ONLINE APPENDIX FOR: MIXED CAUSAL-NONCAUSAL AUTOREGRESSIONS WITH EXOGENOUS REGRESSORS

Alain Hecq

Maastricht University, School of Business and Economics, Department of Quantitative Economics, P.O. Box 616, 6200 MD Maastricht, The Netherlands

<a.hecq@maastrichtuniversity.nl>

João Victor Issler

Graduate School of Economics - EPGE, Getulio Vargas Foundation, Praia de Botafogo 190s. 1100, Rio de Janeiro, RJ 22250-900, Brazil

<Joao.Issler@fgv.br>

Sean Telg

Vrije Universiteit Amsterdam, Department of Econometrics and Operations Research, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands

<j.m.a.telg@vu.nl>

Abstract

This online appendix provides supplementary material for the paper "Mixed Causal-Noncausal Autoregressions with Exogenous Regressors". The material consists of the proofs of the main results, an auxiliary proposition (with proof), the derivation of the approximate likelihood function and how the MARX model can arise from a transfer function model.

A - Graphs of Data

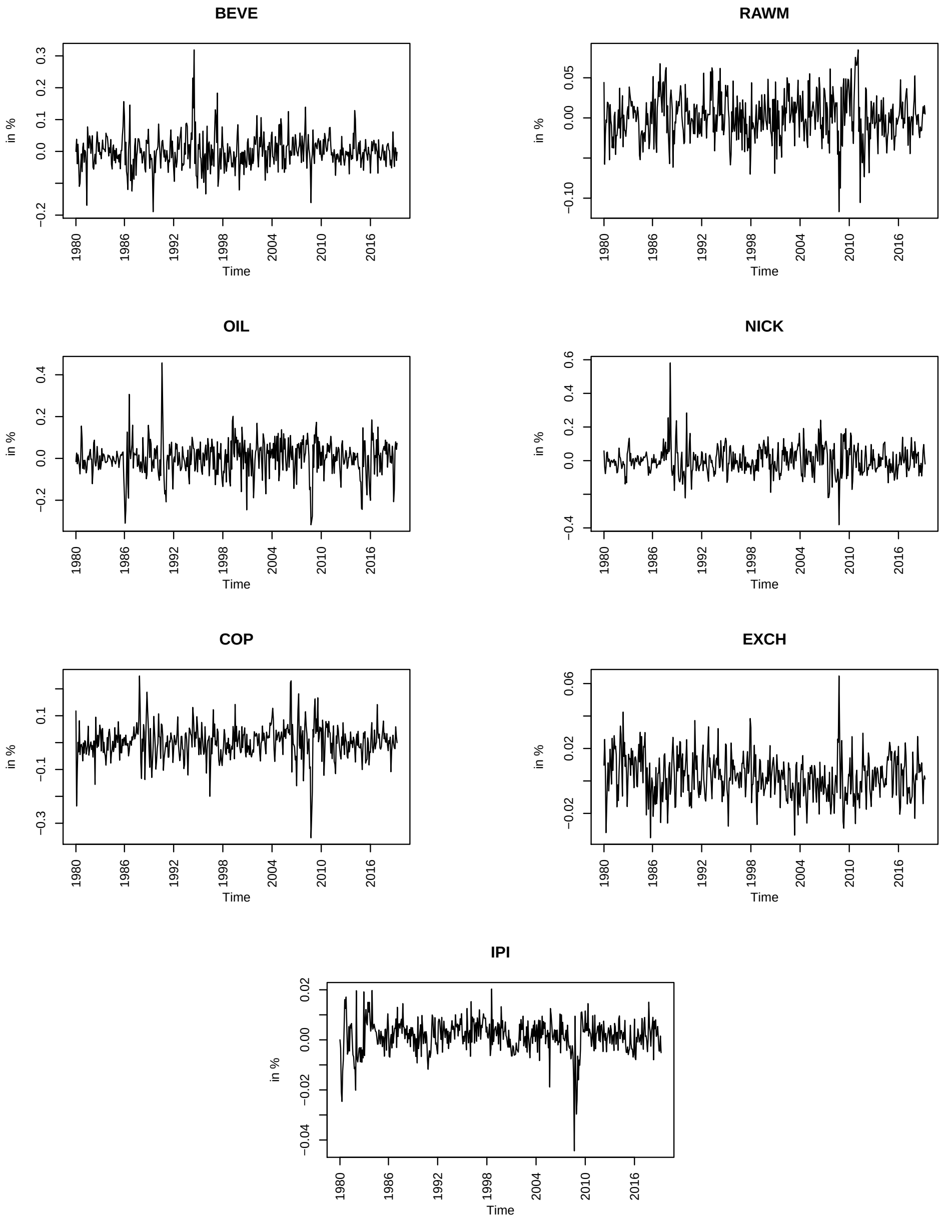

Figure 1: Growth rates of commodity prices, exchange rate and industrial production index

B - Additional Simulation Results

B.1 - Simulation of MARX Processes

A MARX process cannot be simulated directly as simultaneously initial and terminal values are required. If the degree of (at least) one of the polynomials equals 0, the problem is greatly simplified.[16] In the general MARX(r, s, q) setup, filtered values as introduced in Lemma 1 are used to circumvent the problem. Similar to Gouriéroux and Jasiak (2016), we make use of the independence of specific blocks of u, v and y values.

We characterize the simulation procedure as a two-step approach. First, u_t is constructed using the second equality of (3) by taking s terminal values, say $u^*_{T+1}, ..., u^*_{T+s}$ and using simulated sequences for ε_t and $\boldsymbol{X}_t$. Second, y_t is created using the first equality of (3) similar to a conventional causal autoregressive process which requires r starting values for y_t, say $y^*_{-1}, ..., y^*_{-r}$.[17]

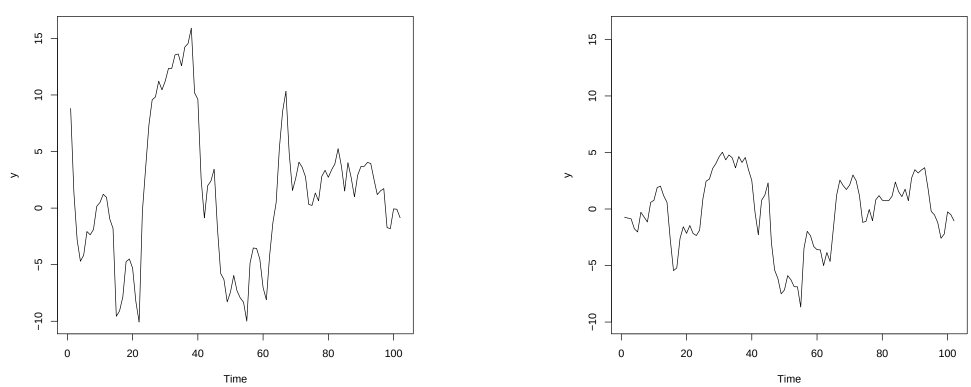

Figure 2: A simulated MARX process (left) and the same process without exogenous variable (right)

Figure 2 shows simulated paths of an MARX(1,1,1) and MAR(1,1) process with $[\phi_1, \varphi_1]' = [0.3, 0.9]'$, $\beta_1 = 0.3$, $x_t \overset{iid}{\sim} t(1, 1)$ and $\varepsilon_t \overset{iid}{\sim} t(3, 1)$ where $t(\nu, \sigma)$ denotes the Student's t distribu-

[16]In that case y_t can easily be simulated directly by generating a sequence of ε_t and choosing starting [terminal] values for y_t and $\boldsymbol{X}_t$ in the causal [noncausal] case.

[17]It is highly advised to allow for a burn-in period in both steps to remove the dependence on the terminal values of u_t and the starting values of y_t. Note that (4) could equivalently be used to simulate an MARX process.

tion with degrees of freedom parameter ν and scale parameter σ. Both processes generally move similarly with the major exception that the MARX process contains more peaks and troughs, which are also more extreme in comparison. This is due to x_t which is standard Cauchy distributed for expository purposes. The MARX specification encompasses shocks that are present because of major changes in explanatory variables at specific points in time.

B.2 - Performance MLE for MARX

To assess the performance of the maximum likelihood estimator, we make use of the DGP introduced at the beginning of Section 4. Four different specifications for x_t are considered: (i) $x_t \overset{iid}{\sim} t(5,1)$, (ii) $x_t \overset{iid}{\sim} \mathcal{N}(0,1)$, (iii) $x_t \overset{iid}{\sim} t(1,1)$ (i.e. Standard Cauchy) and (iv) x_t follows an AR(1) process: $x_t = 0.6x_{1,t-1} + \epsilon_t$ where $\epsilon_t \overset{iid}{\sim} \mathcal{N}(0,5)$. In every replication, processes are simulated and estimated by the ML estimator proposed in Section 3. Table 12 reports the mean and standard deviations of the estimated parameters by MLE over all 10,000 simulations. The scale parameter σ is fixed at 1 for both simulation and estimation.

Results are very similar for different specifications of x_t. The most noticeable difference is that the standard deviations of the parameters are larger for the first two cases especially when T is small. This can be due to the fact that both the $t(5,1)$ and $\mathcal{N}(0,1)$ distribution do not generate large outliers in x_t, making it more difficult to disentangle their contribution to the series from that of lags and leads of y_t. The means of the estimated parameters also lie further away from the true value when compared to the other specifications, but are still close. The most difficult parameter to estimate is ν, which has a large standard deviation for $T = 50$. For larger T, the standard deviations decrease rapidly. In all cases, the estimated mean of all parameters becomes more accurate and standard deviations decline as T grows large. Table 13 shows results for the same simulation study in the infinite variance case, i.e., $\varepsilon_t \overset{iid}{\sim} t(2,1)$. Similar to Hecq et al. (2016) for the MAR model, results suggests the fatter the tails of the error distribution, the more accurate the estimation for all parameters of the MARX model.

		Specification for x_t							
		$x_t \overset{iid}{\sim} t(5,1)$		$x_t \overset{iid}{\sim} \mathcal{N}(0,1)$		$x_t \overset{iid}{\sim} t(1,1)$		$x_t \sim$ AR(1)	
T	Parameter	Mean	Std. dev	Mean	Std. dev	Mean	Std. dev	Mean	Std. dev
50	ϕ_1	0.307	0.167	0.316	0.174	0.298	0.079	0.293	0.079
	φ_1	0.468	0.162	0.462	0.170	0.493	0.071	0.491	0.066
	β_1	0.303	0.152	0.305	0.194	0.299	0.039	0.304	0.036
	ν	5.140	8.562	5.089	8.203	5.050	8.295	5.416	9.906
100	ϕ_1	0.301	0.104	0.302	0.109	0.299	0.039	0.295	0.052
	φ_1	0.488	0.099	0.486	0.104	0.498	0.035	0.497	0.043
	β_1	0.302	0.102	0.302	0.131	0.300	0.018	0.302	0.024
	ν	3.487	1.899	3.487	2.216	3.474	1.754	3.544	2.534
500	ϕ_1	0.299	0.037	0.299	0.038	0.300	0.010	0.299	0.022
	φ_1	0.498	0.037	0.499	0.035	0.500	0.009	0.499	0.018
	β_1	0.301	0.043	0.300	0.055	0.300	0.004	0.300	0.011
	ν	3.057	0.366	3.061	0.367	3.053	0.423	3.060	0.364
1000	ϕ_1	0.300	0.025	0.300	0.026	0.300	0.006	0.300	0.015
	φ_1	0.499	0.023	0.499	0.023	0.500	0.004	0.500	0.013
	β_1	0.300	0.030	0.300	0.039	0.300	0.002	0.300	0.007
	ν	3.031	0.255	3.029	0.254	3.023	0.322	3.026	0.254

Table 12: Finite sample properties of the ML estimator for an MARX(1,1,1) with $\varepsilon_t \overset{iid}{\sim} t(3,1)$

		Specification for $x_{1,t}$							
		$x_{1,t} \overset{iid}{\sim} t(5)$		$x_{1,t} \overset{iid}{\sim} N(0,1)$		$x_{1,t} \overset{iid}{\sim} C(0,1)$		$x_{1,t} \sim$ AR(1)	
T	Parameter	Mean	Std. dev	Mean	Std. dev	Mean	Std. dev	Mean	Std. dev
50	ϕ_1	0.299	0.120	0.303	0.139	0.299	0.081	0.292	0.071
	φ_1	0.481	0.117	0.480	0.138	0.494	0.072	0.494	0.060
	β_1	0.308	0.167	0.313	0.205	0.299	0.043	0.303	0.037
	ν	2.615	4.446	2.594	2.932	2.490	2.470	2.459	3.196
100	ϕ_1	0.298	0.075	0.301	0.071	0.299	0.035	0.295	0.048
	φ_1	0.498	0.067	0.492	0.065	0.500	0.031	0.495	0.041
	β_1	0.300	0.110	0.301	0.139	0.299	0.021	0.302	0.026
	ν	2.129	0.565	2.150	0.533	2.132	0.535	2.173	0.600
500	ϕ_1	0.300	0.024	0.299	0.025	0.300	0.012	0.299	0.018
	φ_1	0.500	0.021	0.500	0.021	0.500	0.007	0.500	0.016
	β_1	0.301	0.046	0.298	0.058	0.300	0.004	0.300	0.010
	ν	2.034	0.187	2.027	0.185	2.012	0.221	2.026	0.190
1000	ϕ_1	0.300	0.015	0.299	0.015	0.300	0.005	0.300	0.012
	φ_1	0.500	0.014	0.500	0.013	0.500	0.005	0.500	0.010
	β	0.300	0.032	0.299	0.041	0.300	0.002	0.300	0.007
	ν	2.007	0.128	2.016	0.136	2.000	0.153	2.014	0.131

Table 13: Finite sample properties of the ML estimator for an MARX(1,1,1) with $\varepsilon_t \overset{iid}{\sim} t(2,1)$

C - Proofs

C.1 Proof of Lemma 2

Define $e_t \equiv \frac{f'_\sigma(\varepsilon_t;\boldsymbol{\lambda})}{f_\sigma(\varepsilon_t;\boldsymbol{\lambda})} \equiv \frac{f'(\varepsilon_t/\sigma;\boldsymbol{\lambda})}{\sigma f(\varepsilon_t/\sigma;\boldsymbol{\lambda})}$, $\tilde{\mathcal{J}} \equiv \sigma^{-2}\mathcal{J}$, $\tilde{\mathcal{I}} \equiv \sigma^{-2}\mathcal{I}$ and $n \equiv (T-p)$. Furthermore, let $x \equiv \varepsilon_t/\sigma$, then we have that

$$\mathbb{E}(e_t^2) = \mathbb{E}\left[\left(\frac{f'_\sigma(\varepsilon_t;\boldsymbol{\lambda})}{f_\sigma(\varepsilon_t;\boldsymbol{\lambda})}\right)^2\right] = \int \left(\frac{f'_\sigma(\varepsilon_t;\boldsymbol{\lambda})}{f_\sigma(\varepsilon_t;\boldsymbol{\lambda})}\right)^2 f_\sigma(\varepsilon_t;\boldsymbol{\lambda})d\varepsilon_t = \sigma^{-2}\int \frac{(f'(x;\boldsymbol{\lambda}))^2}{f(x;\boldsymbol{\lambda})}dx = \tilde{\mathcal{J}},$$

where we used the definitions of the density and $\mathcal{J}$. Also we have that

$$\mathbb{E}(e_t) = \mathbb{E}\left(\frac{f'_\sigma(\varepsilon_t;\boldsymbol{\lambda})}{f_\sigma(\varepsilon_t;\boldsymbol{\lambda})}\right) = \int f'_\sigma(\varepsilon_t;\boldsymbol{\lambda})d\varepsilon_t = \sigma^{-1}f(x)|_{-\infty}^{\infty} = 0,$$

which follows by the definition of the density and assumption (A3) in Breidt et al. (1991). To simplify future computations, we begin by noting that

$$\mathbb{E}(z_s e_t) = \begin{cases} 0, & \text{if } s \neq t, \\ -1, & \text{if } s = t, \end{cases} \tag{16}$$

which follows from the assumptions on the density and strixt exogeneity between all exogeneous regressors and the error term. For $i = 1, ..., r$, we can show that $\mathbb{E}\left(\frac{\partial g_t(\boldsymbol{\theta_0})}{\partial \phi_i}\right) = \mathbb{E}(-e_t v_{t-i}) = -\mathbb{E}\left(e_t \sum_{j=0}^{\infty} \alpha_j z_{t-i-j}\right) = 0$. Hence, we note that $\boldsymbol{V}_{t-1}$ and e_t are still independent as in Lanne and Saikkonen (2011). Consequently, we still find

$$\text{Cov}\left(\frac{\partial g_t(\boldsymbol{\theta_0})}{\partial \boldsymbol{\phi}}\right) = \text{Cov}(-\boldsymbol{V}_{t-1}e_t) = \mathbb{E}(e_t^2)\mathbb{E}(\boldsymbol{V}_{t-1}\boldsymbol{V}'_{t-1}) = \tilde{\mathcal{J}}\boldsymbol{\Gamma}_V,$$

where $\boldsymbol{\Gamma}_V$ denotes the autocovariance matrix of the vector $\boldsymbol{V}_{t-1}$. Because $\boldsymbol{V}_{t-1}e_t$ is uncorrelated, we have $\lim_{n\to\infty}\frac{1}{n}\text{Cov}\left(\sum_{t=r+1}^{T-s}\frac{\partial g_t(\boldsymbol{\theta_0})}{\partial \boldsymbol{\phi}}\right) = \tilde{\mathcal{J}}\boldsymbol{\Gamma}_V$. Symmetrically, by using similar arguments, we can show for $i = 1, ..., s$ that $\mathbb{E}\left(\frac{\partial g_t(\boldsymbol{\theta_0})}{\partial \varphi_i}\right) = \mathbb{E}(-e_t u_{t+i}) = -\mathbb{E}\left(e_t \sum_{j=0}^{\infty}\delta_j z_{t+i+j}\right) = 0$. That is, the independence of e_t and $\boldsymbol{U}_{t+1}$ is preserved through strict exogeneity. Letting $\boldsymbol{\Gamma}_U$ be the autocovariance matrix of $\boldsymbol{U}_{t+1}$, we find that

$$\text{Cov}\left(\frac{\partial g_t(\boldsymbol{\theta_0})}{\partial \boldsymbol{\varphi}}\right) = \text{Cov}(-\boldsymbol{U}_{t+1}e_t) = \mathbb{E}(e_t^2)\mathbb{E}(\boldsymbol{U}_{t+1}\boldsymbol{U}'_{t+1}) = \tilde{\mathcal{J}}\boldsymbol{\Gamma}_U.$$

Because $\boldsymbol{U}_{t+1}e_t$ is uncorrelated, we have $\lim_{n\to\infty}\frac{1}{n}\text{Cov}\left(\sum_{t=r+1}^{T-s}\frac{\partial g_t(\boldsymbol{\theta_0})}{\partial \boldsymbol{\varphi}}\right) = \tilde{\mathcal{J}}\boldsymbol{\Gamma}_U$. Lastly, we can apply the same logic for the parameter vector $\boldsymbol{\beta}$. Since for $i = 1, ..., q$, we have that $\mathbb{E}\left(\frac{\partial g_t(\boldsymbol{\theta_0})}{\partial \beta_i}\right) = 0$ by the strict exogeneity of $x_{i,t}$ and ε_t. If we denote by $\boldsymbol{\Gamma}_X$, the autocovariance matrix of $\boldsymbol{X}_t$, it follows that

$$\text{Cov}\left(\frac{\partial g_t(\boldsymbol{\theta_0})}{\partial \boldsymbol{\beta}}\right) = \text{Cov}(-\boldsymbol{X}_t e_t) = \mathbb{E}(e_t^2)\mathbb{E}(\boldsymbol{X}_t\boldsymbol{X}'_t) = \tilde{\mathcal{J}}\boldsymbol{\Gamma}_X.$$

Because $\boldsymbol{X}_t e_t$ is uncorrelated, we have $\lim_{n\to\infty}\frac{1}{n}\text{Cov}\left(\sum_{t=r+1}^{T-s}\frac{\partial g_t(\boldsymbol{\theta_0})}{\partial \boldsymbol{\beta}}\right) = \tilde{\mathcal{J}}\boldsymbol{\Gamma}_X$. We now characterize the covariances of the partials. To that end, we first notice that

$$\text{Cov}(z_{t-i}e_t, z_{k-j}e_k) = \begin{cases} \mathcal{I} + \tilde{\mathcal{J}}\sum_{m=1}^{q}\beta_m^2\sigma_{x_m}^2 & \text{if } t = k, i = j = 0, \\ \mathcal{J} + \tilde{\mathcal{J}}\sum_{m=1}^{q}\beta_m^2\sigma_{x_m}^2 & \text{if } t = k, i = j \neq 0, \\ 1 & \text{if } t \neq k, i = t-k, j = k-t, \\ 0 & \text{otherwise.} \end{cases} \tag{17}$$

Hence, using (16)-(17), we find that

$$\text{Cov}\left(\frac{\partial g_t(\boldsymbol{\theta_0})}{\partial \phi_i}, \frac{\partial g_k(\boldsymbol{\theta_0})}{\partial \phi_j}\right) = \begin{cases} \gamma_V(i-j)\tilde{\mathcal{J}}, & \text{if } t = k, 1 \leq i \leq j \leq r, \\ 0, & \text{otherwise.} \end{cases}$$

$$\text{Cov}\left(\frac{\partial g_t(\boldsymbol{\theta_0})}{\partial \varphi_i}, \frac{\partial g_k(\boldsymbol{\theta_0})}{\partial \varphi_j}\right) = \begin{cases} \gamma_U(i-j)\tilde{\mathcal{J}}, & \text{if } t=k, 1 \le i \le j \le s, \\ 0, & \text{otherwise.} \end{cases}$$

$$\text{Cov}\left(\frac{\partial g_t(\boldsymbol{\theta_0})}{\partial \beta_i}, \frac{\partial g_k(\boldsymbol{\theta_0})}{\partial \beta_j}\right) = \begin{cases} \sigma_{x_i}^2\tilde{\mathcal{J}}, & \text{if } t=k, i=j \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

Define $Q_m(i,j,a) \equiv \sum_{b=0}^{\infty} \delta_b \gamma_{x_m}(i+j+a+b)$. For the covariance matrix between $\partial g_t(\boldsymbol{\theta_0})/\partial \boldsymbol{\phi}$ and $\partial g_t(\boldsymbol{\theta_0})/\partial \boldsymbol{\varphi}$, first consider for $1 \le i \le r, 1 \le j \le s$:

$$\begin{aligned} \text{Cov}\left(\frac{\partial g_t(\boldsymbol{\theta_0})}{\partial \phi_i}, \frac{\partial g_k(\boldsymbol{\theta_0})}{\partial \varphi_j}\right) &= \text{Cov}\left(\sum_{a=0}^{\infty} \alpha_a z_{t-i-a} e_t, \sum_{b=0}^{\infty} \delta_b z_{k+j+b} e_k\right) \\ &= \sum_{a=0}^{\infty}\sum_{b=0}^{\infty} \alpha_a \delta_b \text{Cov}\left(z_{t-i-a}e_t, z_{k+j+b}e_k\right) \end{aligned}$$

$$= \begin{cases} \alpha_{t-k-i}\delta_{t-k-j}, & \text{for } t > k, \\ \tilde{\mathcal{J}} \sum_{a=0}^{\infty} \alpha_a \sum_{m=1}^{q} \beta_m^2 Q_m(i,j,a) & \text{for } t = k, \\ 0 & \text{for } t < k. \end{cases}$$

The element (i,j) of the matrix $\frac{1}{n}\text{Cov}(\partial L_T(\boldsymbol{\theta_0})/\partial \boldsymbol{\phi}, \partial L_T(\boldsymbol{\theta_0})/\partial \boldsymbol{\varphi})$ is

$$\begin{aligned} & n\text{Cov}\left(\frac{1}{n}\sum_{t=r+1}^{T-s} \frac{\partial g_t(\boldsymbol{\theta_0})}{\partial \phi_i}, \frac{1}{n}\sum_{k=r+1}^{T-s} \frac{\partial g_k(\boldsymbol{\theta_0})}{\partial \varphi_j}\right) \\ =& \frac{1}{n}\sum_{t=r+1}^{T-s}\sum_{k=r+1}^{T-s} \text{Cov}\left(\frac{\partial g_t(\boldsymbol{\theta_0})}{\partial \phi_i}, \frac{\partial g_k(\boldsymbol{\theta_0})}{\partial \varphi_j}\right) \\ =& \frac{1}{n}\sum_{t=r+1}^{T-s}\sum_{k=r+1}^{T-s} \text{Cov}(-v_{t-i}e_t, -u_{k+j}e_k) \\ =& \frac{1}{n}\sum_{t=r+1}^{T-s}\sum_{k=r+1}^{T-s} \left(\mathbb{1}_{\{t>k\}}\alpha_{t-k-i}\delta_{t-k-j} + \mathbb{1}_{\{t=k\}}\tilde{\mathcal{J}}\sum_{a=0}^{\infty}\alpha_a\sum_{m=1}^{q}\beta_m^2 Q_m(i,j,a)\right) \end{aligned}$$

$$= \frac{1}{n}\left(\sum_{k=r+1}^{T-s-1}\sum_{t=0}^{T-s-k-i}\alpha_t\delta_{t+i-j}\right) + \tilde{\mathcal{J}}\sum_{a=0}^{\infty}\alpha_a\sum_{m=1}^{q}\beta_m^2 Q_m(i,j,a)$$

$$\to \sum_{t=0}^{\infty}\alpha_t\delta_{t+i-j} + \tilde{\mathcal{J}}\sum_{a=0}^{\infty}\alpha_a\sum_{m=1}^{q}\beta_m^2 Q_m(i,j,a),$$

where convergence of the first term follows from the geometric decay of the sequences $\{\alpha_t\}$ and $\{\delta_t\}$. Note that $\delta_{t+i-j} = 0$ for $t+i-j < 0$. The equalities follow from results presented earlier, the change of summands follows from imposing $t > k$.

Next, we consider the covariance between the partial derivatives of the log-likelihood with respect to the causal autoregressive parameters $\boldsymbol{\phi}$ and the parameter vector of the exogenous variables $\boldsymbol{\beta}$. That is,

$$\begin{aligned}
\operatorname{Cov}\left(\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial\phi_i}, \frac{\partial g_k(\boldsymbol{\theta}_0)}{\partial\beta_j}\right) &= \operatorname{Cov}(-v_{t-i}e_t, -x_{j,k}e_k) \\
&= \mathbb{E}\left(\sum_{a=0}^{\infty}\alpha_a\varepsilon_{t-i-a}x_{j,k}e_ke_t\right) \\
&+ \mathbb{E}\left(\sum_{a=0}^{\infty}\alpha_a\sum_{m=1}^{q}\beta_m x_{m,t-i-a}x_{j,k}e_ke_t\right) \\
&= \begin{cases} \beta_j\tilde{\mathcal{J}}\sum_{a=0}^{\infty}\alpha_a\gamma_{x_j}(i+a) & \text{for } t = k, 1 \le i \le r, \\ 0 & \text{for } t \ne k, 1 \le i \le r. \end{cases}
\end{aligned}$$

Note that this outcome is independent of time t. Symmetrically, we can compute the covariance between the partial derivatives of the log-likelihood with respect to the noncausal autoregressive parameters $\boldsymbol{\varphi}$ and the parameter vector of the exogenous variables $\boldsymbol{\beta}$:

$$\begin{aligned}
\operatorname{Cov}\left(\frac{\partial g_t(\boldsymbol{\theta}_0)}{\partial\varphi_i}, \frac{\partial g_k(\boldsymbol{\theta}_0)}{\partial\beta_j}\right) &= \operatorname{Cov}(-u_{t+i}e_t, -x_{j,k}e_k) \\
&= \mathbb{E}\left(\sum_{b=0}^{\infty}\delta_b\varepsilon_{t+i+b}x_{j,k}e_ke_t\right)
\end{aligned}$$

$$+\mathbb{E}\left(\sum_{b=0}^{\infty}\delta_b\sum_{m=1}^{q}\beta_m x_{m,t+i+b}x_{j,k}e_k e_t\right)$$
$$=\begin{cases}\beta_j\tilde{\mathcal{J}}\sum_{b=0}^{\infty}\delta_b\gamma_{x_j}(i+b) & \text{for } t=k, 1\leq i\leq s,\\ 0 & \text{for } t\neq k, 1\leq i\leq s.\end{cases}$$

The proof of asymptotic normality is similar to Breidt et al. (1991) and Lanne and Saikkonen (2011). Define $\boldsymbol{M}=\text{diag}(\boldsymbol{\Sigma},\boldsymbol{\Omega})$, $\boldsymbol{W}_t=\frac{\partial g_t(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}}$ and note that $n\equiv(T-p)$. By the Cramér-Wold theorem, it suffices to show that for any vector $\boldsymbol{a}$ of appropriate size,

$$\frac{1}{\sqrt{n}}\sum_{t=r+1}^{T-s}\boldsymbol{a}'\boldsymbol{W}_t\xrightarrow{d}\mathcal{N}(\boldsymbol{0},\boldsymbol{a}'\boldsymbol{M}\boldsymbol{a}). \tag{18}$$

Define the sequence of $(p+q+d+1)$ dimensional random vectors $\{\boldsymbol{W}_{tm}, t\in\mathbb{Z}\}$ to be the partials defined in Appendix A, where v_t, u_t and all $x_{i,t}$ for $i=1,...,q$ are replaced by their representation $u_t=\sum_{j=0}^{\infty}\delta_j z_{t+j}$ and $v_t=\sum_{j=0}^{\infty}\alpha_j z_{t-j}$ and assumption (II) with the sums truncated at a large positive integer m, i.e., $v_t^{(m)}=\sum_{j=0}^{m}\alpha_j z_{t-j}$, $u_t^{(m)}=\sum_{j=0}^{m}\delta_j z_{t+j}$ and $x_{i,t}^{(m)}=c_i+\sum_{j=-m}^{m}\rho_{i,j}\eta_{i,t-j}$.

It can be verified that $\mathbb{E}(\boldsymbol{W}_t)=\boldsymbol{0}$ and $\boldsymbol{\gamma}_{\boldsymbol{W}_t}(0)+2\sum_{j=1}^{\infty}\boldsymbol{\gamma}_{\boldsymbol{W}_t}(j)\neq\boldsymbol{0}$. This result also holds for $\boldsymbol{W}_t$ replaced by $\boldsymbol{W}_{tm}$. Let $\boldsymbol{M}_m$ be the matrix corresponding to $\boldsymbol{M}$, obtained by truncating u_t, v_t and $\boldsymbol{X}_t$. Then the stationary sequence $\{\boldsymbol{W}_{tm}, t\in\mathbb{Z}\}$ is $\max\{m+p, 2m\}$ dependent.[18] Now that we verified the conditions, we can apply Theorem 6.4.2 in Brockwell and Davis (1991) to obtain

$$\frac{1}{\sqrt{n}}\sum_{t=r+1}^{T-s}\boldsymbol{a}'\boldsymbol{W}_{tm}\xrightarrow{d}\mathcal{N}(\boldsymbol{0},\boldsymbol{a}'\boldsymbol{M}_m\boldsymbol{a}).$$

[18] The $m+p$ follows from writing u_t and v_t in their truncated representation, $2m$ follows from the processes in $\boldsymbol{X}_t$ which have a two-sided MA representation truncated by m at both sides.

Now, it follows that for $m \to \infty$, $\boldsymbol{W}_{tm} \to \boldsymbol{W}_t$ (by definition) and thus $\boldsymbol{M}_m \to \boldsymbol{M}$. Because

$$\lim_{m\to\infty}\lim_{n\to\infty} \mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=r+1}^{T-s}\left(\boldsymbol{a}'\boldsymbol{W}_{tm} - \boldsymbol{a}'\boldsymbol{W}_t\right)\right) = 0,$$

the convergence in (18) is immediate from Proposition 6.3.9 in Brockwell and Davis (1991). The positive definiteness of $\boldsymbol{\Sigma}$ can be established similar to the proof in Breidt et al. (1991). In the MARX case, the block matrix $\boldsymbol{\Sigma}$ is given as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{13} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \boldsymbol{\Sigma}_{23} \\ \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{32} & \boldsymbol{\Sigma}_{33} \end{bmatrix}.$$

In a first step, let us focus on the submatrix $\tilde{\boldsymbol{\Sigma}}_1$ given by

$$\tilde{\boldsymbol{\Sigma}}_1 = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \text{ and partition it as } \tilde{\boldsymbol{\Sigma}}_1 = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}' & \boldsymbol{C} \end{bmatrix},$$

where $\boldsymbol{A}$ is $r \times r$, $\boldsymbol{C}$ is $s \times s$ and $\boldsymbol{B}$ is $r \times s$. Consider the vectors $\boldsymbol{P} = [P_1, ..., P_r]'$ and $\boldsymbol{S} = [S_1, ..., S_s]'$ defined by $\boldsymbol{P}_t = \sum_{a=0}^{\infty} \alpha_a z_{-t-a} e_0$ for $t = 1, ..., r$ and $\boldsymbol{S}_t = \sum_{b=0}^{\infty} \delta_b z_{t-b} e_0$ for $t = 1, ..., s$. It can easily be verified that the covariance matrices of $\boldsymbol{P}$ and $\boldsymbol{S}$, denoted $\boldsymbol{\Sigma}_{PP}$ and $\boldsymbol{\Sigma}_{SS}$, are equal to $\boldsymbol{A}$ and $\boldsymbol{C}$. From (17), it follows that

$$\begin{aligned}
\mathrm{Cov}(P_i, S_j) &= \mathrm{Cov}\left(\sum_{a=0}^{\infty} \alpha_a z_{-i-a} e_0, \sum_{b=0}^{\infty} \delta_b z_{j-b} e_0\right) \\
&= \sum_{a=0}^{\infty}\sum_{b=0}^{\infty} \alpha_a \delta_b \mathbb{E}(z_{-i-a} z_{j-b} e_0^2) \\
&= \sum_{a=0}^{\infty} \alpha_a \delta_{b+i-j} (\mathcal{J} + \tilde{\mathcal{J}} \sum_{m=1}^{q} \beta_m^2 \sigma_m^2).
\end{aligned}$$

We know that $\mathcal{J} > 1$ by condition (A5) of Andrews et al. (2006). We have that $\tilde{\mathcal{J}} \sum_{m=1}^{q} \beta_m^2 \sigma_m^2 = \mathcal{J} \left(\frac{\sum_{m=1}^{q} \beta_m^2 \sigma_m^2}{\sigma^2} \right) > 0$, which in turn implies that $(\mathcal{J} + \tilde{\mathcal{J}} \sum_{m=1}^{q} \beta_m^2 \sigma_m^2) > 1$. Similar to Breidt et al. (1991), we exploit that the matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ are positive definite since there is no linear dependence within the vectors $\boldsymbol{P}$ and $\boldsymbol{S}$. We proceed by proving the positive definiteness of $\tilde{\boldsymbol{\Sigma}}$ by showing that the Schur Complement of the block $\boldsymbol{A}$ of the matrix $\tilde{\boldsymbol{\Sigma}}$ given as $\boldsymbol{C} - \boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}$ is positive definite. We know that the covariance matrix of $\boldsymbol{S} - \boldsymbol{\Sigma}_{\boldsymbol{SP}} \boldsymbol{\Sigma}_{\boldsymbol{PP}}{}^{-1}$, i.e. $\boldsymbol{C} - (\mathcal{J} + \tilde{\mathcal{J}} \sum_{m=1}^{q} \beta_m^2 \sigma_m^2)\boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B}$, is positive semidefinite and hence for a nonzero vector $\boldsymbol{c} \in \mathbb{R}^s$ with $\boldsymbol{Bc} \neq \boldsymbol{0}$, we have that

$$\boldsymbol{c}'(\boldsymbol{C} - \boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B})\boldsymbol{c} > \boldsymbol{c}'(\boldsymbol{C} - (\mathcal{J} + \tilde{\mathcal{J}} \sum_{m=1}^{q} \beta_m^2 \sigma_m^2)\boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B})\boldsymbol{c} \geq 0.$$

Alternatively, if $\boldsymbol{Bc} = \boldsymbol{0}$, we have that

$$\boldsymbol{c}'(\boldsymbol{C} - \boldsymbol{B}'\boldsymbol{A}^{-1}\boldsymbol{B})\boldsymbol{c} = \boldsymbol{c}'\boldsymbol{C}\boldsymbol{c} > 0,$$

by the positive definiteness of $\boldsymbol{C}$. Hence, now that we established positive definiteness of $\tilde{\boldsymbol{\Sigma}}_1$, we can repartition the matrix $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \tilde{\boldsymbol{\Sigma}}_1 & \tilde{\boldsymbol{\Sigma}}_2 \\ \tilde{\boldsymbol{\Sigma}}_2' & \tilde{\boldsymbol{\Sigma}}_3 \end{bmatrix},$$

where $\tilde{\boldsymbol{\Sigma}}_1$ is $(r+s) \times (r+s)$, $\tilde{\boldsymbol{\Sigma}}_2 = [\boldsymbol{\Sigma}_{12}, \boldsymbol{\Sigma}_{23}]'$ is $(r+s) \times q$ and $\tilde{\boldsymbol{\Sigma}}_3 = \boldsymbol{\Sigma}_{33}$ is $q \times q$. Since $\boldsymbol{\Sigma}_{33} = \mathrm{diag}(\sigma_1^2, ..., \sigma_m^2)$, we have that for a nonzero vector $\boldsymbol{c} \in \mathbb{R}^q$, $\boldsymbol{c}'\boldsymbol{\Sigma}_{33}\boldsymbol{c} = c_1^2\sigma_1^2 + ... + c_m^2\sigma_m^2 > 0$. Hence, as we know that $\tilde{\boldsymbol{\Sigma}}_1$ and $\tilde{\boldsymbol{\Sigma}}_3$ are positive definite, it is sufficient to show that the Schur complement of the block $\tilde{\boldsymbol{\Sigma}}_1$ of the matrix $\boldsymbol{\Sigma}$ is positive definite, which can be established analogous to the case above. The positive definiteness of $\boldsymbol{\Omega}$ follows from condition (A6) in Andrews et al. (2006).

C.2 Proof of Theorem 1

We first present the second partial derivatives of the function $g_t(\boldsymbol{\theta})$. Set $h(x;\boldsymbol{\lambda}) = f'(x;\boldsymbol{\lambda})/f(x;\boldsymbol{\lambda})$, such that

$$h'(x;\boldsymbol{\lambda}) = \frac{f''(x;\boldsymbol{\lambda})}{f(x;\boldsymbol{\lambda})} - \left(\frac{f'(x;\boldsymbol{\lambda})}{f(x;\boldsymbol{\lambda})}\right)^2,$$

which can easily be verified using the quotient rule. Let $\boldsymbol{Y}_t$ be the $(r \times s)$ matrix with elements y_{t-i+j}. Write $\tilde{v}_t = v_t(\boldsymbol{\varphi})$ and $\tilde{u}_t = u_t(\boldsymbol{\phi})$ and thus $\tilde{\boldsymbol{V}}_{t-1} = [\tilde{v}_{t-1}, ..., \tilde{v}_{t-r}]'$ and $\tilde{\boldsymbol{U}}_{t+1} = [\tilde{u}_{t+1}, ..., \tilde{u}_{t+s}]'$ to simplify notation. Similarly, $\tilde{\varepsilon}_t = \tilde{v}_t - \phi_1\tilde{v}_{t-1} - ... - \phi_r\tilde{v}_{t-r} = \tilde{u}_t - \varphi_1\tilde{u}_{t+1} - ... - \varphi_s\tilde{u}_{t+s}$ denotes ε_t evaluated at an arbitrary point in the permissible parameter space, not the true one. Then, the second partial derivatives in the MARX case can be obtained through differentiation, similar to Lanne and Saikkonen (2011) and Breidt et al. (1991):

$$\begin{aligned}
\partial^2 g_t(\boldsymbol{\theta})/\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}' &= \sigma^{-2}h'(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})\tilde{\boldsymbol{V}}_{t-1}\tilde{\boldsymbol{V}}'_{t-1},\\
\partial^2 g_t(\boldsymbol{\theta})/\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}' &= \sigma^{-2}h'(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})\tilde{\boldsymbol{U}}_{t+1}\tilde{\boldsymbol{U}}'_{t+1},\\
\partial^2 g_t(\boldsymbol{\theta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' &= \sigma^{-2}h'(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})\boldsymbol{X}_t\boldsymbol{X}'_t,\\
\partial^2 g_t(\boldsymbol{\theta})/\partial\sigma^2 &= 2\sigma^{-3}h(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})\tilde{\varepsilon}_t + \sigma^{-4}h'(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})\tilde{\varepsilon}_t^2 + \sigma^{-2},\\
\partial^2 g_t(\boldsymbol{\theta})/\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}' &= \frac{1}{f(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})}\frac{\partial^2 f(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'}\\
&\quad - \frac{1}{f^2(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})}\left(\frac{\partial f(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\right)\left(\frac{\partial f(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\right)',\\
\partial^2 g_t(\boldsymbol{\theta})/\partial\boldsymbol{\phi}\partial\boldsymbol{\varphi}' &= \sigma^{-2}h'(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})\tilde{\boldsymbol{V}}_{t-1}\tilde{\boldsymbol{U}}'_{t+1} + \sigma^{-1}h(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})\boldsymbol{Y}_t,\\
\partial^2 g_t(\boldsymbol{\theta})/\partial\boldsymbol{\phi}\partial\boldsymbol{\beta}' &= \sigma^{-2}h'(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})\tilde{\boldsymbol{V}}_{t-1}\boldsymbol{X}'_t,\\
\partial^2 g_t(\boldsymbol{\theta})/\partial\boldsymbol{\phi}\partial\sigma &= \sigma^{-3}h'(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})\tilde{\varepsilon}_t\tilde{\boldsymbol{V}}_{t-1} + \sigma^{-2}h(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})\tilde{\boldsymbol{V}}_{t-1},\\
\partial^2 g_t(\boldsymbol{\theta})/\partial\boldsymbol{\phi}\partial\boldsymbol{\lambda}' &= -\sigma^{-1}\tilde{\boldsymbol{V}}_{t-1}\partial h(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}',\\
\partial^2 g_t(\boldsymbol{\theta})/\partial\boldsymbol{\varphi}\partial\boldsymbol{\beta}' &= \sigma^{-2}h'(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})\tilde{\boldsymbol{U}}_{t+1}\boldsymbol{X}'_t,
\end{aligned}$$

$$
\begin{aligned}
\partial^2 g_t(\boldsymbol{\theta})/\partial\boldsymbol{\varphi}\partial\sigma &= \sigma^{-3}h'(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})\tilde{\varepsilon}_t\tilde{\boldsymbol{U}}_{t+1} + \sigma^{-2}h(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})\tilde{\boldsymbol{U}}_{t+1},\\
\partial^2 g_t(\boldsymbol{\theta})/\partial\boldsymbol{\varphi}\partial\boldsymbol{\lambda}' &= -\sigma^{-1}\tilde{\boldsymbol{U}}_{t+1}\partial h(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}',\\
\partial^2 g_t(\boldsymbol{\theta})/\partial\boldsymbol{\beta}\partial\sigma &= \sigma^{-3}h'(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})\tilde{\varepsilon}_t\boldsymbol{X}_t + \sigma^{-2}h(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})\boldsymbol{X}_t,\\
\partial^2 g_t(\boldsymbol{\theta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\lambda}' &= -\sigma^{-1}\boldsymbol{X}_t\partial h(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}',\\
\partial^2 g_t(\boldsymbol{\theta})/\partial\sigma\partial\boldsymbol{\lambda}' &= -\sigma^{-2}\tilde{\varepsilon}_t\partial h(\sigma^{-1}\tilde{\varepsilon}_t;\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}'.
\end{aligned}
$$

It can be verified that $\mathbb{E}(\partial^2 g_t(\boldsymbol{\theta_0})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}') = -\mathrm{diag}(\boldsymbol{\Sigma},\boldsymbol{\Omega})$. The proof for consistency is exactly the same as in Lanne and Saikkonen (2011). That is, similar to Andrews et al. (2006), we use the Taylor expansion

$$
\begin{aligned}
\sum_{t=r+1}^{T-s}\left[g_t(\boldsymbol{\theta_0}+T^{-1/2}\boldsymbol{c}) - g_t(\boldsymbol{\theta_0})\right] =& \frac{1}{\sqrt{T}}\sum_{t=r+1}^{T-s}\boldsymbol{c}'\frac{\partial g_t(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}} + \frac{1}{2T}\sum_{t=r+1}^{T-s}\boldsymbol{c}'\frac{\partial^2 g_t(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\boldsymbol{c}\\
&+ \frac{1}{2T}\sum_{t=r+1}^{T-s}\boldsymbol{c}'\left(\frac{\partial^2 g_t(\boldsymbol{\theta_T^*}(\boldsymbol{c}))}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \frac{\partial^2 g_t(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)\boldsymbol{c},
\end{aligned}
$$

where $\boldsymbol{c}\in\mathbb{R}^{p+q+1+d}$ and the argument $\boldsymbol{\theta_T^*}(\boldsymbol{c})$ in the matrix of second partial derivatives means that each row is evaluated at an intermediate point lying between the true parameter value $\boldsymbol{\theta_0}$ and $T^{-1/2}\boldsymbol{c}$. If $\|\cdot\|$ denotes the Euclidian norm we have $\sup_{\boldsymbol{c}\in\boldsymbol{C}}\left\|\boldsymbol{\theta_T^*}(\boldsymbol{c}) - \boldsymbol{\theta_0}\right\| \to 0$ for any compact set $\boldsymbol{C}\subset\mathbb{R}^{p+q+1+d}$. Using the dominance conditions (A7) in Davis et al. (1992), arguments similar to Breidt et al. (1991, p. 186-190) and assumption (I) in this paper, it can be shown that a uniform law of large numbers for stationary ergodic processes applies to $\partial^2 g_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ over any small enough compact neighborhood $\boldsymbol{\theta_0}$. We can conclude that

$$
\frac{1}{T}\sum_{t=r+1}^{T-s}\boldsymbol{c}'\left(\frac{\partial^2 g_t(\boldsymbol{\theta_T^*}(\boldsymbol{c}))}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \frac{\partial^2 g_t(\boldsymbol{\theta_0})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)\boldsymbol{c} \xrightarrow{p} 0,
$$

for $\boldsymbol{c}\in\boldsymbol{C}$. As in the proof of Theorem 1 of Andrews et al. (2006), we can make use of Remark 1 of Davis et al. (1992) and complete the proof.

C.3 Approximate Likelihood Function

Define $\boldsymbol{b} = \boldsymbol{\beta}'\tilde{\boldsymbol{x}}$ such that $\boldsymbol{z} = \boldsymbol{BAy} - \boldsymbol{\beta}'\tilde{\boldsymbol{x}} = \boldsymbol{BAy} - \boldsymbol{b}$. Assume $\boldsymbol{B}$ and $\boldsymbol{A}$ are invertible. We are interested in the inverse transformation, i.e. $\boldsymbol{y} = \boldsymbol{Q}(\boldsymbol{z} + \boldsymbol{b})$, where $\boldsymbol{Q} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$. Let $\boldsymbol{Q}$ be a (2×2) matrix, then we have

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} q_1 & q_2 \\ q_3 & q_4 \end{bmatrix} \left(\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \right), \tag{19}$$

with the following functions $y_1 = g_1(z_1, z_2) = q_1 z_1 + q_2 z_2 + b_1$ and $y_2 = g_2(z_1, z_2) = q_3 z_1 + q_4 z_2 + b_2$. The Jacobian is given as the matrix of all partial derivatives from y to z, i.e.

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial y_1}{\partial z_1} & \frac{\partial y_1}{\partial z_2} \\ \frac{\partial y_2}{\partial z_1} & \frac{\partial y_2}{\partial z_2} \end{bmatrix} = \begin{bmatrix} q_1 & q_2 \\ q_3 & q_4 \end{bmatrix} = \boldsymbol{Q}. \tag{20}$$

Then the joint density of y_1 and y_2 is given as $f_{y_1,y_2}(y_1, y_2) = \frac{1}{|\det(\boldsymbol{Q})|} f_{z_1,z_2}(\boldsymbol{BAy} - \boldsymbol{b})$. This result can be generalized to higher orders (see e.g., Casella and Berger, 2002, p. 185). From Proposition 1 we know that the information sets (i) and (vi) are observationally equivalent. Using this result and $\boldsymbol{Q} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$, we find that the probability density of the process y_t can be represented in the following way:

$$\begin{aligned} f_{y;\boldsymbol{\lambda}}(\boldsymbol{y}) &= \frac{1}{|\det(\boldsymbol{Q})|} f_z(\boldsymbol{BAy} - \boldsymbol{b}; \boldsymbol{\lambda}) \\ &= |\det(\boldsymbol{A})||\det(\boldsymbol{B})| h_V(\boldsymbol{BAy} - \boldsymbol{b}) f_\varepsilon(\boldsymbol{BAy} - \boldsymbol{b}; \boldsymbol{\lambda}) h_U(\boldsymbol{BAy} - \boldsymbol{b}) \\ &= |\det(\boldsymbol{A})| h_V(\boldsymbol{BAy} - \boldsymbol{b}) \left(\prod_{t=r+1}^{T-s} f_\sigma(\boldsymbol{BAy} - \boldsymbol{b}; \boldsymbol{\lambda}) \right) h_U(\boldsymbol{BAy} - \boldsymbol{b}) \\ &= h_V(\varphi(L^{-1})y_1, ..., \varphi(L^{-1})y_r) \left(\prod_{t=r+1}^{T-s} f_\sigma(\phi(L)\varphi(L^{-1})y_t - \boldsymbol{\beta}'\boldsymbol{X}_t; \boldsymbol{\lambda}) \right) \\ &\quad h_U(\phi(L)y_{T-s+1}, ..., \phi(L)y_T)|\det(\boldsymbol{A})|, \end{aligned} \tag{21}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are two nonsingular matrices with $\det(\boldsymbol{B}) = 1$; h_V and h_U are the joint densities of $(v_1, ..., v_r)$ and $(u_{T-s+1}, ..., u_T)$ respectively. Independence of the blocks $(v_1, ..., v_r)$, $(\varepsilon_{r+1}, ..., \varepsilon_{T-s})$ and $(u_{T-s+1}, ..., u_T)$ is applied in the second equality and the definition of the filtered values as presented in (3) and (4) in the fourth equality. Since $\det(\boldsymbol{A})$ is independent of sample size, the density of y_t can be approximated by the second term in (21).

C.4 From Transfer Function Model to MARX

For expository purposes, we consider a single explanatory variable denoted x_t^*. The transfer function model is given by

$$y_t = \psi^*(L)x_t^* + n_t, \tag{22}$$

where n_t is a noise process assumed to follow a stationary AR process, $a(L)n_t = \varepsilon_t^*$. The ARX (or ARDL) model can be motivated from (22) by assuming that the transfer function operator can be expressed in a rational factorization as $\psi^*(z) = a(z)^{-1}\theta^*(z)$. Multiplying (22) by $a(L)$ yields

$$\begin{aligned} a(L)y_t &= \theta^*(L)x_t^* + a(L)n_t \\ &= \theta^*(L)x_t^* + \varepsilon_t^*, \end{aligned} \tag{23}$$

which is the usual ARX(p, k) model representation when $\deg(a(z)) = p$ and $\deg(\theta^*(z)) = k$. If all roots of $a(z)$ lie outside the unit circle, the process is stationary which implies that estimation and inference can directly be conducted. Breidt et al. (1991) consider the more complex case in which r roots lie outside the unit circle and s inside $(r + s = p)$ and propose to factorize the polynomial to obtain

$$\phi(L)\varphi^*(L)y_t = \theta^*(L)x_t^* + \varepsilon_t^*.$$

Lanne and Saikkonen (2011) propose to rewrite $\varphi^*(z)$ in terms of the lead operator and obtain the relation $\varphi^*(z) = -\varphi_s^* z^s \varphi(z^{-1})$. By rearranging terms, we find

$$\begin{aligned}\phi(L)\varphi(L^{-1})y_t &= \left(-\frac{1}{\varphi_s^*} + \frac{\theta_1^*}{\varphi_s^*}L + ... + \frac{\theta_k^*}{\varphi_s^*}L^k\right)x_{t+s}^* - \frac{1}{\varphi_s^*}\varepsilon_{t+s}^* \\ &= \theta(L)x_t + \varepsilon_t. \end{aligned} \tag{24}$$

In case only a contemporaneous value of x_t enters the system, take $\psi^*(z) = a(z)^{-1}\boldsymbol{\beta}$. Note that the derivation can easily be extended to q regressors by defining $\psi^*(z) = [\psi_1^*(z), ..., \psi_q^*(z)]'$ and considering X_t^*. In the distributed lag case take $\psi^*(z) = a(z)^{-1}\theta^*(z)$ with $\theta^*(z) = [\theta_1^*(z), ..., \theta_k^*(z)]'$; in the contemporaneous case define $\psi^*(z) = a(z)^{-1}\boldsymbol{\beta}$ with $\boldsymbol{\beta} \in \mathbb{R}^q$. In case one wants to allow for (mixed) dynamics in the exogenous regressors, it seems more natural to model such a process as a VAR. The mixed VAR model (see e.g., Lanne and Saikkonen, 2013; Davis and Song, 2012) accommodates this structure.

C Auxiliary Proposition with Proof

Proposition 1. *For an MARX(r, s, q) model, the following information sets are equivalent:*

(i) $(y_1, ..., y_T, \boldsymbol{X}_{r+1}, ..., \boldsymbol{X}_{T-s})$

(ii) $(y_1, ..., y_r, u_{r+1}, ..., u_T, \boldsymbol{X}_{r+1}, ..., \boldsymbol{X}_{T-s})$

(iii) $(v_1, ..., v_{T-s}, y_{T-s+1}, ..., y_T, \boldsymbol{X}_{r+1}, ..., \boldsymbol{X}_{T-s})$

(iv) $(y_1, ..., y_r, \varepsilon_{r+1}, ..., \varepsilon_{T-s}, u_{T-s+1}, ..., u_T)$

(v) $(v_1, ..., v_r, \varepsilon_{r+1}, ..., \varepsilon_{T-s}, y_{T-s+1}, ..., y_T)$

(vi) $(v_1, ..., v_r, \varepsilon_{r+1}, ..., \varepsilon_{T-s}, u_{T-s+1}, ..., u_T)$

Additionally, the following information sets are equivalent:

(i') $(y_1, ..., y_T)$

(ii') $(y_1, ..., y_r, u_{r+1}, ..., u_T)$

(iii') $(v_1, ..., v_{T-s}, y_{T-s+1}, ..., y_T)$

Proof of Proposition 1

Let $\sim$ denote equivalence in information sets. To show that $(i), (ii)$ and (iii) are equivalent is similar to showing that $(i'), (ii')$ and (iii') are equivalent. We prove $(i') \sim (ii'), (i') \sim (iii'), (ii) \sim (iv), (iii) \sim (v)$ and $(i) \sim (vi)$.

Case 1: $(i') \sim (ii')$. Using (3), (ii') $(y_1, ..., y_r, u_{r+1}, ..., u_T) = (y_1, ..., y_r, \phi(L)y_{r+1}, ..., \phi(L)y_T)$. Since $u_{r+1} = y_{r+1} - \phi_1 y_r - ... - \phi_r y_1$ with $y_1, ..., y_r$ and u_{r+1} known, y_{r+1} is known. The same reasoning can be recursively applied to u_{r+2} up to u_T, leading to the desired result.

Case 2: $(i') \sim (iii')$. By (4), (iii') $(v_1, ..., v_{T-s}, y_{T-s+1}, ..., y_T) = (\varphi(L^{-1})y_1, ..., \varphi(L^{-1})y_{T-s}, y_{T-s+1}, ..., y_T)$. Since $v_{T-s} = y_{T-s} - \varphi_1 y_{T-s+1} - ... - \varphi_s y_T$ with $y_{T-s+1}, ..., y_T$ and v_{T-s} known, y_{T-s} is known. The same reasoning can be recursively applied to v_{T-s-1} up to v_1, leading to the desired result. Hence, since $(i'), (ii')$ and (iii') are equivalent, we know that $(i), (ii)$ and (iii) are as well (as all information sets are augmented with the same information).

Case 3: $(ii) \sim (iv)$. We have (iv) $(y_1, ..., y_r, \varepsilon_{r+1}, ...\varepsilon_{T-s}, u_{T-s+1}, .., u_T) = (y_1, ..., y_r, \varphi(L^{-1})u_{r+1} - \boldsymbol{\beta}'\boldsymbol{X}_{r+1}, ..., \varphi(L^{-1})u_{T-s} - \boldsymbol{\beta}'\boldsymbol{X}_{T-s}, u_{T-s+1}, ..., u_T)$ by using the second equality in equation (3). Since $\varepsilon_{T-s} = u_{T-s} - \varphi_1 u_{T-s+1} - ... - \varphi_s u_T - \boldsymbol{\beta}'\boldsymbol{X}_{T-s}$ with $u_{T-s+1}, ..., u_T$, $\boldsymbol{X}_{T-s}$ and ε_{T-s} known, u_{T-s} is known. The same reasoning can be recursively applied to u_{T-s-1} up to u_{r+1}.

Case 4: $(iii) \sim (v)$. We have (v) $(v_1, ..., v_r, \varepsilon_{r+1}, ...\varepsilon_{T-s}, y_{T-s+1}, .., y_T) = (v_1, ..., v_r, \phi(L)v_{r+1} - \boldsymbol{\beta}'\boldsymbol{X}_{r+1}, ..., \phi(L)v_{T-s} - \boldsymbol{\beta}'\boldsymbol{X}_{T-s}, y_{T-s+1}, ..., y_T)$ by using the second equality in equation (4). Since $\varepsilon_{r+1} = v_{r+1} - \phi_1 v_r - ... - \phi_r v_1 - \boldsymbol{\beta}'\boldsymbol{X}_{r+1}$ with $v_1, ..., v_r$, $\boldsymbol{X}_{r+1}$ and ε_{r+1} known, v_{r+1} is known. This reasoning can be recursively applied to v_{r+2} up to v_{T-s}.

Case 5: $(i) \sim (vi)$. To show: $(y_1, ..., y_T, \boldsymbol{X}_r, ..., \boldsymbol{X}_{T-s}) \sim (v_1, ..., v_r, \varepsilon_{r+1}, ..., \varepsilon_{T-s}, u_{T-s+1}, ..., u_T)$.

We denote the vector corresponding to the information set (i) by $\tilde{\boldsymbol{y}}$ and (ii) by $\boldsymbol{z}$ and use results similar to Lanne and Saikkonen (2011). Define $\boldsymbol{w} = [v_1, ..., v_{T-s}, u_{T-s+1}, ..., u_T]'$ and $\boldsymbol{y} = [y_1, ..., y_T]'$. Then

$$\begin{bmatrix} v_1 \\ \vdots \\ v_{T-s} \\ u_{T-s+1} \\ \vdots \\ v_T \end{bmatrix} = \begin{bmatrix} y_1 - \varphi_1 y_2 - ... - \varphi_s y_{s+1} \\ \vdots \\ y_{T-s} - \varphi_1 y_{T-s+1} - ... - \varphi_s y_T \\ y_{T-s+1} - \phi_1 y_{T-s} - ... - \phi_r y_{T-s+1-r} \\ \vdots \\ y_T - \phi_1 y_{T-1} - ... - \phi_r y_{T-r} \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} y_1 \\ \vdots \\ y_{T-s} \\ y_{T-s+1} \\ \vdots \\ y_T \end{bmatrix},$$

which can be written as $\boldsymbol{w} = \boldsymbol{A}\boldsymbol{y}$. Now define $\tilde{\boldsymbol{x}} = [\underbrace{0, ..., 0}_{r \text{ times}}, \boldsymbol{X}_{r+1}, ..., \boldsymbol{X}_{T-s}, \underbrace{0, ..., 0}_{s \text{ times}}]'$. Similarly, we form the following system of equations for $\boldsymbol{z}$, since (6) shows that $\boldsymbol{z} = \boldsymbol{B}\boldsymbol{w} - \boldsymbol{\beta}'\tilde{\boldsymbol{x}}$. Combining both systems of equations, we find that $\boldsymbol{z}$ and $\tilde{\boldsymbol{y}}$ are related as $\boldsymbol{z} = \boldsymbol{B}\boldsymbol{A}\boldsymbol{y} - \boldsymbol{\beta}'\tilde{\boldsymbol{x}}$, where $\boldsymbol{y}$ and $\tilde{\boldsymbol{x}}$ combined form the information set $\tilde{\boldsymbol{y}}$. Since $\boldsymbol{B}$, $\boldsymbol{A}$ and $\boldsymbol{\beta}$ only contain known parameters, this shows that (i) and (ii) are equivalent. All cases combined show that information sets $(i) - (vi)$ are equivalent.

$$\begin{bmatrix} v_1 \\ \vdots \\ v_r \\ v_{r+1} - \phi_1 v_r - ... - \phi_r v_1 - \boldsymbol{\beta}'\boldsymbol{X}_{r+1} \\ \vdots \\ v_{T-s} - \phi_1 v_{T-s-1} - ... - \phi_r v_{T-s-r} - \boldsymbol{\beta}'\boldsymbol{X}_{T-s} \\ u_{T-s+1} \\ \vdots \\ u_T \end{bmatrix} = \boldsymbol{B} \begin{bmatrix} v_1 \\ \vdots \\ v_r \\ v_{r+1} \\ \vdots \\ v_{T-s} \\ u_{T-s+1} \\ \vdots \\ u_T \end{bmatrix} - \boldsymbol{\beta}' \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \boldsymbol{X}_{r+1} \\ \vdots \\ \boldsymbol{X}_{T-s} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

References

Andrews, B., Breidt, F. & Davis, R. (2006). Maximum Likelihood Estimation For All-Pass Time Series Models. *Journal of Multivariate Analysis*, 97, 1638–1659. doi: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. doi:10.1016/0047-259X(91)90056-8

Brockwell, P. & Davis, R. (1991). *Time Series: Theory and Methods.* Springer-Verlag New York, Second Edition.

Casella, G. & Berger, R. (2002). *Statistical Inference.* Thomson Learning, Second Edition.

Davis, R., Knight, K. & Liu, J. (1992). M-Estimation for Autoregressions with Infinite Variance. *Stochastic Processes and their Applications*, 40(1), 145–180.

Davis, R. & Song, L. (2012). Noncausal Vector AR Processes with Application to Economic Time Series. *Discussion Paper*, Colombia University.

Gouriéroux, C. & Jasiak, J. (2016). Filtering, Prediction and Simulation Methods for Noncausal Processes. *Journal of Time Series Analysis*, 37, 405–430. doi:10.1111/jtsa.12165

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. doi:10.15609/annaeconstat2009.123-124.0307

Lanne, M. & Saikkonen, P. (2011). Noncausal Autoregressions for Economic Time Series. *Journal of Time Series Econometrics*, 3(3), 1–32. doi:10.2202/1941-1928.1080

Lanne, M. & Saikkonen, P. (2013). Noncausal Vector Autoregression. *Econometric Theory*, 29(3), 447–481. doi:10.1017/S0266466612000448