Memory on Stock Market Fluctuations," *The Review of Economics and Statistics*, 89, 684-700.

[11] COGLEY, T. AND T. J. SARGENT (2001): " 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 ECCC-GARCH Model," *Econometric Theory*, 26, 838-862.

[14] CONRAD, C. AND M. KARANASOS (2015a): "On the Transmission of Memory in GARCH-in-mean Models," *Journal of Time Series Analysis*, 36, 706-720.

[15] CONRAD, C. AND M. KARANASOS (2015b): "Modeling the Link Between US Inflation and Output: the Importance of the Uncertainty Channel," *Scottish Journal of Political Economy*, 2015, 62, 431-453.

[16] CONRAD, C. AND E. MAMMEN (2016): "Asymptotics for Parametric GARCH-in-mean Models," *Journal of Econometrics*, 194, 319-329.

[17] COOLEY, T. F. AND E. C. PRESCOTT (1976): "Estimation in the Presence of Stochastic Parameter Variation," *Econometrica*, 44, 167-184.

[18] CUKIERMAN, A. AND A. H. MELTZER (1986): "A Theory of Ambiguity, Credibility, and Inflation Under Discretion and Asymmetric Information," *Econometrica*, 54, 1099–1128.

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

[20] DICKEY, D. A. AND W. A. FULLER (1981): "Likelihood Ratio Statistics for Autoregressive Time Series With a Unit Root," *Econometrica*, 49, 1057-1072 .

[21] DING, Z., C. W. J. GRANGER, AND R. F. ENGLE (1993): "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83-106.

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

[23] ELLIOTT, G. AND A. TIMMERMANN (2008): "Economic Forecasting," *Journal of Economic Literature*, 46, 3-56.

[24] ENGLE, R. F., D. M. LILIEN, AND R. P. ROBINS (1987): "Estimating Time Varying Risk Premia in the Term Structure: the ARCH-M Model," *Econometrica*, 55, 391-407.

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

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

[27] 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.

[28] FOUNTAS, S. AND M. KARANASOS (2007): "Inflation, Output Growth, and Nominal and Real Uncertainty: Empirical Evidence for the G7," *Journal of International Money and Finance*, 26, 229-250.

[29] FOUNTAS, S., M. KARANASOS, AND J. KIM (2006): "Inflation Uncertainty, Output Growth Uncertainty, and Macroeconomic Performance," *Oxford Bulletin of Economics and Statistics*, 68, 319-343.

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

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

[32] FRIEDMAN, M. (1977): "Nobel Lecture: Inflation and Unemployment," *Journal of Political Economy*, 85, 451–472.

[33] 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.

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

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

[36] GLOSTEN, L. A., R. JAGANNATHAN, AND D. E. RUNKLE (1993): "On the Relation Between the Expected Value and the Volatility of Nominal Excess Return on Stocks," *Journal of Finance*, 46, 1779-1801.

[37] 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.

[38] 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.

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

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

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

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

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

[44] HALLIN, M. (1978): "Mixed Autoregressive-Moving Average Multivariate Processes with Time-Dependent Coefficients," *Journal of Multivariate Analysis*, 8, 567-572.

[45] 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.

[46] HE, C. AND T. TERÄSVIRTA (1999): "Fourth Moment Structure of the GARCH(p, q) Process," *Econometric Theory*, 15, 824-846.

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

[48] 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.

[49] KARANASOS, M. (1998): "A new Method for Obtaining the Autocovariance of an ARMA Model: an Exact Form Solution," *Econometric Theory,* 14, 622-7640.

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

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

[52] KARANASOS, M. AND J. KIM (2006): "A Re-examination of the Asymmetric Power ARCH Model," *Journal of Empirical Finance*, 13, 113-128.

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

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

[55] 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," arXiv.

[56] 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.

[57] 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.

[58] KARANASOS, M. AND S. SCHURER (2008): "Is the Relationship Between Inflation and its Uncertainty Linear?," *German Economic Review*, 9, 265-286.

[59] KIM, C. J., C. R. NELSON, AND J. PIGER (2004) "The Less Volatile U.S. Economy: a Bayesian Investigation of Timing, Breadth, and Potential Explanations," *Journal of Business & Economics Statistics* 22, 80–93.

[60] KOOP, G. AND D. KOROBILIS (2012): "Forecasting Inflation Using Dynamic Model Averaging," *International Economic Review*, 53, 867-886.

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

[62] KOOP, G. AND S. M. POTTER (2007): "Estimation and Forecasting in Models with Multiple Breaks," *The Review of Economic Studies*, 74, 763-789.

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

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

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

[66] 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.

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

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

[69] MILLER, K. S. (1968): *Linear Difference Equations*. Benjamin.

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

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

[72] PARASKEVOPOULOS, A. G. AND M. KARANASOS (2014): "A Closed Form for the General Solution of a Linear Difference Equation with Variable Coefficients,". arXiv:1412.8521.

[73] PERRON, P. (1989): "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica*, 57, 1361-1401.

[74] 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.

[75] PESARAN, M. H., D. PETTENUZZO, AND A. TIMMERMANN (2006): "Forecasting Time Series Subject to Multiple Structural Breaks," *The Review of Economic Studies*, 73, 1057-1084.

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

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

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

[79] PRIMICERI, G. E. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *The Review of Economic Studies*, 72, 821-852.

[80] 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.

[81] 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.

[82] STOCK, J. H. (2001) "Comment on: Evolving Post-World War II U.S. Inflation Dynamics," In: Gertler, M., Rogoff, K. (Eds.), NBER Macroeconomics Annual, pp. 379-387. MIT Press, Cambridge.

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

[84] STOCK, J. H. AND M. W. WATSON (2007): "Why has U.S. inflation Become Harder to Forecast?," *Journal of Money, Credit and Banking*, 39, 3–33.

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

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

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

[88] ZEMKE, J-P. M. (2006): "Hessenberg Eigenvalue-Eigenmatrix Relations," *Linear Algebra and its Applications*, 414, 589-606.

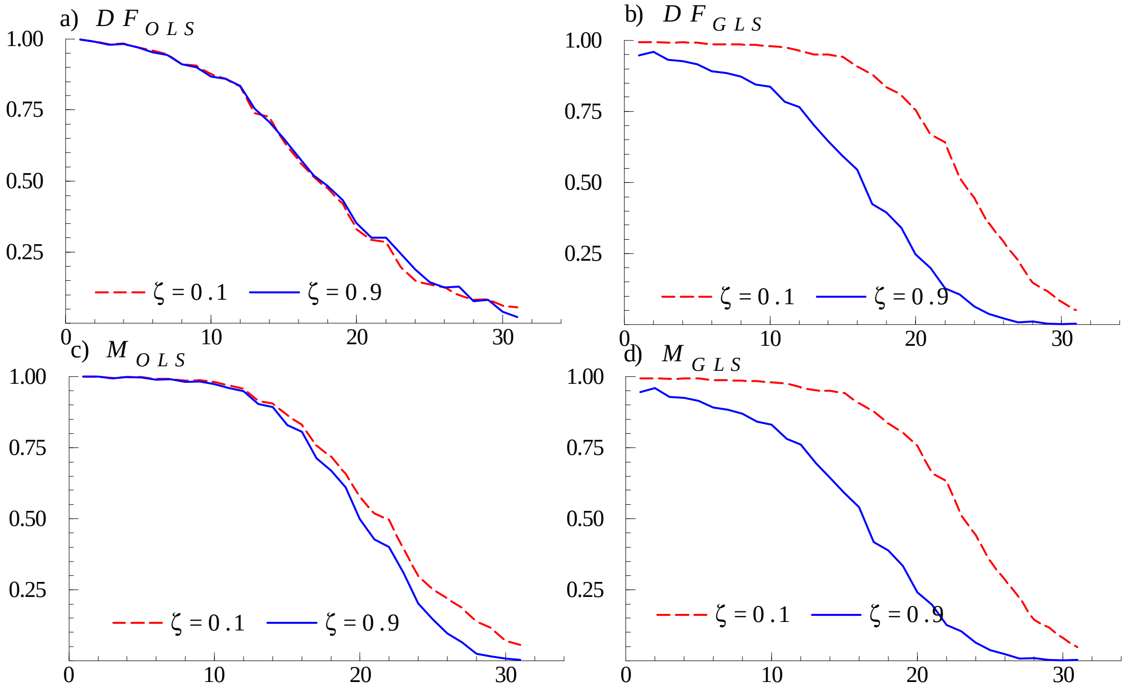
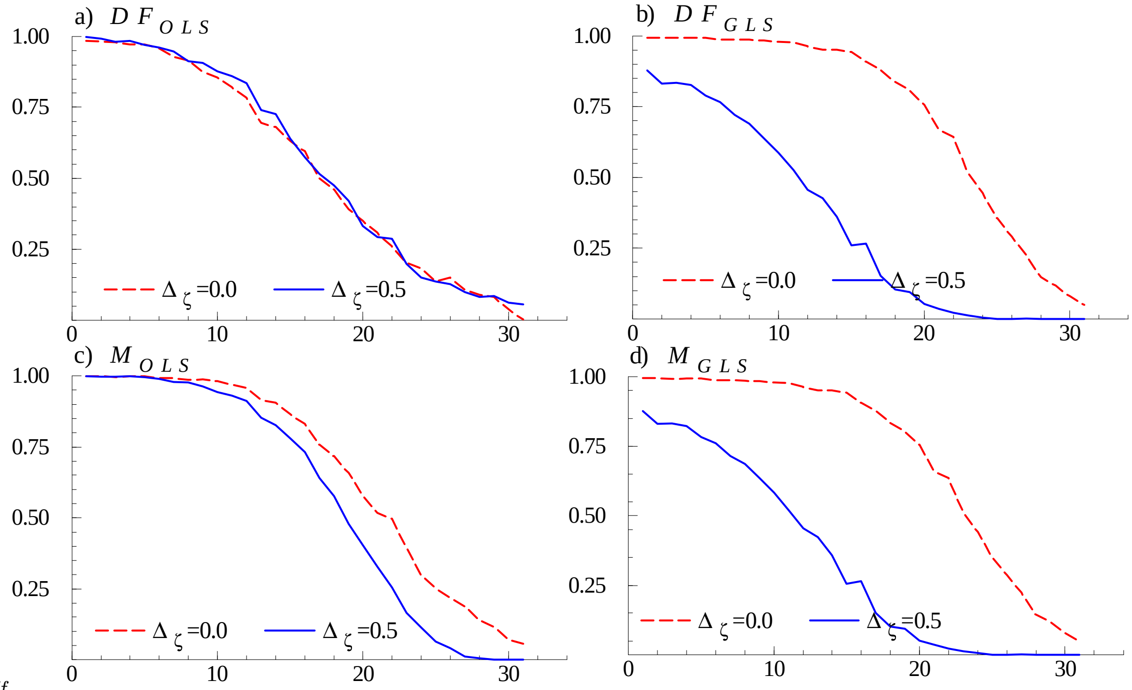


Figure 1. Power of DF and M tests. The DGP is $y_t = 1 + y_{t-1} + \varsigma\sigma_t + \varepsilon_t$ and $\sigma_t = 0.2 + 0.1\,|\varepsilon_{t-1}\sigma_{t-1}| + 0.7\sigma_{t-1}$.

$_T AU01$



$_2.pdf$

Figure 2: Power of DF and M tests. The DGP is $y_t = 1 + y_{t-1} + \varsigma(t)\,\sigma_t + \varepsilon_t$ and $\sigma_t = 0.2 + 0.1\,|\varepsilon_{t-1}\sigma_{t-1}| + 0.7\sigma_{t-1}$, where $\varsigma(\tau) = \varsigma_1$ if $\tau > t - k_1$ or $\tau < t - k_2$, and $\varsigma(\tau) = \varsigma_2 = \varsigma_1 + \Delta_\varsigma$ otherwise with $\varsigma_1 = 0.9$, $k = 1,000$, $k_1 = (k - k_2) = 100$ or $\Delta_k = 800$.

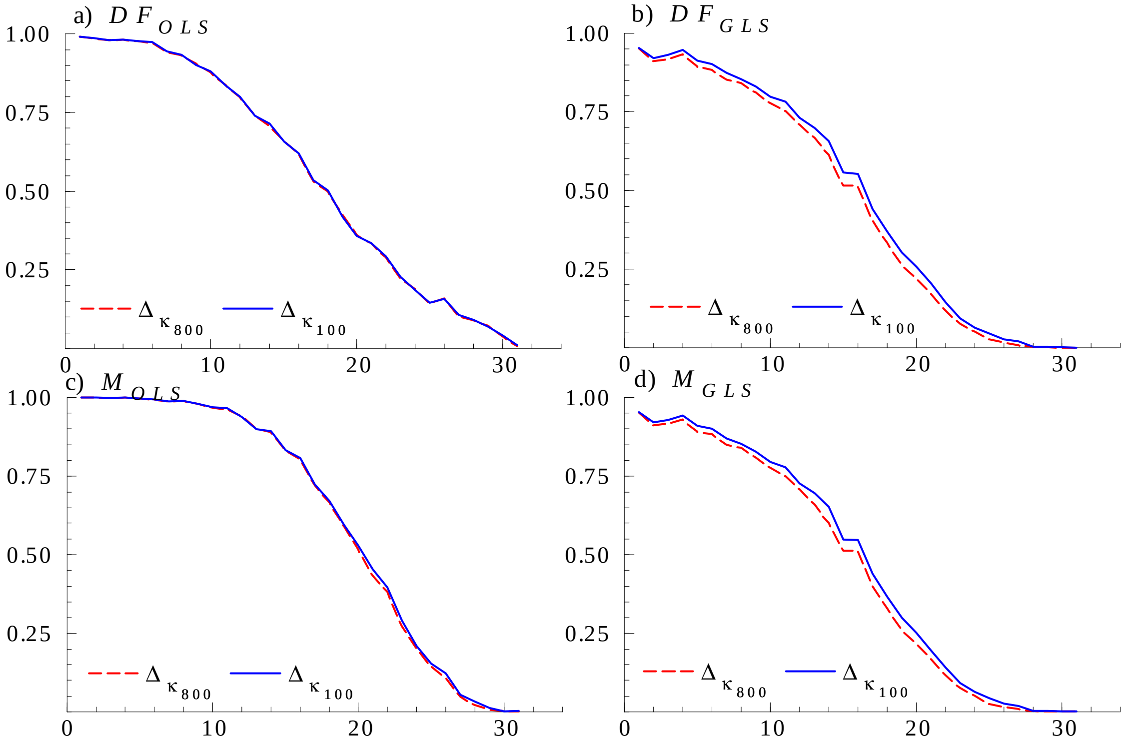


Figure 3. Power of DF and M tests. The DGP is $y_t = 1 + y_{t-1} + \varsigma(t)\,\sigma_t + \varepsilon_t$ and $\sigma_t = 0.2 + 0.1\,|\varepsilon_{t-1}\sigma_{t-1}| + 0.7\sigma_{t-1}$, where $\varsigma(\tau) = \varsigma_1$ if $\tau > t - k_1$ or $\tau < t - k_2$, and $\varsigma(\tau) = \varsigma_2 = \varsigma_1 + \Delta_\varsigma$ otherwise with $\varsigma_1 = 0.0$, $\Delta_\varsigma = 0.07$, $k = 1,000$, $k_1 = (k - k_2) \in \{100, 450\}$ or $\Delta_k \in \{800, 100\}$

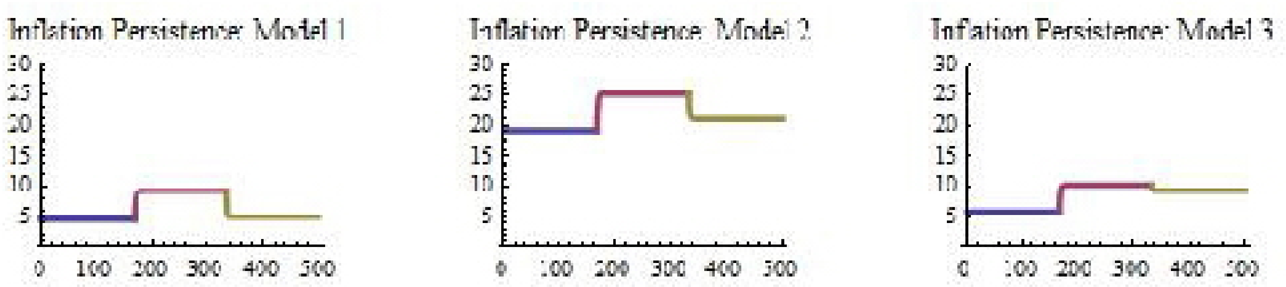


Figure 4. Time Varying Inflation Persistence

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

A APPENDIX

Solution of Ascending Order LDEs

In this appendix we present a method which yields explicit solution expressions for ascending order time varying linear difference equations of index p (ATV-LDE(p)). The class of linear difference equations with variable coefficients (TV-LDEs(p)) which represents TV-ARMA models is a special case of ATV-LDEs(p). The index p coincides with the dimension of the homogeneous solution space for both equation types and generalizes the notion of the order for TV-LDEs(p). Thanks to a special decomposition of the ATV-LDE(p), a non-singular infinite system representation of the equation is derived, making possible the implementation of Cramer's rule and yielding a complete Hessenbergian (determinant of a lower Hessenberg matrix) to the general solution.

An alternative method for the Hessenbergian solution of ATV-LDE(p) is based on the infinite Gaussian elimination process equipped with right pivoting, as demonstrated in (Paraskevopoulos 2014). Both methods lead to the same Hessenbergian form for the general solution of TV-ARMA processes, but in a reverse order. The infinite Gaussian elimination process can be characterized as algorithmic, primarily constructing the fundamental and particular parts of the solution appearing as sequences of Hessenbergian expansions (for an application of this technique to low order periodic models see Karanasos et al., 2014a; this paper illustrates the relative ease with which the infinite Gaussian algorithm can be applied). The general solution is the sum of these two parts. The approach discussed in the present paper can be characterized as heuristic (based on Cramer's rule), which directly yields the general solution as a Hessenbergian. Fundamental and particular solutions are derived from the general solution as special cases. Even-though the infinite Gauss-Jordan elimination algorithm applies to a considerably broader class of infinite systems, comprising all row-finite ones (either singular or non-singular), the solution method developed here yields the general Hessenbergian solution primarily and considerably faster.

An ATV-LDE(p) is given by

$$y_t = \varphi(t) + \sum_{m=\tau-p+1}^{t-1} \phi_{t-m}(t) y_m + u_t, \quad t \in \mathbb{Z}, \tag{A.1}$$

where $\mathbb{I}_\tau = [\tau - p + 1, \tau] \cap \mathbb{Z}$ is the information time-interval and

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

Taking into account that $\tau = t - k$ is fixed, that is as t increases k increases as well, and using the shorthand notation $v_t = \varphi(t) + u_t$, we can also write eq. (A.1) as

$$y_t = \sum_{m=1}^{k+p-1} \phi_m(t) y_{t-m} + v_t. \tag{A.2}$$

A comparison between eq. (1) and eq. (A.2) shows that the greatest time lag between iterates in eq. (1) is the constant p (that is the order of the equation $t - (t - p) = p$), while in eq. (A.2) the corresponding time lag increases as t increases ($t-(t-k-p+1) = t-\tau+p-1$). The latter justifies the ascending order character of eq. (A.2). However, both types of equations yield the same dimension p of the homogeneous solution space.

A.1 Infinite System Representation

For every fixed pair t, k the solution sequence of the ATV-LDE(p)

$$\sum_{m=1}^{p+r-1} \phi_m(t-k+r) y_{t-k+r-m} - y_{t-k+r} = -v_{t-k+r}, \quad r = 1, 2, 3, ... \tag{A.3}$$

comprises all the solution terms of index greater than τ and including t, called *recursively generated solution terms*. Eq. (A.2) is derived from eq. (A.3) by setting $r = k$.

The infinite matrix representation of eq. (A.3) is given by

$$\begin{pmatrix} \phi_p(\tau+1) & ... & \phi_1(\tau+1) & -1 & 0 & 0 & 0 \; ... \\ \phi_{p+1}(\tau+2) & ... & \phi_2(\tau+2) & \phi_1(\tau+2) & -1 & 0 & 0 \; ... \\ \phi_{p+2}(\tau+3) & ... & \phi_3(\tau+3) & \phi_2(\tau+3) & \phi_1(\tau+3) & -1 & 0 \; ... \\ \vdots & \vdots & \vdots & \vdots & & & \end{pmatrix} \begin{pmatrix} y_{\tau+1-p} \\ y_{\tau+2-p} \\ y_{\tau+3-p} \\ \vdots \end{pmatrix} = \begin{pmatrix} -v_{\tau+1} \\ -v_{\tau+2} \\ -v_{\tau+3} \\ \vdots \end{pmatrix}. \tag{A.4}$$

Let us adopt the convention $\phi_0(s) = -1$ for all $s \in \mathbb{Z}$. Writing the coefficient matrix of eq. (A.4) as $\mathbf{H} = (h_{i,j})_{i,j \in \mathbb{Z}^+}$, its entries are given by:

$$h_{i,j} = \phi_{p+i-j}(\tau + i).$$

We may characterize the matrix **H** in (A.4) as infinite lower Hessenberg.

The zero entries above the superdiagonal of **H** will be called trivial and the entries below and including the super-diagonal of **H** will be called non-trivial. The largest difference between the greatest and the least column index of the non-trivial entries of **H** in each row increases by 1, due to the ascending order character of eq. (A.3). The corresponding difference in the case of TV-LDEs(p) is the order p of the equation. The homogeneous solution space of eq. (A.3) coincides with the null space of **H** in eq. (A.4), being of dimension p (see Paraskevopoulos, 2014), which characterizes both ascending and constant order TV-LDEs.

A.2 Equation Decomposition

In this section we consider the decomposition of eq. (1), which yields, unlike eq. (A.4), a non-singular linear system. We divide the sum in the RHS of eq. (A.2) into two parts:

[First Part] $$\phi_k(t)y_{t-k} + \phi_{k+1}(t)y_{t-k-1} + ... + \phi_{k+p-1}(t)y_{t-k+1-p} = \sum_{m=k}^{k+p-1} \phi_m(t)y_{t-m},$$

called *weighed sum of the initial conditions* (WSIC) and

[Second Part] $$\sum_{m=1}^{k-1} \phi_m(t)y_{t-m},$$

called *weighed sum of recursively generated solution terms* (WRSS). Thus eq. (A.2) can be written as

$$\underbrace{\sum_{m=1}^{k-1} \phi_m(t)y_{t-m} - y_t}_{\text{WSRS}} \quad = \quad -v_t \quad \underbrace{- \sum_{m=k}^{k+p-1} \phi_m(t)y_{t-m}}_{\text{WSIC}} . \tag{A.5}$$

The right hand side of eq. (A.5) comprises the forcing term and the initial conditions, that is the quantities in terms of which y_t must be expressed. The left side consists of the recursively generated solution terms exclusively.

A.3 Two Variable Solutions Representation

The two variable representation for the solutions $\xi_{t,k}$ (more generally $y_{t,k}$) enables us to describe solution sequences, while moving the initial condition set to the past, as k increases. This allows us to derive the stability condition and the Wold-Crámer decomposition along with the first and second unconditional moments of TV-HARMA(p, q) models in Appendix C.

Let t, k be fixed. The primary fundamental solution

$$\{\xi_{t-k+1-p,1-p}, \xi_{t-k+2-p,2-p}, ..., \xi_{t-k,0}, ..., \xi_{t-1,k-1}, \xi_{t,k}, \xi_{t+1,k+1}, ...\},$$

can also be described as $\{\xi_{s,s-t+k}\}_{s\geq t-k-p+1}$ (s is the independent variable). For example setting $s = t-k-p+1$, as $1-p = s-t+k$, we conclude that $\xi_{s,s-t+k} = \xi_{t-k-p+1,1-p}$. Taking into account that $(s-t+k)$ is a dependent variable on s, we fall in the one variable representation, using $\{\xi_s\}_{s\geq t-k-p+1}$ in place of $\{\xi_{s,s-t+k}\}_{s\geq t-k-p+1}$.

Following the two variable notation, as in the main body of the paper, eq. (A.3) can be also represented as

$$\sum_{m=1}^{r-1} \phi_m(t-k+r) y_{t-k+r-m,r-m} - y_{t-k+r,r} = -v_{t-k+r} - \sum_{m=r}^{p+r-1} \phi_m(t-k+r) y_{t-k+r-m}, \tag{A.6}$$

for $r = 1, 2, 3, ...$

A.4 Solution Expressions

In this subsection, we provide two expressions of the general solution $y_{t,k}$ of eq. (A.2). The first solution is expressed as a single Hessenbergian and the second is written in terms of the generating solution $\xi_{t,k-r}$.

As $\sum_{m=r}^{p+r-1} \phi_m(t-k+r) y_{t-k+r-m} = \sum_{m=1}^{p} \phi_{r-1+m}(t-k+r) y_{t-k+1-m}$ eq. (A.6) can be written as

$$\sum_{m=1}^{r-1} \phi_m(t-k+r) y_{t-k+r-m,r-m} - y_{t-k+r,r} = -v_{t-k+r} - \sum_{m=1}^{p} \phi_{r-1+m}(t-k+r) y_{t-k+1-m}, \tag{A.7}$$

provided that $\sum_{m=1}^{0} a_m = 0$. Applying eq. (A.7) with $r = k$ we obtain the equation:

$$y_{t,k} = \sum_{m=1}^{k-1} \phi_m(t) y_{t-m,k-m} + \sum_{m=1}^{p} \phi_{k-1+m}(t) y_{t-k+1-m} + v_t. \tag{A.8}$$

The solution sequence of eq. (A.7) also satisfies the infinite system:

$$
\begin{pmatrix}
-1 & 0 & 0 & 0 \ldots \\
\phi_1(t-k+2) & -1 & 0 & 0 \ldots \\
\phi_2(t-k+3) & \phi_1(t-k+3) & -1 & 0 \ldots \\
\vdots & \vdots & \vdots & \vdots
\end{pmatrix}
\begin{pmatrix}
y_{t-k+1,1} \\ y_{t-k+2,2} \\ y_{t-k+3,3} \\ \vdots
\end{pmatrix}
=
\begin{pmatrix}
-v_{t-k+1} - \sum_{m=1}^{p} \phi_m(t-k+1) y_{t-k+1-m} \\
-v_{t-k+2} - \sum_{m=1}^{p} \phi_{1+m}(t-k+2) y_{t-k+1-m} \\
-v_{t-k+3} - \sum_{m=1}^{p} \phi_{2+m}(t-k+3) y_{t-k+1-m} \\
\vdots
\end{pmatrix}. \tag{A.9}
$$

Formally the solution sequence $(y_{t-k+r,r})_r$ of eq. (A.9) is unique, as the system is non-singular.

A.4.1 Hessenbergian Solution Expression

An expression of $y_{t,k}$ in eq. (A.8) in terms of the initial conditions and the forcing term is established in the following proposition.

Proposition A1 *The (t,k) term of the general solution of eq.* (A.8)*, expressed as a single Hessenbergian of order k, is:*

$$
y_{t,k} = \det(\mathbf{Y}_{t,k}), \tag{A.10}
$$

where

$$
\mathbf{Y}_{t,k} =
\begin{pmatrix}
v_{t-k+1} + \sum_{m=1}^{p} \phi_m(t-k+1) y_{t-k+1-m} & -1 & 0 & \ldots & 0 \\
v_{t-k+2} + \sum_{m=1}^{p} \phi_{1+m}(t-k+2) y_{t-k+1-m} & \phi_1(t-k+2) & -1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots \\
v_{t-1} + \sum_{m=1}^{p} \phi_{k-2+m}(t-1) y_{t-k+1-m} & \phi_{k-2}(t-1) & \phi_{k-3}(t-1) & \ldots & -1 \\
v_t + \sum_{m=1}^{p} \phi_{k-1+m}(t) y_{t-k+1-m} & \phi_{k-1}(t) & \phi_{k-2}(t) & \ldots & \phi_1(t)
\end{pmatrix}. \tag{A.11}
$$

Proof. Eq. (A.8) is derived from the matrix multiplication of the kth row of the system (A.9) by the solution sequence. Equivalently, eq. (A.8) is derived from the multiplication of the last row of the non-singular finite system by the solution sequence (we recall that $\tau = t - k$):

$$\begin{pmatrix} -1 & 0 & ... & 0 & 0 \\ \phi_1(\tau+2) & -1 & ... & 0 & 0 \\ \phi_2(\tau+3) & \phi_1(\tau+3) & ... & 0 & 0 \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots \\ \phi_{k-2}(t-1) & \phi_{k-3}(t-1) & ... & -1 & 0 \\ \phi_{k-1}(t) & \phi_{k-2}(t) & ... & \phi_1(t) & -1 \end{pmatrix} \begin{pmatrix} y_{\tau+1,1} \\ y_{\tau+2,2} \\ y_{\tau+3,3} \\ \vdots \\ y_{t-1,k-1} \\ y_{t,k} \end{pmatrix} = \begin{pmatrix} -v_{\tau+1} - \sum_{m=1}^{p} \phi_m(\tau+1) y_{\tau+1-m} \\ -v_{\tau+2} - \sum_{m=1}^{p} \phi_{1+m}(\tau+2) y_{\tau+1-m} \\ \vdots \\ -v_t - \sum_{m=1}^{p} \phi_{k-1+m}(t) y_{\tau+1-m} \end{pmatrix}. \tag{A.12}$$

In particular, the last term of the solution sequence of eq. (A.12) is $y_{t,k}$. Applying Cramer's rule to the above system, we obtain the Hessenbergian expression of $y_{t,k}$ in eq. (A.10). ■

Putting in eq. (A.11) $v(t-k+r) = 0$ for all r, $y_{t-k} = 1$ and $y_{t-k+1-m} = 0$ for all $m > 1$, we obtain the principal determinant $\xi_{t,k} = \det(\mathbf{\Phi}_{t,k})$, where

$$\mathbf{\Phi}_{t,k} = \begin{pmatrix} \phi_1(t-k+1) & -1 & 0 & ... & 0 & 0 \\ \phi_2(t-k+2) & \phi_1(t-k+2) & -1 & ... & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots \\ \phi_{k-1}(t-1) & \phi_{k-2}(t-1) & \phi_{k-3}(t-1) & ... & \phi_1(t-1) & -1 \\ \phi_k(t) & \phi_{k-1}(t) & \phi_{k-2}(t) & ... & \phi_2(t) & \phi_1(t) \end{pmatrix}, \tag{A.13}$$

which is a full lower Hessenberg matrix. That is, in the case of ATV-LDEs(p), the $\xi_{t,k}$ is represented by a full lower Hessenbergian.

Remark 3 *The infinite Gaussian elimination algorithm reduces the system representation coefficient matrix in eq. (A.4), after applying a sequence of row elementary operators into a row-equivalent Hermite form in such a way that the first p columns of the reduced matrix are the fundamental solution sequences* $\{\xi^{(m)}_{t-k+r,r}, r > 0\}$ *not containing the initial condition values* $\{\delta_{m,r}\}_{1\le r\le p}$*, where* $\delta_{m,r}$ *is the Kronecker's delta. Moreover, applying the same sequence of row elementary operations to the RHS column sequence* $[-v_{t-k+r}]_{r\ge1}$ *of eq. (A.4), we obtain a particular solution sequence of the original difference equation.*

A.4.2 Generating and Primary Fundamental Solution Sequences

Following the terminology and notation of Subsection 2.2, the determinant in eq. (A.13) yields the generating sequence $\{\xi_{t,r}\}_{r\ge1-p}$ coupled with the primary fundamental solution $\{\xi_{t-k+r,r}\}_{r\ge1-p}$ associated with ATV-LDEs(p).

Let us consider the generating sequence $\{\xi_{t,r}\}_{r\geq 1-p}$, that is

$$\{\xi_{t,1-p}, \xi_{t,2-p}, ..., \xi_{t,0}, ..., \xi_{t,k}, \xi_{t,k+1}, ...\},$$

along with the primary fundamental solution sequence $\{\xi_{t-k+r,r}\}_{r\geq 1-p}$ that is

$$\{\xi_{t-k+1-p,1-p}, \xi_{t-k+2-p,2-p}, ..., \xi_{t-k,0}, ..., \xi_{t-1,k-1}, \xi_{t,k}, \xi_{t+1,k+1}, ...\}.$$

For $r \leq 0$ and $r = k$ the values of the corresponding terms in the two sequences coincide. For example, $\xi_{t,0} = \xi_{t-k,0} = 1$. Moreover, $\xi_{t,k}$ belongs to both sequences, when applying the formulas for $r = k$. Let us now consider the sequences for the remaining values of r. The corresponding terms of the sequences $\xi_{t,r}$ and $\xi_{t-k+r,r}$ are given by $r \times r$ determinants, but, due to the time varying coefficients, the corresponding matrix entries are evaluated in different moments of time. For example, if $r = 1$, then $\xi_{t,1} = \phi_1(t)$. The corresponding term of the primary fundamental solution sequence is $\xi_{t-k+1,1} = \phi_1(t-k+1)$, whenever $k > 1$. As $k \neq 1$ the values $\phi_1(t)$ and $\phi_1(t-k+1)$ do not coincide in general. It turns out that in the time varying case any two different terms of the generating sequence, say ξ_{t,r_1} and ξ_{t,r_2} such that $r_1 > 0, r_2 > 0$ and $r_1 \neq k, r_2 \neq k$, do not, in general, belong to the same primary fundamental solution sequence. However, in the case of constant coefficients any two different terms of the generating sequence belong to the same primary fundamental solution sequence. For example $\xi_{t-k+1,1} = \phi_1(t-k+1) = \phi_1 = \phi_1(t) = \xi_{t,1}$.

A.4.3 Solution Expression in Terms of the Generating Sequence

In the following proposition we establish a solution in terms of the generating sequence $\xi_{t,k-r}$, that is

$$\xi_{t,k-r} = \det \begin{pmatrix} \phi_1(t-k+r+1) & -1 & 0 & ... & 0 \\ \phi_2(t-k+r+2) & \phi_1(t-k+r+2) & -1 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi_{k-r-1}(t-1) & \phi_{k-r-2}(t-1) & \phi_{k-r-3}(t-1) & ... & -1 \\ \phi_{k-r}(t) & \phi_{k-r-1}(t) & \phi_{k-r-2}(t) & ... & \phi_1(t) \end{pmatrix}, \tag{A.14}$$

for $r = 0, 1, .., k$. For example, if $r = k, k-1, k-2$, then

$$\xi_{t,k-k} = \xi_{t,0} = 1, \ \xi_{t,k-(k-1)} = \xi_{t,1} = \phi_1(t), \text{ and } \xi_{t,k-(k-2)} = \xi_{t,2} = \det \begin{pmatrix} \phi_1(t-1) & -1 \\ \phi_2(t) & \phi_1(t) \end{pmatrix}.$$

Proposition A2 *The general solution of eq.* (A.8) *in terms of the generating sequence* $\{\xi_{t,k-r}\}_{1\le r\le k}$ *is determined by*

$$y_{t,k} = \sum_{m=1}^{p}\sum_{r=1}^{k}\phi_{r-1+m}(t-k+r)y_{t-k+1-m}\xi_{t,k-r} + \sum_{r=1}^{k} v_{t-k+r}\xi_{t,k-r}. \tag{A.15}$$

Proof. Expanding $\det(\mathbf{Y}_{t,k})$ in eq. (A.10) along the first column we get

$$\begin{aligned} y_{t,k} &= [v_{\tau+1} + \sum_{m=1}^{p}\phi_m(\tau+1)y_{\tau+1-m}]\xi_{t,k-1} + ... + [v_t + \sum_{m=1}^{p}\phi_{k-1+m}(t)y_{\tau+1-m}]\xi_{t,0} \\ &= [v_{\tau+1}\xi_{t,k-1} + \sum_{m=1}^{p}\phi_m(\tau+1)y_{\tau+1-m}\xi_{t,k-1}] + ... + [v_t\xi_{t,0} + \sum_{m=1}^{p}\phi_{k-1+m}(t)y_{\tau+1-m}\xi_{t,0}] \\ &= [\sum_{m=1}^{p}\phi_m(\tau+1)y_{\tau+1-m}\xi_{t,k-1} + ... + \sum_{m=1}^{p}\phi_{k-1+m}(t)y_{\tau+1-m}\xi_{t,0}) + (v_{\tau+1}\xi_{t,k-1} + ... + v_t\xi_{t,0}], \end{aligned}$$

that is

$$y_{t,k} = \sum_{r=1}^{k}\sum_{m=1}^{p}\phi_{m+r-1}(t-k+r)y_{t-k+1-m}\xi_{t,k-r} + \sum_{r=1}^{k} v_{t-k+r}\xi_{t,k-r}. \tag{A.16}$$

By reversing the summation order in eq. (A.16) the expression (A.15) follows. ■

A.5 Homogeneous Solution

The homogeneous equation is derived by setting in eq. (A.8) $v_t = 0$:

$$y_{t,k} = \sum_{m=1}^{k-1}\phi_m(t)y_{t-m,k-m} + \sum_{m=1}^{p}\phi_{k-1+m}(t)y_{t-k+1-m}. \tag{A.17}$$

Proposition A3 *The general homogeneous solution of eq.* (A.17) *is determined by*

$$y_{t,k}^{hom} = \sum_{m=1}^{p}\sum_{r=1}^{k}\phi_{r-1+m}(t-k+r)y_{t-k+1-m}\xi_{t,k-r}. \tag{A.18}$$

Proof. Putting $v_{t-k+r} = 0$ for all $r = 1, 2, ...k$ in eq. (A.15), we obtain the solution expression (A.18) of eq. (A.17). ■

A.6 Fundamental Set of Solutions

For every $t \in \mathbb{Z}$ and $k \in \mathbb{Z}^+$, we define, as in Subsection 2.2, the m-th fundamental solution sequence $\{\xi^{(m)}_{t-k+r,r}\}_{r \geq 1-p}$ determined by

$$\xi^{(m)}_{t,k} = \det \begin{pmatrix} \phi_m(t-k+1) & -1 & 0 & \dots & 0 \\ \phi_{m+1}(t-k+2) & \phi_1(t-k+2) & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi_{m+k-2}(t-1) & \phi_{k-2}(t-1) & \phi_{k-3}(t-1) & \dots & -1 \\ \phi_{m+k-1}(t) & \phi_{k-1}(t) & \phi_{k-2}(t) & \dots & \phi_1(t) \end{pmatrix}. \tag{A.19}$$

The $\xi^{(m)}_{t,k}$ is obtained by putting in eq. (A.11) $v_{t-k+r} = 0$ for all r, $y_{t-k+1-m} = 1$ and $y_{t-k+1-r} = 0$ for $r \neq m$. That is $\{\xi^{(m)}_{t-k+r,r}\}_{r>0}$ is a solution of eq. (A.8) subject to the initial conditions $y_{t-k+1-m} = 1$ and $y_{t-k+1-r} = 0$ for $r \neq m$. In particular, if $m = 1$ we get $\xi^{(1)}_{t,k} = \xi_{t,k}$.

Proposition A4 *For every t, k all the fundamental solutions can be expressed in terms of the generating sequence $\xi_{t,k-r}$ as*

$$\xi^{(m)}_{t,k} = \sum_{r=1}^{k} \phi_{m-1+r}(t-k+r)\xi_{t,k-r}, \tag{A.20}$$

for all $m = 1, 2, ..., p$.

Proof. By expanding the determinant in eq. (A.19) along the first column the result follows. ■

Next, we extend the definition of $\xi^{(m)}_{t,k}$ to cover the initial condition values:

$$\xi^{(m)}_{t,r} = \begin{cases} \det(\Phi^{(m)}_{t,r}) & \text{If } r > 0 \\ 1 & \text{If } 1-p \leq r \leq 0 \text{ and } r = 1-m \\ 0 & \text{If } 1-p \leq r \leq 0 \text{ and } r \neq 1-m. \end{cases} \tag{A.21}$$

That is, the initial condition values are the values of $\xi^{(m)}_{t,r}$ for $r \leq 0$:

$$\xi^{(m)}_{t,1-m} = 1 \text{ for all } m, \text{ and } \xi^{(m)}_{t,0} = 0, \text{ whenever } m \neq 1.$$

Theorem A1 *For any arbitrary but fixed pair t, k the set of solution sequences*

$$\Xi_{t,k} = \{\{\xi^{(1)}_{t-k+r,r}\}_{r \geq 1-p}, \{\xi^{(2)}_{t-k+r,r}\}_{r \geq 1-p}, ..., \{\xi^{(p)}_{t-k+r,r}\}_{r \geq 1-p}\},$$

is a fundamental solution set associated with eq. (A.17).

Proof. Definition (A.21) entails that the matrix

$$\mathbf{B}_{t,k} = \begin{pmatrix} \xi^{(p)}_{t-k+1-p,1-p} & \xi^{(p-1)}_{t-k+1-p,1-p} & \cdots & \xi^{(1)}_{t-k+1-p,1-p} \\ \xi^{(p)}_{t-k+2-p,2-p} & \xi^{(p-1)}_{t-k+2-p,2-p} & \cdots & \xi^{(1)}_{t-k+2-p,2-p} \\ \vdots & \vdots & \vdots & \vdots \\ \xi^{(p)}_{t-k,0} & \xi^{(p-1)}_{t-k,0} & \cdots & \xi^{(1)}_{t-k,0} \end{pmatrix}$$

is the identity matrix of order p. Let $\lfloor x \rfloor$ stands for the integer part of a real number $x \in \mathbb{R}$. One needs $\lfloor \frac{p}{2} \rfloor$ column (or row) exchanges in $\mathbf{B}_{t,k}$ to derive the matrix

$$\mathbf{\Xi}_{t,k} = \begin{pmatrix} \xi^{(1)}_{t-k+1-p,1-p} & \xi^{(2)}_{t-k+1-p,1-p} & \cdots & \xi^{(p)}_{t-k+1-p,1-p} \\ \xi^{(1)}_{t-k+2-p,2-p} & \xi^{(2)}_{t-k+2-p,2-p} & \cdots & \xi^{(p)}_{t-k+2-p,2-p} \\ \vdots & \vdots & \vdots & \vdots \\ \xi^{(1)}_{t-k,0} & \xi^{(2)}_{t-k,0} & \cdots & \xi^{(p)}_{t-k,0} \end{pmatrix}$$

Taking into account that for every t, k the (first) Casorati $W_{t,k}(1-p)$ of the set $\Xi_{t,k}$ is $W_{t,k}(1-p) = \det(\mathbf{\Xi}_{t,k})$, we have:

$$W_{t,k}(1-p) = (-1)^{\lfloor \frac{p}{2} \rfloor} \det(\mathbf{B}_{t,k}) = (-1)^{\lfloor \frac{p}{2} \rfloor} \neq 0.$$

As a consequence, for every t, k the set $\Xi_{t,k}$ is linearly independent. Moreover, as the dimension of the homogeneous solution space of eq. (A.3) is p, the set $\Xi_{t,k}$ is a fundamental solution set associated with eq. (A.17). ■

As demonstrated in Subsection A.5, the solution (A.15) in Proposition A2 splits into homogeneous and particular solution parts. In the following theorem, we recover the above solution decomposition, in which the homogeneous solution part is expressed as a linear combination of the fundamental solutions with scalar coefficients the initial condition values. Moreover the particular solution part turns out to be the solution of eq. (A.8) with 0s as initial conditions.

Theorem A2 *The* (t, k) *term of the solution of eq.* (A.1) (*or eq.* (A.2)) *subject to the initial conditions* $y_{t-k+1-m}, m = 1, 2, ..., p,$ *can be expressed as a sum of homogeneous and particular solution parts*

$$y_{t,k} = y^{hom}_{t,k} + y^{par}_{t,k}, \tag{A.22}$$

such that each solution part is expressed as:

$$y_{t,k}^{hom} = \sum_{m=1}^{p} \xi_{t,k}^{(m)} y_{t-k+1-m}, \quad \text{(A.23)}$$

and

$$y_{t,k}^{par} = \sum_{r=0}^{k-1} \xi_{t,r}\varphi(t-r) + \sum_{r=0}^{k-1} \xi_{t,r} u_{t-r}. \quad \text{(A.24)}$$

Proof. We observe that $y_{t-k+1-m}$ in eq. (A.18) does not depend on r. By factoring we can write eq. (A.18) as

$$y_{t,k}^{hom} = \sum_{m=1}^{p} y_{t-k+1-m} \sum_{r=1}^{k} \phi_{m-1+r}(t-k+r)\xi_{t,k-r}. \quad \text{(A.25)}$$

By virtue of eq. (A.20), we can replace the sum $\sum_{r=1}^{k} \phi_{m-1+r}(t-k+r)\xi_{t,k-r}$ in eq. (A.25) by $\xi_{t,k}^{(m)}$ and the solution (A.25) turns into eq. (A.23), as required. By setting $y_{t-k+1-m} = 0$ for all $m = 1, 2, ..., p$ in eq. (A.15), and taking into account the definition of v_t along with the fact that

$$\sum_{r=1}^{k} v_{t-k+r}\xi_{t,k-r} = v_t\xi_{t,0} + v_{t-1}\xi_{t,1} + ... + v_{t-k+1}\xi_{t,k-1} = \sum_{r=0}^{k-1} v_{t-r}\xi_{t,r},$$

we obtain the particular solution:

$$y_{t,k}^{par} = \sum_{r=0}^{k-1} v_{t-r}\xi_{t,r} = \sum_{r=0}^{k-1} [\varphi(t-r) + u_{t-r}]\xi_{t,r} = \sum_{r=0}^{k-1} \xi_{t,r}\varphi(t-r) + \sum_{r=0}^{k-1} \xi_{t,r} u_{t-r}. \quad \text{(A.26)}$$

Finally, as eqs. (A.15) and (A.22) coincide the result follows. ■

Corollary A1 *Given* t, k*, the solution sequence of eq.* (A.1) *subject to the initial conditions* $y_{t-k+1-m}, m = 1, 2, ..., p$ *is given by*

$$[y_{t-k+r,r}]_{r\geq 1-p} = \sum_{m=1}^{p} y_{t-k+1-m}[\xi_{t-k+r,r}^{(m)}]_{r\geq 1-p} + [\sum_{i=0}^{r-1} \xi_{t-k+r,i} v_{t+r-i}]_{r\geq 1-p},$$

where we recall that [.] *stands for column vectors and sequences.*

B APPENDIX

Solution Forms and TV-HARMA Process Stability

In this section we apply the results established in Appendix A to prove statements and formulas introduced in Section 2 and in particular those presented in Theorems 1, 2 and Propositions 1, 2.

Let us consider the TV-LDEs of p-order:

$$y_t = \sum_{m=1}^{p} \phi_m(t) y_{t-m} + v_t. \tag{B.1}$$

Eq. (B.1) implies that for each fixed t the sequence $(\phi_m(t))_m$ is finite with $1 \leq m \leq p$. This allows us to define:

$$\phi_m(t) = 0 \text{ for all } m \text{ such that } m > p \text{ and for all } t. \tag{B.2}$$

All the formulas of Section 2 can be derived by applying the Assumption (B.2) to the formulas obtained in the previous section of this appendix. In particular eq. (B.1) is derived from eq. (A.2).

Proposition B1 *The statement in eq.* (B.2) *can be alternatively expressed as: If $m \in \mathbb{Z}$ such that $1 \leq m \leq p$, then $\phi_{m+r-1}(t-k+r) = 0$ for all $r \in \mathbb{Z}$ such that $r > p - m + 1$.*

Proof. It follows from eq. (B.2) that if $m + r - 1 > p$, then $\phi_{m+r-1}(t-k+r) = 0$. As

$$r > p - m + 1 \iff m + r - 1 > p,$$

the result follows. ■

We first establish the fundamental solutions formulas. The fundamental solution $\xi_{t,k}^{(m)}$ associated with eq. (B.1) is derived from the matrix $\mathbf{\Phi}_{t,k}$ in eq. (A.19) by assigning the zero values to those entries of $\mathbf{\Phi}_{t,k}$ defined by: $\phi_{p+r}(s) = 0$ for all $r \in \mathbb{Z}^+$ and all $s \in \mathbb{Z}$. This yields:

$$\xi_{t,k}^{(m)} = \begin{vmatrix} \phi_m(t-k+1) & -1 & & & & & \\ \phi_{m+1}(t-k+2) & \phi_1(t-k+2) & \ddots & & & & \\ \vdots & \vdots & \ddots & \ddots & & & \\ \phi_p(t-k+p+1-m) & \phi_{p-m}(t-k+p+1-m) & \ddots & \ddots & & & \\ & \vdots & \ddots & \ddots & \ddots & & \\ & \phi_p(t-k+p+1) & \ddots & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & & \phi_p(t-1) & \phi_{p-1}(t-1) & \cdots & -1 \\ & & & & \phi_p(t) & \cdots & \phi_1(t) \end{vmatrix}, \tag{B.3}$$

where $|\mathbf{\Phi}_{t,k}|$ stands for the determinant of $\mathbf{\Phi}_{t,k}$.

The fundamental solution set for TV-HARMA(p, q), as presented in Subsection 2.2 can be derived from the corresponding set $\Xi_{t,k}$ established in Theorem A1 using as $\xi_{t,k}^{(m)}$ the determinants in eq. (B.3).

The expression (7) is derived from Proposition B1 as follows. It follows from $k > p$ that $k > p \geq p - m + 1$ for all m such that $1 \leq m \leq p$. Therefore, we can write eq. (A.20) as

$$\xi_{t,k}^{(m)} = \sum_{r=1}^{k} \phi_{m-1+r}(t-k+r)\xi_{t,k-r} = \sum_{r=1}^{p-m+1} \phi_{m-1+r}(t-k+r)\xi_{t,k-r} + \sum_{r=p-m+2}^{k} \phi_{m-1+r}(t-k+r)\xi_{t,k-r}.$$

By Proposition B1 $\phi_{m+r-1}(t-k+r) = 0$ for all $r \geq p-m+2$, whence $\sum_{r=p-m+2}^{k} \phi_{m-1+r}(t-k+r)\xi_{t,k-r} = 0$, and the result follows.

The expression in (7) can alternatively be obtained by expanding the determinant in eq. (B.3) along the first column.

Proof of Theorem 1. As $p \geq p-m+1$ for $1 \leq m \leq p$, Proposition B1 entails that $\phi_{m+r-1}(t-k+r) = 0$ for all $r > p$. Therefore eq. (A.18) can also be written as

$$y_{t,k}^{hom} = \sum_{m=1}^{p} \sum_{r=1}^{p} \phi_{m-1+r}(t-k+r)\xi_{t,k-r} y_{t-k+1-m}, \tag{B.4}$$

and $\xi_{t,k}^{(m)}$ in eq. (7) as

$$\xi_{t,k}^{(m)} = \sum_{r=1}^{p} \phi_{m-1+r}(t-k+r)\xi_{t,k-r}.$$

By factoring eq. (B.4) and replacing the sum $\sum_{r=1}^{p} \phi_{m+r-1}(t-k+r)\xi_{t,k-r}$ by $\xi_{t,k}^{(m)}$, eq. (B.4) turns into

$$y_{t,k}^{hom} = \sum_{m=1}^{p} y_{t-k+1-m} \sum_{r=1}^{p} \phi_{m-1+r}(t-k+r)\xi_{t,k-r} = \sum_{m=1}^{p} \xi_{t,k}^{(m)} y_{t-k+1-m}. \tag{B.5}$$

This shows the homogeneous solution expression in (7). Using in place of $\xi_{t,k}$ the determinant expression in eq. (B.3) with $m = 1$, the particular solution, as obtained in eq. (A.24), coincides with eq. (8). Finally the sum of the homogeneous solution in eq. (B.5) plus the particular solution in eq. (A.24) yields the general solution as claimed in Theorem 1. ■

Next, we provide a proof for the stability of the DTV-HARMA process associated with non-stochastic coefficients, as presented in Theorem 2. In the case of stochastic coefficients the proof is analogous.

In all that follows we use the notation $\bar{\phi}_m = \sup_t |\phi_m(t)|$ for each m such that $1 \leq m \leq p$.

Lemma B1 *Let* $\lim_{k\to\infty} \xi_{t,k} = 0$ *for all* t. *Then for every* t

$$\lim_{k\to\infty} \xi_{t,k}^{(m)} = 0 \text{ for all } m \text{ such that } 1 \leq m \leq p, \tag{B.6}$$

provided that $\bar{\phi}_m < \infty$.

Proof. In view of eq. (7), on account of $\bar{\phi}_m < \infty$ we have:

$$|\xi_{t,k}^{(m)}| = |\sum_{r=1}^{p-m+1} \phi_{m-1+r}(t-k+r)\xi_{t,k-r}| \leq \sum_{r=1}^{p-m+1} |\phi_{m-1+r}(t-k+r)||\xi_{t,k-r}| \leq \sum_{r=1}^{p-m+1} \bar{\phi}_m|\xi_{t,k-r}| = \bar{\phi}_m \sum_{r=1}^{p-m+1} |\xi_{t,k-r}|.$$

Letting $k \to \infty$ we get:

$$\lim_{k\to\infty} |\xi_{t,k}^{(m)}| \leq \lim_{k\to\infty} \bar{\phi}_m \sum_{r=1}^{p-m+1} |\xi_{t,k-r}| = \bar{\phi}_m \sum_{r=1}^{p-m+1} \lim_{k\to\infty} |\xi_{t,k-r}| = 0.$$

Hence $\lim_{k\to\infty} \xi_{t,k}^{(m)} = 0$ for all t, m. ■

This proposition states that if the first fundamental solution converges to zero, as $k \to \infty$, then all the other fundamental solutions converge to zero, provided that the autoregressive coefficients are bounded sequences. We shall refer to the condition (B.6) as the stability condition.

Considering as initial conditions $y_{t-k+1-m} = c_m$ for $m = 1, 2, ..., p$, the homogeneous solution in eq. (9) must be written as

$$y_{t,k}^{hom} = \sum_{m=1}^{p} c_m \xi_{t,k}^{(m)}. \tag{B.7}$$

We observe that for each new k the evaluation of $y_{t,k}^{hom}$ yields, in the general case, a new forecasting-time outcome, while the values of the initial condition sequence remain the same. To see this, let us call $\mathbf{H}_{t,k}$ the homogeneous solution matrix, obtained by setting to $\mathbf{Y}_{t,k}$ in eq. (A.11): $v_r = 0$ for all r. Then $y_{t,k}^{hom} = \det(\mathbf{H}_{t,k})$. As k is getting larger the information time interval $\mathbb{I}_{t-k} = [t-k-p+1, t-k] \cap \mathbb{Z}$ successively departs from t, while the sequence of initial conditions $\{y_{t-k+1-m}\}_{1\leq m\leq p}$ remains the same, that is $y_{t-k+1-m} = c_m$ for all m. Employing similar arguments as in Subsection A.4.2, the time varying coefficients $\phi_{m+k-r}(t-r+1)$ in $\mathbf{H}_{t,k}$ change as k varies and so does $y_{t,k}^{hom}$, no matter if the initial condition values remain the same.

In view of eq. (B.7), the TV-HARMA process in eq. (1) is globally asymptotically stable (stable for short) if and only if $y_{t,k}^{hom} \to 0$, as $k \to \infty$ for any initial condition sequence of p arbitrary but fixed values $\{c_m\}_{1\leq m\leq p}$. In other words, as the initial condition sequence $\{c_m\}_{1\leq m\leq p}$ moves further to the past its effects on the solution $y_{t,k}^{hom}$ gradually die out. Next, we show the stability Theorem 2.

Proof of Theorem 2.

(Sufficient) In view of eq. (B.7), Lemma B1 along with the stability condition imply:

$$\lim_{k\to\infty} y_{t,k}^{hom} = \lim_{k\to\infty} \sum_{m=1}^{p} c_m\, \xi_{t,k}^{(m)} = \sum_{m=1}^{p} c_m \lim_{k\to\infty} \xi_{t,k}^{(m)} = \sum_{m=1}^{p} c_m \cdot 0 = 0,$$

which shows that $y_{t,k}^{hom}$ is stable, as required.

(Necessary) The formula $y_{t,k}^{hom} = \sum_{m=1}^{p} \xi_{t,k}^{(m)} c_m$ applied with $\{c_1 = 1, c_2 = 0, c_3 = 0, ..., c_p = 0\}$ yields $\xi_{t,k}^{(1)}$ ($\xi_{t,k}$ for short), that is $y_{t,k}^{hom} = \xi_{t,k}$ subject to the initial conditions $\{y_{t-k} = 1, y_{t-k-1} = 0, ..., y_{t-k-p} = 0\}$. This process, for various values of k, yields the generating sequence $\{\xi_{t,1}, \xi_{t,2}, ..., \xi_{t,p}, \xi_{t,p+1}, \xi_{t,p+2}, ...\}$ (or $\{\xi_{t,k}\}_{k\geq 1}$ for short). Taking into account that $\lim_{k\to\infty} y_{t,k}^{hom} = 0$ for any sequence of initial condition values and all t, it follows that $\lim_{k\to\infty} \xi_{t,k} = 0$ for each t, as required. ■

C APPENDIX

Second Moments (Non Stochastic Coefficients)

C.1 First Moments

Next we give a proof for Proposition 3 provided that $\sum_{r=0}^{\infty} |\xi_{t,r}| < \infty$ holds for each t.

Proof. Let us assume that $\tilde{\varphi} = \sup_t |\varphi(t)|$. In order to show that the first unconditional moment exists it suffices to show the absolute convergence of $\sum_{r=0}^{\infty} \xi_{t,r}\varphi(t-r)$. For every t we have:

$$\sum_{r=0}^{k} |\xi_{t,r}\varphi(t-r)| = \sum_{r=0}^{k} |\xi_{t,r}||\varphi(t-r)| \leq \sum_{r=0}^{k} |\xi_{t,r}|\tilde{\varphi} = \tilde{\varphi}\sum_{r=0}^{k} |\xi_{t,r}|, \text{ for all } k. \tag{C.1}$$

Let us call $N = \sum_{r=0}^{\infty} |\xi_{t,r}|$. Letting $k \to \infty$ in eq. (C.1) and taking into account that $\tilde{\varphi}, N \in^{+}$ we have

$$\sum_{r=0}^{\infty} |\xi_{t,r}\varphi(t-r)| \leq \tilde{\varphi}\sum_{r=0}^{\infty} |\xi_{t,r}| = \tilde{\varphi}N < \infty,$$

as required. ■

A proof of Proposition 4 is established below, provided that the sufficient condition $\sum_{r=0}^{\infty} \xi_{t,r}^2 < \infty$ holds for each t.

Proof. For every t, k (we recall that $\sigma_t^2 < M$) we have:

$$\sum_{r=0}^{k} \xi_{t,r}^2\sigma_t^2 < \sum_{r=0}^{k} \xi_{t,r}^2 M = M\sum_{r=0}^{k} \xi_{t,r}^2. \tag{C.2}$$

If follows from the assumption $\sum_{r=0}^{\infty} |\xi_{t,r}| < \infty$ that $\sum_{r=0}^{\infty} \xi_{t,r}^2 < \infty$. Let us call $N = \sum_{r=0}^{\infty} \xi_{t,r}^2$. Letting $k \to \infty$ in eq. (C.2) and taking into account that $M, N \in \mathbb{R}^{+}$, we have

$$\sum_{r=0}^{\infty} \xi_{t,r}^2\sigma_t^2 < M\sum_{r=0}^{\infty} \xi_{t,r}^2 = MN < \infty,$$

as required. ■

In what follows, we prove Theorem 3, that is the Wold-Crámer decomposition of the linear stochastic process $\{y_t\}$, provided that $\{y_t\}$ is a bounded sequence ($|y_t| \leq N$ for all t).

Proof. The homogeneous solution with initial condition values $\{y_{t-k+1-m}\}_{1\leq m\leq p}$ is given by:

$$y_{t,k}^{hom} = \sum_{m=1}^{p} \xi_{t,k}^{(m)} y_{t-k+1-m}. \tag{C.3}$$

Let us call $\tilde{y} = \sup_t |y_t|$. Then taking the limits to both sides of eq. (C.3) we show that $\lim_{k\to\infty} y_{t,k}^{hom} = 0$, that is equivalent to $\lim_{k\to\infty} |y_{t,k}^{hom}| = 0$. The result follows from:

$$\lim_{k\to\infty} |y_{t,k}^{hom}| = \lim_{k\to\infty} |\sum_{m=1}^{p} \xi_{t,k}^{(m)} y_{t-k+1-m}| \leq \lim_{k\to\infty} \sum_{m=1}^{p} |\xi_{t,k}^{(m)}|\tilde{y} = \sum_{m=1}^{p} \tilde{y} \lim_{k\to\infty} |\xi_{t,k}^{(m)}| = \tilde{y} \sum_{m=1}^{p} 0 = 0.$$

On account of eq. (A.24) along with the DTV-HAR(p) hypothesis, that is $u_t = \varepsilon_t$ for all t, we have

$$y_t = \lim_{k\to\infty} y_{t,k} = \lim_{k\to\infty} y_{t,k}^{hom} + \lim_{k\to\infty} y_{t,k}^{par} = \lim_{k\to\infty} y_{t,k}^{par} = \sum_{r=0}^{\infty} \xi_{t,r}\varphi(t-r) + \sum_{r=0}^{\infty} \xi_{t,r} u_{t-r} = \sum_{r=0}^{\infty} \xi_{t,r}\varphi(t-r) + \sum_{r=0}^{\infty} \xi_{t,r}\varepsilon_{t-r},$$

as claimed. ■

C.2 Second Moments and MA part

In order to analyse the covariance structure of the $\{y_t\}$ process, we introduce the solution pair $(y_{t,r+\ell}, y_{t-\ell,r})$. As the time lag $(t - (t-\ell) = \ell)$ between $y_{t,r+\ell}, y_{t-\ell,r}$ equals the corresponding period lag $(r + \ell - r = \ell)$, the following conditions are fulfilled:

a) Both components $y_{t,r+\ell}$ and $y_{t-\ell,r}$ of the pair belong to the same solution sequence for each distinct $\ell \in \mathbb{Z}^+$. In particular, both $y_{t,r+\ell}, y_{t-\ell,r}$ are terms of the solution sequence whose initial condition set is: $\{y_{t-r-\ell+1-m}\}_{1\leq m\leq p}$.

b) As ℓ increases so do (by the same increment) the time lag and the period lag between the components of the solution pair.

Inasmuch as τ varies in a linear fashion with ℓ, according to $\tau = t - r - \ell$, we conclude that:

c) The set of initial conditions moves to the past as ℓ increases.

Next, we give a proof to Proposition 6.

Proof. In the following equalities we use the statements:

i) Since $\{\varepsilon_t\}$ is a martingale difference, it follows that: $\varepsilon_{t-\ell-r_1}\varepsilon_{t-\ell-r_2} = 0$, whenever $r_1 \neq r_2$.

ii) The absolute summability $\sum_{r=0}^{\infty} |\xi_{t,r}| < \infty$ is a sufficient condition for switching expectation with infinite summation.

In view of Theorem 3 and recalling that $u_t = \varepsilon_t$ for all t, the one sided MA representations of y_t and $y_{t-\ell}$ are $y_t = \sum_{r=0}^{\infty} \xi_{t,r+\ell}\varepsilon_{t-r-\ell}$ and $y_{t-\ell} = \sum_{r=0}^{\infty} \xi_{t-\ell,r}\varepsilon_{t-r-\ell}$ for $\ell = 0, 1, 2, ...$ Substituting the expressions of $y_t, y_{t-\ell}$ in $(y_t, y_{t-\ell})$ we have:

$$\begin{aligned}
\gamma_t(\ell) &= \mathbb{E}\left((\sum_{r=0}^{\infty} \xi_{t,\ell+r}\varepsilon_{t-\ell-r})(\sum_{r=0}^{\infty} \xi_{t-\ell,r}\varepsilon_{t-\ell-r}) \right) \\
\text{statement(i)} &= \mathbb{E}(\sum_{r=0}^{\infty} \xi_{t,\ell+r}\xi_{t-\ell,r}\varepsilon^2_{t-\ell-r}) \\
\text{statement(ii)} &= \sum_{r=0}^{\infty} \xi_{t,\ell+r}\xi_{t-\ell,r}\sigma^2_{t-\ell-r},
\end{aligned}$$

as required. ■

TV-*HARMA*(p, q) *Model*

In this Appendix we see how the results in Section 3 are easily modified when there is a MA term.

On account of $u_t = \Theta_t(B)\varepsilon_t$ (see eq. (1)) the second summation in the particular solution part of eq. (9) can be written as:

$$\sum_{r=0}^{k-1} \xi_{t,r}u_{t-r} = \sum_{r=0}^{k-1} \xi^*_{t,r}\varepsilon_{t-r} + \sum_{r=k}^{k-1+q} \xi'_{t,r}\varepsilon_{t-r},$$

where

$$\begin{aligned}
\xi^*_{t,r} &= \xi_{t,r} + \sum_{l=1}^{q} \xi_{t,r-l}\theta_l(t-r+l), \\
\xi'_{t,r} &= \sum_{\ell=r-k+1}^{q} \xi_{t,r-\ell}\theta_\ell(t-r+\ell).
\end{aligned}$$

Clearly, if $q = 0$, that is we have an AR model, then $\xi^*_{t,r} = \xi_{t,r}$ and $\xi'_{t,r} = 0$.

To obtain the k-step ahead predictor for the DTV-HARMA(p, q) model in the right hand side of eq. (10) we have to add $\sum_{r=k}^{k-1+q}\xi'_{t,r}\varepsilon_{t-r}$. The associated forecast error and its MSE in Proposition 2 become

$$\mathbb{FE}(y_t\,|\mathcal{F}_{t-k}) = \sum_{r=0}^{k-1} \xi^*_{t,r}\varepsilon_{t-r}, \quad \mathbb{V}ar[\mathbb{FE}(y_t\,|\mathcal{F}_{t-k})] = \sum_{r=0}^{k-1} (\xi^*_{t,r})^2\sigma^2_{t-r}.$$

Next, we assume that $\sup_t |\theta_l(t)| < \infty$ for all l. The Wold-Crámer decomposition in Theorem 3 is modified as follows:

$$y_t \overset{L_2}{=} \sum_{r=0}^{\infty}[\xi_{t,r}\varphi(t-r) + \xi^*_{t,r}\varepsilon_{t-r}].$$

Similarly, the ℓ order autocovariance function is obtained from eq. (11) in Theorem 5 by replacing the two ξ with the two ξ^*, that is

$$\gamma_t(\ell) = \sum\nolimits_{r=0}^{\infty} \xi^*_{t,\ell+r}\xi^*_{t-\ell,r}\sigma^2_{t-\ell-r}.$$

D APPENDIX

In this appendix we present the proofs of Theorems 4 and 5.

We will prove Theorem 4 by induction with respect to k.

Proof. (**of Theorem** 4). Clearly, it holds for $k + r = 1$: In eq. (19) setting $k = 1$, and thus $r = k_1 = k_2 = 0$ we obtain eq. (17) since $\mathbf{E}(\mathbf{y}_\tau | \mathcal{F}_{\tau-1}) = \varphi + \mathbf{\Phi}\mathbf{y}_{\tau-1} + \mathbf{Z}\varepsilon_{\tau-1}$ and $\mathbf{FE}(\mathbf{y}_\tau | \mathcal{F}_{\tau-k}) = \mathbf{J}\varepsilon_\tau$. Next if we assume that it holds for k, then it will suffice to prove that it also holds for $k+1$. First, rewrite eq. (17) as of time $\tau - k$:

$$\mathbf{y}_{\tau-k} = \varphi(\tau-k) + \mathbf{\Phi}(\tau-k)\mathbf{y}_{\tau-(k+1)} + \mathbf{J}\varepsilon_{\tau-k} + \mathbf{Z}(\tau-k)\varepsilon_{\tau-(k+1)}.$$

Substituting the above equation into eq. (19) using straightforward algebra shows that

$$\mathbf{y}_{\tau,k+1} = \mathbf{E}(\mathbf{y}_\tau \left| \mathcal{F}_{\tau-(k+1)} \right.) + \mathbf{FE}(\mathbf{y}_\tau \left| \mathcal{F}_{\tau-(k+1)} \right.)$$

as claimed. ■

Proof. (**of Theorem** 5). Theorem 4 implies that

$$\mathbf{\Gamma}_\tau = \mathbf{V}ar[\mathbf{E}(\mathbf{y}_\tau | \mathcal{F}_{\tau-k})] + \mathbf{V}ar[\mathbf{FE}(\mathbf{y}_\tau | \mathcal{F}_{\tau-k})]. \tag{D.1}$$

The variance of the optimal predictor is given by

$$\mathbf{V}ar[\mathbf{E}(\mathbf{y}_\tau | \mathcal{F}_{\tau-k})] = \mathbf{\Phi}_1^{k_1+r}\mathbf{\Phi}_2^{k_2-k_1}\mathbf{\Phi}^{k-k_2-1}\mathbf{V}ar(\mathbf{\Phi}\mathbf{y}_{\tau-k} + \mathbf{Z}\varepsilon_{\tau-k})(\mathbf{\Phi}_1^{k_1+r}\mathbf{\Phi}_2^{k_2-k_1}\mathbf{\Phi}^{k-k_2-1})', \tag{D.2}$$

where, under Assumption 1,

$$\mathbf{V}ar(\mathbf{\Phi}\mathbf{y}_{\tau-k} + \mathbf{Z}\varepsilon_{\tau-k}) = \sum_{\ell=0}^{\infty} \mathbf{\Phi}^\ell(\mathbf{\Phi}\mathbf{J} + \mathbf{Z})\mathbf{\Sigma}[\mathbf{\Phi}^\ell(\mathbf{\Phi}\mathbf{J} + \mathbf{Z})]'.$$

Similarly, the variance of the forecast error is given by

$$\mathbf{V}ar[\mathbf{FE}(\mathbf{y}_\tau\,|\mathcal{F}_{\tau-k})] \;=\; \mathbf{J\Sigma J}' + \sum_{\ell=1}^{k_1+r} \mathbf{\Phi}_1^{\ell-1}\mathbf{G}_1\mathbf{\Sigma}(\mathbf{\Phi}_1^{\ell-1}\mathbf{G}_1)' \tag{D.3}$$

$$+\,\mathbf{\Phi}_1^{k_1+r}\sum_{\ell=1}^{k_2-k_1}\mathbf{\Phi}_2^{\ell-1}\mathbf{G}_2\mathbf{\Sigma}(\mathbf{\Phi_1^{k_1+r}}\mathbf{\Phi}_2^{\ell-1}\mathbf{G}_2)' + \mathbf{\Phi}_1^{k_1+r}\mathbf{\Phi}_2^{k_2-k_1}\sum_{\ell=1}^{k-k_2-1}\mathbf{\Phi}^{\ell-1}\mathbf{G}_3\mathbf{\Sigma}(\mathbf{\Phi_1^{k_1+r}}\mathbf{\Phi_2^{k_2-k_1}}\mathbf{\Phi}^{\ell-1}\mathbf{G}_3)',$$

where now

$$\mathbf{G}_n = (\mathbf{\Phi}_n\mathbf{J} + \mathbf{Z}_n),\;\; n = 1,2,3.$$

Since $(\mathbf{XY})^{\otimes 2} = \mathbf{X}^{\otimes 2}\mathbf{Y}^{\otimes 2}$ and $(\mathbf{X}^r)^{\otimes 2} = (\mathbf{X}^{\otimes 2})^r$, using

$$\sum_{\ell=1}^{k_n-k_{n-1}}[(\mathbf{\Phi}_n^{\ell-1})(\mathbf{\Phi}_n\mathbf{J}+\mathbf{Z}_n)]^{\otimes 2} \;=\; \sum_{\ell=1}^{k_n-k_{n-1}}[(\mathbf{\Phi}_n)^{\otimes 2}]^{\ell-1}(\mathbf{\Phi}_n\mathbf{J}+\mathbf{Z}_n)^{\otimes 2}$$
$$= [\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_n^{k_n-k_{n-1}})^{\otimes 2}][\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_n)^{\otimes 2}]^{-1}(\mathbf{\Phi}_n\mathbf{J}+\mathbf{Z}_n)^{\otimes 2},\;\; n = 1,2,3$$

(we recall that $k_0 = -r$ and $k_3 = k$) the vec form of the right hand side of eq. (D.3) gives

$$\begin{aligned}
&\{\mathbf{J}^{\otimes 2} + [\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_1^{k_1+r})^{\otimes 2}][\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_1)^{\otimes 2}]^{-1}(\mathbf{\Phi}_1\mathbf{J}+\mathbf{Z}_1)^{\otimes 2}\\
&+ (\mathbf{\Phi}_1^{k_1+r})^{\otimes 2}[\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_2^{k_2-k_1})^{\otimes 2}][\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}_2)^{\otimes 2}]^{-1}(\mathbf{\Phi}_2\mathbf{J}+\mathbf{Z}_2)^{\otimes 2}\\
&+ (\mathbf{\Phi}_1^{k_1+r})^{\otimes 2}(\mathbf{\Phi}_2^{k_2-k_1})^{\otimes 2}[\mathbf{I}^{\otimes 2} - (\mathbf{\Phi}^{k-k_2-1})^{\otimes 2}](\mathbf{I}^{\otimes 2} - \mathbf{\Phi}^{\otimes 2})^{-1}(\mathbf{\Phi J}+\mathbf{Z})^{\otimes 2}\}\mathbf{s}.
\end{aligned} \tag{D.4}$$

Similarly, the vec form of the right hand side of eq. (D.2) is given by

$$(\mathbf{\Phi}_1^{k_1+r})^{\otimes 2}(\mathbf{\Phi}_2^{k_2-k_1})^{\otimes 2}(\mathbf{\Phi}^{k-k_2-1})^{\otimes 2}(\mathbf{I}^{\otimes 2} - \mathbf{\Phi}^{\otimes 2})^{-1}(\mathbf{\Phi J}+\mathbf{Z})^{\otimes 2}\mathbf{s}. \tag{D.5}$$

Finally, taking the vec form of both sides of eq. (D.1), and using eqs. (D.4) and (D.5), we obtain eq. (21) as claimed. ■

Table A.1. 'First-order persistence' for each of the three periods and models.

	$P(y_t \mid \varepsilon)$			$P(y_t \mid v)$			$P(y_t \mid \varepsilon) + P(y_t \mid v)$		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
1947Q1-1966Q4	1.68	2.20	1.60	1.34	1.610	2.13	3.02	3.81	3.74
1967Q1-2008Q2	3.61	2.20	2.91	2.87	3.89	3.88	6.48	6.09	6.79
2008Q3-2016Q3	1.85	2.20	2.91	1.47	2.49	2.78	3.33	4.69	5.70

Note: We used the limit as $k \to \infty$ of eq. (20) with unit mean and volatility shocks, that is $\overline{\mathbf{\Phi}}_n \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

with $\overline{\mathbf{\Phi}}_n = \mathbf{J} + (\mathbf{I} - \mathbf{\Phi}_n)^{-1}(\mathbf{\Phi}_n \mathbf{J} + \mathbf{Z}_n)$, $n = 1, 2, 3,$ to calculate the corresponding first-order persistence.

For each model and period $P(y_t \mid \varepsilon)$ and $P(y_t \mid v)$ are the $(1, 1)$ and $(1, 2)$ elements of $\overline{\mathbf{\Phi}}_n$.