A Minimal Econometric Interpretation for Monetary and Fiscal Policy Interactions

Job Market Paper

Fei Tan, Indiana University

Manuscript prepared for Fall 2013 Econometrics Workshop
A Road Map...

1. Introduction

2. Structural- & Reduced-Form Models
   - A Prototypical NK Model
   - A Simple Trivariate VAR
   - Marriage of DSGE and VAR

3. Minimal Econometric Interpretation
   - Observational Equivalence of U.S. Policy Mix
   - Some Statistical Evidence of Fiscal Theory

4. Concluding Remarks
Monetary & Fiscal Interactions Matter A Lot

- Canonical macro models: to allow monetary (M) policy’s control of inflation, fiscal (F) backing is always forthcoming [Theory-M].
  - M can and does control inflation by obeying Taylor-type rule;
  - F can and does ensure solvency by making fiscal adjustment;
- This modeling convention makes sense in normal times but ignores how “profligate” F undermines “good” M in abnormal episodes.
  - Bond price-support during wartime: U.S. M policy from 1942 until Treasury-Fed “Accord” of 1951;
  - Recent crisis: M is at or near zero lower bound and F is bouncing between stimulus and austerity.
- Competing story: M pins down price level and F maintains value of government liabilities [Theory-F].
- Two theories are observationally equivalent in theory and identification remains BIG issue for central banks [Tan & Walker (2012)].
Empirically relevant as both stories, Theory-F gets rejected against -M by model fit when embedded and estimated in DSGE’s.

DSGE’s are designed to mimic certain aspects of reality and model selection depends crucially on nature of econometric interpretation.

- strong interpretation: DSGE’s fare badly in providing full probabilistic characterization of observables;
- weak interpretation: DSGE’s fare not too badly in summarizing specified sample moments [Kydland & Prescott (1996)];
- minimal interpretation: DSGE’s fare well in summarizing specified population moments [DeJong et al. (1996) & Geweke (2007)].

False implications from Theory-M and -F DSGE’s lead to misleading model selection if interpreted with strong/weak interpretations.

In practice, how are expectations of M-F interactions anchored under minimal econometric interpretation? [This Work]
Main Results

► We identify a strong asymmetry in model fit between Theory-M and -F under strong econometric interpretation.

► We provide a minimal econometric interpretation for U.S. M-F policy interactions and some statistical evidence of Theory-F.
  ▶ small-scale DSGE model with two decoupled determinacy regions, each indexed by a policy theory, is merged with a VAR;
  ▶ we document quantitative importance of alternative policy theories and characterize their model misspecifications;
  ▶ Theory-M and -F are shown to be equally consistent with U.S. macro data under minimal econometric interpretation;
  ▶ intertemporal comparisons of model fit even suggest minor statistical evidence of Theory-F, especially during recent crisis.

► At more general level, we ought to think of monetary and fiscal policies jointly in macro-policy discussions.
Structural Model: Linearized DSGE

- Generic linear representation of DSGE has form of

\[
\Gamma_0 s_{t+1} = \Gamma_1 s_t + \Psi v_{t+1} + \Pi \eta_{t+1}
\]  

(1)

- Model setup follows standard NK framework. In what follows, we focus on monetary and fiscal policy specifications.

- Linearized monetary policy rule \([\psi^\pi > 1 \text{ or } \psi^\pi < 1]\)

\[
\hat{R}_t = \rho^M \hat{R}_{t-1} + (1 - \rho^M) \psi^\pi \hat{\pi}_t + (1 - \rho^M) \psi^y \hat{y}_t + \theta^M_t
\]  

(2)

- Linearized fiscal policy rule \([\delta^b \approx 0 \text{ or } \delta^b > 0]\)

\[
\hat{\tau}_t = \rho^F \hat{\tau}_{t-1} + (1 - \rho^F) \delta^b b_{t-1} + \psi^F_t
\]  

(3)

- Our model is driven by 4 structural shocks, \(Z - \hat{z}_t\), \(M - \theta^M_t\), \(F - \psi^F_t\), and \(G - \hat{g}_t\), all following AR(1) processes.
Unobservables, Observables, Innovations, and Parameters

- 7 unobservables collected in vector $s_t$:

  $$ s_t = [\hat{y}_t, \hat{\pi}_t, \hat{R}_t, \hat{b}_t, \hat{\tau}_t, \hat{g}_t, \hat{z}_t]' $$

- 3 observables collected in vector $y_t$:

  $$ y_t = [YGR_t, INF_t, INT_t]' $$

- 0 measurement errors and 4 innovations collected in vector $\varepsilon_t$:

  $$ \varepsilon_t = [\varepsilon^z_t, \varepsilon^M_t, \varepsilon^F_t, \varepsilon^G_t]' $$

- 18 DSGE structural parameters collected in vector $\theta$. 
Conventional modeling: raising interest rate is anti-inflationary and fiscal disturbance has no effect on economy [2nd & 3rd columns].
Fiscal theory of price level: raising interest rate is inflationary and fiscal disturbance does affect economy [2nd & 3rd columns].
**Strong Asymmetry in Model Fit**

<table>
<thead>
<tr>
<th>Name</th>
<th>Domain</th>
<th>Density</th>
<th>Para (1)</th>
<th>Para (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\psi^\pi$</td>
<td>$[0, 1)/(1, \infty)$</td>
<td>Beta/Gamma</td>
<td>1.50/0.50</td>
<td>0.25</td>
</tr>
<tr>
<td>$\delta^b$</td>
<td>$\mathbb{R}^+$</td>
<td>Normal(+)</td>
<td>0.20/0.0001</td>
<td>0.05/0.005</td>
</tr>
</tbody>
</table>

**A. Priors [M/F]**

**B. Model Fits**

<table>
<thead>
<tr>
<th>DGP</th>
<th>M-Fit</th>
<th>F-Fit</th>
<th>Bayes</th>
<th>Jefferey</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>$-152.9$</td>
<td>$-258.9$</td>
<td>$106.3$</td>
<td>$&gt;5$</td>
</tr>
<tr>
<td>F</td>
<td>$-342.0$</td>
<td>$-332.9$</td>
<td>$9.2$</td>
<td>$&gt;5$</td>
</tr>
<tr>
<td>55-79</td>
<td>$-511.2$</td>
<td>$-520.6$</td>
<td>$9.4$</td>
<td>$&gt;5$</td>
</tr>
<tr>
<td>82-07</td>
<td>$-401.2$</td>
<td>$-411.2$</td>
<td>$10.0$</td>
<td>$&gt;5$</td>
</tr>
<tr>
<td>08-12</td>
<td>$-103.9$</td>
<td>$-107.0$</td>
<td>$3.1$</td>
<td>$&lt;5$</td>
</tr>
<tr>
<td>55-12</td>
<td>$-1133.0$</td>
<td>$-1147.3$</td>
<td>$14.3$</td>
<td>$&gt;5$</td>
</tr>
</tbody>
</table>

- Misspecified DGP may even replicate true DGP and contaminate model comparison and selection [1st & 2nd rows].
- Strong asymmetry in model fit and actual data indicate true DGP lies closer to Theory-F in model space.
- Need to quantify degrees of model misspecification.
Three Econometric Interpretations of DSGE’s

- **Strong econometric interpretation:** DSGE’s provide *ex ante* predictive distribution $p(y|\theta_D, D)$ for observables $y$.
  - examples: maximum likelihood, Bayesian, etc.;
  - problems: DSGE’s predict exact relations not found in data.

- **Weak econometric interpretation:** DSGE’s provide *ex ante* predictive distribution $p(z|\theta_D, D)$ for selected sample moments $z = f(y)$.
  - examples: calibration, prior predictive analysis, etc.;
  - problems: $p(z|D)$ contains additional unrealistic model dynamics.

- **Minimal econometric interpretation:** DSGE’s provide prior distribution $p(m|D)$ for selected population moments $m = \mathbb{E}(z|\theta_D, D)$.
  - implementation: link up $m$ and $y$ by an econometric model;
  - application: equity premium puzzle dissipates [Geweke (2007)].

- In what follows, we adopt minimal econometric interpretation for M-F policy interactions by marrying DSGE to VAR.
A less restrictive representation of $y_t$ is quarterly trivariate VAR

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + u_t$$

(7)

where $u_t \sim_d N(0, \Sigma_u)$ is vector of one-step-ahead forecast error.

VAR parameterizations are not parsimonious and data availability poses serious constraint on its specification, if not overfitting data.

Solution: constrain parameter estimates toward specific point in parameter space by Bayesian shrinkage estimators, e.g. Minnesota priors [Doan et al. (1984)] and DSGE priors [DeJong et al. (1993)].

To elicit minimal econometric interpretation, we implant DSGE prior on $m$ in VAR likelihood for Theory-M and -F, respectively. This also shifts computational burden from DSGE to VAR.
Likelihood Function

Let $T$ be actual sample size and $n$ be number of observables. Write VAR compactly as $Y = X\Phi + U$.

VAR likelihood function is given by

$$p(Y|\Phi, \Sigma_u, y_0, \ldots, y_{1-p}) = \prod_{t=1}^{T} p(y_t|\Phi, \Sigma_u, y_{t-1}, \ldots, y_{t-p})$$

$$= \prod_{t=1}^{T} \frac{1}{(2\pi)^{n/2} |\Sigma_u|^{-1/2}} \exp \left[ -\frac{1}{2} (y_t - \Phi'x_t)'\Sigma_u^{-1}(y_t - \Phi'x_t) \right]$$

$$\propto |\Sigma_u|^{-T/2} \exp \left\{ -\frac{1}{2} \text{trace} [\Sigma_u^{-1}(Y'Y - \Phi'X'Y - Y'X\Phi + \Phi'X'X\Phi)] \right\}$$

(8)

VAR provides less restrictive representation of same set of observables relative to DSGE.
Let actual sample be augmented with $T^* = \lambda T$ artificial observations $(Y^*, X^*)$ from DSGE conditional on $\theta$.

VAR likelihood for artificial sample is given by

$$p(Y^*(\theta)|\Phi, \Sigma_u, y_0^*, \ldots, y_{1-p}^*)$$

$$\propto |\Sigma_u|^{-\lambda T/2} \exp \left\{ -\frac{1}{2} \text{tr}[\Sigma_u^{-1}(Y^*Y^* - \Phi'X^*Y^* - Y^*X^*\Phi + \Phi'X^*X^*\Phi)] \right\}$$

(9)

Remarks on $p(Y^*(\theta)|\Phi, \Sigma_u, y_0^*, \ldots, y_{1-p}^*)$:

- It can be interpreted as a prior density for $(\Phi, \Sigma_u)$;
- It summarizes information of VAR parameters contained in artificial sample.
- Drawback: repeated application of this prior leads to stochastic variation.
Prior Distribution (Cont’d)

▶ We replace nonstandardized sample moments \((Y^* Y^*, Y^* X^*, X^* X^*)\) by scaled population moments

\[
\begin{align*}
\mathbb{E}_\theta [Y^* Y^*] &= \lambda T \mathbb{E}_\theta [y_t^* y_t^*] := \lambda T \Gamma_{yy}(\theta) \\
\mathbb{E}_\theta [Y^* X^*] &= \lambda T \mathbb{E}_\theta [y_t^* x_t^*] := \lambda T \Gamma_{yx}(\theta) \\
\mathbb{E}_\theta [X^* X^*] &= \lambda T \mathbb{E}_\theta [x_t^* x_t^*] := \lambda T \Gamma_{xx}(\theta)
\end{align*}
\]

▶ After replacement, we obtain DSGE prior given by

\[
p(\Phi, \Sigma_u|\theta) = c(\theta)^{-1} |\Sigma_u|^{-\frac{\lambda T + n + 1}{2}} \exp \left\{ -\frac{1}{2} \text{trace} \left[ \lambda T \Sigma_u^{-1} (\Gamma_{yy}(\theta) - \Phi' \Gamma_{xy}(\theta) - \Gamma_{yx}(\theta) \Phi + \Phi' \Gamma_{xx}(\theta) \Phi) \right] \right\}
\]

(10)

▶ \(\lambda\) determines effective sample size for artificial observations.
Prior Distribution (Cont’d)

- Define functions

\[
\Phi^*(\theta) = \Gamma_{xx}^*(\theta)\Gamma_{xy}^*(\theta)
\]
\[
\Sigma_u^*(\theta) = \Gamma_{yy}^*(\theta) - \Gamma_{yx}^*(\theta)\Gamma_{xx}^{-1}(\theta)\Gamma_{xy}^*(\theta)
\]

- Remarks on \((\Phi^*(\theta), \Sigma_u^*(\theta))\):
  - Were data generated by DSGE conditional on \(\theta\), \(p\)th-order VAR with \(\Phi^*(\theta)\) minimizes one-step-ahead quadratic forecast error loss \(\Sigma_u^*(\theta)\);
  - They trace out subspace of VAR parameter space implied by DSGE implications.

- Given \((\Sigma_u, \theta)\), \(p(\Phi, \Sigma_u|\theta)\) implies conditional prior for \(\Phi\)

\[
p(\Phi|\Sigma_u, \theta) = \exp \left\{-\frac{1}{2} \text{vec}(\Phi - \Phi^*(\theta))' (\Sigma_u^{-1} \otimes \lambda T \Gamma_{xx}^*(\theta)) \text{vec}(\Phi - \Phi^*(\theta))\right\}
\]

implying

\[
\text{vec}(\Phi)|\Sigma_u, \theta \sim_d \mathcal{N}(\text{vec}(\Phi^*(\theta)), \Sigma_u \otimes (\lambda T \Gamma_{xx}^*(\theta))^{-1})
\]
Prior Distribution (Cont’d)

- Using factorization

\[ p(\Phi, \Sigma_u | \theta) = p(\Phi | \Sigma_u, \theta) p(\Sigma_u | \theta) \]

we know

\[ p(\Sigma_u | \theta) \propto |\Sigma_u|^{-\frac{(\lambda T - k) + n + 1}{2}} \exp \left[ -\frac{1}{2} \text{trace}(\lambda T \Sigma_u^{-1} \Sigma_u^*(\theta)) \right] \]

implying

\[ \Sigma_u | \theta \sim_d \mathcal{IW}(\lambda T \Sigma_u^*(\theta), \lambda T - k, n) \] (14)

- Conditional on \( \theta \), DSGE prior has Inverted-Wishart (\( \mathcal{IW} \)) – Normal (\( \mathcal{N} \)) form.

- Overall prior has a hierarchical structure

\[ p(\Phi, \Sigma_u, \theta) = p(\Phi, \Sigma_u | \theta) p(\theta) \] (15)

- We use covariance matrix \( \Sigma_u \otimes (\lambda T \Gamma_{xx}^*(\theta))^{-1} \) to distribute probability mass around \( \Phi^*(\theta) \) and average over \( \theta \) with respect to \( p(\theta) \).
Posterior Distribution

- Decompose joint posterior into posterior of VAR parameters given DSGE parameter and marginal posterior of DSGE parameter

\[
p(\Phi, \Sigma_u, \theta|Y) = p(\Phi, \Sigma_u|Y, \theta)p(\theta|Y) \tag{16}
\]

- Define functions (MLE of (\(\Phi, \Sigma_u\)))

\[
\tilde{\Phi}(\theta) = (\lambda TT^*_{xx}(\theta) + X'X)^{-1}(\lambda TT^*_{xy}(\theta) + X'Y) \tag{17}
\]

\[
\tilde{\Sigma}_u(\theta) = \frac{1}{(\lambda + 1)T}[(\lambda TT^*_{yy}(\theta) + Y'Y) - (\lambda TT^*_{yx}(\theta) + Y'X)(\lambda TT^*_{xx}(\theta) + X'X)^{-1}(\lambda TT^*_{xy}(\theta) + X'Y)] \tag{18}
\]

based on artificial sample and actual sample.

- Because DSGE prior and likelihood function are conjugate, posterior \(p(\Phi, \Sigma_u|Y, \theta)\) is also Inverted Wishart–Normal

\[
\Sigma_u|Y, \theta \sim_d \mathcal{IW}((\lambda + 1)T\tilde{\Sigma}_u(\theta), (\lambda + 1)T - k, n) \tag{19}
\]

\[
\text{vec}(\Phi)|Y, \Sigma_u, \theta \sim_d \mathcal{N}(\text{vec}(\tilde{\Phi}(\theta)), \Sigma_u \otimes (\lambda TT^*_{xx}(\theta) + X'X)^{-1}) \tag{20}
\]
Interpretations of tightness parameter $\lambda$:
- Small $\lambda$ means diffuse prior and actual sample dominates artificial sample; large $\lambda$ means concentrated prior along $(\Phi^*(\theta), \Sigma_u^*(\theta))$;
- It measures extent to which DSGE prior overlap with VAR likelihood;
- It characterizes degree of model misspecification [inverse-U shape].

We use data-driven procedure to determine $\hat{\lambda}$ which maximizes the marginal data density

$$p_\lambda(Y) = \int p_\lambda(Y|\theta)p(\theta)d\theta$$

over some grid $\Lambda = \{\lambda_1, \ldots, \lambda_q\}$.

Model selection with minimal econometric interpretation amounts to comparison of $\hat{\lambda}_M$ and $\hat{\lambda}_F$ [Del Negro & Schorfheide (2004)].
Data Set

- Data source: FED of St. Louis FRED.
- We fit 3 quarterly time series using DSGE-VAR:
  - per capita GDP growth rates (\(YGR\)): \(100 \times \text{LOG diff of GDPC96 (Real Gross Domestic Product) / CLF16OV (Civilian Labor Force)}\);
  - annualized inflation rates (INF): \(100 \times \text{LOG diff of GDPDEF (Gross Domestic Product: Implicit Price Deflator)}\);
  - annualized nominal interest rates (INT): FEDFUNDS (Effective Federal Funds Rate).
- We “shrink” VAR estimates of 3 subsamples toward parameter subspaces implied by Regime-F and -M:
  - pre-Volcker sample: 1955I-1979IV (100 observations);
  - post-Volcker sample: 1982I-2007IV (104 observations);
  - recent recession sample: 2008I-2012IV (20 observations, but fine).
- We confront lack of data availability in recent crisis directly.
Regime-M fares better by model fit but not much; both regimes are equally consistent with data under minimal interpretation $\hat{\lambda}_M \approx \hat{\lambda}_F$. 
Puzzle: Regime-M fares much better by model fit; Regime-F is more consistent with data under minimal interpretation $[\hat{\lambda}_M < \hat{\lambda}_F]$. 
OE of M & F - 2008I:2012IV

- Regime-M & -F fare closely by model fit; both regimes are VERY consistent with data under minimal interpretation $[\hat{\lambda}_M \approx \hat{\lambda}_F \approx 1]$. 
- Regime-M fares better by model fit; both regimes are equally inconsistent with data under minimal interpretation $[\hat{\lambda}_M \approx \hat{\lambda}_F \approx 0]$. 
Fit spread: Regime-M fares much better than -F in normal times (blue) but by only “small” margin in abnormal times (red and black).
Wrapping Up

- Further identification scheme for U.S. M-F policy mix:
  - on structural model - include long-term debt and solve DSGE’s with higher precision;
  - on econometrics model - include debt-to-GDP ratio and use VECM to account for cointegration.

- Future research agenda for DSGE model evaluation:
  - Generic linear representation of DSGE in continuous time

\[
\Gamma_0 ds_t = \Gamma_1 s_t dt + \Psi v_t + \Pi \eta_t 
\] (22)

where \(v\) and \(\eta\) are differentials of martingale processes.

- A natural reduced-form counterpart is martingale regression

\[
dy_t = \mu(y_t, \Phi) dt + du_t 
\] (23)

where \(u_t\) is general martingale.

- Two challenges: use time change to induce normality; extend to multivariate framework.