Priors for the long run

Domenico Giannone
New York Fed

Michele Lenza
European Central Bank

Giorgio Primiceri
Northwestern University

NBP Workshop on Forecasting
Narodowy Bank Polski
Warsaw, 21-22 November 2016
What we do

- Propose a class of prior distributions for VARs that discipline the long-run implications of the model

Priors for the long run
What we do

- Propose a class of prior distributions for VARs that discipline the long-run implications of the model

Priors for the long run

- Properties
  - Based on macroeconomic theory
  - Conjugate → Easy to implement and combine with existing priors

- Perform well in applications
  - Good (long-run) forecasting performance
Outline

- A specific pathology of (flat-prior) VARs
  - Too much explanatory power of initial conditions and deterministic trends
  - Sims (1996 and 2000)

- Priors for the long run
  - Intuition
  - Specification and implementation

- Alternative interpretations and relation with the literature

- Application: macroeconomic forecasting
Simple example

- AR(1): \[ y_t = c + \rho y_{t-1} + \varepsilon_t \]
Simple example

■ AR(1):  
  
  \[ y_t = c + \rho y_{t-1} + \varepsilon_t \]

■ Iterate backwards:  
  
  \[ y_t = \rho^t y_0 + \sum_{j=0}^{t-1} \rho^j c + \sum_{j=0}^{t-1} \rho^j \varepsilon_{t-j} \]
**Simple example**

- **AR(1):**
  \[ y_t = c + \rho y_{t-1} + \varepsilon_t \]

- **Iterate backwards:**
  \[ y_t = \rho^t y_0 + \sum_{j=0}^{t-1} \rho^j c + \sum_{j=0}^{t-1} \rho^j \varepsilon_{t-j} \]

- **Model separates observed variation of the data into**
  - DC: deterministic component, predictable from data at time 0
  - SC: unpredictable/stochastic component
**Simple example**

- **AR(1):**
  \[ y_t = c + \rho y_{t-1} + \epsilon_t \]

- **Iterate backwards:**
  \[ y_t = \rho^t y_0 + \sum_{j=0}^{t-1} \rho^j c + \sum_{j=0}^{t-1} \rho^j \epsilon_{t-j} \]

  DC: deterministic component, predictable from data at time 0
  SC: unpredictable/stochastic component

- **If \( \rho = 1 \), DC is a simple linear trend:**
  \[ DC = y_0 + c \cdot t \]
Simple example

- **AR(1):**
  \[ y_t = c + \rho y_{t-1} + \varepsilon_t \]

- Iterate backwards:
  \[ y_t = \rho^t y_0 + \sum_{j=0}^{t-1} \rho^j c + \sum_{j=0}^{t-1} \rho^j \varepsilon_{t-j} \]
  - **DC:** deterministic component, predictable from data at time 0
  - **SC:** unpredictable/stochastic component

- If \( \rho = 1 \), DC is a simple linear trend:
  \[ DC = y_0 + c \cdot t \]

- Otherwise more complex:
  \[ DC = \frac{c}{1-\rho} + \rho^t \left( y_0 - \frac{c}{1-\rho} \right) \]
Pathology of (flat-prior) VARs (Sims, 1996 and 2000)

- OLS/MLE has a tendency to “use” the complexity of deterministic components to fit the low frequency variation in the data

- Possible because inference is typically conditional on $y_0$
  - No penalization for parameter estimates of implying steady states or trends far away from initial conditions
Deterministic components in VARs

- Problem more severe with VARs
  - implied deterministic component is much more complex than in AR(1) case
Deterministic components in VARs

- Problem more severe with VARs
  - implied deterministic component is much more complex than in AR(1) case

- Example: 7-variable VAR(5) with quarterly data on
  - GDP
  - Consumption
  - Investment
  - Real Wages
  - Hours
  - Inflation
  - Federal funds rate


- Flat or Minnesota prior
"Over-fitting" of deterministic components in VARs

- GDP
- Investment
- Hours
- Investment-to-GDP ratio
- Inflation
- Interest rate

Data | Flat
---|---

Giannone, Lenza, Primiceri

Priors for the long run
“Over-fitting” of deterministic components in VARs

GDP

Investment

Hours

Investment-to-GDP ratio

Inflation

Interest rate

Data  Flat  MN  PLR
Pathology of (flat-prior) VARs (Sims, 1996 and 2000)

- OLS/MLE has a tendency to “use” the complexity of deterministic components to fit the low frequency variation in the data.

- Possible because inference is typically conditional on $y_0$.
  - No penalization for parameter estimates of implying steady states or trends far away from initial conditions.

- Flat-prior VARs attribute an (implausibly) large share of the low frequency variation in the data to deterministic components.
Pathology of (flat-prior) VARs (Sims, 1996 and 2000)

- OLS/MLE has a tendency to “use” the complexity of deterministic components to fit the low frequency variation in the data

- Possible because inference is typically conditional on \( y_0 \)
  - No penalization for parameter estimates of implying steady states or trends far away from initial conditions

- Flat-prior VARs attribute an (implausibly) large share of the low frequency variation in the data to deterministic components

- Need a prior that downplays excessive explanatory power of initial conditions and deterministic component

- One solution: center prior on “non-stationarity”
Outline

- A specific pathology of (flat-prior) VARs
  - Too much explanatory power of initial conditions and deterministic trends
  - Sims (1996 and 2000)

- Priors for the long run
  - Intuition
  - Specification and implementation

- Alternative interpretations and relation with the literature

- Application: macroeconomic forecasting
Prior for the long run

\[ \text{VAR(1)}: \quad y_t = c + By_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0,\Sigma) \]
Prior for the long run

\[ \text{VAR}(1): \quad y_t = c + By_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma) \]

- Rewrite the VAR in terms of levels and differences:

\[
\Delta y_t = c + \Pi y_{t-1} + \varepsilon_t \\
\Pi = B - I
\]
Prior for the long run

\[ VAR(1): \quad y_t = c + By_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma) \]

- Rewrite the VAR in terms of levels and differences:

\[ \Delta y_t = c + \Pi y_{t-1} + \varepsilon_t \]
\[ \Pi = B - I \]

- Prior for the long run  \quad \rightarrow \quad \text{prior on } \Pi \text{ centered at 0}
Prior for the long run

VAR(1): \[ y_t = c + By_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0,\Sigma) \]

- Rewrite the VAR in terms of levels and differences:

\[ \Delta y_t = c + \Pi y_{t-1} + \varepsilon_t \]
\[ \Pi = B - I \]

- Prior for the long run:
  - Prior on \( \Pi \) centered at 0

- Standard approach (DLS, SZ, and many followers)
  - Push coefficients towards all variables being independent random random walks
Prior for the long run

\[ \Delta y_t = c + \Pi y_{t-1} + \varepsilon_t \]

- Rewrite as

\[ \Delta y_t = c + \Pi \underbrace{H^{-1}}_{\Lambda} \underbrace{Hy_{t-1}}_{\tilde{y}_{t-1}} + \varepsilon_t \]
Prior for the long run

$$\Delta y_t = c + \Pi y_{t-1} + \varepsilon_t$$

- Rewrite as

$$\Delta y_t = c + \Pi \left( H^{-1} \Lambda \tilde{y}_{t-1} \right) + \varepsilon_t$$

- Choose $H$ and put prior on $\Lambda$ conditional on $H$
Prior for the long run

\[ \Delta y_t = c + \Pi y_{t-1} + \varepsilon_t \]

- Rewrite as

\[ \Delta y_t = c + \Pi \left( H^{-1} \right)_\Lambda \left( H y_{t-1} \right)_{\tilde{y}_{t-1}} + \varepsilon_t \]

- Choose \( H \) and put prior on \( \Lambda \) conditional on \( H \)

- Economic theory suggests that some linear combinations of \( y \) are less (more) likely to exhibit long-run trends
Prior for the long run

\[ \Delta y_t = c + \Pi y_{t-1} + \epsilon_t \]

- Rewrite as

\[ \Delta y_t = c + \Pi \frac{H^{-1}}{\Lambda} H y_{t-1} + \epsilon_t \]

- Choose \( H \) and put prior on \( \Lambda \) conditional on \( H \)

- Economic theory suggests that some linear combinations of \( y \) are less (more) likely to exhibit long-run trends

- Loadings associated with these combinations are less (more) likely to be 0
Example: 3-variable VAR of KPSW

\[
\Delta y_t = c + \Pi H^{-1} H y_{t-1} + \varepsilon_t
\]

\[
\begin{bmatrix}
1 & 1 & 1 \\
-1 & 1 & 0 \\
-1 & 0 & 1
\end{bmatrix}
\begin{pmatrix}
\text{Output} \\
\text{Consumption} \\
\text{Investment}
\end{pmatrix}
\]
Example: 3-variable VAR of KPSW

\[ \Delta y_t = c + \Pi H^{-1}_\Lambda \left[ \begin{array}{c} H y_{t-1} + \varepsilon_t \\ \tilde{y}_{t-1} \end{array} \right] \]

\[ \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \]

Output
Consumption
Investment

\[
\begin{bmatrix}
\Delta x_t \\
\Delta c_t \\
\Delta i_t
\end{bmatrix} = c + \begin{bmatrix}
\Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\
\Lambda_{21} & \Lambda_{22} & \Lambda_{23} \\
\Lambda_{31} & \Lambda_{32} & \Lambda_{33}
\end{bmatrix} \begin{bmatrix}
x_{t-1} + c_{t-1} + i_{t-1} \\
c_{t-1} - x_{t-1} \\
i_{t-1} - x_{t-1}
\end{bmatrix} + \varepsilon_t
\]
Example: 3-variable VAR of KPSW

\[ \Delta y_t = c + \Pi H^{-1} \begin{bmatrix} \Lambda H y_{t-1} + \epsilon_t \end{bmatrix} \]

\[ \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \]

Output
Consumption
Investment

\[ \begin{bmatrix} \Delta x_t \\ \Delta c_t \\ \Delta i_t \end{bmatrix} = c + \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ \Lambda_{21} & \Lambda_{22} & \Lambda_{23} \\ \Lambda_{31} & \Lambda_{32} & \Lambda_{33} \end{bmatrix} \begin{bmatrix} x_{t-1} + c_{t-1} + i_{t-1} \\ c_{t-1} - x_{t-1} \\ i_{t-1} - x_{t-1} \end{bmatrix} + \epsilon_t \]

Possibly stationary linear combinations
Example: 3-variable VAR of KPSW

\[
\Delta y_t = c + \Pi H^{-1} \left[ \begin{array}{c} \Lambda \\ \hat{y}_{t-1} \end{array} \right] + \varepsilon_t
\]

\[
\begin{bmatrix}
1 & 1 & 1 \\
-1 & 1 & 0 \\
-1 & 0 & 1
\end{bmatrix}
\]

Output
Consumption
Investment

\[
\begin{bmatrix}
\Delta x_t \\
\Delta c_t \\
\Delta i_t
\end{bmatrix} = c + \begin{bmatrix}
\Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\
\Lambda_{21} & \Lambda_{22} & \Lambda_{23} \\
\Lambda_{31} & \Lambda_{32} & \Lambda_{33}
\end{bmatrix} \begin{bmatrix}
x_{t-1} + c_{t-1} + i_{t-1} \\
\end{bmatrix} + \varepsilon_t
\]

Common trend
Possibly stationary linear combinations
Example: 3-variable VAR of KPSW

\[ \Delta y_t = c + \frac{\boldsymbol{H}^{-1}}{\Lambda} H \bar{y}_{t-1} + \varepsilon_t \]

\[ \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \]

Output
Consumption
Investment

\[ \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & \Lambda_{13} \\ \Lambda_{21} & \Lambda_{22} & \Lambda_{23} \\ \Lambda_{31} & \Lambda_{32} & \Lambda_{33} \end{bmatrix} \]

\[ \begin{bmatrix} x_{t-1} + c_{t-1} + i_{t-1} \\ c_{t-1} - x_{t-1} \\ i_{t-1} - x_{t-1} \end{bmatrix} \]

Common trend
Possibly stationary linear combinations
Prior for the long run: specification and implementation

$$\Delta y_t = c + \Pi H^{-1} \underbrace{H y_{t-1}}_{\tilde{y}_{t-1}} + \varepsilon_t$$

- $\Lambda_i \mid H, \Sigma \sim N\left(0, \frac{\Sigma}{(H_i'y_0)^2}\right)$, \quad $i = 1, \ldots, n$
Prior for the long run: specification and implementation

\[ \Delta y_t = c + \Pi \left( H^{-1} \right) \left. H y_{t-1} \right| \tilde{y}_{t-1} + \varepsilon_t \]

- \( \Lambda \mid H, \Sigma \sim N \left( 0, \phi_i^2 \frac{\Sigma}{(H_i \cdot y_0)^2} \right), \quad i = 1, \ldots, n \)
Prior for the long run: specification and implementation

\[ \Delta y_t = c + \prod_{\Lambda} H^{-1}_{\Lambda} \underbrace{H y_{t-1}}_{\tilde{y}_{t-1}} + \varepsilon_t \]

- \( \Lambda_i \mid H, \Sigma \sim N \left( 0, \phi_i^2 \left( \frac{\Sigma}{(H_i y_0)^2} \right) \right), \quad i = 1, \ldots, n \)

- Conjugate
  - Can implement it with Theil mixed estimation in the VAR in levels
Prior for the long run: specification and implementation

\[ \Delta y_t = c + \Pi_H^{-1} \Lambda \tilde{y}_{t-1} + \varepsilon_t \]

- \( \Lambda_i \mid H, \Sigma \sim N \left( 0, \phi_i^2 \frac{\Sigma}{(H_i'y_0)^2} \right), \quad i = 1, \ldots, n \)

- Conjugate
  - Can implement it with Theil mixed estimation in the VAR in levels
  - Can be easily combined with existing priors
Prior for the long run: specification and implementation

\[ \Delta y_t = c + \prod_{\Lambda} H_t^{-1} H y_{t-1} + \varepsilon_t \]

- \[ \Lambda_{i} | H, \Sigma \sim N \left( 0, \phi_i^2 \frac{\Sigma}{(H_i y_0)^2} \right), \quad i = 1, \ldots, n \]

- Conjugate
  - Can implement it with Theil mixed estimation in the \textit{VAR in levels}
  - Can be easily combined with existing priors
  - Can compute the ML in closed form
    - Useful for hierarchical modeling and setting of hyperparameters \( \phi \) (GLP, 2013)
Connections and extreme cases

\[ \Delta y_t = c + \Pi \left[ H^{-1} \frac{H y_{t-1}}{\Lambda} \right] + \varepsilon_t \]

- Rewrite as

\[ \Delta y_t = c + \left[ \Lambda_1 \quad \Lambda_2 \right] \begin{bmatrix} \beta_{t-1} \\ \beta_t \end{bmatrix}' y_{t-1} + \varepsilon_t \]
Connections and extreme cases

\[ \Delta y_t = c + \Pi \Lambda H^{-1} \begin{pmatrix} H & y_{t-1} \end{pmatrix} + \varepsilon_t \]

- Rewrite as

\[ \Delta y_t = c + \begin{bmatrix} \Lambda_1 & \Lambda_2 \end{bmatrix} \begin{bmatrix} \beta_\perp' \\
\beta' \end{bmatrix} y_{t-1} + \varepsilon_t \]

\[ \Delta y_t = c + \Lambda_1 \beta_\perp' y_{t-1} + \Lambda_2 \beta' y_{t-1} + \varepsilon_t \]
Connections and extreme cases

\[ \Delta y_t = c + \Lambda_1 \beta_1' y_{t-1} + \Lambda_2 \beta_2' y_{t-1} + \epsilon_t \]
Connections and extreme cases

\[ \Delta y_t = c + \Lambda_1 \beta_1' y_{t-1} + \Lambda_2 \beta_2' y_{t-1} + \varepsilon_t \]

- Error Correction Model: dogmatic prior on \( \Lambda_1 = 0 \)
Connections and extreme cases

\[ \Delta y_t = c + \Lambda_1 \beta'_1 y_{t-1} + \Lambda_2 \beta'_2 y_{t-1} + \varepsilon_t \]

- Error Correction Model: dogmatic prior on \( \Lambda_1 = 0 \)

- KPSW, CEE
  - fix \( \beta \) based on theory
  - flat prior on \( \Lambda_2 \)
Connections and extreme cases

\[ \Delta y_t = c + \Lambda_1 \beta' y_{t-1} + \Lambda_2 \beta' y_{t-1} + \varepsilon_t \]

- Error Correction Model: dogmatic prior on \( \Lambda_1 = 0 \)
  - KPSW, CEE
    - fix \( \beta \) based on theory
    - flat prior on \( \Lambda_2 \)
  - Cointegration
    - estimate \( \beta \)
    - flat prior on \( \Lambda_2 \)
    - EG (1987)
Connections and extreme cases

\[ \Delta y_t = c + \Lambda_1 \beta'_{\perp} y_{t-1} + \Lambda_2 \beta' y_{t-1} + \varepsilon_t \]

- Error Correction Model: dogmatic prior on \( \Lambda_1 = 0 \)

- **KPSW, CEE**
  - fix \( \beta \) based on theory
  - flat prior on \( \Lambda_2 \)

- **Cointegration**
  - estimate \( \beta \)
  - flat prior on \( \Lambda_2 \)
  - EG (1987)

- **Bayesian cointegration**
  - uniform prior on \( sp(\beta) \)
  - KSvDV (2006)
Connections and extreme cases

\[ \Delta y_t = c + \Lambda_1 \beta'_\perp y_{t-1} + \Lambda_2 \beta' y_{t-1} + \varepsilon_t \]

- Error Correction Model: dogmatic prior on \( \Lambda_1 = 0 \)
  
  - KPSW, CEE
    - fix \( \beta \) based on theory
    - flat prior on \( \Lambda_2 \)
  
  - Cointegration
    - estimate \( \beta \)
    - flat prior on \( \Lambda_2 \)
    - EG (1987)

  - Bayesian cointegration
    - uniform prior on \( \text{sp}(\beta) \)
    - KStV (2006)

- VAR in first differences: dogmatic prior on \( \Lambda_1 = \Lambda_2 = 0 \)
Connections and extreme cases

\[ \Delta y_t = c + \Lambda_1 \beta' y_{t-1} + \Lambda_2 \beta' y_{t-1} + \epsilon_t \]

- Error Correction Model: dogmatic prior on \( \Lambda_1 = 0 \)
  - KPSW, CEE
    - fix \( \beta \) based on theory
    - flat prior on \( \Lambda_2 \)
  - Cointegration
    - estimate \( \beta \)
    - flat prior on \( \Lambda_2 \)
    - EG (1987)
  - Bayesian cointegration
    - uniform prior on \( \text{sp}(\beta) \)
    - KSwDV (2006)

- VAR in first differences: dogmatic prior on \( \Lambda_1 = \Lambda_2 = 0 \)

- Sum-of-coefficients prior (DLS, SZ)
  - \[ [ \beta' \beta' ]' = H = I \]
  - shrink \( \Lambda_1 \) and \( \Lambda_2 \) to 0
Empirical results

- Deterministic component in 7-variable VAR

- Forecasting
  - 3-variable VAR
  - 5-variable VAR
  - 7-variable VAR
Empirical results

- Deterministic component in 7-variable VAR
Empirical results

- Deterministic component in 7-variable VAR
  - GDP, Consumption, Investment, Real Wages, Hours, Inflation, Interest Rate
Empirical results

- Deterministic component in 7-variable VAR
  - GDP, Consumption, Investment, Real Wages, Hours, Inflation, Interest Rate

\[
\mathbf{H} = \begin{bmatrix}
Y & C & I & W & H & \pi & R \\
1 & 1 & 1 & 1 & 0 & 0 & 0 \\
-1 & 1 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 1 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 \\
\end{bmatrix}
\]

Interpretation of \( \mathbf{H} \mathbf{y} \)
- Real trend
- Consumption-to-GDP ratio
- Investment-to-GDP ratio
- Labor share
- Hours
- Real interest rate
- Nominal trend
Empirical results

- **Deterministic component in 7-variable VAR**
  - GDP, Consumption, Investment, Real Wages, Hours, Inflation, Interest Rate

- **Forecasting**
  - 3-variable VAR
  - 5-variable VAR
  - 7-variable VAR

- **\( H = \)**

\[
\begin{bmatrix}
Y & C & I & W & H & \pi & R \\
1 & 1 & 1 & 1 & 0 & 0 & 0 \\
-1 & 1 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 1 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 \\
\end{bmatrix}
\]

**Interpretation of \( Hy \)**
- Real trend
- Consumption-to-GDP ratio
- Investment-to-GDP ratio
- Labor share
- Hours
- Real interest rate
- Nominal trend
Empirical results

Deterministic component in 7-variable VAR
- GDP, Consumption, Investment, Real Wages, Hours, Inflation, Interest Rate

Forecasting
- 3-variable VAR
- 5-variable VAR
- 7-variable VAR

\[
\begin{bmatrix}
Y & C & I & W & H & \pi & R \\
1 & 1 & 1 & 1 & 1 & 0 & 0 \\
-1 & 1 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 1 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 \\
\end{bmatrix}
\]

Interpretation of \( H y \)
- Real trend
- Consumption-to-GDP ratio
- Investment-to-GDP ratio
- Labor share
- Hours
- Real interest rate
- Nominal trend
Empirical results

- **Deterministic component in 7-variable VAR**
  - GDP, Consumption, Investment, Real Wages, Hours, Inflation, Interest Rate

- **Forecasting**
  - 3-variable VAR
  - 5-variable VAR
  - 7-variable VAR

\[ H = \begin{bmatrix}
Y & C & I & W & H & \pi & R \\
1 & 1 & 1 & 1 & 0 & 0 & 0 \\
-1 & 1 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & 1 & 0 & 0 & 0 & 0 \\
-1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 \\
\end{bmatrix} \]

**Interpretation of** \( H y \)
- \( Y \rightarrow \) Real trend
- \( C \rightarrow \) Consumption-to-GDP ratio
- \( I \rightarrow \) Investment-to-GDP ratio
- \( W \rightarrow \) Labor share
- \( H \rightarrow \) Hours
- \( \pi \rightarrow \) Real interest rate
- \( R \rightarrow \) Nominal trend
Deterministic components in VARs

- GDP
- Investment
- Hours
- Investment-to-GDP ratio
- Inflation
- Interest rate

Data | Flat | MN | PLR
Deterministic components in VARs with Prior for the Long Run

- GDP
- Investment
- Hours
- Investment-to-GDP ratio
- Inflation
- Interest rate

Data
Flat
MN
PLR

Giannone, Lenza, Primiceri

Priors for the long run
Forecasting results with 3-, 5- and 7-variable VARs

- Recursive estimation starts in 1955:I
3-variable VAR: MSFE (1985-2013)

![Graphs showing forecast error variance for different variables and combinations over quarters ahead.]

Legend:
- **MN**
- **SZ**
- **Naive**
- **PLR**
3-variable VAR: MSFE (1985-2013)
3-var VAR: Mean Squared Forecast Errors (1985-2013)

The graphs depict the Mean Squared Forecast Errors (MSFE) for different variables and combinations of variables over 40 quarters ahead, with the error rates increasing as the forecast horizon extends. The variables include Y, C, I, Y + C + I, C - Y, and I - Y. Different lines represent different priors: MN (blue), SZ (green), Naive (purple), and PLR (red).
Consumption- and Investment-to-GDP ratios

C - Y

I - Y

Actual
Forecasts (5 years ahead)

C - Y

I - Y

Actual
Naive
Forecasts (5 years ahead)

C - Y

I - Y

Actual  Naive  PLR
A bivariate example

\[ \Delta y_t = c + \prod H^{-1} \left[ \Lambda H_{t-1} \right] \tilde{y}_{t-1} + \varepsilon_t \]

- Simple example, VAR(1) for output \((x_{\downarrow t})\) and investment \((i_{\downarrow t})\)

- SOC prior corresponds to \(\text{vec}(\pi)|\Sigma \sim N(0, [\mu_2/x_{02} & 0 @ 0 & 0]

- Economic theory: \((i_{\downarrow t} + x_{\downarrow t})\) is a common trend while \((i_{\downarrow t} - x_{\downarrow t})\) is stationary

- What do the SOC and the PLR priors imply for these linear combinations of the variables?
A bivariate example

- The SOC prior implies \( \text{vec}(\Lambda)|H,\Sigma \sim N(0, [\mu \mu / x, \mu / i, \mu / x, \mu / i] \otimes \Sigma) \)

- “Same” shrinkage for linear combinations that are more likely to be trends or stationary!

- Instead the PLR prior implies \( \text{vec}(\Lambda)|H,\Sigma \sim N(0, [\phi \phi / (i, x, 0) - 0, 0, \phi / (i, x, 0)] \otimes \Sigma) \)

- Idea: \( (i, x) \) is likely to be a “much larger number” than \( (i \) - \( x) \), much less shrinkage on the loadings of combinations more likely to be stationary!
5-variable VAR: MSFE (1985-2013)
7-variable VAR: MSFE (1985-2013)
Nominal Neutrality
Invariance to rotations of the “stationary” space

- Our baseline prior depends on the choice of a specific $H$ matrix

$$H = [\begin{bmatrix} \beta \downarrow \downarrow \uparrow \uparrow & @ \beta \uparrow \uparrow \end{bmatrix}$$
Our baseline prior depends on the choice of a specific $H$ matrix

$$H = \begin{bmatrix} \beta_{\bot\bot} \, \beta' @ \beta' \end{bmatrix}$$

Economic theory is useful, but not sufficient to uniquely pin down $H$

- Macro models are typically informative about $\beta_{\bot\bot}$ and $sp(\beta)$
Invariance to rotations of the “stationary” space

- Our baseline prior depends on the choice of a specific $H$ matrix
  
  $H = [\begin{bmatrix} \beta_{\perp} & \beta_{\uparrow} \end{bmatrix}]

- Economic theory is useful, but not sufficient to uniquely pin down $H$
  
  - Macro models are typically informative about $\beta_{\perp}$ and $sp(\beta)$

- Extension of our PLR that is invariant to rotations of $\beta$
Invariance to rotations of the “stationary” space

- Our baseline prior depends on the choice of a specific $H$ matrix
  $H = \begin{bmatrix} \beta \perp \ \beta \parallel \end{bmatrix}$

- Economic theory is useful, but not sufficient to uniquely pin down $H$:
  - Macro models are typically informative about $\beta \perp$ and $sp(\beta)$

- Extension of our PLR that is invariant to rotations of $\beta$

Baseline PLR: $\Lambda \downarrow i \cdot (H \downarrow i \cdot y \downarrow 0) | H, \Sigma \sim N(0, \phi \downarrow i \parallel 2 \Sigma), \quad i = 1, \ldots, n$
Invariance to rotations of the “stationary” space

- Our baseline prior depends on the choice of a specific $H$ matrix
  \[ H = [\begin{bmatrix} \beta \downarrow \downarrow \uparrow \& \beta \uparrow \uparrow \end{bmatrix} ] \]

- Economic theory is useful, but not sufficient to uniquely pin down $H$
  - Macro models are typically informative about $\beta \downarrow \downarrow$ and $sp(\beta)$

- Extension of our PLR that is invariant to rotations of $\beta$

Baseline PLR: \[ \Lambda \downarrow i \cdot (H \downarrow i \cdot y \downarrow 0 ) | H, \Sigma \sim N(0, \phi \downarrow i \uparrow 2 \Sigma), \quad i=1,...,n \]

Invariant PLR: \[ \{ \Lambda \downarrow i \cdot (H \downarrow i \cdot y \downarrow 0 ) | H, \Sigma \sim N(0, \phi \downarrow i \uparrow 2 \Sigma), \quad i=1,...,n \}
- \[ -r \sum_{i=n-r+1}^{n} \Lambda \downarrow i \cdot (H \downarrow i \cdot y \downarrow 0 ) | H, \Sigma \sim N(0, \phi \downarrow n-r+1 \uparrow 2 \Sigma) \]
7-variable VAR: Forecasting results with “invariant” PLR

[Graphs showing the comparison between PLR baseline and PLR invariant for different variables (Y, C, I, H, π, R) with MSFE over quarters ahead (0 to 40).]

Giannone, Lenza, Primiceri

Priors for the long run
$H_y$ in the data
7-variable VAR: Forecasting results with “invariant” PLR

- **Y**
- **C**
- **I**
- **H**
- **π**
- **R**

<table>
<thead>
<tr>
<th>MSFE</th>
<th>0</th>
<th>0.002</th>
<th>0.004</th>
<th>0.006</th>
<th>0.008</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Y</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>C</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>I</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>H</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>π</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>R</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PLR baseline

PLR invariant

PLR invariant (except C-Y)
Strengths and weaknesses

- Strengths
  - Imposes discipline on long-run behavior of the model
  - Based on robust lessons of theoretical macro models
  - Performs well in forecasting (especially at longer horizons)
  - Very easy to implement
Strengths

- Imposes discipline on long-run behavior of the model
- Based on robust lessons of theoretical macro models
- Performs well in forecasting (especially at longer horizons)
- Very easy to implement

“Weak” points

- Non-automatic procedure → need to think about it
- Might prove difficult to set up in large-scale models → might require too much thinking