

*Which Measure of Aggregate Output Should We Use?**

Aggregate output is a focal point of macroeconomics. In the empirical literature, a number of variables are used as a proxy for aggregate output. In this paper, we compare the temporal behavior of four monthly measures of aggregate output: namely GDP, ICI, IP, and XCI. These four measures of output are found to have a common long-run permanent component but distinct short-run cyclical patterns. The monetary effects on each of these four output variables are used to illustrate the implications of the choice of an output measure on empirical research. It is found that test results can be driven by the choice of the proxy for aggregate output.

1. Introduction

Aggregate output dynamics is an active and broad research topic. There are ongoing efforts, both theoretical and empirical, to investigate the temporal fluctuations and determinants of aggregate output. Aggregate output also plays an important role in modeling the behavior of other economic variables. Even though aggregate output is a focal point of macroeconomics, it is measured in several different ways in empirical research. For example, data on both the gross national/domestic product (GNP/GDP) and the industrial production (IP) index are used to examine output dynamics (Baxter and Stockman 1989; Campbell and Mankiw 1989). Friedman and Kuttner (1992) use GNP data to study the causal effect of money on output while Stock and Watson (1989a) use IP data. In the exchange rate literature, both GDP and IP are used as a proxy for output in empirical exchange rate determination models (Meese and Rogoff 1983; Mark 1995). Some studies also use the index of coincident indicators (ICI) compiled by the Bureau of Economic Analysis as an empirical aggregate output variable (Phillips, Vargas and Zarnowitz 1996; Walsh and Wilcox 1995). Recently, an experimental

*We would like to thank two anonymous referees, Doug McMillin (the Editor), Menzie Chinn, James Wilcox, and Carl Walsh for their helpful comments and suggestions. Carl also kindly provided us with the data on ICI and GDP.

coincident index (XCI) was constructed to measure the overall aggregate cyclical movement (Stock and Watson 1989b and 1991).

Given the alternative measures of aggregate output and the differences in their coverages, which one of these measures should be used in empirical studies? In principle, the choice of an empirical output measure should be guided by the theory being examined. However, as indicated above, different researchers tend to use different output measures to investigate the same economic hypothesis. There seems to be no apparent consensus on the appropriate choice of aggregate output in empirical studies. This raises the question, Do the empirical results depend on the choice of output data?

If different measures of aggregate output have essentially the same dynamic properties, then it does not matter which one is used. However, given the differences in coverage and construction method, different output variables may measure quite different things and have their own unique characteristics and dynamics. That is, different output variables can lead to different interpretations of economic theory. In fact, Romer (1986a, 1986b) shows that the relative volatility of historical output depends on the method used to construct these output data. Cheung, Chinn and Tran (1995) report that the persistence of per capita output data is sensitive to whether the output is valued in domestic currency terms or in some international numeraire.

This paper attempts to provide a systematic comparison of some common output measures. Specifically, GDP, ICI, IP, and XCI data are considered. In terms of its constituting components, GDP is the variable closest to the concept of aggregate output. Until recently, GDP data were available quarterly but not monthly (Bernanke and Gertler 1995; Bernanke and Mihov 1995). This makes the IP a convenient choice among researchers when monthly data are considered. ICI, which has a wider coverage than IP and is also available at the monthly frequency, is gaining popularity as a proxy for aggregate output. The XCI is a recently developed economic index to capture the unobservable state of the aggregate economy.

In comparing the four output variables, we examine both their long-run and short-run dynamics. Specifically, the Johansen technique is used to test for a common long-term component among the output series. The presence of common short-run cyclical patterns in these data is investigated using the common feature test. In addition, an example of the money and output relationship is used to illustrate the implications of using these different output measures in empirical studies.

The remainder of the manuscript is organized as follows: Section 2 contains a brief description of the data. The unit root test results are also reported in this section. Section 3 examines the long-term comovement, compares the short-term dynamics, and tests for the presence of a common

cycle. Section 4 presents the results of using the four different aggregate output measures to study the interactions between money and output. Some concluding remarks are given in Section 5.

2. Data and Preliminary Analysis

Monthly observations of the four aggregate real output series—GDP, ICI, IP, and XCI—are considered. The monthly GDP data are constructed by Bernanke and Gertler (1995), the data on IP are retrieved from the Citibase data set, the ICI data are provided by the Bureau of Economic Analysis, and the XCI data are obtained from the NBER website (<http://www.nber.org/~jstock>). For comparison purposes, we consider the sample 1964:1 to 1993: 12 because monthly real GDP data are available only during that period. The aggregate output data are seasonally adjusted at the sources and expressed in logarithmic terms.

Data Description

Perhaps GDP is the most preferable measure of aggregate output. However, the official data are only available quarterly. To fill the void, Bernanke and Gertler (1995) use components of GDP that are available on the monthly basis and the quarterly data to construct the monthly GDP series. Specifically, the Chow and Lin (1971) method is employed to extract the information to establish the monthly GDP series. For example, monthly inventory investment is derived from data on “inventory investment, manufacturing,” “inventory investment, retail trade,” and “inventory investment, wholesale trade.” See Bernanke and Mihov (1995, Appendix 1) for further details on the construction of the monthly GDP data. As documented by Wilcox (1983), the Chow and Lin method can successfully recover the essential dynamic characteristics of a data series, including autocorrelation structure and turning points.

Among the four output series, IP has the narrowest coverage of the economy-wide activity. The IP index is prepared by the Federal Reserve Board of Governors. It captures economic activity in the manufacturing, mining, and utilities sectors. Physical quantities of output in these sectors rather than their dollar sales are used to construct the index. Specifically, individual components of the index are measured in physical products, production-worker hours, and electric kilowatt hour usage. The weights of these individual components used to update the aggregate IP index are revised periodically. A more detailed discussion of the IP index is given in, for example, Board of Governors of the Federal Reserve System (1986) and Corrado, Gilbert, and Raddock (1997). Data on IP are commonly used as a proxy for aggregate output in a monthly model.

TABLE 1. *Unit Root Test Results*

series	level series			lag p	differenced series		
	lag p	ADF-GLS statistic	5% Critical Value		lag p	ADF-GLS statistic	5% Critical Value
GDP	3	-1.247	-2.887	3		-9.966	-2.887
ICI	4	-1.403	-2.883	2		-7.961	-2.890
XCI	3	-1.536	-2.883	2		-7.351	-2.890
IP	3	-1.637	-2.887	2		-8.929	-2.890

NOTES: GDP, ICI, IP, and XCI are the monthly gross domestic product, index of coincident indicators, industrial production index, and experimental coincident index. The sample period is from 1964:01–1993:12. All variables are in natural logarithms and in real terms. The Schwarz Bayesian criterion is used to select the lag parameter p with a maximum lag of 12 allowed. The 5% finite sample critical values are obtained from Cheung and Lai (1995).

The ICI, compiled by the Bureau of Economic Analysis and published by the Conference Board, provides a broader measure of aggregate economic activity. The ICI is devised to track the general state of the economy and the business cycle, and it is found to be closely related to GDP (Stock and Watson 1989b). Green and Beckman (1993) describe the most recent revision and the computation of this index series. The ICI consists of four components: IP, the number of employees on nonagricultural payrolls, personal income less transfer payments in constant dollars, and manufacturing and trade sales in constant dollars. While ICI has a narrower coverage than GDP, it provides a better representation of economic activity than IP does.

The XCI series is derived from the same four components of ICI. The weights of individual components are estimated from a single-index (dynamic factor) model assuming that the component series have a common unobserved element, which is shared by other macroeconomic variables. Stock and Watson (1991) describe the technical detail and show that the XCI estimated from the single-index model is very similar to the ICI, especially at business cycle frequencies.

Preliminary Data Analysis

Table 1 presents the unit root test results of the four output series. While there is not a strictly uniformly most powerful invariant test for the unit root hypothesis, Elliott, Rothenberg and Stock (1996) derive a modified Dickey-Fuller test, called the ADF-GLS test, which is approximately the uniformly most powerful invariant against the local alternatives. The ADF-GLS test is based on the regression equation

$$(1 - L)y_t^\tau = a_0 y_{t-1}^\tau + \sum_{j=1}^p a_j (1 - L)y_{t-j}^\tau + \varepsilon_t, \quad (1)$$

where $\{y_t^\tau\}$ is the locally detrended process under the local alternative of $\bar{\alpha}$ and is obtained by

$$y_t^\tau = y_t - \tilde{\gamma}' z_t, \quad (2)$$

with $z_t = (1, t)'$. $\tilde{\gamma}$ is the least squares regression coefficient of \tilde{y}_t on \tilde{z}_t where $(\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_T) = (y_1, (1 - \bar{\alpha}L)y_2, \dots, (1 - \bar{\alpha}L)y_T)$ and $(\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_T) = (z_1, (1 - \bar{\alpha}L)z_2, \dots, (1 - \bar{\alpha}L)z_T)$ with L being the lag operator. The local alternative $\bar{\alpha}$ is defined by $\bar{\alpha} = 1 + \bar{c}/T$ for which \bar{c} is set to -13.5 as recommended by Elliott, Rothenberg and Stock (1996). The ADF-GLS test statistic is given by the usual t-statistics for $a_0 = 0$ against the alternative of $a_0 < 0$. See Elliott, Rothenberg and Stock (1996) for a more detailed discussion.

The choice parameter p is determined by the following procedures. First the maximum autoregressive (AR) lag is set to 12 and the Schwarz Bayesian criterion (SBC) is used to determine the first estimate of p . Then, the residuals from the selected model are checked for serial correlations. If there is no significant serial correlation in the estimated residuals, the number of lags determined by the SBC is used to conduct the test. Otherwise, the lag parameter will be increased by one until the resulting specification successfully removes serial correlation in the residuals.

The ADF-GLS test results are reported in Table 1. It is evidence that the levels of the series do not reject the unit root hypothesis. On the other hand, the unit root hypothesis is soundly rejected by the first differenced data. The results indicate the presence of unit root persistence in these output data. This finding is consistent with the existing empirical results that it is difficult to find statistical evidence against the existence of a unit root in the postwar output data. Hence, in the subsequent discussion, these monthly output series are assumed to be difference-stationary.

The log data and their first differences are graphed in Figure 1 and Figure 2, respectively. Broadly speaking, the four output series tend to move together (see Figure 1). On the other hand, a visual inspection of Figure 2 indicates that there are some dissimilarities among the differenced data series. Thus, the four output series may have different short-run behavior even though they appear to follow a general trend. A formal analysis of output comovement is given in the next section.

3. Long-Term and Short-Term Interactions

In the previous section we found that there is a unit root in the GDP, ICI, IP, and XCI data and the four series appear to follow a common trend.

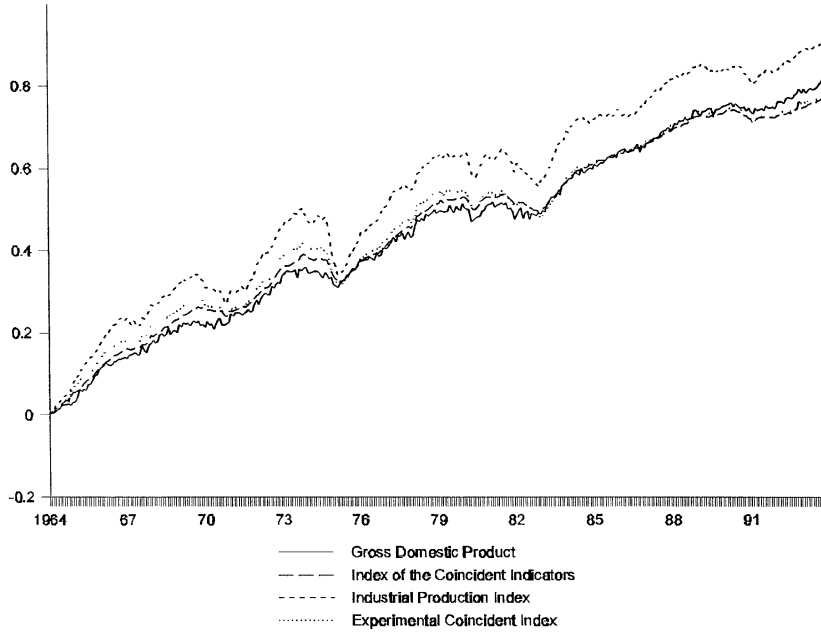


Figure 1.

Gross Domestic Product, Index of Coincident Indicators, Industrial Production Index, and Experimental Coincident Index: Log Levels

To provide a formal analysis of the long-term comovement, the Johansen procedure is used to test for cointegration in these four output series. The Johansen (1991) test is based on the regression equation

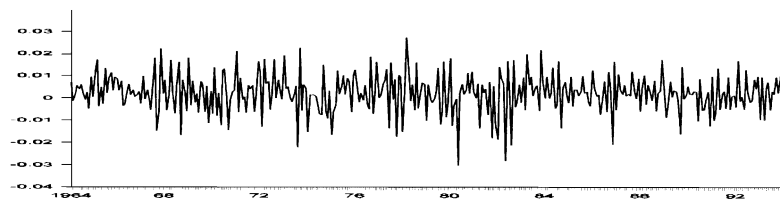
$$\Delta y_t = \mu + \Gamma_1 \Delta y_{t-1} + \dots + \Gamma_{k-1} \Delta y_{t-k+1} + \pi y_{t-k} + u_t, \quad (3)$$

where y_t is a 2×1 vector consisting of two of the four aggregate output measures, μ is a constant vector, and u_t is the error term. $\Delta \equiv (1 - L)$ is the first-difference operator. The test is based on the reduced rank regression technique as follows. First, regress Δy_t and y_{t-k} separately on a constant and $\Delta y_{t-1}, \dots, \Delta y_{t-k+1}$ to obtain the residuals \hat{u}_{1t} and \hat{u}_{2t} , respectively. Then, define the product moment matrices $S_{ij} = T^{-1} \sum_{t=1}^T \hat{u}_{it} \hat{u}_{jt}$ and compute the eigenvalues, $\lambda_1 \geq \lambda_2$, of $S_{21} S_{11}^{-1} S_{12}$ with respect to S_{22} . The hypothesis that there are at most r cointegrating vectors is tested using the trace statistic

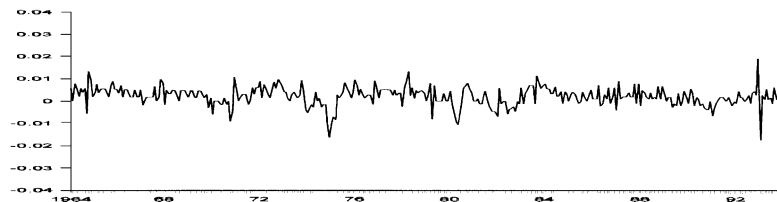
$$\lambda_{trace}(r) = -T \sum_{j=r+1}^2 \ln(1 - \lambda_j). \quad (4)$$

Which Measure of Aggregate Output Should We Use?

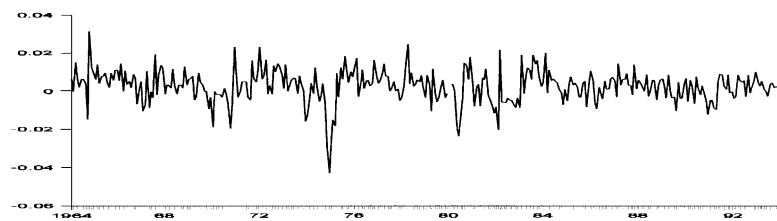
2.A Gross Domestic Product



2.B Index of the Coincident Indicators



2.C Industrial Production Index



2.D Experimental Coincident Index

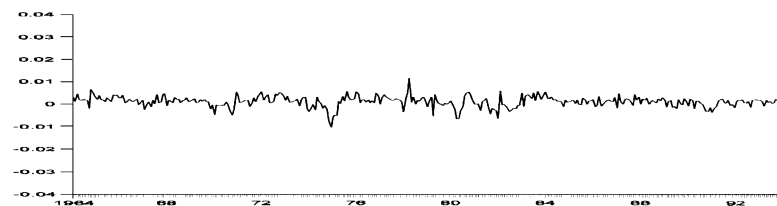


Figure 2.

Gross Domestic Product, Index of Coincident Indicators, Industrial Production Index, and Experimental Coincident Indicators: First Log Differences

TABLE 2. *Cointegration Test Results, 1964:01–1993:12*

A. Bivariate Cointegration						
	(GDP, ICI)	(GDP, IP)	(GDP, XCI)	(ICI, IP)	(ICI, XCI)	(IP, XCI)
Max. eigenvalue statistics						
$r = 1$	0.00	0.24	0.52	5.19	5.58	2.46
$r = 0$	20.63*	20.79*	20.14*	29.37*	22.69*	18.36*
Trace statistics						
$r = 1$	0.00	0.24	0.522	5.19	5.58	2.46
$r = 0$	20.63*	21.03*	20.19*	34.56*	28.27*	20.83*
Lag p	3	3	6	3	6	4
B. Four-variable (GDP, ICI, IP, XCI) Cointegration						
Maximum eigenvalue statistics						
$r = 3$			1.36			
$r = 2$			8.85			
$r = 1$			27.87*			
$r = 0$			35.86*			
Trace statistics						
$r = 3$			1.36			
$r = 2$			10.21			
$r = 1$			38.08*			
$r = 0$			73.95*			
Lag p			6			

NOTES: The maximum eigenvalue and trace statistics of the Johansen cointegration test are reported for the indicated pairs of output series in Panel A and for all four output series in Panel B. The sample period is from 1964:01–1993:12. The lag parameter p is selected using the Schwarz Bayesian criterion. If it is necessary, additional lags are added to remove remaining serial correlation in the estimated residuals. The finite sample critical values from Cheung and Lai (1993) are used to evaluate the significance of the statistics. “*” indicates significance at the 5% level.

One can also test the hypothesis of r cointegrating vectors against the alternative of $r + 1$ cointegrating vectors using the maximum eigenvalue statistic

$$\lambda_{\max}(r, r + 1) = -T \ln(1 - \lambda_{r+1}) . \quad (5)$$

Table 2 presents the results of bivariate cointegration analysis, which provide better information on whether one output measure is cointegrated with another output measure. Similar to the procedures used in the previous section, both the SBC and serial correlation in the estimated residuals are used to determine the lag parameter. For each of the six pairs of output series, both the trace and maximum eigenvalue statistics reject the no-cointegration null hypothesis. However, the null hypothesis of, at most, one

cointegration vector is not rejected. The evidence indicates the presence of one cointegration vector in each of the six pairs of output series. The result of applying the cointegration test to the system of four output measures is reported in the lower panel of Table 2. Again, both the trace and maximum eigenvalue statistics suggest the series are cointegrated.

Even though individual series are integrated of order one and do not converge to a specific value, they are cointegrated. That is, these output series tend to move together in the long run. This is an encouraging finding. If the stochastic trend represents the long-term component, the cointegration results indicate that the four output series share a common long-term output component and do not drift apart over time.

While the four output series have a common long-term component, the preliminary analysis in the previous section suggests that these series may have distinct short-run behavior. The common feature test (Engle and Kozicki 1993) offers a formal procedure to test if there exists, for example, a common serial correlation pattern in any pair of these output series. If the output series have a common stochastic component, then an appropriate linear combination of the series will eliminate the effect of the common component. It means linear combinations of the series can be used to test for the existence of a common feature. Specifically, Engle and Kozicki show that the test for common serial correlation can be conducted as follows. First, we estimate the equation

$$y_{it} = c + \delta y_{jt} + \omega_t, \quad (6)$$

with the instrument variables $\{c, \Omega_{t-1}\}$, where $\Omega_{t-1} \equiv (y_{it-k}, y_{jt-k}; k = 1, \dots, p)$ and y_{it} and y_{jt} are any two of the GDP, ICI, IP, and XCI series. $(1, \delta)$ is the normalized common feature vector. Then, the TR^2 test statistic is computed from the regression of $\hat{\omega}_t$ on Ω_{t-1} . The TR^2 statistic has an asymptotic $\chi^2(2p - 1)$ distribution under the null hypothesis of a common serial correlation feature relative to the information set Ω_{t-1} .

Since it is possible that two output series, but not four, have a common feature, we focus on the bivariate common feature analysis. The results of the serial correlation common feature test are reported in Table 3. For each of the six possible pairs of output series, the lag parameter p is set to 6. Given the lag structures reported in previous sections, $p = 6$ can capture the salient correlation patterns in the data. Further, the cases $p = 4$ and $p = 12$ yield essentially the same results. The entries below the diagonal are the TR^2 statistics with the series listed under the first column as the dependent variable and those above the diagonal have these series as the right-hand-side variable. For the log data, the null hypothesis that there is a common serial correlation feature is rejected at the conventional levels of

TABLE 3. *Common Feature Test Results*

	GDP	ICI	XCI	IP
A. Level Series				
GDP	—	34.68	32.59	39.32
ICI	37.87	—	107.46	185.28
XCI	40.52	99.84	—	84.29
IP	48.27	198.77	88.02	—
B. Differenced Series				
GDP	—	116.18	126.76	102.60
ICI	66.79	—	68.69	24.75
XCI	92.74	72.10	—	6.86
IP	70.96	23.71	6.74	—

NOTES: Results of testing for a common serial correlation feature in a pair of output series are presented. The lag parameter is set to 6. Statistics computed from regressions with variables listed under the first column as the dependent variable are given below the diagonal while statistics with these variables as the right-hand-side variables are given above the diagonal. With the exception of the IP-XCI pair in Panel B, the null hypothesis of a common serial correlation feature is rejected in all cases.

significance in all the six bivariate systems (Panel A). The results of applying the test to the first differenced data are reported in Panel B. With the exception of the IP-XCI pair, the null hypothesis of a common cyclical feature is still clearly rejected.

The preceding analysis confirms that even though these output variables have a common stochastic trend and do not drift apart in the long-run, each variable has its own unique short-term movement pattern which is not presented in other variables. Thus, the choice of output variables can have nontrivial implications for empirical studies related to the short-run output behavior.

4. An Example

There are similarities and differences between the temporal dynamics of aggregate output variables GDP, ICI, IP, and XCI. While these output variables share a common permanent component, they exhibit idiosyncratic short-run growth patterns. Thus, for empirical analyses, different choices of aggregate output can lead to substantially different results.

To investigate the potential implication, we compare the causal effect of money on the four aggregate output variables. The empirical issue of money-output causality is crucial to the debate on the monetary against the

real theories of aggregate output fluctuations. Effects of money on real output are routinely interpreted as supportive evidence for the effectiveness of monetary policy. In the literature, the money-output causality is investigated using different model specifications, various measures of money and output, and data from different sample periods. Our intention is neither to review the massive literature on this topic nor to conduct a thorough test for monetary effects. Instead, we would like to illustrate the implication of using different variables as a proxy for aggregate output on the money-output regression.

We adopt the four-variable equation devised in Stock and Watson (1989a) to investigate the monetary effect on aggregate output. The four-variable equation is given by

$$\Delta y_t = f(t) + \sum_{i=1}^6 \psi_i \Delta m_{t-i} + \sum_{i=1}^{12} \phi_i \Delta y_{t-i} + \sum_{i=1}^{12} \zeta_i \Delta p_{t-i} + \sum_{i=1}^{12} \eta_i \Delta r_{t-i} + u_t, \quad (7)$$

where y is aggregate output, m is the nominal money supply, p is the Producer Price Index, and r is the secondary market rate on 90-day U.S. treasury bills. These variables, with the exception of r , are in logs, $f(t)$ gives the trend term, u is an error term, and Δ is the first-difference operator. The null hypothesis that money does not cause output is evaluated by testing if all the ψ_i coefficients are jointly insignificant. We follow Stock and Watson in setting the lag parameters and defining the alternative trend terms. Specifically the lag parameters are set to six for the money variable and to twelve for the others. Three alternative trend specifications are considered: (a) $f(t) = c$, (b) $f(t) = c + \alpha t$, and (c) $f(t) = c + \alpha_1 t + \alpha_2 t^2$.

The results of testing the hypothesis that money does not Granger cause output are present in Table 4. In Table 4, the results are based on the original Stock and Watson (1989a) specification; that is, M1 is the m variable and the sample is 1964:01 to 1985:12. When the IP data are used as a proxy for aggregate output, there is significant evidence that money Granger causes output for all three $f(t)$ specifications. The result is similar to that reported by Stock and Watson, who also used IP data in their analysis. However, a totally different conclusion is obtained if data on GDP, ICI, or XCI are used to test the same hypothesis. The null hypothesis that money does not Granger cause output cannot be rejected when output is measured by either GDP, ICI, or XCI. That is, the causation result reported in Stock and Watson (1989a) seems to be the artifact of using IP as a measure of aggregate output.

In recent years, M2, instead of M1, is commonly used to test for

TABLE 4. *Money-Output Causality Test Results*

	GDP	ICI	IP	XCI
A. With M1, 1964:01–1985:12				
<i>No time trend: $f(t) = c$</i>				
<i>F</i> -statistic	1.363 (0.230)	1.171 (0.322)	1.800 (0.100)	0.947 (0.462)
<i>Linear trend: $f(t) = c + \alpha t$</i>				
<i>F</i> -statistic	1.690 (0.124)	1.680 (0.127)	2.214 (0.043)	1.422 (0.207)
<i>Quadratic trend: $f(t) = c + \alpha_1 t + \alpha_2 t^2$</i>				
<i>F</i> -statistic	1.671 (0.129)	1.652 (0.134)	2.171 (0.047)	1.389 (0.220)
B. With M2, 1964:01–1993:12				
<i>No time trend: $f(t) = c$</i>				
<i>F</i> -statistic	1.929 (0.075)	2.121 (0.050)	3.071 (0.006)	2.379 (0.029)
<i>Linear trend: $f(t) = c + \alpha t$</i>				
<i>F</i> -statistic	1.547 (0.162)	1.830 (0.092)	2.760 (0.012)	2.053 (0.058)
<i>Quadratic trend: $f(t) = c + \alpha_1 t + \alpha_2 t^2$</i>				
<i>F</i> -statistic	1.966 (0.070)	2.091 (0.054)	3.674 (0.001)	2.401 (0.027)

NOTES: The table reports the *F*-statistics for testing the null hypothesis that money does not cause output when different measures of output are used. The model (Stock and Watson, 1989a)

$$\Delta y_t = f(t) + \sum_{i=1}^6 \psi_i \Delta m_{t-i} + \sum_{i=1}^{12} \varphi_i \Delta y_{t-i} + \sum_{i=1}^{12} \zeta_i \Delta p_{t-i} + \sum_{i=1}^{12} \eta_i \Delta r_{t-i} + u_t,$$

is used to test the hypothesis. Panel A contains the test statistics when M1 is used with the sample period of 1964:01–1985:12. Panel B reports the results when M2 is employed for 1964:01–1993:12. The *p*-values are given in parentheses below the statistics.

money-output causality. The second panel of Table 4 reports test results based on M2 and the sample 1964:01 to 1993:12. Compared with M1, M2 has a more significant effect on output measures. The *p*-values in Panel 4B are smaller than the corresponding ones in Panel 4A. Again, the level of significance appears to be related to the choice of output variables. M2 has the most statistically significant *p*-values when IP is used as the aggregate output variable and the least significant ones when GDP is used. The pattern is the same as that in Panel 4A. The results suggest that money supply tends to have different impacts on different measures of aggregate output.

To further compare the effects of money, we compute the impulse

Which Measure of Aggregate Output Should We Use?

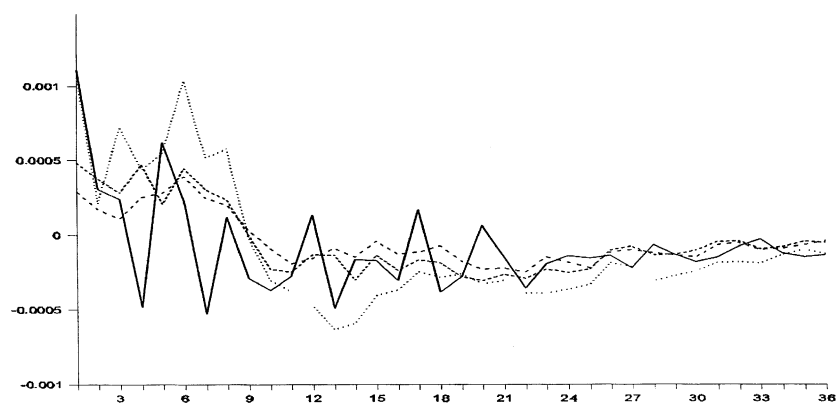


Figure 3A.
Impulse Responses to a Choleski M1 Shock, 1964:01–1985:12

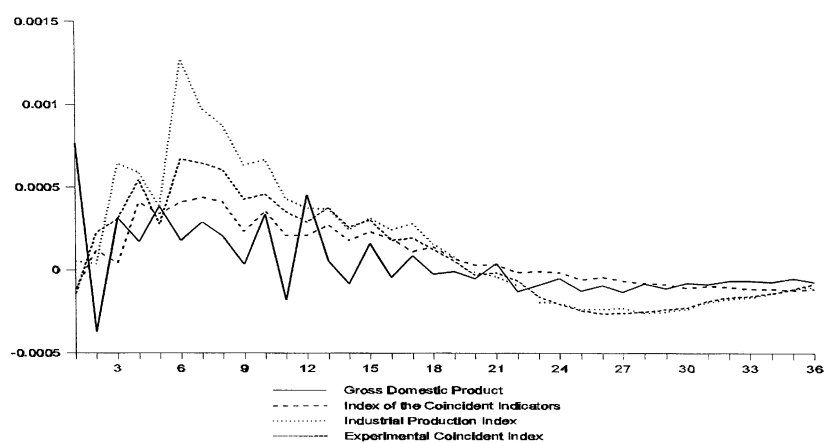


Figure 3B.
Impulse Responses to a Choleski M2 Shock, 1964:01–1993:12

responses of these four output measures to a Choleski-decomposed monetary shock using the Stock and Watson (1989a) system. The results are graphed in Figure 3. Figure 3A is based on the estimation results reported in Table 4A and Figure 3B is based on those in Table 4B. The patterns of impulse responses appear to converge after the first twelve periods (Walsh and Wilcox 1995; Christiano, Eichenbaum and Evans 1998). During the first twelve periods, nonetheless, the impulse responses look quite distinct. In

addition to the differences in magnitude, the four impulse response functions display dissimilar turning points. While the long-term monetary effects appear alike, the short-term effects are specific to individual aggregate output variables. The impulse response analysis reinforces the observation that different measures of aggregate output display considerable differences in short-run dynamics.

5. Concluding Remarks

Arguably aggregate output is the most important variable in macroeconomics. Output dynamics and the causes of output fluctuations are hotly contested research topics. Output is also an important element in other economic models. In the empirical literature, however, a number of variables are employed as a proxy for the theoretical aggregate output. In this paper we investigate the similarities and differences of four monthly measures of output: namely GDP, ICI, IP, and XCI. These four measures of output are found to share some common characteristics and, at the same time, to exhibit their own idiosyncrasies. Specifically, these output variables have a common long-run permanent component but distinct short-run cyclical patterns. The implication of distinct short-run cyclical patterns on empirical analysis is illustrated using the money-output causality regression. It is found that test results can be driven by the choice of the proxy for aggregate output.

Our results indicate that the four output measures react differently to monetary shocks. Money growth is found to significantly affect IP, which offers the most limited coverage of economic activity and ignores the growing service sector. On the other hand, there is weaker evidence that the more broadly defined GDP, ICI, and XCI are influenced by money growth. Impulse response analysis also indicates that the monetary effect depends on, among other things, which aggregate output variable is considered. Specifically, the choice of aggregate output variable has consequential implications on the output's short-term impulse responses to monetary shocks.

Our findings underscore the differences between alternative aggregate output measures and the sensitivity of regression results to the choice of an empirical output variable. Which measure of aggregate output should we use? We need some guidance from the theory. Ideally, the choice should be dictated by the economic issue under investigation. For instance, if the theoretical output variable refers to the general state of the economy, then a broad measure of economic activity should be considered. Under such a circumstance, *IP* is not the most preferable variable to use, especially for developed countries with growing significance of the service sector and declining importance of industrial production. From a theoretical perspective, GDP, ICI, and XCI provide a better empirical observation of aggregate

output as the first variable includes most economic activities and the other two are designed to track the general business cycle. The money-output regression exercise shows that the narrowly defined output variable *IP* can yield empirical results substantially different from those using GDP, ICI, and XCI. A potential future research topic is to investigate if a similar sensitivity issue is relevant for other economic variables and the related empirical analysis.

Received: June 1998

Final version: July 1999

References

- Baxter, Marianne, and Alan Stockman. "Business Cycles and the Exchange-Rate Regime: Some International Evidence." *Journal of Monetary Economics* 23 (May 1989): 377–400.
- Bernanke, Ben S., and Mark Gertler. "Inside the Black Box: The Credit Channel of Monetary Policy Transmission." *Journal of Economic Perspectives* 9 (Fall 1995): 27–48.
- Bernanke, Ben S., and Ilian Mihov. "Measuring Monetary Policy." *NBER Working Paper Series* 5145, 1995.
- Board of Governors of the Federal Reserve System. *Industrial Production—1986 Edition with a Description of the Methodology*. Washington, D.C.: Board of Governors of the Federal Reserve System, 1986.
- Campbell, John Y., and N. Gregory Mankiw. "International Evidence on the Persistence of Economic Fluctuations." *Journal of Monetary Economics* 23 (March 1989): 319–33.
- Cheung, Yin-Wong, Menzie Chinn and Tuan Tran. "How Sensitive Are Trends to Data Definition? Results for East Asian and G-5 Countries." *Applied Economic Letters* 2 (January 1995): 1–6.
- Cheung, Yin-Wong, and Kon S. Lai. "Finite-Sample Sizes of Johansen's Likelihood Ratio Tests for Cointegration." *Oxford Bulletin of Economics and Statistics* 55 (August 1993): 313–28.
- . "Lag Order and Critical Values of a Modified Dickey-Fuller Test." *Oxford Bulletin of Economics and Statistics* 57 (August 1995): 411–19.
- Chow, Gregory, and An-loh Lin. "Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series." *Review of Economics and Statistics* 53 (November 1971): 372–75.
- Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. "Monetary Policy Shocks: What Have We Learned and to What End?" *NBER Working Paper Series* 6400, 1998.
- Corrado, Carol, Charles Gilbert, and Richard Raddock. "Industrial Produc-

- tion and Capacity Utilization: Historical Revisions and Recent Developments." *Federal Reserve Bulletin* 83 (February 1997): 67–92.
- Elliott, Graham, Thomas J. Rothenberg, and James H. Stock. "Efficient Tests for an Autoregressive Unit Root." *Econometrica* 64 (July 1996): 813–36.
- Engle, Robert F., and Sharon Kozicki. "Testing for Common Features." *Journal of Business and Economic Statistics* 11 (October 1993): 369–80.
- Friedman, Benjamin M., and Kenneth N. Kuttner. "Money, Income, Prices, and Interest Rates." *American Economic Review* 82 (June 1992): 472–92.
- Green, George R., and Barry A. Beckman. "Business Cycle Indicators: Upcoming Revision of the Composite Indexes." *Survey of Current Business* 73 (October 1993): 44–51.
- Johansen, Søren. "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models." *Econometrica* 59 (November 1991): 1551–80.
- Mark, Nelson. "Exchange Rates and Fundamentals: Evidence on Long Horizon Predictability." *American Economic Review* 85 (March 1995): 201–18.
- Meese, Richard A., and Kenneth Rogoff. "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics* 14 (February 1983): 3–24.
- Phillips, Keith R., Lucinda Vargas, and Victor Zarnowitz. "New Tools for Analyzing the Mexican Economy: Index of Coincident and Leading Economic Indicators." Federal Reserve Bank of Dallas *Economic Review* (Second quarter 1996): 2–15.
- Romer, Christina D. "Spurious Volatility in Historical Unemployment Data." *Journal of Political Economy* 94 (February 1986a): 1–37.
- . "Is the Stabilization of the Postwar Economy a Figment of the Data?" *American Economic Review* 79 (June 1986b): 314–34.
- Stock, James H., and Mark W. Watson. "Interpreting the Evidence on Money-Income Causality." *Journal of Econometrics* 40 (January 1989a): 161–81.
- . "New Indexes of Coincident and Leading Economic Indicators." In *NBER Macroeconomics Annual 1989*, edited by Olivier J. Blanchard and Stanley, 351–94. Fischer. Cambridge: The MIT Press, 1989b.
- . "A Probability Model of the Coincident Economic Indicators." In *Leading Economic Indicators: New Approaches and Forecasting Records*, edited by Kajal Lahiri and Geoffrey H. Moore, 63–89. Cambridge: Cambridge University Press, 1991.
- Walsh, Carl E., and James A. Wilcox. "Bank Credit and Economic Activity."

*Which Measure of Aggregate Output Should We Use?*⁹

In *Is Bank Lending Important for the Transmission of Monetary Policy?*, edited by Joe Peek and Eric S. Rosengren, 83–112. Boston: Federal Reserve Bank of Boston, 1995.

Wilcox, James A. “Disaggregating Data Using Related Series.” *Journal of Business and Economic Statistics* 1 (July 1983): 187–91.