Improving GDP Measurement: A Measurement-Error Perspective

Boragan Aruoba, University of Maryland
Francis X. Diebold, University of Pennsylvania
Jeremy Nalewaik, Federal Reserve Board
Frank Schorfheide, University of Pennsylvania
Dongho Song, University of Pennsylvania

April 28, 2014
Two U.S. GDP Estimates: $GDP_E$ and $GDP_I$

Both are available for U.S.

- $GDP_E$ used routinely
- $GDP_I$ may also be valuable

We provide a superior estimate, $GDP_*$
$GDP_E$ vs. $GDP_I$
(Nalewaik 2010, ...)

Dynamic factor models and optimal signal extraction
(..., Fleischman and Roberts 2011, ...)

Data revision properties
(..., Faust-Rogers-Wright 2005, ...)

Forecast combination
(..., Timmermann 2006, ...)
Warm-up: The Forecast-Error Approach to Combining (Pooling Noisy GDP “Forecasts”)

\[ GDP_c = \lambda GDP_E + (1 - \lambda) GDP_I \]

\[ \lambda = \frac{1 - \phi \rho}{1 + \phi^2 - 2\phi \rho} \]

\[ (\phi = \sigma_E^2 / \sigma_I^2, \ \rho = \text{corr}(e_E, e_I)) \]
Optimal Combining Weights are Far From 0 and 1

Figure: $\lambda$ vs. $\phi$ for Various $\rho$ Values. Reference at $\lambda = 0.50$. 
Gains From Combining Are Huge

Figure: $\frac{\sigma_C^2}{\sigma_E^2}$ for $\lambda \in [0, 1]$. We set $\phi = 1.10$ and $\rho = 0.45$. Reference at $\lambda = 0.50$. 
The Measurement-Error Approach to Combining (Smoothing Noisy GDP Measurements)
“Two-Equation Model”

\[
\begin{bmatrix}
GDP_{et} \\
GDP_{lt}
\end{bmatrix} =
\begin{bmatrix}
1 \\
1
\end{bmatrix} GDP_t +
\begin{bmatrix}
\epsilon_{et} \\
\epsilon_{lt}
\end{bmatrix}
\]

\[
GDP_t = \mu(1 - \rho) + \rho GDP_{t-1} + \epsilon_{Gt},
\]

\[
(\epsilon_{Gt}, \epsilon_{Et}, \epsilon_{lt})' \sim iid \ N(0, \Sigma)
\]

0 \leq \rho < 1

- Both GDP_E and GDP_I are noisy measures of latent true GDP
- Optimal smoothing for GDP (over space and time)
- Estimation rather than calibration
- Interesting hypotheses regarding the form of \(\Sigma\)
Hypotheses of Interest

Diagonal-$\Sigma$: (“standard”)

$$\Sigma = \begin{bmatrix} \sigma_{GG}^2 & 0 & 0 \\ 0 & \sigma_{EE}^2 & 0 \\ 0 & 0 & \sigma_{II}^2 \end{bmatrix}$$

Block-Diagonal-$\Sigma$: (captures overlap in counts)

$$\Sigma = \begin{bmatrix} \sigma_{GG}^2 & 0 & 0 \\ 0 & \sigma_{EE}^2 & \sigma_{EI}^2 \\ 0 & \sigma_{IE}^2 & \sigma_{II}^2 \end{bmatrix}$$

Unrestricted-$\Sigma$: (motivated by Nalewaik, 2010, *inter alia*)

$$\Sigma = \begin{bmatrix} \sigma_{GG}^2 & \sigma_{GE}^2 & \sigma_{GI}^2 \\ \sigma_{EG}^2 & \sigma_{EE}^2 & \sigma_{EI}^2 \\ \sigma_{IG}^2 & \sigma_{IE}^2 & \sigma_{II}^2 \end{bmatrix}$$
Identification

Diagonal-\(\Sigma\) model is identified

Block-Diagonal-\(\Sigma\) model is identified

Unrestricted-\(\Sigma\) model is \textit{unidentified}

(We can increase the volatility of true \textit{GDP} innovations and the measurement errors, but decrease the covariance between true \textit{GDP} innovations and the measurement errors, without changing the distribution of observables.)

Identification requires fixing any element of \(\Sigma\)
A Useful Re-Parameterization

Recall:

\[ GDP_t = \mu (1 - \rho) + \rho GDP_{t-1} + \epsilon_G t \]

\[ \Sigma = \begin{bmatrix}
\sigma_{GG}^2 & \sigma_{GE}^2 & \sigma_{GI}^2 \\
\sigma_{EG}^2 & \sigma_{EE}^2 & \sigma_{EI}^2 \\
\sigma_{IG}^2 & \sigma_{IE}^2 & \sigma_{II}^2
\end{bmatrix} \]

Reparameterize in terms of the ratio of \( GDP \) variance to \( GDP_E \) variance:

\[ \zeta = \frac{\frac{1}{1-\rho^2} \sigma_{GG}^2}{\frac{1}{1-\rho^2} \sigma_{GG}^2 + 2 \sigma_{GE}^2 + \sigma_{EE}^2} \]

A \( \zeta \) value less than, but close to, 1 seems most natural.

We take \( \zeta = 0.80 \) as our benchmark.
The “Three-Equation Model”

Add an additional observable variable $U_t$ with a certain structure:

$$
\begin{bmatrix}
GDP_{Et} \\
GDP_{It} \\
U_t
\end{bmatrix} =
\begin{bmatrix}
0 \\
0 \\
\kappa
\end{bmatrix} +
\begin{bmatrix}
1 \\
1 \\
\lambda
\end{bmatrix} GDP_t +
\begin{bmatrix}
\epsilon_{Et} \\
\epsilon_{It} \\
\epsilon_{Ut}
\end{bmatrix}
$$

$$GDP_t = \mu(1 - \rho) + \rho GDP_{t-1} + \epsilon_{Gt},$$

where $(\epsilon_{Gt}, \epsilon_{Et}, \epsilon_{It}, \epsilon_{Ut})' \sim iid N(0, \Omega)$, with

$$\Omega =
\begin{bmatrix}
\sigma_{GG}^2 & \sigma_{GE}^2 & \sigma_{GI}^2 & \sigma_{GU}^2 \\
\sigma_{EG}^2 & \sigma_{EE}^2 & \sigma_{EI}^2 & 0 \\
\sigma_{IG}^2 & \sigma_{IE}^2 & \sigma_{II}^2 & 0 \\
\sigma_{UG}^2 & 0 & 0 & \sigma_{UU}^2
\end{bmatrix}$$
What to Use for $U$?

We take $U$ to be the change in the unemployment rate

– Clearly unemployment rate changes load on $GDP$ growth

– Unemployment data are constructed from household surveys, and very little household survey data are used to construct $GDP_E$ and $GDP_I$

– Hence unemployment measurement errors are reasonably assumed to be orthogonal to those of $GDP_E$ and $GDP_I$
Empirics I

Estimation
Bayesian Analysis of the Dynamic Factor Model

Carter-Kohn multi-move Gibbs sampling:

1. Update parameter values using random-walk Metropolis-Hastings

2. Filter latent state using the Kalman filter

– Markov chains burned in for 25,000 steps and then sampled for 25,000 steps.
For the 2-equation model with $\zeta = 0.80$, we have

$$GDP_t = 3.08 \ (1 - 0.57) + 0.57 \ GDP_{t-1} + \epsilon_{Gt}$$

$$\Sigma = \begin{bmatrix}
7.09 & -0.69 & -0.38 \\
[6.54,7.70] & [-1.15,-0.29] & [-0.74,-0.04] \\
-0.69 & 3.90 & 1.29 \\
[-1.15,-0.29] & [3.14,4.77] & [0.80,1.85] \\
-0.38 & 1.29 & 2.36 \\
[-0.74,-0.04] & [0.80,1.85] & [1.98,2.82]
\end{bmatrix}$$
For the 3-equation model, we have

\[
\begin{bmatrix}
GDP_{Et} \\
GDP_{It} \\
U_t
\end{bmatrix} =
\begin{bmatrix}
0 & 0 & 1 \\
0 & 1 & 1.62 \\
[1.53,1.71] & [−0.55,−0.50] & 
\end{bmatrix}
\begin{bmatrix}
GDP_t \\
\epsilon_{Et} \\
\epsilon_{It} \\
\epsilon_{Ut}
\end{bmatrix}
\]

\[
GDP_t = 2.78 \ (1 − 0.58) + 0.58 \ GDP_{t−1} + \epsilon_{Gt}
\]

\[
\Omega =
\begin{bmatrix}
6.96 & −1.10 & −0.82 & 1.46 \\
[6.73,7.35] & [−1.27,−0.84] & [−1.03,−0.59] & [1.27,1.66] \\
−1.10 & 4.57 & 1.95 & 0 \\
[−1.27,−0.84] & [4.17,4.79] & [1.70,2.12] & \\
−0.82 & 1.95 & 3.07 & 0 \\
[−1.03,−0.59] & [1.70,2.12] & [2.54,3.27] & \\
1.46 & 0 & 0 & 0.59 \\
[1.27,1.66] & \\
\end{bmatrix}
\]
On the Selection of $\zeta$

We view the “3-equation identification” as a useful complement to the “$\zeta$-identification”

We can even view the 3-equation approach as a device for implicitly selecting $\zeta$

$\zeta^* = 0.82$ minimizes the Frobenius matrix-norm divergence between $\hat{\Sigma}_\zeta$ and $\hat{\Sigma}_3$
Empirics II

The Importance of $GDP_i$
Kalman Gains

Solid lines indicate 90% posterior coverage ellipsoids. Stars indicate posterior median values.
Sample Path Properties of $GDP^*$
In each panel we show the sample path of $GDP_*$ in red together with a light-red posterior interquartile range, and we show one of the competitor series in black. We obtain $GDP_*$ from the 2-equation model with $\zeta = 0.80$. 
In each panel we show the sample path of $GDP_*$ in red together with a light-red posterior interquartile range, and we show one of the competitor series in black. We obtain $GDP_*$ from the 2-equation model with $\zeta = 0.80$. 

\textbf{GDP} \textit{vs. GDP}_E \textit{and GDP}_I \textit{Sample Paths, 2007Q1-2009Q4}
Empirics IV

$(\rho, \sigma^2_{GG})$ for $GDP_*$ vs. $GDP_E$ and $GDP_I$
$(\hat{\rho}, \hat{\sigma}_{GG}^2)$ Pairs Across MCMC Draws

Solid lines indicate 90% $(\sigma_{GG}^2, \rho)$ posterior coverage ellipsoids for the various models. Stars indicate posterior median values. The sample period is 1960Q1-2011.Q4. For comparison we show $(\sigma^2, \rho)$ values corresponding to $AR(1)$ models fit to $GDP_E$ alone and $GDP_I$ alone.
Conclusion and Future Research

Conclusion:

\( GDP_* \) is an obvious benchmark U.S. \( GDP \) estimate

Now produced by Federal Reserve Bank of Philadelphia

Future:

Levels analysis with co-integration
Prediction
Real-time
Serially-correlated measurement error
Additional identifying variables