Likelihood-based Estimation of Stochastically Singular DSGE Models using Dimensionality Reduction

Michal Andrle¹

Computing in Economics and Finance, Prague June 2012

¹The views expressed herein are those of the author and should not be attributed to the International Monetary Fund, its Executive Board, or its management.
Outline of the Talk

- Motivation
- Stochastic Singularity
  - Definition & Problem
  - Data & Models
- Estimating Dimensionality reduction
  - Principal Components Transform
  - Dynamic Principal Components Transform
- Examples...
Motivation

- How many shocks do we need to explain an economy’s dynamics?
- Many dynamic economic models feature only few truly structural shocks, yet have implications for a handful of macro variables
- Macroeconomic data feature a great amount of regularity
- Adding back newly defined ‘regression error’ to DSGE models
Stochastic Singularity – Definition

Assume a model is cast in terms of linear, time invariant state-space form

\[ Y_t = ZX_t \]
\[ X_t = TX_{t-1} + R\varepsilon_t, \]

The model is ‘stochastically singular’ if a spectral density of observed variables \( S_{yy}(\lambda) \) is rank-deficient at almost all frequencies.

Implications:

- Number of shocks is less than or equal to number of observables (sufficient condition)
- A combination of elements in \( Y \) is perfectly correlated.
Stochastic Singularity – Problem

- Macroeconomic and financial data are never perfectly correlated
- The model cannot be exact data-generating process for real data
- It precludes use of likelihood-based methods and Kalman filter
- Bayesian-likelihood methods are of no help here, of course ;)}
Stochastic Singularity – Common Solutions (I)

A. Methods of moments (Ruge-Murcia, 2004)
   ▶ What moment restriction are suitable?
   ▶ How to estimate structural shocks and unobserved variables?

B. Restricting the set of observed variables
   ▶ Loss of valuable information. Which variables to retain?

C. Introducing more ‘structural shocks’
   ▶ Are many shocks used in the literature really structural?
Stochastic Singularity – Common Solutions (II)

D. Adding uncorrelated measurement error (Altug 1989, Sargent 1989)
   ▶ $Y_t = ZX_t + K\nu_t$
   ▶ Cross-correlations driven only by the model
   ▶ Restrictive specification, misspecification issues
   ▶ What variables are noisy?

E. Adding dynamic measurement error process (Ireland, 2004)
   ▶ $Y_t = ZX_t + D(L)\nu_t$
   ▶ Increases the complexity and number of parameters of the model
   ▶ Ad-hoc dynamics may overrule the structural model
Appeal of Stochastic Singularity

Models:
- Theory underpins only few truly structural driving forces
- Sharper economic identification of structural shocks

Data:
- Macroeconomic data feature robust regularities and co-movement
- At business cycle frequency, real economic variables are driven by few factors (Andrle and Brůha, 2012)
- Real and nominal cyclical comovement is very strong (Andrle, 2012)
Real Comovements – Principal Components Analysis

United States...

Source: Andrle and Brůha (2012)
Real Comovements – Principal Components Analysis

Japan...

Source: Andrle and Brůha (2012)
Real Comovements – Principal Components Analysis

Spain...

Source: Andrle and Brůha (2012)
Real and Nominal Comovements

Source: Andrle (2012)
New Estimation Method for Singular Models

Our models and estimation method should exploit data regularities!

Underlying intuitive approach:

1. Find the principal components and principal subspace of the data (subspace of ‘maximum covariance’)
2. Use at most $n_e$ dimensions, where $n_e$ is number of shocks
3. Rotate the model into this subspace
4. Estimate transformed model using likelihood-based methods
5. Estimate (filter) unobserved variables and structural shocks

The estimator penalizes models incompatible with robust feature of the data.
Principal Components – Static vs. Dynamic

Stochastic singularity relates to dynamic rank of the model/data.

Consider a well-known example, a simple process $U(t)$ driven by serially uncorrelated stationary process $\nu(t) \sim N(0, \sigma^2)$

\[
U_t = \begin{bmatrix} \nu_t \\ \alpha \nu_{t-1} \end{bmatrix}, \quad \Gamma_0^U = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \alpha^2 \sigma^2 \end{bmatrix}, \quad \Gamma_1^U = \begin{bmatrix} 0 & 0 \\ \alpha \sigma^2 & 0 \end{bmatrix}, \quad (3)
\]

with rank-deficient spectral density $S_U(\lambda)$

\[
S_U(\lambda) = \begin{bmatrix} 1 & \alpha e^{-i\lambda} \\ \alpha e^{i\lambda} & \alpha^2 \end{bmatrix}. \quad (4)
\]

Static rank (PCA) is 2.
Dynamic rank (DPCA) is 1 – only one driving force.
PC-MLE – Classic Principal Components (i)

Let $\Sigma_\mathcal{Y}$ be a covariance matrix of observed data $\mathcal{Y}$.

Using principal component analysis find $\mathbf{P}$ and $\Lambda$

$$\Sigma_\mathcal{Y} = \mathbf{P} \Lambda \mathbf{P}' \quad r = \text{rank}(\Sigma_\mathcal{Y}),$$

where $\mathbf{P}, \Lambda$ denote matrix of eigenvectors and diagonal matrix of eigenvalues, respectively.

Define $\mathbf{P}^k$ to be a projection matrix on a subspace spanned by $k \leq n_e \leq r$ first principal components of the data, hence

$$\mathbf{F}_t^k = \mathbf{P}^k \mathcal{Y}_t, \quad \hat{\mathcal{Y}}_t = \bar{\mathbf{P}}^k \mathbf{F}_t^k = \bar{\mathbf{P}}^k \mathbf{P}^k \mathcal{Y}_t$$

where $\mathbf{F}_t^k$ is a $(k \times 1)$ vector and $\mathbf{P}^k$ is a $(k \times n_y)$ projection matrix. $\hat{\mathcal{Y}}_t$ is a ‘recovered’ signal based on low-rank dimensionality transform.
The approximation error determines the measurement error process:

\[ e_t = \mathcal{Y}_t - \hat{\mathcal{Y}}_t \]  

(7)

with a spectral density \( \mathbf{S}_e(\lambda) = \mathbf{L} \mathbf{S}_\mathcal{Y}(\lambda) \mathbf{L}' \) where \( \mathbf{L} = (\mathbf{I} - \bar{\mathbf{P}}^k \mathbf{P}^k) \).

Transform the model – project it into factor space:

\[ \mathbf{F}_{S,t}^k = \mathbf{K} \mathbf{X}_t \]  

(8)

\[ \mathbf{X}_t = \mathbf{T} \mathbf{X}_{t-1} + \mathbf{R} \varepsilon_t, \]  

(9)

where \( \mathbf{F}_{m,t}^k \equiv \mathbf{P}^k \mathbf{Y}_t \), with \( \text{dim}(\mathbf{F}_{S,t}) = k \leq n_e \), and \( \mathbf{K} = \mathbf{P}^k \mathbf{Z} \), which is \((k \times n_x)\).

Conditional on \( \mathbf{P}^k \), form standard log-likelihood criterion and use conventional Kalman filter.
The problem is to determine \((k \times n_y)\) linear filter \(P^k(L)\) and \((n_y \times k)\) filter \(\bar{P}^k(L)\) such as to minimize

\[
E||\epsilon|| = E||\mathcal{Y}_t - \hat{\mathcal{Y}}_t|| = E||\mathcal{Y}_t - \bar{P}^k(L)P^k(L)\mathcal{Y}_t|| \tag{10}
\]

Boils down to PCA on spectral density at each frequency.

Data pre-filtering using DPCA filter determines the measurement error process:

\[
S_\epsilon(\lambda) = [I - A(\lambda)]S_\mathcal{Y}(\lambda)[I - A(\lambda)]^H, \tag{11}
\]

where \(A(\lambda) = \bar{P}^k(\lambda)P^k(\lambda)\) is a ‘purifying filter’.
Denote $F_{D,t}$ be the vector of dynamic principal components using DPCA filter $P^k(z)$ if

$$F_{D,t}^k = P^k(L)Y_t \quad \hat{Y} = P^k(L)F_{D,t}^k \quad P^k(z) = \sum_{i=-\infty}^{\infty} P^k_i z^i. \quad (12)$$

Complication: $P^k(z)$ is possibly infinite!

The transformation of the model and estimation is carried out in frequency-domain, where the problem is well defined.
DPC-MLE – Dynamic Principal Components (iii)

Rotate (transform) the model into DPC space:

\[ F_{D,t}^k = P^k(L)Z X_t \]  \hspace{1cm} (13)

\[ X_t = T X_{t-1} + R \varepsilon_t. \]  \hspace{1cm} (14)

Formulate Whittle likelihood

\[ \log \mathcal{L} \propto -\frac{1}{2} \sum_{j=0}^{T-1} \log[\det S_F(\lambda_j)] \]  \hspace{1cm} (15)

\[ -\pi \times \text{tr} \sum_{j=0}^{T-1} [S_F(\lambda_j)^{-1} I_{\mathcal{F}_D}(\lambda_j)], \]

Very efficient & fast way of evaluating likelihood and identification checks.
Structural shocks cannot be estimated using the Kalman filter.

Use a Wiener-Kolmogorov filter implied by the transformed model:

\[ \hat{X}_{t|T} = W(L)P^k(L)\gamma_t \quad W(z) = S_{X,F_D}(z)S_{F_D}(z)^{-1}, \quad (16) \]

where \( S_{X,F_D}(\lambda) \) is a cross-spectral density matrix and \( S_{F_D}(\lambda) \) is a spectral density matrix implied by the parameterized infinite-dimensional state-space model.
DPC-MLE – Dynamic Principal Components (v)

Some important questions & considerations:

- Principal component estimates are not scale invariant.
- Should one compute components on data or of the model?
- ‘Good’ estimates of sample spectral density are not easy to get
Simple Example (I.)

Simple model:

\[ x_t = \rho_x x_{t-1} + \varepsilon_x \]
\[ u_t = \rho_u u_{t-1} + (1 - \rho_u) \alpha_{x,u} x_t + \varepsilon_u, \]  \hspace{1cm} (17) \hspace{1cm} (18)

Parameterized: \( \rho_x = 0.90, \rho_u = 0.4, \alpha_{x,u} = 0.4, \sigma_x = \sigma_u = 0.1 \)

Observed \( u_t, T=200, N=600 \) replications.
RBC Example – One Component/Factor (i)

Simple, standard RBC model:

- Single structural shock (technology)
- Measurement errors added…
- One component/factor computed using cons, inv & wages
- Can we recover the unobserved components well?
RBC Example – One Component/Factor (iii)

Filter estimates — Rental rate

true
reduced–rank filter
filter with cons only
Histogram of estimates for $\rho$
RBC Example – One Component/Factor (v)

Identification/sensitivity for $\rho$
Issues & Caveats

- The best thing is to incorporate most of data considerations directly to the model.
- There are many reasons not to base estimation on likelihood, or unconditional moments only in general. . .
- Likelihood assumes your model is a correct one, pure likelihood methods are not very robust.
Conclusions

- Stochastic singularity can be a useful thing in a model
- Thinking about the data, number of shocks and ‘how much’ one wants to ‘fit’ is useful
- Dimensionality reduction methods provide one way how to extract important features in the data and comovements
Thank you for your patience...