Autoregressive Models for Matrix-Valued Time Series

Rong Chen, Han Xiao, Dan Yang

Journal of Econometrics, 2021

Motivations

- In many applications (e.g., economics, finance), data naturally appear as **matrices** observed over time:
  - Rows: different types of indicators (e.g., GDP, inflation)
  - Columns: different entities (e.g., countries)
- **Traditional approach:** vectorize the matrix and use a VAR model.
- **Problems** with vectorization:
  - Leads to many parameters and reduced interpretability.
  - Ignores meaningful row-wise and column-wise relationships.
- **Goal:** Develop a model that preserves matrix structure and reduces complexity.

Motivations

- Proposed model: **Matrix Autoregressive** (MAR(1)):

$$X_t = AX_{t-1}B' + E_t$$

  where X_t is an $m \times n$ design matrix, while A and B are respectively $m \times m$ and $n \times n$ matrices capturing rows and columns interactions. Instead E_t is the $m \times n$ white-noise matrix.

- **Advantages:**
  - Keeps matrix form: interpretable row/column dynamics.
  - Fewer parameters: only $m^2 + n^2$ vs $m^2 n^2$ in VAR.
  - Naturally extends to higher order: MAR(p).
- **Extension:**
  - Structured error covariance (Kronecker product) to reduce dimensionality.
- **Applications:** Cross-country economic indicators, transport networks, multi-site sensors.

Key Structural Features

- **VAR Representation:**

$$\mathrm{vec}(X_t) = (B \otimes A)\,\mathrm{vec}(X_{t-1}) + \mathrm{vec}(E_t)$$

  Uses only $m^2 + n^2$ parameters instead of $m^2 n^2$ of an unstructured VAR.

- **Identifiability Constraint:**
  - Model is invariant under scaling: $(cA)X_{t-1}(B'/c) = AX_{t-1}B'$.
  - Resolved by fixing $\|A\|_F = 1$ (Frobenius norm).
- **Error Covariance Structure:**
  - $\mathrm{vec}(E_t) \sim \mathcal{N}(0, \Sigma_c \otimes \Sigma_r)$.
  - Σ_r: row-wise covariances.
  - Σ_c: column-wise covariances.
  - Greatly reduces number of parameters in $\mathrm{Cov}(\mathrm{vec}(E_t))$.

Interpretation

- **Model:** $X_t = AX_{t-1}B' + E_t$
  - $A \in \mathbb{R}^{m\times m}$: models **row-wise interactions** (e.g. between indicators)
  - $B \in \mathbb{R}^{n\times n}$: models **column-wise interactions** (e.g. between countries)
- **Special Case:** $A = I$
  - $X_t = X_{t-1}B' + E_t$
  - Each indicator in a country depends on that indicator's past values across all countries
  - *(e.g., US GDP depends on past GDPs from US, UK, DE, etc.)*
- **Special Case:** $B = I$
  - $X_t = AX_{t-1} + E_t$
  - Each country's indicators depend on past values of all indicators within that country
  - *(e.g., US GDP depends on past US GDP, inflation, interest rate, etc.)*
- **General MAR(1):** combines both indicator-wise and country-wise effects in a bilinear way.

Probabilistic Properties

- **Causality Condition:**
  - MAR(1) is **stationary and causal** if:

$$\rho(A) \cdot \rho(B) < 1$$

    where $\rho(\cdot)$ is the spectral radius (max eigenvalue modulus).
  - Ensures the time series does not diverge over time.
- **Causal Moving Average Representation:**

$$\text{vec}(X_t) = \sum_{k=0}^{\infty} (B^k \otimes A^k)\, \text{vec}(E_{t-k})$$

  - Expresses X_t as a function of past shocks, where k is the lag.
  - The more stable the process (i.e., the smaller (A)(B)), the faster the weights decay to zero.
  - Basis for deriving theoretical properties.

Estimation: MLE under a Structured Covariance Tensor

Model: $X_t = AX_{t-1}B' + E_t, \quad \text{with} \quad \mathrm{vec}(E_t) \sim \mathcal{N}(0, \Sigma_c \otimes \Sigma_r)$

Log-Likelihood Function:

$$
\begin{aligned}
& -m(T-1)\log|\Sigma_c| - n(T-1)\log|\Sigma_r| \\
& \quad - \sum_t \mathrm{tr}\left(\Sigma_r^{-1}(X_t - AX_{t-1}B')\Sigma_c^{-1}(X_t - AX_{t-1}B')'\right).
\end{aligned}
$$

Estimation Procedure:

- Iteratively update gradient conditions for A, B, Σ_r, and Σ_c while fixing the others.
- Use closed-form updates from setting partial derivatives to zero.
- Apply normalization: $\|A\|_F = 1$, $\|\Sigma_r\|_F = 1$ to ensure identifiability.

Estimation: MLE vs LSE

- Most **efficient** estimator when $\text{Cov}(\text{vec}(E_t)) = \Sigma_c \otimes \Sigma_r$.
- Achieves smaller estimation error in simulations.
- Computationally intensive.
- Sensitive to model misspecification (esp. if Σ is not truly separable)

Iterated Least square:

$$\min_{A,B} \sum_t \|\mathbf{X}_t - A\mathbf{X}_{t-1}B'\|_F^2$$

- Assumes the entries of E_t are i.i.d. normal with mean zero and a constant variance.
- Less efficient but robust when structure is unknown.
- Lower computational cost.

Asymptotic Properties

Under stationarity and identifiability conditions:

- Both estimators are:
  - **Consistent**: converge to the true (A, B) as $T \to \infty$
  - **Asymptotically normal**:

$$\sqrt{T}\left(\begin{bmatrix}\mathrm{vec}(\hat{A})\\ \mathrm{vec}(\hat{B}')\end{bmatrix} - \begin{bmatrix}\mathrm{vec}(A)\\ \mathrm{vec}(B')\end{bmatrix}\right) \xrightarrow{d} \mathcal{N}(0, \Omega)$$

$$\sqrt{T}\left[\mathrm{vec}(\widehat{\mathbf{B}}') \otimes \mathrm{vec}(\widehat{\mathbf{A}}) - \mathrm{vec}(\mathbf{B}') \otimes \mathrm{vec}(\mathbf{A})\right] \Rightarrow \mathcal{N}(0, V\Omega V')$$

- Asymptotic variances Ω are derived explicitly for:
  - **MLE:** $\Omega = \tilde{H}^{-1}\mathbb{E}(W_t\Sigma^{-1}W_t^{-1})\tilde{H}$, where $\tilde{H} = \mathbb{E}(W_t\Sigma^{-1}W_t') + \gamma\gamma'$, $W_t^{'} = [(BX_t^{'} \otimes I, I \otimes (AX_t)]$, $V = [\beta \otimes I, I \otimes \alpha]$, $\gamma = (\alpha', 0')$, $\alpha = vec(A)$ and $\beta = vec(B')$.
  - **LSE:** $\Omega = H^{-1}\mathbb{E}(W_t\Sigma W_t^{-1})H$, where $H = \mathbb{E}(W_t W_t') + \gamma\gamma'$ and $W_t' = [(BX_t^{'}) \otimes I : I \otimes (AX_t)]$.

Simulations

Objective:
Evaluate and compare the performance of three estimators for matrix-valued time series:

- **MAR(1)** with **LSE**
- **MAR(1)** with **MLE**
- **VAR(1)**

Simulation Design:

- Data generated from MAR(1): $X_t = AX_{t-1}B' + E_t$
- Matrix sizes: $(m, n) = (3, 2), (6, 4), (9, 6)$
- Sample sizes: $T = 100, 200, 400, 5000$
- Error covariance settings:
  - Setting I: $\Sigma = I$ (iid entries).
  - Setting II: Full arbitrary Σ (arbitrarily correlated entries).
  - Setting III: $\Sigma = \Sigma_c \otimes \Sigma_r$ (structured entries).

Evaluation metric and Results

The evaluation metric is the **logarithm of the Frobenius norm of the estimation error**:

$$\log(\|\widehat{\mathbf{B}} \otimes \widehat{\mathbf{A}} - \mathbf{B} \otimes \mathbf{A}\|_F^2)$$

Key Findings:

- **VAR(1)** has consistently the worst performance due to overparameterization.
- **LSE** performs best in Setting I as it aligns with the true likelihood.
- **MLE** and **LSE** performs surprisingly similarly under setting II. MLE shows efficiency even under model misspecification.
- **MLEs** is optimal under Setting III (Kronecker structured covariance).

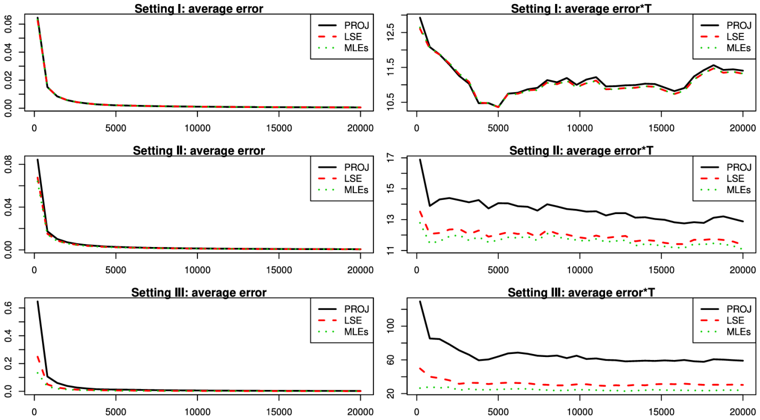

Fig. 5. Comparison of the asymptotic efficiencies of three estimators, PROJ, LSE, and MLEs, under three settings. The left column shows the average error over 100 repetitions for $\|\hat{\boldsymbol{B}} \otimes \hat{\boldsymbol{A}} - \boldsymbol{B} \otimes \boldsymbol{A}\|_F^2$ and the right for $T\|\hat{\boldsymbol{B}} \otimes \hat{\boldsymbol{A}} - \boldsymbol{B} \otimes \boldsymbol{A}\|_F^2$.

Confidence Intervals

Table 1
Percentage of coverages of 95% confidence intervals.

	Setting	I			II			III		
	Estimator	PROJ	LSE	MLEs	PROJ	LSE	MLEs	PROJ	LSE	MLEs
$(\mathrm{vec}'(\hat{\boldsymbol{A}}), \mathrm{vec}'(\hat{\boldsymbol{B}}))'$	T = 100	0.926	0.934	0.932	0.913	0.935	0.923	0.872	0.906	0.947
	T = 200	0.938	0.941	0.941	0.937	0.944	0.932	0.915	0.934	0.950
	T = 1000	0.950	0.951	0.951	0.947	0.947	0.933	0.946	0.949	0.953
$\mathrm{vec}(\hat{\boldsymbol{B}}) \otimes \mathrm{vec}(\hat{\boldsymbol{A}})$	T = 100	0.915	0.923	0.921	0.905	0.922	0.911	0.860	0.885	0.936
	T = 200	0.935	0.938	0.937	0.930	0.939	0.928	0.903	0.923	0.945
	T = 1000	0.950	0.952	0.951	0.946	0.945	0.932	0.942	0.944	0.950

Application: Economic Indicators

Data:

- Quarterly observations from 1991–2016
- 4 indicators: Interest rate, GDP growth, Manufacturing production, Core inflation
- 5 countries: US, UK, Canada, Germany, France
- Forms a 4×5 matrix at each time point

RSS:

Table 2
Residual sum of squares of MAR(1) model using three different estimators and the stacked VAR(1) estimator; and the total residual sum of squares of fitting univariate AR(1) and AR(2) to each individual time series; and the total sum of squares of the original (normalized) data.

MAR(1) PROJ	MAR(1) LSE	MAR(1) MLEs	VAR(1)	iAR(1)	iAR(2)	Original
1828	1318	1332	1028	1585	1515	2076

Application: Economic Indicators

Table 3
Estimated left coefficient matrix $\boldsymbol{A}$ of MAR(1) using LS method. Standard errors are shown in the parentheses. The right panel indicates the positively significant, negatively significant and insignificant parameters at 5% level using symbols $(+, -, 0)$, respectively.

	Int	GDP	Prod	CPI	Int	GDP	Prod	CPI
Int	0.177	0.215	0.132	−0.171	+	+	0	−
	(0.061)	(0.082)	(0.088)	(0.063)				
GDP	−0.19	0.341	0.346	−0.08	−	+	+	0
	(0.05)	(0.086)	(0.081)	(0.062)				
Prod	−0.223	0.318	0.424	−0.095	−	+	+	0
	(0.054)	(0.092)	(0.087)	(0.068)				
CPI	−0.028	0.048	−0.045	0.502	0	0	0	+
	(0.05)	(0.07)	(0.078)	(0.052)				

Table 4
Estimated right coefficient matrix $\boldsymbol{B}$ of MAR(1) using LS method. Standard errors are shown in parentheses. The right panel indicates the positively significant, negatively significant and insignificant parameters at 5% level using symbols $(+, -, 0)$, respectively.

	USA	DEU	FRA	GBR	CAN	USA	DEU	FRA	GBR	CAN
USA	0.878	−0.044	0.15	0.359	−0.043	+	0	0	+	0
	(0.134)	(0.202)	(0.138)	(0.132)	(0.156)					
DEU	0.722	0.072	0.801	0.308	−0.212	+	0	+	+	−
	(0.076)	(0.124)	(0.083)	(0.078)	(0.092)					
FRA	0.44	0.064	0.438	0.208	0.024	+	0	+	0	0
	(0.12)	(0.197)	(0.136)	(0.125)	(0.148)					
GBR	0.545	0.032	0.272	0.406	−0.018	+	0	+	+	0
	(0.089)	(0.153)	(0.101)	(0.101)	(0.118)					
CAN	0.553	0.023	−0.002	0.531	0.324	+	0	0	+	+
	(0.079)	(0.13)	(0.087)	(0.085)	(0.1)					

Application: Economic Indicators

Table 5
Sum of out-of-sample prediction error squares of MAR(1) model using three different estimators and the stacked VAR(1) estimator, and the total sum of out-of-sample prediction error squares of fitting univariate AR(1) and AR(2) to each individual time series.

MAR(1) PROJ	MAR(1) LSE	MAR(1) MLEs	iAR(1)	iAR(2)	VAR(1)
148.05	142.03	137.29	143.82	150.80	181.64

Takeaways:

- **Forecasting:** MAR(1) with LSE or MLEs outperforms both VAR(1) and univariate AR models in out-of-sample prediction.
- **VAR(1)** overfits in-sample and underperforms out-of-sample, confirming inefficiency in high dimensions.

Conclusion

- The MAR(1) model provides a structured and interpretable framework for matrix-valued time series.
- It significantly reduces parameter dimensionality compared to traditional VAR models.
- Simulation studies validate theoretical properties and show that MLEs is most efficient under structured errors, while LSE is robust.
- Real data analysis demonstrates the model's practical effectiveness in capturing economic dynamics and improving forecasts.

Final Thought:
MAR(1) is a promising tool for high-dimensional time series analysis where preserving matrix structure matters.

Thank You!