Mixed causal–noncausal autoregressions with exogenous regressors

Hecq, Alain; Issler, Joao Victor; Telg, Sean

published in
Journal of Applied Econometrics
2020

DOI (link to publisher)
10.1002/jae.2751

document version
Publisher's PDF, also known as Version of record

Received: 12 September 2017 | Revised: 8 December 2019

DOI: 10.1002/jae.2751

RESEARCH ARTICLE

WILEY Journal of APPLIED ECONOMETRICS

Mixed causal–noncausal autoregressions with exogenous regressors

Alain Hecq[1] | **Joao Victor Issler**[2] | **Sean Telg**[3]

[1]Department of Quantitative Economics, School of Business and Economics, Maastricht University, Maastricht, The Netherlands

[2]Graduate School of Economics—EPGE, Getulio Vargas Foundation, Rio de Janeiro, Brazil

[3]Department of Econometrics and Operations Research, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Correspondence
Sean Telg, Department of Econometrics and Operations Research, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands.
Email: j.m.a.telg@vu.nl

Summary

Mixed causal–noncausal autoregressive (MAR) models have been proposed to model time series exhibiting nonlinear dynamics. Possible exogenous regressors are typically substituted into the error term to maintain the MAR structure of the dependent variable. We introduce a representation including these covariates called MARX to study their direct impact. The asymptotic distribution of the MARX parameters is derived for a class of non-Gaussian densities. For a Student t likelihood, closed-form standard errors are provided. By simulations, we evaluate the MARX model selection procedure using information criteria. We examine the influence of the exchange rate and industrial production index on commodity prices.

1 | INTRODUCTION

The usefulness of mixed causal–noncausal autoregressive (MAR) models—that is, time series specifications with both lag and lead components—can be explained by at least three findings. First, Gouriéroux and Zakoïan (2016), Fries and Zakoian (2019), and Hecq, Lieb, and Telg (2016) demonstrate how noncausal autoregressive terms can generate dynamics such as speculative bubbles and asymmetric cycles that instead could only be generated using complex nonlinear models. Second, Lanne, Luoto, and Saikkonen (2012) and Lanne, Nyberg, and Saarinen, (2012) show that noncausal models might improve forecast performances. Third, the MAR representation of an economic variable can be seen as solutions of rational expectation models, which are prevalent in macroeconomics (Lanne & Saikkonen, 2011). The MAR model explicitly allows for nonfundamental outcomes, which proves extremely useful, as some economic models do not have a fundamental solution by construction (Alessi, Barigozzi, & Capasso, 2011).

Our motivation for introducing the MARX model, which introduces covariates in the MAR, is to further increase the relevance of the MAR specification for modeling and forecasting economic processes. First, the MARX model offers more flexibility, as the inclusion of exogenous components with potentially marginal leptokurtic distributions enriches the nonlinear patterns. The model can be seen as a generalization of not only the autoregressive model with exogenous regressors (ARX), but also the class of autoregressive distributed lag (ARDL) models and marginal equations within a class of vector autoregressive (VAR) models. In terms of interpretation, practitioners are able to study directly the impact of exogenous forcing variables on the variable of interest. In the existing literature (see, e.g., Lanne & Luoto, 2013), exogenous regressors are included in the error term on which an MAR structure is assumed. This might be beneficial in modeling structural equations, but it hides the direct link between key economic variables. Second, the MARX model allows us to investigate the claim that when controlling for additional variables the lead part might become insignificant. It has been argued that the noncausal component acts as a proxy for the omission of important variables that are not included in the model (Lof, 2013; Lof & Nyberg, 2017). Since noncausal and causal models with covariates are contained in the class of MARX models,

we can study conditions under which the practitioner favors one specification over the other. Third, MARX processes have the feature that, contrary to MAR processes, one can discriminate between causal, noncausal and mixed specifications based on second-order properties. The presence of cross-covariances leads to asymmetries in the autocovariance functions of the different models, which allows for a confirmatory analysis. We study commodity prices that are expected to (i) show signs of noncausality (see, e.g., Karapanagiotidis, 2014; Lof & Nyberg, 2017) due to their forward-looking nature and (ii) depend on an indicator of economic activity (Issler, Rodrigues, & Burjack, 2014) and the US dollar exchange rate.

The paper is organized as follows. In Section 2 we introduce the MARX model, propose a model selection and forecasting procedure, and show that MARX models have the appealing feature to be (potentially) identifiable by second-order properties. Section 3 shows how MARX parameters can be estimated by maximum likelihood (ML) for non-Gaussian densities as defined by Andrews, Breidt, and Davis (2006). For the case of the Student t likelihood, we provide a method to compute closed-form solutions of the corresponding standard errors. A forecasting procedure is presented. By means of Monte Carlo simulations, we evaluate in Section 4 our model selection method, the identifiability of MARX processes, and the forecasting performance of MARX models. Section 5 consists of the empirical application in which we highlight the main features of the MARX model and their usefulness for practitioners. Section 6 summarizes and concludes.[1]

2 | THE MARX MODEL

Let y_t be the variable of interest, which is observed over the time period $t = 1, \dots, T$. Let $x_{i,t}$ $(i = 1, \dots, q)$ be the ith variable in a set of q explanatory regressors and $\boldsymbol{\beta} \in \mathbb{R}^q$ a vector of parameters. Then we can define $\boldsymbol{X}_t = [x_t, \dots, x_{q,t}]' \in \mathbb{R}^q$ as the vector of all exogenous variables at time t.[2] The MARX(r, s, q) for a stationary time series y_t can be represented as

$$\phi(L)\varphi\left(L^{-1}\right) y_t - \boldsymbol{\beta}'\boldsymbol{X}_t = \varepsilon_t, \tag{1}$$

where $\phi(L)$ is a lag polynomial of order r, $\varphi(L^{-1})$ a lead polynomial of order s and we denote $r + s = p$. The operator L is the lag operator when raised to positive powers; that is, $L^i y_t = y_{t-i}$, and interpreted as a lead operator when raised to negative powers: $L^{-i} y_t = y_{t+i}$. All roots of $\phi(z)$ and $\varphi(z^{-1})$ lie outside the unit circle.[3] The sequence ε_t is assumed to be i.i.d. non-Gaussian. In order to ensure identifiability of the parameter vector $\boldsymbol{\beta}$ and to prove consistency of the maximum likelihood (ML) estimator, we make the following assumptions on $\boldsymbol{X}_t$.

Assumption 1. The processes in $\boldsymbol{X}_t$ are ergodic, (strictly) stationary, and strictly exogenous w.r.t. ε_t; that is, $\mathbb{E}\left(\varepsilon_t | x_{i,t-j}\right) = 0$ for all $i = 1, \dots, q$ and all $j \in \mathbb{Z}$.

Assumption 2. The processes in $\boldsymbol{X}_t$ are of the form $x_{i,t} = c_i + \sum_{j=-\infty}^{\infty} \rho_{i,j}\eta_{i,t-j}$ with $\eta_{i,t} \overset{\text{i.i.d.}}{\sim} \left(0, \sigma_{\eta_i}^2\right)$ for all $i = 1, \dots, q$.

Assumption 3. The processes in $\boldsymbol{X}_t$ are linearly independent.

Assumption 1 is necessary for the central limit theorem for m-dependent processes (theorem 6.4.2 in Brockwell & Davis, 1991). Whereas this zero conditional mean assumption restricts the economic relationships to be studied, we consider it as a natural extension to the exogeneity assumption imposed on the well-known ARX model. Assumption 2 defines the dynamic structure the processes in $\boldsymbol{X}_t$ can assume. By allowing for a two-sided moving average process, purely causal and noncausal as well as mixed ARMA processes are included. By Assumption 3 the matrix of explanatory variables has full column rank. The concepts of causality and noncausality are defined in terms of the strictly stationary solution of the model. When $q = 0$, the process in Equation (1) reduces to an MAR process that has a two-sided moving average representation consisting of past, current, and future values of ε_t under appropriate moment conditions for ε_t and summability conditions for the sequence of moving average coefficients (Gouriéroux & Zakoïan, 2016). When $q > 0$, the process no

[1]The Appendix discusses the properties of the approximate ML estimator. Proofs and additional material are collected in the Supplementary Material (Supporting Information). Methods proposed in this paper are implemented in the R package MARX (https://CRAN.R-project.org/package=MARX), which is discussed in detail by Hecq, Lieb, and Telg (2017).

[2]For now, we only consider contemporaneous values of $\boldsymbol{X}_t$. However, the MARX model can also take the form of a mixed autoregressive distributed lag (MARDL) model.

[3]We replace the operator L by the complex variable z when considering the properties of polynomials.

longer has a strictly stationary solution solely in terms of ε_t, but involves both $\boldsymbol{X}_t$ and ε_t. That is,

$$y_t = \pi\left(L, L^{-1}\right)\left[\varepsilon_t + \boldsymbol{\beta}'\boldsymbol{X}_t\right] = \sum_{j=-\infty}^{\infty} \pi_j z_{t-j}, \tag{2}$$

where $z_{t-j} = \varepsilon_{t-j} + \sum_{i=1}^{q} \beta_i x_{i,t-j}$ and $\pi(z, z^{-1})$ is a polynomial satisfying $\pi\left(z, z^{-1}\right)\phi(z)\varphi\left(z^{-1}\right) = 1$. The process y_t exists almost surely under rather weak conditions, such as $\sum_{j=-\infty}^{\infty} |\pi_j|^{\delta} < \infty$ and $\mathbb{E}|z_t|^{\delta} < \infty$ for $\delta > 0$, where $\phi(z)$ and $\varphi\left(z^{-1}\right)$ are invertible (Gouriéroux & Jasiak, 2016). We observe that y_t has a two-sided moving average representation augmented with a second part involving linear combinations of past, current, and future values of $\boldsymbol{X}_t$. This latter part can be interpreted as the sum of q processes $\beta_i x_{i,t}$ that are passed through a two-sided linear filter with coefficients resulting from inverting the product $\left[\phi(z)\varphi\left(z^{-1}\right)\right]$ to $\pi\left(z, z^{-1}\right)$.[4]

To make a comparison of the MARX and MAR model, we have to be more explicit regarding the role of $\boldsymbol{X}_t$ in Equation (1). Following Assumption 2, the covariates have a representation as a two-sided moving average. Hence let us assume z_t can be written as a mixed causal–noncausal process $\rho(L)\gamma\left(L^{-1}\right) z_t = \varepsilon_t^*$, where ε_t^* is an i.i.d. error term. After inverting the polynomials, we can substitute z_t for $\boldsymbol{\beta}'\boldsymbol{X}_t + \varepsilon_t$ in Equation (1) to obtain $\phi(L)\varphi\left(L^{-1}\right) y_t = \left[\rho(L)\gamma\left(L^{-1}\right)\right]^{-1}\varepsilon_t^*$, which yields

$$\phi^*(L)\varphi^*\left(L^{-1}\right) y_t = \varepsilon_t^*,$$

with $\phi^*(L) \equiv \phi(L)\rho(L)$ and $\varphi^*\left(L^{-1}\right) \equiv \varphi\left(L^{-1}\right)\gamma\left(L^{-1}\right)$. Hence, if we want to represent the MARX by an MAR model, the dynamics of z_t, which are determined by the fluctuations in $\boldsymbol{X}_t$, might increase the degree of the causal and noncausal polynomial of the MAR. In other words, given the dynamics of z_t, we have that $\deg[\phi^*(z)] \geq \deg[\phi(z)]$ and $\deg[\varphi^*(z)] \geq \deg[\varphi(z)]$.

Lemma 1. *From Equation (1), we can construct unobserved noncausal and causal components (u, v) similar to Lanne and Saikkonen (2011) and Gouriéroux and Jasiak (2016) and obtain*

$$u_t \equiv \phi(L)y_t \leftrightarrow \varphi\left(L^{-1}\right) u_t - \boldsymbol{\beta}'\boldsymbol{X}_t = \varepsilon_t, \tag{3}$$

$$v_t \equiv \varphi\left(L^{-1}\right) y_t \leftrightarrow \phi(L)v_t - \boldsymbol{\beta}'\boldsymbol{X}_t = \varepsilon_t. \tag{4}$$

Defining $\left[\varphi\left(z^{-1}\right)\right]^{-1} \equiv \delta\left(z^{-1}\right)$ and $[\phi(z)]^{-1} \equiv \alpha(z)$, the second equalities in Equation (3)- (4) can be written as $u_t = \sum_{j=0}^{\infty}\delta_j z_{t+j}$ and $v_t = \sum_{j=0}^{\infty}\alpha_j z_{t-j}$, which are called unobserved noncausal and causal components respectively. The first equations show that (u, v) can be interpreted as a noncausal (resp. causal) "error term" of a purely causal (resp. noncausal) autoregression. The relations in Lemma 1 can be used to simulate MARX processes in two steps.[5]

2.1 | Interpretation of the MARX model

The MARX model allows practitioners to study the direct impact of exogenous regressors on the variable of interest. We argue that there are more reasons to study the MARX model. In particular, consider the autoregressive distributed lag (ARDL) model of the form[6]

$$y_t = \sum_{i=1}^{p} a_i y_{t-i} + \sum_{i=0}^{k} \boldsymbol{b}_i'\boldsymbol{X}_{t-i} + \epsilon_t. \tag{5}$$

We show that the MARX model allows for different representations of Equation (5), which can exhibit richer dynamics. Define $a(L) \equiv \left(1 - \sum_{i=1}^{p} a_i L^i\right)$ and suppose, similarly to Breidt, Davis, Lii, and Rosenblatt (1991), that the pth-order polynomial $a(z)$ has r well-behaved roots (i.e., outside the unit circle) and s ill-behaved roots (i.e., inside the unit circle) with $p = r + s$. Following Lanne and Saikkonen (2011), write $a(L) = \phi(L)\varphi^*(L)$ with $\phi(z)$ and $\varphi^*(z)$ being polynomials of orders r and s respectively. The polynomial $\phi(z)$ has the well-behaved roots, whereas $\varphi^*(z)$ possesses explosive roots

[4]The effects of two-sided linear filters, with a focus on seasonal adjustment, on the identification of mixed causal–noncausal models was studied by Hecq, Telg, and Lieb (2017).
[5]More detailed information on how to simulate MARX processes can be found in the Supplementary Material.
[6]For notational convenience, we fix the lag order k to be the same for all variables contained in $\boldsymbol{X}_t$.

in direct time. Considering $\varphi^*(z)$ in reverse time, all roots are inverted and thus outside the unit circle. Straightforward algebra gives $\varphi^*(z) = -\varphi_s^* z^s \varphi(z^{-1})$, which yields

$$\phi(L)\varphi(L^{-1})y_t = \left[-\frac{1}{\varphi_s^*}L^{-s}\right]\left(\sum_{i=0}^{k}\boldsymbol{b}_i'\boldsymbol{X}_{t-i} + \epsilon_t\right).$$

Denoting $\varepsilon_t = \left(-\frac{1}{\varphi_s^*}\epsilon_{t+s}\right)$ and $\boldsymbol{\beta}_i' = \left(-\frac{1}{\varphi_s^*}\boldsymbol{b}_i'\right)$, we obtain

$$\phi(L)\varphi(L^{-1})y_t = \sum_{i=0}^{k}\boldsymbol{\beta}_i'\boldsymbol{X}_{t-i+s} + \varepsilon_t, \tag{6}$$

which is similar to Equation (1), but explicitly allows for the inclusion of lags and leads of the exogenous regressors. As explained in footnote 2, the MARX model is allowed to include such terms and can be labeled as MARDL. Lanne and Saikkonen (2011) discuss the condition ($\varphi_s \neq 0$) for which a one-to-one correspondence between Equations (5) and (6) exists for the MAR model. The latter specification has the same advantages in the MARX case: It is easier to set up conventional likelihood tests to determine r and s, and the model parameters are orthogonal to the distributional parameters, which implies asymptotic independence of the corresponding ML estimates.

Example 1. Consider the process $y_t = a_1 y_{t-1} + a_2 y_{t-2} + b_1 x_{t-1} + \epsilon_t$, which is nested in Equation (5). Following the algebra above, this model can equivalently be written as an MARX model of the form

$$(1 - \phi_1 L)\left(1 - \varphi_1 L^{-1}\right) y_t = \beta_1 x_t + \varepsilon_t. \tag{7}$$

The MARX broadens the set of models in Equation (5) with counterparts that include a noncausal component. This encompasses models that consider Granger causality from x_t to y_t (as in Example 1) and single equations within a VAR.

Example 2. The VAR(2) model is given by

$$\begin{aligned} y_t &= a_{11}y_{t-1} + a_{12}x_{t-1} + b_{11}y_{t-2} + b_{12}x_{t-2} + \epsilon_{y,t}, \\ x_t &= a_{21}y_{t-1} + a_{22}x_{t-1} + b_{21}y_{t-2} + b_{22}x_{t-2} + \epsilon_{x,t}. \end{aligned}$$

The exogeneity assumption (Assumption 1) implies that x_t is not determined by y_t at any displacement in time. Hence the MARX model is compatible with a situation in which $a_{21} = b_{21} = 0$ and $\text{corr}(\epsilon_{y,t}, \epsilon_{x,t}) = 0$. Under these restrictions, the MARX represents the first equation of the VAR. Whenever the exogeneity assumption cannot be credibly made, a better modeling approach might be to compare causal, mixed causal–noncausal and purely noncausal VAR models.

Another observation can be made from Example 1. To compare MARX models that are theoretically equivalent, the time subscript of the exogenous regressor has to be adjusted by the degree of noncausality. That is, the ARDL considers x_{t-1}, whereas the MARX model contains x_t as the noncausal order is increased by one. This consideration might be important in cases of structural modeling, as it is crucial to identify the "correct" reduced form. For prediction, the main point of interest is to select a model that fits the data best. In Section 5 we observe that no time shift of the exogenous regressors typically results in higher log-likelihood values.

Lastly, we show that the MARX is compatible with the rational-expectation models studied in Lanne and Luoto (2013), with a caveat. After some algebra, Equation (7) can be rewritten as

$$y_t = \frac{\phi_1}{1 + \phi_1\varphi_1}y_{t-1} + \frac{\varphi_1}{1 + \phi_1\varphi_1}y_{t+1} + \frac{\beta_1}{1 + \phi_1\varphi_1}x_t + \varepsilon_t. \tag{8}$$

Calling these coefficients γ_f, γ_b, λ and replacing y_{t+1} by $\mathbb{E}(y_{t+1})$ plus a martingale difference, we obtain an equation similar to Lanne and Luoto (2013; Equation (4)) expressed in regression form. The caveat alluded to above is that an MARX model must obey the restriction imposed by Assumption 1. Thus, unless we already start with a strictly exogenous x_t in our analysis, an MARX representation is not obtained in the end. There are plausible examples covered by the rational-expectation literature where Assumption 1 may fail. However, this is not true in general, and whether one obtains

an MARX model is essentially an empirical question, guided by economic theory. In Section 5 we present empirical applications where Assumption 1 is credibly matched.

2.2 | Model selection

The findings in the previous section provide important considerations to perform MARX model selection. We propose a model selection procedure and illustrate it by an example afterwards.

- *Step 1: Determining total autoregressive order.* In a first step, we identify the so-called pseudo-causal model (see, e.g., Hecq et al., 2016) by estimating ARDL models as in Equation (5) by ordinary least squares (OLS) and identify by information criteria such as Akaike (AIC), Bayesian (BIC), and Hannan–Quinn (HQ) the autoregressive order p that makes the residuals free of serial correlation.[7] Additionally, we find the lowest lag orders of the exogenous regressors to consider in the estimation of all models. No leads can be selected in the ARDL, as it violates the exogeneity condition in Assumption 1.
- *Step 2: From causal to noncausal.* From the identified ARDL model, we estimate the theoretically equivalent mixed and noncausal models with $p = r + s$ where the time subscript of the exogenous regressors is adjusted by the degree of noncausality. The proposed estimation method is the non-Gaussian approximate maximum likelihood (AML) estimator, which is discussed in Section 3. As the exogeneity condition is respected in Step 1, it also holds here and controls the maximum allowed time shift in the noncausal model.
- *Step 3: From noncausal to causal.* We estimate a purely noncausal model of order p with contemporaneous exogenous regressors together with all theoretically equivalent MARX models with $p = r + s$ by non-Gaussian AML.
- *Step 4: Remaining models.* The models in steps 2 and 3 are MARX models that consider the exogenous regressors at different points in time. In this step, we propose to estimate models that consider the exogenous regressors at time periods between those found at steps 2 and 3 by non-Gaussian AML.
- *Step 5: Highest log-likelihood.* The model that attains the highest value for the log-likelihood at its estimated parameters is chosen to be the final model. The set of considered models are those estimated in steps 2, 3, and 4, as well as the original ARDL, which is reestimated by AML.

The MARX model selection procedure is more complicated compared to the MAR. The first step for the MAR consists of estimating causal AR models by OLS, which is important to consistently estimate the lag order that makes the residuals free of serial correlation. In the MARX case, the significance of the exogenous regressors might change depending on the type of model. For example, the regressors might be significant in causal models, while they are not so in the noncausal model. To deal with such situations, steps 2 and 3 consider the causal and noncausal model as separate starting points. Step 4 examines remaining models for completeness. Thus these extra steps take into account that exogenous regressors create asymmetries based on the direction of the time line considered. The procedure is best characterized by an example.

Example 3. Suppose we have one exogenous regressor x_t and identify the following ARDL model in the first step: y_t on y_{t-1}, y_{t-2}, x_t. Table 1 displays the proposed model selection.

Remark 1. By simulations, we find that if the coefficients of the exogenous regressors are zero in the true data-generating process (DGP), then the OLS estimates of these coefficients are insignificant. We justify estimating ARDL by OLS in the first step, as it consistently estimates the lag order p, while the non-Gaussian ML estimator might underestimate the autoregressive order in the presence of heavy-tailed distributions (Gouriéroux & Jasiak, 2018).

2.3 | Identifiability

It is theoretically possible to distinguish between different specifications based on second-order properties. Contrary to the MAR model, the presence of cross-covariances in the MARX models allows us to achieve identification without distributional assumptions.

Example 4. Suppose the true DGP is the MARX(1, 0, 1) given by $y_t = \phi_1 y_{t-1} + \beta_C x_t + \varepsilon_t$. The corresponding noncausal model is the MARX(0, 1, 1) represented as $y_t = \varphi_1 y_{t+1} + \beta_{\text{NC}} x_{t+1} + \epsilon_t$. In the case of the MARX(1, 0, 1) model, the OLS

[7]It is advised to test whether additional lags are needed to remove autocorrelation. The null of normality should be tested on the residuals of the pseudo-causal model to justify the non-Gaussian ML estimator and to look for signs of noncausality. The model selection procedure for MAR models can be found in Hecq et al. (2016).

TABLE 1 Model selection example

Step 1	Step 2	Step 3
Selected ARDL model:	Estimate by non-Gaussian AML	Noncausal starting point:
C: y_t on y_{t-1}, y_{t-2}, x_t	theoretically equivalent models:	NC: y_t on y_{t+1}, y_{t+2}, x_t
	M: y_t on $y_{t-1}, y_{t+1}, x_{t+1}$	Theoretically equivalent models:
	NC: y_t on $y_{t+1}, y_{t+2}, x_{t+2}$	M: y_t on $y_{t-1}, y_{t+1}, x_{t-1}$
		C: y_t on $y_{t-1}, y_{t-2}, x_{t-2}$
Step 4	Step 5	
Remaining models:	Reestimate ARDL of Step 1 by	
C: y_t on $y_{t-1}, y_{t-2}, x_{t-1}$	non-Gaussian AML and compare	→ Select model with
M: y_t on y_{t-1}, y_{t+1}, x_t	log-likelihood of models in all	highest log-likelihood
NC: y_t on $y_{t+1}, y_{t+2}, x_{t+1}$	steps.	

Note. C, M, and NC stand for causal, mixed, and noncausal respectively.

estimator fulfills

$$\hat{\phi}_{1,T} = \frac{\sum_{t=r+1}^{T-s} y_t y_{t-1}}{\sum_{t=r+1}^{T-s} y_{t-1}^2} - \hat{\beta}_{C,T} \frac{\sum_{t=r+1}^{T-s} y_{t-1} x_t}{\sum_{t=r+1}^{T-s} y_{t-1}^2} \xrightarrow{p} \frac{\gamma_y(1)}{\gamma_y(0)} - \beta_C \frac{\gamma_{xy}(1)}{\gamma_y(0)},$$

$$\hat{\beta}_{C,T} = \frac{\sum_{t=r+1}^{T-s} x_t y_t}{\sum_{t=r+1}^{T-s} x_t^2} - \hat{\phi}_{1,T} \frac{\sum_{T=r+1}^{T-s} y_{t-1} x_t}{\sum_{t=r+1}^{T-s} x_t^2} \xrightarrow{p} \frac{\gamma_{xy}(0)}{\gamma_x(0)} - \phi_1 \frac{\gamma_{xy}(1)}{\gamma_x(0)},$$

where $\gamma_x(k)$ and $\gamma_y(k)$ denote the autocovariance function of x_t and y_t at order k and $\gamma_{xy}(k)$ the cross-covariance between x_t and y_{t-k}. For the MARX(0, 1, 1) model, results are slightly different[8]:

$$\hat{\varphi}_{1,T} = \frac{\sum_{t=r+1}^{T-s} y_t y_{t+1}}{\sum_{t=r+1}^{T-s} y_{t+1}^2} - \hat{\beta}_{\mathrm{NC},T} \frac{\sum_{t=r+1}^{T-s} y_{t+1} x_{t+1}}{\sum_{t=r+1}^{T-s} y_{t+1}^2} \xrightarrow{p} \frac{\gamma_y(1)}{\gamma_y(0)} - \beta_{\mathrm{NC}} \frac{\gamma_{xy}(0)}{\gamma_y(0)},$$

$$\hat{\beta}_{\mathrm{NC},T} = \frac{\sum_{t=r+1}^{T-s} x_{t+1} y_t}{\sum_{t=r+1}^{T-s} x_{t+1}^2} - \hat{\varphi}_{1,T} \frac{\sum_{T=r+1}^{T-s} y_{t+1} x_{t+1}}{\sum_{t=r+1}^{T-s} x_{t+1}^2} \xrightarrow{p} \frac{\gamma_{xy}(1)}{\gamma_x(0)} - \varphi_1 \frac{\gamma_{xy}(0)}{\gamma_x(0)}.$$

For estimation of the autoregressive parameter, information about the cross-covariance between x_t and y_t is contained that is different for both specifications. This is not explained by x_t entering at different time periods in both models. If x_t is considered at the same point in time, the OLS estimates still differ from one another as for cross-covariances $\gamma_{xy}(k) \neq \gamma_{xy}(-k)$ generally ($k = 1, 2, 3, \ldots$). In the same spirit, Cubadda, Hecq, and Telg (2019) show that reduced rank restrictions allow identification of purely causal and noncausal VAR models in a Gaussian framework.

In practice, discriminating between different specifications depends on sample size T and the value of the covariate's coefficient. OLS estimation can be used whenever we restrict our attention to pure models. OLS cannot be applied in the mixed setting due to the multiplicative structure of the polynomials. In that case, we propose the Gaussian ML estimator. The discussion can be generalized to processes with higher order dynamics and multiple regressors with possibly infinite unconditional mean and/or variance. Since the identifiability of the MARX model by either OLS or Gaussian maximum likelihood estimation (MLE) depends on the significance of the coefficients of the exogenous regressors, we propose to (initially) use a non-Gaussian ML estimator.

[8]Without shifted covariate,x_{t+1} should be replaced byx_t and$\gamma_{xy}(0)$ by$\gamma_{xy}(-1)$ in bothequations below.

Remark 2. A problem of model identification by second-order properties can occur when the set of exogenous regressors is not significant in all time directions. Suppose a purely noncausal model (without exogenous regressors) is the true DGP. It can be shown that the residual sum of squares (RSS) of purely causal and noncausal models without exogenous regressors are asymptotically equal. Since noncausal models can create time series with heteroskedasticity in direct time (Gouriéroux & Zakoïan, 2016), the inclusion of volatile exogenous regressors in causal models might improve the fit of the model. Based on RSS, the misspecified causal model with covariates might be preferred. However, as the number of regressors is different in both models, it is better to consider a measure that takes this into account - for example, information criteria.

3 | ESTIMATION AND FORECASTING

3.1 | Approximate maximum likelihood estimation

Maximum likelihood estimation of noncausal autoregressive models has been studied by Breidt, Davis, Lii, and Rosenblatt (1991), Andrews et al. (2006), and Lanne and Saikkonen (2011).[9] They show that ML estimators are consistent and asymptotically normal under general conditions. This section establishes similar results for the MARX model. Similar to Lanne and Saikkonen (2011), we assume that the density function $f_\sigma(x;\boldsymbol{\lambda}) = \sigma^{-1} f(\sigma^{-1}x;\boldsymbol{\lambda})$ satisfies the regularity conditions of Andrews et al. (2006). The permissible parameter space of $\boldsymbol{\lambda}$, which collects the distributional parameters (except for the scale parameter), is denoted by Λ, and is some subset of $\mathbb{R}^d$. The scale parameter σ only takes positive values. The permissible space of $\boldsymbol{\phi}$ and $\boldsymbol{\varphi}$ is defined by the stationarity condition: $\phi(z)$ and $\varphi(z)$ have all roots outside the unit circle. Using the independence of the blocks $(v_1, \dots, v_r)$, $(\varepsilon_{r+1}, \dots, \varepsilon_{T-s})$ and $(u_{T-s+1}, \dots, u_T)$, the density of the process y_t is the product of the densities of these three blocks. Since the densities of the first and third blocks do not depend on sample size T, we approximate the density of y_t by the density of the second block. Using Equation (1) and taking logs, we obtain the following log-likelihood function:

$$L_T(\theta) = \sum_{t=r+1}^{T-s} \ln f_\sigma(\phi(L)\varphi(L^{-1})y_t - \boldsymbol{\beta}'\boldsymbol{X}_t;\boldsymbol{\lambda}) = \sum_{t=r+1}^{T-s} g_t(\boldsymbol{\theta}), \tag{9}$$

where $\boldsymbol{\theta} = [\boldsymbol{\phi}', \boldsymbol{\varphi}', \boldsymbol{\beta}', \boldsymbol{\lambda}', \sigma]'$. We denote the "approximate" sample size $(T-p)$ to compute the log-likelihood by n. We use the definition of the filtered values in Equations (3)–(4) to write u_t and v_t as functions of the parameters; that is, $u_t(\boldsymbol{\phi})$ and $v_t(\boldsymbol{\varphi})$. Then we can characterize $g_t(\boldsymbol{\theta})$ as follows:

$$\begin{aligned} g_t(\theta) &= \ln f\{\sigma^{-1}[v_t(\boldsymbol{\varphi}) - \phi_1 v_{t-1}(\boldsymbol{\varphi}) - \dots - \phi_r v_{t-r}(\boldsymbol{\varphi}) - \boldsymbol{\beta}'\boldsymbol{X}_t];\boldsymbol{\lambda}\} - \ln(\sigma) \\ &= \ln f\{\sigma^{-1}[u_t(\boldsymbol{\phi}) - \varphi_1 u_{t+1}(\boldsymbol{\phi}) - \dots - \varphi_s u_{t+s}(\boldsymbol{\phi}) - \boldsymbol{\beta}'\boldsymbol{X}_t];\boldsymbol{\lambda}\} - \ln(\sigma). \end{aligned}$$

Maximizing $L_T(\boldsymbol{\theta})$ over permissible values of $\boldsymbol{\theta}$ gives an approximate ML estimator of $\boldsymbol{\theta}$. We assume for now that the orders r and s are known and denote the true value of $\boldsymbol{\theta}$ by $\boldsymbol{\theta_0}$ (similarly for its components). Furthermore, assume that $\boldsymbol{\lambda_0}$ is an interior point of Λ. We derive the limiting distribution of the approximate ML estimator. Details are provided in the Appendix.

Lemma 2. *If conditions (A1)–(A7) of Andrews et al. (2006) and Assumptions 1–3 hold, then*

$$\frac{1}{\sqrt{n}} \sum_{t=r+1}^{T-s} \frac{\partial}{\partial \boldsymbol{\theta}} g_t(\boldsymbol{\theta_0}) \xrightarrow{d} \mathcal{N}[\boldsymbol{0}, diag(\Sigma, \Omega)].$$

Moreover, the matrices Σ *and* Ω*, as defined in the Appendix, are positive definite.*

Block diagonality of the covariance matrix of the limiting distribution follows from the formulation of the model in terms of lag and lead operator (Lanne & Saikkonen, 2011) and ensures that the model parameters are orthogonal to the

[9] Breidt et al. (1991) studied an autoregressive model that had roots inside the unit circle, while Andrews et al. (2006) considered all-pass models. Lanne and Saikkonen (2011) introduced the lead operator for the noncausal part.

distributional parameters. The positive definiteness of Ω is assumed through condition (A6) of Andrews et al. (2006). The positive definiteness of Σ follows, similar to the MAR case, from the condition $\mathcal{J} \equiv \int \frac{[f'(x;\boldsymbol{\lambda_0})]^2}{f(x;\boldsymbol{\lambda_0})} dx > 1$.

Theorem 1. *If conditions (A1)–(A7) of Andrews et al. (2006) and Assumptions 1–3 hold, there exists a sequence of local maximizers* $\hat{\theta}_{ML} = [\hat{\boldsymbol{\phi}}', \hat{\boldsymbol{\varphi}}', \hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\lambda}}', \hat{\sigma}]'$ *of* $L_T(\boldsymbol{\theta})$ *in Equation (9) such that*

$$\sqrt{n}(\hat{\theta}_{ML} - \boldsymbol{\theta_0}) \xrightarrow{d} \mathcal{N}\left[\boldsymbol{0}, diag\left(\Sigma^{-1}, \Omega^{-1}\right)\right].$$

3.2 | Computing the covariance matrix

A conventional estimator of Σ is based on the Hessian of the log-likelihood, but nonlinear optimization of this function often involves complicated numerical methods, which are relatively unstable. Similar to Hecq et al. (2016), we provide an alternative way to approximate standard errors based on the asymptotic distribution of the Student t MLE estimated parameters in the finite variance framework. If $\nu > 2$, the MLE is $\sqrt{n}$ consistent and asymptotically normal. Define the $(n \times 1)$ series $\boldsymbol{u} \equiv \boldsymbol{U}_t^* = [u_{r+1}, \ldots, u_{T-s}]'$ up to $\boldsymbol{U}_{t+s}^* = [u_{r+s+1}, \ldots, u_T]'$, $\boldsymbol{V}_{t-r}^* = [v_1, \ldots, v_{T-p}]'$ up to $\boldsymbol{v} \equiv \boldsymbol{V}_t^* = [v_{r+1}, \ldots, v_{T-s}]'$, $\boldsymbol{X}_{i,t} = [x_{i,r+1}, \ldots, x_{i,T-s}]'$ and $\boldsymbol{\epsilon} = [\varepsilon_{r+1}, \ldots, \varepsilon_{T-s}]'$. We construct $\boldsymbol{Z} = \left[\boldsymbol{U}_{t+1}^*, \ldots, \boldsymbol{U}_{t+s}^*, \boldsymbol{X}_{1,t}, \ldots, \boldsymbol{X}_{q,t}\right]$ and similarly $\boldsymbol{Q} = [\boldsymbol{V}_{t-1}^*, \ldots, \boldsymbol{V}_{t-r}^*, \boldsymbol{X}_{1,t}, \ldots, \boldsymbol{X}_{q,t}]$, which are of dimensions $[n \times (s+q)]$ and $[n \times (r+q)]$ respectively. Using this notation, we can write the autoregressions in Equations (3) and (4) in matrix notation by $\boldsymbol{u} = \boldsymbol{Z\zeta} + \boldsymbol{\epsilon}$ and $\boldsymbol{v} = \boldsymbol{Q\xi} + \boldsymbol{\epsilon}$ with $\boldsymbol{\zeta} = [\boldsymbol{\varphi}', \boldsymbol{\beta}']' \in \mathbb{R}^{s+q}$ and $\boldsymbol{\xi} = \left[\boldsymbol{\phi}', \boldsymbol{\beta}'\right]' \in \mathbb{R}^{r+q}$. Conditional on the unobserved causal and noncausal components, it can be shown that in the case of an MARX(r, s, q) model

$$\sqrt{n}(\hat{\boldsymbol{\zeta}}_{\mathrm{ML}} - \boldsymbol{\zeta}_0) \xrightarrow{d} \mathcal{N}\left(\boldsymbol{0}, \frac{\nu+3}{\nu+1}\sigma^2 \boldsymbol{\Upsilon}_{\boldsymbol{\phi}}^{-1}\right), \tag{10}$$

$$\sqrt{n}(\hat{\boldsymbol{\xi}}_{\mathrm{ML}} - \boldsymbol{\xi}_0) \xrightarrow{d} \mathcal{N}\left(\boldsymbol{0}, \frac{\nu+3}{\nu+1}\sigma^2 \boldsymbol{\Upsilon}_{\boldsymbol{\varphi}}^{-1}\right). \tag{11}$$

We use the notation $\boldsymbol{\Upsilon}_{\boldsymbol{\varphi}} = \mathbb{E}[\boldsymbol{Q}'\boldsymbol{Q}]$ and $\boldsymbol{\Upsilon}_{\boldsymbol{\phi}} = \mathbb{E}[\boldsymbol{Z}'\boldsymbol{Z}]$, where $\boldsymbol{\varphi}$ and $\boldsymbol{\phi}$ signify the relation between the unobserved values $\boldsymbol{u}, \boldsymbol{v}$ and $\boldsymbol{y}$ as defined in Equations (3)–(4). These quantities can be estimated consistently by $(1/n)\sum_{i=1}^{n} \boldsymbol{Q}_i'\boldsymbol{Q}_i$ and $(1/n)\sum_{i=1}^{n} \boldsymbol{Z}_i'\boldsymbol{Z}_i$, where $\boldsymbol{Q}_i$ [resp. $\boldsymbol{Z}_i$] denotes the ith row of the matrix $\boldsymbol{Q}$ [resp. $\boldsymbol{Z}$]. For large ν, model parameters cannot be consistently estimated as $L_T(\boldsymbol{\theta})$ approaches the Gaussian log-likelihood. Using the definitions above, results similar to those of Hecq et al. (2016) can be derived for the least absolute deviations estimator.

3.3 | Forecasting

By adapting the methodology of Lanne, Luoto, and Saikkonen (2012), we can compute forecasts for the MARX model. The mean-square sense optimal one-step-ahead forecast is given by

$$\mathbb{E}(y_{T+1}|\Omega_T) = \phi_1 y_T + \ldots + \phi_r y_{T-r+1} + \mathbb{E}\left(\sum_{j=0}^{\infty} \delta_j z_{t+j+1} \mid \Omega_T\right), \tag{12}$$

where z_{t+j} is defined as in Equation (2). To obtain a close approximation, the last term (i.e., the infinite sum) is truncated by a sufficiently large number M. The forecasting method can be described as follows. From the conditional distribution $(z_{T+1}, \ldots, z_{T+M}|y_1, \ldots, y_T)$ we simulate N mutually independent realizations of $(z_{T+1}, \ldots, z_{T+M})$, which are plugged into Equation (12) to obtain the one-step-ahead forecast y_{T+1}.[10] We assume that M future values of $\boldsymbol{X}_t$ are known, such that they enter the density function as a constant. Abandoning this assumption requires a further extension of the forecasting procedure, as one needs the joint density of $(z_{T+1}, \ldots, z_{T+M})$. Evaluation of the distribution of z_t is outside the scope of this paper. In the case of ex post forecasting, realized future values of $\boldsymbol{X}_t$ can be used to forecast y_t. Whenever we want to perform an ex ante forecasting exercise, one needs auxiliary equations to forecast $\boldsymbol{X}_t$ and subsequently y_t. We study

[10] As proposed by Lanne, Luoto, and Saikkonen (2012), we set $M = 50$ and $N = 10,\ 000$ in this paper.

TABLE 2 Frequency (%) with which the order p is selected for the MARX model

	$T = 100$			$T = 200$			$T = 500$			$T = 1,\ 000$		
p	AIC	BIC	HQ	AIC	BIC	HQ	AIC	BIC	HQ	AIC	BIC	HQ
0	0.3	0.3	0.3	0.3	0.3	0.3	0.0	0.0	0.0	0.0	0.0	0.0
1	61.6	82.4	72.6	55.3	74.8	65.3	30.0	47.0	34.9	0.2	4.1	0.7
2	33.5	16.4	25.2	41.8	24.1	32.9	67.1	52.9	63.8	96.4	95.5	97.6
3	4.6	0.9	1.9	2.6	0.8	1.6	2.9	0.1	1.3	3.4	0.4	1.7
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

TABLE 3 Frequency (%) with which the order p is selected for the MAR model

	$T = 100$			$T = 200$			$T = 500$			$T = 1,\ 000$		
p	AIC	BIC	HQ	AIC	BIC	HQ	AIC	BIC	HQ	AIC	BIC	HQ
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	41.0	67.7	53.3	26.3	56.8	39.0	3.5	19.6	8.1	0.0	1.8	0.2
2	52.7	30.1	43.2	63.8	40.8	56.0	86.9	78.9	87.3	84.7	96.5	93.4
3	6.3	2.2	3.5	10.0	2.4	5.0	9.6	1.5	4.6	15.3	1.7	6.4
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

both forecasting setups by simulation in the next section. An alternative forecasting procedure for noncausal models is the sample-based method proposed by Gouriéroux and Jasiak (2016), and an extensive comparison of the two methods can be found in Hecq and Voisin (2019).

4 | MONTE CARLO SIMULATION STUDY

We study three different cases of interest for MARX processes: (i) the model selection procedure, (ii) identifiability under Gaussianity, and (iii) forecasting. Each table in this simulation study reports results for 10,000 replications. For cases (i) and (ii), the DGP is the MARX(1, 1, 1):

$$(1 - 0.3L)(1 - 0.5L^{-1})y_t - \beta_1 x_t = \varepsilon_t, \tag{13}$$

with the error term $\varepsilon_t \overset{\text{i.i.d.}}{\sim} t(3, 1)$. For (i), $\beta_1 = 0.3$, the value of β_1 takes various values in (ii) and x_t follows different specifications. The DGP for (iii) is detailed in that respective section.[11]

4.1 | Model selection

We evaluate the proposed model selection procedure in Section 2.2. We simulate Equation (13) with $x_t \overset{\text{i.i.d.}}{\sim} t(2, 1)$ and limit our attention to the estimation of purely causal ARX(p, 1) models by Gaussian MLE, where $p = 0, \ldots, 4$. Table 2 shows the percentages with which AIC, BIC, and HQ select a certain order p (true order equals 2). The same exercise has been done on the MAR model, for which we use the same specification, only without exogenous variable x_t; that is, $\beta_1 = 0$. The corresponding frequencies for the MAR model can be found in Table 3.

All information criteria tend to underestimate the true lag order (especially BIC) in small samples. Performance improves when T grows, but at $T = 500$ we still only observe a correct lag order selection in around 65% of the cases. Information criteria might not perform optimally in finite samples (Hurvich & Tsai, 1989), but the performance in the MARX setup is considerably worse compared to MAR models for most T. This stresses the usage of diagnostic tests to discover model fit improvements. Therefore, we test for autocorrelation (Box–Pierce and LM tests) in the empirical section after identifying the pseudo-causal model. We increase the lag length p if necessary (see Fries & Zakoian, 2019, for new Box–Pierce tests for MAR models). Further, we suppose the order $p = 2$ is known and examine the selection of MARX(r, s, 1) and MAR(r, s) models with $r + s = 2$ based on the highest log-likelihood. We consider two cases for the MARX model: (i) shifting x_t along the lines of Section 2.2 and (ii) considering x_t at the same point in time for all models. In Table 4 we observe that the model selection procedure improves with sample size for all models considered. The MARX model with shifted x_t, denoted MARX-s, performs best and shows extremely satisfactory results already at $T = 100$. Identification is likely to be achieved more swiftly as there are two types of information in the DGP: the type of dynamics (i.e., causal, noncausal, or mixed) and how x_t is linked with respect to y_t in time. The MAR and the MARX model without

[11]The Supplementary Material contains additional simulations—for example, on the performance of the ML estimator.

TABLE 4 Lag–lead order (r, s) selected by highest log-likelihood for $p = 2$ (in %)

	$T = 100$			$T = 200$			$T = 500$			$T = 1,\ 000$		
	(2, 0)	(1, 1)	(0, 2)	(2, 0)	(1, 1)	(0, 2)	(2, 0)	(1, 1)	(0, 2)	(2, 0)	(1, 1)	(0, 2)
MARX-s	1.22	98.55	0.23	0.02	99.98	0.00	0.00	100.00	0.00	0.00	100.00	0.00
MARX	0.24	87.82	11.94	0.01	96.11	3.88	0.00	99.70	0.30	0.00	100.00	0.00
MAR	2.42	89.54	8.04	0.31	97.71	1.98	0.00	99.91	0.09	0.00	100.00	0.00

TABLE 5 Frequency (%) with which the correct model is chosen

β_1	$T = 100$	$T = 200$	$T = 500$	$T = 1,\ 000$
0	41.27	41.72	27.28	33.35
0.3	69.37	84.51	94.63	99.56
0.5	89.20	97.32	99.70	99.97
0.8	98.17	99.59	99.98	99.99

TABLE 6 Frequency (%) with which the correct model is chosen

	$T = 100$	$T = 200$	$T = 500$	$T = 1,\ 000$
OLS	85.41	94.74	99.26	99.87
Gaussian MLE	85.57	94.81	99.26	99.88
t-MLE	98.96	99.98	100.00	100.00

shift do not contain the second source of information, leading to more doubt between the true mixed or the misspecified noncausal specification.

The model selection procedure is sensitive to the values of the parameters and the error distribution chosen in the DGP. Consider the case in which $\phi_1 = 0.1$ and $\varphi_1 = 0.7$. Since the value of ϕ_1 is low, first-order models are selected too often in the first step of the model selection procedure. As $\varphi_1 \gg \phi_1$, a noncausal model will be selected in the second step. If we consider $\phi_1 = 0.5$ and $\varphi_1 = 0.7$ instead, it is less likely that first-order models are selected too often in the first step. Hence it is also more probable that the "correct" MAR model is chosen in the second step. This suggests the use of complementary analysis (e.g., bootstrap or cross-validation).

4.2 | Identifiability

We study whether MARX processes can be identified using solely second-order properties. We simulate an MARX(1, 1, 1) process from Equation (13), where $x_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$.[12] The coefficient β_1 takes several values. After the process is generated, three models are estimated by Gaussian MLE: MARX(2, 0, 1), MARX(1, 1, 1), and MARX(0, 2, 1). The selected model is that which attains the highest log-likelihood. The results can be found in Table 5. We use Gaussian MLE to estimate the models due to the multiplicative structure of polynomials in the MARX(1, 1, 1) specification.

When $\beta_1 = 0$, second-order properties cannot distinguish backward- and forward-looking behavior as the process reduces to an MAR. All three models are chosen with approximately equal probability. Identification of the correct model increases for larger β_1 and T. Frequencies are very high already when $\beta_1 = 0.3$ for most T. Identification of MARX models using Gaussian MLE is dependent on: (i) the value of the coefficients of the covariates and (ii) the type of dynamics they exhibit. Next, we study identification using different estimation methods based on the following DGP:

$$y_t = 0.6y_{t+1} + 0.5x_t + \varepsilon_t, \tag{14}$$

with $\varepsilon_t \overset{\text{i.i.d.}}{\sim} t(3, 1)$ and x_t standard normal. We estimate the "true" model in Equation (14) and the pseudo-causal model (with shift in x_t) by OLS, Gaussian MLE, and t-MLE. The model with the highest log-likelihood is chosen to be the final model. Table 6 shows the frequency with which the true noncausal model is chosen. Two main conclusions can be drawn: (i) t-MLE outperforms OLS as well as Gaussian MLE and (ii) frequencies are above 85% already for $T = 100$ for all estimation procedures. Results are even better when x_t exhibits dynamics. As simulation results differ depending on chosen parameter values, we take rather moderate values for both parameters. The misspecified models reveal different behavior based on, among others: (i) the value chosen for φ_1, (ii) the value chosen for β_1, (iii) the type of process x_t (e.g., static, dynamic), and (iv) the distribution of x_t and ε_t. Gouriéroux and Jasiak (2018) studied the exact link between the DGP and corresponding misspecified MAR models, but this is outside the scope of this paper.

[12]Simulation results where x_t follows an AR(1) process are not reported here, as they are qualitatively similar.

TABLE 7 RMSFE for different model specifications

	Low persistence ($\gamma = 0.3$)		High persistence ($\gamma = 0.8$)	
	$w_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$	$w_t \overset{\text{i.i.d.}}{\sim} t(3,1)$	$w_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$	$w_t \overset{\text{i.i.d.}}{\sim} t(3,1)$
MARX1	2.040	1.923	1.983	1.965
MARX2	2.386	2.766	2.540	4.626
MAR1	2.441	2.842	2.543	4.703
MAR2	2.400	2.798	2.499	4.993

4.3 | Forecasting

The DGP is the MARX(0, 1, 1) where the covariate x_t follows a causal AR(1) process; i.e.,

$$y_t = 0.7y_{t+1} + 0.8x_t + \epsilon_t, \text{with } x_t = \gamma x_{t-1} + w_t, \tag{15}$$

with $\epsilon_t \overset{\text{i.i.d.}}{\sim} t(3,1)$. We examine a low- ($\gamma = 0.3$) and a high-persistence ($\gamma = 0.8$) case and vary the distribution for the error term w_t in the auxiliary regression equation. Two models are considered for forecasting y_t: the MARX(0, 1, 1) and MAR(0, 1). We use $T = 200$ observations to estimate the model and forecast $h = 8$ observations using one-step-ahead forecasts. To make a broad comparison of the forecasting performance of MARX and MAR models, we consider several setups. For the MARX model, we examine ex post and ex ante forecasts. For the first case, we use realized future values of x_t (MARX1), while we forecast the covariate directly as a causal AR(1) under a known parameter value for the second case (MARX2). For the MAR model, we also study two setups. In the first case, we fit an MAR model to y_t and produce forecasts based on these parameter estimates (MAR1). In the second case, forecasts are computed based on the first equation of Equation (15) without taking the exogenous regressor into account (MAR2). In this way, we assess the effect of an omitted variable on the forecasting performance.

Table 7 shows average root mean squared forecast errors (RMSFE) over all replications for the four setups. In all cases, MARX1 has the lowest RMSFE, which is sensible as future information is used to compute forecasts of y_t. The RMSFE for the other models follow at a respectable distance. MARX2 has forecasting advantages when x_t contributes largely to y_t—for example, when w_t follows a fat-tailed distribution. Benefits are larger in the high-persistence case, as forecasts for x_t return to their unconditional mean more slowly. MAR1 outperforms MAR2 in such instances, as the predictable effect of the covariate is absorbed in the noncausal parameter. When x_t contributes minimally to y_t, MAR forecasts can outperform MARX2. The RMSFE of these three are relatively close and there is no clear evidence in favor of the MAR or MARX.

5 | EMPIRICAL APPLICATION

5.1 | The data

We study nonseasonally adjusted monthly commodity prices $\text{CP}_{i,t}$ from 1980:M1 to 2019:M4—that is, 472 observations for $i = 1, \ldots, 5$ indexes released by the IMF.[13] These are benchmark prices determined by the largest exporter of a given commodity and thus representative of the global market. IMF releases many individual commodity prices but we only take the following indexes as variables of interest: BEVE (beverage price index), COP (copper price index), NICK (nickel price index), OIL (crude oil price index), and RAWM (agricultural raw materials index). As exogenous regressors, we consider the trade weighted US dollar index (S_t) and the industrial production index (IP_t) from the Federal Reserve Bank of St. Louis database.[14] We justify strict exogeneity of the industrial production index and the dollar index by the small size of the US economy on trade for almost all of these commodities. A potential problem in this statement is that the classifications of the commodity indices do not exactly match those of trade statistics. However, we find a very good fit in almost all cases. For example, regarding RAWM, for 2016, the US share of raw material exports and imports were 9.42% and 8.43% respectively. For OIL, in 2017, US exports represented a mere 2.6% of the total. The method to detrend series before identifying MAR models is an ongoing debate. Hencic and Gouriéroux (2015) fit a cubic trend to Bitcoin data. We rely on unit root analysis and, using ADF tests, we do not reject a unit root at the 5% significance level in each

[13]IMF primary commodity prices; see http://www.imf.org/external/np/res/commod/index.aspx.
[14]Series named TWEXBMTH and INDPRO at https://fred.stlouisfed.org.

TABLE 8 Estimation results for pseudo-causal models by OLS, HCSE in (·)

	Commodities				
	BEVE	**COP**	**NICK**	**OIL**	**RAWM**
p_{HQ}	1	1	1	1	2
c	0.001	0.004	0.003	0.004	0.002
	(0.002)	(0.003)	(0.004)	(0.004)	(0.001)
a_1	0.270	0.263	0.299	0.274	0.194
	(0.069)	(0.058)	(0.044)	(0.063)	(0.046)
a_2					0.169
					(0.045)
b_1	-0.554	-1.652	-1.564	-1.613	-0.524
	(0.180)	(0.280)	(0.333)	(0.321)	(0.109)
b_2	0.571	0.849	1.267	0.965	0.240
	(0.289)	(0.402)	(0.488)	(0.534)	(0.192)
$\bar{R}^2$	0.104	0.236	0.176	0.165	0.162
JB	<0.001	<0.001	<0.001	<0.001	0.069
LM ARCH[1,2]	<0.001	0.009	0.871	<0.001	0.008

TABLE 9 Values for the log-likelihood of different models (highest value in bold)

Candidate model	**BEVE**	**OIL**	**NICK**	**COP**
MARX$(1,0,2)$ with $\boldsymbol{X}_t$	776.623	563.314	**585.320**	**732.792**
MARX$(0,1,2)$ with $\boldsymbol{X}_{t+1}$	781.880	549.099	566.490	701.597
MARX$(0,1,2)$ with $\boldsymbol{X}_t$	**784.588**	**565.910**	577.364	723.192
MARX$(1,0,2)$ with $\boldsymbol{X}_{t-1}$	772.447	546.849	570.342	706.584

series. We consequently work with monthly growth rates (in %): $\Delta \text{cp}_{i,t} \equiv 100(1-L)\ln \text{CP}_{i,t}$, $\Delta \text{ip}_t \equiv 100(1-L)\ln \text{IP}_t$ and $\Delta \text{s}_t \equiv 100(1-L)\ln \text{S}_t$.

5.2 | Identification of pseudo-causal models

Let $y_{i,t} = \Delta \text{cp}_{i,t}$ and $\boldsymbol{X}_t = [\Delta s_t, \Delta \text{ip}_t]'$. As outlined in Section 2.3, the first step of our modeling strategy identifies the pseudo-causal model for the dependent variable $y_{i,t}$. Since variables are not demeaned, we include an intercept in all models. The first part of Table 8 reports the lag order chosen by HQ after estimating ARDL models up to order $p_{\max} = 8$ (both for $y_{i,t}$ and $\boldsymbol{X}_t$). We rely on HQ because of its good results in Section 4.1. We check correlograms and perform LM tests for the null of no autocorrelation on the residuals. These results confirm the lag order chosen by HQ. We do not reject the null of linear independence in a simple regression between the two exogenous variables in $\boldsymbol{X}_t$ and no lags are selected for these covariates.[15]

The second part of Table 6 reports the estimation results, with heteroskedasticity-consistent standard errors (HCSE) in brackets. Commodity prices depend on their own lags, the exchange rate, and industrial production (except for OIL and BEVE) at the 5% significance level. The highest negative effect of the growth rate of the exchange rate is on the COP, OIL, and NICK indexes. This result is plausible, as these products heavily depend on exports and are thus more negatively influenced by an increase in the US dollar. The last rows of Table 8 report adjusted R^2, p-values of the Jarque–Bera test for normality, and the LM ARCH[1,2] test for no autocorrelation up to order two in the squared residuals. Both tests follow a χ^2 distribution with 2 degrees of freedom under the null. We reject the null of normality in every equation (except for RAWM).

5.3 | Identification and estimation of MARX

For the four commodities that have an order equal to one, we characterize the remaining steps of the procedure presented in Section 2.2. Table 9 displays the values of the log-likelihood for four different models. The first two rows represent models that are theoretically equivalent using the ARDL as starting point; the last two rows are equivalent models with the noncausal model as initial choice. There are no remaining models to be tested as specified in Step 4 of the algorithm.

For BEVE and OIL a noncausal model with contemporaneous $\boldsymbol{X}_t$ is selected, while NICK and COP admit a causal representation. Since RAWM has a different autoregressive order from the other commodities, it is excluded from Table 9, as the set of candidate models is larger. Since $p = 2$, we follow the procedure as presented in Example 2 to find the model

[15]This step can be done using, for example, EViews ARDL estimation features or the R package MARX.

TABLE 10 Estimation results for MARX models by Student t MLE, SE in (·)

	Commodities				
	BEVE	**COP**	**NICK**	**OIL**	**RAWM**
	MARX(0, 1, 2)	**MARX**(1, 0, 2)	**MARX**(1, 0, 2)	**MARX**(0, 1, 2)	**MARX**(0, 2, 2)
α	-0.118	0.398	0.045	0.771	0.162
	(0.190)	(0.211)	(0.300)	(0.394)	(0.119)
ϕ_1		0.278	0.296		
		(0.034)	(0.037)		
φ_1	0.267			0.231	0.202
	(0.037)			(0.037)	(0.043)
φ_2					0.155
					(0.043)
β_1	-0.335	-1.295	-1.290	-1.526	-0.496
	(0.144)	(0.160)	(0.227)	(0.230)	(0.090)
β_2	0.222	0.783	1.451	0.364	0.287
	(0.279)	(0.310)	(0.441)	(0.447)	(0.174)
ν	3.965	3.831	5.016	4.058	13.114
	(0.761)	(0.767)	(1.048)	(0.812)	(7.240)
σ	3.471	3.840	5.627	5.573	2.393
	(0.202)	(0.238)	(0.295)	(0.330)	(0.128)

with the highest log-likelihood. The selected model is the MARX(0, 2, 2) with contemporaneous $\boldsymbol{X}_t$ and a log-likelihood value of 1,039.573. However, the causal specification lies close, with a value of 1,039.312. As the null of normality cannot be rejected on the residuals of the RAWM series, an identification problem occurs. For RAWM, the presence of exogenous regressors does not create a clear asymmetry in different time directions, which complicates identification for the MARX.

Table 10 reports the final results for each commodity; standard errors (SE) are in parentheses. Distributions are leptokurtic: The smallest value for the degrees of freedom parameter is $\hat{\nu} = 3.833$ for COP; the largest value is $\hat{\nu} = 13.186$ for RAWM, which is in line with the result of the JB test. The negative impact of the exchange rate is more pronounced for OIL, and the smallest value is for the BEVE commodity index. The coefficient of the industrial production index is not significantly different from zero in the OIL and BEVE equations. For all commodity series we do not reject the null of no autocorrelation, but we still observe nonnormality and heteroskedasticity using standard diagnostic tests on the residuals of the identified MARX models. Tests for homoskedasticity proposed by Gouriéroux and Zakoïan (2016) and Cavaliere, Nielsen, and Rahbek (2020) assume that the true model is noncausal with i.i.d. Cauchy error term, which is not the case here, as most series are heteroskedastic both in direct and reverse time.

Lanne and Saikkonen (2011) claim that the errors in mixed causal–noncausal models contain effects of omitted variables that are predictable by the considered series. Suppose the variables $\boldsymbol{X}_t$ are indeed part of the DGP of the commodity prices $y_{i,t}$ but we fail to include them in our model. The identified models can be found in Table 11. We find that, in all cases except for RAWM, the causal and noncausal orders of the MAR and MARX are the same. However, the estimated values of both the causal and noncausal coefficients are larger in absolute value for the MAR model. Additionally, the degrees of freedom parameters are lower for the MAR models. Small differences indicate that the MARX models involve only minor benefits compared to the corresponding MAR model. A special case is the RAWM series, for which the degrees of freedom parameter is estimated to be much lower in the MAR specification ($\hat{\nu} = 6.830$, compared to $\hat{\nu} = 13.114$ for the MARX). Values of the log-likelihood for the different MAR models lie relatively far from another, which allows clear identification. The MARX, however, shows that the fat-tailedness of the series is mostly due to the omission of explanatory variables.

A few observations are in order. Successful identification of the MARX using second-order properties depends on factors such as the type of process generating the exogenous regressors as well as the value of their coefficients. Besides, we discussed that the degrees of the polynomials might increase depending on the dynamics of the exogenous regressors. These orders probably did not increase as the exchange rate resembles a white noise series. We do not claim that these models capture all the dynamics in the series. It could be that the inclusion of more exogenous regressors is needed or that a system of equations is a more adequate modeling approach. However, the inclusion of exogenous regressors clearly affects the autoregressive parameters and thus caution is required when interpreting the values of the structural parameters in Equation (8). The main debate in the literature evolves around the importance of forward- and backward-looking behavior. Results suggest that this comparison is distorted when relevant covariates are omitted.

TABLE 11 Estimation results for MAR models by Student *t* MLE, SE in (·)

	Commodities				
	BEVE	**COP**	**NICK**	**OIL**	**RAWM**
	MAR(0, 1)	**MAR**(1, 0)	**MAR**(1, 0)	**MAR**(0, 1)	**MAR**(1, 1)
α	-0.178	0.176	-0.144	0.389	0.058
	(0.190)	(0.225)	(0.310)	(0.318)	(0.117)
ϕ_1		0.329	0.338		-0.334
		(0.036)	(0.038)		(0.041)
φ_1	0.284			0.287	0.561
	(0.037)			(0.038)	(0.036)
ν	3.887	3.824	4.889	4.380	6.830
	(0.695)	(0.708)	(1.024)	(0.913)	(2.083)
σ	3.465	4.110	5.805	5.893	2.261
	(0.201)	(0.239)	(0.310)	(0.340)	(0.126)

6 | CONCLUSION

This paper proposes to estimate mixed causal–noncausal models including exogenous regressors by non-Gaussian ML. A detailed description of the method to estimate and select MARX models is given, highlighting the implications of covariates. We find that the presence of exogenous regressors makes it potentially possible to discriminate between different MARX specifications using second-order properties. Monte Carlo studies assess the model selection procedure in finite samples and compare the forecasting performance of MARX and MAR models in several setups, favoring the former in the case of ex post forecasting. We provide an empirical study on commodity prices, the exchange rate, and the industrial production index. We identify noncausal models for BEVE, OIL, and RAWM, and causal models for NICK and COP based on maximizing the log-likelihood. A comparison is made with MAR models: The same causal and noncausal orders are selected in most cases, but the estimated autoregressive parameters are larger in absolute value, while the estimates for the degrees of freedom parameter are lower. This suggests that the MAR model compensates for the omission of relevant regressors through these channels.

ACKNOWLEDGMENTS

This work was partly written while Sean Telg visited the CREST in Paris, Alain Hecq EPGE/FGV in Rio de Janeiro, and João Victor Issler Maastricht University. We thank all institutions for hosting us. We would like to express gratitude to Christian Francq and Jean-Michel Zakoïan for stimulating and fruitful discussions. We also thank participants of CFE (Seville, 2016), SNDE (Paris, 2017), EcoSta (Hong Kong, 2017), IAAE (Sapporo, 2017), and ESEM (Lisbon, 2017), as well as two anonymous referees for valuable comments and remarks. João Victor Issler acknowledges the financial support of CNPq, FAPERJ, and CAPES on different grants. This study was partly financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES; Finance Code 001). The authors declare no conflict of interest.

OPEN RESEARCH BADGES

This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at [http://qed.econ.queensu.ca/jae/datasets/hecq001/].

REFERENCES

Alessi, L., Barigozzi, M., & Capasso, M. (2011). Non-fundamentalness in structural econometric models: A review. *International Statistical Review*, *79*, 16–47. https://doi.org/10.1111/j.1751-5823.2011.00131.x

Andrews, B., Breidt, F., & Davis, R. (2006). Maximum likelihood estimation for all-pass time series models. *Journal of Multivariate Analysis*, *97*, 1638–1659. https://doi.org/10.1016/j.jmva.2006.01.005

Breidt, F., Davis, R., Lii, K., & Rosenblatt, M. (1991). Maximum likelihood estimation for noncausal autoregressive processes. *Journal of Multivariate Analysis*, *36*, 175–198. https://doi.org/10.1016/0047-259X(91)90056-8

Brockwell, P., & Davis, R. (1991). *Time series theory and methods* (2nd ed.). New York, NY: Springer.

Cavaliere, G., Nielsen, H. B., & Rahbek, A. (2020). Bootstrapping noncausal autoregressions: With applications to explosive bubble modeling. *Journal of Business and Economic Statistics*, *38*(1), 55–67. https://doi.org/10.1080/07350015.2018.1448830

Cubadda, G., Hecq, A., & Telg, S. (2019). Serial correlation common noncausal features. *Oxford Bulletin of Economics and Statistics*, *81*, 697–715. https://doi.org/10.1111/obes.12281

Fries, S., & Zakoian, J.-M. (2019). Mixed causal–noncausal AR processes and the modelling of explosive bubbles. *Econometric Theory*, *35*, 1234–1270. https://doi.org/10.1017/S0266466618000452

Gouriéroux, C., & Jasiak, J. (2016). Filtering, prediction and simulation methods for noncausal processes. *Journal of Time Series Analysis*, *37*, 405–430. https://doi.org/10.1111/jtsa.12165

Gouriéroux, C., & Jasiak, J. (2018). Misspecification of noncausal order in autoregressive processes. *Journal of Econometrics*, *205*, 226–248. https://doi.org/10.1016/j.jeconom.2018.03.012

Gouriéroux, C., & Zakoïan, J. M. (2016). Local explosion modelling by non-causal process. *Journal of the Royal Statistical Society, Series B*, *79*, 737–756. https://doi.org/10.1111/rssb.12193

Hecq, A., Lieb, L., & Telg, S. (2016). Identification of mixed causal–noncausal models in finite samples. *Annals of Economics and Statistics*, *123*(124), 307–331. https://doi.org/10.15609/annaeconstat2009.123-124.0307

Hecq, A., Lieb, L., & Telg, S. (2017). Simulation, estimation and selection of mixed causal-noncausal autoregressive models: The MARX package. (*Working paper*) Available at SSRN: https://ssrn.com/abstract=3015797

Hecq, A., Telg, S., & Lieb, L. (2017). Do seasonal adjustments induce noncausal dynamics in inflation rates? *Econometrics*, *5*, 48. https://doi.org/10.3390/econometrics5040048

Hecq, A., & Voisin, E. (2019). Forecasting bubbles with mixed causal–noncausal autoregressive models. (*MPRA Paper No. 92734*). Munich, Germany: University Library of Munich.

Hencic, A., & Gouriéroux, C. (2015). Noncausal autoregressive model in application to Bitcoin/USD exchange rates. In V. N. Huynh, V. Kreinovich, S. Sriboonchitta, & K. Suriya (Eds.), *Econometrics of risk: Studies in computational intelligence*. Berlin, Germany: Springer, pp. 17–40.

Hurvich, M., & Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*, 297–307. https://doi.org/10.1093/biomet/76.2.297

Issler, J. V., Rodrigues, C., & Burjack, R. (2014). Using common features to understand the behavior of metal-commodity prices and forecast them at different horizons. *Journal of International Money and Finance*, *42*, 310–335. https://doi.org/10.1016/j.jimonfin.2013.08.017

Karapanagiotidis, P. (2014). Dynamic modeling of commodity futures prices. (*MPRA Paper No. 56805*). Munich, Germany: University Library of Munich.

Lanne, M., & Luoto, J. (2013). Autoregression-based estimation of the New Keynesian Phillips curve. *Journal of Economic Dynamics and Control*, *37*, 561–570. https://doi.org/10.1016/j.jedc.2012.09.008

Lanne, M., Luoto, J., & Saikkonen, P. (2012). Optimal forecasting of noncausal autoregressive time series. *International Journal of Forecasting*, *28*, 623–631. https://doi.org/10.1016/j.ijforecast.2011.08.003

Lanne, M., Nyberg, H., & Saarinen, E. (2012). Does noncausality help in forecasting economic time series? *Economics Bulletin*, *32*(4), 2849–2859.

Lanne, M., & Saikkonen, P. (2011). Noncausal autoregressions for economic time series. *Journal of Time Series Econometrics*, *3*(3), 1–32. https://doi.org/10.2202/1941-1928.1080

Lof, M. (2013). Noncausality and asset pricing. *Studies in Nonlinear Dynamics and Econometrics*, *17*, 211–220. https://doi.org/10.1515/snde-2012-0035

Lof, M., & Nyberg, H. (2017). Noncausality and the commodity currency hypothesis. *Energy Economics*, *65*, 424–433. https://doi.org/10.1016/j.eneco.2017.05.024

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Hecq A, Issler J V, Telg S. Mixed causal–noncausal autoregressions with exogenous regressors. *J Appl Econ*. 2020;35:328–343. https://doi.org/10.1002/jae.2751

APPENDIX A: DEFINITIONS FOR LEMMA 2

Consider the score of θ evaluated at true parameter values. Define $\boldsymbol{V}_{t-1} = [v_{t-1}, \dots, v_{t-r}]'$ and $\boldsymbol{U}_{t+1} = [u_{t+1}, \dots, u_{t+s}]'$, where u_t and v_t are defined in terms of true parameter values; that is, $u_t = \sum_{j=0}^{\infty} \delta_{0j} \left(\sum_{i=1}^{q} \beta_{0i} x_{i,t+j} + \varepsilon_{t+j} \right)$ and $v_t =$

$\sum_{j=0}^{\infty}\alpha_{0j}\left(\sum_{i=1}^{q}\beta_{0i}x_{i,t-j}+\varepsilon_{t-j}\right)$. By direct differentiation of Equation (9), we obtain

$$\frac{\partial}{\partial\boldsymbol{\phi}}g_t(\boldsymbol{\theta_0})=-\frac{f'\left(\sigma_0^{-1}\varepsilon_t;\boldsymbol{\lambda_0}\right)}{\sigma_0 f(\sigma_0^{-1}\varepsilon_t;\boldsymbol{\lambda_0})}\boldsymbol{V}_{t-1},\quad \frac{\partial}{\partial\boldsymbol{\varphi}}g_t(\boldsymbol{\theta_0})=-\frac{f'\left(\sigma_0^{-1}\varepsilon_t;\boldsymbol{\lambda_0}\right)}{\sigma_0 f\left(\sigma_0^{-1}\varepsilon_t;\boldsymbol{\lambda_0}\right)}\boldsymbol{U}_{t+1},$$

and

$$\frac{\partial}{\partial\boldsymbol{\beta}}g_t(\boldsymbol{\theta_0})=-\frac{f'\left(\sigma_0^{-1}\varepsilon_t;\boldsymbol{\lambda_0}\right)}{\sigma_0 f\left(\sigma_0^{-1}\varepsilon_t;\boldsymbol{\lambda_0}\right)}\boldsymbol{X}_t,$$

where $f'(x;\boldsymbol{\lambda})=\partial f(x;\boldsymbol{\lambda})/\partial x$ and use has been made of the fact that $\varphi_0(L^{-1})u_t-\boldsymbol{\beta}_0'\boldsymbol{X}_t=\varepsilon_t=\phi_0(L)v_t-\boldsymbol{\beta}_0'\boldsymbol{X}_t$. Similarly, for the distributional parameters,

$$\frac{\partial}{\partial\sigma}g_t(\boldsymbol{\theta_0})=-\sigma_0^2\left(\frac{f'\left(\sigma_0^{-1}\varepsilon_t;\boldsymbol{\lambda_0}\right)}{f\left(\sigma_0^{-1}\varepsilon_t;\boldsymbol{\lambda_0}\right)}+\sigma_0\right),\quad \frac{\partial}{\partial\boldsymbol{\lambda}}g_t(\boldsymbol{\theta_0})=\frac{1}{f\left(\sigma_0^{-1}\varepsilon_t;\boldsymbol{\lambda_0}\right)}\frac{\partial}{\partial\boldsymbol{\lambda}}f\left(\sigma_0^{-1}\varepsilon_t;\boldsymbol{\lambda_0}\right).$$

Lemma 2 presents the asymptotic distribution of the score vector. Let $\mathcal{J}\equiv\int\frac{[f'(x;\boldsymbol{\lambda_0})]^2}{f(x;\boldsymbol{\lambda_0})}dx>1$ and $\mathcal{I}\equiv\int x^2\frac{[f'(x;\boldsymbol{\lambda_0})]^2}{f(x;\boldsymbol{\lambda_0})}dx-1$. The first inequality follows from Remark 2 in Andrews et al. (2006). Furthermore, set Σ as a 3×3 block matrix with entries Σ_{kl} for $k,l=1,2,3$. The matrix Σ is symmetric and has the matrices $\Sigma_{11}=\sigma_0^{-2}\mathcal{J}\boldsymbol{\gamma}_V$, $\Sigma_{22}=\sigma_0^{-2}\mathcal{J}\boldsymbol{\gamma}_U$ and $\Sigma_{33}=\sigma_0^{-2}\mathcal{J}\boldsymbol{\gamma}_X$ on the diagonal, where $\boldsymbol{\gamma}_V$ and $\boldsymbol{\gamma}_U$ are the autocovariance matrices of $\boldsymbol{V}_{t-1}$ and $\boldsymbol{U}_{t+1}$. $\boldsymbol{\gamma}_X$ is the cross-covariance matrix of $\boldsymbol{X}_t$, which is diagonal under Assumption 3. Σ_{12} is an $(r\times s)$ matrix where the (i,j)th element equals: $\sum_{t=0}^{\infty}\alpha_t\delta_{t+i-j}+\tilde{\mathcal{J}}\sum_{a=0}^{\infty}\sum_{b=0}^{\infty}\alpha_a\delta_b\sum_{m=1}^{q}\beta_m^2\gamma_{x_m}(i+j+a+b)$. The Σ_{13} matrix has size $(r\times q)$ with the (i,j)th element $\beta_j\sigma^{-2}\mathcal{J}\sum_{a=0}^{\infty}\alpha_a\gamma_{x_j}(i+a)$, while this element equals $\beta_j\sigma^{-2}\mathcal{J}\sum_{b=0}^{\infty}\delta_b\gamma_{x_j}(i+b)$ for Σ_{23}. The $(d+1)\times(d+1)$ matrix Ω is defined as

$$\Omega=\begin{bmatrix}\omega_\sigma^2 & \boldsymbol{\omega}_{\sigma\lambda}\\ \boldsymbol{\omega}_{\lambda\sigma} & \Omega_{\lambda\lambda}\end{bmatrix},$$

where $\Omega_{\lambda\lambda}\equiv\int\frac{1}{f(x;\boldsymbol{\lambda_0})}\left[\frac{\partial}{\partial\boldsymbol{\lambda}}f(x;\boldsymbol{\lambda})\right]\left[\frac{\partial}{\partial\boldsymbol{\lambda}}f(x;\boldsymbol{\lambda})\right]'dx$, $\boldsymbol{\omega}_{\lambda\sigma}\equiv-\sigma_0\int x\frac{f'(x;\boldsymbol{\lambda_0})}{f(x;\boldsymbol{\lambda_0})}\frac{\partial}{\partial\boldsymbol{\lambda}}f(x;\boldsymbol{\lambda_0})dx=\boldsymbol{\omega}_{\sigma\lambda}'$ and $\omega_\sigma^2\equiv\omega_0^{-2}\mathcal{I}$.