Econometric Theory, **0**, 2025, 1–54.
doi:10.1017/S0266466624000306

A UNIFIED THEORY FOR ARMA MODELS WITH VARYING COEFFICIENTS: ONE SOLUTION FITS ALL

Menelaos Karanasos
Department of Economics and Finance, Brunel University London

Alexandros G. Paraskevopoulos
University of Piraeus

Tassos Magdalinos
University of Southampton

Alessandra Canepa
University of Turin

A new explicit solution representation is provided for ARMA recursions with drift and either deterministically or stochastically varying coefficients. It is expressed in terms of the determinants of banded Hessenberg matrices and, as such, is an explicit function of the coefficients. In addition to computational efficiency, the proposed solution provides a more explicit analysis of the fundamental properties of such processes, including their Wold–Cramér decomposition, their covariance structure, and their asymptotic stability and efficiency. Explicit formulae for optimal linear forecasts based either on finite or infinite sequences of past observations are provided. The practical significance of the theoretical results in this work is illustrated with an application to U.S. inflation data. The main finding is that inflation persistence increased after 1976, whereas from 1986 onward, the persistence declines and stabilizes to even lower levels than the pre-1976 period.

The first two author names are in alphabetical order. We thank the Editor, the Co-Editor, and two anonymous referees for their thoughtful comments and constructive suggestions. We gratefully acknowledge the helpful conversations we had with L. Giraitis, and G. Kapetanios in the preparation of the paper. We have also benefited from the comments given by participants (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) at various seminars and conferences. We would like to thank S. Dafnos, who contributed to the interpretation and synthesis in the initial version of the paper, and P. Koutroumpis, who provided the empirical section in an earlier version. Magdalinos gratefully acknowledges financial support by the British Academy: grant SRG2324\241667. Alessandra Canepa acknowledges financial support under the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.1, Call for tender No. 104 published on February 2, 2022 by the Italian Ministry of University and Research (MUR), funded by the European Union—NextGenerationEU—Project Title 20223725WE—Methodological and computational issues in large-scale time series models for economics and finance – CUP J53D23003960006—Grant Assignment Decree No. 967 adopted on June 30, 2023 by the Italian Ministry of Ministry of University and Research (MUR). Address correspondence to Tassos Magdalinos, University of Southampton, Southampton, UK; e-mail: a.magdalinos@soton.ac.uk.

1. INTRODUCTION

Modeling time series processes with variable coefficients has received considerable attention in recent years in the wake of several financial crises and high volatility due to frequent abrupt changes in the market. Justification for the use of such structures in econometric modeling can be found in Cavaliere and Taylor (2005), Cavaliere and Georgiev (2008), Giraitis et al. (2014), Harvey et al. (2018), and Chambers and Taylor (2020). Models with time-varying coefficients (hereafter time-varying models) are extensively applied by practitioners, and their importance is widely recognized (see, e.g., Petrova, 2019). Crucial advances in both the theory and the empirics for these structures are the works by Whittle (1965), Abdrabbo and Priestley (1967), Rao (1970), Hallin (1979, 1986), Singh and Peiris (1987), Kowalski and Szynal (1991), and Grillenzoni (1993, 2000).

This paper provides a general framework for the study of nonstationary autoregressive moving average models with time-varying coefficients and drift as well as heteroscedastic errors (hereafter TV-ARMA; see eq. (1)). It comprises explicit solution representations along with analogous representations for the fundamental properties of these models, whereas the scope of the useful tool of characteristic polynomial representations is diminished when variable coefficients are present (see, for details, Hallin, 1986, and Grillenzoni, 1990). There are two large classes of stochastic models: the ones with deterministically and those with stochastically varying coefficients. Both have been widely applied in many fields of research, such as economics, finance, and engineering, but traditionally they have been examined separately. The new framework unifies the study of these models, as the title of the paper suggests.

For the standard ARMA(p, q) models with constant parameters, it is well known that the coefficients in the Wold representation can be expressed either in terms of the roots of the autoregressive (AR) polynomial or, equivalently, as determinants of banded Toeplitz–Hessenberg matrices[1] in which the elements of each nonzero diagonal are the constant AR coefficients (see the matrix in eq. (8)).[2] In the case of time-varying models, the coefficients in a generalization of the Wold decomposition, referred to as the Wold–Cramér decomposition,[3] can be expressed in terms of domain restrictions of the so-called one-sided Green function (see Supplementary Appendix E2), referred to in the rest of the paper simply as *the Green function*.[4] Following Miller (1968), in a series of papers, Hallin (1979, 1984,

[1] In the open literature, the Hessenbergian attribute of such matrices does not usually appear.

[2] Toeplitz matrices are square matrices in which all elements along a diagonal have the same value. They are banded because above the superdiagonal and below the pth subdiagonal the elements of the matrix are zero. Moreover, the elements of the superdiagonal are (-1)s.

[3] Cramér's generalization to the Wold decomposition is also referred to as the Wold–Cramér representation.

[4] A detailed account of the one-sided Green function and its restriction is presented in Paraskevopoulos and Karanasos, 2021. For a discussion on the alternative wording regarding the "Green function" or "Green's function," see Wright (2006).

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

1986), Singh and Peiris (1987), and Kowalski and Szynal (1991) employ the Green function to describe the fundamental properties of models with deterministically varying coefficients and zero mean. However, the lack of an explicit representation of the Green function led to recursive methods for its computation (see also Grillenzoni, 2000; Azrak and Mélard, 2006).

The standard banded Toeplitz matrix formulation of ARMA models with constant coefficients is replaced here by banded Hessenberg matrices and their determinants (called banded Hessenbergians) will play a key role in the new closed-form solutions provided by this paper (see, for comparison, eqs. (4) and (8); for details on Hessenbergians and their properties, see, e.g., the book of Vein and Dale, 1999 and the references cited therein). More specifically, a banded Hessenberg matrix in which the nonzero diagonals are occupied by the time-varying AR coefficients, evaluated at consecutive time instances, is called *principal matrix* (see eq. (4)). Its determinant is called *principal determinant*. This is a fundamental solution of the difference equation associated with a TV-ARMA model, as highlighted by the title phrase "one solution fits all," which explicitly represents the abovementioned domain restriction of the Green function (see Paraskevopoulos and Karanasos, 2021). A method that constructs the principal determinant is grounded on prior research established by Paraskevopoulos (2012, 2014). This method applied to infinite systems associated with linear difference equations and the linear computational complexity of banded Hessenbergians are discussed in Supplementary Appendix F1.

The main contribution of this work is a novel explicit solution representation of time-varying process recursion in terms of the principal determinant (see eq. (16)). This is a fruitful result of using Hessenberg matrix determinant properties. To the best of our knowledge, no prior studies have established fully explicit solution representations of TV-ARMA models and of their fundamental properties due to the lack of analogous representations of the Green function. Some consequences are discussed in what follows.

Under an absolute summability assumption of the principal determinant function (condition (17)), our new representation of the solution yields a variety of important results summarized below. Condition (17) along with the boundedness of the drift and of the moving average (MA) coefficients guarantee the existence of a unique asymptotically stable second order solution process, that is, the Wold–Cramér decomposition of the model in L_2 (see Theorem 2). This result is generalized by Theorem 3 showing the existence and uniqueness of second-order solution processes without invoking the boundedness of the drift, but provided that a p-dimensional nonzero first moment vector has been estimated by the time series data. In both formulations, the demeaned processes (purely nondeterministic) have the same Wold–Cramér representation.

Explicit forms for the first two unconditional moments along with sufficient and necessary conditions for their existence are obtained in Propositions 2 and 3. Singh and Peiris (1987), Kowalski and Szynal (1991), and Grillenzoni (2000) provided regularity conditions ensuring that various mean-zero processes with deterministically varying coefficients are second-order.[5] Replacing the Green function involved in the formulas established in the above cited references with the principal determinant, these formulas turn into fully explicit expressions. A potential application of the Wold–Cramér decomposition to asymptotic theory, based on a decomposable Bernoulli shift, is discussed in Remark 1 of Section 4.1.

The solution representations of the model obtained in Theorems 1–3 yield explicit optimal linear predictor formulas along with analogous forecasting and associated mean square errors, when either infinite or finite sequences of data are observed (see Section 5). The invertibility of a second-order solution process is guaranteed by the absolute summability condition, which is also explicitly expressed by a banded Hessenbergian, but now its entries are occupied by the MA coefficients of the model (see eq. (G.2) in Supplementary Appendix G1). We illustrate formally one of the focal points in Hallin's (1986) analysis concerning the asymptotic efficiency of such models. Namely, that in a time-varying setting, two forecasts with identical forecasting horizons, but at different times, yield different mean squared errors.

Various processes with stochastically varying coefficients are treated within our unified framework in Section 6. Another goal of this work is to show that, in the case of AR models with random coefficients, when the principal determinant function converges to zero almost surely, then the process converges in distribution (see Theorem 4). Furthermore, the double stochastic autoregressive model is also employed to formulate some of its fundamental properties.

The paper concludes with an empirical application to inflation persistence in the United States over the time period from the last quarter of 1963 up to the beginning of 2018 (see Section 7), which employs a time-varying model of inflation dynamics grounded in statistical theory. In particular, we estimate an AR process with abrupt structural breaks and we compute an alternative measure of second-order time-dependent persistence, which distinguishes between changes in the dynamics of inflation and its volatility as well as their persistence. Our main conclusion is that persistence increased after 1976, whereas from 1986 onward, it declines and stabilizes to even lower levels than the pre-1976 period. Our results are in line with those in Cogley and Sargent (2002), who find that the persistence of inflation in the United States rose in the 1970s and remained high during this decade, before starting a gradual decline from the 1980s until the early 2000s.

The outline of the paper is as follows. Section 2 introduces the notation and the main assumptions used throughout the paper, followed by the principal determinant. The next section presents the explicit representation for an extensive family of time-varying ARMA models, based on the general solution of the associated linear difference equation. Section 4 presents explicit formulas for the second-order properties of the model, including the Wold–Cramér decomposition, the first two unconditional moments and the autocovariance function. In the next section, explicit optimal linear forecasts are established. Section 6 is concerned

[5]That is, although they are nonstationary, their second moments exist being finite and time-varying.

with AR models in which the coefficients are stochastically varying. The next section presents an empirical study for inflation persistence. The final section of the paper contains some concluding remarks and future work.

Throughout the paper, the proofs of the statements, formulated in the main body, are included in the Appendices and a few of them are deferred to the Supplementary Material.

2. TIME-VARYING ARMA

The main notation and assumptions associated with TV-ARMA processes are presented in Section 2.1. In Section 2.2, the principal determinant is introduced.

2.1. The Model

Throughout the paper, we adopt the following notation: $\mathbb{Z}$ will stand for the set of integers and $\mathbb{Z}_a$ for its subset defined by $\mathbb{Z}_a = \{z \in \mathbb{Z} : z \geq a\}$ for $a \in \mathbb{Z}$. The set of real numbers (resp. positive and nonnegative real numbers) is denoted by $\mathbb{R}$ (resp. $\mathbb{R}_{>0}$ and $\mathbb{R}_{\geq 0}$). Moreover, $(\Omega, \mathcal{F}, P)$ stands for a probability space and $L_2(\Omega, \mathcal{F}, P)$ (in short L_2) stands for the Hilbert space of real random variables with finite first two moments.

Let $p \in \mathbb{Z}_1$ and $q \in \mathbb{Z}_1$. A TV-ARMA(p,q) model is a stochastic process $\{y_t\}$ satisfying, for each $t \in \mathbb{Z}$,

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

with moving average term u_t given by

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

where $\varphi(t)$, $\phi_m(t)$, and $\theta_l(t)$ are deterministic (real-valued functions) or stochastic with $\phi_p(t)\theta_q(t) \neq 0$ for all $t \in \mathbb{Z}$, $\{\varepsilon_t\}$ is a mean-zero random process ($\mathbb{E}(\varepsilon_t) = 0$ for all t) such that $\mathbb{E}(\varepsilon_t \varepsilon_s) = 0$ for $s \neq t$ (uncorrelatedness condition) with time-varying variance $\mathbb{E}(\varepsilon_t^2) = \sigma^2(t)$ such that $0 < \sigma^2(t) \leq M < \infty$ for all t. The above conditions ensure that $\varepsilon_t \in L_2$ and $\varepsilon_t, \varepsilon_s$ are orthogonal, whenever $s \neq t$.

A solution $\{y_t\}$ to the recursive process (1) requires p initial condition values. These are considered as realizations of the prescribed random variables $y_{s+1-p}, \dots, y_s$ for any $s < t$, which are taken from the time series data in the recent or remote past. An explicit representation of y_t is expressed directly in terms of $y_{s+1-p}, \dots, y_s$ for any $s < t$, as established in eq. (16).

If the AR coefficients, $\phi_m(t)$, the MA coefficients, $\theta_l(t)$, and the drift, $\varphi(t)$, are deterministic (resp. stochastic), we shall refer to eq. (1) as DTV-ARMA

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

(resp. STV-ARMA: Further assumptions for these processes are considered in Section 6).[6]

The forcing term υ_t is assigned to be the time-varying drift plus the MA term:

$$\upsilon_t \stackrel{\text{def}}{=} \varphi(t) + u_t. \qquad \textbf{(2)}$$

The associated linear difference equation of eq. (1) (briefly TV-LDEs(p)) is defined by

$$y_t = \sum_{m=1}^{p} \phi_m(t) y_{t-m} + \upsilon_t, \text{ for all } t \in \mathbb{Z}. \qquad \textbf{(3)}$$

The general solution of eq. (3) is decomposed into two parts: the general homogeneous solution (see eq. (10)) plus the particular solution (see eq. (12)).

In this work, both assumptions of stationarity and homoscedasticity have been relaxed (see also, among others, Singh and Peiris, 1987; Kowalski and Szynal, 1991; Azrak and Mélard, 2006), which is likely to be violated in practice and we allow $\{\varepsilon_t\}$ to follow, for example, a stochastic volatility or a time-varying GARCH type of process (see, e.g., Karanasos et al., 2014b; Canepa et al., 2022) or we allow for abrupt structural breaks in the variance of ε_t (see Section 7). In the available open literature, the theoretical properties of time-varying processes are demonstrated within the framework of mean-zero processes. In this study, the presence of a nonzero drift is essential to ensure the existence of nonzero time-varying first-order moments.

The general model nests both the AR one as a special case when $q = 0$ and the specification in which the drift, the AR and MA coefficients, and the variances are all constants, adopting the conventional identifications for this purpose: $\varphi(t) = \varphi,\ \phi_m(t) = \phi_m,\ \theta_l(t) = \theta_l, \sigma^2(t) = \sigma^2$ for all t.

The relation between the process under consideration and its innovations is essentially described by the Wold–Cramér decomposition (see Section 4.1), which is the main analytical tool for studying the asymptotic efficiency of the model (see Karanasos et al., 2020). In this case, the latest time point of the observable random variables, denoted here by s, moves to the remote past ($s \to -\infty$), while the forecast time point, denoted here by t, is kept fixed.[7]

The product of companion matrices is commonly used (see, e.g, Kowalski and Szynal, 1991) to derive the Green function associated with eq. (1), but without providing an explicit form for the entries of the matrix product. Paraskevopoulos and Karanasos (2021) capitalized on the connection between the product of

[6]Notice that in our setting, the time-varying coefficients can depend on the length of the series as well, as in Azrak and Mélard (2006); see also Examples D.2 and D.3 in Supplementary Appendix D1.

[7]The forward asymptotic efficiency of the model (so-called by Hallin, 1986; Granger-Andersen) is strongly related to the forecasting problem. It directs attention to the asymptotic properties of the mean square forecasting error (MSE for short), as the time t moves to the far future, while s is kept fixed. Due to space considerations, results on the forward asymptotic efficiency of TV-ARMA(p, q) models are not presented here, but are available in Karanasos et al. (2020, Sect. 6.3).

companion matrices and time-varying stochastic difference equations but in the opposite direction. That is, they went the other way around, and by finding an explicit and compact representation of the fundamental solutions associated with TV-ARMA models, they obtained an analogous representation for the elements of the associated companion matrix product. Some consequences are presented in Supplementary Appendix F3 of this paper.

2.2. The Principal Determinant

To distinguish scalars from vectors, we adopt lower and uppercase boldface symbols within square brackets for column vectors and matrices, respectively: $\mathbf{x} = [x_i]$, $\mathbf{X} = [x_{ij}]$. Row vectors are indicated within round brackets and usually appear as transpositions of column vectors: $\mathbf{x}' = (x_i)$. For every pair $(t,s) \in \mathbb{Z}^2$ such that $s < t$, the *principal matrix* associated with eq. (1) is defined by

$$\boldsymbol{\Phi}_{t,s} \overset{\text{def}}{=} \begin{bmatrix} \phi_1(s+1) & -1 & & & & & & \\ \phi_2(s+2) & \phi_1(s+2) & \ddots & & & & & \\ \vdots & \vdots & \ddots & \ddots & & & & \\ \phi_p(s+p) & \phi_{p-1}(s+p) & \ddots & \ddots & \ddots & & & \\ & \phi_p(s+p+1) & \ddots & \ddots & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & \phi_p(t-1) & \phi_{p-1}(t-1) & \cdots & \phi_1(t-1) & -1 \\ & & & & \phi_p(t) & \cdots & \phi_2(t) & \phi_1(t) \end{bmatrix}. \tag{4}$$

Here and in what follows, empty spaces in a matrix have to be replaced by zeros. $\boldsymbol{\Phi}_{t,s}$ is a lower Hessenberg matrix of order $k = t - s$. It is also a banded 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 its nonzero superdiagonals), and lower bandwidth $p-1$ (the number of its nonzero subdiagonals). In particular, the elements of $\boldsymbol{\Phi}_{t,s}$ are: (-1) occupying the entries of the superdiagonal, the values of the first AR coefficient $\phi_1(\cdot)$ (from time $s+1$ to time t), occupying the entries of the main diagonal, the values of the $(r+1)$th AR coefficient $\phi_{r+1}(\cdot)$ for $r = 1, 2, \ldots, p-1$ (from time $s+1+r$ to time t), occupying the entries of the rth subdiagonal, and zero entries elsewhere. If the order k of $\boldsymbol{\Phi}_{t,s}$ is less than or equal to p, that is, $k \le p$, then $\boldsymbol{\Phi}_{t,s}$ is a full lower Hessenberg matrix.

For every pair $(t,s) \in \mathbb{Z}^2$ with $s < t$, the *principal determinant* associated with eq. (4) is defined by

$$\xi(t,s) \overset{\text{def}}{=} \det(\boldsymbol{\Phi}_{t,s}). \tag{5}$$

Formally $\xi(t,s)$ is a banded Hessenbergian (determinant of a lower Hessenberg matrix; for recent developments on Hessenbergians, see Jeerawat and Daowsud, 2022 and the references cited therein).

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

We further extend the definition of $\xi(t,s)$ on the entire $\mathbb{Z}^2$ by assigning the additional values:

$$\xi(t,s) \overset{\text{def}}{=} \begin{cases} 1, & \text{if} \quad t = s, \\ 0, & \text{if} \quad t < s. \end{cases} \tag{6}$$

Eq. (6) indicates the initial conditions located before and including the time point s. The difference $k = t - s \in \mathbb{Z}_1$ stands for the *forecasting horizon*. In Proposition A3 (see Appendix A.2), it is shown that $\xi(t,s)$ is the solution of the homogeneous linear difference equation associated with eq. (3), that is,

$$y_t = \sum_{m=1}^{p} \phi_m(t) y_{t-m}, \tag{7}$$

taking on the initial values $y_{s+1-p} = 0, \ldots, y_{s-1} = 0, y_s = 1$. As the initial value problem solution is unique, $\xi(t,r)$ must coincide with the Green function restriction for $t \in \mathbb{Z}_{s+1-p}$ and $s \leq r \leq t-1+p$ (see Supplementary Appendix E2). Moreover, $\xi(t,r)$ is the first fundamental solution (see Section 3.1), which determines all the others (see eq. (9)).

Example 1. The AR polynomial $\Phi(B) = 1 - \sum_{m=1}^{p} \phi_m B^m$ associated with eq. (7), whenever $\phi_m(t) = \phi_m$ (constant AR coefficients), is explicitly expressed in terms of the characteristic values as $\Phi(B) = \prod_{m=1}^{p}(1 - \lambda_m B)$. In this case, the banded Hessenbergian $\xi(t,s)$ turns into a banded Toeplitz–Hessenberg matrix determinant (for details on Toeplitz matrices, see, e.g., the book by Gray, 2006)[8] and satisfies the identity (usually called Widom's determinant formula; see Widom, 1958), that is,

$$\xi(k) = \begin{vmatrix} \phi_1 & -1 & & & & & & & \\ \phi_2 & \phi_1 & -1 & & & & & & \\ \phi_3 & \phi_2 & \phi_1 & \ddots & & & & & \\ \vdots & \vdots & \vdots & \ddots & \ddots & & & & \\ \phi_p & \phi_{p-1} & \phi_{p-2} & \ddots & \ddots & \ddots & & & \\ & \phi_p & \phi_{p-1} & \ddots & \ddots & \ddots & \ddots & & \\ & & \phi_p & \phi_{p-1} & \phi_{p-2} & \cdots & \phi_1 & -1 & \\ & & & \phi_p & \phi_{p-1} & \cdots & \phi_2 & \phi_1 & -1 \\ & & & & \phi_p & \cdots & \phi_3 & \phi_2 & \phi_1 \end{vmatrix}$$

$$= \sum_{m=1}^{p} \frac{\lambda_m^{k+p-1}}{\prod_{\substack{n=1 \\ n \neq m}}^{p} (\lambda_m - \lambda_n)}, \tag{8}$$

[8]For the use of Toeplitz matrices on double-differenced AR(1) models, see Han (2007).

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

where $k = t - s$ indicates the order of the matrix. Clearly, $\xi(t,s)$ does not depend on t,s but on their difference and is therefore denoted as $\xi(k)$ with independent variable the forecasting horizon k. The second equality in eq. (8) follows (only if $\lambda_m \neq \lambda_n$) from standard results in ARMA models (see, e.g., Karanasos, 1998, 2001, and in particular, eq. (2.6) in the latter citation).

3. EXPLICIT SOLUTION REPRESENTATIONS

In the upcoming subsections, we shall use the principal determinant to describe explicitly the general homogeneous and particular solutions of TV-LDE(p), the sum of which yields the general solution of eq. (3).[9] It leads to an explicit and computationally tractable solution representation of TV-ARMA processes, described in Theorem 1. A useful decomposition of the innovation part of the solution, especially for forecasting, is presented in Proposition 1.

3.1. Homogeneous Solution

We start by establishing a fundamental set of p solutions, being all banded Hessenbergian functions with entries the AR coefficients of the model. This is a crucial result (see for further details Appendix A.1), which enables one to obtain the Green function explicitly (see the introductory notes to Supplementary Appendix F3 and for more details see Paraskevopoulos and Karanasos, 2021) as well as the general homogeneous solution of TV-LDEs(p).

In a similar manner with the principal matrix and determinant eqs. (A.1) and (A.2) in Appendix A.1, define another $p-1$ banded Hessenbergian solutions of eq. (7) (see Proposition A1(ii)), which along with the principal determinant, which is identified as $\xi^{(1)}(t,s)$, form a fundamental set of solutions

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

that is, a set of p linearly independent solution sequences associated with eq. (7) (see Proposition A2). We show in Proposition A1(i) that $\xi^{(m)}(t,s)$ can also be expressed exclusively in terms of the principal determinant and the AR coefficients, that is,

$$\xi^{(m)}(t,s) = \sum_{r=1}^{p+1-m} \phi_{m-1+r}(s+r)\xi(t,s+r). \tag{9}$$

[9] 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 a compact and explicit solution for the aforementioned equations, but it appears not to be computationally tractable. 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." More recently, an alternative explicit and compact solution representation, which is computationally tractable, is presented in Paraskevopoulos and Karanasos (2021). This is based on the infinite Gauss–Jordan elimination method (see Supplementary Appendix F1 and the references cited there).

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

As a consequence of the superposition principle of TV-LDEs, the general solution of eq. (7) can be expressed as a linear combination of the elements of Ξ_s (see Proposition A3 in Appendix A.2)

$$y_{t,s}^{hom} = \sum_{m=1}^{p} \xi^{(m)}(t,s) y_{s+1-m} \text{ for } t \in \mathbb{Z}_{s+1-p}, \tag{10}$$

where $\{y_{s+1-p}, ..., y_s\}$ is a sequence of p initial condition values. The two-variable notation for the homogeneous solution, $y_{t,s}^{hom}$, is consistent with that used for the principal determinant function (or the Green function), as the latter is the solution of eq. (7) under the initial values $y_s = 1$ and $y_{s-n} = 0$ for $n = 1, ..., p-1$. This is a suitable notation for the asymptotic properties of TV-ARMA(p,q) processes.

Applying eq. (9) to eq. (10), we obtain the homogeneous solution entirely in terms of the principal determinant:

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

Expanding the determinant $\xi(t, s+r)$ in eq. (11), $y_{t,s}^{hom}$ can be ultimately expressed in terms of the AR coefficients, exclusively. This explicit representation of the homogeneous solution in terms of the principal determinant function in eq. (10) was recorded in earlier versions of the paper (see, e.g., Karanasos et al., 2020, and the references cited there).[10]

3.2. A Particular Solution and Its Decomposition

Recalling that $\upsilon_t = \varphi(t) + u_t$ (see eq. 2), a particular solution to eq. (3) subject to the initial condition values $y_s = y_{s-1} = \cdots = y_{s+1-p} = 0$ is given by

$$y_{t,s}^{par} = \sum_{r=s+1}^{t} \xi(t,r) \upsilon_r. \tag{12}$$

A proof of the above formula is demonstrated in Proposition A4 of Appendix A.2, noting that the solution in eq. (12) depends both on t and s. This has to be compared with the equivalent result presented in Miller (1968, p. 40, eqs. (2.8) and (2.9)), but now in a fully explicit and directly computable representation. The sum of eqs. (10) (or (11)) and (12) yields the general solution of eq. (3).

In Proposition 1 below, we introduce a decomposition of the innovation part in eq. (12), which will be used throughout this paper. But first we introduce the following definition.

[10] An analogous result for the multivariate case was given by Miller (1968) and Hallin (1986), but without explicitly giving the Green matrix. An explicit representation of the Green matrix can be found in Supplementary Appendix F3.

DEFINITION 1. *Define the function on* $\mathbb{Z}^2$ *by*

$$\xi_q(t,r) \stackrel{\text{def}}{=} \xi(t,r) + \sum_{l=1}^{q} \xi(t,r+l)\theta_l(r+l). \tag{13}$$

For each arbitrary but fixed $s \in \mathbb{Z}$*, we define the function on* $\mathbb{Z}^2$ *by*

$$\xi_{s,q}(t,r) \stackrel{\text{def}}{=} \sum_{l=s+1-r}^{q} \xi(t,r+l)\theta_l(r+l). \tag{14}$$

As $\xi_q(t,r)$ is equal to $\xi(t,r)$ plus a sum of terms consisting of the first q instances of $\xi(t,r+l)$ multiplied by corresponding MA coefficients, it can also be expressed as a banded Hessenbergian (the proof is deferred to Supplementary Appendix E1). The same holds, for $\xi_{s,q}(t,r)$. We shall refer to $\xi_q(t,r)$ and $\xi_{s,q}(t,r)$ as *banded Hessenbergian coefficients*, which are fully explicit (expressed in terms of the model coefficients) and more compact, compared with the corresponding coefficients "g(t, s)" in eq. (2.2), defined by Peiris (1986) (see also eq. (2.3)'s coefficients "G(t, s)" in Singh and Peiris, 1987).

PROPOSITION 1. *The innovation part of the particular solution in eq. (12) can be decomposed into two parts as follows:*

$$\sum_{r=s+1}^{t} \xi(t,r)u_r = \sum_{r=s+1}^{t} \xi_q(t,r)\varepsilon_r + \sum_{r=s+1-q}^{s} \xi_{s,q}(t,r)\varepsilon_r. \tag{15}$$

A formal proof of this result is provided in Appendix A.3. The integer interval $\{s+1,...,t\}$, in the relevant summations of the above equation, consists of $k = t-s$ points, known as the length of the interval that coincides with the forecasting horizon. In the first summation of the right-hand side, the errors whose index ranges over $\{s+1,...,t\}$ are unobservable. In the second one, the time interval extends from $s+1-q$ to s and therefore the errors whose index ranges over $\{s+1-q,...,s\}$ are observable.

3.3. The Main Result

The aforementioned results on linear difference equations are applied herein to obtain an explicit representation to the solution of eq. (1) in the following theorem. This is a consequence of the general solution representation of a TV-LDE(p) in eq. (3) as a sum of homogeneous and particular solutions, established in Sections 3.1 and 3.2, that is,

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

for all s with $s < t$. Notice that for each fixed t, the random variable y_t is independent of the past time point s, because it satisfies eq. (1) for any s, while both homogeneous and particular solution functions depend both on t and s. Applying eq. (15) to the second sum of the above expression of y_t, the upcoming theorem follows immediately.

THEOREM 1. *An alternative explicit representation of y_t in eq. (1) in terms of the prescribed sequence $\{y_r, s+1-p \leq r \leq s\}$ for any $s \in \mathbb{Z}$ such that $s < t$ is given by*

$$y_t = \underbrace{\sum_{m=1}^{p} \xi^{(m)}(t,s) y_{s+1-m}}_{\text{Homogeneous Solution Part}} + \underbrace{\sum_{r=s+1}^{t} \xi(t,r)\varphi(r)}_{\substack{\text{Particular Solution:} \\ \text{Drift Part}}} + \underbrace{\sum_{r=s+1}^{t} \xi_q(t,r)\varepsilon_r + \sum_{r=s+1-q}^{s} \xi_{s,q}(t,r)\varepsilon_r}_{\text{Particular Solution: Innovation Part}}. \tag{16}$$

The right-hand side of eq. (16) comprises four summation parts all involving the principal determinant. In view of Proposition 1, the sum of its last three parts is the particular solution given by eq. (12). More analytically, the first sum (the homogeneous solution in eq. (10)) is a product of the fundamental solutions multiplied by prescribed random variables. The second sum (the drift part of the particular solution in eq. (12)) is formed by products involving the principal determinant $\xi(t,r)$ multiplied by the drift $\varphi(r)$. The terms of the third sum (the first part of the "MA decomposition," see eq. (15)), are the banded Hessenbergian coefficients $\xi_q(t,r)$ in eq. (13) multiplied by the unobservable errors. Finally, the terms of the fourth sum (the second part of the "MA decomposition") are the banded Hessenbergian coefficients $\xi_{s,q}(t,r)$ in eq. (14), multiplied by observable errors. In the available open literature, there are no fully explicit solution representations of eq. (1), in which all the coefficients are easily handled and computationally tractable expressions (as being banded Hessenbergians). A compact representation of eq. (16) and its equivalence with the single determinant representation obtained by Kittappa (1993) are established in Paraskevopoulos and Karanasos (2021).

The methodology presented in this section can be used in the study of infinite-order autoregression models as well as in the case of the fourth-order moments for time-varying GARCH models (see, e.g., Canepa et al., 2022).

Another advantage of our TV-ARMA representation 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 nonstationary) or deterministic. Therefore, no restrictions are imposed on their functional form. In the deterministic case, the model allows for known abrupt changes, smooth changes, and mixtures of them. If the changes are smooth, the coefficients can depend on an exogenous variable, say x_t, or t or both. In the stochastic case, the model includes the generalized random coefficient (GRC) AR specification (see, e.g., Glasserman and Yao, 1995; Hwang

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

and Basawa, 1998) as a special case. In the aforementioned case, the model also allows for periodicity. We should also mention that the solution includes the case where the variable coefficients depend on the length of the series (see Examples D.2 and D.3 in Supplementary Appendix D1). Another consequence is an efficient approach to optimal linear forecasts based on a finite set of past observations, presented in Section 5. As in the case of AR(1) processes, the latter result follows from the explicit solution representation for the family of TV-ARMA(p,q) models, described by eq. (16).

3.4. Asymptotic Stability

The asymptotic stability[11] problem is to provide sufficient conditions which ensure that a class of stochastic processes solving eq. (1) approaches a solution independently of the prescribed p initial conditions (the effect of the initial conditions is gradually dying out) as $s \to -\infty$, that is, when the homogeneous solution in eq. (16) tends to zero, under a prescribed type of convergence. The explicit representation of the homogeneous solution in eq. (10) makes it possible to provide such types of conditions in Theorem 2 in the upcoming section (see, e.g., the absolute summability condition in (17)).[12] Stability characterizes 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.[13]

4. SECOND-ORDER STRUCTURE

Having specified a general explicit solution formula for the TV-ARMA types of model, we turn our attention to the explicit representation of their Wold–Cramér decomposition along with their second-order structure and the optimal linear forecast formulas associated with them. In this section, we shall restrict ourselves to the treatment of DTV-ARMA processes. In Section 6, we consider STV-AR processes.

[11] As pointed out by Grillenzoni (2000) stability is a useful feature of stochastic models because it is a sufficient (although not necessary) condition for optimal properties of parameter estimates and forecasts. Since model (1) can be expressed in Markovian form, stability conditions are necessary for many other significant properties, such as irreducibility, recurrence, regularity, nonevanescence, and tightness (see Grillenzoni, 2000, for details).

[12] Due to space considerations, alternative conditions on the asymptotic stability of TV-ARMA(p,q) models are not presented here, but are available in Karanasos et al. (2020, Sects. 4 and 5). In particular, the conditions in Theorem 2, established therein, include the "bounded random walk" of Giraitis et al. (2014), also used by Petrova (2019).

[13] 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, and depend on a finite number of parameters which need to be estimated. Other researchers have also considered the statistical properties of maximum likelihood estimators for very general nonstationary models. For example, Dahlhaus (1997) has obtained asymptotic results for a new class of locally stationary processes, which includes TV-ARMA processes (see Azrak and Mélard, 2006, and the references therein).

4.1. Wold–Cramér Decomposition

In Theorems 2 and 3, we show the existence and uniqueness of the Wold–Cramér decomposition (see Cramér, 1961)[14] and, therefore, impulse response functions (IRFs), for the DTV version of the model in eq. (1). In particular, we provide an explicit condition, that is,

$$\sum_{r=-\infty}^{t} |\xi(t,r)| < \infty, \text{ for all } t, \qquad \text{(absolute summability condition)} \tag{17}$$

which, along with the boundedness hypothesis for the MA coefficients, enables us to introduce fully explicit second-order formulas for the Wold–Cramér decomposition in eqs. (18a) and (20a).

Peiris (1986), Singh and Peiris (1987), and Kowalski and Szynal (1991), working with processes of zero drift, derive their solution representation as a purely nondeterministic process of mean zero in terms of the Green function. The presence of a drift yields a more complete and realistic model, in which the solution representation can be decomposed into two orthogonal parts: a deterministic part and a random one. We start with a more restrictive, but equally important result on the class of asymptotically stable second-order solution processes.

THEOREM 2. *Let the condition in (17) hold. Also let the drift $\varphi(t)$ and MA coefficients $\theta_l(t)$, $1 \leq l \leq q$, be bounded functions in t. Then there exists a unique asymptotically stable second-order solution process y_t of eq. (1), given explicitly by*

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

A proof of the theorem is given from first principles in Appendix B.1, supported by the results demonstrated in: Lemma B1, Corollary B1, and Proposition B1. Formally y_t in eq. (18a) is decomposed into two orthogonal parts: a nonrandom part (the unconditional expectation of the process) and a mean-zero random one.

In all that follows, we shall use the notation $z_t \overset{\text{def}}{=} \sum_{r=-\infty}^{t} \xi_q(t,r)\varepsilon_r$. Under the conditions of Theorem 2, z_t exists, being the random part of y_t in eq. (18a). Moreover, eq. (18a) can also be viewed as the mean square limit of eq. (16), as $s \to -\infty$. This follows from two facts, both shown in Appendix B.1: (i) $\sum_{m=1}^{p} \xi^{(m)}(t,s) y_{s+1-m} \to 0$, as $s \to -\infty$, which establishes that $\lim_{s\to-\infty} y_{t,s}^{hom} = 0$ or equivalently that the process y_t is asymptotically stable (see the proof of Theorem 2) and (ii) $\sum_{r=s+1-q}^{s} \xi_{s,q}(t,r)\varepsilon_r \to 0$, as $s \to -\infty$ (see Corollary B2).

[14] Since a nonstationary generalization of Wold's result was given by Cramér, it is referred to either as Wold–Cramér decomposition or as Cramér–Wold decomposition (see Nelson, 2008).

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

Proposition B1(iv) in Appendix B.1 shows that y_t in eq. (18a) can be rewritten as $y_t = \sum_{r=-\infty}^{t} \xi(t,r)\upsilon_r$. Thus, eq. (12) implies

$$y_t = \lim_{s\to-\infty} y_{t,s}^{par}. \tag{18b}$$

In the following remark, we discuss a potential application of Theorem 2.[15]

Remark 1. Theorem 2 could be useful to design or perform (asymptotic) inference. In particular, eq. (18a) is potentially very helpful even for those who wish to work on asymptotic theory. For example, a convenient way to study the approximation of partial sums of y_t, would be to show that for the very general class of TV-ARMA models, y_t is a decomposable Bernoulli shift. In Supplementary Appendix F4, we summarize what such a causal process is (see, also for more details, Massacci and Trapani, 2022). More specifically, if y_t satisfies a decomposable Bernoulli shift, then many results follow, for example, a functional central limit theorem (FCLT) or a strong invariance principle (SIP) of the form:

$$\max_{1\le k\le T} \frac{1}{k^{\zeta}} \left| \sum_{t=1}^{k} y_t - W(k) \right| = O_P(1),$$

where $W(k)$ is a Wiener process and $\zeta < \frac{1}{2}$. This result would be extremely helpful to study structural breaks for example.

The main result of Theorem 2 is generalized by Theorem 3 below. It is shown that second-order solutions of eq. (1) exist, without invoking the boundedness of the drift, but all other conditions of Theorem 2 remain the same. This is grounded on Proposition F.1(i) (see Supplementary Appendix F2.2), which shows that any second-order solution process y_t of eq. (1) is decomposed into two orthogonal parts: a mean-zero random process $\{y_t - \mathbb{E}(y_t)\}$ and a first moment process $\{\mathbb{E}(y_t)\}$, simultaneously satisfying the following equations:

$$z_t = \sum_{m=1}^{p} \phi_m(t) z_{t-m} + u_t, \tag{19a}$$

$$\mu_t = \varphi(t) + \sum_{m=1}^{p} \phi_m(t)\mu_{t-m}, \tag{19b}$$

for $z_t = y_t - \mathbb{E}(y_t)$ and $\mu_t = \mathbb{E}(y_t)$, respectively.

In all that follows, $\mathcal{M}(\varepsilon)$ (resp. $\mathcal{M}_t(\varepsilon)$) stands for the closed linear subspace of L_2 spanned by $\{\varepsilon_t, t\in\mathbb{Z}\}$ (resp. $\{\varepsilon_s, -\infty < s \le t\}$). On account of $\mathcal{M}_t(\varepsilon) \subset \mathcal{M}(\varepsilon)$, if $z_t \in \mathcal{M}_t(\varepsilon)$ for all $t\in\mathbb{Z}$, then $\{z_t\}$ is a process in $\mathcal{M}(\varepsilon)$.

THEOREM 3 (Generalization). *Let the absolute summability in (17) and the boundedness of MA coefficients $\theta_l(t)$, $1\le l\le q$, hold. Given a first moment process*

[15] We are indebted to an anonymous reviewer of this paper who provided us all necessary information to produce this remark.

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

μ_t, which satisfies eq. (19b), the process

$$y_t = \mu_t + \sum_{r=-\infty}^{t} \xi_q(t,r)\varepsilon_r \tag{20a}$$

is a unique second-order solution of eq. (1) such that $\mathbb{E}(y_t) = \mu_t$ and

$$\mathbb{E}(y_t^2) = \sum_{r=-\infty}^{t} \xi_q^2(t,r)\sigma^2(r) + \left(\mathbb{E}(y_t)\right)^2 < \infty.$$

The proof of Theorem 3 is presented in Appendix B.1. The second-order moment of y_t, given in the above equality, is shown in Section 4.2 (see the discussion below Proposition 2). Under the conditions of Theorem 3, eq. (20a) entails that, for every $t \in \mathbb{Z}$,

$$y_t - \mu_t = z_t = \sum_{r=-\infty}^{t} \xi_q(t,r)\varepsilon_r, \tag{20b}$$

which is a unique mean-zero solution process of eq. (19a) in $\mathcal{M}(\varepsilon)$. Formally, y_t in eq. (20a) satisfies eq. (1), when y_t takes on p prescribed random variables of the form $y_{s+1-m} = \mathbb{E}(y_{s+1-m}) + z_{s+1-m}$ for $1 \le m \le p$ and for any $s < t$, and therefore y_t also satisfies eq. (16). As the demean eq. (20a) is future-independent, that is, z_t can be expressed in terms of the current and past values of ε_r, we shall also refer to y_t as a *causal* solution of DTV-ARMA models. The uniqueness of z_t (see the proof of Theorem 3) entails that two causal solutions must differ in their means, exclusively (for further details, see the discussion below the proof of Proposition F.1 in Supplementary Appendix F2.2). The conditions of Theorem 3, which yield the causal solution in eq. (20a), will be referred to as a *causal environment*. Certainly, the conditions of Theorem 2 furnish a stronger causal environment.

In practice, a vector $\boldsymbol{\mu}_s = [\mu_s, \mu_{s-1}, ..., \mu_{s+1-p}]$ of p consecutive values of the first moment process $\{\mu_t\}$ can be estimated by the time series data via regression techniques (see Fuller, 1996). Employing $\boldsymbol{\mu}_s$ as the initial condition vector, the whole process $\{\mu_t\}$ in eq. (19b) can be constructed (see eq. (F.13) in Supplementary Appendix F3). In the above cited references (stated below (17)) and more recently in Grillenzoni (2000) and Azrak and Mélard (2006), the Wold–Cramér decomposition is computed via recursion. In sharp contrast, eqs. (18a) and (20a) provide more complete and explicit representations for the solution of eq. (1). To the extent of our knowledge, in the accessible literature on TV-ARMA processes, there is no fully explicit representation of the Wold–Cramér decomposition. Moreover, Theorem 3 (resp. Theorem 2) can be rephrased by replacing condition (17) and the boundedness of the MA coefficients with the regularity conditions (see Singh and Peiris, 1987; Kowalski and Szynal, 1991). Accordingly, regularity conditions also yield an explicit solution representation of eq. (1), which is given by eq. (20a) (resp. by eq. (18a), whereas in this case the boundedness hypothesis of the drift is a prerequisite). Moreover, the solution formulas in Theorems 2 and 3 explicitly recover the mean-zero second-order solution processes of demean DTV-ARMA(p,q) models ($\varphi(t) = 0$ for all t), established in the above cited references.

The regularity conditions noted above are not, however, necessary for $y_t \in L_2$, since they do not cover the case of periodic coefficients (see Grillenzoni, 1990 or Karanasos et al., 2014a, 2022). A necessary condition for the regularity conditions to hold is the square summability of $\{\xi_q^2(t,r)\}_{r\leq t}$, that is, $\sum_{r=-\infty}^{t} \xi_q^2(t,r) < \infty$, for all $t \in \mathbb{Z}$ (see Supplementary Appendix F5).

4.2. Unconditional Moments

In this subsection, we present explicit formulae for the first and second unconditional moments of the Wold–Cramér decomposition in eq. (18a) for DTV-ARMA processes coupled with sufficient and necessary conditions for their existence, as the following proposition demonstrates.

PROPOSITION 2. *Let the conditions of Theorem 2 hold. Then the unconditional mean of the asymptotically stable process y_t in eq. (18a) is given by*

$$\mathbb{E}(y_t) = \sum_{r=-\infty}^{t} \xi(t,r)\varphi(r) \tag{21a}$$

and its unconditional variance is given by

$$\mathbb{V}ar(y_t) = \sum_{r=-\infty}^{t} \xi_q^2(t,r)\sigma^2(r). \tag{21b}$$

Necessary conditions for the process y_t in eq. (18a) to be first- and second-order, on account of eqs. (21a) and (21b), are, respectively,

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

Moreover, the condition $\lim_{s\to-\infty} \xi(t,s) = 0$ is sufficient for the above two limits to exist, due to the boundedness of $\varphi(r)$, $\theta_l(r)$ and $\sigma^2(r)$.

A proof of the formulas in eqs. (21a) and (21b) is included in Proposition B1(iii) of Appendix B.1, provided that the conditions of Theorem 2 hold. Three illustrative examples for the $\mathbb{V}ar(y_t)$ are presented in Supplementary Appendix D1. The unconditional mean $\mathbb{E}(y_t)$ is of the same form for both the AR and the ARMA processes. Moreover, both solution processes in eqs. (18a) and (20a) share identical unconditional variances. The logical connections between the conditions that are described in the above proposition are summarized in the following commutative diagrams (see Appendix B.2 for a proof):

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

$$
\begin{array}{ccccc}
\sum_{r=-\infty}^{t} \xi(t,r)\varphi(r) \in \mathbb{R} & \Longleftarrow & \sum_{r=-\infty}^{t} |\xi(t,r)| < \infty & \Longrightarrow & \sum_{r=-\infty}^{t} \xi_q^2(t,r)\sigma^2(r) < \infty \\
\Downarrow & \swarrow & \Downarrow & \searrow & \Downarrow \\
\lim_{s\to-\infty} \xi(t,s)\varphi(s) = 0 & \Longleftarrow & \lim_{s\to-\infty} \xi(t,s) = 0 & \Longrightarrow & \lim_{s\to-\infty} \xi_q^2(t,s)\sigma^2(s) = 0.
\end{array}
$$

Commutative Diagrams

4.3. Autocovariance Function

In the following proposition, we state an explicit expression of the covariance structure for the Wold–Cramér decomposition of DTV-ARMA(p,q) processes.

PROPOSITION 3. *Let the conditions of Theorem 3 hold. Then the time-varying ℓ-order autocovariance function $\gamma_t(\ell) = \mathbb{C}ov(y_t, y_{t-\ell})$, $\ell \in \mathbb{Z}_0$, of a second-order solution process $\{y_t\}$ given in eq. (20a), exists and is given explicitly by*

$$\gamma_t(\ell) = \sum_{r=-\infty}^{t-\ell} \xi_q(t,r)\xi_q(t-\ell,r)\sigma^2(r). \tag{22}$$

Moreover, $\lim_{\ell\to\infty} \gamma_t(\ell) = 0$.

A proof of Proposition 3 is presented in Appendix B.3. The time-varying variance of y_t in eq. (21b) is recovered by applying $\gamma_t(\ell)$ for $\ell = 0$, that is, $\gamma_t(0) = \mathbb{V}ar(y_t)$.

From a computational viewpoint, the covariance structure of the process y_t can be numerically evaluated by computing the banded Hessenbergian coefficients, $\xi_q(t,r)$ in eq. (13) and substituting these in eq. (22). The next remark highlights the importance of the existence of finite second moments.

Remark 2. Azrak and Mélard (2006) considered the asymptotic properties of QML estimators for the DTV-ARMA models, where the coefficients depend not only on t but on the sample size T too (see Alj et al., 2017, for the multivariate case). In their Theorem and Lemma 1, the existence of finite second-order moments was required. They also showed there that the dependence of the model on 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).

5. FORECASTING

The invertibility is crucial for obtaining reliable approximate forecasts based on a finite sequence of past observations (see, e.g., Hamilton, 1994, p. 85). The

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

inversion of a second-order mean-zero solution process is discussed in this section along with forecasting. We derive two generic explicit forms for the k-step-ahead optimal linear predictor (in L_2 sense), based on, respectively, infinite and finite sequences of observable random variables coupled with their forecast and mean square errors.

Recalling that $z_t = \sum_{r=-\infty}^{t} \xi_q(t,r)\varepsilon_r$ stands for the unique mean-zero solution process of eq. (19a) in $\mathcal{M}(\varepsilon)$.

5.1. Invertibility

The invertibility of $\{z_t\}$ is discussed in some detail in Supplementary Appendix G1. By analogy with the definition of the principal matrix, we there define the banded Hessenberg matrix $\mathbf{\Theta}_{t,s}$ (see eq. (G.2)), whose entries are the opposite sign moving average coefficients of eq. (1). The associated principal determinant is defined by $\vartheta(t,s) = \det(\mathbf{\Theta}_{t,s})$. By analogy with the definition of $\xi_q(t,r)$ in eq. (13), the associated banded Hessenbergian coefficients are $\vartheta_p(t,r) = \vartheta(t,r) - \sum_{m=1}^{p} \vartheta(t,r+m)\phi_m(r+m)$. We show in Theorem G.1 that in a causal environment the additional condition $\sum_{r=-\infty}^{t} |\vartheta(t,r)| < \infty$ is sufficient for the invertibility of z_t and its explicit representation is given by

$$\varepsilon_t = \sum_{r=-\infty}^{t} \vartheta_p(t,r) z_r.$$

Let $\mathcal{M}_t(z)$ be the closed linear subspace of L_2 spanned by $\{z_r, -\infty < r \leq t\}$. Under both absolute summability conditions, that is, $\sum_{r=-\infty}^{t} |\xi_q(t,r)| < \infty$ and $\sum_{r=-\infty}^{t} |\vartheta(t,r)| < \infty$, the representation of z_t in terms of ε_r coupled with the above representation of ε_t in terms of z_r ensures that $\mathcal{M}_t(z) = \mathcal{M}_t(\varepsilon)$, for all $t \in \mathbb{Z}$. Replacing the aforementioned absolute summability conditions with the AR and MA regularity conditions, respectively, they also ensure that $\mathcal{M}_t(z) = \mathcal{M}_t(\varepsilon)$, for all $t \in \mathbb{Z}$ (see, e.g., Singh and Peiris, 1987). The conditions ensuring the invertibility of the solution process y_t will be referred to as *invertible environment*.

5.2. Forecasts Based on Infinite Observations

The following approach is appropriate for large sample optimal linear forecasts. In what follows, t stands for the forecast time point, $k = t - s$ for $s < t$ denotes the forecast horizon, and $\{\varepsilon_r\}_{r \leq s}$ is the sequence of observable random variables. The orthogonal projection of z_t on the closed subspace $\mathcal{M}_s(\varepsilon)$ of $\mathcal{M}(\varepsilon)$ is the k-step-ahead optimal linear predictor of z_t, based on $\mathcal{M}_s(\varepsilon)$, which is given explicitly by

$$\hat{P}(z_t \,|\, \mathcal{M}_s(\varepsilon)) = \sum_{r=-\infty}^{s} \xi_q(t,r)\varepsilon_r.$$

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

Replacing the absolute summability conditions with the AR and MA regularity conditions, established in the references cited in Section 4.1, identical forecasting formulas for the process $\{z_t\}$ are obtained, but now the banded Hessenbergian coefficients $\xi_q(t,r)$ are explicitly expressed directly in terms of the coefficients of the model in eq. (1), without invoking recursion for their computation.

Assuming in addition that $\{\varepsilon_t\}$ is a white noise martingale difference sequence, relative to $\{z_{s-r}\}_{0\leq r<\infty}$, that is, $\mathbb{E}(\varepsilon_t|z_s, z_{s-1}, ...) = 0$ for all t, the above stated linear predictor coincides with the conditional expectation of z_t based on $\mathcal{M}_s(\varepsilon)$, that is, the k-step-ahead optimal predictor minimizing the corresponding mean square error (see Brockwell and Davis, 2016, p. 334).

Following Hamilton (1994, p. 74), we shall denote by $\hat{\mathbb{E}}\big(y_t\,|\,\mathcal{M}_s(\varepsilon)\big)$ the orthogonal projection of $y_t \in L_2$ on the closed linear subspace of L_2 spanned by the sequence $\{1, \varepsilon_s, \varepsilon_{s-1}, ...\}$, which is also an orthogonal sequence, and contains all constants. Let $\{\mu_t\}$ be the estimated first moment process generated by eq. (19b). An explicit representation of the linear predictor of y_t in eq. (20a) for $t > s$ is given by adding the estimated mean $\mathbb{E}(y_t) = \mu_t$ to the foregoing forecast $\hat{P}(z_t\,|\,\mathcal{M}_s(\varepsilon))$, that is,

$$\hat{\mathbb{E}}(y_t\,|\,\mathcal{M}_s(\varepsilon)) = \mathbb{E}(y_t) + \sum_{r=-\infty}^{s} \xi_q(t,r)\varepsilon_r.$$

The forecast error $\mathbb{FE}_{t,s} \overset{\text{def}}{=} y_t - \hat{\mathbb{E}}(y_t\,|\,\mathcal{M}_s(\varepsilon)) = z_t - \hat{\mathbb{E}}(z_t\,|\,\mathcal{M}_s(\varepsilon))$ and the associated mean square error $\mathbb{MSE}_{t,s} \overset{\text{def}}{=} \mathbb{V}ar(\mathbb{FE}_{t,s})$ are also explicitly expressed in the forthcoming eqs. in (23b), being identical to the counterpart errors obtained therein.

In an invertible environment, since $\mathcal{M}_t(z) = \mathcal{M}_t(\varepsilon)$ for all $t \in \mathbb{Z}$, the sequence $\{z_r\}_{r\leq s}$ can also be identified as the sequence of observable random variables. In this environment, the optimal linear predictor formula is presented in Supplementary Appendix G2.

If the conditions of Theorem 2 hold, then $\mathbb{E}(y_t)$ is given by eq. (21a) and therefore the k-step-ahead optimal linear predictor formula, stated above, is modified by replacing $\mathbb{E}(y_t)$ with $\sum_{r=-\infty}^{t} \xi(t,r)\varphi(r)$.

5.3. Forecasts Based on Finite Observations

In what follows, we discuss a more realistic approach to forecasting, focusing on finite size sample forecasts. If we have a finite number of observations, say N, at our disposal, we can truncate the series of $\hat{\mathbb{E}}(y_t\,|\,\mathcal{M}_s(\varepsilon))$ up to and including the $(s+1-N)$ term to obtain approximations to the k-step-ahead optimal linear forecasts obtained in the previous paragraph.

An alternative optimal linear forecasting approach to the one based on infinite observations is obtained by taking advantage of the second-order solution in eq. (20a), but written as in eq. (16) (see the discussion below eq. (20b) in Section 4.1). Let $\mathcal{K}_s$ for $s < t$ stand for the linear subspace of L_2 spanned by the

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

set: $\{y_s, y_{s-1}, \ldots, y_{s+1-p}\} \cup \{\varepsilon_s, \ldots, \varepsilon_{s+1-q}\} \cup \{1\}$. Formally, $\mathcal{K}_s$ is closed, as being the span of a finite set of elements of L_2. In this case, y_{s-i}, for $0 \le i \le p-1$, and ε_{s-i}, for $0 \le i \le q-1$, are the observable random variables. The optimal linear predictor of y_t in eq. (16), based on $\mathcal{K}_s$, along with the forecast and mean square errors are explicitly expressed in the following proposition.

PROPOSITION 4. *The k-step-ahead optimal linear predictor of $y_t \in L_2$ in eq. (16), based on $\mathcal{K}_s$, is*

$$\hat{\mathbb{E}}(y_t \,|\, \mathcal{K}_s) = \sum_{m=1}^{p} \xi^{(m)}(t,s) y_{s+1-m} + \sum_{r=s+1}^{t} \xi(t,r)\varphi(r) + \sum_{r=s+1-q}^{s} \xi_{s,q}(t,r)\varepsilon_r. \tag{23a}$$

The forecast error $\mathbb{FE}_{t,s} = y_t - \hat{\mathbb{E}}(y_t|\mathcal{K}_s)$ and its associated mean square error (its variance) are

$$\mathbb{FE}_{t,s} = \sum_{r=s+1}^{t} \xi_q(t,r)\varepsilon_r, \ \mathbb{MSE}_{t,s} = \sum_{r=s+1}^{t} \xi_q^2(t,r)\sigma^2(r). \tag{23b}$$

As the expectation of $\mathbb{FE}_{t,s}$ is zero, the forecast is unbiased.

A proof of Proposition 4 is deferred to Supplementary Appendix G2. Since the expressions of $\mathbb{FE}_{t,s}$ and $\mathbb{MSE}_{t,s}$ are independent of ys, we conclude that both error formulas in eq. (23b) remain invariant for any second-order solution process. More specifically, for any two second-order solutions of eq. (1), say y_t and y_t^*, it follows that $y_t - \hat{\mathbb{E}}(y_t|\mathcal{K}_s) = \sum_{r=s+1}^{t} \xi_q(t,r)\varepsilon_r = y_t^* - \hat{\mathbb{E}}(y_t^*|\mathcal{K}_s)$. These error formulas are identical to those associated with an infinite sequence of past observations, as already stated in the preceding paragraph.

In the case of ARMA models with constant parameters, by using $t-r$ in place of (t,r), the coefficient $\xi^{(m)}(t,s)$ in eq. (9) becomes $\xi^{(m)}(t-s) = \sum_{r=1}^{p+1-m} \phi_{m-1+r}\xi(t-s-r)$ and the banded Hessenbergian coefficient in eq. (14) becomes $\xi_{s,q}(t-r) = \sum_{l=s+1-r}^{q} \xi(t-r-l)\theta_l$, where $\xi(r)$ is given by eq. (8). In this case, the k-step-ahead optimal linear predictor in eq. (23a) takes the form

$$\hat{\mathbb{E}}(y_t \,|\, \mathcal{K}_s) = \sum_{m=1}^{p} \xi^{(m)}(t-s) y_{s+1-m} + \varphi \sum_{r=s+1}^{t} \xi(t-r) + \sum_{r=s+1-q}^{s} \sum_{l=s+1-r}^{q} \xi(t-r-l)\theta_l\varepsilon_r. \tag{23c}$$

We should highlight the fact that eq. (23c) coincides with the formula obtained by Karanasos (2001), in his Theorem 1, eq. (2.7) (see, for more details, Supplementary Appendix D2).

5.4. Some Further Consequences

For each time point t, the forecast error in eq. (23b) converges in $\mathcal{M}_t(\varepsilon)$, as $s \to -\infty$, to the mean-zero random part of eq. (20a) (or eq. (18a)), that is,

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

$$\lim_{s\to-\infty} \mathbb{FE}_{t,s} = \sum_{r=-\infty}^{t} \xi_q(t,r)\varepsilon_r.$$

Moreover, the linear optimal forecast, based on $\mathcal{M}_s(\varepsilon)$, and the mean square error converge, respectively, to the first moment and to the unconditional variance, as $s \to -\infty$, that is,

$$\lim_{s\to-\infty} \hat{\mathbb{E}}(y_t \,|\, \mathcal{M}_s(\varepsilon)) = \mathbb{E}(y_t),$$

$$\lim_{s\to-\infty} \mathbb{MSE}_{t,s} = \sum_{r=-\infty}^{t} \xi_q^2(t,r)\sigma^2(r) = \mathbb{V}ar(y_t).$$

We remark that the explicit form of the variance for a Gaussian process is essential for the determination of the confidence intervals for $\hat{\mathbb{E}}(y_t \,|\mathcal{K}_s)$.

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

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

Changing the summation limits ($[s_1 + 1..t_1]$ and $[s_2 + 1..t_2]$), both to $[0..k-1]$, we get

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

Accordingly, in a time-varying coefficient setting, a comparison between $\mathbb{MSE}_{t_1,s_1}$ and $\mathbb{MSE}_{t_2,s_2}$, whenever $t_1 - s_1 = t_2 - s_2$, entails that $\mathbb{MSE}_{t_1,s_1} \neq \mathbb{MSE}_{t_2,s_2}$. In contrast, in a constant parameter setting, the two MSEs coincide, since $\mathbb{MSE}_{t_1,s_1} = \mathbb{MSE}_k = \mathbb{MSE}_{t_2,s_2}$, where

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

6. STOCHASTIC COEFFICIENTS

In Sections 4 and 5, we restricted ourselves to the treatment of DTV models. In this section, we examine processes with stochastically varying coefficients. For simplicity, instead of ARMA processes, we will concentrate on the AR specification. In particular, we will investigate two models: the random coefficients one and the double stochastic AR process.[16] But first we will express the STV-AR

[16] Some results concerning the GRC model are not presented here, but are available at Karanasos et al. (2020, Sect. 7.3).

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

model in a companion matrix form. The proofs of the present section are provided in Appendix C.

6.1. Companion Matrix Form

In this subsection, we show how to utilize the principal determinant (Green function) and the mth fundamental solution in order to obtain a compact and explicit representation of the companion matrix.

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

$$y_t = \phi_{0t} + \boldsymbol{\phi}_t' \mathbf{y}_{t-1} + \varepsilon_t, \tag{24}$$

where $\mathbf{y}_{t-1} = (y_{t-1}, y_{t-2}, \dots, y_{t-p})'$ is a $p \times 1$ vector of preceding random variables of y_t, and $\boldsymbol{\phi}_t = (\phi_{1t}, \phi_{2t}, \ \dots, \phi_{pt})'$ is a $p \times 1$ vector of the autoregressive random coefficients. Notice that we denote the stochastically varying coefficients, including the drift ϕ_{0t}, by ϕ_{mt}, $m = 0, \dots, p$, instead of $\phi_m(t)$, which was the notation used for the deterministic ones.

It is well known that model (24) can be written in a companion matrix form

$$\mathbf{y}_t = \boldsymbol{\phi}_{0t} + \boldsymbol{\Phi}_t \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{25}$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_t 0 \dots 0)'$, $\boldsymbol{\phi}_{0t} = (\phi_{0t} 0 \dots 0)'$, and the companion (square) matrix $\boldsymbol{\Phi}_t$ of order p associated with the vector $\boldsymbol{\phi}_t$ is given by

$$\boldsymbol{\Phi}_t = \begin{pmatrix} \phi_{1t} & \phi_{2t} & \dots & \phi_{p-1,t} & \phi_{pt} \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}. \tag{26}$$

That is, the STV-AR(p) process is converted to a p-dimensional vector STV-AR(1) model. For any set of p prescribed random variables $\mathbf{y}_s$, iterating eq. (25) yields

$$\mathbf{y}_t = \mathbf{C}_{t,s} \mathbf{y}_s + \sum_{r=s+1}^{t} \mathbf{C}_{t,r} (\boldsymbol{\phi}_{0r} + \boldsymbol{\varepsilon}_r), \tag{27}$$

where $\mathbf{C}_{t,s} = \prod_{r=s+1}^{t} \boldsymbol{\Phi}_r$ is the product of companion matrices with initial square matrix $\mathbf{C}_{t,t} = \mathbf{I}$ of order p. It follows directly from eq. (27) and Theorem 1 (see, for more details, Paraskevopoulos and Karanasos, 2021; see also, Supplementary Appendix F3) that the p-dimensional square matrix $\mathbf{C}_{t,s}$ is given by

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

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

In other words, the element occupying the $(n+1,m)$th entry of the matrix $\mathbf{C}_{t,s}$ ($n = 0,\ldots,p-1$) is the mth fundamental solution $\xi^{(m)}(t-n,s)$. We recall that $\xi^{(1)}(t,s)$ is given in eq. (5), where now $\phi_m(t)$ in eq. (4) is replaced by ϕ_{mt}, and $\xi^{(m)}(t,s)$ is given in eq. (A.2).

6.2. Random Coefficients AR Model

In this subsection, we examine the random coefficient AR(p) model (with acronym RC-AR(p)), which is given by eq. (24), using for this the following notation and specifications: $\boldsymbol{\phi}_t^* = (\phi_{0t}\ \boldsymbol{\phi}_t)'$, $t = s+1, s+2\ldots$, is an $i.i.d.(p+1)$-dimensional random vector of the coefficients, and the $i.i.d.$ errors $\{\varepsilon_t, t \geq s+1\}$ are independent of the random drift and autoregressive coefficients. Also let ε_s be a random variable independent of everything else, being the initial state.

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

Condition 1. $\xi(t,s) \overset{a.s}{\to} 0$ as $t \to \infty$.

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

THEOREM 4. *Consider the RC-AR(p) model. Under Condition 1, the following are equivalent:*

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

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

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

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

v) $\sup_{t\geq s+1}\left|\mathbf{C}_{t-1,s}\boldsymbol{\varepsilon}_t\right| < \infty$ *a.s.*

As pointed out by Erhardsson (2014), the implications $(ii) \implies (iii) \implies (iv) \implies (v)$ remain valid even if condition $\left|\left|\mathbf{C}_{t,s}\right|\right| \overset{a.s.}{\to} 0$ as $t\to\infty$ does not hold.

To the best of our knowledge, no prior studies have provided an analogous condition to Condition 1 for the RC-AR(p) model.

6.3. Double Stochastic AR Models

In this subsection, we examine the more general case where the autoregressive coefficients follow AR processes. We show that for this model the unconditional variance exists in $\mathbb{R}_{>0}$ provided that the associated Green function convergences in L_2, a result which is in line with Theorem 2(ii) in Karanasos et al. (2020). In other

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

words, we investigate the double stochastic AR model, hereafter termed DS-AR (for double stochastic processes and in particular ARMA processes with ARMA coefficients, see, e.g., Grillenzoni, 1993, and the references therein). This model is defined by eq. (24), but in this case, the coefficients, ϕ_{mt} for $m = 1, \ldots, p$, follow AR processes:

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

where β_{m0} and β_{ml} are constant coefficients and $p_m \in \mathbb{Z}_{\geq 1}$ for all $m : 1 \leq m \leq p$. $\{e_{mt}\}$ are martingale difference sequences defined on L_2, where e_{mt} and $\varepsilon_{t\pm b}$, $b \in \mathbb{Z}$, are mutually independent for all m, b, and t. We will also assume that the drift in eq. (24) is time invariant, that is, $\phi_{0t} = \phi_0$ for all t.

The results in Sections 4 and 5 can be easily modified to cover DS-AR models by replacing the fundamental solutions with their respective (unconditional and conditional) expectations. More specifically, we present two theorems followed by two propositions (their proofs essentially repeat the arguments of the proofs of those in Sections 4 and 5).

There are two sufficient conditions ensuring the Wold–Cramér decomposition of DS-AR(p) models, and therefore the existence of the first two unconditional moments:

$$\left.\begin{aligned} &\sum_{r=-\infty}^{t} |\mathbb{E}\big(\xi(t,r)\big)| < \infty \ \text{ for all } t \text{ (first-order absolute summability)}, \\ &\sum_{r=-\infty}^{t} \mathbb{E}\big(\xi^2(t,r)\big) < \infty \ \text{ for all } t \ \text{ (second-order summability)}. \end{aligned}\right\} \tag{29}$$

Remark 3. Generally, it is very difficult to verify if the two summability conditions are fulfilled. Only some special cases allow for explicit solutions (see, Anděl, 1991, and the references therein). A sufficient condition for the absolute summability to hold is that $\{\sum_{m=1}^{p} \phi_{mt}\}$ belongs with probability one to the interval $(-1, 1)$, nearly everywhere, that is, with the exception, at most, of a finite number of t (see, e.g., Grillenzoni, 1993). Similarly, a sufficient condition for the square summability to hold is that with probability one $\lambda_t^{(\max)}[\mathbf{\Phi}_t^{\otimes 2}] < 1$, nearly everywhere, where $\lambda_t^{(\max)}[\mathbf{\Phi}_t^{\otimes 2}]$ refers to the modulus of the largest eigenvalue of $\mathbf{\Phi}_t^{\otimes 2} = \mathbf{\Phi}_t \otimes \mathbf{\Phi}_t$ and $\otimes$ stand for the Kronecker product; we recall that $\mathbf{\Phi}_t$ is the companion matrix (see eq. (26)).

THEOREM 5. *Let the two summability conditions in (29) hold. The Wold–Cramér decomposition (in L_2 sense) is a solution of the DS-AR(p) model in eq. (24), where its AR coefficients are given by eq. (28), being of the form*

$$y_t = \sum_{r=-\infty}^{t} \xi(t,r)(\phi_0 + \varepsilon_r). \tag{30}$$

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

THEOREM 6. *If all the AR coefficients, ϕ_{mt}, $m = 1, \ldots, p$, are strictly stationary, then eq. (24) has a stationary solution of the type (30) if and only if*

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

In what follows, we present explicit formulae for the first and second unconditional moments for the DS-AR family of processes coupled with sufficient and necessary conditions for their existence.

PROPOSITION 5. *Under the two conditions in (29), it follows from Theorem 5 that the unconditional mean of the DS-AR(p) process in eq. (30) is given by*

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

A necessary condition for the absolute summability to hold is

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

Moreover, the unconditional variance of the process is given by

$$\mathbb{V}ar(y_t) = \phi_0^2\, \mathbb{V}ar\Big(\sum_{r=-\infty}^{t} \xi(t,r)\Big) + \sigma_\varepsilon^2 \sum_{r=-\infty}^{t} \mathbb{E}\big(\xi^2(t,r)\big).$$

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

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

We notice that $\lim_{s\to-\infty} \mathbb{E}\big(\xi^2(t,s)\big) = 0$ is equivalent to $\lim_{s\to-\infty} ||\xi(t,s)||_{L_2}^2 = 0$, which, in turn, is equivalent to $\lim_{s\to-\infty} ||\xi(t,s)||_{L_2} = 0$. In this latter case, we write $\lim_{s\to-\infty} \xi(t,s) \overset{L_2}{=} 0$.

PROPOSITION 6. *Following the notation of Proposition 4, let $\mathcal{K}_s$ be the smallest closed linear subspace of L_2 spanned by the finite observable sequence*

$$\{y_s, \ldots, y_{s+1-p}\} \cup \left(\bigcup_{i=1}^{m} \{\phi_{i,s}, \phi_{i,s-1}, \ \ldots, \phi_{i,s+1-p_i}\}\right) \cup \{1\}.$$

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

$$\hat{\mathbb{E}}(y_t \,|\, \mathcal{K}_s) = \sum_{m=1}^{p} \hat{\mathbb{E}}\big(\xi^{(m)}(t,s)|\mathcal{K}_s\big) y_{s+1-m} + \phi_0 \sum_{r=s+1}^{t} \hat{\mathbb{E}}\big(\xi(t,r) \,|\, \mathcal{K}_s\,\big).$$

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

Moreover, $\lim_{s\to-\infty}\hat{\mathbb{E}}(y_t|\mathcal{K}_s)=\mathbb{E}(y_t)$*, which is given by eq. (31). The forecast error* $\mathbb{FE}_{t,s}$ *for the above k-step-ahead predictor, is given by*

$$\mathbb{FE}_{t,s}=\phi_0\sum_{r=s+1}^{t}\Big(\xi(t,r)-\hat{\mathbb{E}}(\xi(t,r)\,|\mathcal{K}_s)\Big)+\sum_{r=s+1}^{t}\xi(t,r)\varepsilon_r+\sum_{m=1}^{p}\big(\xi^{(m)}(t,s)\\ -\hat{\mathbb{E}}(\xi^{(m)}(t,s)|\mathcal{K}_s)\big)y_{s+1-m},$$

and its variance, $\mathbb{V}ar(y_t|\mathcal{K}_s)$*, is given by*

$$\mathbb{V}ar(y_t|\mathcal{K}_s)=\mathbb{V}ar\Bigg(\big(\phi_0\sum_{r=s+1}^{t}\xi(t,r)+\sum_{m=1}^{p}\xi^{(m)}(t,s)y_{s+1-m}\big)|\mathcal{K}_s\Bigg)\\ +\sigma_\varepsilon^2\sum_{r=s+1}^{t}\hat{\mathbb{E}}\big(\xi^2(t,r)|\mathcal{K}_s\big)$$

(we recall that $\xi(t,r)$ *have been introduced in eqs. (4) and (5), where* $\phi_m(t)$ *are now replaced by* ϕ_{mt}*). Finally,* $\lim_{s\to-\infty}\mathbb{V}ar(y_t|\mathcal{K}_s)=\mathbb{V}ar(y_t)$*, which is given in Proposition 5.*

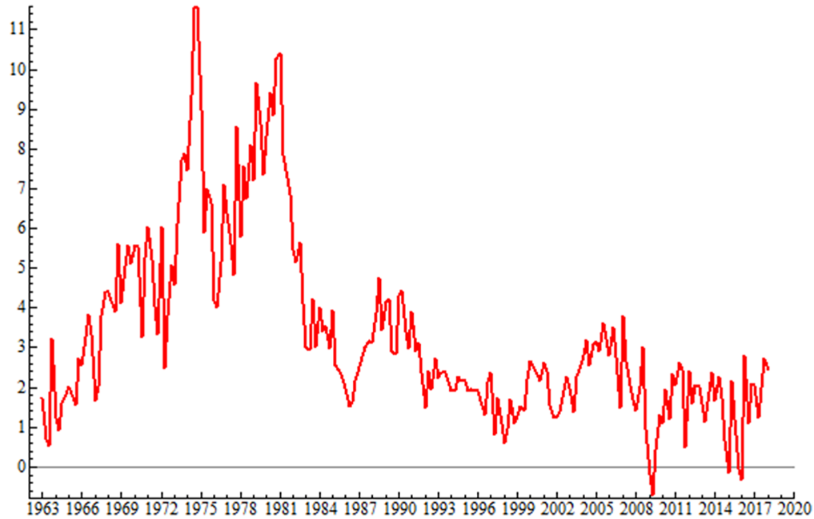
Remark 4. If the AR coefficients are deterministically varying, then the results in Proposition 6 coincide with those in Proposition 4 for the AR model.

7. MODELLING INFLATION

In this section, we directly link econometric theory with empirical evidence. In our empirical application, we consider the possible presence of structural breaks in inflation in the United States. Although the empirical example serves as an illustration of the practical usefulness of the proposed approach and a further examination of the structural breaks versus long memory debate[17] is beyond the scope of this paper, we should highlight the fact that the unified theory can be applied to infinite-order autoregressions with either constant or variable coefficients. Since a special case of such a process is the time-varying long memory specification with structural breaks, our methodology unifies the concurring views of long memory and structural breaks. We use quarterly data on the GDP deflator as the measure of the price level. The dataset consists of observations from 1963Q4 to 2018Q1. The figure below plots the data used for the empirical investigation. Inflation is calculated as the quarterly change of price level at an annualized rate calculated as $\pi_t=400(\ln(P_t/P_{t-1})$.

[17] Sibbertsen (2004) reviews the literature on misspecifying structural breaks as long-range dependence and proposes various methods for distinguishing both these phenomena. As pointed out by Sibbertsen in many situations, it is not clear whether the observed dependence structure is real long memory or an artefact of some other phenomena such as structural breaks. Long memory in the data would have strong consequences (i.e., for forecasting future events it is important to know whether the data exhibits long-range dependence or if it is an artefact of structural breaks). However, distinguishing both of these phenomena is difficult because their finite sample properties are rather similar.

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press



In terms of inflation modelling, the period under consideration is of particular interest as it covers the boom-time inflation of the late 1960s, the stagflation in the 1970s, and the double-digit inflation of the early 1980s. During this period, substantial shifts in monetary policy occurred, most notably the Fed's radical step of switching policy from targeting interest rates to targeting the money supply in the early 1980s. Therefore, when modelling inflation, it is important to allow for time-varying parameters. In what follows, we estimate AR(p) models with abrupt structural breaks.[18]

The optimal model is an AR(2) process with two deterministic abrupt breaks [DAB-AR(2; 2)] at fixed points of time t_1 and t_2, where $t_1 > t_2$. The process is defined by

$$y_t = \begin{cases} \varphi_1 + \phi_{1,1} y_{t-1} + \phi_{2,1} y_{t-2} + \sigma_1 e_t & \text{if} \quad t > t_1, \\ \varphi_2 + \phi_{1,2} y_{t-1} + \phi_{2,2} y_{t-2} + \sigma_2 e_t & \text{if} \quad t_2 < t \leq t_1, \\ \varphi_3 + \phi_{1,3} y_{t-1} + \phi_{2,3} y_{t-2} + \sigma_3 e_t & \text{if} \quad t \leq t_2, \end{cases} \tag{32}$$

where $e_t \sim i.i.d.\ (0,1)$ for all t and $0 < \sigma_i^2 \leq M < \infty$, $i = 1,2,3$. An explicit solution representation is given by eq. (I.1) in Supplementary Appendix I1.

The choice of the number of lags was based on the modified AIC (Akaike information criterion) and the Bayesian information criteria. The break points are

[18] Various unit-root tests show that, in general, we can reject the null hypothesis of a unit root in inflation series (see, for details, Supplementary Appendix H1). Many studies have also investigated the orders of integration of the Fisher equation variables. In particular, Sun and Phillips (2004) find that the three Fisher components (i.e., nominal and real interest rates, and expected inflation) are integrated of the same order, with memory parameter in the range (0.75, 1). There is a voluminous literature on the long memory versus structural break debate (see, e.g., the book by Goerg, 2010). For a recent discussion of long range dependence and multiple structural changes in the persistence of European inflation rate series, see Karanasos et al. (2016).

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

TABLE 1. Structural break test and estimation results.

Panel A: Bai and Perron tests of $L+1$ vs. L sequentially determined breaks

Null hypotheses	F-statistic	Critical value
H_0 : 0 vs 1	57.96**	13.98
H_0 : 1 vs 2	18.13**	15.72
H_0 : 2 vs 3	13.57	16.83

Panel B: Model estimation and misspecification tests

Period	φ_i	$\phi_{1,i}$	$\phi_{2,i}$	σ_i
1964Q2–1976Q3	0.496** (0.224)	0.470* (0.108)	0.376* (0.102)	1.077* (0.367)
1976Q4–1986Q2	3.637* (0.954)	0.710* (0.119)	0.127 (0.112)	2.300* (0.689)
1986Q3–2018Q1	2.859* (0.396)	0.247* (0.082)	−0.314* (0.077)	2.160* (0.489)
R^2		0.614		
Breusch–Godfrey test		2.055 (0.561)		
White test		3.103 (0.376)		
Log-likelihood		−482.91		

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

treated as unknown. Note that breaks in the variance are permitted provided that they occur on the same dates as the break in the autoregressive parameters.

Coming to the estimation procedure, the first step is to identify possible points of parameter changes. In order to do so the Bai and Perron (2003) sequential test on inflation rates is used to identify possible breaks during the sample period. The Bai–Perron test (for details, see Supplementary Appendix H2) concludes that there are two structural breaks. The results of the structural break test are summarized in Panel A of Table 1.

A possible limitation of structural break models is that they are typically sample-period-dependent. Therefore, their forecasting performance may be affected by the assumption that an abrupt break at one point in time is a one off shock to inflation and therefore not persistent (for a detailed discussion on the weaknesses of the structural break approach in analyzing and forecasting time series such as inflation, see Sun and Phillips, 2004; Phillips, 2005). Note that in the literature, fractionally integrated specifications have often been used to model inflation along with those with structural breaks. In this respect, taking into account fractional integration may produce better forecasting properties.[19]

[19]Empirical studies on fractional integration analysis include, among others, Baillie et al. (2002), Hsu (2005), Canarella and Miller (2017), and Iacone et al. (2019). However, as Diebold and Inoue (2001) point out, structural breaks in time series can induce a strong persistence in the autocorrelation function and hence generate spurious long

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

The first break occurred in the mid-1970s, when the Fed tightened monetary policy to fight the high inflation rate after the end of the Bretton Woods period. The second break occurred in 1986, when the Fed embarked on an aggressive policy to reduce inflation, which reached unusually high levels starting from the 70s. As a result, inflation fell from 10.5% at the end of 1980 to 1.1% in 1986Q2, which is also the date of the estimated break.[20]

7.1. Estimation Results

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

Considering now the estimated autoregressive parameters, $\phi_{1,i}$ and $\phi_{2,i}$, according to the estimates until 1986, the inflation process had a high *intrinsic* persistence ($\phi_{1,3} + \phi_{2,3} = 0.846 \simeq \phi_{1,2} + \phi_{2,2} = 0.837$), but it has fallen ever since. These results are consistent with the findings in Cogley and Sargent (2002) (see also Brainard and Perry, 2000; Taylor, 2000). With respect to the variance parameter σ_i, we see that the volatility of the innovation was relatively high during the decade 1976–1986 ($\sigma_2 = 2.30$) and it has fallen slightly in the last 30 years ($\sigma_1 = 2.160$). However, it did not go back to the relatively low level before 1976 ($\sigma_3 = 1.077$). This is probably due to the fact that the last period included the turmoil of the financial crisis that started in 2005 (see, e.g., Stock and Watson, 2009).

Our estimated model confirms that changes in inflation dynamics can be explained by changes in the drift, the *intrinsic* persistence, and the variance parameter. To summarize our results, we find evidence that the parameters in the models capturing persistence change over time. Therefore, not allowing for time-varying

memory (see also Granger and Hyung, 2004). In the same vein, Perron and Qu (2006) show how a stationary short memory process with level shifts can generate spurious long memory. In this respect, Sun and Phillips (2004) propose a robust semiparametric estimator that explicitly allows for the presence of short memory noise in the data (see also Phillips, 2007). Since the suggested Whittle estimator makes it possible to separate low-frequency component from high-frequency behavior, it may have better forecasting properties with respect to the model in eq. (32). The use of fractional integration (and especially a time-varying type) is certainly promising and it will be considered in future work.

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

parameters in the estimation procedure would result in a less accurate modeling of the inflation process. This, in light of the simulation results in Supplementary Appendix H3, may lead to poor forecasting. Finally, the misspecification tests are reported at the bottom of Panel B. It turns out that the Breusch–Godfrey for autocorrelation does not reject the null hypothesis of no serial correlation. Similarly, the White test for heteroscedasticity does not reject the null hypothesis of homoscedasticity, therefore indicating that the model does not suffer from misspecification.

7.2. Inflation Persistence

The model presented in eq. (32) can be used as a base for a new measure of inflation persistence. In the empirical literature, a common approach for modeling inflation persistence is to estimate a univariate AR(p) specification, where the sum of the estimated autoregressive parameters is used to approximate the sluggishness with which the inflation process responds to macroeconomic shocks and/or apply unit root tests (see, e.g., Fuhrer and Moore, 1995; Angeloni et al., 2006; Devpura et al., 2021). In an influential paper, Pivetta and Reis (2007) applied a Bayesian approach to produce a time-varying measure of inflation persistence. Estimating the persistence of inflation over time using different measures and procedures is beyond the scope of this paper.[21] In this section, we depart from their study in an important way, that is, we contribute to the measurement of inflation persistence over time by taking a different approach to the problem and estimate a DAB-AR model of inflation dynamics grounded on econometric theory, and we compute an alternative measure of persistence, that is, the second-order persistence (using the methodology in Supplementary Appendix I3), which distinguishes between changes in the dynamics of inflation and its volatility (and their persistence).

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

Table 2 presents the within each period time-invariant first- and second-order measures of persistence for all three periods.

[21] 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 foresee what the persistence of inflation will be. They also estimated forward-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. In Supplementary Appendix H1, we employ a number of unit-root tests.

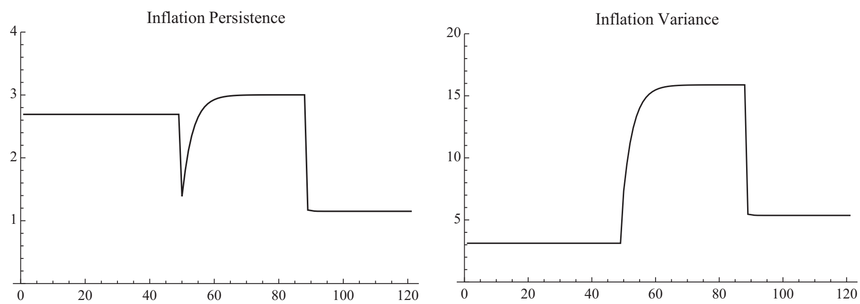
TABLE 2. Persistence for each of the three periods/models.

First and second-order measures of persistence						
Period	First-order			Second-order		
	LAR	1/(1−SUM)	$\mathbb{E}(\pi_t)$	S_0	$P_2(\pi_t\,\vert\,\varepsilon_t)$	$\mathbb{V}ar(\pi_t)$
$1964Q_2$–$1976Q_3$	0.892	6.493	3.221	7.784	2.692	3.122
$1976Q_4$–$1986Q_2$	0.858	6.135	22.313	31.688	3.002	15.881
$1986Q_3$–$2018Q_1$	0.560	0.937	2.679	0.652	1.150	5.365

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

The first three columns report the three first-order measures of persistence: LAR (largest autoregressive root), $1/(1 - SUM)$ (where SUM is the sum of the AR coefficients), and $\mathbb{E}(\pi_t)$). For the first two measures, Period 1 yields the highest persistence. In particular, the persistence (measured by $1/(1 - SUM)$) decreases by 5.5% in the post-1976 period and it decreases further by 85% in the post-1986 period. The mean of inflation, $\mathbb{E}(\pi_t)$, increases by 59.3% in the second period, and it decreases by 88% in the third period. The last three columns of Table 2 report the three second-order measures of persistence, that is, S_0,[22] $P_2(\pi_t\,|\,\varepsilon_t) \overset{\text{def}}{=} \frac{\mathbb{V}ar(\pi_t)}{\mathbb{V}ar(\varepsilon_t)}$ and $\mathbb{V}ar(\pi_t)$; see Supplementary Appendices I1 and I2, and in particular Proposition I1, to see how the variance is calculated. For two out of the three measures, the post-1986 period exhibits the lowest persistence, whereas the persistence is the highest in the second period. The variance of inflation, $Var(\pi_t)$, from 1976 to 1986 is almost five times the variance of inflation of the pre-1976 period and it is almost three times the variance of the post-1986 period.

The following couple of graphs[23] depict the measures $P_2(\pi_t\,|\,\varepsilon_t)$ and $\mathbb{V}ar(\pi_t)$, reflecting the dynamics of the second-order time-varying inflation persistence.



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

Second-Order time-varying Persistence

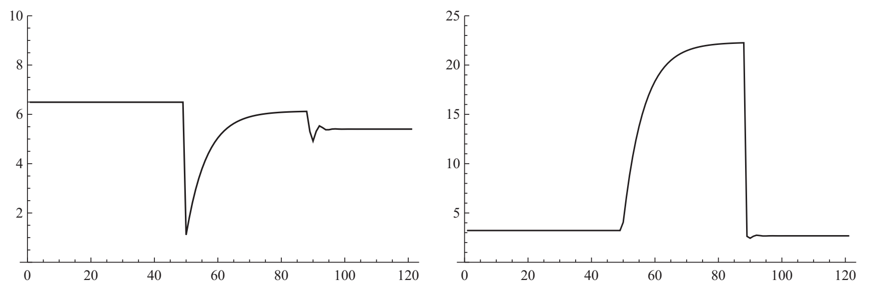
[22] Cogley and Sargent (2002) measured persistence by the spectrum at frequency zero, S_0. As an example, for the time invariant AR(2) model, this will be given by $S_0 = \frac{\sigma_\varepsilon^2}{2\pi(1-\phi_1-\phi_2)^2}$.

[23] The graphs have been designed and plotted using Mathematica drawing tools.

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

In the x-axis, each unit represents a year-quarter starting with $1964Q_2$, chosen as the first. In particular, $1976Q_3 = 49$ (49th year-quarter) and $1986Q_2 = 88$ (88th year-quarter). Some key features for the graph of the inflation variance $Var(\pi_t)$ are discussed in what follows. (i) In the pre-76 period, the graph is constant: $Var(\pi_t) = 3.122$. (ii) Within the post-76 and pre-86 period, the graph increases abruptly next to the quarter $1976Q_4$, but at a decreasing rate, in the end reaching the highest value $Var(\pi_t) = 15.881$. (iii) In the post-86 period, the graph stabilizes to $Var(\pi_t) = 5.365$, after an abrupt drop next to the quarter $1986Q_3$. Analogous statements can be addressed for the inflation persistence graph $P_2(\pi_t|\varepsilon_t)$. As illustrated above, the main difference between the shapes of the two graphs is due to the abrupt drop next to the quarter $1976Q_4$ followed shortly afterward by an abrupt increase at a decreasing rate. Details of how we construct the graphs are presented in Supplementary Appendix J.

The following couple of graphs depict the measures $P_1(\pi_t|\varepsilon_t) \stackrel{\text{def}}{=} \frac{\mathbb{E}(\pi_t)}{\varphi(t)}$ and $\mathbb{E}(\pi_t)$ for the first-order time-varying persistence.



(a) $P_1(\pi_t|\varepsilon_t)$ (b) $\mathbb{E}(\pi_t)$

First-Order time-varying Persistence

In sum, our main conclusion is that for our chosen specification (DAB-AR model), the preferred measure of persistence, that is, the second-order persistence, as measured by the $P_2(\pi_t|\varepsilon_t)$ of inflation, increased considerably from 1976 onward, whereas in the post-1986 period, the persistence falls to even lower levels than the pre-1976 period. Our results are in line with those in Cogley and Sargent (2002), who find that the persistence of inflation in the United States rose in the 1970s and remained high during this decade, before starting a gradual decline from the 1980s to the early 2000s (similar to the results of Brainard and Perry, 2000; Taylor, 2000). Stock and Watson (2002) found no evidence of a change in persistence in U.S. inflation. However, they found strong evidence of a fall in volatility. Therefore, their results are in agreement with ours.

8. CONCLUSIONS AND FUTURE WORK

The unified scheme presented in the present paper covers the subclass of stochastic processes, which belong to the large family of TV-ARMA models endowed with

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

a variable drift. In the available literature, the Green function representations of the second-order properties of this type of model, based on an analogous Wold–Cramér representation, have been extensively reported. Notwithstanding, they are not fully explicit due to the lack of an analogous representation of the Green function. Along these research lines, a theoretical gap emerges between LDEs with constant coefficients and higher-order TV-LDEs, concerning the explicit representation of their general solution. Due to the principal determinant representation of the Green function, we have filled in the aforesaid gap, which made it possible to recover and extend the fundamental properties of time-varying processes, establishing fully explicit representations for them.

Our methodology is a practical tool that can be applied to many dynamic problems. As an illustration, we constructed an AR specification with abrupt breaks, which is grounded on econometric theory. The second moment structure of this construction was employed to obtain a new time-varying measure of second-order persistence. With the help of a few examples, including smooth transition AR processes, periodic and cyclical formulations, we have demonstrated how to encompass various time series processes within our unified framework (see Supplementary Appendix D).

To summarize, we have identified a lack of an effective and broadly applicable approach to time-varying models with counterpart nonzero mean. Responding to this challenge, we explicitly obtain a solution representation of such models for any sequence of p consecutive prescribed random variables. This enables us to treat ARMA processes with variable coefficients within a unified scheme, including the cases of deterministic and stochastic ones. We further obtain a banded Hessenbergian representation of their unique Wold–Cramér decomposition, highlighting its strong connection with their asymptotic stability. We derive the second moments of these processes along with necessary and sufficient conditions for their existence, which (in the case of the deterministically varying coefficients) are prerequisites for the quasi maximum likelihood and central least squares estimation. We also present conditions for the invertibility of such processes, followed by optimal linear forecasts based on infinite and finite sets of observations. Finally, a sufficient condition for their asymptotic efficiency grounded on the boundedness of the mean square error (see Karanasos et al., 2020) is deduced.

We developed this new technique, which can be applied virtually unchanged to any "ARMA" environment, that is, to the even larger family of time-varying models, with ARMA representations (i.e., GARCH type of [or stochastic] volatility and Markov switching processes). Thus our results can be applied to TV-GARCH models too, without any significant modifications. This generic framework releases us from the need to work with characteristic polynomials and enables us to examine a variety of specifications and to solve a number of problems, helping us to deepen our familiarity with their distinctive features.

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

The usefulness of our unified theory is apparent from the fact that it enables us to analyze an abundance of models and solve a plethora of problems. In addition, an extension of the methodology developed in this paper enables us to (just to mention a few consequences) (i) examine in depth infinite-order autoregressions with either constant or variable coefficients, since it releases us from the need to work with characteristic polynomials; (ii) obtain the fourth moments of TV-GARCH models, which themselves follow linear time-varying difference equations of infinite order, taking advantage of the fact that various GARCH formulations have weak ARMA representations (see, e.g., Karanasos, 1999); (iii) work out the fundamental time series properties of time-varying linear VAR systems (since it can be easily modified and applied to a multivariate setting; see, e.g., Karanasos et al., 2014b); and (iv) derive explicit formulas for the nonnegativity constraints and the second moment structure of both constant and time-varying multivariate GARCH processes (thus extending the results in He and Teräsvirta, 2004; Conrad and Karanasos, 2010; Karanasos et al., 2022a).

Hallin (1986) applied recurrences in a multivariate context to obtain the Green matrices (see also Mélard, 2024, and the references therein). Work is at present continuing on the multivariate case. When this has been completed, one should be able to apply the methods of this paper to multivariate TV ARMA and GARCH models without any major alterations. Spectral factorization is another important problem that can be solved by our new representations.

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

APENDICES

In the appendices, we provide proofs for the statements and formulas reported earlier in the paper including some supporting material. The standard notation used in the main body of the paper is followed throughout the appendices.

A. Time-Varying ARMA

In this section, we present an autonomous procedure for the proofs of the statements of Section 3. The origins of the main tool of our analysis (principal determinant) is based on the work of Paraskevopoulos (2012, 2014), and discussed in Supplementary Appendix F1. In particular, we show there how the Hessenbergian representations of the fundamental solutions and therefore of the Green function are constructed by the infinite Gauss–Jordan elimination method.

The banded Hessenberg matrix $\mathbf{\Phi}_{t,s}$ in eq. (4) whose determinant is the so-called principal determinant is reduced to the well-known banded Toeplitz matrix, when the coefficients of eq. (1) are constants (see eq. (8)). Compact representations of banded Hessenbergians, established in Marrero and Tomeo (2012, 2017) and Paraskevopoulos and Karanasos (2021), can be applied to derive analogous representations for the principal determinant. These results endow compact representations to the fundamental properties of TV-ARMA models, presented in the main body of the paper.

A.1. A Fundamental Set of Solutions

In the literature, a fundamental set of solutions play a crucial role for the solution representation of linear difference equations with variable coefficients in terms of the Green function. Their existence is theoretically established by the fundamental theorem of linear difference equations (see Elaydi, 2005, p. 72). On account of the superposition principle (see the previously cited reference), the general solution of eq. (7) can be expressed as a linear combination of the elements of a fundamental set of solutions.

In this subsection, we define the banded Hessenbergian solutions $\xi^{(m)}(t,s)$ for $1 \leq m \leq p$ of eq. (7) (see eqs. (A.1) and (A.2)) and show that they form a fundamental set of solutions (Ξ_s). In Section 2.2, we have introduced the principal determinant $\xi(t,s)$. In this subsection, we also show that the principal determinant can also be used to generate the remaining $p-1$ elements of Ξ_s (see Section 3.1). In particular, we derive the formula in eq. (A.3), which shows that every fundamental solution $\xi^{(m)}(t,s)$ can be exclusively expressed in terms of the principal determinant $\xi(t,s)$ and the autoregressive coefficients.

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

$$\mathbf{\Phi}^{(m)}_{t,s} = \tag{A.1}$$

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

$\mathbf{\Phi}^{(m)}_{t,s}$ for $m \geq 2$, is derived from $\mathbf{\Phi}_{t,s}$ (see eq. (4)), by replacing its first column with the column vector:

$$\left(\phi_m(s+1), \quad \phi_{m+1}(s+2), \quad \ldots \quad , \phi_p(s+p+1-m), \quad 0, \quad \ldots, \quad 0\right)'.$$

In other words, each matrix in the sequence $\{\mathbf{\Phi}^{(m)}_{t,s}\}_{1 \leq m \leq p}$, differs from any other matrix in this sequence only in the first column. Formally, $\mathbf{\Phi}^{(m)}_{t,s}$ is a banded Hessenberg matrix of order $k = t-s$. The principal matrix $\mathbf{\Phi}_{t,s}$ (resp. principal determinant $\xi(t,s)$) is identified

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

with $\mathbf{\Phi}_{t,s}^{(1)}$ (resp. $\xi^{(1)}(t,s)$). For notational convenience, we will interchangeably use $\mathbf{\Phi}_{t,s}^{(1)}$ (resp. $\xi^{(1)}(t,s)$) in place of $\mathbf{\Phi}_{t,s}$ (resp. $\xi(t,s)$).

The sequences $\{\xi^{(m)}(t,s)\}_{t\geq s-p+1}$ for $1\leq m\leq p$, that is, the elements of the set Ξ_s, are defined by

$$\xi^{(m)}(t,s)=\begin{cases}\det(\mathbf{\Phi}_{t,s}^{(m)}), & \text{if } t>s,\\ 1, & \text{if } t=s+1-m,\\ 0, & \text{if } s+1-p\leq t\leq s \text{ and } t\neq s+1-m.\end{cases} \quad \textbf{(A.2)}$$

Applying eq. (A.2) with $s=t-1$, we conclude that $\xi^{(m)}(t,t-1)=\phi_m(t)$.

In Lemma A1(i) (resp. (ii)), we give the cofactors of the elements of the first column (resp. last row) of $\mathbf{\Phi}_{t,s}^{(m)}$ (see, for a proof, Paraskevopoulos and Karanasos, 2021, Lem. 1).

LEMMA A1. i) *The cofactor of the coefficient* $\phi_{m+n}(s+1+n)$ *in the first column of* $\mathbf{\Phi}_{t,s}^{(m)}$ *coincides with* $\xi(t,s+1+n)$ *for* $n=0,...,p-m$.

ii) *The cofactor of the coefficient* $\phi_n(t)$*, in the last row of* $\mathbf{\Phi}_{t,s}^{(m)}$*, coincides with* $\xi^{(m)}(t-n,s)$*,* $n=1,...,p$.

The following proposition is a direct consequence of Lemma A1 (for further details, see the previously cited reference).

PROPOSITION A1. i) *The cofactor expansion of* $\xi^{(m)}(t,s)$ *along the first column of* $\mathbf{\Phi}_{t,s}^{(m)}$ *is given by*

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

which coincides with the expression of $\xi^{(m)}(t,s)$ *in eq. (9).*

ii) *The cofactor expansion of* $\xi^{(m)}(t,s)$ *along the last row of* $\mathbf{\Phi}_{t,s}^{(m)}$ *gives*

$$\xi^{(m)}(t,s)=\sum_{n=1}^{p}\phi_n(t)\xi^{(m)}(t-n,s). \quad \textbf{(A.4)}$$

Eq. (A.4) entails that the sequence $\xi^{(m)}(t,s)$, $t\geq s+1-p$ is the solution of eq. (7) under the initial values $\xi^{(m)}(s+1-m,s)=1$ and $\xi^{(m)}(s+1-n,s)=0$, whenever $n=1,2,...,p$ and $n\neq m$ (for a proof, see Paraskevopoulos and Karanasos, 2021, Prop. 2). The linear independence of the p solution sequences $\xi^{(m)}(t,s)$, $t\geq s+1-p$, for $1\leq m\leq p$, is verified in the following proposition.

PROPOSITION A2. *For any arbitrary but fixed* $s\in\mathbb{Z}$*, the set*

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

is a fundamental set of solutions associated with eq. (7).

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

Proof. For each fixed s, consider the sequence matrices in t associated with the set Ξ_s:

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

The definition in eq. (A.2) entails that the matrix $\boldsymbol{\Xi}_{s,s}$ is the identity matrix of order p. Therefore, Ξ_s is $|\boldsymbol{\Xi}_{s,s}| = 1 \neq 0$. It turns out that $|\boldsymbol{\Xi}_{t,s}| \neq 0$ for all $t \geq s$ and the set Ξ_s is linearly independent (see Elaydi, 2005, Cor. 2.14. p. 69). Moreover, as the dimension of the homogeneous solution space of eq. (7) is p, the set Ξ_s is a fundamental set of solutions associated with eq. (7). □

$\boldsymbol{\Xi}_{t,s}$ is known as the Casorati matrix associated with the fundamental solution set Ξ_s, that is, $\boldsymbol{\Xi}_{t,s}$ coincides with the product of companion matrices, first expressed in a fully explicit form in earlier versions of the paper (see, e.g., Karanasos et al., 2020). This matrix form plays a central role both in the explicit representation of the Green function (see Paraskevopoulos and Karanasos, 2021) as well as in the explicit solution representation in the vector case, as illustrated in Supplementary Appendix F3.

A.2. Homogeneous and Particular Solutions

The following proposition provides an expression of the homogeneous solution as a linear combination of the fundamental solutions, the coefficients of which are the initial condition values.

PROPOSITION A3. *The solution of eq. (7), taking on the prescribed initial values* $\{y_{s+1-m}\}_{0\leq m\leq p}$, *is given by eq. (10).*

Proof. As Ξ_s, defined in Proposition A2, is a fundamental set of solutions, every solution of eq. (7) can be expressed as $y^{hom}_{t,s} = \sum_{m=1}^{p} a_m \xi^{(m)}(t,s)$. Fixing the initial values at $y_{s+1-m} = c_m$, for $m = 1,2,...,p$, it remains to show that $c_m = a_m$ for all $m: \ 1 \leq m \leq p$. Taking into account that

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

for each m such that $1 \leq m \leq p$, we have

$$c_m = y_{s+1-m} = y^{hom}_{s+1-m,s} = \sum_{r=1}^{p} a_r \xi^{(r)}(s+1-r,s) = a_m \xi^{(m)}(s+1-m,s) = a_m.$$

This completes the proof of Proposition A3. □

Recalling that $\upsilon_r = \varphi(r) + u_r$, the following proposition provides a particular solution of eq. (3).

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

Proposition A4. *The expression of* $y_{t,s}^{par}$ *in eq. (12) is a particular solution of eq. (3) taking on zero initial values, that is,* $y_{s+1-i}=0$ *for* $1\le i\le p$ *and it can be expressed as a single Hessenbergian:*

$$y_{t,s}^{par}= \tag{A.5}$$

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

Proof. Let us call $y_{t,s}^*$ the determinant on the right-hand side of eq. (A.5). In order to show that $y_{t,s}^{par}=\sum_{r=s+1}^{t}\xi(t,r)\upsilon_r$ solves eq. (3) subject to zero initial values, it suffices to show the following statements: (*i*) $y_{t,s}^*$ solves eq. (3), (*ii*) $y_{t,s}^{par}=y_{t,s}^*$, and (*iii*) the solution $y_{t,s}^{par}$ holds for zero initial conditions, that is, whenever $y_{s+1-m}=0$ for all $1\le m\le p$. In what follows, we sketch the main steps for the proof of the abovementioned statements (for further details, see Paraskevopoulos and Karanasos, 2021).

i) Working with elementary properties of determinants (similarly to the argument in Lemma A1(ii)), we first give the cofactors of the elements occupying the last row of the determinant $y_{t,s}^*$. At this aim, we consider the following two cases:

i_a) The cofactor of υ_t for the determinant $y_{t,s}^*$ is $y_{s,s}^*=1$.
i_b) The remaining nonzero entries of the last row of $y_{t,s}^*$ are the AR coefficients $\phi_m(t)$ for $1\le m\le p$, each of which has $y_{t-m,s}^*$ as its cofactor.

The cofactor expansion of $y_{t,s}^*$ along the last row gives

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

This shows that $y_{t,s}^*$ satisfies eq. (3).

ii) Working along the first column of $y_{t,s}^*$, the cofactor of υ_{s+i} is $\xi(t,s+i)$ for $1\le i\le k$. Thus, the cofactor expansion of $y_{t,s}^*$ along the first column yields

$$y_{t,s}^*=\sum_{r=s+1}^{t}\xi(t,r)\upsilon_r=y_{t,s}^{par},$$

as required.

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

iii) Having proved eq. (A.5) in (*ii*), we apply it for $t = s + r$, $r = 1, ..., p$ to get

$$y^{par}_{s+r,s} = \sum_{m=1}^{r-1} \phi_m(s+r) y^{par}_{s+r-m,s} + \upsilon_{s+r}.$$

We can write

$$y^{par}_{s+i,s} = \sum_{m=1}^{i-1} \phi_m(s+i) y^{par}_{s+i-m,s} + \upsilon_{s+i} + \sum_{m=i}^{p} \phi_m(s+i) y_{s+i-m},$$

whenever

$$\sum_{m=i}^{p} \phi_m(s+i) y_{s+i-m} = 0,$$

for all i, m such that $1 \leq i \leq m \leq p$. Since $s + i - m \leq s$ for any i, m such that $1 \leq i \leq m \leq p$, it follows that y_{s+i-m} are initial condition values. The latter equation can be expressed as

$$\begin{bmatrix} \phi_p(s+p) & 0 & 0 & \ldots & 0 \\ \phi_{p-1}(s+p-1) & \phi_p(s+p-1) & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \phi_1(s+1) & \phi_2(s+1) & \phi_3(s+1) & \ldots & \phi_p(s+1) \end{bmatrix} \begin{bmatrix} y_s \\ y_{s-1} \\ y_{s-2} \\ \vdots \\ y_{s+1-p} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

which is a triangular linear system with nonzero elements in the main diagonal and therefore nonsingular. It follows directly that $\{y_{s+1-p} = 0, ..., y_s = 0\}$ is a solution. Moreover, since the system is nonsingular, the solution is unique. Finally, expanding the determinant in eq. (A.5) along the first column, the expression in eq. (12) follows immediately. □

A.3. Decomposition

In this subsection, we prove Proposition 1, which provides the reported decomposition of the innovation part of eq. (1).

Proof of Proposition 1. Let us write u_r in eq. (1) as $u_r = \sum_{l=0}^{q} \theta_l(r) \varepsilon_{r-l}$, provided that $\theta_0(r) \overset{\text{def}}{=} 1$ for all t. The left side of eq. (15) can be expressed as

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

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

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

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

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

Substituting the above sum into eq. (A.6), we get

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

or equivalently

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

Using the definition of $\xi_q(t,r)$ in eq. (13), eq. (A.7) can be rewritten as

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

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

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

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

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

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

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

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

Substituting the defining formula of $\xi_{s,q}(t,r)$ (see eq. (14)) into eq. (A.10), the latter takes the form

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

that is, eq. (15), as required. □

B. Second-Order Structure

In this section, we show the Cramér–Wold decomposition of DTV-ARMA(p,q) processes along with their second-order properties, reported in Section 4.1.

B.1. Wold–Cramér Decomposition

In Lemma B1 and Proposition B1 below, we provide some essential statements for the proof of the main results of this subsection given by Theorems 2 and 3. We remark that, for each fixed $t \in \mathbb{Z}$, the deterministic drift $\varphi(t)$ along with the coefficients $\phi_m(t)$ and $\theta_l(t)$ are constants in $\mathbb{R}$, since they are values of corresponding functions, while ε_t, y_t, u_t are all random variables. Since the results valid for the process y_t under the conditions of Theorem 3 are more general, they are also valid for the process y_t under the conditions of Theorem 2.

LEMMA B1. *Let the condition (17) and the boundedness of the MA coefficients hold. Then the following statements hold true:*

i) $\sum_{r=-\infty}^{t} |\xi_q(t,r)| < \infty$ *for all t.*

ii) *Assuming in addition that the drift is a bounded real valued function in* $t \in \mathbb{Z}$, *then the first infinite sum of* y_t *in eq. (18a) converges in* $\mathbb{R}$ *and its expectation is*

$$\mathbb{E}\Big(\sum_{r=-\infty}^{t} \xi(t,r)\varphi(r)\Big) = \sum_{r=-\infty}^{t} \xi(t,r)\varphi(r) \in \mathbb{R} \text{ (finite) } \textit{for all } t \in \mathbb{Z}.$$

iii) *Recalling that* $\upsilon_r = \varphi(r) + u_r$ *(see eq. (2)), we have* $\mathbb{E}(\upsilon_r) = \varphi(r)$. *Additionally, if the drift is a bounded real-valued function in* $t \in \mathbb{Z}$, *then* $\{\upsilon_r\}$ *is a bounded process in* L_2, *that is,* $\sup_r \mathbb{E}(\upsilon_r^2) \le V < \infty$ *for* $V \in \mathbb{R}_{>0}$.

A proof of Lemma B1 is deferred to Supplementary Appendix F2.1. As the absolute summability implies square summability, Corollary B1 below follows directly.

COROLLARY B1. *If the conditions of Lemma B1 hold, then*

$$\sum_{r=-\infty}^{t} \xi_q^2(t,r) < \infty \;\; \textit{for all} \;\; t \in \mathbb{Z}. \tag{B.1}$$

In the next proposition, we present some results, which support the statements reported in Section 4.1 including those involved in the proof of Theorem 2. These results are supplemented by Proposition F.1 in Supplementary Appendix F2.2, including those involved in the proof of Theorem 3. Remark F.1 in Supplementary Appendix F2.2 summarizes some well-known results used in the proofs of Proposition B1 and Theorem 3 below. The uncorrelatedness assumption (see Section 2) entails that $\{\varepsilon_s, s \le t\}$ is an orthogonal basis of $\mathcal{M}_t(\varepsilon)$.

PROPOSITION B1. *Let* $e_r = \varepsilon_r ||\varepsilon_r||_{L_2}^{-1}$. *The following statements hold true:*

i) *If* $\sum_{r=-\infty}^{t} \xi_q^2(t,r)\sigma^2(r) < \infty$, *then* $\boldsymbol{e}_t = \{e_r\}_{r \le t}$ *is an orthonormal basis of* $\mathcal{M}_t(\varepsilon)$ *and the Fourier representation of* z_t *in terms of* $\boldsymbol{e}_t$ *is given by* $\sum_{r=-\infty}^{t} \xi_q(t,r)\sigma(r)e_r$ *with*

$$\mathbb{E}(z_t^2) = \sum_{r=-\infty}^{t} \xi_q^2(t,r)\sigma^2(r).$$

Moreover, for every arbitrary but fixed $t \in \mathbb{Z}$, *the process* $z_t = \sum_{r=-\infty}^{t} \xi_q(t,r)\varepsilon_r$ *exists in* $\mathcal{M}_t(\varepsilon)$ *if and only if* $\sum_{r=-\infty}^{t} \xi_q^2(t,r)\sigma^2(r) < \infty$.

ii) *Let* $0 < m \le \sigma^2(t) \le M < \infty$ *for all t. Then the weaker condition in eq. (B.1), compared to the above condition in (i) is necessary and sufficient for* z_t *to exist in* $\mathcal{M}_t(\varepsilon)$.

iii) *If the conditions of Theorem 3 hold, then* z_t *also exists in* $\mathcal{M}_t(\varepsilon)$ *and* $\mathbb{E}(z_t) = 0$ *for all t. The unconditional variance of the processes* y_t *in eq. (20a) is given by eq. (21b).*

iv) *If the conditions of Theorem 2 hold, the process* y_t *in eq. (18a) can be rewritten as* $y_t = \sum_{r=-\infty}^{t} \xi(t,r)\upsilon_r$ *and its unconditional mean can be explicitly expressed by eq. (21a).*

v) *Both processes in eqs. (18a) and (20a) have the same unconditional variance given in eq. (21b), which also coincides with the unconditional variance of* z_t.

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

Proof. i) Since $\mathbb{E}(\varepsilon_t^2) = \sigma^2(t) > 0$, it turns out that $||\varepsilon_t||_{L_2} \neq 0$ for all t, whence $\varepsilon_t \neq 0$. By the definition of e_t, we infer $||e_t||_{L_2} = 1$. It follows that $\mathbb{E}(e_r e_j) = \begin{cases} 0, & \text{if } r \neq j \\ 1, & \text{if } r = j \end{cases}$ and $\mathbf{e}_t = \{e_r\}_{r \leq t}$ also spans $\mathcal{M}_t(\varepsilon)$. Taking into account that a closed linear subspace of a Hilbert space is a separable Hilbert space, we conclude that the sequence $\mathbf{e}_t$ is an orthonormal basis of $\mathcal{M}_t(\varepsilon)$. On account of

$$\xi_q(t,r)\varepsilon_r = ||\varepsilon_r||_{L_2}\xi_q(t,r)e_r = \sigma(r)\xi_q(t,r)e_r,$$

it follows directly that z_t can be rewritten as

$$z_t = \sum_{r=-\infty}^{t} \sigma(r)\xi_q(t,r)e_r.$$

Thereby $\sigma(r)\xi_q(t,r)$ are the unique Fourier coefficients of the representation of z_t in terms of the orthonormal basis $\mathbf{e}_t$. Accordingly, $||z_t||_{L_2}^2 = \mathbb{E}(z_t^2) = \sum_{r=-\infty}^{t} \xi_q^2(t,r)\sigma^2(r)$. The equivalence

$$z_t \in \mathcal{M}_t(\varepsilon) \iff \sum_{r=-\infty}^{t} \xi_q^2(t,r)\sigma^2(r) < \infty$$

follows immediately (see Remark F.1 in Supplementary Appendix F).

ii) *Sufficiency*: The following implications show that the condition in (B.1) is sufficient for $z_t \in \mathcal{M}_t(\varepsilon)$:

$$\begin{aligned} \sum_{r=-\infty}^{t} \xi_q^2(t,r) < \infty &\implies M \sum_{r=-\infty}^{t} \xi_q^2(t,r) < \infty \\ &\implies \sum_{r=-\infty}^{t} \sigma^2(r)\xi_q^2(t,r) \leq \sum_{r=-\infty}^{t} M\xi_q^2(t,r) < \infty \\ &\implies z_t \in \mathcal{M}_t(\varepsilon). \end{aligned}$$

Necessity: As $0 < m \leq \sigma^2(r)$ for all r with $r \leq t$, the following implications hold:

$$\begin{aligned} z_t \in \mathcal{M}_t(\varepsilon) &\implies \mathbb{E}(z_t^2) = \sum_{r=-\infty}^{t} \xi_q^2(t,r)\sigma^2(r) < \infty \\ &\implies m \sum_{r=-\infty}^{t} \xi_q^2(t,r) \leq \sum_{r=-\infty}^{t} \xi_q^2(t,r)\sigma^2(r) < \infty. \end{aligned}$$

On account of $m \neq 0$, it follows from $m\sum_{r=-\infty}^{t} \xi_q^2(t,r) < \infty$ that $\frac{1}{m}(m\sum_{r=-\infty}^{t} \xi_q^2(t,r)) < \infty$, that is, $\sum_{r=-\infty}^{t} \xi_q^2(t,r) < \infty$, as required.

iii) As the absolute summability implies square summability (see Corollary B1), z_t also exists in $\mathcal{M}_t(\varepsilon)$ for all $t \in \mathbb{Z}$. Moreover, as the absolute summability is sufficient for

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

switching expectation with infinite summation, on account of $\mathbb{E}(\varepsilon_t) = 0$ for all $t \in \mathbb{Z}$, the first-order moment of z_t follows from

$$\mathbb{E}(z_t) = \mathbb{E}\left(\sum_{r=-\infty}^{t} \xi_q(t,r)\varepsilon_r\right) = \sum_{r=-\infty}^{t} \xi_q(t,r)\mathbb{E}(\varepsilon_r) = 0.$$

Moreover, the expression of the unconditional variance and of the second-order moment of z_t coincides and therefore, by statement (i), is given by

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

iv) As $r+l > t$ for all l, r such that $1 \leq l \leq q$ and $t-l+1 \leq r \leq t$, it follows that $\xi(t, r+l) = 0$. In the case, when $l = 0$, it follows that $\sum_{r=t+1}^{t} \xi(t, r+l)\theta_l(r+l)\varepsilon_r = \sum_{r=t+1}^{t} \xi(t,r)\varepsilon_r = 0$ too. Therefore, we conclude that

$$\sum_{l=0}^{q} \sum_{r=t-l+1}^{t} \xi(t, r+l)\theta_l(r+l)\varepsilon_r = 0. \tag{B.2}$$

Call $\tilde{\theta}_l = \sup_r |\theta_l(l+r)| \in \mathbb{R}_{>0}$ for each $l = 1, ..., q$. In what follows, we shall use the notation: $\Theta = \max_{0 \leq l \leq q} \tilde{\theta}_l$. As the absolute summability condition in (17) implies that $\sum_{r=-\infty}^{t} |\xi(t, r+l)| < \infty$ (see eq. (F.3) in Supplementary Appendix F2.1), we conclude that

$$\sum_{r=-\infty}^{t} |\xi(t, r+l)\theta_l(r+l)| \leq \Theta \sum_{r=-\infty}^{t} |\xi(t, r+l)| < \infty.$$

Starting with the definition of $\xi_q(t,r)$ in eq. (13), the following chain of equalities holds:

$$\begin{aligned}
z_t = \sum_{r=-\infty}^{t} \xi_q(t,r)\varepsilon_r \quad &= \sum_{r=-\infty}^{t} \sum_{l=0}^{q} \xi(t, r+l)\theta_l(r+l)\varepsilon_r \\
\text{(switching summation)} \quad &= \sum_{l=0}^{q} \sum_{r=-\infty}^{t} \xi(t, r+l)\theta_l(r+l)\varepsilon_r \\
\text{(spliting summation)} \quad &= \sum_{l=0}^{q} \sum_{r=-\infty}^{t-l} \xi(t, r+l)\theta_l(r+l)\varepsilon_r \\
&\quad + \sum_{l=0}^{q} \sum_{r=t-l+1}^{t} \xi(t, r+l)\theta_l(r+l)\varepsilon_r \\
\text{(removing zero terms, see eq. (B.2))} \quad &= \sum_{l=0}^{q} \sum_{r=-\infty}^{t-l} \xi(t, r+l)\theta_l(r+l)\varepsilon_r \\
\text{(changing the summation limits)} \quad &= \sum_{l=0}^{q} \sum_{r=-\infty}^{t} \xi(t,r)\theta_l(r)\varepsilon_{r-l}
\end{aligned}$$

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

$$\text{(switching summation)} \quad = \sum_{r=-\infty}^{t} \sum_{l=0}^{q} \xi(t,r)\theta_l(r)\varepsilon_{r-l}$$

$$\text{(by definition of } u_r) \quad = \sum_{r=-\infty}^{t} \xi(t,r)u_r.$$

The above results are summarized by

$$z_t = \sum_{r=-\infty}^{t} \xi(t,r)u_r \ \ (\text{in } L_2). \tag{B.3}$$

Let us call $x_t = \sum_{r=-\infty}^{t} \xi(t,r)\upsilon_r$. It follows from eq. (B.3) that

$$x_t = \sum_{r=-\infty}^{t} \xi(t,r)\upsilon_r = \sum_{r=-\infty}^{t} \xi(t,r)(\varphi(r)+u_r) = \sum_{r=-\infty}^{t} \xi(t,r)\varphi(r) + \sum_{r=-\infty}^{t} \xi(t,r)u_r$$
$$= \sum_{r=-\infty}^{t} \xi(t,r)\varphi(r) + \sum_{r=-\infty}^{t} \xi_q(t,r)\varepsilon_r = y_t \ (\text{in } L_2).$$

It turns out that $y_t = x_t$ in L_2, as claimed. Finally, as $\mathbb{E}(z_t) = 0$, taking the expectations on both sides of eq. (18a), on account of the infinite sum convergence in Lemma B1(ii), the unconditional expectation of y_t in eq. (21a) follows.

(*v*) As the y_ts either in eq. (18a) or in eq. (20a) differ from z_t in $\mathbb{E}(y_t)$, the unconditional variances of y_ts are identical with $\mathbb{V}ar(z_t)$. Hence, eq. (21b) follows. This completes the proof of the proposition. □

Proof of Theorem 2. Under the conditions of Theorem 2, we have established in Lemma B1(ii) and Proposition B1(iii) the existence of both terms of y_t in eq. (18a), respectively. It remains to be shown that the process y_t in eq. (18a) is (i) a solution of eq. (1), (ii) asymptotically stable, and (iii) unique.

i) The compact representation of y_t, that is, $y_t = \sum_{r=-\infty}^{t} \xi(t,r)\upsilon_r$, is used to show that $\{y_t\}$ solves eq. (1). Applying the expression $y_{t-m} = \sum_{r=-\infty}^{t-m} \xi(t-m,r)\upsilon_r$ for $m = 0, 1, ..., p$ to eq. (3), it suffices to show that

$$\sum_{r=-\infty}^{t} \xi(t,r)\upsilon_r = \sum_{m=1}^{p} \phi_m(t) \sum_{r=-\infty}^{t-m} \xi(t-m,r)\upsilon_r + \upsilon_t. \tag{B.4}$$

Let $p = 1$. Applying eq. (A.4) for $p = m = 1$, it takes the form $\xi(t,r) = \phi_1(t)\xi(t-1,r)$. Starting with the right-hand side of eq. (B.4), we have

$$\phi_1(t) \sum_{r=-\infty}^{t-1} \xi(t-1,r)\upsilon_r + \upsilon_t = \sum_{r=-\infty}^{t-1} \Big(\phi_1(t)\xi(t-1,r)\Big)\upsilon_r + (1\ \upsilon_t)$$
$$= \sum_{r=-\infty}^{t-1} \xi(t,r)\upsilon_r + \xi(t,t)\upsilon_t$$
$$= \sum_{r=-\infty}^{t} \xi(t,r)\upsilon_r,$$

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

as required. It remains to establish that eq. (B.4) holds for $p \geq 2$. If $t-m+1 \leq r \leq t-1$ and $2 \leq m \leq p$, then $r > t-m$, which implies that $\xi(t-m,r) = 0$. Moreover, if $m = 1$, it follows that $\sum_{r=t}^{t-1} \phi_m(t)\xi(t-m,r)\upsilon_r = 0$. Therefore, we conclude that

$$\sum_{m=1}^{p} \sum_{r=t-m+1}^{t-1} \phi_m(t)\xi(t-m,r)\upsilon_r = 0. \tag{B.5}$$

The following equalities hold:

$$\begin{aligned}
\sum_{m=1}^{p} \phi_m(t) \sum_{r=-\infty}^{t-m} \xi(t-m,r)\upsilon_r &= \sum_{m=1}^{p} \sum_{r=-\infty}^{t-m} \phi_m(t)\xi(t-m,r)\upsilon_r \\
\text{(adding some zero terms, see eq. (B.5))} &= \sum_{m=1}^{p} \sum_{r=-\infty}^{t-m} \phi_m(t)\xi(t-m,r)\upsilon_r + \sum_{m=1}^{p} \sum_{r=t-m+1}^{t-1} \phi_m(t)\xi(t-m,r)\upsilon_r \\
\text{(condensed sum)} &= \sum_{m=1}^{p} \sum_{r=-\infty}^{t-1} \phi_m(t)\xi(t-m,r)\upsilon_r \\
\text{(switching summation)} &= \sum_{r=-\infty}^{t-1} \left(\sum_{m=1}^{p} \phi_m(t)\xi(t-m,r) \right) \upsilon_r \\
\text{(applying eq. (A.4) for m = 1)} &= \sum_{r=-\infty}^{t-1} \xi(t,r)\upsilon_r = \sum_{r=-\infty}^{t} \xi(t,r)\upsilon_r - \upsilon_t.
\end{aligned}$$

This shows that eq. (B.4) holds (in L_2 sense), as required.

ii) Taking into account that $y_{t,s}^{hom} = y_t - y_{t,s}^{par}$ and that the solution process y_t in eq. (18a) can be equivalently expressed as

$$y_t = \lim_{s \to -\infty} y_{t,s}^{par}$$

(see eq. (18b)), the following equivalences hold:

$$\begin{aligned}
y_t = \lim_{s \to -\infty} y_{t,s}^{par} &\iff \lim_{s \to -\infty} ||y_t - y_{t,s}^{par}||_{L_2} = 0 \\
&\iff \lim_{s \to -\infty} ||y_{t,s}^{hom}||_{L_2} = 0 \iff \lim_{s \to -\infty} y_{t,s}^{hom} \overset{L_2}{=} 0,
\end{aligned}$$

that is, $\{y_t\}$ is asymptotically stable, as required.

iii) Let $\{x_t\}$ be an asymptotically stable stochastic process in L_2, which solves eq. (1). Then we shall show that x_t coincides with y_t in eq. (18a) for all t. In view of eq. (16), we can write x_t in a more condensed form as

$$x_t = \sum_{m=1}^{p} \xi^{(m)}(t,s)x_{s+1-m} + \sum_{r=s+1}^{t} \xi(t,r)\upsilon_r,$$

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

for each t, s with $s < t$. Thus,

$$||x_t - \sum_{r=s+1}^{t} \xi(t,r)\upsilon_r||_{L_2} = ||\sum_{m=1}^{p} \xi^{(m)}(t,s)x_{s+1-m}||_{L_2} \text{ for all } s: s < t. \tag{B.6}$$

Taking into account that $x_{t,s}^{hom} = \sum_{m=1}^{p} \xi^{(m)}(t,s)x_{s+1-m}$, it follows from the asymptotic stability assumption on $\{x_t\}$ that $\lim_{s\to-\infty} ||\sum_{m=1}^{p} \xi^{(m)}(t,s)x_{s+1-m}||_{L_2} = 0$. Thus, taking the limits to both sides of eq. (B.6) as $s \to -\infty$, it follows that

$$\lim_{s\to-\infty} ||x_t - \sum_{r=s+1}^{t} \xi(t,r)\upsilon_r||_{L_2} = 0,$$

that is, $\lim_{s\to-\infty} \sum_{r=s+1}^{t} \xi(t,r)\upsilon_r = x_t$ or equivalently $\sum_{r=-\infty}^{t} \xi(t,r)\upsilon_r = x_t$ in L_2. Proposition B1(iv) entails that $y_t = \sum_{r=-\infty}^{t} \xi(t,r)\upsilon_r$, whence $x_t = y_t$ in L_2, as asserted. This completes the proof of the theorem. □

COROLLARY B2. *Let the conditions of Theorem 2 hold. Then the last summation in eq. (16) tends to zero in L_2, as $s \to -\infty$, that is,*

$$\lim_{s\to-\infty} \sum_{r=s+1-q}^{s} \xi_{s,q}(t,r)\varepsilon_r \overset{L_2}{=} 0.$$

A proof of Corollary B2 is a combination of eqs. (15) and (B.3) (see Supplementary Appendix F2.1).

Proof of Theorem 3. Let $\{\mu_t\}$ be any estimated first moment solution process generated by eq. (19b) and $\{y_t\}$ be defined by eq. (20a). Recalling that $z_t = y_t - \mathbb{E}(y_t)$, we shall show that (i) the first moment of y_t is $\mathbb{E}(y_t) = \mu_t$ for all t, (ii) $\{y_t\}$ is second order, (iii) $\{y_t\}$ solves eq. (1), and (iv) $\{y_t\}$ is unique.

i) It follows from $\sum_{r=-\infty}^{t} \xi_q(t,r) < \infty$ (see Lemma B1(i)) that the expectation operator, $\mathbb{E}$, commutes with infinite summation. Use the latter result along with the linearity of $\mathbb{E}$ and the fact that $\mathbb{E}(\varepsilon_r) = 0$ to conclude that

$$\mathbb{E}(z_t) = \sum_{r=-\infty}^{t} \xi_q(t,r)\mathbb{E}(\varepsilon_r) = 0.$$

Apply $\mathbb{E}$ to both sides of eq. (20a) to conclude that $\mathbb{E}(y_t) = \mu_t$, as required.

ii) As $\{\mu_t\}$ is deterministic and $z_t \in \mathcal{M}_t(\varepsilon) \subset L_2$ (see Proposition F.1(ii) in Supplementary Appendix F2.2), it follows that $y_t \in L_2$.

iii) It is established in Proposition F.1(ii) that $\{z_t\}$ solves eq. (19a). Since $\{\mathbb{E}(y_t)\}$ solves eq. (19b), it follows from Proposition F.1(i) that $\{y_t\}$ in eq. (20a) solves eq. (1).

iv) To see the uniqueness of $\{y_t\}$, consider any other second-order process, say $\{y_t^*\}$, which solves eq. (1) with $\mathbb{E}(y_t^*) = \mu_t$. It follows from Proposition F.1(i) that $\{y_t^* - \mathbb{E}(y_t^*)\}$ is a mean-zero process which solves eq. (19a). Moreover, it follows from the uniqueness of $\{z_t\}$ (see Proposition F.1(ii)), that $y_t^* - \mathbb{E}(y_t^*) = z_t$. Taking into account that $\mathbb{E}(y_t^*) = \mu_t = \mathbb{E}(y_t)$, we conclude that $y_t^* = \mathbb{E}(y_t) + z_t$, which coincides with y_t (in the L_2 sense). This completes the proof of Theorem 3. □

Any second-order solution process of eq. (1) can be expressed as $y_t = \mathbb{E}(y_t) + z_t$, where $z_t \in \mathcal{M}_t(\varepsilon)$. As the inner product $\langle z_t, \mathbb{E}(y_t)\rangle = \mathbb{E}(z_t\,\mathbb{E}(y_t)) = \mathbb{E}(y_t)\mathbb{E}(z_t) = 0$, it follows that $y_t - z_t = \mathbb{E}(y_t) \in \mathcal{M}_t^{\perp}(\varepsilon)$.

B.2. Unconditional Moments

In this subsection, we give a proof for the existence of the first and second unconditional moments, described in Proposition 2, completed by the logical implications that render the associated diagrams commutative.

Proof of Proposition 2. The explicit form for the first unconditional moment of y_t in eq. (18a) has already been proven in Lemma B1(ii). Moreover, the unconditional variance of y_t is shown in Proposition B1(iii). The expressions of the two unconditional moments imply the existence of the following limits $\lim_{s\to-\infty}\xi(t,s)\varphi(s) = 0$ and $\lim_{s\to-\infty}\xi_q^2(t,s)\sigma^2(s) = 0$, respectively, which, in turn, are necessary conditions for the existence of these moments. Finally, in order to show that the associated diagrams commute, it remains to verify the following implications:

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

The first implication in the diagram (B.7) is well known. As the product of a null sequence $(\xi(t,s))_s$ times a bounded sequence $\varphi(s)$ is also a null sequence, the upper branch implication in diagram (B.7) follows. As $\theta_l(t)$ are bounded, it follows that $|\xi_q(t,r)| \le \Theta\sum_{l=0}^{q}|\xi(t,r+l)|$, for some $\Theta \in \mathbb{R}_{\ge 0}$ (see eq. (F.2) in Supplementary Appendix F2.1). Therefore, the following implications hold:

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

where M is an upper bound of $\sigma^2(t)$. It follows from the squeeze theorem that

$$0 \le \lim_{s\to-\infty}\xi_q^2(t,s)\sigma^2(s) \le \lim_{s\to-\infty}M\xi_q^2(t,s) = 0.$$

The latter establishes the lower branch implication in diagram (B.7) and the proof is complete. □

B.3. Autocovariance Function

We recall the following result. The inner product of $X, Y \in L_2$ is

$$\langle X, Y\rangle = \mathbb{E}(X\,Y) = \mathbb{C}ov(X,Y) - \mathbb{E}(X)\mathbb{E}(Y).$$

Under the conditions of Theorem 3, $y_t = \mathbb{E}(y_t) + z_t$ is the second-order solution process of eq. (1). Taking into account that $\mathbb{E}(y_t)$ has no impact upon variances and covariances, it follows from $\mathbb{E}(z_t)\mathbb{E}(z_{t-\ell}) = 0$ that the autocovariance function of y_t can be rewritten as

$$\gamma_t(\ell) = \mathbb{C}ov(y_t, y_{t-\ell}) = \mathbb{C}ov(z_t, z_{t-\ell}) = \mathbb{E}(z_t z_{t-\ell}).$$

Proof of Proposition 3. As $\sum_{j=-\infty}^{t} |\xi_q(t,j)| < \infty$, it follows that z_t and $z_{t-\ell}$ exist in L_2. First step: The uncorrelatedness of ε_t implies $\mathbb{E}(\varepsilon_j \varepsilon_r) = 0$, whenever $-\infty < r \leq t-\ell$ and $j \geq t-\ell+1$, since $j \neq r$. Thereby,

$$\mathbb{E}\left(z_{t-\ell} \sum_{j=t-\ell+1}^{t} \xi_q(t,j)\varepsilon_j\right) = \sum_{r=-\infty}^{t-\ell} \sum_{j=t-\ell+1}^{t} \xi_q(t-\ell,r)\xi_q(t,j)\mathbb{E}(\varepsilon_j\varepsilon_r) = 0. \quad \textbf{(B.8)}$$

Second step: The following chain of equalities holds:

$$\mathbb{E}\left(\sum_{j=-\infty}^{t-\ell} \xi_q(t,j)\varepsilon_j \, z_{t-\ell}\right) = \mathbb{E}\left(\sum_{j=-\infty}^{t-\ell} \sum_{r=-\infty}^{t-\ell} \xi_q(t,j)\xi_q(t-\ell,r)(\varepsilon_j \, \varepsilon_r)\right)$$

$$\text{(switch twice expectation} = \sum_{j=-\infty}^{t-\ell} \sum_{r=-\infty}^{t-\ell} \xi_q(t,j)\xi_q(t-\ell,r)\mathbb{E}(\varepsilon_j \, \varepsilon_r)$$

with infinite summation)

$$\text{(uncorrelatedness)} = \sum_{r=-\infty}^{t-\ell} \xi_q(t,r)\xi_q(t-\ell,r)\mathbb{E}(\varepsilon_r^2)$$

$$\text{(model assumption)} = \sum_{r=-\infty}^{t-\ell} \xi_q(t,r)\xi_q(t-\ell,r)\sigma^2(r). \quad \textbf{(B.9)}$$

Third step: The formula in eq. (22) is a result of the following chain of equalities:

$$\gamma_t(\ell) = \mathbb{E}(z_t \, z_{t-\ell}) = \mathbb{E}\left(\sum_{j=-\infty}^{t} \xi_q(t,j)\varepsilon_j \, z_{t-\ell}\right)$$

$$\text{(spliting summation)} = \mathbb{E}\left(\left(\sum_{j=-\infty}^{t-\ell} \xi_q(t,j)\varepsilon_j + \sum_{j=t-\ell+1}^{t} \xi_q(t,j)\varepsilon_j\right) z_{t-\ell}\right)$$

$$\text{(distributive law)} = \mathbb{E}\left(\sum_{j=-\infty}^{t-\ell} \xi_q(t,j)\varepsilon_j \, z_{t-\ell} + \sum_{j=t-\ell+1}^{t} \xi_q(t,j)\varepsilon_j \, z_{t-\ell}\right)$$

$$\text{(by eq. (B.8))} = \mathbb{E}\left(\sum_{j=-\infty}^{t-\ell} \xi_q(t,j)\varepsilon_j \;\; z_{t-\ell}\right)$$

$$\text{(by eq. (B.9))} = \sum_{r=-\infty}^{t-\ell} \xi_q(t,r)\xi_q(t-\ell,r)\sigma^2(r).$$

https://doi.org/10.1017/S0266466624000306 Published online by Cambridge University Press

Let $C_t = \sum_{r=-\infty}^{t} |\xi_q(t,r)|$. Since $C_t < \infty$ for all t, we have

$$|\gamma_t(\ell)| \leq \sum_{r=-\infty}^{t-\ell} |\xi_q(t,r)\xi_q(t-\ell,r)|\sigma^2(r) \leq M \sum_{r=-\infty}^{t-\ell} |\xi_q(t,r)||\xi_q(t-\ell,r)|$$

$$\leq M \sum_{j=-\infty}^{t} |\xi_q(t,j)| \sum_{r=-\infty}^{t-\ell} |\xi_q(t-\ell,r)| \leq M C_t C_{t-l} < \infty.$$

Thus, $\gamma_t(\ell)$ exists in $\mathbb{R}$.

Finally, for each fixed t, on account of $\lim_{\ell\to\infty} C_{t-\ell} = \lim_{\ell\to\infty} \sum_{r=-\infty}^{t-\ell} |\xi_q(t-\ell,r)| = 0$, we have $\lim_{\ell\to\infty} |\gamma_t(\ell)| \leq \lim_{\ell\to\infty} MC_tC_{t-l} = 0$ for all $t \in \mathbb{Z}$. This completes the proof of Proposition 3. □

C. Stochastic Coefficients

The proofs of Propositions 5 and 6 and Theorem 5 are omitted, as they are similar to those presented for the deterministically varying coefficient case.

Proof of Theorem 6. The two summability conditions in (29) must hold for all integers t. Since all the AR coefficients are strictly stationary, the principal determinant is strictly stationary, and, therefore, the left-hand sides of the two aforementioned conditions do not change when we subtract $(t-r)$ from each index (see also Anděl, 1991). □

SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at https://doi.org/10.1017/S0266466624000306.

REFERENCES

Abdrabbo, N. A., & Priestley, M. B. (1967). On the prediction of non-stationary processes. *Journal of the Royal Statistical Society*, 29, 570–585.

Alj, A., Azrak, R., Ley, C., & Mélard, G. (2017). Asymptotic properties of QML estimators for VARMA models with time-dependent coefficients. *Scandinavian Journal of Statistics*, 44, 617–635.

Anděl, J. (1991). On stationarity of a multiple doubly stochastic model. *Kybernetica*, 2, 114–119.

Angeloni, I., Aucremanne, L., Ehrmann, M., Galí, J., Levin, A., & Smets, F. (2006). New evidence on inflation persistence and price stickiness in the euro area: Implications for macro modelling. *Journal of European Economic Association*, 4, 562–574.

Azrak, R., & Mélard, G. (2006). Asymptotic properties of quasi-maximum likelihood estimators for ARMA models with time-dependent coefficients. *Statistical Inference for Stochastic Processes*, 9, 279–330.

Bai, J., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18, 1–22.

Baillie, R. T., Han, Y. W., & Kwon, T. G. (2002). Further long memory properties of inflationary shocks. *Southern Economic Journal*, 68, 496–510.

Brainard, W. C., & Perry, G. L. (2000). Making policy in a changing world. In G. L. Perry, & J. Tobin (Eds.), *Economic events, ideas, and policies: The 1960s and after* (pp. 43–82). Brookings Institution Press.

Brockwell, P. J., & Davis, R. A. (2016). *An introduction to time series and forecasting*. Springer.
Canarella, G., & Miller, S. F. (2017). Inflation targeting: New evidence from fractional integration and cointegration. *Journal of Economics and Business*, 92, 45–62.
Canepa, A., Karanasos, M., Paraskevopoulos, A. G., & Zanetti Chini, E. (2022). *Forecasting inflation: A GARCH-in mean-level model with time-varying predictability (Publication No. 11)* [*Working paper*]. University of Bergamo.
Cavaliere, G., & Georgiev, I. (2008). Regime switching autoregressive coefficients and the asymptotics for unit root tests. *Econometric Theory*, 24, 1137–1148.
Cavaliere, G., & Taylor, A. M. R. (2005). Stationarity tests under time-varying variances. *Econometric Theory*, 21, 1112–1129.
Chambers, M. J., & Taylor, A. M. R. (2020). Deterministic parameter change models in continuous and discrete time. *Journal of Time Series Analysis*, 41, 134–145.
Cogley, T., & Sargent, T. J. (2002). Evolving post-World War II U.S. inflation dynamics. In M. Getler, & K. Rogoff (Eds.), *NBER Macroeconomics Annual* (pp. 331–373). MIT Press.
Conrad, C., & Karanasos, M. (2010). Negative volatility spillovers in the unrestricted ECCC-GARCH model. *Econometric Theory*, 26, 838–862.
Cramér, H. (1961). On some classes of nonstationary stochastic processes. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 2, 57–78.
Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics*, 25, 1–37.
Devpura, N., Sharma, S. S., Harischandra, P. K. G., & Pathberiya, L. R. C. (2021). Is inflation persistent? Evidence from a time-varying unit root model. *Pacific-Basin Finance Journal*, 68(C), Article 101577.
Diebold, F. X., & Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics*, 105, 131–159.
Elaydi, S. (2005). *An introduction to difference equations*. Springer.
Erhardsson, T. (2014). Conditions for convergence of random coefficient AR(1) processes and perpetuities in higher dimensions. *Bernoulli*, 20, 990–1005.
Fuhrer, J., & Moore, G. (1995). Inflation persistence. *The Quarterly Journal of Economics*, 110, 127–159.
Fuller, W. A. (1996). *Introduction to statistical time series*. John Wiley & Sons.
Giraitis, L., Kapetanios, G., & Yates, T. (2014). Inference on stochastic time-varying coefficient models. *Journal of Econometrics*, 179, 46–65.
Glasserman, P., & Yao, D. D. (1995). Stochastic vector difference equations with stationary coefficients. *Journal of Applied Probability*, 32, 851–866.
Goerg, G. M. (2010). *Time series analysis of long memory versus structural breaks: A time-varying memory approach*. VDM Verlag Dr. Müller.
Granger, C. W. J., & Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance*, 11, 399–421.
Gray, R. M. (2006). Toeplitz and circulant matrices: A review. *Foundation and Trends in Communications and Information Theory*, 2, 155–239.
Grillenzoni, C. (1990). Modelling time-varying dynamical systems. *Journal of the American Statistical Association*, 85, 499–507.
Grillenzoni, C. (1993). ARIMA processes with ARIMA parameters. *Journal of Business and Economic Statistics*, 11, 235–250.
Grillenzoni, C. (2000). Time-varying parameters prediction. *Annals of the Institute of Statistical Mathematics*, 52, 108–122.
Hallin, M. (1979). Mixed autoregressive-moving average multivariate processes with time-dependent coefficients. *Journal of Multivariate Analysis*, 8, 567–572.
Hallin, M. (1984). Spectral factorization of nonstationary moving average processes. *Annals of Statistics*, 12, 172–192.
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.

Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press.
Han, C. (2007). Determinants of covariance matrices of differenced AR (1) processes. *Econometric Theory*, 23, 1248–1253.
Harvey, D., Leybourne, S., Georgiev, I., & Taylor, A. M. R. (2018). Testing for parameter instability in predictive regression models. *Journal of Econometrics*, 204, 101–118.
He, C., & Teräsvirta, T. (2004). An extended constant conditional correlation GARCH model and its fourth-moment structure. *Econometric Theory*, 20, 904–926.
Hsu, C. C. (2005). Long memory or structural changes: An empirical examination on inflation rates. *Economics Letters*, 88, 289–294.
Hwang, S. Y., & Basawa, I. V. (1998). Parameter estimation for generalized random coefficient autoregressive processes. *Journal of Statistical Planning and Inference*, 68, 323–337.
Iacone, F., Leybourne, S. J., & Taylor, A. M. R. (2019). Testing the order of fractional integration of a time series in the possible presence of a trend break at an unknown point. *Econometric Theory*, 35, 1201–1233.
Jeerawat, U., & Daowsud, K. (2022). Determinants of some Hessenberg matrices with generating functions. *Special Matrices*, 11, 1–8.
Karanasos, M. (1998). A new method for obtaining the autocovariance of an ARMA model: An exact form solution. *Econometric Theory*, 14, 622–640.
Karanasos, M. (1999). The second moment and the autocovariance function of the squared errors of the GARCH model. *Journal of Econometrics*, 90, 63–76.
Karanasos, M. (2001). Prediction in ARMA models with GARCH in mean effects. *Journal of Time Series Analysis*, 22, 555–576.
Karanasos, M., Hu, Y., & Yfanti, S. (2022a). *Constrained QML for vector MEM: The practicality of matrix inequality constraints*. Unpublished paper.
Karanasos, M., Koutroumpis, P., Karavias, Y., Kartsaklas, A., & Arakelian, V. (2016). Inflation convergence in the EMU. *Journal of Empirical Finance*, 39, 241–253.
Karanasos, M., Paraskevopoulos, A. G., & Canepa, A. (2020). *A unified theory for the large family of time-varying models with ARMA representations: One solution fits all*. Cognetti de Martiis Department of Economics and Statistics.
Karanasos, M., Paraskevopoulos, A. G., & Dafnos, S. (2014a). *A univariate time-varying analysis of periodic ARMA processes*. Preprint, arXiv:1403.4803.
Karanasos, M., Paraskevopoulos, A. G., & Dafnos, S. (2022). The fundamental properties of AR models with deterministically varying coefficients. In M. Pempetzoglou, & D. Hristou-Varsakelis (Eds.), *Essays in economic theory and policy in honor of Professor Stella Karagianni*. Springer.
Karanasos, M., Paraskevopoulos, A. G, Karoglou, M., Menla Ali, F., & Yfanti, S. (2014b). Modelling returns and volatilities during financial crises: A time-varying coefficient approach. *Journal of Empirical Finance*, 29, 113–128.
Kittappa, R. K. (1993). A representation of the solution of the *n*th order linear difference equation with variable coefficients. *Linear Algebra and Its Applications*, 193, 211–222.
Kowalski, A., & Szynal, D. (1991). On a characterization of optimal predictors for nonstationary ARMA processes. *Stochastic Processes and Their Applications*, 37, 71–80.
Kozicki, S., & Tinsley, P. (2002). *Alternative sources of the lag dynamics of inflation* (Publication No. 02-12) [Working paper]. FRB Kansas City Research.
Levin, A., & Piger, J. (2004) *Is inflation persistence intrinsic in industrial economics?* (Publication No. 202-023A) [Working paper]. Federal Reserve Bank of St. Louis.
Lim, A., & Dai, J. (2011). On product of companion matrices. *Linear Algebra and Its Applications*, 435, 2921–2935.
Mallik, R. K. (1998). Solutions of linear difference equations with variable coefficients. *Journal of Mathematical Analysis and Applications*, 222, 79–91.
Marrero, A. J., & Tomeo, V. (2012). On the closed representation for the inverse of Hessenberg matrices. *Journal of Computational and Applied Mathematics*, 236, 2962–2970.

Marrero, A. J., & Tomeo, V. (2017). On compact representations for the solutions of linear difference equations with variable coefficients. *Journal of Computational and Applied Mathematics*, 318, 411–421.

Massacci, D., & Trapani, L. (2022). High dimensional threshold regression with common stochastic trends. *SSRN*. https://ssrn.com/abstract=4133488; https://doi.org/10.2139/ssrn.4133488

McConnell, M., & Perez-Quiros, G. (2000). Output fluctuations in the United States: What has changed since the early 1980's? *American Economic Review*, 90, 1464–1476.

Mélard, G. (2024) Time-dependent time series models: Comments on Marc Hallin's early contributions and a pragmatic view on estimation. In: M. Barigozzi, S. Hörmann and D. Paindaveine (Eds.), *Recent advances in econometrics and statistics: Festschrift in honour of Marc Hallin*. Springer.

Miller, K. S. (1968). *Linear difference equations*. Benjamin.

Nelson, D. (2008). *The Penguin dictionary of mathematics* (4th ed.). Penguin Books Limited.

Paraskevopoulos, A. G. (2012). *The infinite Gauss–Jordan elimination on row-finite $\omega \times \omega$ matrices*. Preprint, arXiv:1201.2950.

Paraskevopoulos, A. G. (2014). The solution of row-finite linear systems with the infinite Gauss–Jordan elimination. Preprint, arXiv:1403.2624.

Paraskevopoulos, A. G., & Karanasos, M. (2021). Explicit and compact representations for the Green function and the solution of linear difference equations with variable coefficients. Preprint, arXiv:1906.07130.

Peiris, S. (1986). On prediction with time dependent ARMA models. *Communications in Statistics – Theory and Methods*, 15, 3659–3668.

Perron, P., & Qu, Z. (2006). Estimating restricted structural change models. *Journal of Econometrics*, 134, 373–399.

Petrova, K. (2019). A quasi-Bayesian local likelihood approach to time-varying parameter VAR models. *Journal of Econometrics*, 212, 286–306.

Phillips, P. C. B. (2005). Econometric analysis of Fisher's equation. *The American Journal of Economics and Sociology*, 64, 125–168.

Phillips, P. C. B. (2007). Unit root log periodogram regression. *Journal of Econometrics*, 138, 104–124.

Pivetta, F., & Reis, R. (2007). The persistence of inflation in the United States. *Journal of Economic Dynamics and Control*, 31, 1326–1358.

Rao, S. T. (1970). The fitting of non-stationary time-series models with time-dependent parameters. *Journal of the Royal Statistical Society*, 32, 312–322.

Sibbertsen, P. (2004). Long memory versus structural breaks: An overview. *Statistical Papers*, 45, 465–515.

Singh, N., & Peiris, M. S. (1987). A note on the properties of some nonstationary ARMA processes. *Stochastic Processes and Their Applications*, 24, 151–155.

Stock, J. H., & Watson, M. W. (2002). Has the business cycle changed and why? In M. Gertler, & K. Rogoff (Eds.), *NBER Macroeconomics Annual* (pp. 159–218). MIT Press.

Stock, J. H., & Watson, M. W. (2009). Phillips curve inflation forecasts. In J. C. Fuhrer, Y. Kodrzycki, J. Little, & G. Olivei (Eds.), *Understanding inflation and the implications for monetary policy* (pp. 99–202). MIT Press.

Sun, Y., & Phillips, P. C. B. (2004). Understanding the Fisher equation. *Journal of Applied Econometrics*, 19, 869–886.

Taylor, J. (2000). Low inflation, pass through, and the pricing power of firms. *European Economic Review*, 44, 1389–1408.

Vein, R., & Dale, P. (1999). *Determinants and their application in mathematical physics*. Springer.

Whittle, P. (1965). Recursive relations for predictors of non-stationary processes. *Journal of the Royal Statistical Society, Series B*, 27, 523–532.

Widom, H. (1958). On the eigenvalues of certain Hermitian operators. *Transactions of the American Mathematical Society*, 88, 491–522.

Wright, M. (2006). Green function or Green's function? *Nature Physics*, 2, Article 646.