Priors for the long run

Domenico Giannone
New York Fed

Michele Lenza
European Central Bank

Giorgio Primiceri
Northwestern University

9th ECB Workshop on Forecasting Techniques
June 3, 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):**
  \[ y_t = c + \rho y_{t-1} + \epsilon_t \]

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

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

  \[ \text{DC} \]

  \[ \text{SC} \]

- Model separates observed variation of the data into
  - 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

- **Model separates observed variation of the data into**
  - 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  MN  PLR
“Over-fitting” of deterministic components in VARs

DATA

Flat

MN

PLR

Giannone, Lenza, Primiceri

Priors for the long run
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
VAR(1): \( y_t = c + By_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0,\Sigma) \)
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

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

- 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

\[ \text{prior on } \Pi \text{ centered at 0} \]
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  
  \[ \text{prior on } \Pi \text{ centered at 0} \]

- 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

\[ \Delta y_t = c + \Pi \left( H^{-1} \right) \Lambda \left( H y_{t-1} \right) + \varepsilon_t \]
Prior for the long run

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

- Rewrite as

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

- 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

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

- 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

\[ \Delta y_t = c + \Pi \frac{H^{-1}}{\Lambda} \frac{Hy_{t-1}}{\tilde{y}_{t-1}} + \epsilon_t \]

- 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

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

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

Output
Consumption
Investment
Example: 3-variable VAR of KPSW

\[ \Delta y_t = c + \prod \left[ H^{-1} \right] \Lambda \left[ H_y_{t-1} \right] + \epsilon_t \]

\[ \left[ \begin{array}{c} 1 \\ -1 \\ -1 \end{array} \right] \quad \text{Output} \]
\[ \left[ \begin{array}{c} 1 \\ 1 \\ 0 \end{array} \right] \quad \text{Consumption} \]
\[ \left[ \begin{array}{c} 1 \\ 0 \\ 1 \end{array} \right] \quad \text{Investment} \]

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

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

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

Output Consumption Investment

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

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

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

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

Output
Consumption
Investment

Common trend

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

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

\[
\Delta y_t = c + \Pi H^{-1} \cdot H\hat{y}_{t-1} + \varepsilon_t
\]

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

Output

Consumption

Investment

Common trend

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

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

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

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

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

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

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

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

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

\[ \Delta y_t = c + \Pi H^{-1} H \hat{y}_{t-1} + \epsilon_t \]

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

- Conjugate
  - Can implement it with Theil mixed estimation in the VAR in levels
  - Can be easily combined with existing priors

\[ \Delta y_t = c + \Pi H^{-1} H \hat{y}_{t-1} + \epsilon_t \]
Prior for the long run: specification and implementation

\[ \Delta y_t = c + \Pi \underbrace{H^{-1}}_{\Lambda} \underbrace{Hy_{t-1}}_{\tilde{y}_{t-1}} + \epsilon_t \]

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

Conjugate

- Can implement it with Theil mixed estimation in the 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)
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

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

Interpretation of \(Hy\)
- \(\rightarrow\) Real trend
- \(\rightarrow\) Consumption-to-GDP ratio
- \(\rightarrow\) Investment-to-GDP ratio
- \(\rightarrow\) Labor share
- \(\rightarrow\) Hours
- \(\rightarrow\) Real interest rate
- \(\rightarrow\) 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

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

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 \( H y \)

- Real trend
- Consumption-to-GDP ratio
- Investment-to-GDP ratio
- Labor share
- Hours
- Real interest rate
- 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
Forecasting results with 3-, 5- and 7-variable VARs

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

The graphs depict the Mean Squared Forecast Error (MSFE) for different combinations of variables over 40 quarters ahead, from 1985 to 2013. The variables include Y (output), C (consumption), and I (investment). The plots show the MSFE for each variable and their combinations, with different priors represented by different lines: MN, SZ, Naive, and PLR.
3-variable VAR: MSFE (1985-2013)
Consumption- and Investment-to-GDP ratios

C - Y

I - Y

Actual

Giannone, Lenza, Primiceri

Priors for the long run
Forecasts (5 years ahead)

C - Y

Actual

I - Y

Actual

Naive
Forecasts (5 years ahead)
5-variable VAR: MSFE (1985-2013)
7-variable VAR: MSFE (1985-2013)
Invariance to rotations of the “stationary” space

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

$$H = \begin{bmatrix} \beta' \\ \beta' \end{bmatrix}$$
Invariance to rotations of the “stationary” space

- Our baseline prior depends on the choice of a specific $H$ matrix
  \[ H = \begin{bmatrix} \beta_1' \\ \beta_2' \end{bmatrix} \]

- Economic theory is useful, but not sufficient to uniquely pin down $H$
  - Macro models are typically informative about $\beta_1$ and $sp(\beta)$
Invariance to rotations of the “stationary” space

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

$$H = \begin{bmatrix} \beta' \\ \beta' \end{bmatrix}$$

- 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
  \[ H = \begin{bmatrix} \beta_1' \\ \beta' \end{bmatrix} \]

- 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: \[ \Lambda_i \cdot (H_i.\bar{y}_0)|H, \Sigma \sim N(0, \phi_i^2 \Sigma), \quad i = 1, \ldots, n \]
Invariance to rotations of the “stationary” space

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

- 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: $\Lambda_i \cdot (H_i.\bar{y}_0)|H, \Sigma \sim N(0, \phi_i^2 \Sigma), \quad i = 1, ..., n$

Invariant PLR:
\[
\begin{cases}
\Lambda_i \cdot (H_i.\bar{y}_0)|H, \Sigma \sim N(0, \phi_i^2 \Sigma), & i = 1, ..., n - r \\
\sum_{i=n-r+1}^{n} \Lambda_i \cdot (H_i.\bar{y}_0)|H, \Sigma \sim N(0, \phi_{n-r+1}^2 \Sigma)
\end{cases}
\]
7-variable VAR: Forecasting results with “invariant” PLR

- **Y**
  - MSFE: 0.002 to 0.01
  - Graph shows MSFE increasing with time horizon.

- **C**
  - MSFE: 0.002 to 0.01
  - Graph shows MSFE increasing with time horizon.

- **I**
  - MSFE: 0.0005 to 0.02
  - Graph shows MSFE increasing with time horizon.

- **H**
  - MSFE: 0.0005 to 0.002
  - Graph shows MSFE increasing with time horizon.

- **π**
  - MSFE: 1e-05 to 2e-05
  - Graph shows MSFE increasing with time horizon.

- **R**
  - MSFE: 2e-05 to 1e-04
  - Graph shows MSFE increasing with time horizon.

**Key:**
- Red line: PLR baseline
- Dotted line: PLR invariant

**Legend:**
- Quarters Ahead
- MSFE: Mean Squared Forecast Error

**Notes:**
- The graphs illustrate the forecasting performance of different variables using PLR baseline and PLR invariant priors.
- The MSFE values are shown for different time horizons (0 to 40 quarters).

**References:**
- Giannone, Lenza, Primiceri
- Priors for the long run
Hy in the data

- C-Y
- I-Y
- H
- W-Y
- R+ π
- R- π

Giannone, Lenza, Primiceri

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

- **“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
Connections and extreme cases

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

- Rewrite as

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

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

- Rewrite as

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

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

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

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

- Error Correction Model: dogmatic prior on \( \Lambda_1 = 0 \)
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 \( \Lambda_2 \)
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)
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)

- **Bayesian cointegration**
  - uniform prior on \( sp(\beta) \)
  - KSvDV (2006)
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)
  - Bayesian cointegration
    - uniform prior on $\text{sp}(\beta)$
    - KSvDV (2006)

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

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

- 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 \( \text{sp}(\beta) \)
    - KSvDV (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
3-var VAR: Mean Squared Forecast Errors (1985-2013)