ONLINE APPENDIX:

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

M. Karanasos, A. G. Paraskevopoulos, T. Magdalinos, A. Canepa

July 21, 2024

This document is an online Appendix to "A unified theory for ARMA models with varying coefficients: One solution fits all". It is a continuation of the Appendices, included in the main body of the paper. We shall here follow the notation as well as the numbering of equations, Sections, Theorems, etc., employed therein. For example eq. (1) (or eq. (B.1)) and Theorem 1 (or Proposition B1) are referred to corresponding equations and statements in the main body of the paper. Some supportive supplements as well as the proofs of statements already mentioned in the main body of the paper are presented here.

D Examples

In this Section we apply our results to three specific models, included in Subsection D1 and to the ARMA(p, q) model with constant parameters in Subsection D2.

D1 Specific Models

The following models are built upon the TV-AR(1) process defined by

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

where $\phi(t)$ is identified with $\phi_1(t)$ of eq. (1), $\{\varepsilon_t\}$ is an orthogonal sequence (uncorrelated) defined on L_2 with constant variance σ^2. The principal matrix $\mathbf{\Phi}_{t,r}$ consists solely of two nonzero diagonals: the superdiagonal whose elements are (-1)s and the main diagonal whose elements are $\phi(r+1), \phi(r+2), ..., \phi(t)$. Therefore the principal determinant is the product of the elements of the main diagonal, that is:

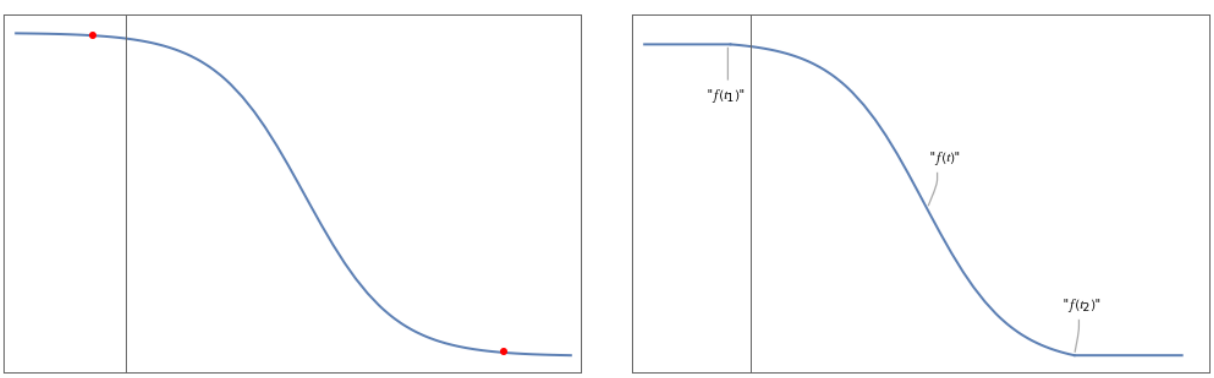
$$\xi(t,r) = \prod_{m=r+1}^{t} \phi(m). \tag{D.1}$$

The Wold-Crámer representation formula in eq. (18a) must be applied with $\xi_q(t,r) = \xi(t,r)$ for all r (since $\varphi(r) = 0$, $u_r = \varepsilon_r$ and $\theta_l(r) = 0$ for any $l \in [1..q]$).

Example D.1 is concerned with the Logistic Smooth Transition AR (LSTAR) model (see, for example, Teräsvirta, 1994). The other two are taken from Azrak and Mélard (2006), dealing with coefficients of periodic and exponential functions, respectively. In all these specifications we explicitly formulate the principal determinant function $\xi(t,r)$ and we show that the conditions of Theorems 2 and 3 are fulfilled. These results enable us to derive the Wold-Crámer representations explicitly. Besides, in view of Proposition 2, the unconditional variance $\mathbb{V}ar(y_t)$ yields a closed form representation.

Example D.1 (Logistic Smooth Transition Model) In the LSTAR(1) model, the AR coefficient $\phi(t)$ is build up as follows. First, define the logistic function: $F(t;\gamma,\tau) = [1 + e^{\gamma(t-\tau)}]^{-1}$, where $\gamma \in \mathbb{R}_{\geq 0}$ is the parameter of the logistic growth rate and $\tau \in \mathbb{Z}$ is the t value of the function's midpoint parameter. Second, define the function: $f(t) = \phi_1 F(t;\gamma,\tau) + \phi_2[1 - F(t;\gamma,\tau)]$, where $|\phi_1| < 1$, $\phi_2 < \phi_1$. It is decomposed into two regimes: $\phi_1 F(t;\gamma,\tau)$ (regime 1) and $\phi_2[1 - F(t;\gamma,\tau)]$ (regime 2). $F(t;\gamma,\tau)$ is a decreasing sigmoid function of time t with $F(\tau;\gamma,\tau) = 0.5$. Moreover, $F(t;\gamma,\tau) \to 1$, as $t \to -\infty$, $F(t;\gamma,\tau) \to 0$, as $t \to \infty$, and $0 < F(t;\gamma,\tau) < 1$. Accordingly, $f(t) \to \phi_1$, as $t \to -\infty$ and $f(t) \to \phi_2$, as $t \to \infty$ and $\phi_2 < \phi(t) < \phi_1$ for all $t \in \mathbb{Z}$. The latter implies that $f(t)$ is a bounded function of time with horizontal asymptotes at $y = \phi_1$ and $y = \phi_2$. Choosing t_1 small enough ($t_1 \ll \tau$), then $F(t_1;\gamma,\tau) \approx 1$ and regime 1 prevails, that is $f(t) \approx \phi_1$ for $t \leq t_1$, whereas choosing $t_2 \gg \tau$, then $F(t_2;\gamma,\tau) \approx 0$ and regime 2 prevails, that is $f(t) \approx \phi_2$ for $t \geq t_2$. Clearly $t_2 > t_1$, since $F(t;\gamma,\tau)$ is a decreasing function in t. Finally, the coefficient $\phi(t)$ of the LSTAR model, followed by relevant graphs, is:

$$\phi(t) = \begin{cases} f(t_1) & \text{if} \quad t \leq t_1 \quad \text{(constant function)}, \\ f(t) & \text{if} \quad t_1 \leq t \leq t_2, \\ f(t_2) & \text{if} \quad t_2 \leq t \quad \text{(constant function)}, \end{cases}$$



(a) The graph of $f(t)$ along with $(t_i, f(t_i)), i = 1, 2$.

(b) The graph of the coefficient $\phi(t)$.

LSTAR(1) Model

In all that follows we assume that $r \leq t_1 < t_2 \leq t$. The elements in the main diagonal of $\mathbf{\Phi}_{t,r}$ are:

$$\begin{array}{lccccccccc} & \phi(r+1), & ..., & \phi(t_1), & \phi(t_1+1), & ..., & \phi(t_2-1), & \phi(t_2), & ..., & \phi(t), \\ \text{or} & \phi_1, & ..., & \phi_1, & \phi(t_1+1), & ..., & \phi(t_2-1), & \phi_2, & ..., & \phi_2. \end{array}$$

The principal determinant in eq. (D.1) takes the form:

$$\xi(t,r) = \phi_1^{t_1-r+1}\Big(\prod_{j=t_1+1}^{t_2-1}\phi(j)\Big)\phi_2^{t-t_2+1}.$$

On account of $\sum_{r=-\infty}^{t_1}|\phi_1|^{t_1-r+1} < \infty$, since $|\phi_1| < 1$, the absolute summability condition in (17) holds, that is for any fixed t:

$$\sum_{r=-\infty}^{t}|\xi(t,r)| = |\phi_2|^{t-t_2+1}\prod_{j=t_1+1}^{t_2-1}|\phi(j)|\sum_{r=-\infty}^{t_1}|\phi_1|^{t_1-r+1} < \infty.$$

As the conditions of Theorem 2 are fulfilled, the Wold-Crámer representation formula in eq. (18a), applied with $\varphi_t = 0$ (zero drift) and $\theta_l(t) = 0$ (zero moving average coefficients), yields an asymptotically stable MA representation, which is of the form:

$$y_t = \sum_{r=-\infty}^{t}\xi(t,r)\varepsilon_r = \phi_2^{t-t_2+1}\prod_{j=t_1+1}^{t_2-1}\phi(j)\sum_{r=-\infty}^{t_1}\phi_1^{t_1-r+1}\varepsilon_r.$$

By virtue of Proposition 2, the unconditional variance $\mathbb{V}ar(y_t)$, can be derived from eq. (21b) using σ in place of $\sigma(r)$, that is

$$\mathbb{V}ar(y_t) = \sum_{r=-\infty}^{t}\xi^2(t,r)\sigma^2 = \sigma^2\phi_2^{2(t-t_2+1)}\prod_{j=t_1+1}^{t_2-1}\phi^2(j)\sum_{r=-\infty}^{t_1}\phi_1^{2(t_1-r+1)}.$$

A closed form representation of $\mathbb{V}ar(y_t)$ is obtained in what follows. Taking into account that $|\phi_1| < 1$, it follows that

$$\begin{aligned}\sum_{r=-\infty}^{1}\phi_1^{2(t_1-r+1)} &= \phi_1^{2t_1} + \phi_1^{2t_1+2} + \phi_1^{2t_1+4} + ... = \phi_1^{2t_1} + \phi_1^2\phi_1^{2t_1} + \phi_1^4\phi_1^{2t_1} + ... \\ &= \phi_1^{2t_1}(1 + \phi_1^2 + \phi_1^4 + ...) = \frac{\phi_1^{2t_1}}{1-\phi_1^2}.\end{aligned}$$

Without loss of generality, we assume that $t_1 \geq 1$. It follows that:

$$\sum_{r=-\infty}^{t_1}\phi_1^{2(t_1-r+1)} = \sum_{r=-\infty}^{1}\phi_1^{2(t_1-r+1)} + \sum_{r=2}^{t_1}\phi_1^{2(t_1-r+1)} = \sum_{r=2}^{t_1}\phi_1^{2(t_1-r+1)} + \frac{\phi_1^{2t_1}}{1-\phi_1^2}.$$

Substituting the above result in the formula of $\mathbb{V}ar(y_t)$, the closed form of the latter is:

$$\mathbb{V}ar(y_t) = \sigma^2\phi_2^{2(t-t_2+1)}\prod_{j=t_1+1}^{t_2-1}\phi^2(j)\Big(\sum_{r=2}^{t_1}\phi_1^{2(t_1-r+1)} + \frac{\phi_1^{2t_1}}{1-\phi_1^2}\Big).$$

Example D.2 (Periodic) Consider the AR(1) model with time-varying coefficient $\phi(t) = \beta_{t-n\lfloor t/n \rfloor}$ (see Example 1 in Azrak and Mélard, 2006), where $t \in \mathbb{Z}$, $\lfloor t/n \rfloor$ is the integer part of the division t/n, $n \in \mathbb{Z}_2$ and $\beta_j \in \mathbb{R}$ for $0 \leq j \leq n-1$; both n, β_j remain fixed throughout the example. It is also assumed that $|\prod_{j=0}^{n-1} \beta_j| < 1$. We first show that the above specification is an equivalent version to the traditional periodic (PAR(1)) one which is grounded on the Euclidean division identity. In particular, applying Euclid's division Lemma for $n \in \mathbb{Z}_2$, it follows that for any $t \in \mathbb{Z}$ there exists a unique pair of integers (T, R), called the quotient and the remainder of the division, respectively, such that:

$$\left.\begin{aligned} &t = nT + R \\ &0 \leq R \leq n-1. \end{aligned}\right\} \tag{D.2}$$

The remainder (inequality) condition in eq. (D.2) is essential, as it ensures the uniqueness of the pair (T, R). In particular, eq. (D.2) meets the traditional interpretation of a PAR(1) model in which the time t is expressed in terms of seasons (e.g., quarters), $n \in \mathbb{Z}_2$ is the number of seasons in a whole period (e.g., $n = 4$ is the number of quarters per year), $T \in \mathbb{Z}_0$ stands for the number of whole periods (e.g., years), $R = t - nT$ is the number of remaining seasons (e.g., quarters), after the elimination of all whole periods included in t (nT) and $k = t - r > 0$ is the total number of seasons (or periods) from season $r + 1$ up to and including season t (e.g., quarters).

Since n is positive and $\lfloor t/n \rfloor$ is the greatest integer less than or equal to t/n, the Euclidean division identity in eq. (D.2) can be equivalently expressed by a single identity:

$$t = n\lfloor t/n \rfloor + R. \tag{D.3}$$

This is due to the fact that the identity in eq. (D.3) necessarily implies both: $0 \leq R \leq n-1$ and $T = \lfloor t/n \rfloor$. As t/n is a rational number (division of integers), writing $t - n\lfloor t/n \rfloor = R$, it follows from eq. (D.3) that $\lfloor t/n \rfloor$ and R are respectively the quotient and the remainder of the Euclidean division of t by n. This establishes the equivalence of these models. For example if $n = 4$ and $t = 7$ (quarters), then $T = 1$ (year) and $R = 3$ (quarters). If $n = 4$ and $t = 8$, then $T = 2$ and $R = 0$. These are in line with $\phi(7) = \beta_{7-4\lfloor 7/4 \rfloor} = \beta_{7-4\cdot 1} = \beta_3$ and $\phi(8) = \beta_{8-4\lfloor 8/4 \rfloor} = \beta_{8-4\cdot 2} = \beta_0$.

Let $\beta = \prod_{j=0}^{n-1} \beta_j$. We shall refer to β as a whole period product. Trivially $\prod_{j=i}^{n+i-1} \beta_j$ for any $i \in \mathbb{Z}_0$, is also a whole period product, since $\prod_{j=i}^{n+i-1} \beta_j = \beta$. The elements in the main diagonal of $\mathbf{\Phi}_{t,r}$ are

$$\beta_{r+1-n\lfloor (r+1)/n \rfloor}, \beta_{r+2-n\lfloor (r+2)/n \rfloor} \cdots, \beta_{t-1-n\lfloor (t-1)/n \rfloor}, \beta_{t-n\lfloor t/n \rfloor}$$

or in reverse order $\{\beta_{t-j-n\lfloor \frac{t-j}{n} \rfloor}, j = 0, ..., t-r-1\}$ and the number of these elements is $(t - r)$. In view of eq. (D.1), the principal determinant $\xi(t, r)$ (the product of the main diagonal elements) comprises $\lfloor \frac{t-r}{n} \rfloor$ whole period products and the number of their elements is $n\lfloor \frac{t-r}{n} \rfloor$.

Consequently, we have:

$$\xi(t,r)=\left(\prod_{j=0}^{t-r-1}\beta_{t-j-n\lfloor\frac{t-j}{n}\rfloor}\right)=\left(\prod_{j=0}^{t-r-1-n\lfloor\frac{t-r}{n}\rfloor}\beta_{t-j-n\lfloor\frac{t-j}{n}\rfloor}\right)\left(\prod_{j=t-r-n\lfloor\frac{t-r}{n}\rfloor}^{t-r-1}\beta_{t-j-n\lfloor\frac{t-j}{n}\rfloor}\right).$$

As $t-r-1-(t-r-n\lfloor\frac{t-r}{n}\rfloor)+1=n\lfloor\frac{t-r}{n}\rfloor$, it follows that the number of terms in the second of the above products coincide with the number of elements of whole period products contained in the main diagonal of $\mathbf{\Phi}_{t,r}$, thereby:

$$\prod_{j=t-r-n\lfloor\frac{t-r}{n}\rfloor}^{t-r-1}\beta_{t-j-n\lfloor\frac{t-j}{n}\rfloor}=\beta^{\lfloor\frac{t-r}{n}\rfloor}.$$

We thus conclude that:

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

For each fixed t, it follows from the assumption $|\beta|<1$ that $\beta^{\lfloor\frac{t-r}{n}\rfloor}\to 0$, as $r\to-\infty$. Additionally

$$\prod_{j=0}^{t-r-1-n\lfloor\frac{t-r}{n}\rfloor}\beta_{t-j-n\lfloor\frac{t-j}{n}\rfloor}=\begin{cases}1 & \text{if}\quad r=t,t-n,...,t-2n,...\\ \beta_{t-n\lfloor\frac{t}{n}\rfloor} & \text{if}\quad r=t-1,t-n-1,...,t-2n-1,...\\ \vdots & \vdots\quad\vdots\\ \prod_{j=0}^{n-2}\beta_{t-j-n\lfloor\frac{t-j}{n}\rfloor} & \text{if}\quad r=t-n+1,t-2n+1,...,t-3n+1,...\end{cases},$$

(where we use: $\prod_{j=0}^{-1}(\cdot)=1$) that is, there are only $(n-1)$ possible outcomes of $\prod_{j=0}^{t-r-1-n\lfloor\frac{t-r}{n}\rfloor}\beta_{t-j-n\lfloor\frac{t-j}{n}\rfloor}$ for all $r\leq t$. Since $|\beta|<1$, the absolute summability condition in (17) is fulfilled, as shown in what follows:

$$\begin{aligned}\sum_{r=-\infty}^{t}|\xi(t,r)|=&\sum_{r=-\infty}^{t}|\beta|^{\lfloor\frac{t-r}{n}\rfloor}\prod_{j=0}^{t-r-1-n\lfloor\frac{t-r}{n}\rfloor}|\beta_{t-j-n\lfloor\frac{t-j}{n}\rfloor}|\\
&=|\beta|^0\cdot 1+|\beta|^0\cdot|\beta_{t-n\lfloor\frac{t}{n}\rfloor}|+...+|\beta|^0\cdot\prod_{j=0}^{n-2}|\beta_{t-j-n\lfloor\frac{t-j}{n}\rfloor}|\\
&+|\beta|\cdot 1+|\beta|\cdot|\beta_{t-n\lfloor\frac{t}{n}\rfloor}|+...+|\beta|\prod_{j=0}^{n-2}|\beta_{t-j-n\lfloor\frac{t-j}{n}\rfloor}|\\
&+|\beta|^2\cdot 1+|\beta|^2\cdot|\beta_{t-n\lfloor\frac{t}{n}\rfloor}|+...+|\beta|^2\prod_{j=0}^{n-2}|\beta_{t-j-n\lfloor\frac{t-j}{n}\rfloor}|+...\\
&=|\beta|^0\left(1+|\beta_{t-n\lfloor\frac{t}{n}\rfloor}|+...+\prod_{j=0}^{n-2}|\beta_{t-j-n\lfloor\frac{t-j}{n}\rfloor}|\right)\\
&+|\beta|\left(1+|\beta_{t-n\lfloor\frac{t}{n}\rfloor}|+...+\prod_{j=0}^{n-2}|\beta_{t-j-n\lfloor\frac{t-j}{n}\rfloor}|\right)\\
&+|\beta|^2\left(1+|\beta_{t-n\lfloor\frac{t}{n}\rfloor}|+...+\prod_{j=0}^{n-2}|\beta_{t-j-n\lfloor\frac{t-j}{n}\rfloor}|\right)+...\\
&=(1+|\beta|+|\beta|^2+...)\left(1+|\beta_{t-n\lfloor\frac{t}{n}\rfloor}|+...+\prod_{j=0}^{n-2}|\beta_{t-j-n\lfloor\frac{t-j}{n}\rfloor}|\right)\\
&=\frac{1}{1-|\beta|}\sum_{r=t-n+1}^{t}\prod_{j=0}^{t-r-1-n\lfloor\frac{t-r}{n}\rfloor}|\beta_{t-j-n\lfloor\frac{t-j}{n}\rfloor}|<\infty.\end{aligned}$$

As the conditions of Theorem 2 are fulfilled, the Wold-Crámer representation formula in eq. (18a) for a PAR(1) model, applied with $\varphi_t = 0$ and $\theta_l(t) = 0$, yields an asymptotically stable MA representation, which is of the form:

$$y_t = \sum_{r=-\infty}^{t} \xi(t,r)\varepsilon_r = \sum_{r=-\infty}^{t} \beta^{\lfloor \frac{t-r}{n} \rfloor} \left(\prod_{j=0}^{t-r-1-n\lfloor \frac{t-r}{n} \rfloor} \beta_{t-j-n\lfloor \frac{t-j}{n} \rfloor} \right) \varepsilon_r.$$

Proposition 2 entails that the unconditional variance is given by:

$$\mathbb{V}ar(y_t) = \sigma^2 \sum_{r=-\infty}^{t} \xi^2(t,r) = \sigma^2 \sum_{r=-\infty}^{t} \beta^{2\lfloor \frac{t-r}{n} \rfloor} \left(\prod_{j=0}^{t-r-1-n\lfloor \frac{t-r}{n} \rfloor} \beta_{t-j-n\lfloor \frac{t-j}{n} \rfloor} \right)^2.$$

Replacing $|\xi(t,r)|$ with $\xi^2(t,r)$ in the proof of the above formula of $\sum_{r=-\infty}^{t} |\xi(t,r)|$, we can similarly show that:

$$\sum_{r=-\infty}^{t} \xi^2(t,r) = \frac{1}{1-\beta^2} \sum_{r=t-n+1}^{t} \prod_{j=0}^{t-r-1-n\lfloor \frac{t-r}{n} \rfloor} \beta^2_{t-j-n\lfloor \frac{t-j}{n} \rfloor}.$$

Therefore a closed form representation of the unconditional variance is given by:

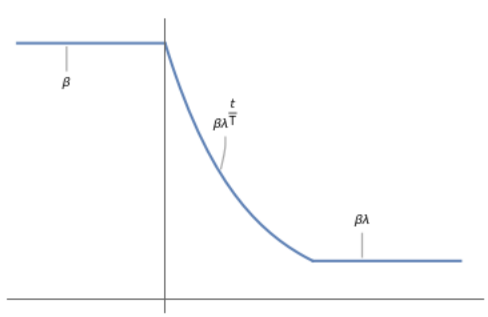
$$\mathbb{V}ar(y_t) = \frac{\sigma^2}{1-\beta^2} \sum_{r=t-n+1}^{t} \left(\prod_{j=0}^{t-r-1-n\lfloor \frac{t-r}{n} \rfloor} \beta_{t-j-n\lfloor \frac{t-j}{n} \rfloor} \right)^2.$$

This result coincides with eq. (4.2) in Azrak and Mélard, 2006 (under the aforementioned assignment).

Example D.3 (Exponential) In this example (see Example 2 in Azrak and Mélard, 2006), the coefficient of an AR(1) model is a decreasing exponential function of time in the interval $[0, T]$ and constant elsewhere

$$\phi(t) = \begin{cases} \beta & \text{if} \quad t \le 0, \\ \beta\lambda^{t/T} & \text{if} \quad 0 \le t \le T, \\ \beta\lambda & \text{if} \quad T \le t, \end{cases}$$

where $T \in \mathbb{Z}_1$ is the sample size, $0 < \beta < 1$ and the true value of λ ranges over $(0, 1)$ (as pictured below).



Exponential model

Clearly $\beta\lambda < \beta$ and so $\beta\lambda \leq \phi(t) \leq \beta$ for all t, that is $\phi(t)$, $t \in \mathbb{Z}$ is a bounded function.
If $t \geq T$ and $r \leq 0$, then the elements of the main diagonal of $\mathbf{\Phi}_{t,r}$ are

$$\begin{array}{cccc} & \phi(r+1), ..., \phi(0), & \phi(1), ..., \phi(T), & \phi(T+1), ..., \phi(t), \\ \text{or} & \underbrace{\beta, \;\; ..., \quad\; \beta,}_{-r} & \underbrace{\beta\lambda^{\frac{1}{T}}, \; ..., \; \beta\lambda,}_{T} & \underbrace{\beta\lambda, \;\; ..., \quad\; \beta\lambda,}_{t-T} \end{array}$$

the product of which yields the principal determinant $\xi(t,r)$, given by (D.1). More specifically, we consider the following cases:
If $T \leq r \leq t$, then: $\xi(t,r) = \phi(r+1)\ldots\phi(t) = (\beta\lambda)^{t-r}$.
If $0 \leq r \leq T-1 < T \leq t$, then $1 \leq r+1 \leq T \leq t$ and:

$$\begin{aligned} \xi(t,r) =& \big(\phi(r+1)\ldots\phi(T)\big)\big(\phi(T+1)\ldots\phi(t)\big) = \Bigg(\prod_{i=r+1}^{T} \beta\lambda^{\frac{i}{T}}\Bigg)(\beta\lambda)^{t-T} = \beta^{T-r}\lambda^{\frac{(r+1)+\ldots+T}{T}}\beta^{t-T}\lambda^{t-T} \\ =& \beta^{t-r}\lambda^{t-T}\lambda^{\frac{(T-r)(T+r+1)}{2T}} = \beta^{t}\beta^{-r}\lambda^{t}\lambda^{\frac{T^2+T-r^2-r}{2T}-T} = (\beta\lambda)^{t}\beta^{-r}\lambda^{(-\frac{2T^2}{2T}+\frac{T^2+T-r^2-r}{2T})} \\ =& (\beta\lambda)^{t}\beta^{-r}\lambda^{\frac{-T^2+T-r^2-r}{2T}} = (\beta\lambda)^{t}\beta^{-r}\lambda^{-\frac{T(T-1)+r(r+1)}{2T}} = (\beta\lambda)^{t}\beta^{-r}\lambda^{-(\frac{T-1}{2}+\frac{r(r+1)}{2T})}. \end{aligned}$$

If $-\infty < r \leq 0 < 1 \leq T-1 < T \leq t$, since $\xi(t,0) = \Big(\prod_{i=1}^{T} \beta\lambda^{\frac{i}{T}}\Big)(\beta\lambda)^{t-T}$, it follows that

$$\xi(t,r) = \beta^{-r}\Bigg(\prod_{i=1}^{T} \beta\lambda^{\frac{i}{T}}\Bigg)(\beta\lambda)^{t-T} = \beta^{-r}\xi(t,0).$$

The above results are summarized as follows:

$$\xi(t,r) = \begin{cases} (\beta\lambda)^{t-r} & \text{if} \quad T \leq r \leq t, \\ (\beta\lambda)^{t}\beta^{-r}\lambda^{-(\frac{T-1}{2}+\frac{r(r+1)}{2T})} & \text{if} \quad 0 \leq r \leq T, \\ \beta^{-r}\xi(t,0) & \text{if} \quad r \leq 0. \end{cases}$$

On account of $0 < \beta < 1$, the absolute summability condition in (17) is fulfilled by:

$$\begin{aligned} \sum_{r=-\infty}^{t} \xi(t,r) =& \sum_{r=T}^{t} \xi(t,r) + \sum_{r=1}^{T-1} \xi(t,r) + \sum_{r=-\infty}^{0} \xi(t,r) \\ =& \sum_{r=0}^{t-T}(\beta\lambda)^{r} + \sum_{r=1}^{T-1}(\beta\lambda)^{t}\beta^{-r}\lambda^{-(\frac{T-1}{2}+\frac{r(r+1)}{2T})} + \xi(t,0)(1+\beta+\beta^2+...) \\ =& \frac{1-(\beta\lambda)^{t-T+1}}{1-\beta\lambda} + (\beta\lambda)^{t}\sum_{r=1}^{T-1}\beta^{-r}\lambda^{-(\frac{T-1}{2}+\frac{r(r+1)}{2T})} + \frac{\xi(t,0)}{1-\beta} < \infty. \end{aligned}$$

As the conditions of Theorem 2 are fulfilled, the Wold-Crámer representation formula in eq. (18a) for an exponential AR(1) model, applied with $\varphi_t = 0$ and $\theta_l(t) = 0$, yields an asymptotically stable MA representation, which is of the form:

$$\sum_{r=-\infty}^{t} \xi(t,r)\varepsilon_r = \sum_{r=0}^{t-T}(\beta\lambda)^r\varepsilon_r + (\beta\lambda)^t \sum_{r=1}^{T-1} \beta^{-r}\lambda^{-\left(\frac{T+1}{2}+\frac{r(r+1)}{2T}\right)}\varepsilon_r + \sum_{r=-\infty}^{0} \beta^{-r}\xi(t,0)\varepsilon_r.$$

Moreover, the unconditional variance yields a closed form representation:

$$\begin{aligned}\mathbb{V}ar(y_t) = \sum_{r=-\infty}^{t} \xi^2(t,r)\sigma^2 &= \Bigg(\sum_{r=T}^{t}\xi^2(t,r) + \sum_{r=1}^{T-1}\xi^2(t,r) + \sum_{r=-\infty}^{0}\xi^2(t,r)\Bigg)\sigma^2 \\ &= \Bigg(\frac{1-(\beta\lambda)^{2(t-T+1)}}{1-(\beta\lambda)^2} + (\beta\lambda)^{2t}\sum_{r=1}^{T-1}\beta^{-2r}\lambda^{-(T+1+\frac{r(r+1)}{T})} + \frac{\xi^2(t,0)}{1-\beta^2}\Bigg)\sigma^2.\end{aligned}$$

As pointed out by Azrak and Mélard in the above cited reference the use of variable coefficients, which depend on the length of the series, is compatible with the approach of Dahlhaus (1996).

D2 ARMA with Constant Parameters

In this Subsection, we deal with ARMA(p,q) models with constant parameters:

$$y_t = \varphi + \sum_{m=1}^{p}\phi_m y_{t-m} + \varepsilon_t + \sum_{l=1}^{q}\theta_l\varepsilon_{t-l} \text{ for all } t \in \mathbb{Z}. \tag{D.4}$$

In this case, the principal determinant is a determinant of a banded Toeplitz-Hessenberg matrix. We have shown that the optimal linear predictor formula in eq. (23a) is of the form:

$$\hat{\mathbb{E}}(y_t \,|\, \mathcal{K}_s) = \sum_{m=1}^{p}\xi^{(m)}(t-s)y_{s+1-m} + \varphi\sum_{r=s+1}^{t}\xi(t-r) + \sum_{r=s+1-q}^{s}\xi_{s,q}(t-r)\varepsilon_r \tag{D.5}$$

(see the analysis in Proposition 4). In what follows we show that eq. (D.5) coincides with the formula established by Karanasos (2001), in his Theorem 1, eq. (2.7), that is

$$\hat{\mathbb{E}}(y_t \,|\, \mathcal{K}_s) = \sum_{m=1}^{p}\sum_{l=1}^{p}\zeta_{lk}\gamma_{lm}y_{s+1-m} + \varphi\left(\frac{1}{\Phi(1)} - \sum_{m=1}^{p}\frac{\zeta_{mk}}{(1-\lambda_m)}\right) + \sum_{r=s+1-q}^{s}\sum_{m=1}^{p}\sum_{l=s+1-r}^{q}\zeta_{m0}\lambda_m^{t-r-l}\theta_l\varepsilon_r \tag{D.6}$$

where $\{\lambda_1, ..., \lambda_p\}$ are the distinct roots of the characteristic equation associated with eq. (D.4), that is $\Phi(B) = 0$, where $\Phi(B) = 1 - \sum_{m=1}^{p}\phi_m B^m = \prod_{m=1}^{p}(1-\lambda_m B)$ and

$$\zeta_{lk} = \frac{\lambda_l^{k+p-1}}{\prod\limits_{\substack{j=1\\ j\neq l}}^{m}(\lambda_l - \lambda_j)}, \quad \gamma_{lm} = (-1)^{m-1}\bigwedge_{j=1}^{m-1}\sum_{\substack{i_j=i_{j-1}+1\\ i_j\neq l}}^{p-m+j+1}\prod_{n=1}^{m-1}\lambda_{i_n},$$

with $\gamma_{l1} = 1$ and $k = t-s$.

First, we introduce some additional notation. For each fixed m such that $2 \leq m \leq p$, we define the nested sum operator

$$\bigwedge_{j=1}^{m-1} \sum_{i_j=i_{j-1}+1}^{p-m+j+1} \stackrel{\text{def}}{=} \sum_{i_1=1}^{p-m+2} \sum_{i_2=i_1+1}^{p-m+3} \cdots \sum_{i_{m-1}=i_{m-2}+1}^{p}$$

By expressing the AR parameters in terms of the roots of the AR polynomial, it is not difficult to show that

$$\phi_{m-1} = (-1)^m \bigwedge_{j=1}^{m} \sum_{i_j=i_{j-1}+1}^{p-m+j} \prod_{n=1}^{m} \lambda_{i_n}$$

and that $\sum_{l=1}^{p} \zeta_{lk}\gamma_{lm} = \sum_{r=1}^{p+1-m} \phi_{m-1+r}\xi(k-r)$, where $\xi(k)$ is given by eq. (8). Now, it follows from eq. (9) that:

$$\xi^{(m)}(t-s) = \sum_{r=1}^{p+1-m} \phi_{m-1+r}\xi(t-s-r) = \sum_{l=1}^{p} \zeta_{lk}\gamma_{lm}. \tag{D.7a}$$

Similarly, it is straightforward to show that:

$$\sum_{r=s+1}^{t} \xi(t-r) = \left[\frac{1}{\Phi(1)} - \sum_{m=1}^{p} \frac{\zeta_{mk}}{(1-\lambda_m)}\right]. \tag{D.7b}$$

Finally, eq. (8) entails that $\xi(t-r-l) = \sum_{m=1}^{p} \zeta_{m0}\lambda_m^{t-r-l}$. It follows directly that:

$$\xi_{s,q}(t,r) = \sum_{l=s+1-r}^{q} \xi(t-r-l)\theta_l = \sum_{m=1}^{p} \sum_{l=s+1-r}^{q} \zeta_{m0}\lambda_m^{t-r-l}\theta_l. \tag{D.7c}$$

Applying eqs. (D.7a), (D.7b) and (D.7c) to eq. (D.5), we obtain eq. (D.6), as asserted. We should highlight the fact that eq. (D.5), unlike eq. (D.6), includes the case of multiple characteristic roots.

E Hessenbergians and the Green Function

In this Section, we show that the functions, $\xi_q(t,r)$ and $\xi_{s,q}(t,r)$, defined by eqs. (13) and (14), can be expressed as banded Hessenbergians (for a discussion of Hessenbergians and their application to LDEs with variable coefficients, see Paraskevopoulos and Karanasos (2021) and the references cited there). Moreover, we discuss the Green function and its restriction involved in the solution of TV-LDEs(p).

E1 Hessenbergians

Formally, we can rephrase our results with $n = \max\{p,q\}$ in place of p and q. Therefore, without loss of generality, in the following Proposition we shall assume that $q = p$.

Proposition E.1 *$\xi_q(t,r)$ defined in eq.* (13) *can be expressed as a banded Hessenbergian of order $t-r+1$, whenever $t-r+1 > p$, which is given by:*

$$
\xi_q(t,r) = \tag{E.1}
$$

$$
\left|\begin{array}{ccccccccc}
1 & -1 & & & & & & & \\
\theta_1(r+1) & \phi_1(r+1) & -1 & & & & & & \\
\theta_2(r+2) & \phi_2(r+2) & \phi_1(r+2) & \ddots & & & & & \\
\vdots & \vdots & \vdots & \ddots & \ddots & & & & \\
\theta_q(r+q) & \phi_p(r+p) & \phi_{p-1}(r+p) & \ddots & \ddots & \ddots & & & \\
 & & \phi_p(r+p+1) & \ddots & \ddots & \ddots & \ddots & & \\
 & & & \ddots & \ddots & \ddots & \ddots & \ddots & \\
 & & & & \phi_p(t-1) & \phi_{p-1}(t-1) & \cdots & \phi_1(t-1) & -1 \\
 & & & & & \phi_p(t) & \cdots & \phi_2(t) & \phi_1(t)
\end{array}\right| .
$$

Proof. The cofactor of the first element in the first column of $\xi_q(t,r)$ in eq. (E.1) coincides with $\xi(t,r)$. Similarly the cofactor of the element $\theta_l(r+l)$, $l=1,\dots,q$, in the first column of $\xi_q(t,r)$ coincides with $\xi(t,r+l)$ (see Lemma A1). As a consequence, the cofactor expansion of $\xi_q(t,r)$ along its first column is

$$
\xi_q(t,r) = \xi(t,r) + \sum_{l=1}^{q} \xi(t,r+l)\theta_l(r+l),
$$

which coincides with eq. (13). ■

Proposition E.2 *$\xi_{s,q}(t,r)$ in eq. (14) can be expressed as a banded Hessenbergian of order k, whenever $k>p$:*

$$
\xi_{s,q}(t,r) =
$$

$$
\left|\begin{array}{ccccccccc}
\theta_{s+1-r}(s+1) & -1 & & & & & & & \\
\theta_{s+2-r}(s+2) & \phi_1(s+2) & -1 & & & & & & \\
\theta_{s+3-r}(s+3) & \phi_2(s+3) & \phi_1(s+3) & \ddots & & & & & \\
\vdots & \vdots & \vdots & \ddots & \ddots & & & & \\
\theta_q(r+q) & \phi_p(r+p) & \phi_{p-1}(r+p) & \ddots & \ddots & \ddots & & & \\
 & & \phi_p(r+p+1) & \ddots & \ddots & \ddots & \ddots & & \\
 & & & \ddots & \ddots & \ddots & \ddots & \ddots & \\
 & & & & \phi_p(t-1) & \phi_{p-1}(t-1) & \cdots & \phi_1(t-1) & -1 \\
 & & & & & \phi_p(t) & \cdots & \phi_2(t) & \phi_1(t)
\end{array}\right| .
$$

The proof of Proposition E.2 is similar to the proof of Proposition E.1 and, thus, it is omitted.

E2 Green Function

Let $s \in \mathbb{Z}$. Also let the AR coefficients $\phi_m(t)$ of eq. (1) be defined on $\mathbb{Z}_s$. The Green function $G(t,r)$ associated with eq. (7) is a two variable function defined on the domain $\mathbb{Z}_{s+1-p} \times \mathbb{Z}_s$ and takes the form of a ratio of two determinants whose matrix entries are elements of any fundamental set of solutions associated with eq. (7) (see Miller, 1968, p. 39 and Agarwal, 2000, p. 77). Having at our disposal the fundamental solution set Ξ_s (see Section 3.1), an explicit representation of $G(t,r)$ over the entire domain $\mathbb{Z}_{s+1-p} \times \mathbb{Z}_s$, is established by Paraskevopoulos and Karanasos (2021) (see their Theorem 4, eq. (54)). Both determinants involved in the ratio formula of $G(t,r)$ are expressed in terms of the elements $\xi^{(m)}(t,r)$ of Ξ_s and, ultimately, by eq. (9), in terms of the principal determinant $\xi(t,r)$ exclusively. It is also shown there that the restriction of the Green function for $t \in \mathbb{Z}_{s+1-p}$ and $s \leq r \leq t-1+p$, referred to as $H(t,r)$, coincides with the principal determinant function $\xi(t,r)$.

F Origins and Proofs

This Section is primarily devoted to the origins and history of the central notion of the paper, the principal determinant function, identified as a homogeneous solution to an infinite row-finite linear system. Some proofs reported in Sections 4 and 5 are also presented in the remaining Subsections of this Section.

F1 The Origins of the Principal Determinant

Linear difference equations with varying coefficients of order p (in brief TV-LDEs(p)) associated with TV-ARMA(p,q) processes (see eq. (3)) can be represented by infinite linear systems whose coefficient matrix is row-finite, as shown below[1]

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

or in more compact form as: $\mathbf{A}_s \; \mathbf{y}_s = \mathbf{v}_s$. The elements of the coefficient matrix $\mathbf{A}_s$ are the AR coefficients of eq. (1) at consecutive instances with starting instance: $s+1$. Formally $\mathbf{A}_s$ is an $\mathbb{N} \times \mathbb{N}$ row-finite matrix in generalized row echelon form (see Definition 2, in Paraskevopoulos, 2004).

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

Row-finite systems, in their general form, were first studied by Toeplitz (1909). Some results on finite linear systems were extended there to cover infinite row-finite ones. The representation of their solution was further developed by Fulkerson (1951). There he devised and proved the existence of a reduced form (generalizing finite matrices in row-reduced echelon form) for any arbitrary row-finite matrix with the aid of which the general solution of the system can be formulated. However, Fulkerson's proof was based on the countable axiom of choice, which is not constructive. The lack of a method for transforming row-finite matrices into their row reduced echelon form has been highlighted in Paraskevopoulos (2012), who in responding to this challenge has introduced a modified version of the Gauss-Jordan elimination algorithm equipped with a new pivot elimination strategy (called Infinite Gauss-Jordan elimination algorithm, briefly denoted as IGJEA). This strategy is grounded on the rightmost pivot nonzero elements of the matrix, yielding a lower row reduced form of the original row-finite matrix. This turns out to be essential for preserving the solutions of eq. (F.1) in the reduced system.[2] In a companion paper, Paraskevopoulos (2014) has further developed the IGJEA, focusing on the type and form of the general solution of row-finite linear systems. The algorithm is effectively applied to the infinite system representation of TV-LDEs(p), as in eq. (F.1), constructing the lower row reduced echelon form of $\mathbf{A}_s$, which is denoted as $\mathbf{LRREF}(\mathbf{A}_s)$ ($\mathbf{L}(\mathbf{A}_s)$ for short) and is given by

$$\mathbf{L}(\mathbf{A}_s) = \begin{bmatrix} -\xi^{(p)}(s+1,s) & -\xi^{(p-1)}(s+1,s) & \ldots & -\xi^{(1)}(s+1,s) & 1 & 0 & 0 & \ldots \\ -\xi^{(p)}(s+2,s) & -\xi^{(p-1)}(s+2,s) & \ldots & -\xi^{(1)}(s+2,s) & 0 & 1 & 0 & \ldots \\ -\xi^{(p)}(s+3,s) & -\xi^{(p-1)}(s+3,s) & \ldots & -\xi^{(1)}(s+3,s) & 0 & 0 & 1 & \ldots \\ \vdots & \vdots & \vdots\vdots\vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}.$$

This is a unique lower matrix form (called *row canonical*) of $\mathbf{A}_s$, which preserves the homogeneous solution of eq. (F.1), that is $\mathbf{A}_s = \mathbf{0} \iff \mathbf{L}(\mathbf{A}_s) = \mathbf{0}$. The first p opposite signed columns in $\mathbf{L}(\mathbf{A}_s)$ are the p fundamental (or linearly independent) solution sequences of the homogeneous system $\mathbf{A}_s\ \mathbf{y}_s = \mathbf{0}$, which represents eq. (7) and they are denoted as $\xi^{(m)}(t,s)$ for $t \geq s+1$, taking on the initial values $\xi^{(m)}(s+1-m,s) = 1$ and $\xi^{(m)}(s+1-j,s) = 0$ for $1 \leq j \leq p$ with $j \neq m$. The principal determinant is: $\xi(t,s) \stackrel{\text{def}}{=} \xi^{(1)}(t,s)$. An alternative method to the generation of $\xi(t,s)$ can be obtained by Cramer's rule (see Singh, 1980 and the references cited therein).

As an illustration, applying the IGJEA for $p = 2$, a few first terms of the principal determinant sequence $\{\xi(t,s)\}_{t\geq s+1}$, taking on the initial values $y_{s-1} = 0, y_s = 1$, are given below[3]:

$$y_{s+1} = \phi_1(s+1),\ y_{s+2} = \begin{vmatrix} \phi_1(s+1) & -1 \\ \phi_2(s+2) & \phi_1(s+2) \end{vmatrix},\ y_{s+3} = \begin{vmatrix} \phi_1(s+1) & -1 & \\ \phi_2(s+2) & \phi_1(s+2) & -1 \\ & \phi_2(s+3) & \phi_1(s+3) \end{vmatrix}, \ldots$$

[2]The property preserving the solutions is known as *left association* and in the finite dimensional case as *row-equivalence*.
[3]The IGJEA produces direct expansions of the principal determinant sequence.

Using the above values of y's, each outcome of the matrix product $\mathbf{A}_s\ \mathbf{y}_s$ is easily verified to be $\mathbf{0}$. The aforementioned solution sequence is the result of a rightmost pivot elimination strategy. As a counter example, employing the TV-LDE(1) (or TV-AR(1) model), it is shown in the above cited reference that the traditional (leftmost pivot) elimination strategy applied to the Gauss-Jordan algorithm fails to preserve the general homogeneous solution of the original TV-LDE(1). This was the main barrier which previously prevented researchers from implementing the Gauss-Jordan elimination algorithm for solving TV-LDEs and more generally row-finite systems. Linear difference equations with variable coefficients of irregular order are conveniently treated by the IGJEA too (see Paraskevopoulos, 2014).

Applying the same sequence of elementary operations, which reduce $\mathbf{A}_s$ to its $\mathbf{L}(\mathbf{A}_s)$, to the sequence of forcing terms $\{-\upsilon_{s+i}\}_{i\geq 1-p}$, a particular solution sequence is constructed. This is also represented by a lower Hessenbergian given by eq. (12) (see Proposition A.4).

The general homogeneous solution of eq. (F.1) turns out to be a linear combination of the obtained fundamental solutions with coefficients of any set of p initial condition values, say $\{c_m \in \mathbb{R}, 1 \leq m \leq p\}$. These initial condition values occupy the first p terms of the solution sequence $\{y_t\}_{t\geq s-p+1}$, that is $\{y_{s-p+1} = c_1, y_{s-p+2} = c_2, ...y_{s-1} = c_{p-1}, y_s = c_p\}$. The general (nonhomogeneous) solution of eq. (F.1) is equal to the sum of the general homogeneous solution plus the particular one both of which have been obtained above.

As the computational complexity for the calculation of banded-matrix determinants of order k (which coincides with the forecasting horizon) is $O(k)$, we conclude that the principal determinant representation of the Green function restriction, involved in the solution of the associated TV-LDE, is computationally tractable.[4] This is due to the Gaussian elimination algorithm, which requires approximately $\dfrac{k(p+1)^2}{4}$ multiplies, where $(p+1)$ is the bandwidth of the principal matrix, to reduce the aforementioned $k \times k$ matrix into its reduced row echelon form (see Thorson, 1979). Linear $O(k)$ computational complexity is comparable with the complexity of algorithms that calculate the Green function by recursion.

F2 Wold-Cramér Decomposition

In this Subsection, we show some results reported in Section 4.1 and in Appendix B.1.

F2.1 Proofs of Lemma B1 and Corollary B2

Proof of Lemma B1. *i)* Let us call $\tilde{\theta}_l = \sup_r |\theta_l(l+r)| \in \mathbb{R}_{\geq 0}$ for each $l = 1, ..., q$ and $\Theta = \max\limits_{0\leq l\leq q} \tilde{\theta}_l$, where $\theta_0(t) \stackrel{\text{def}}{=} 1$ for all t. Now $\xi_q(t,r)$ in eq. (13) can be rewritten as $\xi_q(t,r) = \sum_{l=0}^{q} \xi(t, r+l)\theta_l(r+l)$, whence:

$$|\xi_q(t,r)| \leq \left|\sum_{l=0}^{q} \xi(t, r+l)\tilde{\theta}_l\right| \leq \Theta \sum_{l=0}^{q} |\xi(t, r+l)|\,. \tag{F.2}$$

[4] An explicit expression evaluated in polynomial running time is referred to as computationally tractable.

Since $\xi(t,r) = 0$, whenever $r > t$, it follows that $\sum_{r=t+1}^{t+l} |\xi(t,r)| = 0$ for all t and any l such that $0 \leq l \leq q$. The latter result along with condition (17), that is $\sum_{r=-\infty}^{t} |\xi(t,r)| < \infty$, imply that

$$\sum_{r=-\infty}^{t} |\xi(t,r+l)| = \sum_{r=-\infty}^{t+l} |\xi(t,r)| = \sum_{r=-\infty}^{t} |\xi(t,r)| + \sum_{r=t+1}^{t+l} |\xi(t,r)| = \sum_{r=-\infty}^{t} |\xi(t,r)| < \infty \tag{F.3}$$

for all t and any $l : 0 \leq l \leq q$. In view of (F.3), it follows that $\sum_{l=0}^{q} \sum_{r=-\infty}^{t} |\xi(t,r+l)| < \infty$. Combining the latter result with (F.2), it follows that:

$$\sum_{r=-\infty}^{t} |\xi_q(t,r)| \leq \Theta \sum_{r=-\infty}^{t} \sum_{l=0}^{q} |\xi(t,r+l)| = \Theta \sum_{l=0}^{q} \sum_{r=-\infty}^{t} |\xi(t,r+l)| < \infty \quad \text{for all } t. \tag{F.4}$$

Statement (i) of the Lemma follows from the inequalities in (F.4).

ii) The absolute summability condition in (17) and the boundedness of the drift-process, that is $|\varphi(r)| \leq N < \infty$ for some $N \in \mathbb{R}_{\geq 0}$ and all $r \in \mathbb{Z}$, imply the convergence of $\sum_{r=-\infty}^{t} |\xi(t,r)\varphi(r)|$, since:

$$\sum_{r=-\infty}^{t} |\xi(t,r)\varphi(r)| \leq N \sum_{r=-\infty}^{t} |\xi(t,r)| < \infty.$$

Thus $\sum_{r=-\infty}^{t} \xi(t,r)\varphi(r)$ is a convergent numerical series in $\mathbb{R}$ for all $t \in \mathbb{Z}$ and therefore

$$\mathbb{E}\big(\sum_{r=-\infty}^{t} \xi(t,r)\varphi(r)\big) = \sum_{r=-\infty}^{t} \xi(t,r)\varphi(r) \in \mathbb{R} \text{ (finite)}$$

for all t, as required.

iii) Recalling that $u_r = \sum_{l=0}^{q} \theta_l(r)\varepsilon_{r-l}$, and $\mathbb{E}(\varepsilon_r) = 0$ the linearity of $\mathbb{E}$ entails that $\mathbb{E}(u_r) = 0$, whence

$$\mathbb{E}(\upsilon_r) = \mathbb{E}(\varphi(r)) + \mathbb{E}(u_r) = \varphi(r) \quad \text{for all } r. \tag{F.5}$$

As $\mathbb{E}(\varepsilon_r^2) = \sigma^2(r)$ and $\{\varepsilon_r\}$ are uncorrelated, it follows that $\mathbb{E}(u_r^2) = \sum_{l=0}^{q} \theta_l^2(r)\sigma^2(r-l)$.[5] Moreover, as $0 < \sigma^2(r) \leq M < \infty$, we conclude that

$$\mathbb{V}ar(\varphi(r) + u_r) = \mathbb{V}ar(u_r) = \mathbb{E}(u_r^2) = \sum_{l=0}^{q} \theta_l^2(r)\sigma^2(r-l) \leq M \sum_{l=0}^{q} \theta_l^2(r) \leq M \sum_{l=0}^{q} \Theta^2 = M(q+1)\Theta^2 \tag{F.6}$$

for all r, whence $\sup_r \mathbb{V}ar(\upsilon_r) \leq M(q+1)\Theta^2 < \infty$. Combining eqs. (F.5) with (F.6), we infer:

$$\mathbb{E}(\upsilon_r^2) = \mathbb{V}ar(\upsilon_r) + (\mathbb{E}(\upsilon_r))^2 < M(q+1)\Theta^2 + N^2 \text{ for all } r.$$

[5]By definition $u_r = \sum_{l=0}^{q} \theta_l(r)\varepsilon_{r-l}$. Taking expectations to both sides of the well known algebraic identity

$$\Big(\sum_{l=0}^{q} \theta_l(r)\varepsilon_{r-l}\Big)^2 = \sum_{l=0}^{q} \theta_l^2(r)\varepsilon_{r-l}^2 + 2\sum_{l=0}^{q}\sum_{j=0}^{l-1} \theta_l(r)\theta_j(r)\varepsilon_{r-l}\varepsilon_{r-j}$$

on account of $\mathbb{E}(\varepsilon_{r-l}^2) = \sigma^2(r-l)$ and $\langle \varepsilon_{r-l}, \varepsilon_{r-j} \rangle = \mathbb{E}(\varepsilon_{r-l}\varepsilon_{r-j}) = 0$, whenever $j \leq l-1$, the expectation of the double summation in the right-hand side of the above identity becomes zero and the result follows.

Letting $V = M(q+1)\Theta^2 + N^2$, the latter inequality shows that $\sup_t \mathbb{E}(\upsilon_t^2) \leq V < \infty$, as required. ■

Proof of Corollary B2. By virtue of eq. (B.3), $\sum_{r=-\infty}^{t} \xi(t,r)u_r - z_t = 0$ or equivalently $\sum_{r=-\infty}^{t} \big(\xi(t,r)u_r - \xi_q(t,r)\varepsilon_r\big) = 0$ in L_2, that is

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

Thereby

$$\lim_{s\to-\infty} || \sum_{r=s+1}^{t} \xi(t,r)u_r - z_t||_{L_2}^2 = 0. \tag{F.7}$$

In view of eq. (15), the following equalities hold

$$\begin{aligned}
\lim_{s\to-\infty} \mathbb{E}\left(\sum_{r=s+1-q}^{s} \xi_{s,q}(t,r)\varepsilon_r\right)^2 &= \lim_{s\to-\infty} \mathbb{E}\left(\sum_{r=s+1}^{t} \xi(t,r)u_r - z_t\right)^2 \\
\text{(the norm definition of } L_2\text{)} &= \lim_{s\to-\infty} || \sum_{r=s+1}^{t} \xi(t,r)u_r - z_t||_{L_2}^2 \\
\text{(by eq. (F.7))} &= 0,
\end{aligned}$$

as claimed. ■

F2.2 Remark F.1 and Proposition F.1

In the following Proposition, we show some results reported in Section 4.1, which support the proof of Theorem 3. We start with the following Remark, which presents some well known results on L_2 spaces.

Remark F.1 For each fixed $t \in \mathbb{Z}$, the closed span of a subset $\mathbf{e}_t = \{e_r, -\infty < r \leq t\}$ of the Hilbert space L_2, denoted here by $\mathcal{M}_t(e)$, is the smallest closed subspace of L_2, which contains $\mathbf{e}_t$. In all that follows $\mathbf{e}_t$ is assumed to be an orthonormal set. Therefore $\mathcal{M}_t(e)$ turns into a separable Hilbert space and $\mathbf{e}_t$ is called the orthonormal basis of $\mathcal{M}_t(e)$, that is every $x_t \in \mathcal{M}_t(e)$ has a unique representation

$$x_t = \sum_{r=-\infty}^{t} \langle x_t, e_r\rangle e_r,$$

where the inner product of x_t and e_r, i.e., $\langle x_t, e_r\rangle$, are the Fourier coefficients of the orthonormal expansion of x_t. Moreover, the second order moment of x_t is:

$$||x_t||_{L_2}^2 = \mathbb{E}(x_t^2) = \sum_{r=-\infty}^{t} |\langle x_t, e_r\rangle|^2$$

(see, for example, Brockwell and Davis, 1991, Theorem 2.4.2).

As a consequence, let $\{\lambda_r, r \leq t\}$ be any sequence in $\mathbb{R}$. Then $x_t = \sum_{r=-\infty}^{t} \lambda_r\, e_r \in \mathcal{M}_t(e)$ if and only if $\sum_{r=-\infty}^{t} \lambda_r^2 < \infty$. In this case $\lambda_r = \langle x_t, e_r\rangle$ for all $r \leq t$ and $\mathbb{E}(x_t^2) = \sum_{r=-\infty}^{t} \lambda_r^2$.

Proposition F.1 *i) Let $\{y_t\}$ be a second order process. Then the process $\{y_t\}$ satisfies eq.* (1) *if and only if the processes $\{y_t - \mathbb{E}(y_t)\}$ and $\{\mathbb{E}(y_t)\}$ simultaneously satisfy eqs.* (19a) *and* (19b)*, respectively, where $\{y_t - \mathbb{E}(y_t)\}$ is a zero mean random process and $\{\mathbb{E}(y_t)\}$ is a nonrandom process. ii) Under the conditions of Theorem 3, $y_t - \mathbb{E}(y_t) = \sum_{r=-\infty}^{t} \xi_q(t,r)\varepsilon_r \in \mathcal{M}_t(\varepsilon)$, which yields a mean zero process $\{y_t - \mathbb{E}(y_t)\}$ in $\mathcal{M}(\varepsilon)$ and uniquely solves eq.* (19a)*, whereas $\{\mathbb{E}(y_t)\}$ solves eq.* (19b) *for $\mu_t = \mathbb{E}(y_t)$.*

Proof. *i*) We start with the direct implication. Since $y_t \in L_2$, it follows that $\mathbb{E}(y_t)$ exists. Moreover, as $\mathbb{E}(\varepsilon_r) = 0$, it follows from the linearity of the expectation operator that:

$$\mathbb{E}(u_t) = \mathbb{E}(\varepsilon_t + \sum_{l=1}^{q} \theta_l(t)\varepsilon_{t-l}) = 0.$$

Taking expectations to both sides of eq. (1), on account of $\mathbb{E}(\varphi(t)) = \varphi(t)$, it follows immediately that

$$\mathbb{E}(y_t) = \varphi(t) + \sum_{m=1}^{p} \phi_m(t)\mathbb{E}(y_{t-m}) \tag{F.8a}$$

and so $\mathbb{E}(y_t)$ solves eq. (19b), when applied for $\mu_t = \mathbb{E}(y_t)$. Let us rewrite eq. (19b) (or eq. (F.8a)) as: $\mathbb{E}(y_t) - \sum_{m=1}^{p} \phi_m(t)\mathbb{E}(y_{t-m}) = \varphi(t)$. Then eq. (1) can be rewritten as

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

or equivalently as

$$y_t - \mathbb{E}(y_t) = \sum_{m=1}^{p} \phi_m(t)\Big(y_{t-m} - \mathbb{E}(y_{t-m})\Big) + u_t. \tag{F.8b}$$

The latter shows that the process $\{y_t - \mathbb{E}(y_t)\}$ solves eq. (19a) and the direct implication follows.

To see the converse implication add eqs. (F.8a) and (F.8b) to get y_t in eq. (1) and the equivalence is complete. Finally since

$$\mathbb{E}(y_t - \mathbb{E}(y_t)) = \mathbb{E}(y_t) - \mathbb{E}(y_t) = 0,$$

$\{y_t - \mathbb{E}(y_t)\}$ is a mean zero random process, which satisfies (19a).

ii) As we have shown in Proposition B1(iii), under the assumptions of Theorem 3, the solution y_t of eq. (1) is second order and $z_t = \sum_{r=-\infty}^{t} \xi_q(t,r)\varepsilon_r \in \mathcal{M}_t(\varepsilon)$ yields a mean zero process in $\mathcal{M}(\varepsilon)$. To show that z_t solves eq. (19a), we proceed as follows. In view of eq. (B.3), proved in the main paper, we can rewrite $z_t = \sum_{r=-\infty}^{t} \xi(t,r)u_r$. Recalling that $v_t = \varphi(t) + u_t$ and taking into account that $\varphi(t) = 0$, we can replace v_t with u_t in eq. (B.4) to get

$$\sum_{r=-\infty}^{t} \xi(t,r)u_r = \sum_{m=1}^{p} \phi_m(t) \sum_{r=-\infty}^{t-m} \xi(t-m,r)u_r + u_t,$$

which shows that $\{z_t\}$ also solves eq. (19a) in $\mathcal{M}(\varepsilon)$, that is

$$z_t = \sum_{m=1}^{p} \phi_m(t) z_{t-m} + u_t.$$

To show the uniqueness of the solution $z_t = \sum_{r=-\infty}^{t} \xi_q(t,r)\varepsilon_r$ consider any other mean zero process, say x_t, which also solves eq. (19a), that is

$$x_t = \sum_{m=1}^{p} \phi_m(t) x_{t-m} + u_t.$$

It follows that

$$z_t - x_t = \sum_{m=1}^{p} \phi_m(t)\big(z_{t-m} - x_{t-m}\big),$$

that is $\{z_t - x_t\}$ is a nonrandom process and so $\mathbb{V}ar(z_t - x_t) = 0$. Moreover, as $\mathbb{E}(z_t) = \mathbb{E}(x_t) = 0$, it follows that $\mathbb{E}(z_t - x_t) = 0$. As a consequence, we have:

$$||z_t - x_t||_{L_2}^2 = \mathbb{E}\Big(\big(z_t - x_t\big)^2\Big) = \mathbb{V}ar(z_t - x_t) + \Big(\mathbb{E}(z_t - x_t)\Big)^2 = 0.$$

We thus conclude that $||z_t - x_t||_{L_2} = 0$, whence $z_t = x_t$ (in L_2 sense). ■

The uniqueness result in Theorem 3 can be equivalently rephrased as follows. Under the conditions of Theorem 3, any two distinct second order processes which solve eq. (1), say $\{y_t\}, \{y_t^*\}$, must differ only in their first order moments. This follows from the fact that their demean random processes coincide (in the L_2 sense) on $\mathcal{M}(\varepsilon)$, that is: $y_t - \mathbb{E}(y_t) = y_t^* - \mathbb{E}(y_t^*) = z_t$ for all t, whenever both processes $\{\mathbb{E}(y_t)\}, \{\mathbb{E}(y_t^*)\}$ satisfy eq. (19b).

F3 Explicit Representation of First Moment Vector Processes

The explicit form of the elements of the Casorati matrix $\boldsymbol{\Xi}_{t,s}$[6] associated with the fundamental solution set Ξ_s of eq. (7) is employed to obtain computationally feasible formulas for any p-dimensional vector with components consecutive first moment solution $\{\mu_t\}$ of eq. (19b), provided that a p-dimensional initial value first moment vector has been estimated from the information data. The explicit form of $\boldsymbol{\Xi}_{t,s}$ has some computational advantages, discussed below eq. (F.12).

[6] According to Agarwal's terminology and notation (see Agarwal, 2000, p.56) $\boldsymbol{\Xi}_{t,s}$ is called the principal fundamental matrix and is denoted there as $\mathcal{U}(k,k_0)$, that is $\mathcal{U}(k,k_0) = \boldsymbol{\Xi}_{k,k_0}$ for $k \geq k_0$. However, unlike the elements of $\boldsymbol{\Xi}_{k,k_0}$, the elements of $\mathcal{U}(k,k_0)$ are not explicitly expressed. This has some remarkable consequences discussed in this Section.

Let the estimated vector be $\boldsymbol{\mu}_s = [\mu_s, \mu_{s-1}, ..., \mu_{s+1-p}]$. In particular, it is shown below that one can directly calculate the first moment vector $\boldsymbol{\mu}_t$ for all $t \in \mathbb{Z}$, whose elements are generated by eq. (19b), taking on the components of $\boldsymbol{\mu}_s$ as initial values (see the explicit formulas in eq. (F.13) below). In doing so we use the explicit form of $\boldsymbol{\Xi}_{t,i}$ (see Proposition A2, in the main paper), which is given by

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

where $i \leq t$ and $\xi^{(m)}(t,i)$ are banded Hessenbergian coefficients, with entries the AR coefficients of the model (see eq. (A.2)). Notice that $\xi^{(m)}(t,i)$, and therefore $\boldsymbol{\Xi}_{t,i}$, are explicitly defined for $t, i \in \mathbb{Z}$.

In the following two paragraphs, entitled "Forward" and "Backward", we assume that $t, i \in \mathbb{Z}$ and $t \geq i$. Moreover, $\boldsymbol{\varphi}_i = [\varphi(i), 0, ..., 0]$, where $\varphi(i)$ is the time-varying drift. Recall that $\boldsymbol{\Xi}_{t,t}$ is the identity matrix for all $t \in \mathbb{Z}$.

Forward: Given $\boldsymbol{\mu}_s = [\mu_s, \mu_{s-1}, ..., \mu_{s+1-p}]$, the vector $\boldsymbol{\mu}_t$ for $t > s$ is explicitly represented by:

$$\boldsymbol{\mu}_t = \boldsymbol{\Xi}_{t,s}\boldsymbol{\mu}_s + \sum_{i=s+1}^{t} \boldsymbol{\Xi}_{t,i}\boldsymbol{\varphi}_i. \tag{F.9}$$

To see this, we can rewrite eq. (19b), applied for $t = s+1$, in vector form as

$$\boldsymbol{\mu}_{s+1} = \boldsymbol{\varphi}_{s+1} + \boldsymbol{\Gamma}_{s+1}\boldsymbol{\mu}_s, \tag{F.10}$$

where $\boldsymbol{\Gamma}_i$ is the $p \times p$ companion matrix, which is given by:

$$\boldsymbol{\Gamma}_i = \begin{bmatrix} \phi_1(i) & \phi_2(i) & ... & \phi_{p-1}(i) & \phi_p(i) \\ 1 & 0 & ... & 0 & 0 \\ 0 & 1 & ... & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & ... & 1 & 0 \end{bmatrix}.$$

Following Paraskevopoulos and Karanasos (2021) (see Theorem 3 therein) we can write:

$$\boldsymbol{\Xi}_{t,s} = \begin{cases} \boldsymbol{\Gamma}_t\boldsymbol{\Gamma}_{t-1}...\boldsymbol{\Gamma}_{s+1}, & \text{if } t \geq s+1 \\ \mathbf{I}, & \text{if } t = s. \end{cases} \tag{F.11}$$

Applying recursion to eq. (F.10), we obtain the sequence of vector means $\boldsymbol{\mu}_{s+2}, \boldsymbol{\mu}_{s+3}, ...\boldsymbol{\mu}_t, ...$ as follows:

$$\begin{aligned}
\boldsymbol{\mu}_t &= \boldsymbol{\Gamma}_t\boldsymbol{\Gamma}_{t-1}...\boldsymbol{\Gamma}_{s+1}\boldsymbol{\mu}_s + \sum_{i=s+1}^{t-1} \boldsymbol{\Gamma}_t\boldsymbol{\Gamma}_{t-1}...\boldsymbol{\Gamma}_{i+1}\boldsymbol{\varphi}_i + \boldsymbol{\varphi}_t \\
\text{(by eq. (F.11))} &= \boldsymbol{\Xi}_{t,s}\boldsymbol{\mu}_s + \sum_{i=s+1}^{t-1} \boldsymbol{\Xi}_{t,i}\boldsymbol{\varphi}_i + \boldsymbol{\Xi}_{t,t}\boldsymbol{\varphi}_t \\
(\text{as } \boldsymbol{\Xi}_{t,t} = \mathbf{I}) &= \boldsymbol{\Xi}_{t,s}\boldsymbol{\mu}_s + \sum_{i=s+1}^{t} \boldsymbol{\Xi}_{t,i}\boldsymbol{\varphi}_i
\end{aligned}$$

as stated in eq. (F.9).

Backward: Given $\boldsymbol{\mu}_t$, the vector $\boldsymbol{\mu}_s$ for $s < t$ is derived by solving eq. (F.9) for $\boldsymbol{\mu}_s$, that is by applying the inverse of $\boldsymbol{\Xi}_{t,s}$ to both sides of eq. (F.9), which is given by:

$$\boldsymbol{\mu}_s = \boldsymbol{\Xi}_{t,s}^{-1}\boldsymbol{\mu}_t - \sum_{i=s+1}^{t} \boldsymbol{\Xi}_{t,s}^{-1}\boldsymbol{\Xi}_{t,i}\boldsymbol{\varphi}_i.$$

As shown in Paraskevopoulos and Karanasos (2021), the matrix product $\boldsymbol{\Xi}_{t,s}\boldsymbol{\Xi}_{i,s}^{-1}$ for $t \geq s$ and $i \geq s$, coincides with the Green matrix $\mathbf{G}(t,i)$ associated with eq. (7). Let $s \leq i \leq t$. Then, we have:

$$\begin{aligned}
\boldsymbol{\Xi}_{t,s}\boldsymbol{\Xi}_{i,s}^{-1} &= (\boldsymbol{\Gamma}_t\boldsymbol{\Gamma}_{t-1}...\boldsymbol{\Gamma}_{i+1}\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_{i-1}...\boldsymbol{\Gamma}_{s+1})(\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_{i-1}...\boldsymbol{\Gamma}_{s+1})^{-1} \\
\text{(elementary property of invertible matrices)} &= \boldsymbol{\Gamma}_t\boldsymbol{\Gamma}_{t-1}...\boldsymbol{\Gamma}_{i+1}\boldsymbol{\Gamma}_i\boldsymbol{\Gamma}_{i-1}...\boldsymbol{\Gamma}_{s+1}\boldsymbol{\Gamma}_{s+1}^{-1}...\boldsymbol{\Gamma}_{i-1}^{-1}\boldsymbol{\Gamma}_i^{-1} \\
(\text{since } s \leq i \leq t) &= \boldsymbol{\Gamma}_t\boldsymbol{\Gamma}_{t-1}...\boldsymbol{\Gamma}_{i+1} \\
\text{(by eq. (F.11))} &= \boldsymbol{\Xi}_{t,i}. \qquad \text{(F.12)}
\end{aligned}$$

Applying eq. (F.12) to eq. (F.9), we equivalently obtain the standard variation of constants formula (see Agarwal, 2000, eq. (2.5.1)), that is:

$$\boldsymbol{\mu}_t = \boldsymbol{\Xi}_{t,s}\boldsymbol{\mu}_s + \sum_{i=s+1}^{t} \boldsymbol{\Xi}_{t,s}\boldsymbol{\Xi}_{i,s}^{-1}\boldsymbol{\varphi}_i.$$

Thanks to the explicit form of $\boldsymbol{\Xi}_{t,i}$ for all $i \leq t$, its replacement by $\boldsymbol{\Xi}_{t,s}\boldsymbol{\Xi}_{i,s}^{-1}$, as in eq. (F.12), is not required. This is due to the fact that both the matrix $\boldsymbol{\Xi}_{t,i}$ and its elements are functions of t, i, which releases us from the need to work within the same set of fundamental solutions (Ξ_s), all of which started at a fixed time point s. Accordingly, we can directly apply the fully explicit form in eq. (F.9), while t, i vary, saving a significant amount of computational time.

As a consequence, the estimated vector $\boldsymbol{\mu}_s = [\mu_s, \mu_{s-1}, ..., \mu_{s+1-p}]$ can be used as the initial condition vector to produce forward and backward vector means. Their components are elements of the process μ_t

for $t \in \mathbb{Z}$ in eq. (19b), given by the unified formula:

$$\boldsymbol{\mu}_t = \begin{cases} \boldsymbol{\Xi}_{t,s}\boldsymbol{\mu}_s + \sum_{i=s+1}^{t} \boldsymbol{\Xi}_{t,i}\boldsymbol{\varphi}_i & \text{if } \ t \geq s \ \text{ (forward)} \\ \boldsymbol{\Xi}_{s,t}^{-1}\boldsymbol{\mu}_s - \sum_{i=t+1}^{s} \boldsymbol{\Xi}_{s,t}^{-1}\boldsymbol{\Xi}_{s,i}\boldsymbol{\varphi}_i & \text{if } \ t < s \ \text{ (backward).} \end{cases} \tag{F.13}$$

The first (resp. second) branch of eq. (F.13) coincides with eq. (2.5.1) (resp. (2.5.2)) in Agarwal (2000).[7]

F4 Decomposable Bernoulli Shift

In Remark 1 (Section 4.1.), we discussed an application of y_t in eq. (18a) in Theorem 2, which potentially generates a decomposable Bernoulli shift. In this Subsection, we provide the Definition of an L_v-decomposable Bernoulli shift, as introduced in Massacci and Trapani (2022).[8]

Definition F.1 *The sequence $\{y_t, -\infty < t < \infty\}$ forms an L_v-decomposable Bernoulli shift if and only if it holds that: $y_t = g(\varepsilon_t, \varepsilon_{t-1}, ...)$, where: $g(\cdot) : \mathbb{S}^\infty \to \mathbb{R}$ is a nonrandom measurable function; $\{\varepsilon_t, -\infty < t < \infty\}$ is an i.i.d. sequence with values in a measurable space $\mathbb{S}$; $\mathbb{E}(y_t) = 0$, $\mathbb{E}(|y_t|^v) < \infty$; $\left| y_t - y^*_{t,l} \right|_v \leq c_0 l^{-a}$, for some $c_0 > 0$ and $a > 0$, where $y^*_{t,l} = g(\varepsilon_t, \varepsilon_{t-1}, ..., \varepsilon_{t-l+1}, \varepsilon^*_{t-l,t,l}, \varepsilon^*_{t-l-1,t,l}, \ldots)$, with $\{\varepsilon^*_{i,j,l}, -\infty < i, j, l < \infty\}$ i.i.d. copies of ε_0 independent of $\{\varepsilon_t, -\infty < t < \infty\}$.*

As pointed out in the above cited reference (see p.5 in their paper), "Since the seminal works by Wu (2005) and Berkes et al. (2011) (see also Hörmann, 2009), decomposable Bernoulli shifts have proven a very convenient way to model and study dependent time series, mainly due to their generality and to the fact that it is much easier to verify whether a sequence forms a decomposable Bernoulli shift than e.g., verifying mixing conditions. Virtually all the most commonly employed DGPs in econometrics and statistics can be shown to generate decomposable Bernoulli shifts: [...], nonlinear time series models (such as e.g., random coefficient autoregressive models and threshold models)."

F5 Regularity Conditions

Originating with the pioneering work of Neimi (1983) on nonstationary ARMA processes with constant coefficients in a series of papers, AR and MA regularity conditions (regularity conditions for short) were introduced to cover time-varying demean models, guaranteeing the existence and uniqueness of the second order purely nondeterministic solution of eq. (19a) (see Singh and Peiris, 1987, Kowalski and Szynal,

[7]Due to the conventional notation employed in the definition of the principal fundamental matrix $\mathcal{U}(k, k_0)$ in the second branch of eq. (2.4.2) in Agarwal (2000), it turns out that $\mathcal{U}(k, k_0)$ for $k_0 > k$ coincides with the inverse of $\boldsymbol{\Xi}_{k_0,k}$, that is $\mathcal{U}(k, k_0) = \boldsymbol{\Xi}^{-1}_{k_0,k}$ for $k_0 > k$. Taking the inverses in the latter matrix equality, we conclude that $\mathcal{U}^{-1}(k, k_0) = \boldsymbol{\Xi}_{k_0,k}$ for $k_0 > k$. These verify the equivalence of eq. (2.5.2) with the second branch of eq. (F.13), when the latter is applied for $s = k_0$, $t = k$ and $i = \ell$.

[8]Thanks to an anonymous reviewer of this paper who provided us with the suitable references that cover a great deal of the relevant material.

1991). In the former of the last two references, the regularity conditions imply UBLS (Uniformly Bounded Linear Stationary), while in the latter, being more general, they do not. However, the Green function solution representations in both references coincide and so do the Hessenbergian solution representations. As a consequence, under the regularity conditions, we can obtain explicit solution formulas in L_2, in terms of banded Hessenbergians.

In what follows we consider eq. (1) with $\varphi(t) = 0$ for all $t \in \mathbb{Z}$, which in a causal environment, generates mean zero random solution processes described by eq. (19a), thus in this case using the identification: $y_t = z_t = \sum_{r=-\infty}^{t} \xi_q(t,r)\varepsilon_r$. The latter is also an explicit representation of the solution derived from the regularity conditions (see, for example eq. (2.5), in Singh and Peiris, 1987). We have shown in Proposition B1(ii) in Appendix B.1 that the condition $\sum_{r=-\infty}^{t} \xi_q^2(t,r) < \infty$ is necessary and sufficient for $y_t \in L_2$, provided that $0 < m \leq \sigma^2(r) \leq M < \infty$ for all $r \in \mathbb{Z}$. The above results are summarized in the following implications

$$\{\text{Regularity Conditions}\} \implies y_t \in L_2 \iff \sum_{r=-\infty}^{t} \xi_q^2(t,r) < \infty,$$

which show that $\sum_{r=-\infty}^{t} \xi_q^2(t,r) < \infty$ for all $t \in \mathbb{Z}$ is necessary for the regularity conditions to hold.

G Forecasting

The deferred proofs, reported in the homonymous matching Section of the main body of the paper, are included herein.

G1 Invertibility

A second order mean zero DTV-ARMA(p,q) process z_r is invertible if and only if ε_t can be expressed as a convergent series of the present and past random variables z_r ($r \leq t$) (see Brockwell and Davis, 2016, p.76). The main result of this Section is presented in Theorem G.1 below.

Eq. (1) applied with $\varphi(t) = 0$ (see also eq. (19a)) for all t can be rewritten as

$$\varepsilon_t = z_t - \sum_{l=1}^{q} \theta_l(t)\varepsilon_{t-l} - \sum_{m=1}^{p} \phi_m(t) z_{t-m}, \tag{G.1}$$

where $z_t = y_t - \mathbb{E}(y_t)$. The principal matrix associated with the moving average part is defined by

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

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

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

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

Formally $\vartheta(t,s)$ (and similarly to $\xi(t,s)$), is a banded Hessenbergian. We further extent its definition by assigning the initial values: $\vartheta(s,s) = 1$ and $\vartheta(t,s) = 0$ for $t < s$. Accordingly, $\vartheta(t,s)$ is the Green function associated with the MA(q) operator.

By analogy with the definition of $\xi_q(t,r)$ (see eq. (13)), we define:

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

In the following Theorem we give a sufficient condition for a causal mean zero DTV-ARMA(p,q) process determined by z_t in $\mathcal{M}(\varepsilon)$ to be invertible.

Theorem G.1 *Let the absolute summability condition in* (17) *hold. Also let the absolute summability condition* $\sum_{r=-\infty}^{t} |\vartheta(t,r)| < \infty$ *hold for all* t. *Then the process* $\{z_t\}$ *is invertible, that is*

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

and solves eq. (G.1).

The proof of Theorem G.1 essentially repeats the arguments of the proof in Proposition B.1(iii), switching the roles of z_r and ε_r, $\xi(t,r)$ and $\vartheta(t,r)$ and $\xi_q(t,r)$ and $\vartheta_p(t,r)$. Theorem G.1 recovers the formula devised by Hallin (1979) (see eq. (16) in his Theorem 3), but this time in a fully explicit form.

G2 Optimal Linear Forecasts

Forecasts Based on Infinite Observations. In an invertible environment, the inversion of z_t in Theorem G.1, yields an optimal linear predictor, based on $\mathcal{M}_s(z)$, which is explicitly expressed exclusively in terms of z_r for $r \leq s$ as

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

Accordingly, the k-step-ahead optimal linear predictor of y_t takes the form:

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

Now the observable random variables are $\{z_r\}_{r\leq s}$.

If the conditions of Theorem 2 hold, $\{y_t\}$ is a unique asymptotically stable second order solution process given by eq. (18a). In this case the k-step-ahead optimal linear predictor formulas are modified by replacing $\mathbb{E}(y_t)$ with $\sum_{r=-\infty}^{t} \xi(t,r)\varphi(r)$.

Forecasts Based on finite Observations. Let $\{\mu_t\}$ be the estimated first moment process generated by eq. (19b) and $\mathbb{E}(y_t) = \mu_t$. Taking into account that z_t is the unique mean zero solution process of eq. (19a) in $\mathcal{M}(\varepsilon)$ and that $y_t = \mu_t + z_t$, the optimal linear predictor in eq. (23a) takes the form:

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

As a consequence of the latter expression of $\hat{\mathbb{E}}(y_t \,|\, \mathcal{K}_s)$, the observable random variables y_{s+1-m} in eq. (23a) are now replaced with the demean random variables z_{s+1-m} for $1 \leq m \leq p$. In an invertible environment the q past observations of ε_r in the above equation are realizations of the prescribed random variables $\varepsilon_r = \sum_{i=-\infty}^{r} \vartheta_p(r,i)z_i$ for $s+1-q \leq r \leq s$.

In what follows we show Proposition 4. Recall that a second order solution process of eq. (1) can be expressed as $y_t = \mathbb{E}(y_t) + \sum_{r=-\infty}^{t} \xi_q(t,r)\varepsilon_r$ (see eq. (20a)).

Proof of Proposition 4. Consider any arbitrary element $S(\varepsilon_j, y_m, 1)$ of $\mathcal{K}_s$, that is any linear combination of the past observations: $S(\varepsilon_j, y_m, 1) = \sum_{m=1}^{p} a_m y_{s+1-m} + \sum_{j=1}^{q} b_j \varepsilon_{s+1-j} + c$, where a_m, b_j, c are arbitrary scalars. Subtracting eq. (23a) from eq. (16), we have $y_t - \hat{\mathbb{E}}(y_t \,|\, \mathcal{K}_s) = \sum_{r=s+1}^{t} \xi_q(t,r)\varepsilon_r$, whence $\mathbb{FE}_{t,s} = \sum_{r=s+1}^{t} \xi_q(t,r)\varepsilon_r$. As $\hat{\mathbb{E}}(y_t \,|\, \mathcal{K}_s) \in \mathcal{K}_s$, it suffices to show that $\big(y_t - \hat{\mathbb{E}}(y_t \,|\, \mathcal{K}_s)\big) \perp S(\varepsilon_j, y_m, 1)$, that is $\mathbb{FE}_{t,s} \perp S(\varepsilon_j, y_m, 1)$. In other words, we must show that the inner product $\langle \mathbb{FE}_{t,s}, S(\varepsilon_j, y_m, 1)\rangle = 0$

or equivalently that $\mathbb{E}\big(\mathbb{FE}_{t,s}\ S(\varepsilon_j, y_m, 1)\big) = 0$. Expanding the product

$$\mathbb{FE}_{t,s}S(\varepsilon_j, y_m, 1) = \Big(\sum_{r=s+1}^{t} \xi_q(t,r)\varepsilon_r\Big)\Big(\sum_{m=1}^{p} a_m y_{s+1-m} + \sum_{j=1}^{q} b_j \varepsilon_{s+1-j} + c\Big)$$

we obtain:

$$\mathbb{FE}_{t,s}S(\varepsilon_j, y_m, 1) = \sum_{m=1}^{p}\sum_{r=s+1}^{t} a_m \xi_q(t,r)\varepsilon_r y_{s+1-m} + \sum_{j=1}^{q}\sum_{r=s+1}^{t} b_j \xi_q(t,r)\varepsilon_r \varepsilon_{s+1-j} + c \sum_{r=s+1}^{t} \xi_q(t,r)\varepsilon_r.$$

Taking expectations to both sides of the above formula, it follows from the linearity of the expectation operator that:

$$\begin{aligned}\mathbb{E}\big(\mathbb{FE}_{t,s}\ S(\varepsilon_j, y_m, 1)\big) &= \sum_{m=1}^{p}\sum_{r=s+1}^{t} a_m \xi_q(t,r)\ \mathbb{E}(\varepsilon_r y_{s+1-m}) + \sum_{j=1}^{q}\sum_{r=s+1}^{t} b_j \xi_q(t,r)\ \mathbb{E}(\varepsilon_r \varepsilon_{s+1-j}) \\ &\quad + c \sum_{r=s+1}^{t} \xi_q(t,r)\mathbb{E}(\varepsilon_r). \end{aligned} \tag{G.3}$$

Since $\mathbb{E}(\varepsilon_r) = 0$ for all r and $\mathbb{E}(\varepsilon_r \varepsilon_{s+1-j}) = 0$ for all $r \geq s+1$ and $1 \leq j \leq q$, it follows that the last two terms on the right-hand side of eq. (G.3) are zero. It remains to show that the first term in the right-hand side is also zero. It suffices to show that $\mathbb{E}(\varepsilon_r\ y_{s+1-m}) = 0$ for all $r \geq s+1$ and $1 \leq m \leq p$. Substituting $\mathbb{E}(y_{s+1-m}) + \sum_{j=-\infty}^{s+1-m} \xi_q(s+1-m, j)\varepsilon_j$ for y_{s+1-m} and using the linearity of the expectation operator, we have:

$$\begin{aligned}\mathbb{E}(\varepsilon_r\ y_{s+1-m}) &= \mathbb{E}\big(\varepsilon_r \mathbb{E}(y_{s+1-m})\big) + \mathbb{E}\big(\varepsilon_r \sum_{j=-\infty}^{s+1-m} \xi_q(s+1-m, j)\varepsilon_j\big) \\ &= \mathbb{E}(y_{s+1-m})\mathbb{E}(\varepsilon_r) + \sum_{j=-\infty}^{s+1-m} \xi_q(s+1-m, j)\mathbb{E}\big(\varepsilon_r \varepsilon_j\big) = \ 0. \end{aligned}$$

Accordingly, all the terms in the right-hand side of eq. (G.3) are zero, as claimed. ■

H Modelling Inflation

In this Section we present some supplementary results reported in Section 7.

H1 Unit Root Tests

In Table H.1 below, a number of common unit roots tests are reported: ADF (Augmented Dickey–Fuller), ERS, by Elliott *et al.* (1996), and MZ GLS, suggested by Perron and Ng (1996) and Ng and Perron (2001). As recommended by Ng and Perron (2001), the choice of the number of lags is based on the modified

Akaike information criterion (AIC).

Table H.1. Unit Root Tests.

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

The notations *,** indicate the statistical significance at 1% and 5%, respectively.

Table H.1 shows that, in general, we can reject the null hypothesis of a unit root in inflation series.

H2 Bai and Perron Methodology

For each l partition $(T_1, ..., T_l)$ the DAB-AR$(2; l)$ model can be estimated using the least-squared principle by minimizing the sum of the squared residuals where the minimization is taken over all partitions. Since the break points are discrete parameters and can only take a finite number of values they can be estimated by grid search using dynamic programming (see Bai and Perron, 2003, for more details).

Bai and Perron (2003) propose an F-type test for l versus $l+1$ breaks, which we will refer to as $\sup F_t(l+1|l)$. The testing procedure allows for a specific to general modelling strategy for the determination of the number of breaks in each series. The test is applied to each segment containing the T_{i-1} to T_i $(i = 1, ..., l+1)$. In particular, the procedure involves using a sequence of $(l+1)$ tests, where the conclusion of a rejection in favour of a model with $(l+1)$ breaks if the overall minimal value of the sum of squared residuals is sufficiently smaller than the sum of the squared residuals from the l break model. Note that the sum of the squared residuals is calculated over all segments where an additional break is included and compared with the residuals from the l model. Therefore, the break date selected is the one associated with the overall minimum.

Bai and Perron (1998) address the problem of the estimation of the break dates and present an efficient algorithm to obtain global minimizers of the sum of squared residuals based on the principle of dynamic programming, which requires at most least squares operations of order $\mathrm{O(T)}^2$ for any number of breaks. The limit distribution theory for inference about the break dates is considered in Bai and Perron (1998). Note that a valid alternative test to use for point detection changes is proposed by Horváth and Trapani (2021). The authors propose a family of CUSUM based statistics to detect the presence of change points in the deterministic part of the AR parameter in a random coefficient autoregressive sequence. An interesting feature of the test is that the inference procedure allows for heteroskedasticity of unknown form.

The results of the structural break test are summarized in Panel A of Table 1. The first column in Panel A compares the null hypothesis of l breaks against the alternative hypothesis of $l+1$ breaks, the second column reports the calculated value of the statistics and the third column the critical value of the test. Observing the calculated values of the test, it appears that the null hypothesis zero versus one break is rejected in favour of the alternative hypothesis. Similarly, the hypothesis of one break versus two breaks is rejected. However, the null hypothesis of two versus three breaks is not rejected, therefore we conclude that there are two structural breaks.[9]

H3 Forecasting

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

The evaluation of the out-of-sample forecast exercise does not rely on a single criterion; for robustness we compare the results of three different forecasting measures, namely, the root mean square forecast error (RMSE), the mean absolute error (MAE) and Theil's Inequality Coefficient (U Coefficient).

Table H.2 demonstrates the results of the forecasting exercise.[10]

Table H.2. Forecasting inflation in the United States: point predictive performances.

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

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

In columns 1 and 2 the forecasting horizon and the performance measure are shown, respectively. Columns 3-5 show the forecasting results. It follows directly from the results of Table H.2 that according to the RMSE and MAE criteria the DAB-AR (2;2) model performs better than its misspecified counterpart. According to these two performance measures Model 1 has forecasting properties in line with those

[9]Note that breaks in the variance are permitted provided that they occur on the same dates as the break in the AR parameters. Benati (2008) also used an AR model allowing for time-varying volatility. Cogley and Sargent (2005) also estimated a model in which the variance of innovations can vary over time.

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

obtained using the true model, Model 2. However, looking at the U coefficient measure the results are more mixed with Model 3 outperforming Model 1 in the short term and Model 1 having superior performance in the longer term.

Having investigated the out-of-sample forecasting performance of the DAB-AR-(2;2) model, in Section 7.2 we investigate whether inflation and its volatility are highly persistent.

I Structural Breaks

The results on structural breaks reported in Subsection 7 (in the main body of the paper) are discussed in the present Section.

I1 Solution Representation

Applying the results of Theorem 1, the following Corollary provides the solution representation (at time $t_1 + l$, $l \in \mathbb{Z}_0$) of eq. (32). But first, we define the following matrices.

Definition I.1 *Let the two tridiagonal matrices of order $r \in \mathbb{Z}_1$, denoted by $\mathbf{\Phi}_r^{(j)}$, $j = 1, 2$, be defined as:*

$$\mathbf{\Phi}_r^{(j)} = \begin{bmatrix} \phi_{1,j} & -1 & & & & \\ \phi_{2,j} & \phi_{1,j} & -1 & & & \\ & \phi_{2,j} & \phi_{1,j} & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \phi_{2,j} & \phi_{1,j} & -1 \\ & & & & \phi_{2,j} & \phi_{1,j} \end{bmatrix}.$$

Definition I.2 *i) The tridiagonal matrix (of order $l - r$) $\mathbf{\Phi}_{t_1+l,t_1+r}$, for $r = 1, \ldots, l-1$ and $l \geq 1$, is defined as:*

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

where its determinant is $\xi(t_1 + l, t_1 + r) = |\mathbf{\Phi}_{t_1+l,t_1+r}|$ with initial values $\xi(t_1 + l, t_1 + l) = 1$ and $\xi(t_1, t_1 + r) = 0$.

ii) The Hessenberg matrix $\mathbf{\Phi}_{t_1+l,t_1-r}$, for $r = 0, \ldots, t_1 - t_2$ and $r + l \geq 1$, is defined as:

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

where (for $r, l \neq 0$) $\bar{\mathbf{0}}$ is an $r \times l$ matrix of zeros except for -1 in its $r \times 1$ entry, and $\widetilde{\mathbf{0}}$ is an $l \times r$ matrix of zeros except for $\phi_{2,1}$ in its $1 \times r$ entry. Notice that $\mathbf{\Phi}_{t_1+l,t_1-r}$ is a block square matrix of order $l + r$.

Its determinant is $\xi(t_1+l, t_1-r) = |\mathbf{\Phi}_{t_1+l,t_1-r}|$ *with initial value* $\xi(t_1,t_1) = 1$.

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

Corollary I.1 *The explicit representation of* y_{t_1+l} *in eq.* (32) *in terms of the prescribed values* y_{t_2}, y_{t_2-1}, *is given by*

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

where $\varphi(r) = \begin{cases} \varphi_2 & \text{if } r \leq t_1 \\ \varphi_1 & \text{if } r > t_1. \end{cases}$

I2 Second Moment Structure

In this Section we will examine the second moment structure of the DAB-AR$(2;2)$ model. To obtain the time-varying variance of y_{t_1+l}, we will directly apply Corollary I.1.

First, let $1-\phi_{1,i}B-\phi_{2,i}B^2 = (1-\lambda_{1,i}B)(1-\lambda_{2,i}B)$, for $i = 1,2,3$.

Assumption I.1 *[Second-Order]*: $|\lambda_{m,i}| < 1$, $m = 1,2$ *and* $i = 1,3$.
The above condition implies that the DAB-AR$(2;2)$*-process is second-order.*

The following Proposition states expressions for the time-varying variance of y_{t_1+l} in eq. (I.1).

Proposition I.1 *Consider the model in eq.* (32). *Under Assumption* I.1*, the variance* $\mathbb{V}ar(y_{t_1+l})$ *is given by*

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

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

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

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

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

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

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

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

I3 Second-Order Persistence

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

In what follows, we introduce a time-varying second-order (or variance) persistence measure that is able to take into account the presence of breaks not only in the mean but in the variance as well. Fiorentini and Sentana (1998) argue that any reasonable measure of shock persistence should be based on the IRF's. For a univariate process $\{x_t\}$ with *i.i.d.* errors $\{e_t\}$, they define the persistence of a shock at a time point t as $P_2(x_t\,|e_t\,) \overset{\text{def}}{=} \dfrac{\mathbb{V}ar(x_t)}{\mathbb{V}ar(e_t)}$. Clearly $P_2(x_t\,|e_t\,)$ will take its minimum value of 1, whenever x_t is white noise, while it will not exist (will be infinite) for an I(1) process. It follows directly from eq. (I.2) that $P_2(y_{t_1+l}\,|\varepsilon_{t_1+l}\,) = \dfrac{\mathbb{V}ar(y_{t_1+l})}{\sigma_1^2}$, is given by

$$P_2(y_{t_1+l}\,|\varepsilon_{t_1+l}\,) = A_{t_1+l} + B_{t_1+l}\frac{\sigma_2^2}{\sigma_1^2} + C_{t_1+l}\frac{\sigma_3^2}{\sigma_1^2}. \tag{I.3}$$

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

Having derived explicit formulas for time-varying second-order (or variance) persistence measures, in Section 7, we show the empirical relevance of these results using U.S. inflation data.

J Time-Varying Persistence

In this Section we explain how we constructed the two pairs of Graphs in Section 7.2, starting with the first pair concerning the time-varying second-order persistence of US inflation: $P_2(\pi_t\,|\varepsilon_t\,) = \mathbb{V}ar(\pi_t)/\mathbb{V}ar(\varepsilon_t)$ and $\mathbb{V}ar(\pi_t)$.

1. First case, when $t \leq t_2$. The persistence is constant and it is given by

$$P_2(\pi_{t_2}\,|\varepsilon_{t_2}\,) = \frac{\mathbb{V}ar(\pi_{t_2})}{\sigma_3^2} = \frac{(1-\phi_{2,3})}{(1+\phi_{2,3})[(1-\phi_{2,3})^2 - \phi_{1,3}^2]},$$

where $\sigma_3 = 1.077,\ \phi_{1,3} = 0.470,\ \phi_{2,3} = 0.376.$

2. Second case, when $t_2 < t \leq t_1$: $t = t_2 + l$, $l = 1, \ldots, t_1 - t_2$, where $t_1 - t_2 = 39$. The persistence, for each l, is given by

$$P_2(\pi_{t_2+l}\,|\varepsilon_{t_2+l}) = \frac{\mathbb{V}ar(\pi_{t_2+l})}{\sigma_2^2} = A_{t_2+l} + C_{t_2+l}\frac{\sigma_3^2}{\sigma_2^2},$$

where $\sigma_2 = 2.300$, and

$$\begin{aligned} A_{t_2+l} &= \sum_{r=1}^{l} \xi^2(t_2+l, t_2+r), \\ C_{t_2+l} &= \frac{\left[(1-\phi_{2,3})\left(\xi^2(t_2+l,t_2) + \phi_{2,2}^2\xi^2(t_2+l,t_2+1)\right) + 2\phi_{1,3}\xi(t_2+l,t_2)\phi_{2,2}\xi(t_2+l,t_2+1)\right]}{(1+\phi_{2,3})[(1-\phi_{2,3})^2 - \phi_{1,3}^2]}, \end{aligned}$$

in which $\xi(t_2+l, t_2+r)$ is an $(l-r)$-th order tridiagonal determinant given by

$$\xi(t_2+l, t_2+r) = \begin{vmatrix} \phi_{1,2} & -1 & & & & \\ \phi_{2,2} & \phi_{1,2} & -1 & & & \\ & \phi_{2,2} & \phi_{1,2} & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \phi_{2,2} & \phi_{1,2} & -1 \\ & & & & \phi_{2,2} & \phi_{1,2} \end{vmatrix}, \tag{J.1}$$

with initial value $\xi(t_2+l, t_2+l) = 1$, where $\phi_{1,2} = 0.710, \phi_{2,2} = 0.127$.

A few values of A_{t_2+l} for $l = 1, 2, 3$ are given below:

$$\begin{aligned} A_{t_2+1} &= 1,\ A_{t_2+2} = \sum_{r=1}^{2} \xi^2(t_2+2, t_2+r) = 1 + \phi_{1,2}^2, \\ A_{t_2+3} &= \sum_{r=1}^{3} \xi^2(t_2+3, t_2+r) = 1 + \phi_{1,2}^2 + \begin{vmatrix} \phi_{1,2} & -1 \\ \phi_{2,2} & \phi_{1,2} \end{vmatrix}^2, \\ A_{t_2+4} &= \sum_{r=1}^{4} \xi^2(t_2+4, t_2+r) = 1 + \phi_{1,2}^2 + \begin{vmatrix} \phi_{1,2} & -1 \\ \phi_{2,2} & \phi_{1,2} \end{vmatrix}^2 + \begin{vmatrix} \phi_{1,2} & -1 & \\ \phi_{2,2} & \phi_{1,2} & -1 \\ & \phi_{2,2} & \phi_{1,2} \end{vmatrix}^2. \end{aligned}$$

3. Third case, when $t_1 < t$: $t = t_1 + l$, $l = 1, 2, \ldots$. The persistence, for each l, is given by

$$P_2(\pi_{t_1+l}\,|\varepsilon_{t_1+l}) = \frac{\mathbb{V}ar(\pi_{t_1+l})}{\sigma_1^2} = A_{t_1+l} + B_{t_1+l}\frac{\sigma_2^2}{\sigma_1^2} + C_{t_1+l}\frac{\sigma_3^2}{\sigma_1^2},$$

(see eq. (I.3)) where $\sigma_1 = 2.160$ and i)

$$A_{t_1+l} = \sum_{r=1}^{l} \xi^2(t_1+l, t_1+r),$$

in which $\xi(t_1+l, t_1+r)$ is an $(l-r)$-th order tridiagonal determinant given by

$$\xi(t_1+l, t_1+r) = \begin{vmatrix} \phi_{1,1} & -1 & & & & \\ \phi_{2,1} & \phi_{1,1} & -1 & & & \\ & \phi_{2,1} & \phi_{1,1} & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \phi_{2,1} & \phi_{1,1} & -1 \\ & & & & \phi_{2,1} & \phi_{1,1} \end{vmatrix}, \tag{J.2}$$

with initial value $\xi(t_1+l, t_1+l) = 1$, where $\phi_{1,1} = 0.247, \phi_{2,1} = -0.314$.

As an example for $l = 4$ the value of A_{t_1+4} is given by:

$$A_{t_1+4} = \sum_{r=1}^{4} \xi^2(t_1+4, t_1+r) = 1 + \phi_{1,1}^2 + \begin{vmatrix} \phi_{1,1} & -1 \\ \phi_{2,1} & \phi_{1,1} \end{vmatrix}^2 + \begin{vmatrix} \phi_{1,1} & -1 & \\ \phi_{2,1} & \phi_{1,1} & -1 \\ & \phi_{2,1} & \phi_{1,1} \end{vmatrix}^2.$$

ii)

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

where $\xi(t_1+l, t_1-r)$ is an $(r+l)$-th order Hessenbergian of the form

$$\xi(t_1+l, t_1-r) = \begin{vmatrix} \phi_{1,2} & -1 & & & & & & \\ \phi_{2,2} & \phi_{1,2} & -1 & & & & & \\ & \ddots & \ddots & & & & & \\ & & \phi_{2,2} & \phi_{1,2} & -1 & & & \\ & & & \phi_{2,1} & \phi_{1,1} & -1 & & \\ & & & & \ddots & \ddots & \ddots & \\ & & & & & \phi_{2,1} & \phi_{1,1} & -1 \\ & & & & & & \phi_{2,1} & \phi_{1,1} \end{vmatrix}, \tag{J.3}$$

in which the last l rows and columns form a tridiagonal submatrix whose nonzero elements are $\phi_{2,1}$, $\phi_{1,1}$ and (-1); the first r rows and columns form a tridiagonal submatrix whose nonzero elements are $\phi_{2,2}$, $\phi_{1,2}$ and (-1); the first r rows and the last l columns form a submatrix of zeros except for the (-1) in its

$r \times 1$ entry; the last l rows and the first r columns form a submatrix of zeros except for $\phi_{2,1}$ in its $1 \times r$ entry.

As an example for $l = 2$ the value of B_{t_1+2} is given by

$$B_{t_1+2} = \underbrace{\left|\begin{matrix} \phi_{1,1} & -1 \\ \phi_{2,1} & \phi_{1,1} \end{matrix}\right|^2}_{r=0} + \underbrace{\left|\begin{matrix} \phi_{1,2} & -1 & \\ \phi_{2,1} & \phi_{1,1} & -1 \\ & \phi_{2,1} & \phi_{1,1} \end{matrix}\right|^2}_{r=1} + \underbrace{\left|\begin{matrix} \phi_{1,2} & -1 & & \\ \phi_{2,2} & \phi_{1,2} & -1 & \\ & \phi_{2,1} & \phi_{1,1} & -1 \\ & & \phi_{2,1} & \phi_{1,1} \end{matrix}\right|^2}_{r=2} +$$

$$+ \cdots + \underbrace{\left|\begin{matrix} \phi_{1,2} & -1 & & & & & & \\ \phi_{2,2} & \phi_{1,2} & -1 & & & & & \\ & \ddots & \ddots & \ddots & & & & \\ & & \phi_{2,2} & \phi_{1,2} & -1 & & & \\ & & & \phi_{2,2} & \phi_{1,2} & -1 & & \\ & & & & \phi_{2,1} & \phi_{1,1} & -1 \\ & & & & & \phi_{2,1} & \phi_{1,1} \end{matrix}\right|^2}_{r=t_1-t_2-1=38},$$

where the last Hessenbergian in the above summation is of order $2 + 38 = 40$.

iii)

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

where $\xi(t_1+l,t_2)$ and $\xi(t_1+l,t_2+1)$ are as in case (ii), derived by $\xi(t_1+l,t_1-r)$ for $r = t_1 - t_2$ and $r = t_1 - t_2 - 1$, respectively.

Finally, we explain how we have constructed the second pair of graphs, concerning the time-varying first-order persistence of US inflation, as measured by $\mathbb{E}(\pi_t)$ and $P_1(\pi_t\,|\varepsilon_t) = \mathbb{E}(\pi_t)/\varphi(t)$, where

$$\varphi(t) = \begin{cases} \varphi_3 & \text{if} \quad t \le t_2 \\ \varphi_2 & \text{if} \quad t_2 < t \le t_1 \\ \varphi_1 & \text{if} \quad t_1 < t. \end{cases}$$

Case I: $t \le t_2$. The persistence is constant and it is given by

$$\mathbb{E}(\pi_t) = \frac{\varphi_3}{1 - \phi_{1,3} - \phi_{2,3}},$$

where

$$\varphi_3 = 0.496, \; \phi_{1,3} = 0.470, \; \phi_{2,3} = 0.376.$$

Case II: $t_2 < t \leq t_1$. That is $t = t_2 + l$, $l = 1, \ldots, t_1 - t_2$ and $t_1 - t_2 = 39$. The persistence, for each l, is given by

$$E(\pi_t) = A_{t_2+l}\varphi_2 + C_{t_2+l}\varphi_3,$$

where $\varphi_2 = 3.637$, and

$$\begin{aligned} A_{t_2+l} &= \sum_{r=1}^{l} \xi(t_2 + l, t_2 + r), \\ C_{t_2+l} &= \xi(t_2 + l, t_2) + \phi_{2,2}\xi(t_2 + l, t_2 + 1), \end{aligned}$$

in which $\xi(t_2 + l, t_2 + r)$ has been defined in eq. (J.1).

Case III: $t_1 < t$. That is $t = t_1 + l$, $l = 1, 2, ...$ The persistence, for each l, is given by

$$E(\pi_t) = A_{t_1+l}\varphi_1 + B_{t_1+l}\varphi_2 + C_{t_1+l}\varphi_3,$$

where $\varphi_1 = 2.859$ and

$$\begin{aligned} A_{t_1+l} &= \sum_{r=1}^{l} \xi(t_1 + l, t_1 + r), \\ B_{t_1+l} &= \sum_{r=0}^{t_1-t_2-1} \xi(t_1 + l, t_1 - r), \\ C_{t_1+l} &= \xi(t_1 + l, t_2) + \phi_{2,1}\xi(t_1 + l, t_2 + 1), \end{aligned}$$

in which $\xi(t_1 + l, t_1 + r)$ and $\xi(t_1 + l, t_1 - r)$ are given by eqs. (J.2) and (J.3), respectively.

REFERENCES

Agarwal, R.P. (2000) *Difference Equations and Inequalities: Theory, Methods, and Applications.* Marcel Dekker.

Azrak, R. and 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. and Perron, P. (1998) Estimating and testing linear models with multiple structural changes. *Econometrica*, 66, pp. 47 – 78.

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

Benati, L. (2008) Investigating inflation persistence across monetary regimes. *The Quarterly Journal of Economics* 123, 1005 – 1060.

Berkes, I., Hörmann, S. and Schauer, J. (2011) Split invariance principles for stationary processes. *Annals of Probability* 39, 2441 – 2473.

Brockwell, P.J. and Davis, R.A. (1991) *Time Series Theory and Methods.* Springer.

Brockwell, P.J. and Davis, R.A. (2016) *An Introduction to Time Series and Forecasting.* Springer.

Cogley, T. and Sargent, T.J. (2005) The conquest of US inflation: Learning and robustness to model uncertainty. *Review of Economic Dynamics* 8, 528 – 563.

Dahlhaus, R. (1996) Maximum likelihood estimation and model selection for locally stationary processes. *Journal of Nonparametric Statistics* 6, 171 – 191.

Elliott, G., Rothenberg, T.J. and Stock, J.H. (1996) Efficient tests for an autoregressive unit root. *Econometrica* 75, 813 – 836.

Fiorentini, G. and Sentana, E. (1998) Conditional means of time series processes and time series processes for conditional means. *International Economic Review* 39, 1101 – 1018.

Fulkerson, D.R. (1951) Quasi-Hermite forms of row-finite matrices. Ph.D. Thesis. University of Wisconsin, URL: http://www.researchgate.net/publication/36218809.

Hallin, M. (1979) Mixed autoregressive-moving average multivariate processes with time-dependent coefficients. *Journal of Multivariate Analysis* 8, 567 – 572.

Hörmann, S. (2009). Berry-Esseen bounds for econometric time series. *ALEA Latin American Journal of Probability and Mathematical Statistics* 6, 377 – 397.

Horváth, L. and Trapani, L. (2021) Changepoint detection in random coefficient autoregressive models. Preprint, arXiv:2104.13440.

Karanasos, M. (2001) Prediction in ARMA models with GARCH in mean effects. *Journal of Time Series Analysis* 22, 555 – 576.

Massacci, D. and Trapani, L. (2022) High Dimensional Threshold Regression with Common Stochastic Trends. SSRN: https://ssrn.com/abstract=4133488 or http://dx.doi.org/10.2139/ssrn.4133488.

Miller, K.S. (1968) *Linear Difference Equations.* Benjamin.

Neimi, H. (1983) On the effects of a non-stationary noise on ARMA models. *Scandinavian Journal of Statistics* 10, 11 – 17.

Ng, S. and Perron, P. (2001) Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519 – 1554.

Paraskevopoulos, A.G. (2004) A recursive approach to the solution of abstract linear equations and the tau method. *Comput. Math. with Appl.*, 47, 1753 – 1774.

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. and 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.

Perron, P. and Ng, S. (1996) Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. *The Review of Economic Studies* 63, 435 – 463.

Pesaran, M.H. and Timmermann, A. (2005) Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics* 129, 183 – 217.

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

Singh, V.N. (1980) Solution of a general homogeneous linear difference equation. *Journal of the Australian Mathematical Society.* 22, 53 – 57.

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

Teräsvirta, T. (1994) Specification, estimation and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* 89, 208 – 218.

Thorson, J. (1979) Gaussian elimination on a banded matrix. Stanford Exploration Project, Stanford University.

Toeplitz, O. (1909) Über die Auflösung unendich vieler linearer Gleichungen mit unendlich vielen unbekannten. *Rend. Circ. Mat. Palermo* 28, 88 – 96.

Wu, W.B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America* 102, 14150 – 14154.