High-Dimensional Vector Autoregressive Modeling via Tensor Decomposition

Mini-Project - MVA 2025/2026

Antoine Le Maguet Alexandre Mallez

Master MVA - ENS Paris-Saclay

17 Décembre 2025

Plan de la présentation

Le problème de la haute dimension dans les VAR

- **Modèle Standard (VAR) :**

$$Y_t = \sum_{k=1}^{P} A_k Y_{t-k} + \epsilon_t$$

- **Problème de dimensionnalité :**
  - Le nombre de paramètres croît en $O(N^2 P)$ (quadratique en N).
- **Solution proposée (Wang et al., 2020) :**
  - Réarranger les matrices de transition en un tenseur d'ordre 3.
  - Appliquer une décomposition de Tucker pour réduire l'espace des paramètres.

$$Y_t = \mathcal{A}_{(1)} x_t + \epsilon_t$$

Reformulation Tensorielle

Décomposition de Tucker : Le tenseur de transition $\mathcal{A}$ est décomposé en un cœur $\mathcal{G}$ et trois matrices facteurs U_1, U_2, U_3. (HOSVD)

$$\mathcal{A} = \mathcal{G} \times_1 U_1 \times_2 U_2 \times_3 U_3$$

Interprétation des modes :

1. **Mode 1 (U_1) :** Facteurs de réponse (réduit $N \to r_1$).
2. **Mode 2 (U_2) :** Facteurs prédicteurs (réduit $N \to r_2$).
3. **Mode 3 (U_3) :** Facteurs temporels (réduit les lags $P \to r_3$).

Avantage : Le nombre de paramètres devient linéaire en N et P.

1. Estimateur MLR (Multilinear Low-Rank)

Adapté pour les dimensions fixes. Minimisation des moindres carrés :

$$\hat{\mathcal{A}}_{MLR} = \arg \min_{\mathcal{G}, U_1, U_2, U_3} \sum_{t=1}^{T} ||y_t - (\mathcal{G} \times_1 U_1 \times_2 U_2 \times_3 U_3)_{(1)} x_t||_2^2$$

Propriétés théoriques :

- Consistance et Normalité Asymptotique.
- Variance asymptotique inférieure à celle des OLS et RRR.

Algorithme : Alternating Least Squares (ALS). On minimimise alternativement U_1, U_2, U_3 et $\mathcal{G}$.

2. Estimateur SHORR (Sparse Higher-Order)

En général dans des données réelles, les matrices U_i sont parsimonieuses. Surtout à haute dimension ($N, P \to \infty$). Ajout d'une pénalité de sparsité l_1.

$$\hat{\mathcal{A}}_{SHORR} = \arg\min_{\mathcal{G},U} \left(L(\mathcal{G}, U) + \lambda ||U_3 \otimes U_2 \otimes U_1||_1 \right)$$

Caractéristiques :

- Impose la sparsité et l'orthogonalité des facteurs.
- Erreur bornée proportionnelle à $\sqrt{S \log(N^2 P)/T}$.

Algorithme : ADMM (Alternating Direction Method of Multipliers).

- Complexe à implémenter (contraintes d'orthogonalité non convexes + pénalité l_1).
- Necessite des termes de régularisation en entrée

Sélection de Rang (Ridge-Type Ratio)

Les rangs (r_1, r_2, r_3) sont inconnus en pratique.
Méthode : "Ridge-Type Ratio Estimator". Basée sur l'identification de la chute des valeurs singulières d'un estimateur initial (ex: OLS ou Norme Nucléaire).

$$\hat{r}_i = \arg \min_{1 \leq j \leq d_i} \frac{\sigma_{j+1} + c}{\sigma_j + c}$$

Résultats expérimentaux :

- La probabilité de sélectionner le bon rang converge vers 1.
- **Condition :** Nécessite un "gap suffisant" entre les valeurs singulières pour ne pas avoir l'impact du bruit.

Simulations Numériques

Données synthétiques stationnaires :

- Génération de processus stables (rayon spectral < 1).

Validations :

1. **MLR (Efficacité) :** La variance empirique converge vers la borne théorique calculée ($\Sigma_{MLR} \leq \Sigma_{RRR} \leq \Sigma_{OLS}$).
2. **SHORR (Convergence) :** L'erreur d'estimation diminue avec T même en haute dimension.

Limitations techniques

L'algorithme ADMM (SHORR) est coûteux en temps de calcul, limitant le nombre de réplications Monte Carlo.

Résultats

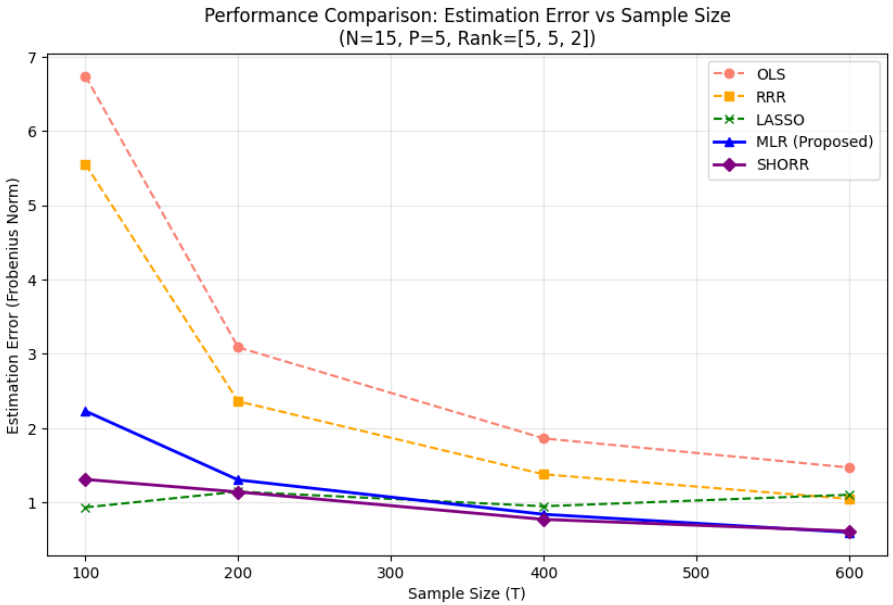

Fig 1. Erreur vs Taille d'échantillon

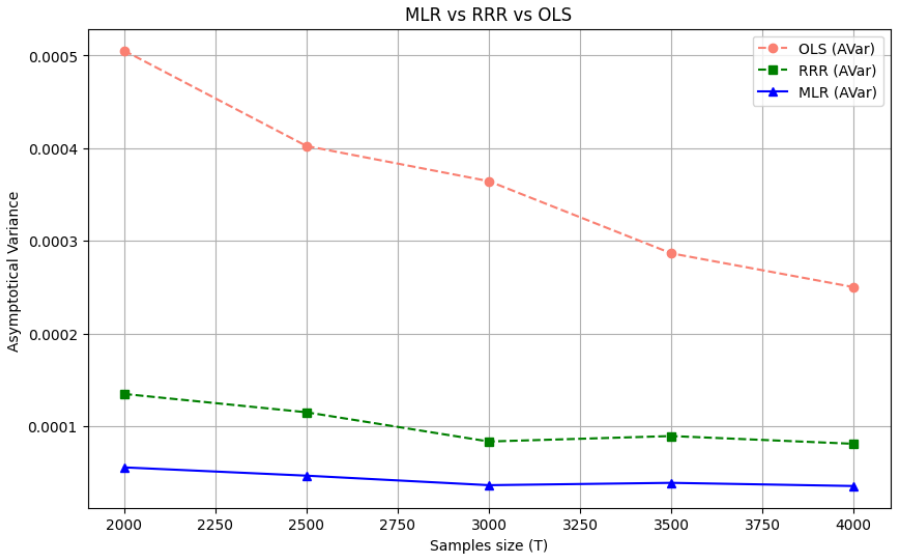

Fig 2. Variance asymptotique : MLR vs RRR vs OLS.

- **Observation :** La variance empirique de l'estimateur MLR (bleu) est nettement inférieure à celle des OLS et RRR.
- **Validation :** La convergence vers la borne théorique confirme la validité de l'algèbre tensorielle implémentée.

Application Réelle : Macroéconomie US

Données :

- Dataset : Données Macro-économique .
- 11 indicateurs (PIB, Inflation, Taux d'intérêt, etc.).
- Prétraitement pour stationnarité.

Comparaison des prédictions :

Horizon	**Meilleur Modèle**
Court terme ($t+1$)	**MLR** (Capture les variations locales)
Long terme	**SHORR** (Plus robuste, moins d'explosivité)

Observation : SHORR généralise mieux grâce à la régularisation dans un environnement bruité.

Résultats

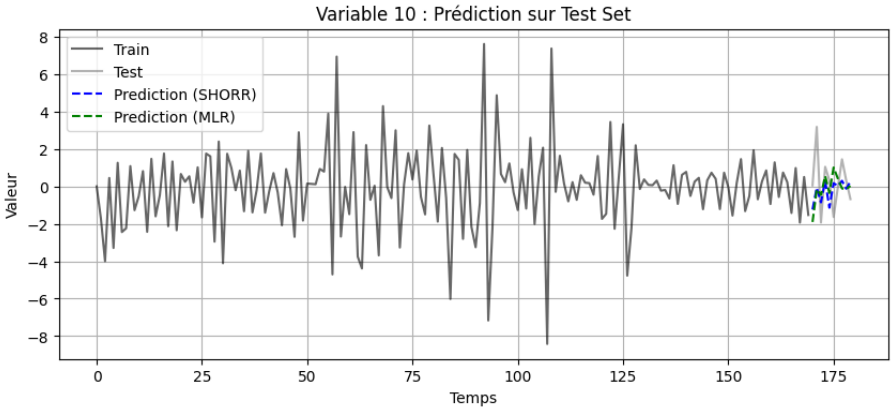

Fig 1. Prédiction d'une série temporelle

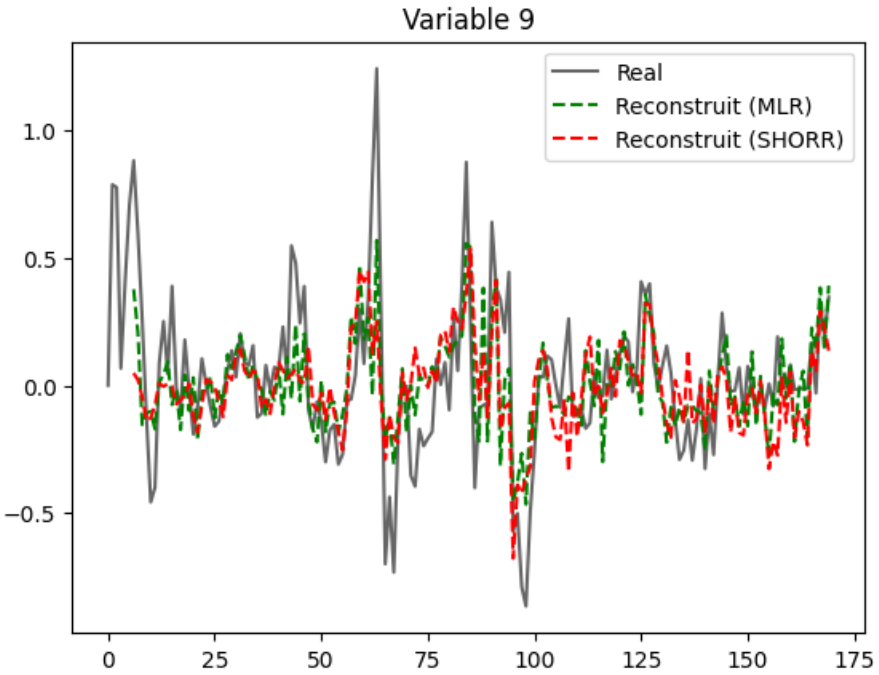

Fig 2. Reconstruction du signal après estimations des paramètres.

Conclusion et Contributions

Réalisations :

- Implémentation **100% from scratch** (Python, Numpy, Tensorly pour opérations de base).
- Validation numérique des bornes théoriques de variance.
- Application sur données économiques réelles.

Limitations :

- Coût computationnel de l'ADMM.
- Sensibilité de la sélection de rang au gap des valeurs singulières.
- Hypothèse de stationnarité parfois forte pour les données réelles.
- Choix des termes de régularisation

Pistes futures :

- Modèles à coefficients variables dans le temps (Tenseurs d'ordre 4).
- Grid search/BIC pour le choix de termes de régularisation optimaux

Algorithm: ALS (MLR Estimator)

Algorithm 1 Alternating Least Squares (ALS)

Initialization: Choose $\mathcal{A}^{(0)}$ (e.g., via OLS) and compute HOSVD.
repeat
    Update Factors U_n**:**
    $U_1^{(k+1)} \leftarrow \arg\min_{U_1} \sum_t \|y_t - (\dots)\mathrm{vec}(U_1)\|_2^2$
    $U_2^{(k+1)} \leftarrow \arg\min_{U_2} \sum_t \|y_t - (\dots)\mathrm{vec}(U_2')\|_2^2$
    $U_3^{(k+1)} \leftarrow \arg\min_{U_3} \sum_t \|y_t - (\dots)\mathrm{vec}(U_3)\|_2^2$
    Update Core $\mathcal{G}$**:**
    $\mathcal{G}^{(k+1)} \leftarrow \arg\min_{\mathcal{G}} \sum_t \|y_t - (\dots)\mathrm{vec}(\mathcal{G}_{(1)})\|_2^2$
    Reconstruction: $\mathcal{A}^{(k+1)} \leftarrow \mathcal{G}^{(k+1)} \times_1 U_1^{(k+1)} \times_2 U_2^{(k+1)} \times_3 U_3^{(k+1)}$
until convergence
Finalization: Re-orthogonalize via HOSVD on $\hat{\mathcal{A}}$.
return $\hat{\mathcal{A}}_{MLR} = [\![\hat{\mathcal{G}}; \hat{U}_1, \hat{U}_2, \hat{U}_3]\!]$

Algorithm: ADMM (SHORR) - Part 1

Algorithm 2 ADMM for SHORR (Factors Update)

Init: $\mathcal{A}^{(0)}$ via Nuclear Norm. Get ranks (r_1, r_2, r_3).
repeat
    Update Factors (Sparse & Orthogonal)
    $U_1^{(k+1)} \leftarrow \arg\min_{U^T U = I} \{L(\dots) + \lambda \|U_1\|_1 \dots\}$
    $U_2^{(k+1)} \leftarrow \arg\min_{U^T U = I} \{L(\dots) + \lambda \|U_2\|_1 \dots\}$
    $U_3^{(k+1)} \leftarrow \arg\min_{U^T U = I} \{L(\dots) + \lambda \|U_3\|_1 \dots\}$
    Continued on next slide...
until End of loop (see next slide)

Note: Chaque mise à jour de facteur est un sous-problème non convexe résolu via linéarisation et projection sur l'espace de Stiefel. C'est la méthode SOC (splitting orthogonality constraint method)

Algorithm: ADMM (SHORR) - Part 2

Algorithm 3 ADMM for SHORR (Core & Duals)

...Continuation of the loop
Update Core (All-Orthogonal Constraint)
$\mathcal{G}^{(k+1)} \leftarrow \arg\min_{\mathcal{G}} \{L(\dots) + \sum \rho_i \|\mathcal{G}_{(i)} - D_i V_i' + \mathcal{C}_i\|_F^2\}$
Update Auxiliary (D, V) & Dual $(\mathcal{C})$ variables
for $i = 1, 2, 3$ **do**
    $D_i^{(k+1)} \leftarrow \arg\min_{D \text{ diag}} \|\mathcal{G}_{(i)}^{(k+1)} - D V_i^{(k)'} + (\mathcal{C}_i^{(k)})_{(i)}\|_F^2$
    $V_i^{(k+1)} \leftarrow \arg\min_{V'V=I} \|\mathcal{G}_{(i)}^{(k+1)} - D_i^{(k+1)} V' + (\mathcal{C}_i^{(k)})_{(i)}\|_F^2$
    $(\mathcal{C}_i^{(k+1)})_{(i)} \leftarrow (\mathcal{C}_i^{(k)})_{(i)} + \mathcal{G}_{(i)}^{(k+1)} - D_i^{(k+1)} V_i^{(k+1)'}$
end for
Reconstruction: $\mathcal{A}^{(k+1)} \dots$
convergence

ADMM Sub-problem: Projection on Stiefel

Solving for factors involves a specific quadratic problem with orthogonality constraints:

$$\min_{B^T B = I} \left\{ \|y - X\mathrm{vec}(B)\|_2^2 + \kappa \|B - \mathrm{Target}\|_F^2 \right\}$$

1. **Step 1: Unconstrained Solution**
   We solve the Ridge-like system to get an intermediate matrix G:

$$\mathrm{vec}(G) = \left(X^T X + \kappa I \right)^{-1} \times \mathrm{RHS}$$

2. **Step 2: Projection (Procrustes)**
   To enforce $B^T B = I$, we compute the SVD of G:

$$G = U \Sigma V^T \quad \Longrightarrow \quad B^* = U V^T$$

Rang

Contexte : Consistence du rank .

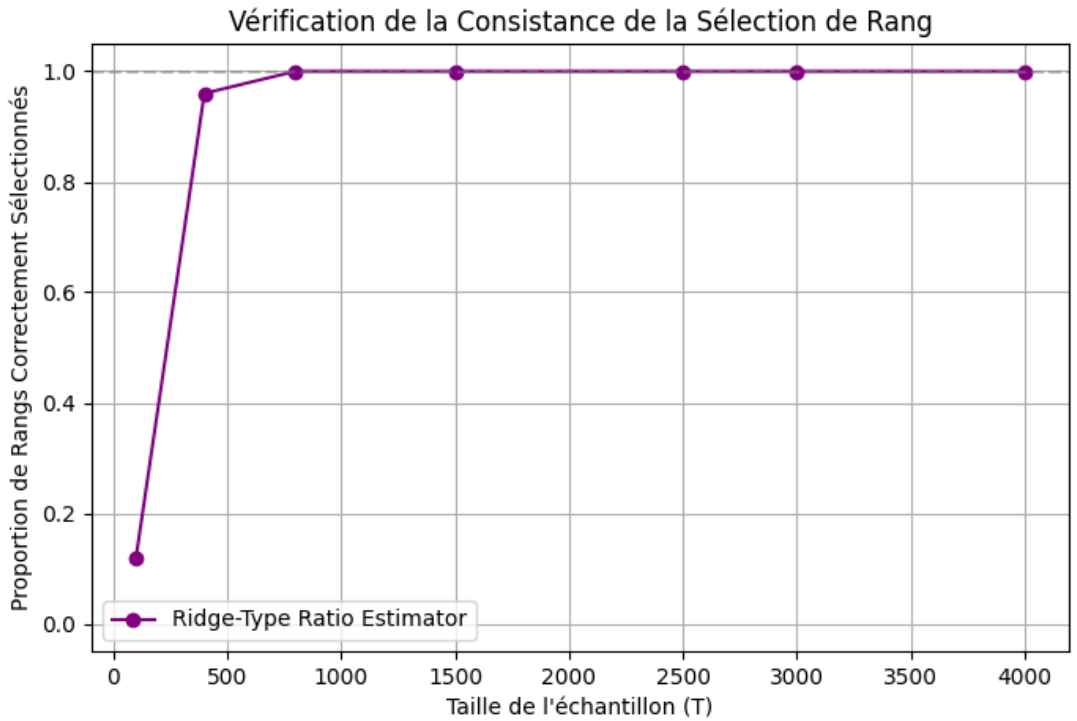

Figure: Vérification que les rangs sont bien estimés