

The Econometric Analysis of Business Cycles

James H. Stock

Jan Tinbergen's pioneering work on empirical macroeconomic models has shaped business cycle research ever since and thus has framed our current understanding of the business cycle. His models, first of the Dutch economy and then of the U.S. economy, have several essential features that are present in many modern models.

In Tinbergen's models, business cycles were treated as the outcome of shocks, or impulses, that propagate through the economy over time to result in complicated dynamic patterns; even though the individual equations of the model were linear with simple, typically single-period lag structures, the resulting system could exhibit cyclical dynamics. The individual equations of the model were motivated by economic theory, and the model itself provided a framework for linking a large number of variables. His work, as described in the first volume of his report to the League of Nations, entailed many important details that continue to be part of modern macroeconomic methodology. Notably, his emphasis on testing business cycle theories led to the evaluation of these models by their forecasting ability and to checking for the stability of their parameters over time. Finally, Tinbergen used his models both for positive and normative analysis, that is, both to evaluate economic theories and to provide a tool for the analysis of macroeconomic policy.

This essay looks at one aspect of modern business cycle analysis in the light of Tinbergen's early contributions: the current state of knowledge about the structural stability of macroeconomic relations. In Tinbergen's time, methods for detecting structural changes entailed reestimating parameters using small subsets of the data and, absent measures of sampling uncertainty, drawing expert judgments about the results. Econometricians now have a large set of statistical tools for analyzing the stability of equations and systems of equations, and these tools have been applied to macroeconomic data sets. As I discuss below, the result has been an empirical finding of widespread instability in macroeconomic relations. This in turn poses substantial problems for practical forecasters and policy analysis.

Changes in the Postwar Business Cycle

Over the past fifteen years, econometricians have produced a collection of methods for modeling structural breaks and time-varying parameters in economic relations. Two such methods are tests for structural breaks, notably the so-called Quandt (1960) likelihood ratio (QLR) test and analysis of stability based on pseudo out-of-sample forecasts. I

show that, for the U.S., both methods suggest widespread evidence of structural instability among macroeconomic relations. I then turn to a particularly interesting change in the modern business cycle: the recent moderation in the cyclical volatility of aggregate output in many developed economies.

Methods for Identifying Structural Changes

The QLR test for structural breaks.

The first step towards identifying a structural break in a macroeconomic time series is having a reliable test for a structural break, that is, a test that has controlled size under the null of no break and good power against the alternative of a break. One such test is the Quandt likelihood ratio test. To be concrete, consider the linear regression model with a single stationary regressor X_t ,

$$Y_t = \beta_0 + \beta_{1t}X_t + u_t \quad (1)$$

where u_t is a serially uncorrelated error term, Y_t is the dependent variable, and β_0 and β_{1t} are regression coefficients.

The unusual feature of is that the slope coefficient can change over time. In particular, suppose that the regression coefficient changes at some date τ :

$$\beta_{1t} = \begin{cases} \beta_1 & t \leq \tau \\ \beta_1 + \delta & t > \tau \end{cases} \quad (2)$$

If the break date τ is known, then the problem of testing the null hypothesis of no break (that is, $\delta = 0$) against the alternative of a nonzero break ($\delta \neq 0$) is equivalent to testing the hypothesis that the coefficient δ is zero in the augmented version of

$$Y_t = \beta_0 + \beta_{1t}X_t + \delta Z_t(\tau) + u_t \quad (3)$$

where $Z_t(\tau) = X_t$ if $t > \tau$ and $Z_t = 0$ otherwise. This test can be computed using a conventional t -statistic from estimating by ordinary least squares; calling this $t(\tau)$, the hypothesis of no break is rejected at the 5% significance level if $|t(\tau)| > 1.96$.

In practice, τ is typically unknown so the test in the pre-

Series	p -value	break date
GDP (growth rate)	0.98	.
Production of goods (total) (growth rate)	0.92	.
Production of nondurable goods (growth rate)	0.09	.
Production of durable goods (growth rate)	0.77	.
Production of services (growth rate)	0.00	1968:3
Production of structures (growth rate)	0.02	1991:3
Nonagricultural employment (growth rate)	0.03	1981:2
Price inflation (GDP deflator)	0.00	1973:2
90-day T-bill rate (first difference)	0.00	1981:1
10-year T-bond rate (first difference)	0.02	1981:1

Table 1: QLR Test for Changes in Autoregressive Parameters: Quarterly U.S. Data, 1959 - 2002.

Notes: The first column show the p -value for the QLR sttistic testing the stability of all the autoregressive coefficients in an AR(4). The second column shows the least squares estimate of the break date when the QLR statistic is significant at the 5% level. Source: Stock and Watson (2002a, Table 3).

ceding paragraph cannot be implemented. However, the t -statistic can be computed for all possible values of τ in some range. If the largest value of the absolute t -statistic exceeds some critical value, then the hypothesis of no break can be rejected. The difficulty with this method is that the critical value is not 1.96, because you are providing yourself with multiple opportunities to reject the null hypothesis. The distribution has, however, been worked out (Andrews, 1993), and if the values of τ considered are those in the central 70% of the sample then the relevant 5% critical value is 2.95.

To extend this method to multiple regressors, one needs to compute the F -statistics testing the hypothesis that the additional regressors are zero for each value of τ , then compute the maximum of these. This is commonly called the QLR (or sup-Wald) test statistic for a structural break; for a table of critical values, see Stock and Watson (2003, Table 12.5).

Pseudo out-of-sample forecast comparisons

An alternative approach to assessing structural stability builds on Tinbergen's intuition that one way to test a model is to look at its forecasting performance: a change in its forecasting performance over the sample suggests changes in the model coefficients. Although true forecasting occurs in real time, historical forecasting experience can be simulated by first estimating the model using data through an earlier date (for example, 1990), then making a forecast one or more periods ahead (for example, for 1991), then moving forward one period and repeating this exercise. This produces a sequence of forecasts and forecast errors that simulate the forecasts that would have been made using the model in real time. If the resulting pseudo out-of-sample performance of the model changes, relative to a benchmark forecasting model, then this can be taken as evidence of a shift in the parameters of the model. When this comparison is made using non-nested models, formal statistical comparisons can be made using the statistics proposed in West (1996).

Evidence of Instability in Reduced-Form Forecasting Models

What happens when these methods are applied to relations that are used for macroeconomic forecasting? Are the coefficients of econometric models stable, or do they appear to change over time?

One way to address these questions is to examine Tinbergen's 'final equation' for econometric models, which (absent exogenous variables) is the univariate autoregressive representation of the time series entering the model. Table 1 presents the results of applying QLR tests to quarterly U.S. data from 1959 - 2002 on major macroeconomic variables. The first column reports the p -value of the test of stability of all the autoregressive parameters (so a p -value $\leq .05$ indicates rejection at the 5% significance level). If the test rejects at the 5% significance level, the second column reports the estimated break date, where the break date is estimated by least squares estimation of the model augmented by the interaction variables (as in (3)), where τ is estimated along with the regression coefficients by least squares. The results in Table 1 point towards substantial instability in these autoregressions; although the hypothesis of stability is not rejected for GDP or the goods production series, it is rejected for other production measures, employment, inflation, and interest rates.

A similar conclusion of widespread instability is reached when pseudo out-of-sample forecasting methods are used. In Stock and Watson (2003b), we review the ability of asset prices to forecast real output growth and inflation over the past four decades in seven developed economies. Using pseudo out-of-sample forecasts, we find that good forecasting performance of an asset price in one period or in one country does not imply that the same asset price will be a good predictor in another time period or another country.

The results in Table 2 are indicative of the findings in the broader study of Stock and Watson (2003b). That table reports the root mean squared forecast error for pseudo out-of-sample forecasts, relative to a benchmark model, where the benchmark is a univariate autoregression with the number of

Predictor	1971-1984	1985-1999
Federal Funds rate (level)	0.78	1.42
90-day Treasury bill rate (level)	0.85	1.06
Term spread (long bond rate minus Federal Funds rate)	0.48	2.51
Real stock return	0.90	1.27
Percentage change in real M2	0.57	1.41

Table 2: Mean Square Forecast Errors for Pseudo Out-of-Sample Forecasts of Real GDP Growth, 4 quarters ahead (Univariate autoregression = 1.0): Quarterly U.S. Data, 1959 - 1999.

Notes: Each entry is the mean squared pseudo out-of-sample forecast errors (MSFE) for a candidate model with lags of GDP growth and lags of the additional predictor, relative to the corresponding MSFE for a univariate autoregression; all lags are chosen by BIC. An entry less than 1.0 means that the candidate model outperformed the univariate benchmark during this period. Source: Stock and Watson (2003b, Table 3).

lags chosen using the Bayes information criterion (BIC); a value of 1.0 means that the forecast errors of the candidate model is the same as that of the benchmark autoregression. For the series shown in Table 2, all the candidate models outperformed the autoregressive benchmark during 1971 - 1984, but all produced relatively poor forecasts during the later period of 1985 - 1999. In some cases, such as forecasts based on the term spread, both the early gain and the later deterioration are very large. In some cases, not reported in Table 2, forecast performance was better in the second period than in the first. The general pattern in Table 2 of changes in relative forecast performance is typical for other predictors, other forecast horizons, and countries other than the U.S.

Recent Changes in the Business Cycle

The discussion so far has focused on changes in the coefficients of reduced form models, which implies that some important coefficients of the original structural model have been changing. Recently, there has been considerable interest in a different apparent change in the postwar business cycle: a marked decline in the volatility of overall economic activity in a number of developed economies, that is, an overall moderation in the business cycle.

This moderation, originally noted for the U.S. by Kim and Nelson (1999) and McConnell and Perez-Quiros (2000), can be seen by examining the evolution of the standard deviation of annual GDP growth. As is summarized in Table 3, GDP growth was relatively quiescent in the 1960s, was more volatility in the 1970s and 1980s, and during the 1990s was even less volatile than during the 1960s. Kim and Nelson (1999) and McConnell and Perez-Quiros (2000) independently concluded that there was a break in the volatility of GDP in the mid 1980s, a conclusion that they reached using different methods (Kim and Nelson (1999) used a Bayesian stochastic volatility model and a nonGaussian smoother, whereas McConnell and Perez-Quiros (2000) used classical break tests like the QLR). Although there has been some debate about whether this reduction in volatility is actually a sharp break or, as suggested by Blanchard and Simon (2001), part of a longer trend to increasingly moderate cycles, there is little debate about whether there actually was a decline in volatility. Moreover, as shown by van Dijk, Osborn and

Sensier (2002), this moderation is evident in other developed economies as well.

Although there is little disagreement about the fact that this moderation occurred, there is no consensus concerning its source. As summarized in Stock and Watson (2002a), a variety of explanations have been proposed for this reduction in volatility. These explanations include changes in the structure of modern economies, such as reduced credit constraints among consumers or changes in production technologies; changes in policy regimes, in particular a switch towards more steadfast monetary policies that result in reduced sensitivity of the economy to economic shocks; and simply a reduction in the magnitude of recent economic shocks. In Tinbergen's framework, some of these explanations entail changes in the magnitude of the impulses to the macroeconomy, while others involve changes in the propagation mechanism. It is important to understand which of these proposed explanations is actually right. If the source of the moderation is improved policy or changes in the structure of the economy, then we might expect this moderation to persist. On the other hand, if this moderation is merely reflects a decade of good luck – that is, small macroeconomic shocks – then macroeconomic managers and forecasters should be ready for a return to the more turbulent business cycles of the past as soon as this period of good luck ends.

Summary

The findings of Tables 1 - 3 are typical of broader evidence of instability in macroeconomic relations; for addition-

Sample period	Standard deviation (%)
1960-1969	2.0
1970-1979	2.7
1980-1989	2.6
1990-2001	1.5

Table 3: Summary statistics for Four-Quarter Growth in Real GDP, 1960 - 2001

Notes: Summary statistics are shown for $100\ln(GDP_t/GDP_{t-4})$, where GDP_t is the quarterly value of real GDP. Source: Stock and Watson (2002a, Table 1).

al references and results, see Stock and Watson (1996) and Hansen (2001). As the results in Table 2 make clear, this instability poses a substantial practical problem for real-time forecasters and for other users of macroeconomic models. Developing methods for detecting this instability in real time, for modeling this instability, and for making forecasts that are robust to this instability continues to be an important line of research that follows in the path of Tinbergen's early checks for coefficient stability.

Although I have focused on just one of aspect of Tinbergen's legacy, important research in business cycle modeling continues in many other areas that Tinbergen pioneered. One of these areas is the development of models with a large number of variables. By the standards of his day, Tinbergen's models included very many variables. To the extent that there are many sources of shocks or impulses to the economy, in theory increasing the number of variables should increase the amount of information in the model and thus has the potential to improve forecast performance. In practice, however, the performance of structural macroeconomic models has not improved with their size, perhaps because of specification problems, compounding of estimation error, or model instability. Thus researchers have recently started to explore a second route towards high-dimensional modeling, one of high-dimensional multivariate reduced form models. One strand of this research is to model a large number - hundreds - of macroeconomic variables as having most their comovements arising from a small number of common macroeconomic sources, or factors, plus individual idiosyncratic disturbances or errors in the series such as might arise from series-specific measurement error. This reasoning leads to what is known as a dynamic factor model. Recent research (Forni, Hallin, Lippi, and Reichlin (2000, 2001) and Stock and Watson (1999, 2002b) suggests that large dynamic factor models can provide a coherent framework for economic forecasting and for obtaining improved measures of the underlying business cycle in one or more economies.

A key objective of macroeconomic modelers in Tinbergen's tradition has been to develop quantitative tools for policy analysis. Achieving this goal has, however, proven more elusive than some of the other objectives of Tinbergen's research program, for it requires estimation of dynamic causal effects from macroeconomic data. We now understand that many of the relations appearing in large simultaneous equations macroeconomic models are not really structural, but really are reduced-form equations that do not in general lead to consistent estimates of dynamic causal effects. Much academic work on estimation of dynamic causal effects in macroeconomics now takes one of two approaches: either single-equation (instrumental variable) estimation or system estimation using structural vector autoregression, where the structural elements come from imposing an economic model on the forecast errors. There is an ongoing debate about whether these methods are scientifically convincing or simply quantify the predispositions of the researchers (see Rudebusch, 1998). It is here, in my view, that the most work remains if

we are to achieve Tinbergen's vision of a stable quantitative structural macroeconomic model that can be used to design macroeconomic policies to control inflation and the business cycle.

About the author

James Stock is Professor of Economics at Harvard University, Cambridge.

References

- Andrews, D.W.K. 1993. "Tests for Parameter Instability and Structural Change with Unknown Change Point", *Econometrica*, 61, 821 – 856.
- Blanchard, O. and J. Simon. 2001. "The Long and Large Decline in U.S. Output Volatility," *Brookings Papers on Economic Activity* 2001:1, 135 – 164.
- Dijk, D. van, D.R. Osborn, and M. Sensier. 2002. "Changes in Variability of the Business Cycle in the G7 Countries," manuscript, Erasmus University Rotterdam.
- Forni, M., M. Hallin, M. Lippi and L. Reichlin. 2000. "The Generalized Factor Model: Identification and Estimation," *Review of Economics and Statistics*, 82, 540–554.
- Forni, M., M. Hallin, M. Lippi and L. Reichlin. 2001. "Do Financial Variables Help Forecasting Inflation and Real Activity in the EURO Area?," CEPR Working Paper
- Hansen, B.E. 2001. "The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity," *Journal of Economic Perspectives* 15, no. 4, 117 – 128.
- Kim, C.-J. and C.R. Nelson. 1999. "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle," *The Review of Economics and Statistics* 81, 608 – 616.
- McConnell, M.M and G. Perez-Quiros. 2000. "Output Fluctuations in the United States: What has Changed Since the Early 1980's" *American Economic Review*, Vol. 90, No. 5, 1464-1476.
- Rudebusch, G.D. 1998. "Do Measures of Monetary Policy in a VAR Make Sense?" *International Economic Review*, 39, 907 – 931.
- Stock, J.H. and M.W. Watson. 1996. "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics* 14, pp. 11 – 29.
- Stock, J.H. and M.W. Watson. 1999. "Forecasting Inflation," *Journal of Monetary Economics* 44, 293–335.
- Stock, J.H. and M.W. Watson. 2002a. "Has the Business Cycle Changed and Why?," forthcoming, *NBER Macroeconomics Annual 2002*. Cambridge: MIT Press.
- Stock, J.H. and M.W. Watson. 2002b. "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, Vol. 20 No. 2, 147-162.
- Stock, J.H. and M.W. Watson. 2003a. *Introduction to Econometrics*. Boston: Addison Wesley Longman.
- Stock, J.H. and M.W. Watson. 2003b. "Forecasting Output and Inflation: The Role of Asset Prices," forthcoming, *Journal of Economic Literature*.
- West, K.D. 1996. "Asymptotic Inference about Predictive Ability," *Econometrica* 64, pp. 1067 – 1084.