The Statistical Discrepancy

William D. Nordhaus
Yale University

BEA Advisory Committee
May 2010

Samuelson on the Statistical Discrepancy

Statisticians must always work with incomplete reports and fill in data gaps by estimation. Just as measurements in a chemistry lab differ from the ideal, so, in fact, do errors creep into both [income and expenditure] estimates. These are reconnected by an item called the "statistical discrepancy."

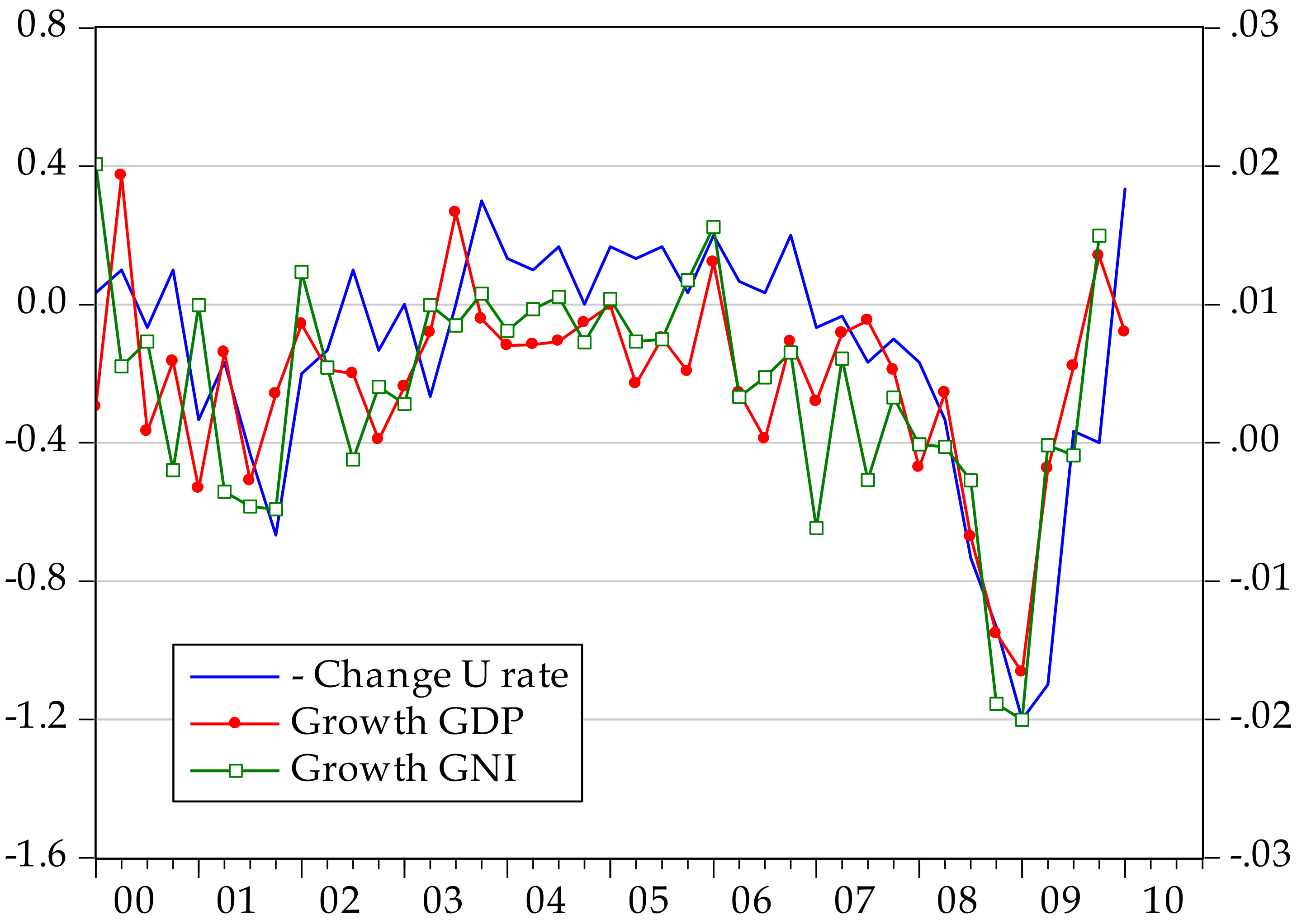
Along with civil servants who are heads of units called "Wages," "Interest," and so forth, there is actually someone with the title of "Head of the Statistical Discrepancy." If data were perfect, that individual would be out of a job; but because real life is never perfect, that person's task of reconciliation is one of the hardest of all."

[Paul Samuelson, *Economics*, 1960 with edits.]

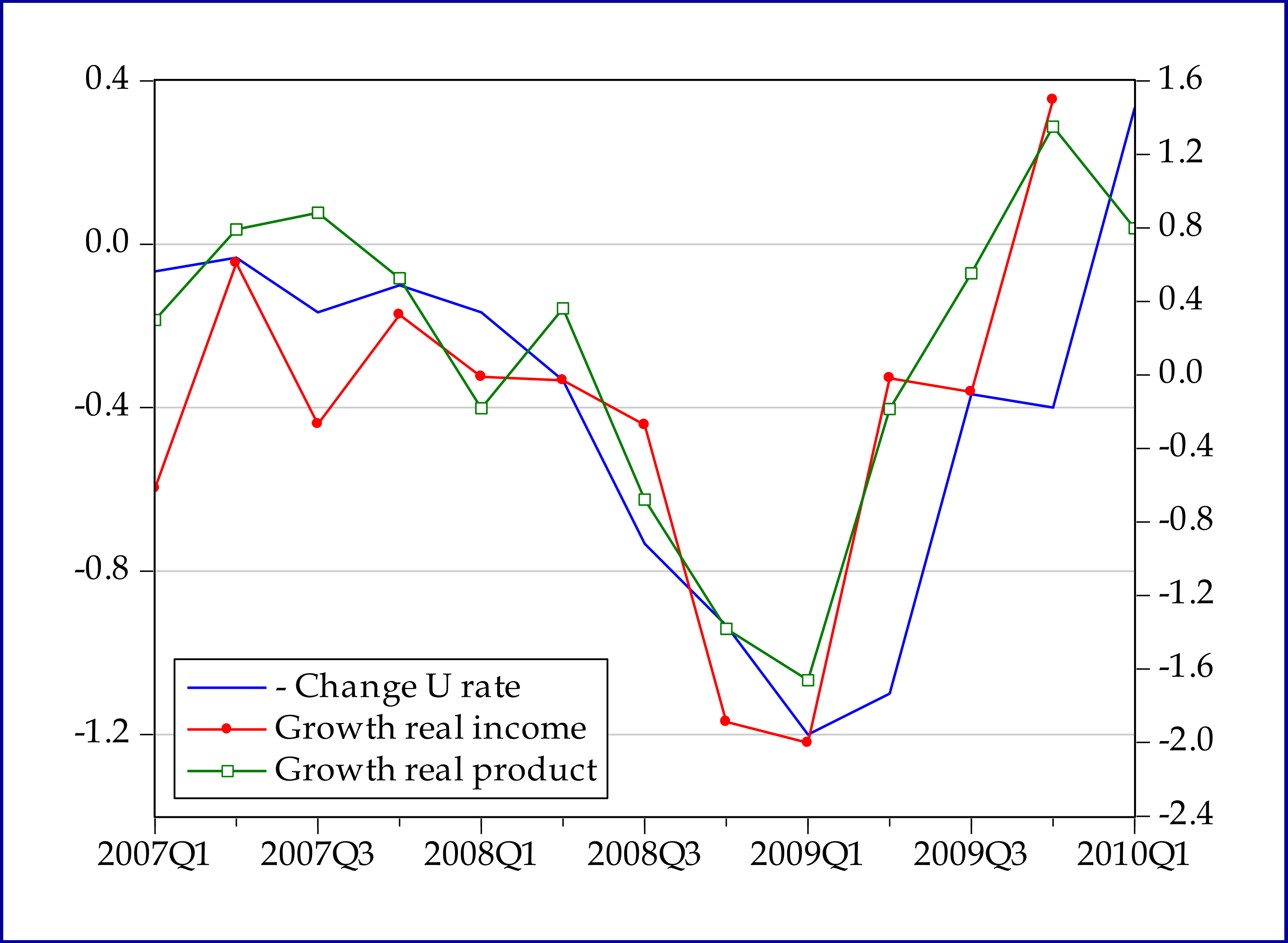
A signal extraction problem

- We have two noisy signals (product and income side)
- Have a true value (true GDP)
- How can we extract true value from noisy signals?
- I show an example using Okun’s Law and the unemployment rate.

Business cycle basics



Business cycle basics



Technical details

$$g_t^{income} = g_t^{true} + \varepsilon_t^{income} \qquad [g_t^* = \text{ growth real output}]$$

$$g_t^{product} = g_t^{true} + \varepsilon_t^{product} \qquad [\varepsilon_t^* = \text{ measurement error}]$$

$$g_t^{true} = \lambda\left(g_t^{income} - \varepsilon_t^{income}\right) + (1-\lambda)\left(g_t^{product} - \varepsilon_t^{product}\right)$$

$$g_t^{true} = \lambda g_t^{income} + (1-\lambda)g_t^{product} + \left[\lambda\varepsilon_t^{income} + (1-\lambda)\varepsilon_t^{product}\right]$$

Technical details

$$g_t^{income} = g_t^{true} + \varepsilon_t^{income}$$ [$g_t^* =$ growth real output]

$$g_t^{product} = g_t^{true} + \varepsilon_t^{product}$$ [$\varepsilon_t^* =$ measurement error]

$$g_t^{true} = \lambda\left(g_t^{income} - \varepsilon_t^{income}\right) + (1-\lambda)\left(g_t^{product} - \varepsilon_t^{product}\right)$$

$$g_t^{true} = \lambda g_t^{income} + (1-\lambda) g_t^{product} + \left[\lambda \varepsilon_t^{income} + (1-\lambda)\varepsilon_t^{product}\right]$$

$$\Delta U_t = \alpha + \beta g_t^{true} + \theta_t$$ [Okun's Law]

Technical details

$$g_t^{income} = g_t^{true} + \varepsilon_t^{income} \qquad [g_t^* = \text{ growth real output}]$$

$$g_t^{product} = g_t^{true} + \varepsilon_t^{product} \qquad [\varepsilon_t^* = \text{ measurement error}]$$

$$g_t^{true} = \lambda\left(g_t^{income} - \varepsilon_t^{income}\right) + (1-\lambda)\left(g_t^{product} - \varepsilon_t^{product}\right)$$

$$= \lambda g_t^{income} + (1-\lambda) g_t^{product} + \left[\lambda \varepsilon_t^{income} + (1-\lambda)\varepsilon_t^{product}\right]$$

$$\Delta U_t = \alpha + \beta g_t^{true} + \theta_t \qquad \text{[Okun's Law]}$$

$$\Delta U_t = \alpha + \beta\left[\lambda g_t^{income} + (1-\lambda) g_t^{product}\right] + \xi_t$$

Technical details

$$g_t^{income} = g_t^{true} + \varepsilon_t^{income} \qquad [g_t^* = \text{ growth real output}]$$

$$g_t^{product} = g_t^{true} + \varepsilon_t^{product} \qquad [\varepsilon_t^* = \text{ measurement error}]$$

$$g_t^{true} = \lambda\left(g_t^{income} - \varepsilon_t^{income}\right) + (1-\lambda)\left(g_t^{product} - \varepsilon_t^{product}\right)$$

$$= \lambda g_t^{income} + (1-\lambda) g_t^{product} + \left[\lambda \varepsilon_t^{income} + (1-\lambda)\varepsilon_t^{product}\right]$$

$$\Delta U_t = \alpha + \beta g_t^{true} + \theta_t \qquad \text{[Okun's Law]}$$

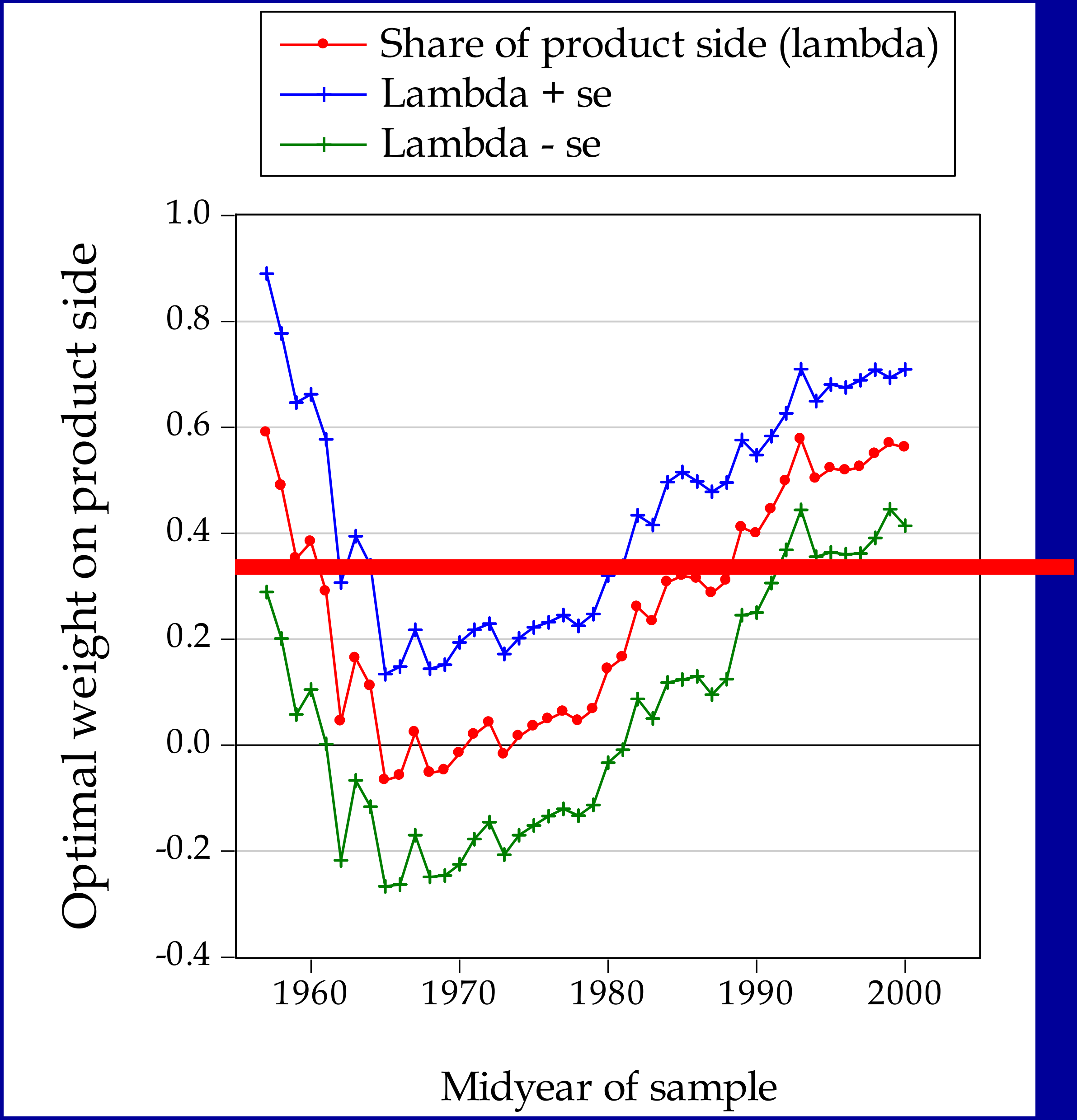
$$\Delta U_t = \alpha + \beta\left[\lambda g_t^{income} + (1-\lambda) g_t^{product}\right] + \xi_t$$

where $\xi_t = \theta_t - \beta\left[\lambda \varepsilon_t^{income} + (1-\lambda)\varepsilon_t^{product}\right]$

λ is optimal weight on income (to be estimated).

This procedure would work for any variable that is orthogonal to errors and is highly correlated with "true" output.

Estimates in rolling regressions
(centered 20 year regressions)



1960-2010:
1-λ= 0.35 (± .12)

1-λ= optimal weight on product side

Regression for 1960-2009

Dependent Variable: D(LHUR)
Method: Least Squares
Sample (adjusted): 1960Q1 2009Q4
Included observations: 200 after adjustments
Convergence achieved after 13 iterations
D(LHUR)=C(1)+C(2)*(C(3)*GPROD+(1 -C(3))*GINC)+[AR(1)=C(5)]

	Coefficient	Std. Error	t-Statistic	Prob.
Constant	0.188543	0.033669	5.599889	0.0000
Okun	-21.41886	2.041418	-10.49215	0.0000
1-lambda	0.349802	0.122263	2.861054	0.0047
AR1	0.495493	0.063144	7.847027	0.0000

R-squared	0.622317	Mean dependent var	0.022167
Adjusted R-squared	0.616536	S.D. dependent var	0.342655
S.E. of regression	0.212187	Akaike info criterion	-0.242899
Sum squared resid	8.824582	Schwarz criterion	-0.176933
Log likelihood	28.28991	Hannan-Quinn criter.	-0.216203
F-statistic	107.6513	Durbin-Watson stat	2.012821
Prob(F-statistic)	0.000000		

Inverted AR Roots .50

Other tests

- Product side does relatively better after 1990.
- Product side does much worse when unemployment is rising.
- Product side does better in NBER recessions and when unemployment is very high.

Conclusion

Elementary considerations would say that we should combine income and product to get best measure.

Recent data suggest that they should have approximately equal weights using unemployment as instrument.

Need other orthogonal and highly correlated proxies to test (employment problem because of use in output)

Complications:

- Statistical problem: Is the SD managed? If so, big statistical problems in determining optimal weights.
- Communications problem: Can this new concept be managed? [Yes. A lot easier than Fisher constructs.]