In This Issue:

Consumer Attitudes:
King for a Day


Path-Dependent Options: Valuation and Applications

Review Essay
Consumer sentiment is often cited as an explanation for economic events that standard variables have failed to predict. This article seeks to determine whether consumer attitude surveys are informative, independently of other available information, about future economic conditions. From his empirical tests the author concludes that, when used alone, unanticipated changes in consumer sentiment are correlated with movements in industrial production and unemployment. This correlation becomes very weak, however, when a small set of readily available financial information is included in the tests. The author cautions that relying on sentiment can distract policymakers from focusing on the underlying causes of economic fluctuations.

The rapid pace of urban growth during this century has created a great demand for regional economic models for use in urban and regional planning, forecasting, and policy-making. This article critically reviews the literature on the most popular regional model—the economic base model. From his own empirical work and his survey of research, including recent work that has applied modern macroeconomic time-series techniques to the economic base model, the author concludes that the model has severe limitations, especially for longer-term economic planning and policy analysis.

Since being introduced more than a decade ago, path-dependent options have filled several niches in derivative securities markets. This article discusses the valuation of two popular kinds of path-dependent options—lookback and average-rate options—and, in this context, gives insights into their advantages and risks.

The State of Macroeconomics
Edited by Seppo Honkapohja
Consumer Attitudes: King for a Day

Eric M. Leeper

Every so often economic observers latch onto consumer attitudes as an explanation for economic events that seem otherwise inexplicable. The recession that started in the fall of 1990 is but the most recent example of this phenomenon. When Iraq invaded Kuwait in August 1990, leading surveys indicated that consumer sentiment had plummeted to its lowest level in a decade. Paralleling the decline in sentiment was a sharp drop in economic activity and the end of nearly eight years of American economic expansion. The minutes of the Federal Open Market Committee meeting in December 1990 list the depressed level of sentiment first among factors that contributed to weak consumer spending. By the summer of 1991 policymakers announced that the recession had likely bottomed out in the spring. Coinciding with the announcements, sentiment had risen to well above its preinvasion level. When the level of consumer confidence fell again in October 1991, the decline made headlines, and analysts sought to determine if it portended a prolonged recession.

This article looks at the role that consumer sentiment plays in forecasting economic activity. Although the idea that sentiment is highly significant is intuitively appealing, it turns out that sentiment adds very little to such forecasts. Other easily available variables tend to dominate the information contained in sentiment, even when the changes in consumer attitudes are large and come as a surprise, so the changes are not merely reflecting past economic conditions. Careful statistical analysis strips away the casual associations typically made between sentiment and economic events to reveal that consumer attitudes are a “king-for-a-day” variable that was enthroned when all the usual variables failed to predict the recent recession. Attitudinal data have since fallen from grace after falsely forecasting both a strong recovery...
early last year and a double-dip recession late the same year. Empirical evidence suggests that the de-throning was well deserved.

**From Animal Spirits to Impulse Responses**

In their analyses of business cycle fluctuations, policymakers, economic forecasters, and macroeconomists ascribe one of at least three roles to consumer attitudes: a causal variable, a catalyst for other shocks, and a summary statistic that simply reflects information contained in other economic data. Rather than trying to distinguish which of these roles consumer sentiment surveys play, this article seeks to determine whether attitudinal data are informative, independently of other available information, about future economic conditions. If so, then survey results may be valuable to economic forecasters and policymakers. If, however, the information in attitudinal surveys is already contained in other variables regularly used to forecast economic conditions, the survey results are redundant and probably can be ignored.

Among economists who view attitudes as a causal variable is A.C. Pigou, who lists one “initiating impulse” of business cycles as “psychological causes . . . [which] are changes that occur in men’s attitudes of mind, so that, on a constant basis of fact, they do not form a constant judgement” (1927, 30). John Maynard Keynes’s “animal spirits,” which induce “sudden and violent changes” in investment, are generated in part by changes in consumers’ tastes and by expectations of the strength of future aggregate demand (1936, 315; also see 137-62). George Katona and Eva Mueller developed the survey that underlies the University of Michigan Index of Consumer Sentiment to measure “those factors which are capable of giving rise to independent variation in the rate of consumer spending and saving, namely, changes in people’s perceptions, attitudes, motivations, and expectations” (1953, 1).

The Michigan survey data are included in some econometric models. The Data Resources Inc./McGraw-Hill (DRI) model, for example, allows consumer sentiment to affect a wide variety of private decisions directly, and it “indirectly affects virtually all other sectors of the economy” (David Kelly 1990, 18). During the Persian Gulf War, DRI ran a variety of simulations that involved, among other things, assuming an “optimistic” path for consumer sentiment with the prediction that “consumer and business confidence recover quickly, aborting the recession after only one quarter” (DRI/McGraw-Hill 1991, 4). These simulations amount to treating consumer attitudes as an exogenous variable that causes shifts in certain behavioral relations. Indeed, Roger Brinner, director of research at DRI/McGraw-Hill, was quoted as saying, “If consumers hadn’t panicked [in August 1990] there wouldn’t have been a recession” (Sylvia Nasar 1991a, D4).

A slight twist on the causal view is that consumer attitudes are a catalyst for shocks that generate business cycles. Gottfried von Haberler summarizes “psychological theories” of the business cycle, which include those of Pigou and Keynes: “Optimism and pessimism are regarded as causal factors which tend to induce or intensify the rise and fall of investment which are characteristic of the upswing and downswing . . . ” (1938, 136) (emphasis added). In this view, increases in income that do not coincide with enhanced consumer optimism may not produce the expected increase in economic activity. The DRI model is specified so that consumer sentiment can be treated as a catalyst for economic activity. In the model, variables such as employment, interest rates, and oil prices influence spending through their effect on consumer sentiment, so fluctuations in these variables that do not coincide with changes in consumer sentiment will have a less pronounced effect on the economy (see Otto Eckstein 1983 or Kelly 1990).

Finally, surveys of consumer attitudes are widely thought to predict economic conditions. Policymakers sometimes cite fluctuations in consumer confidence surveys as providing evidence of changes in future economic activity: “Consumer and business attitudes were seen as a critical factor bearing on the prospective performance of the economy” (Board of Governors 1991b, 7). In addition, the U.S. Department of Commerce includes the University of Michigan Index of Consumer Expectations in its Composite Index of 11 Leading Indicators. By this view, attitudinal surveys are useful because they summarize prevailing economic conditions, even though attitudes may have no independent causal effect on behavior.

Whatever one’s view of the role of consumer attitudes, the survey data are taken seriously by a wide range of economic observers. Each of the three roles ascribed to consumer attitudes should help predict economic conditions. Moreover, if the predictive value is strong enough to warrant the prominent position that sentiment has occupied in recent discussions, the predictive value should not disappear when a small set of readily available data is taken into account.
Most empirical work adopts a Keynesian perspective by assuming that attitudes directly affect individuals’ consumption behavior. In contrast, this article concentrates on how well attitudes forecast industrial production and unemployment. The Index of Consumer Expectations is based on questions about the respondents’ expectations of general economic conditions rather than about their consumption expenditure plans per se, so there is not necessarily a direct connection between the index and consumption behavior. Moreover, even if attitudes work through consumption to influence output and employment, these variables will nonetheless be related to attitudes, so there is no harm in skipping the link to consumption.

This article adopts an agnostic view of how consumer attitudes may help to predict economic conditions. The empirical tests do not force consumer attitudes to influence economic conditions according to any particular theory of economic behavior. Although the purely statistical perspective of the empirical analysis is silent on exactly how attitudes influence economic behavior, the results can show whether changes in attitudes are associated with subsequent changes in economic variables.

Practical policy and forecasting exercises use a wide variety of information to generate forecasts. For consumer attitudes to contribute to these exercises they must contain information about future economic conditions that is independent of the data regularly used. To some extent consumer attitudes will always reflect the state of the economy. When interest rates and unemployment have been high consumers are likely to feel dissatisfied; a recent history of low rates and strong economic growth is likely to leave consumers optimistic. Such swings in attitudes that are related to recent economic performance will be predictable. Some changes in attitudes, however, will not be predictable from past economic information. For example, the large drop in sentiment at the beginning of the Persian Gulf conflict was in response to noneconomic events. It is these unexpected changes in attitudes that potentially contain news that alters consumers’ behavior and the subsequent course of the economy.

The first task of the empirical evaluation of the Index of Consumer Expectations is to isolate its unanticipated changes. These surprises are defined as changes in the index that cannot be predicted using historical values of economic variables.

With the unanticipated changes identified, the evaluation next asks whether such changes actually are systematically related to industrial production and unemployment. There are three parts to the analysis. First, the historical correlations between surprises in sentiment and subsequent movements in economic variables are summarized. Second, the historical correlations are used to produce forecasts of industrial production and unemployment, which are compared to the actual paths of the forecasted variables. Finally, the evaluation asks whether “big” surprises in attitudes are strongly associated with fluctuations in economic variables.

It turns out that unanticipated changes in consumer attitudes are correlated with large subsequent movements in industrial production and unemployment when the estimated econometric model includes only these three variables. This correlation all but disappears, however, once information embodied in real (inflation-adjusted) stock prices and short-term nominal interest rates is taken into account; these financial variables appear to absorb the correlation between sentiment and the measures of economic activity. Accuracy of production and unemployment forecasts is sometimes improved by adding consumer attitudes when financial variables are absent. When the forecasts are based on the information in stock prices and the interest rate, adding consumer attitudes often increases the forecast errors. Finally, there is no systematic evidence that unanticipated “big” changes in consumer attitudes are associated with movements in industrial production and unemployment once the information contained in financial variables is factored into the analysis.

A Careful Look at the Evidence

The data set examined in this article includes quarterly observations from 1954:1 to 1992:1 of total industrial production, the civilian unemployment rate, the real Standard and Poor’s (S&P) 500 stock price index, the three-month Treasury bill rate, and the Index of Consumer Expectations (hereafter referred to as “attitudes” or “sentiment”). The index is plotted in Chart 1, which also marks the National Bureau of Economic Research’s business cycle peaks and troughs. Many, but not all, economic downturns have been associated with declines in the index. More importantly from the perspective of a forecaster using consumer attitudes, there have been a number of declines in the index that did not coincide with the peak of a business cycle.
Historical Correlations. Informal discussions of attitude surveys' forecasting ability frequently graph consumer attitudes along with some economic variable to illustrate that the two series are correlated (see Richard T. Curtin [1990] or Alan C. Garner 1991). One way to formalize this procedure, which emphasizes the "news" aspects of changes in consumer attitudes, is to estimate a vector autoregression (VAR) that includes consumer attitudes, industrial production, and the unemployment rate. The VAR is a procedure that summarizes the average correlations among the variables at different points in time. Mechanically, the VAR consists of an equation for each variable in which the equations are estimated by regressing each of the three variables against lagged values of all three variables. By not imposing any particular theoretical connection among the variables, the VAR will capture any correlations that exist in the data.

Panel A of Table 1 reports the results of $F$-tests that coefficients on lagged values of each variable are jointly zero. If, for example, all the coefficients on past values of industrial production in the attitudes equation are not significantly different from zero, then industrial production may not help to predict attitudes. This result would be reported in Table 1 as a number greater than approximately .10 in the cell associated with the sentiment row and the industrial production column. According to the table, past consumer attitudes predict industrial production and unemployment, but past values of production and unemployment do not individually help to predict sentiment. Sentiment is not statistically exogenous, in the sense that it is predicted only by its own history, because past values of production and unemployment jointly predict sentiment.

A better way to evaluate the predictive power of consumer attitudes is to calculate the responses of all the variables to "typical" one-standard-deviation impulses (or innovations) in each variable. These responses are a type of correlation between variables at different times. If surprises in sentiment are highly correlated with future industrial production and unemployment, then sentiment innovations will be followed
Table 1
F-Tests of Exclusion Restrictions for the VARs

A: VAR without Financial Variables

<table>
<thead>
<tr>
<th>Equation</th>
<th>Sentiment</th>
<th>Industrial Production</th>
<th>Unemployment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentiment</td>
<td>.000</td>
<td>.536</td>
<td>.319</td>
</tr>
<tr>
<td>Industrial Production</td>
<td>.040</td>
<td>.000</td>
<td>.021</td>
</tr>
<tr>
<td>Unemployment</td>
<td>.009</td>
<td>.002</td>
<td>.000</td>
</tr>
</tbody>
</table>

The null hypothesis that sentiment depends only on its own past values can be rejected at the .039 level of marginal significance.

B: VAR with Financial Variables

<table>
<thead>
<tr>
<th>Equation</th>
<th>Stock Prices</th>
<th>T-Bill</th>
<th>Sentiment</th>
<th>Industrial Production</th>
<th>Unemployment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock Prices</td>
<td>.000</td>
<td>.193</td>
<td>.534</td>
<td>.643</td>
<td>.614</td>
</tr>
<tr>
<td>T-Bill</td>
<td>.063</td>
<td>.000</td>
<td>.131</td>
<td>.657</td>
<td>.495</td>
</tr>
<tr>
<td>Sentiment</td>
<td>.506</td>
<td>.410</td>
<td>.000</td>
<td>.257</td>
<td>.621</td>
</tr>
<tr>
<td>Industrial Production</td>
<td>.006</td>
<td>.247</td>
<td>.692</td>
<td>.000</td>
<td>.020</td>
</tr>
<tr>
<td>Unemployment</td>
<td>.028</td>
<td>.105</td>
<td>.420</td>
<td>.004</td>
<td>.000</td>
</tr>
</tbody>
</table>

The null hypothesis that sentiment depends only on its own past values can be rejected at the .038 level of marginal significance.

---

by large movements in the two variables. The solid lines in Chart 2 are the 95th percentile probability bands for the response functions for twelve quarters after the innovation.\(^{13}\) When the bands do not straddle the zero axis, there is a high probability the responses are nonzero. Each row in the chart shows the response of that row’s variable to an unanticipated increase in each of the variables listed in the columns.

In the impulse response functions sentiment is allowed to influence production and unemployment contemporaneously, but production and unemployment influence sentiment with a one-quarter lag. Treating consumer sentiment as predetermined in this way is consistent with the fact that survey results are known before output and employment data are released. Positive innovations in sentiment are associated with substantially higher production and lower unemployment, as shown in the first column of the chart. Comparing the responses across rows, the production and unemployment responses to sentiment innovations are at least as large as the responses to their own innovations. This result indicates that, given the
Chart 2

Historical Correlations: Impulse Responses from VARs

Responses to Innovations in:

- Sentiment
- Industrial Production
- Unemployment

The rows graph the responses of sentiment, industrial production, and unemployment to a one-standard-deviation innovation in each of the three variables for twelve quarters following the innovation. The solid lines show bands for the VAR without financial variables; the dashed lines show bands for the VAR with financial variables. The bands are 95th percentile probability bands based on 1,000 draws of a Monte Carlo experiment. Sentiment is allowed to influence industrial production and unemployment within the quarter, and, in the five-variable system, real stock prices and the Treasury-bill rate are allowed to influence sentiment, industrial production, and unemployment within the quarter. Sentiment and industrial production are measured in percentages; unemployment is measured in percentage points. The three-variable VAR is estimated from 1956:1 to 1992:1, and the five-variable VAR is estimated from 1955:3 to 1992:1. The responses of sentiment and industrial production are multiplied by 100. (Responses of stock prices and the interest rate and the responses to innovation in these variables are not pictured.)
information embodied in the three-variable VAR, unan-
ticipated changes in sentiment systematically signal
strong subsequent movements in industrial production
and unemployment. Sentiment responds strongly to
production innovations and less strongly to unemploy-
ment innovations, as shown in the first row of the
chart. This chart is consistent with the casual view that
sentiment is correlated with economic conditions, with
the strongest correlations arising a year or more after a
surprise change in sentiment.

One argument for why attitudes are a useful variable
for forecasting is that survey results are available with
minimal delay. Of course, a wide variety of financial
variables are also readily available, and for forecasting
purposes one would like to know what the marginal
contribution of attitudes is, given other available data.
To answer this, a real stock price index and the three-
month Treasury-bill rate are added to the VAR.

Panel B of Table 1 reports that lagged values of con-
sumer sentiment no longer help predict industrial pro-
duction and unemployment. Real stock prices and, to
a lesser extent, interest rates appear to have absorbed
the predictive power of attitudes.

The predictive power of unanticipated changes in
sentiment also diminishes in the expanded VAR with fi-
nancial variables, indicating that there is little informa-
tion in consumer attitudes about economic conditions
that is independent of financial information. The dashed
lines in Chart 2 are the 95th percentile probability bands
for the impulse response functions from the five-variable
system. (The chart reports only the response functions for
the three variables of interest.) The results do not take a
position on whether financial variables reflect consumer
sentiment. Instead, the responses to sentiment innova-
tions report the correlations of sentiment with production
and unemployment, given information on current stock
prices and interest rates and past values of all the vari-
bles. Looking down the first column of the chart, senti-
ment innovations are associated with small increases in
industrial production for only one quarter following the
innovation. Unemployment falls for three quarters but by
only one-third as much as it does in the VAR without
financial variables. As seen in the second row of the
chart, relatively more of the variability of unanticipated
changes in industrial production is now associated with
production innovations than with sentiment innovations.

If consumer sentiment is a summary statistic for
other variables, its predictive power would be expected
to diminish as we control for other available infor-
mation. The marked decline in predictive power shown in Chart 2, which occurs after controlling for
only two more variables, is surprising.

Forecasts Using Sentiment. To be a useful addition
to a forecaster's arsenal, consumer sentiment should
improve forecasts. The contribution of attitudes is eval-
uated by first generating forecasts with the two VARs
without the attitudes variable. The VARs with con-
sumer sentiment are then reestimated and reforecasted,
using actual future values of sentiment in the forecast.
This technique adds information to the forecasts that is
not embedded in the estimated VAR and ensures that
the errors in forecasting industrial production and un-
employment do not arise from the model's poor fore-
casts of sentiment. The forecasts are calculated from
four different stages of the business cycle: a peak, a
trough, a mid-expansion, and shortly before a peak.

Chart 3 reports the results of forecasting industrial
production with and without sentiment. The forecasts
in the first column are produced by a VAR without fi-
nancial variables. The solid lines are the actual values
of production, and the unevenly dashed lines are the
forecasts from a model with production and unem-
ployment but without consumer sentiment. The even-
ly dashed line in each graph is the forecast from a
VAR with sentiment using the actual path of consumer
sentiment over the forecast period to produce the
forecast. Chart 4 presents analogous results for unem-
ployment.

In some periods knowledge of consumer sentiment
improves the accuracy of the forecast. For example,
Chart 4 shows that adding sentiment allows the model to
track unemployment closely following the business cy-
cle trough in 1982:4. In the recent period from 1990:1 to
1992:1, however, the VAR with sentiment consistently
overpredicts production and underpredicts unem-
ployment, completely missing the timing of their turning
points; indeed, the forecasts from the model with only
production and unemployment tend to dominate those
from the VAR that includes consumer sentiment.

The second columns of the charts report results
from VARs that include financial variables. Once fi-
nancial variables are added to the model, the forecasts
from the system without consumer sentiment are fre-
quently better. As shown in Chart 3, the forecasts of
industrial production are better without sentiment in
three of the four forecast periods. The results for fore-
casting unemployment are more mixed, but in none of
the forecast periods does knowledge of sentiment lead
to superior forecasts at all horizons. Of course, the fi-
nancial variables also do not help to predict the busi-
ness cycle turning point in late 1990.

“Big” Changes in Consumer Attitudes. The historical
correlations reported by the impulse responses in
Chart 2 use the average correlations estimated over the
Chart 3
Forecasts of Industrial Production

The solid lines show actual data, the unevenly dashed lines show forecasts without sentiment, and the evenly dashed lines show forecasts with sentiment.

*Industrial production is measured in logarithms. Each set of four graphs is from VARs estimated using data through 1981:2 (business cycle peak), 1982:4 (business cycle trough), 1985:4 (mid-expansion), and 1990:1 (near a business cycle peak). Forecasts without sentiment are produced by VARs with industrial production and unemployment (and possibly real stock prices and the three-month Treasury-bill rate). Forecasts with sentiment add sentiment to the VAR and are conditioned on actual future values of consumer sentiment.
Forecasts of Unemployment

Forecasts without financial variables

Forecasts with financial variables

The solid lines show actual data, the unevenly dashed lines show forecasts without sentiment, and the evenly dashed lines show forecasts with sentiment.

Unemployment is measured in percentage points. Each set of four graphs is from VARs estimated using data through 1981:2 (business cycle peak), 1982:4 (business cycle trough), 1985:4 (mid-expansion), and 1990:1 (near a business cycle peak). Forecasts without sentiment are produced by VARs with industrial production and unemployment (and possibly real stock prices and the three-month Treasury-bill rate). Forecasts with sentiment add sentiment to the VAR and are conditioned on actual future values of consumer sentiment.
full sample to calculate the new time paths of industrial production and unemployment following a “typical” (or average) unanticipated increase in consumer sentiment. The VAR is constructed so that the unanticipated changes in variables average out to zero over time. A typical surprise in sentiment is modeled as a one-quarter increase with no subsequent surprise in sentiment. If, as Katona (1975) emphasizes, surveys of consumer attitudes are most useful when changes in attitudes are large and persistent, then the responses to such a typical unanticipated change in sentiment may be misleading. Much of Katona’s analysis of attitudinal surveys takes the form of event studies of particular historical episodes during which attitudes changed dramatically.

Specific historical episodes can be analyzed by decomposing actual production and unemployment at some date, $T + k$, into their $k$-period-ahead predicted values, based on information at time $T$, and the contribution of actual unanticipated (and unforecasted) changes in the variables over the $k$-period prediction horizon. Because the VAR coefficients are estimated over the full sample period, they are estimates of the average correlations between the variables. The innovations during the historical period are the actual prediction errors during the prediction period; these errors may be large and may arise in sequences of large positive numbers or large negative numbers.

Chart 5, which plots the actual changes and the innovations in attitudes implied by the five-variable VAR with financial variables, is used to isolate three periods of big changes in consumer sentiment. The chart also plots a band marking two standard deviations of the VAR innovations. Big changes in attitudes occurred in the early to mid-1970s, the early 1980s, and in 1990 and 1991. Except for 1990, most big positive changes in attitudes were forecasted by the VAR, and the big negative changes in the 1970s and in 1990 were unanticipated.

Charts 6 and 7 plot the actual values of the variables (solid line) along with their predictions using...
Chart 6
Predictions of Industrial Production during “Big” Changes in Sentiment

The solid lines show actual industrial production. The evenly dashed lines show predicted industrial production plus the contribution of actual sentiment innovations during the prediction period to the predicted value of industrial production. The unevenly dashed lines show predicted industrial production, setting all future innovations to zero.

---

* The solid lines show actual industrial production. The evenly dashed lines show predicted industrial production plus the contribution of actual sentiment innovations during the prediction period to the predicted value of industrial production. The unevenly dashed lines show predicted industrial production, setting all future innovations to zero.
Chart 7
Predictions of Unemployment during “Big” Changes in Sentiment

Without financial variables


With financial variables


The solid lines show actual unemployment. The evenly dashed lines show predicted unemployment plus the contribution of actual sentiment innovations during the prediction period to the predicted value of unemployment. The unevenly dashed lines show predicted unemployment, setting all future innovations to zero.
only past information (unevenly dashed line) and their predictions plus the contribution of future unanticipated changes in consumer sentiment (evenly dashed line). The plots for industrial production are presented in Chart 6; the plots for unemployment, in Chart 7. In both cases, the contribution of consumer sentiment is evaluated in models with and without financial variables. Whenever the predictions that include actual sentiment innovations lie between a variable’s actual value and its predicted value using only information up to date $T$, innovations in sentiment improve the model’s $k$-period-ahead prediction of the variable.

In general, sentiment innovations are more likely to improve predictions in the VAR without financial variables. For example, looking down the first columns of the charts, sentiment innovations somewhat improve accuracy of predictions of industrial production in 1972, 1975-76, 1981, and in 1990-92. Unemployment predictions are improved in 1972, 1975-76, 1981, and throughout the 1990-92 period. In fact, during 1990-92, including the actual prediction errors in sentiment allows the model to pick out the turning points of output and unemployment. Once financial variables are included in the VAR, however, the contribution of sentiment innovations diminishes substantially, as shown in the second columns of the charts. Even in the recent period, sentiment innovations do not improve the predictions of industrial production at all horizons. Actual sentiment innovations continue to improve the predictions of unemployment in 1990-92, although the improvement is smaller in the system with financial variables.

**The King Is Dead**

The recession of 1990-91 is only the most recent example of a long tradition of policymakers’ turning to consumer attitude surveys as indicators of economic conditions. In 1971 then–Federal Reserve Board Chairman Arthur F. Burns cited weak consumer confidence as one of the “grounds for concern . . . with regard to some features of the recovery” (Burns 1978, 120).

Although there is ample informal theoretical precedent for taking consumer attitudes seriously, the empirical grounds for viewing attitudes as having an important independent influence appear to be somewhat barren. Used alone, unanticipated changes in consumer sentiment are correlated with movements in industrial production and unemployment. But when a small set of readily available financial information is taken into account, innovations in sentiment appear to be largely redundant, having very weak correlations with industrial production and unemployment. Even when these financial variables are ignored, knowledge of past, current, and future values of sentiment does not systematically improve the forecasts of industrial production and unemployment.

Finally, the empirical work suggests a reason that economic observers latched onto consumer sentiment as an explanation of the most recent economic downturn. Surprises in attitudes do appear to be associated with unpredictable fluctuations in unemployment and, to a lesser extent, production during the 1990:1-to-1992:1 period. This association, however, is not systematic because it fails to hold during other episodes of large and persistent changes in attitudes.

When economic events are confounding it is always tempting to find some event-specific fact that appears to dispel our ignorance. To be sure, a thorough enough search will always uncover some variable that suits the needs of the day. In the fall of 1990 the hoopla surrounding consumer attitudes swept aside careful analysis and pointed toward the obvious coincidence of the plunge in sentiment with the decline in activity as prima facie evidence that depressed consumers caused the recession. The danger in this king-for-a-day economic analysis is that it can easily sidetrack policymakers from focusing on the true underlying causes of economic fluctuations.

**Notes**

1. See, for example, statements by Michael J. Boskin, chairman of the Council of Economic Advisers, reported by Nasar (1991b, 1991c), and the testimony of Alan Greenspan (1991), chairman of the Federal Reserve Board.

2. Substantial effort has been devoted to estimating consumption functions that include consumer sentiment, with mixed results. A chronological list of some studies is Tobin (1959), Mueller (1963), Friend and Adams (1964), Adams

3. Whether sentiment is treated as an exogenous variable or as a catalyst depends on how the simulation is performed. When an “optimistic” path of sentiment is assumed, the equation that determines sentiment is thrown out and replaced with the assumed optimistic path. When the focus of the simulation is not on attitudes per se, then the time path of sentiment evolves endogenously according to some estimated relationship.

4. The minutes of the Federal Open Market Committee meeting on December 18, 1990, seem to reflect the view of consumer attitudes as a catalyst: “Even under the assumption that the Persian Gulf situation would be more settled and oil prices lower, restoration of the degree of confidence needed to induce substantial upturn in spending was not assured” (Board of Governors 1991a, 7-8).

5. Katona and Mueller deny this summary statistic view: “The economic attitudes of consumers originate in a variety of political and economic developments. It is inadequate to assume that consumer attitudes are merely a reflection of recent trends in, say, incomes or prices and that these can serve as substitutes for data on attitudes . . .” (1953, 2).

6. The Organisation for Economic Cooperation and Development publishes surveys of businesses in Main Economic Indicators, and the European Community publishes surveys of businesses and consumers in European Economy, Supplement B.

7. Of course, a statistical finding that consumer attitudes strongly predict economic conditions after controlling for other available information would cry out for a coherent economic explanation.

8. If, based on historical data, this month’s change in consumer attitudes were fully anticipated, then decisionmakers who are basing their decisions on the historical data would have already incorporated this data, and, therefore, this month’s consumer attitudes, into their decision processes. The announced change in attitudes contains no news that will change decisionmakers’ perceptions of current or expected future economic conditions, so there is no reason for decisionmakers to alter their behavior.

9. All quarterly data are averages of the monthly series. Survey data from 1954 to 1965 are not available for every quarter, so missing quarters are assigned the previous quarter’s value. For 1978 on, quarterly observations of the survey are averages of their monthly values. The monthly real stock price index is the nominal S&P 500 index deflated by the previous month’s consumer price index for all urban wage earners. Industrial production, unemployment, and consumer prices are seasonally adjusted at the source.

10. Letting $s$ be sentiment, $y$ be industrial production, and $u$ be unemployment, the three-variable VAR estimated with one lag would be written as

\[
s_t = a_{11}s_{t-1} + a_{12}y_{t-1} + a_{13}u_{t-1} + \epsilon_{s,t}
\]

\[
y_t = a_{21}s_{t-1} + a_{22}y_{t-1} + a_{23}u_{t-1} + \epsilon_{y,t}
\]

\[
u_t = a_{31}s_{t-1} + a_{32}y_{t-1} + a_{33}u_{t-1} + \epsilon_{u,t}
\]

The $t$ and $t-1$ subscripts represent the dates at which the variables are measured. The error terms in the VAR—the $\epsilon$’s—are called “innovations.” Innovations represent changes in the left-hand-side variables that are unanticipated based on past information, so they are one-step-ahead forecast errors.

11. Likelihood ratio tests unambiguously called for estimating the three-variable VAR with eight lags of each variable. The VAR is estimated over the full sample with a constant term using logarithms of the attitudinal survey data and industrial production and the level of the unemployment rate.

12. The impulse response analysis is a better way to evaluate the predictive value of sentiment because the analysis takes account of how all three equations interact, whereas the $F$-tests reported in Table 1 hold only equation-by-equation. It is quite possible, for example, for the $F$-tests to indicate that the coefficients on lagged sentiment in the industrial production equation are zero, yet innovations in sentiment can be correlated with future production indirectly through their influence on unemployment. It is also possible, though less likely, for the $F$-tests to indicate that sentiment predicts production, even though sentiment innovations are not followed by changes in industrial production. This result can occur if the indirect influence of sentiment on production through unemployment offsets the direct influence of sentiment on industrial production. The likelihood that $F$-tests will give misleading results increases as the number of variables in the VAR increases.

13. The probability bands are generated by the Bayesian Monte Carlo procedure described in Doan (1991). The procedure takes draws from the posterior distribution of the VAR coefficients. After taking 1,000 random draws, the impulse response functions are ordered and the 95th percentile bands are extracted. When a band lies entirely above (below) the zero axis of the chart there is a 95 percent likelihood that the response is positive (negative).

14. Survey results are released one week after the end of the survey month.

15. The five-variable VAR was estimated with six lags, as suggested by likelihood ratio tests. The VAR is estimated with a constant term using logarithms of consumer attitudes, industrial production, and real stock prices and the levels of the interest rate and the unemployment rate.


17. Robert E. Hall, professor of economics at Stanford University, was quoted in early April as saying, “None of the usual financial predecessors of recession have appeared to this day” (Nasar 1991a).

18. Keynes emphasizes this point about periods of extreme uncertainty: “In abnormal times . . . the market will be subject to waves of optimistic and pessimistic sentiment, which are unreasoning and yet in a sense legitimate when no solid basis exists for a reasonable calculation” (1964, 154).

19. This approach does not allow the correlations between sentiment and industrial production and unemployment to be different between periods of large changes in consumer attitudes and periods of small changes in attitudes.
20. The three periods are 1972:1 to 1976:4, 1980:1 to 1984:4, and 1990:1 to 1992:1. Each of these—the oil embargo, the change in Federal Reserve operating procedures, and the Persian Gulf War, respectively—coincides with Keynes's notion of extreme uncertainty.

21. As with the impulse responses, the results with financial variables report the marginal contribution of sentiment innovations to predictions of industrial production and unemployment, given current stock prices and interest rates and past values of all the variables.

22. As shown in Charts 3 and 4, however, this result does not carry over to the forecasts.

References


Regional economic models are used in a variety of decision-making contexts. Government officials use them to prepare annual budgets. Businesses rely on them for producing short-run market demand forecasts and for analyzing longer-term growth strategies. Urban planners and transportation officials use them to develop long-range plans for urban and regional development. Finally, state and local policymakers turn to them to get new ideas for programs and policies to promote long-run regional growth.

Although it would be convenient if a single model had been developed to serve all these purposes simultaneously, no such model is ever likely to exist. Instead, regional models tend to be highly specialized in terms of the issues that they are able to address and the time horizons over which their analytical results are most reliable. For example, a short-run forecasting model might serve the needs of state or local government officials engaged in the annual budgeting process, but it would contribute little information relevant to long-run local economic development issues confronting planners and policymakers. Only rarely is a regional model able to perform well in more than one of these distinct decision-making contexts.¹

The rapid pace of urban growth during this century, along with the challenge it has presented for planners trying to anticipate and influence this growth, has ensured a healthy demand for regional economic models, particularly since 1945. Unfortunately, models supplied have been inadequate.

Andrew C. Krikelas

¹The author is an economist in the regional section of the Atlanta Fed’s research department.
At the beginning of the postwar period, the economic base model was probably the only such instrument generally available for regional economic analysis. This model focuses on regional export activity as the primary determinant of local-area growth; it is one of the oldest and most durable theories of regional growth, with origins extending at least as far back as the early 1900s. However, economic base theory received the greatest amount of attention from scholars in regional science between 1950 and 1985. Despite the model's acceptance over such a long period, when the noted regional scientist Harry W. Richardson, writing for a special twenty-fifth anniversary issue of the Journal of Regional Science, reflected upon the more than forty years of research conducted within this paradigm, he concluded that the findings on economic base models are conclusive. The spate of recent research has done nothing to increase confidence in them... The literature would need to be much more convincing than it has been hitherto for a disinterested observer to resist the conclusion that economic base models should be buried, and without prospects for resurrection (1985, 646).

Like Richardson, others over the years have expressed concern with the narrow focus of economic base theory on exports—just one portion of the demand side of the regional growth equation—to the exclusion of important supply-side factors and constraints. Many have suggested that economic base theory, its analytical and methodological techniques, and the public policies that it promotes should be abandoned in favor of other, more comprehensive theories of regional growth and development.

Nevertheless, economic base research continues. Most notably, James P. Lesage and J. David Reed (1989) and Lesage (1990) have provided empirical evidence in support of the economic base hypothesis as both a short-run and long-run theory of regional growth. These authors suggest that their models could be used both for short-term forecasting of regional employment, income, and product and for long-range regional economic planning and policy analysis. If these claims were valid, then the economic base model, rather than being of little value, would be one of the few regional models that might be useful in each of these very different but crucially important decision-making contexts.

Because regional economic models play such an important role in planning and policy discussions, it is important to have a clear understanding of their strengths and weaknesses. Limitations of the economic base model in particular, because it tends to be widely used, should be recognized. Recent research has provided evidence suggesting substantial improvement in traditionally static economic base model specifications through the adoption of techniques routinely employed in the macroeconomics time-series literature. However, this author's research suggests that these studies may have overstated the usefulness of these new economic base model specifications (Andrew C. Krikelas 1991).

The purpose of this article, therefore, is twofold. First, a concise analytical history of the old and extensive economic base literature generated by a variety of professional and academic disciplines is provided in order to place recent research in perspective. The discussion then turns to the central question addressed in Krikelas (1991): Can techniques borrowed from statistical time-series literature successfully breathe new life into the traditional economic base model?

### Definition of the Economic Base Concept

As originally formulated, the economic base model focused on regional export activity as the primary source of local-area growth. According to this theory total economic activity, $E_I$, is assumed to be dichotomous, with a distinction being made between basic economic activity, $E_B$ (activities devoted to the production of goods and services ultimately sold to consumers outside the region), and nonbasic economic activity, $E_{NB}$, which includes activities involved in producing goods and services consumed locally:

$$E_I = E_B + E_{NB}.$$ (1)

This division of regional economic activity into these two distinct sectors is the central concept of the model. A serious empirical concern is raised by this approach; however, because appropriate export data are available at any subnational level only at high cost and with long lags. Various alternative measures have been proposed and analyzed in the literature over the years, but none has been found entirely adequate. Data problems, therefore, have always complicated economic base research.

While the central concept of the economic base model is the duality of regional economic activity, its fundamental behavioral assumption is that nonbasic economic activity depends on basic economic activity. In this perspective, external demand for a region's
exportable goods and services injects income into the regional economy, in turn augmenting local demand for nonexportable goods and services. The model assumes that the income injected into the regional economy and the accompanying potential for developing locally oriented, nonbasic industries are in proportion to the size of a region’s export base. Static and demand-oriented, the model ignores factors that affect the supply of a region’s output and other changes, such as the introduction of new products, that affect demands.

\[ E_{NB} = f(E_B) = \alpha + \beta \cdot E_B. \quad (2) \]

Equations (1) and (2) can then be combined into the reduced-form expression in equation (3), which indicates that total economic activity is primarily a function of basic activity:

\[ E = \alpha + (1 + \beta) \cdot E_B. \quad (3) \]

The expression \((1 + \beta)\) is commonly referred to as the economic base multiplier, and the parameter, \(\beta\), is called the economic base ratio.

When applied to analyzing regional growth, the economic base model suggests that the growth process will be led by industries that export goods and services beyond regional boundaries. It even offers a prediction, captured in the multiplier, of the total regional impact likely to result from a change in basic economic activity generated outside the region. Understanding the future path of a regional economy, the model implies, requires simply concentrating on the prospects for the base industries. These few important industries are often dubbed “engines of regional growth.”

This simple model captures the essence of economic base theory. Although the model has been enhanced over the years to include additional variables as well as to capture more explicitly the dynamic nature of the regional growth process, most changes have been made within the scope of this simple demand-oriented specification. In general, economic base models have not evolved to acknowledge the potential impact of many important variables that may affect regional growth—interregional capital flows; labor migration patterns; changes in products, tastes, and production processes; demographic shifts; and changes in state and local tax laws, to name a few. Because these issues are generally too important to ignore, many regional scientists have concluded that economic base theory lacks the complexity to provide a useful framework for analyzing many regional economic issues and policies. The following review of the development and testing of the model will summarize where the debate on this topic stands at this point.

**History of the Economic Base Literature**

Five fairly distinct chronological periods characterize the history of the economic base literature: (1) the origin of the concept, 1916-21; (2) early development, 1921-50; (3) the first round of serious debate, 1950-60; (4) the second round of debate, 1960-85; and (5) a third and perhaps final round of debate begun in 1985 and continuing today. Decades of research within the economic base paradigm have created a body of conventional wisdom concerning the uses and limitations of the model, both in theory and in practice. Nonetheless, as yet another round of discussion has begun, it seems that few lessons of the past have been learned and that a brief summary of the history of this literature might be useful.

**Origin of the Economic Base Concept.** The essential duality of regional economic activity that is central to the simple model expressed in the equations above was first articulated in 1916 by the German sociologist Werner Sombart, who wrote of “actual city founders,” identified as the “active, originative, or primary city formers”—those whose positions of authority, wealth, or occupation allowed them to draw income from outside the city—and the “passive or derived or secondary city formers,” whose livelihood depended on the city formers (Günther Krumme 1968, 114).

In 1921 M. Arrousseau made a similar observation in commenting on the relationship between what he distinguished as a town’s primary and secondary occupations: “The primary occupations are those directly concerned with the functions of the town. The secondary occupations are those concerned with the maintenance of the well-being of the people engaged in those of primary nature” (John W. Alexander 1954, 246). Also in 1921, landscape architect Frederick Law Olmsted distinguished between what he called primary and ancillary economic activity in an urban area (Alexander 1954, 246).

Thus, although Sombart was apparently the first to observe formally the seeming duality of urban and regional economic activity, the remarks of his contemporaries Arrousseau and Olmsted make it abundantly clear that the concept was ripe for expression. By the early 1920s, therefore, the economic base concept had generally surfaced as a potential theory for explaining the regional growth process.
Early Development of the Theory. Following establishment of the theory, the next logical step should have been the empirical testing of the validity of the model's central hypothesis. However, this step was almost universally ignored and the model adopted as useful as the rapid growth of cities early in the century pressured state and local officials to improve the way in which they developed plans for urban expansion and the provision of public infrastructure and government services. The economic base model provided a much-desired framework for developing such plans, and studies designed to identify and measure basic industries—economic base studies—quickly became primary tools employed in acquiring information for long-range planning.

After identifying a region's export base, economic base studies calculate a local-area economic base ratio, $\beta$. Once calculated, the economic base ratio can be used with forecasts of the future growth of the region's export base industries to predict the region's overall growth. The study's focus on the smaller number of industries identified as regional export industries helps streamline the process of forecasting total regional economic activity. In addition, by identifying those industries considered most important to the regional growth process, an economic base study provides information that adds insight to discussions of regional industrial policies and programs.

Sombart's analysis of the Berlin economy, published in 1927, was the first economic base study conducted during this period. Sombart, complaining that "nobody makes the effort to sit down with a pencil and figure out with the help of occupational statistics how much there actually is of a city-forming industry in a city such as Berlin," developed an empirical approach for dividing an urban economy into its dual parts (Krumme 1968, 116).

Lacking detailed information on regional export activity, Sombart relied upon industry employment data collected in Berlin in 1907 to estimate the basic and nonbasic sectors of the city's economy. Relying mainly upon his personal judgment, Sombart estimated that approximately 262,000 of Berlin's total work force of 543,000 were employed in export base industries (Krumme 1968, 113). These calculations placed Berlin's nonbasic/basic ratio, $\beta$, at 1.07, an approximately one-to-one relationship.

Although Sombart did not use this information to forecast Berlin's growth, he could have done so. Making a more limited forecast of the prospects for those industries he had identified as being part of the city's export base and multiplying that total by the city's economic base multiplier $(1 + \beta)$ of 2.07 (assuming that the city's base ratio had remained relatively stable in the intervening twenty years since the census was conducted) would have provided a forecast of the change in total economic activity expected in Berlin as a result of some externally generated change in demand for its export product.

The reliance on secondary data sources for Sombart's study of Berlin's economic base is typical of most such research. As pointed out earlier, even today the appropriate regional export data required to conduct an adequate economic base study are available only at relatively high cost. The comprehensive economic analysis of the city of Oskaloosa, Iowa, published in Fortune magazine in 1938 illustrates this point ("Oskaloosa..." 1938).

Although published in a popular magazine, this study represents an important contribution to research on the economic base theory. The magazine staff conducted a complete census of the town's 3,000 families in order to determine the origin and destination of income flows within the city. They also conducted a census of the town's businesses, including an accounting of the destination of their output and the source and value of the most important inputs into the local production process.

The results of the study indicated that in 1937 Oskaloosa was a net exporter of goods and services to the rest of the world and that manufactured goods and professional services were the town's leading export industries. The study's findings are interesting because they were based upon a census that provides a relatively accurate portrayal of Oskaloosa's export activity during the year studied. Even by present standards this study represents one of the most thorough economic analyses of a small community ever published.

The great effort required to collect these data, however, explains why a survey- or census-oriented approach to economic base identification generally has been abandoned for the nonsurvey identification techniques made popular by Homer Hoyt in the late 1930s. Working with the Federal Housing Administration during the mid-1930s, Hoyt developed and employed an economic base methodology for producing forecasts of local housing market demand. His techniques became known to a wide audience with the original publication of his textbook, Principles of Urban Real Estate (coauthored with Arthur M. Weimer in 1939), which Richard B. Andrews called the first "complete statement of the theory of the economic base." In commenting on the impact of this work, Andrews continued, "This statement included much material that was new
outside of technical reports. For example, it introduced in formal fashion the idea of a mathematical relation between basic employment and service employment. . . . Hoyt considered the economic base idea to be a tool that might be employed in analyzing the economic background of cities with the objective of forecasting the future of the entire city” (1953a, 163).

In this text Weimer and Hoyt distinguished between “urban growth” and “urban service” industries, suggesting that a region’s potential for growth depended primarily upon the prospects for the region’s urban growth industries. They provided a six-step procedure for identifying such industries. Using relatively accessible income and employment data, the authors developed a methodology that represented a combination of what has become known as the assignment technique and the location-quotient technique of economic base identification. The assignment technique is essentially identical to Sombart’s methodology, in which personal judgment is used to assign industries within a particular regional economy to basic and nonbasic sectors. The location-quotient technique, on the other hand, relies upon regional economic data to make such distinctions.

Location-quotient methodology compares a region’s concentration of economic activity in a particular industry with that of a benchmark economy, usually the entire country in which the region is located. If the regional concentration, measured in terms of the industry’s share of total regional employment or income, exceeds the benchmark economy’s concentration in that industry, the surplus level of employment or income is assumed to measure regional export activity. For example, if an industry accounts for 6 percent of regional employment but only 2 percent of national employment, two-thirds of that industry’s employment would be called basic. (If the regional activity in an industry is less than that at the national level, the industry is categorized as nonbasic.) Making this identification requires only industry employment or income data for the region and a similar set of data for an appropriate benchmark economy.

Although Weimer and Hoyt were not the first to propose using the location quotient and assignment techniques as nonsurvey methodologies for dividing regional economic activity into its basic and nonbasic components, dissemination of the techniques through their textbook introduced these shortcuts to a wide audience. With these methodologies available it became feasible for local development officials to adopt the economic base paradigm for purposes of analyzing specific urban and regional economies. During the latter half of the 1940s, once these techniques had become more widely known, a much larger number of cities and states began to use the economic base model in urban and regional planning and economic analysis.7

Theoretical Debate. By 1950 economic base theory and its methodological techniques had become established as the primary tools of regional planning. The theory itself had been accepted, uncritically, as an explanation of local-area growth and economic development. Between 1950 and 1960, however, discussion at the theoretical and methodological level turned directly to the question of the validity of the economic base hypothesis. Unfortunately, only a handful of empirical tests were reported during this entire decade.

The earliest and most cogent critique of economic base theory was presented by George Hildebrand and Arthur Mace (1950) in their analysis of the Los Angeles metropolitan area. This important contribution identified the theoretical model upon which the economic base paradigm was founded and performed an empirical test that provided evidence supporting the validity of the economic base hypothesis, at least for short-run forecasting.

Hildebrand and Mace’s most significant contribution was their explicit formulation of economic base theory as a testable behavioral hypothesis. Their results, which demonstrated a statistically significant short-run relationship between basic and nonbasic employment in Los Angeles, represented the first empirical confirmation of the economic base hypothesis. Furthermore, the authors formulated their tests within the context of an explicitly Keynesian national income model and then outlined the inherent limitations of such a model.

Consider the familiar Keynesian relationship:

\[ Y = C + I + G + (X - M), \]  

where total regional income, \( Y \), is divided into a number of distinct sectors, including consumption, \( C \); investment, \( I \); government expenditures, \( G \); and exports minus imports, \( X - M \). The reduced-form expression of this model would include some smaller set of exogenous variables, only one of which would be regional exports. (Other exogenous variables would include the autonomous components of consumption, investment, government expenditures, and imports; marginal propensities to consume locally, to invest locally, and to import; and local and federal tax policies.) It is this set of exogenous factors that would determine, theoretically, a region’s total income level, \( Y \).
The economic base model focuses on one particular aspect of this relationship, regional export activity, \( X \) (\( E_p \) in equation [1] above), and can be considered a special case of the more general Keynesian model in equation (4). Given this interpretation, it becomes clear that for exports to be considered the only exogenous determinant of regional growth, all other relevant factors, related to both demand and supply, must remain fairly constant or be functions of export activity. Although this might be a tenable assumption in the short run, it probably is an extremely poor one in the long run. Hildebrand and Mace made this observation explicit and suggested that the model was most appropriate for anticipating regional economic trends over a short time horizon. In addition, they listed some of the other variables that they thought should be taken into account in developing a more comprehensive model of regional economic activity: population levels and interregional migration patterns, regional capital investment levels and annual flows, state and local tax policies, and changes in the cost of transportation to reach external markets. Despite these reservations, Hildebrand and Mace offered a fairly encouraging assessment of the prospects for this type of research, based on the availability of additional census data and further empirical analysis across a ten-year span.

Unfortunately, the lessons contained in Hildebrand and Mace’s study were not widely disseminated. Hildebrand and Mace were among the first economists to contribute to the economic base literature. Their article was published in a journal not normally read by geographers and urban planners, who, before 1950, had played a dominant role in the research conducted within the economic base paradigm. Therefore, rather than playing the role of a seminal article to a further body of empirical research, the Hildebrand and Mace article remained relatively unknown. The debate of the 1950s brought many of their important insights to the attention of geographers and urban planners, but it took nearly a decade for all of these contributions to be uncovered.

Most of the 1950s’ debate on economic base theory was conducted in the geography and planning literatures. The origin of this debate can be traced to a series of nine articles published by Andrews between 1953 and 1956 (see reference list). These articles provided a careful exposition of economic base theory and the methodologies that had been developed to analyze urban and regional economic activity. The author’s stated purpose was to explore and evaluate the entire concept. “We have operated far too long on a set of ideas which appear valid but which, despite substantial conceptual omissions and difficulties of application, seem to be accepted all too blithely,” he wrote, calling for “more fundamental thinking on and questioning of the reality and utility of base theory as presently conceived” (1953a, 167).

While Andrews was somewhat critical in his assessment of the economic base paradigm, he clearly was a proponent of its inherent validity and usefulness. Instead of suggesting the abandonment of the model as a tool for urban and regional economic analysis, he identified ways in which it could be improved to serve such purposes better. His recommendation included better efforts at basic industry identification and measurement, improvements in the collection of regional data, and modifications in the way in which economic base concepts were used.

When applied to analyzing regional growth, the economic base model suggests that the growth process will be led by industries that export goods and services beyond regional boundaries.

Given Andrews’s criticism of the state of the economic base research prior to 1950, it is surprising to note he did not address one of the most fundamental shortcomings of this research: the lack of empirical verification of the underlying hypothesis. Krikelas (1991) identified only five empirical tests of the economic base hypothesis conducted during the 1950s. Three of those studies, including that of Hildebrand and Mace, supported the validity of the economic base hypothesis, at least in the short run, and two provided evidence against it. A decade of research, therefore, provided insufficient empirical evidence for determining the validity of the model’s central hypothesis.

Instead, most of the debate of the 1950s centered on questions related to theory and practice rather than testing. Hans Blumenfeld (1955) was critical of the economic base model’s narrow focus on export activity as the primary source of regional growth. While he agreed that this model might do well to explain economic growth in small or highly specialized
economies, he argued that it was inadequate to explain the growth of complex urban economies. Blumenfeld was also critical of the policy implications of the model; these focused almost exclusively on supporting existing export industries at the expense of other reasonable alternatives, such as fostering the establishment and development of industries that would compete with imported goods and services.

Charles M. Tiebout (1956a, 1956b) and Douglass C. North (1955, 1956) engaged in a short but lively debate over the short-run versus long-run applicability of the economic base model. Tiebout, explicitly recognizing the Keynesian roots of the economic base model, supported Hildebrand and Mace’s (1950) contention that the economic base model was most appropriate for short-run economic analysis. He also argued that the economic base model minimized the important contribution that nonbasic economic activity made to local area growth and development. He wrote that, although export activity was important, “in terms of causation, the nature of the residential industries will be a key factor in any possible development. Without the ability to develop residential activities, the cost of development of export activities will be prohibitive” (1956a, 164).

North, however, objected to the characterization of the economic base model as an adaptation of the demand-oriented Keynesian model. Instead, he argued that the most important determinant of a region’s long-run growth potential was its ability to attract capital and labor into the region from outside. Such supply-enhancing flows in turn would respond quite favorably to profit opportunities offered by regions engaged in high levels of export activity. North observed that historically “it was frequently the opportunities in manufacturing for the United States market which led to immigration of labor and capital into a region. The important point is that the pull of economic opportunity as a result of a comparative advantage in producing goods and services in demand in existing markets was the principal factor in the differential rates of growth of regions” (1956, 166).

The economic base model proposed by North explicitly recognized the important role of supply factors in determining the nature and growth potential of a region’s export base. In practice, however, most economic base models of this and subsequent periods have maintained a fairly strict demand orientation. This demand-oriented model is also the one to which Tiebout raised so many objections. As a result, although Tiebout and North found themselves on different sides concerning the validity of the model as a long-run theory of regional growth, both ultimately agreed that supply factors needed to be added to the model in order to make it relevant for long-run regional economic analysis.

One additional advance in the theoretical literature of this period that called into question the adequacy of economic base modeling techniques was the development of regional input-output models. Before 1950 the economic base model represented the primary tool available to regional planners for analyzing the impacts of anticipated changes in regional economic activity. During the first half of the 1950s, however, input-output modeling techniques first developed by Wassily W. Leontief (1951) were adapted for purposes of regional economic analysis. While a regional input-output model could distinguish between the differential regional impacts that might be associated with, for example, the construction of a specialty steel manufacturer versus a mail-order catalog facility—two very different kinds of basic economic activity—the simple two-sector economic base model could not make such a distinction. Given this limitation, many urban planners began to advocate input-output techniques as more appropriate for forecasting anticipated changes in regional economic activity.

The debate of the 1950s also focused on several important methodological issues. Papers by John M. Mattila and Wilbur R. Thompson (1955) and Charles L. Leven (1956) considered the adequacy of the location-quotient technique’s ability to identify a region’s economic base industries. While suggesting certain improvements to the traditional formulation of the location quotient, Mattila and Thompson concluded that “if used with care, the index of surplus workers in both its absolute and relative form should prove to be...
a highly useful tool in regional economic base studies” (1955, 227).\(^{10}\) Leven, on the other hand, arrived at the opposite conclusion, stating that “the shortcomings of this technique render it useless as a quantitative measure of basic activity in an area” (1956, 256).

The issue of the appropriate measure to be used for calculating location quotients was also discussed. Because employment data were more readily available than wage or income data, most economic base studies of this period used employment in identifying regional export activity. This measure, however, has some serious drawbacks. In addition to placing equal weight upon part-time and full-time employment and failing to adjust adequately for productivity and wage differences between workers employed in different industries, employment data do not provide any measure of the impact that transfer payments and other sources of unearned income, such as interest payments, rents, and profits, have upon a regional economy.

Recognizing the serious weaknesses associated with the use of employment data for purposes of identifying a region’s economic base, Andrews (1954a), Leven (1956) and Tiebout (1956c) all suggested the adoption of alternative measures of regional economic activity. Andrews and Tiebout advocated the use of income received by residents of the region, and Leven argued for a value-added measure. Income and value-added data, however, generally are not available for regional economies, especially at the substate level, except with long lags.

By the beginning of the 1960s professionals engaged in urban and regional economic analysis had divided into three distinct camps concerning the conduct of research within the economic base paradigm: those who still considered the economic base model to be a reasonable framework for urban and regional economic analysis; those who questioned its validity but sought more empirical evidence before abandoning the paradigm; and those who rejected the validity of the hypothesis, instead turning to the investigation of other methods of regional economic analysis, including regional input-output models. Whereas the debate of the 1950s was conducted primarily at the theoretical level, the quarter-century between 1960 and 1985 was filled with empirical examinations of a wide range of theoretical and methodological questions related to the economic base model.

**Empirical Debate.** Between 1960 and 1985 a large number of articles and several books were published on the economic base model.\(^{11}\) Yet while the question of the empirical relevance of the economic base hypothesis was arguably the most important issue facing the profession on the heels of the debate of the 1950s, only a quarter of these contributions actually addressed it.

To provide some perspective on the extensive literature of this period, Krikelas (1991) developed a taxonomy. The six categories listed represent distinct facets of the economic base literature of this period: (1) identification of export base activity, (2) calibration studies, (3) extensions of the base model, (4) case studies, (5) theoretical works, and (6) tests of the economic base hypothesis.

A thorough discussion of the contributions that fall into each of these categories is beyond the scope of this article. However, a summary of the major developments in each category should yield insights. It should be noted that the majority of the research published during this period—that is, categories (1)-(4)—assumed, at least implicitly, the validity of the economic base hypothesis.

**Identification of Export Base Activity.** The most contentious issue facing researchers using the economic base model is the identification of regional export activity. Much attention has been paid to the development of nonsurvey techniques, and during this period seventeen studies were devoted to creating new or improving old methodologies. Edward L. Ullman and Michael F. Dacey (1960) and Vijay K. Mathur and Harvey S. Rosen (1974) introduced two completely new nonsurvey methods for identifying regional export activity, and several other researchers suggested refinements for improving both the location-quotient and assignment methods of economic base identification. Andrew M. Isserman (1980) offers an excellent survey of the developments of this period, including a critique of each methodology.

**Calibration Studies.** Calibration studies are research designed to test the adequacy of competing nonsurvey identification techniques. Researchers either compare nonsurvey estimates of regional exports with benchmark survey or census data on regional exports or simply compare results of several nonsurvey techniques. Another seventeen studies conducted between 1960 and 1985 can be classified as calibration studies, and Isserman provides an excellent summary of such research, concluding that although efforts to develop and refine the nonsurvey methods had been substantial, “the situation is lamentable” (1980, 178-79).

**Extensions of the Base Model.** During this period at least two important extensions were made to the simple economic base model. In the first, additional variables other than basic economic activity were added to the original specification in order to investigate...
their effects on the regional growth process. Stanislaw Czamanski’s (1965) study represents the first of several in which a demographic variable—population—was explicitly included in the model specification. Paul E. Polzin (1977), on the other hand, developed a model designed to capture the effects of local-area labor supply conditions on regional economic activity, and Ron E. Shaffer (1983) and Shahin Shahidsaless, William Gillis, and Shaffer (1983) included variables designed to measure the contribution of both demographic and geographic factors. Given the fact that these authors generally found the additional variables to be very important determinants of regional growth, it is somewhat surprising that relatively few studies focused on this issue.

A third period of debate on the economic base model centers on the question of whether new techniques borrowed from macroeconomics time-series literature can revive the traditional economic base model.

A second innovation, which gained a much broader acceptance in the literature, was the disaggregation of basic activity into more than one sector—manufacturing, construction, services, and government, for example. This work was stimulated by the challenge posed by regional input-output models and their clear demonstration that changes in regional activity in different export industries were likely to have very different effects upon a regional economy. Steven J. Weiss and Edwin C. Gooding (1968) provide the first example of a multisectoral economic base model, and their work was repeated and extended in many subsequent studies. However, while the literature of this period reported the results of numerous multisectoral economic base models, the maximum number of sectors for which multipliers can be estimated has always been limited by the length of available data series, usually to ten sectors or fewer. As a result, no economic base model has ever been able to reproduce the level of industry disaggregation available in most regional input-output models.

Case Studies. In most instances the main purpose of these base studies was the calculation of multisectoral economic base multipliers intended to demonstrate the significant impact of the sectors under consideration. Early studies had focused mainly on the role of manufacturing in the regional growth process. Many of these later works were instead devoted to showing the important contribution that the trade and service sectors could also play in regional growth.12

Theoretical Works. Several contributions during this period were devoted exclusively to advancing the theoretical foundations of the economic base paradigm. Edwin F. Terry (1965) explicitly derived the linkage between the economic base model and the Keynesian model. John Mutti (1981), on the other hand, demonstrated the close relationship between economic base and international trade models. And finally, Wolfgang Mayer and Saul Pleeter (1975) and F.J.B. Stillwell and B.D. Boatwright (1971) developed economic base theoretic models that demonstrated that the location-quotient and minimum-requirements methods of export industry identification could be derived from, and were consistent with, economic base theory. While these and other contributions provided a formal statement of the theoretical underpinnings of the economic base model and its methodological techniques, they did not provide empirical evidence in support of the theory’s central hypothesis.

Tests of the Economic Base Hypothesis. In considering the empirical results of studies published during this period, it is important to distinguish between dynamic and static tests of the economic base hypothesis. Although the economic base paradigm generally has been used, implicitly, to analyze dynamic regional economic events, most specifications of the model, like that in equations (1)-(4), have been explicitly static in nature. This point was made clear first by Charles E. Ferguson (1960). Subsequently, one of the major contributions of this period was the more explicit consideration of the dynamic properties of the economic base model. Researchers began using time-series modeling and other econometric techniques to analyze the short-run versus long-run applicability of the economic base model as well as to develop practical regional forecasting models.

The majority of these studies, however, were still predicated upon explicitly static model specifications. Even some of the studies that ostensibly attempted to capture the dynamic properties of the economic base model failed to do so adequately.13 Given that the utility of an economic base study depends upon its use for analyzing dynamic economic events, it is
Unfortunately—and surprising—that relatively few of these empirical studies were specified in such a way as to explore this issue.

In reviewing the literature of this period, Krikelas (1991) examined twenty-three studies that reported the results of tests of the economic base hypothesis. Eleven were static tests; twelve, dynamic. Of these, six static tests and seven dynamic tests provided results consistent with the economic base hypothesis. Many of the dynamic tests of the hypothesis were further designed to explore the issue of the short-run versus long-run validity of the economic base hypothesis. Of these, six static tests and seven dynamic tests provided results, which were based upon an appropriately specified dynamic model, were more likely to be attributable to the authors' choice of bifurcation methodology than to the existence of a long-run economic relationship between basic and nonbasic employment. Thus, while a narrow majority of the test results reported during this twenty-five-year period provided evidence in support of the validity of the economic base hypothesis, at least in the short run, very little empirical evidence suggested that the model could also perform well in the long run.

By 1985 the most definite and positive comment the literature could support about an economic base model was that it would perform best in providing relatively short-term forecasts of total regional economic activity. More than fifty years of research had failed to provide any substantial evidence in support of the model as a long-run theory of regional growth—a serious limitation in light of the fact that policymakers are generally more interested in long-run growth issues. It should be clear that the economic base model, because it fails to account for some of the fundamental determinants of the regional growth process, should not be adopted for long-range planning and policy analysis. These are the results that led to Richardson's call (cited earlier) for burying economic base models “without prospects for resurrection” (1985, 646).

Third Period of Debate. Despite Richardson’s impassioned warning, research continues to be performed within the framework of the economic base paradigm. Recently, a resurgence in such research has been fueled by a recognition that some sophisticated econometric techniques used in analysis of macroeconomics time-series may be applied to the economic base model. In particular, it has been demonstrated that the essential features of the economic base model can be captured within the context of a bivariate vector autoregression (VAR) linking basic and nonbasic economic activity. Once specified, such a VAR can be subjected to the time-series econometric tests and analytical procedures that have been developed over the years. Granger causality tests can be formulated in order to test the validity of the economic base hypothesis. Impulse-response functions (the response of a variable to an unanticipated increase in other variables) can be derived and given a natural interpretation as dynamic base multipliers. Forecasting competitions can be held in order to assess how well competing models improve the accuracy of a given forecast. Finally, co-integration tests can be performed in order to assess whether there might be a long-run relationship between basic and nonbasic economic activity.

Using such techniques, Lesage and Reed (1989) and Lesage (1990) found empirical evidence in support of the economic base hypothesis. Lesage and Reed reported Granger causality test results that were generally consistent with the economic base hypothesis, at least in the short run. Proceeding further, the authors used their VAR model specifications to derive impulse-response functions describing the dynamic relationships between basic and nonbasic employment in eight metropolitan statistical areas (MSAs) in Ohio. The reasonable nature of the multipliers calculated from this experiment led the authors to conclude that this methodology offered promise for regional economic forecasting and policy analysis purposes. When Lesage (1990) reported the results of co-integration tests that demonstrated a long-run economic relationship between basic and nonbasic employment in several of these MSAs, the combined results of this research effort seemed to provide evidence that such empirical work was both justified and could prove fruitful.

The results of Lesage and Reed's (1989) and Lesage's (1990) studies are already being cited in the literature. David S. Kraybill and Jeffrey Dorfman (1992), for example, used these authors' methodology to estimate a three-sector model for the state of Georgia. These and other recent contributions represent examples of what has become a third period of debate on the economic
base model, centered on the question of whether new techniques borrowed from macroeconomics time-series literature can revive the traditional economic base model.

Replicating and expanding this research, this author conducted extensive time-series econometric tests of the economic base hypothesis on models specified for the state of Wisconsin (Krikelas 1991). The results of this research, based upon a large number of two-sector and multisector model specifications, suggest that these new techniques do not provide the convincing evidence to support revival of the economic base model for purposes of long-term forecasting or planning context.

First and foremost, the fundamental problems associated with deriving adequate estimates of regional export activity remain unresolved. Although Lesage and Reed (1989) claimed that their dynamic location-quotient technique “provides a more accurate decomposition of local area employment” (1989, 616), this claim seems to be overstated. Krikelas (1991) confirms the results reported by Iserman (1980) and several others who have found that the location-quotient technique tends to underestimate the level of regional export activity and, consequently, lend an upward bias to export base multiplier estimates.

Second, in order to assess the stability of multiplier estimates derived from a bivariate VAR, Krikelas (1991) calculated impulse response functions for models that were based upon data generated from a variety of alternative sample separation techniques. The results of this experiment show that small changes in the way in which a given data set is divided into its basic and nonbasic components can lead to large changes in multiplier estimates. These results call into question the usefulness of the dynamic multipliers derived from a bivariate economic base VAR for even short-run regional impact analysis.

Finally, Krikelas (1991) explored the possibility of deriving multipliers from multisectoral VAR specifications and found similar difficulties. As the number of sectors included in a VAR is expanded, establishing identifying restrictions required in order to derive multiplier estimates becomes so arbitrary as to call into question the credibility of the multipliers derived from such specifications. As a result, any policy implica-

tions that might be implicit in a finding of significant differences between sectoral multiplier estimates would also be questionable.

More fundamentally, however, Krikelas concludes that the new techniques employed in Lesage and Reed and similar research do nothing to broaden the economic base paradigm’s focus on the demand side of the regional growth equation. Past research has clearly indicated that economic base models that fail to account for important supply-side factors and constraints do not perform as well as models that try to incorporate such relationships. Labor migration patterns, interregional capital flows, and state and local tax policies all have important effects upon regional economic growth and development and need to be incorporated into regional economic model specifications for the model to have value for anything other than short-term forecasting. Although it is possible to expand the bivariate economic base VAR to include some of these important supply-side variables, this author has concluded that such research would be largely in vain because other problems would remain (see Krikelas 1991). The recent attempt to breathe new life into the economic base model seems to have failed to resuscitate the patient.

**Conclusion**

Given the fact that several authors have begun to report empirical results in support of the validity of the economic base hypothesis, a third round of debate on the model seems already under way in the literature. An examination of some of the claims made by the proponents of these new dynamic economic base models, however, indicates that they are apparently unaware of the scope of the literature preceding their efforts. This brief analytical history should be sufficient to convince users that the economic base model has severe limitations, especially for economic planning and policy analysis, and to help make this next and perhaps final round of debate a relatively short-lived one.
Notes

1. Structural econometric models are often used for purposes of both forecasting and policy analysis. However, the great expense required to specify and maintain such models has generally led economists either to develop less complex models that focus narrowly on a small set of policy issues or to develop atheoretical time-series models that perform well for purposes of short-run economic forecasting.

2. Besides the terms basic and nonbasic, a number of others have been proposed to distinguish between the two types of economic activity: town builders/town fillers, exchange production/own production, primary/ancillary, export/local, as well as others. Andrews (1953b) directly addresses the issue of the profligate and confusing terminology of the economic base paradigm.

3. Krumme was translating Werner Sombart’s *Der Moderne Kapitalismus, Erster Band: Die Vorkapitalistische Wirtschaft*, 2nd rev. ed. (Munich: Duncker and Humblot, 1916). Sombart identified the city formers as “a king who collects taxes; a landlord who receives rent payments; a merchant who profits from trade with outsiders; a craftsman, a manufacturer, who sells industrial products to the outside; an author, whose writings are being bought outside the gates; a physician, who has clients in the countryside; a student, who is supported by his parents in another place, etc. These are the people who live and let live.”

4. Alexander cites a letter dated February 21, 1921, to John M. Glenn, a member of the New York Regional Planning Committee in which Olmsted wrote, “The multiplicity of their productive occupations may be roughly divided into those which can be considered primary, such as carrying on the marine shipping business of the port and manufacturing goods for general use (i.e., not confined to use within the community itself), and those occupations which may be called ancillary, such as are devoted directly or indirectly to the service and convenience of the people engaged in the primary occupations.”

5. According to Krumme’s translation, Sombart wrote, “It is necessary to find out for each trade how much of it is engaged in work for local consumption and how much in work for exports out of the city. This figure then is the city-forming ratio for the individual trade. Naturally, the ratio can be found accurately only with the assistance of an extensive enquete (survey). However, one could gain at least an approximate impression of the shares of the export industries in the total gainful employment by a careful investigation of the results of the occupational census” (1968, 116). The empirical study cited by Krumme was published for the first time in the second revised edition of Sombart’s *Der Moderne Kapitalismus, Dritter Band: Das Wirtschaftsleben im Zeitalter des Hochkapitalismus*, in 1927. Krumme, however, was quoting from the third printing of this edition, published in Berlin in 1955.

6. The following list identifies a few of the communities that performed economic base studies during the 1940s, the individuals or institutions that performed these analyses, and the base ratios ($\beta$) calculated, respectively: New York, The Regional Plan Association Inc., 2.1; Detroit, Detroit City Plan Commission, 1.1; Cincinnati, Victor Roterus and the staff of Cincinnati City Planning Commission, 1.7; Washington D.C., National Capitol Park and Planning, 1.1; Brockton, Massachusetts, Homer Hoyt, 0.8; the state of New Jersey, Homer Hoyt, 1.1; and Albuquerque, New Mexico, Federal Reserve Bank of Kansas City, 0.9. This information was originally compiled by Edward Ulman and published in the third edition of Weimer and Hoyt’s text in 1954 and was reprinted in Pfouts (1960, 30).

7. Hildebrand and Mace wrote, “The forthcoming Census of 1950 will permit further advances in this research. Recalculation of location quotients and comparisons with 1940 will indicate changes in external markets and locational concentrations during the war decade, particularly in communities undergoing large gains or losses in population. With monthly statistics of insured employment, a current record of employment in non-localized industries can be maintained. Improved multiplier analysis, with current local labor force statistics, should then permit more precise depiction of local unemployment problems and attainment of more adequate policies at the over-all and community levels” (1950, 249).

8. Perhaps the most often-cited contribution to the early regional input-output literature was an article coauthored by Isard and Kuenne (1953).

9. The index of surplus workers is simply a measure of the number of workers in excess of that which would be required if the region’s employment profile matched the national average.

10. Krikelas (1991) identified eighty-four contributions to the literature during this period.

11. Some of the sectoral multiplier studies conducted and the region or project for which they were calculated, include the following, respectively: retail trade multipliers calculated by Friedly (1965) for Redondo Beach, California; trade and service sector multipliers calculated by Terry (1965) for St. Louis, Missouri; defense industry multipliers calculated by Billings (1970) for the state of Arizona and by Erickson (1977) for the Badger Ammunition Plant, near Baraboo, Wisconsin; rural area multipliers calculated by Garrison (1972) for five nonmetropolitan counties in Kansas; and university sector multipliers calculated by Wilson (1977) for Tulsa, Oklahoma.


13. A VAR model consists of an equation for each variable in which the equations are estimated by regressing each of the variables against lagged values of all the variables. By not imposing any particular theoretical connection among the
variables, the VAR will capture any correlations that exist in the data. In this sense, VARs are distinct from traditional structural models, which typically include a large number of variables that are theoretically linked. 

15. Lesage, for example, reported on one of the few empirical tests recorded in the history of the literature that supports the economic base hypothesis as a long-run theory of regional growth and wrote that “this finding would not be particularly surprising to most regional economists” (1990, 309). His is one of several comments published recently that have pointed toward the need for presentation of a comprehensive history of the extensive body of literature that exists.

References


“Oskaloosa vs. The United States.” *Fortune*, April 1938, 58ff.


Path-Dependent Options: Valuation and Applications

William C. Hunter and David W. Stowe

Path-dependent options, unlike most claims whose value depends on the behavior of some other assets, are contracts entitling their holders to a cash flow that depends on the price path taken by the asset (stock, bond, commodity, and the like) underlying the contract. A relatively new class of options, their popularity has grown dramatically over the last decade, since they were first traded in 1982. That year, the trading of lookback options—so-named because at expiration the owner can “look back” over the life of the option and choose to buy or sell the underlying asset at the most favorable price that had occurred—demonstrated the value for investors of such path-dependent options as speculative and hedging instruments and proved their viability as traded securities.1

Standard European call and put options (giving the right to buy or sell, respectively, an underlying asset only on a particular expiration date) written on shares of common stock that pay no dividends have what are termed path-independent payoff structures. That is, their payoff is not influenced by the changes in market price of the underlying common stock between the date the option is written and its maturity date. In contrast, the payoff structures of path-dependent options are directly related to the price path followed by the option’s underlying asset over the life of the option. For example, a standard American put option (one that allows its owner to sell the underlying asset [exercise the option] at any time during the life of the option) written on a common stock has a path-dependent payoff structure.

Such path-dependent options have generated interest by filling several niches in derivative securities markets. Investors find them attractive because their design matches that of some financial contracts, giving them a kind of built-in insurance feature that makes it possible to limit potential losses and gains, and because they allow investors to better use their knowledge.
dependent options—the lookback option described above and the average-rate or Asian option—and their valuation using hedge portfolio and risk-neutral pricing techniques. Used most extensively in the foreign exchange markets, the average-rate option is European in structure and has a strike price based on the geometric or arithmetic average of the price of its underlying instrument over a specific period. The essential characteristics of these two types of path-dependent options along with the history of their development and uses as investment vehicles are described in a complementary article in the March/April 1992 issue of this Review. In addition to describing some basic features of the pricing models used to value these options, the sections that follow illustrate how these pricing models are implemented in practice.

The discussion also includes a brief presentation of the basic tenets of option pricing using the modern risk-neutral pricing technology developed by John Cox and Stephen Ross (1976). In addition, it explains how Monte Carlo analysis can be used to price Asian options written on the arithmetic average as well as how these options can be used to hedge foreign exchange risks from the viewpoint of individual investors and multinational corporations.

Valuation of Path-Dependent Options

Path-dependent options, like most contingent claims (that is, those whose value is tied to some other asset's behavior), can be priced using the hedge portfolio valuation methodology developed by Black and Scholes (1973) mentioned above. In simple terms, this approach implies that the cash flow obligation(s) involved in, for example, a lookback call option can be exactly met by the payoff from another portfolio—a hedge or replicating portfolio. More technically, if stocks and bonds can be used to construct a portfolio investment strategy that would provide the same cash flows as the contingent claim at the same points in time, then at any point throughout the lifetime of the contingent claim the claim's price must equal the value of the stock-bond portfolio at that time. This stock-bond portfolio hedge strategy forms the basis of the risk-neutral valuation methodology. The fundamentals of this methodology are presented in Box 1. More thorough discussions of option pricing fundamentals can be found in John Hull (1989) and Robert W. Kolb (1991). (Readers might also find helpful the standard options pricing notation collected in Box 2.)

The following discussion examines the valuation or pricing of lookback and average-rate or Asian options.

Valuing a Lookback Call Option. The lookback option is often referred to as a "minimize regret" option because it gives the purchaser the right to buy an asset at its lowest price or sell it at its highest price attained over the option's life. On a more sophisticated level, because its value is determined by the high or

Box 1

Valuing Standard European Call Options

Like many analytical models in finance, the basic model for pricing a standard European call option written on a share of common stock—the Black-Scholes option pricing model (named after its developers, Fischer Black and Myron Scholes 1973)—is based on a set of assumptions which, though abstract, work to simplify the valuation process. That is, it is assumed that

- the risk-free interest rate, \( r \), is constant and the same for all maturities.

At the time a call option on a stock expires, its worth is the greater of either the difference between the stock price at that time and the option's exercise price or zero, represented as \( \max(S^t - X, 0) \). Without this condition a riskless arbitrage opportunity would arise. To price the option today when there is time remaining before expiration, \( T-t > 0 \), the option's terminal price has to be approximated and the present value of this price computed using an appropriate discount rate.

The movements of stock prices over time can be modeled as following a random or stochastic process called geometric Brownian motion, which means that the stock price returns, which are defined as the natural logarithm of

\[ \log(S^t) - \log(S^0) \]

of asset price volatility in investing and hedging. Although path-dependent options do offer certain benefits, factors such as design mismatches and cost or risk mean that they will not satisfy the needs of every investor. The discussion of their pricing that follows will consider in greater depth both the advantages and risks of using these instruments.
the ratio of successive stock prices, $\ln(S_t/S_{t-1})$, are log-normally distributed. This model of stock price movement can be generalized to a continuous-time Markov process known as an Ito process. In simple terms, a Markov process is a stochastic process in which the observed value of the stock price (state variable) tomorrow depends only on its observed value today. The Ito process is characterized by a smooth predictable component—for example, the expected rate of return on the stock—and a highly erratic component that adds uncertainty or noise to the stock price movement. (See Box 3 for further discussion).

Given the properties of a log-normal distribution, the expected stock value at the option's expiration date, $S^*$, can be determined using the following equation:

$$E(S^*) = S_0 e^{\mu - \frac{\sigma^2}{2} T - \sigma \sqrt{T} \epsilon},$$  \hspace{1cm} \text{(B1)}$$

That is, the stock price is expected to rise continuously from today, $t$, until time period $T$, by its instantaneous continuously compounded expected rate of return $\mu$, less one-half the stock’s variance. The term $\sigma$ represents the standard deviation of the expected returns, and the term $(\mu - \frac{\sigma^2}{2})$ is called the drift rate of the stock price process.

A key aspect of the Black-Scholes model, however, is that one does not have to be concerned with the risk-adjusted expected return on a stock, $\mu$. It has been shown that if an investor held a portfolio containing a long position in a proportionate share of stock, $\Delta$, and a short position in one call option on the stock, $C$, then the investor would be perfectly hedged, with the portfolio generating a riskless rate of return. Perfectly hedged, the portfolio is equivalent to investing in a risk-free bond, $B$. This portfolio is given by

$$B = \Delta S - C. \hspace{1cm} \text{(B2)}$$

As is true for a risk-free bond, arbitrage will force the hedge portfolio to earn the riskless rate of return (Cox and Ross 1976). Therefore, an option's value will not be affected by a particular stock's expected rate of return, $\mu$, since it can be replaced with the riskless interest rate, $r$, using the Cox-Ross risk-neutral valuation framework.

If a call option at the time of expiration has a payoff of $\max(S^* - X, 0)$, the call option's value today is the present value of this expected payoff at expiration. That is,

$$c = e^{-rT} E(\max(S^* - X, 0)), \hspace{1cm} \text{(B3)}$$

where the term $e^{-rT}$ represents continuous discounting of the terminal payoff. The call value can also be derived using equation (B2) since the payoff from investing in a call option can be replicated by purchasing a proportionate share of stock, $\Delta S$, and borrowing at the risk-free rate, that is, short selling a Treasury security. Thus,

$$c = \Delta S - B. \hspace{1cm} \text{(B4)}$$

Black and Scholes derived an equation that essentially consolidates all of the steps required to compute the discounted value of a call option’s expected payoff at expiration. This equation can be written as follows (see Robert Jarrow and Andrew Rudd 1983):

$$c = e^{-rT} E(S^* \mid S^* > X) \times \text{prob}(S^* > X) \times e^{-r(T-t)} \times \text{prob}(S^* > X).$$  \hspace{1cm} \text{(B5)}$$

The first term in equation (B5), $e^{-rT} E(S^* \mid S^* > X)$, represents the present value of the expected stock price at the time of the option’s expiration, given that $S^*$ is greater than the exercise price. The second term, $\text{prob}(S^* > X)$, represents the probability that the stock price will be greater than the exercise price at expiration. The third term, $e^{-r(T-t)} \times \text{prob}(S^* > X)$, is the present value of the exercise price times the probability that the stock price will be greater than the exercise price at expiration. In short, the call option is worth the value of receiving the stock at expiration, conditional on the stock price being higher than the exercise price, minus the present value of paying the exercise price, conditional on exercising the option.

The general Black-Scholes model for computing an option's value over a longer period of time can be written in the following, more convenient closed form:

$$c = SN(d_1) - e^{-rT} SN(d_2), \hspace{1cm} \text{(B6)}$$

where

$$d_1 = \frac{\ln(S/X) + (r + \frac{\sigma^2}{2})(T-t)}{\sigma \sqrt{T-t}}$$

and

$$d_2 = d_1 - \sigma \sqrt{T-t}. \hspace{1cm} \text{(B7)}$$

In short, this model gives the discounted expected value of a call option at expiration. In this model, $SN(d_1)$ is the cumulative standard normal distribution function giving the probability that a random variable would be less than or equal to the value $d_1$. It reflects the uncertainty regarding the stock’s value on the option’s expiration date. Given a calculation of $d_1$, the value of $SN(d_1)$ can be found in the tables in the back of most statistics textbooks or approximated numerically. Compared to equation (B5), the term $SN(d_1)$ in equation (B6) represents $\Delta S$, or the proportionate share of stock, and the term $e^{-rT}SN(d_2)$ represents the remainder, $c - \Delta S$, or $B$.

The Black-Scholes formula shows that, to value a European call option, one needs to know only the current stock price, the exercise price, the time to expiration, the volatility of the stock’s returns, and the risk-free interest rate. With the exception of the volatility parameter, all of these variables can be directly observed. The volatility parameter can be estimated from historical data.

Although the Black-Scholes model employs some restrictive and likely unrealistic assumptions, it has nevertheless been shown to be quite robust when the underlying assumptions are modified.
Box 2

Guide to Options Pricing Notation

$T^*$ Terminal price of underlying asset at the time the option expires, $T$.

$S$ Spot price today. For foreign currency options, $S$ is the spot exchange rate, which is the price in U.S. dollars to buy one unit of a foreign currency today.

$L$ For lookback options only, it is the achieved minimum of the underlying asset’s price from the date of the option’s inception until $t$.

$C$ Call value at time of expiration.

$c$ Call value today.

$r$ Risk-free interest rate—for example, the three-month U.S. Treasury-bill yield. Assumed to be known and constant.

$r_f$ Foreign “risk-free” interest rate.

$\delta$ $r - r_f$, the interest rate differential.

$\lambda$ $2\delta/\sigma^2$. For the strike bonus option (lookback option model), it is the “speed” parameter.

$\sigma$ The volatility of a stock’s return per annum—standard deviation. It is assumed to be known and constant.

$T$ Expiration date of the option.

$t$ Today.

$T - t$ Time remaining until option expires expressed relative to 365 days.

$e$ The natural number 2.78.

$\mu$ The instantaneous expected return on the stock.

$\epsilon$ A random drawing from a standardized normal distribution with mean 0 and standard deviation of 1.

$E(S^*)$ Expected value of stock price at time of option’s expiration in a risk-neutral world.

$\Delta$ Delta; the change in the value of the call option given a $1 change in the value of the stock.

$N(*)$ The cumulative standard normal distribution.

The low price of the underlying asset over the life of the contract, the lookback option has the advantage of allowing investors to better use their knowledge about an underlying asset’s price volatility. As noted earlier, the key to pricing this instrument in the Black-Scholes framework is that the lookback option must be hedgeable.

In their 1979 article, M. Barry Goldman, Howard B. Sosin, and Mary Ann Gatto developed a model to value the lookback option using the Black-Scholes option pricing methodology and showed that a hedge portfolio could be constructed, allowing the lookback option to be valued without regard to the risk premium in the underlying asset’s expected return. As the example in Box 1 illustrates, the domestic risk-free interest rate, $r$, can be used in estimating the expected terminal stock price as well as to discount the option’s terminal value to the present. Goldman, Sosin, and Gatto showed that when the risk-free interest rate is equal to exactly one-half the asset’s volatility (as measured by variance), the lookback call option is identical to the purchase of a straddle (a portfolio of puts and calls on the same asset with the same strike price) on the asset. In this case, the writers of lookback calls can simply hedge their obligation by purchasing a straddle on the same underlying asset.

The examples that follow employ a more general model developed by Mark B. Garman (1989) to illustrate the lookback option’s essential features. The Garman model, while based on the work of Goldman, Sosin, and Gatto (1979), is simpler and offers additional insight into the option’s valuation by separating it into two underlying options. Moreover, unlike the earlier model, it can be used to value a European option on a “dividend-paying” asset such as a foreign currency option.

The Garman (1989) model for valuing a European lookback call option on a foreign currency, presented in equation (3) below, is a combination of two separate models: a European call option (see Garman and Steven W. Kohlhagen 1983), given by equation (1), and what Garman refers to as a strike-bonus option, shown in equation (2).

$$c = Se^{r_f(T-t)} N(d_1) - Xe^{-r_f(T-t)} N(d_2); \quad (1)$$

$$d_1 = \frac{\ln(S/X) + (r - r_f + \frac{\sigma^2}{2})(T-t)}{\sigma \sqrt{T-t}};$$

$$d_2 = d_1 - \sigma \sqrt{T-t}.$$
Equation (1) is a simple Black-Scholes valuation model for a foreign currency. The intuition behind this component of the model is explained in Box 1, with the difference that the foreign risk-free interest rate, \( r_f \), has been introduced to represent the continuous “dividend” or interest on the foreign currency. Equation (2) is similar to equation (1) except that the strike price is replaced by the underlying asset’s achieved minimum value, \( L \), from the time the option originated, along with other modifications. In the basic Black-Scholes model for a nondividend-paying stock, the valuation is carried out in a risk-neutral setting so that a stock’s expected rate of return is equal to the risk-free interest rate, \( r \). With a foreign currency, however, the return provided by the foreign interest rate, which does involve risk, must also be taken into account. The expected proportional growth rate of the spot exchange rate, \( S_t \), must be reduced by \( r_f \) to allow for the fact that the foreign currency can be invested to earn the foreign interest rate.

Equation (1) can be thought of as the value of an option to buy the asset at its minimum value achieved to date, from time \( t \), to \( T \). The strike-bonus option represented by equation (2) captures and prices or values the right to buy the underlying asset (the foreign currency) at the new minimum value likely to be achieved over the remaining life of the option, from time \( t \) until the expiration date \( T \)—that is, \( (T - t) \).

In general, the lookback call option is similar to a basic call option. For example, as it is for the basic call option, the value of the lookback call is positively correlated with the spot price of the underlying asset, the domestic risk-free interest rate, the time remaining until the option expires, and, most importantly, the volatility of the underlying asset. In addition, in both cases the option is negatively correlated with the contract’s exercise price and the foreign risk-free interest rate.

Charts 1 and 2 illustrate the price behavior of a basic European call option and a strike-bonus option, respectively, over various initial strike prices ranging from $0.80 to $1.00. In these charts the initial strike price used for the basic call and the strike-bonus option is the minimum value, \( L \), achieved by the underlying asset since the option’s origination. Chart 3 gives the value of the lookback call option (the sum of the value of the European call option and the strike-bonus option) over the same range of initial strike prices.

Although Charts 1 and 3 appear similar, the values of the basic call and lookback options (on the y-axis) differ as \( L \) is increased. The two options have approximately the same value when \( L \) is very low relative to the spot price in dollars.

---

**Chart 1**

The Price Behavior of a Basic European Call Option

- **Chart**
- **S** = $1.00
- **R** = 9.5%
- **V** = 4.0%
- **R_f** = 4.9%
- **t** = 1 Year

---

Economic Review

July/August 1992
Chart 2
The Price Behavior of a Strike-Bonus Option

Price in Dollars

Achieved Minimum Value (L) in Dollars

$ S = \$1.00 $
$ R = 9.5\% $
$ V = 4.0\% $
$ R_f = 4.9\% $
$ t = 1 Year $

Chart 3
The Price Behavior of a Lookback Call Option

Price in Dollars

Achieved Minimum Value (L) in Dollars

$ S = \$1.00 $
$ R = 9.5\% $
$ V = 4.0\% $
$ R_f = 4.9\% $
$ t = 1 Year $
price, \( S \). At this point, the call option is “deep in-the-money”—the price of the underlying asset exceeds the strike or exercise price, and the lookback’s value is derived primarily from the basic call option’s value. The strike-bonus option is relatively worthless. However, the value of the basic call option declines further than that of the lookback call option as its strike price, \( L \), is increased to $1.00 (and is said to be “at-the-money”). Hence, the closer the achieved minimum asset price is to the spot price, the higher the value of the strike-bonus option. As the call option moves closer to being at-the-money, the value of the strike-bonus option pushes the lookback’s price above the price of the basic call.

Thus, while the relationship between the lookback call option’s value and the initial strike price, \( L \), is similar to that of the basic call option, the relationship between the value of the lookback’s strike-bonus component and the strike price is opposite that of the basic call option. That is, other things being equal, the value of the strike-bonus option is positively correlated with the initial strike price.

As stated earlier, the strike-bonus option gives the holder the right to buy the asset at a new minimum value anticipated from the perspective of time \( t \). Given a very low initial minimum value, \( L \), relative to the spot price, the probability is low that the asset’s price will fall to a new minimum value and then rise again before the option expires, so the strike-bonus option has little value. On the other hand, if the option were at-the-money—the achieved minimum equals the spot price—the asset very likely would establish a new minimum price and subsequently rise before expiration. The strike-bonus option derives its value from the probability that the asset’s price will achieve such a new minimum value. Therefore, for options having initial at-the-money prices, the lookback call option’s value can be substantially higher than that of a comparable basic call option.

Besides the relationship between the exercise or strike price and the value of the underlying asset—the moneyness of the option—another factor influencing the value of both the basic call and the lookback call option is the underlying asset’s volatility. While either option’s value is positively correlated with volatility, for a lookback call option this factor is more significant because it affects the probability that the price of the underlying asset will achieve a new minimum prior to the expiration date. For instance, holding other factors constant, the value of the basic at-the-money call option priced above increased by 1.5 percent given an increase from 2.0 percent to 2.1 percent in the estimated variance of the underlying asset. In this same situation, however, the value of the strike-bonus option increased by 2.34 percent.

**Valuing an Average-Rate (Asian) Option.** There are two types of average-rate or Asian options: the fixed-strike and the floating-strike. The fixed-strike average-rate option is one for which the terminal payoff is the maximum of either the difference between the average value of the underlying asset and a fixed strike price or zero: \( \max(Avg_s - X, 0) \). The floating-strike Asian call option is similar to the lookback option given that its payoff at expiration is equal to the greater of the difference between the underlying asset’s terminal spot price and its average value over the life of the option or zero—that is, \( \max(S - Avg_s, 0) \).

The floating-strike average-rate option is comparable to a lookback call option for which the strike price is the average value of the underlying asset, as opposed to its minimum value. The value of this option can never be greater than the value of a regular lookback call option, whose strike price is the achieved minimum of the asset, because the average is never equal to an extreme value such as the minimum unless all of the values are equal. Thus, the price of a regular lookback option sets an upper bound on the value of the floating-strike average-rate option.

The fixed-strike average-rate option is more commonly used in practice than the floating-strike option and will be the focus of this discussion. For both the floating- and fixed-strike average-rate options, the average can be computed as either the geometric or the arithmetic average. Although in practice the arithmetic average is typically used, no closed-form equation has been developed for pricing the average-rate option written on the arithmetic average because of mathematical complexities. Thus, a numerical approximation technique must be used to value an option written on the arithmetic average of the underlying asset’s price (see A.G.Z. Kemna and A.C.F. Vorst 1990). In the study presented here Monte Carlo simulation was the numerical approximation valuation methodology chosen. (See Box 3.)

**Limiting Bounds on the Value of an Arithmetic Average-Rate Option.** Theoretically, the value of an option written on an asset’s arithmetic average should lie between the values of a comparable European call option and an average-rate option written on the geometric average of the asset’s price. The primary difference between the fixed-strike average rate option and a comparable basic call option is the volatility of the underlying spot price. Because the variance used to compute the value of the basic European call option exceeds the variance used to compute the value of the
Monte Carlo simulation is a numerical procedure used to approximate the expected value of a random variable or vector. The procedure approximates the expected value by generating random variables or vectors with a given probability density or joint probability density and, using the law of large numbers, takes the average of these values as an estimate of the expected value.

The Monte Carlo method is used to compute the expected value of a European-style call option. The value is computed by simulating the path taken by the price of the asset underlying the option over time. By simulating numerous such price paths, the technique allows computing the expected value or price of the option with increasing precision as the number of iterations or runs of the simulation are increased. The example that follows illustrates the technique.

In this exercise the objective is to compute the expected payoff of a European-style call option as of its expiration date, \( E[\max(S^n - X, 0)] \) and then to compute the present value of this quantity—that is, \( e^{-r(T-t)} E[\max(S^n - X, 0)] \). Given a model for stock price movements over time, the Monte Carlo simulation technique will allow computation of the price expected at expiration.

Each iteration or sample run of the Monte Carlo simulation gives a stock price at time \( T \). Increasing the number of runs of the simulation increases the reliability of the estimate for \( S^n \). (For example, 100,000 iterations would be reasonable.) The expected value of \( S^n \) is an average of the values obtained over the number of iterations.

The stock’s price path can be modeled using the following equation:

\[
S_{t+1} = S_t e^{-\sigma^2/2\Delta t} + \sigma \varepsilon \sqrt{\Delta t},
\]

where \( r \) is the risk-free interest rate, \( \sigma \) is the standard deviation of stock returns, \( \Delta t \) represents a very small change in time (a day, for example), and \( \varepsilon \sqrt{\Delta t} \) is a discrete approximation to an increment in a Wiener process, in which \( \varepsilon \) represents a random drawing from a standardized normal distribution (a normal distribution with mean 0 and variance 1).

In equation (M1), \( S_t \) is the value of the stock price today, it is the risk-free interest rate, \( \sigma \) is the standard deviation of stock returns, \( \Delta t \) represents a very small change in time (a day, for example), and \( \varepsilon \sqrt{\Delta t} \) is a discrete approximation to an increment in a Wiener process, in which \( \varepsilon \) represents a random drawing from a standardized normal distribution (a normal distribution with mean 0 and variance 1).

Suppose, for example, that the stock price today, \( S_t \), is $50.00, \( \Delta t \) is 1/365 years, and the annual risk-free interest rate and standard deviation of the stock’s expected return are 8 percent and 20 percent, respectively. Then, using the model in equation (M1), the stock’s price one day hence can be approximated by

\[
S_{t+1} = 50e^{-0.08/365} + 0.20 \varepsilon \sqrt{\frac{1}{365}}.
\]

To complete the approximation of \( S_{t+1} \), a value of \( \varepsilon \) needs to be randomly selected from a standardized normal distribution. (This task can typically be accomplished with an internal function available in many computer program languages.) If, for example, the random selection for \( \varepsilon \) is 2.1, the stock’s price one day hence would be expected to be $51.11. Another random sampling for \( \varepsilon \) would yield a different stock price for \( S_{t+1} \).

In practice, these steps would be repeated numerous times to get a stock price path from \( S_t \) to \( S_{t+\nu} \), where \( S_{t+\nu} \) would be the stock price as of \( \nu \) days from now—that is, \( S^n \) at expiration. The option’s terminal value would then be computed and this amount would be discounted back to today using the formula

\[
Call = e^{-r(T-t)} \max(S^n - K, 0).
\]

These steps would complete one sample iteration of this Monte Carlo simulation.

After performing the appropriate number of runs, the expected call option value, \( Call \), is computed as the average of the estimates of the call option’s value obtained over the iterations. Next, the standard deviation of the estimates of \( Call \) is computed. Finally, given an expected value of \( Call \) and the standard deviation of the expected value, a confidence interval—a range within which there is some level of assurance that the actual value will fall—can be constructed for the value of the option. For normally-distributed variables there is a 95 percent probability that the actual value of the variable will lie within a range of ± 2 standard deviations from its sample mean.

Suppose, for example, that a European-style call option on a stock has eighty-seven days until expiration and an exercise price of $35.00. Today’s stock price is $40.00, its annual volatility is 20 percent, and the yield on a Treasury bill with just over eighty-seven days until maturity is 2.96 percent. The Monte Carlo simulation consists of 10,000 iterations, giving 10,000 approximations for the present value of the call option. The mean value of the call is $5.361, and the standard deviation of the estimate was $0.0335. A 95 percent confidence interval for the value of the call can be constructed as follows:

\[
$5.361 \pm 2(0.0335) = [$5.29, $5.43].
\]

That is, there is a 95 percent probability that the actual value of this call option is within the computed range. This confidence interval can be tightened—thus obtaining a more reliable estimate—by increasing the number of iterations in the simulation or by employing a variance reduction technique (see note 4). The desired level of accuracy depends on a particular application.

In short, the Monte Carlo method simulates the random movement of stock prices and provides a probabilistic solution to the option pricing problem. Thus, by the technique’s very nature, the final value derived using the

---

**Box 3**

**Monte Carlo Simulation: A Tutorial**
Monte Carlo simulation will not be exact, no matter how accurate the equations in the model.

The advantage of using this simulation technique is that it can accommodate complex payoffs such as those that depend on the history of the underlying asset's price over time—for example, lookback and average-rate options. However, the technique has the disadvantage of being time-consuming, depending on the number of iterations required to obtain a reasonably accurate approximation, and it may therefore be infeasible for some practical applications.

call option written on the average price of the same underlying asset, the value of the call option on the asset's average value will never be greater than the value of a standard European call option on the asset. Furthermore, because the arithmetic average is always greater than the corresponding geometric average, the value of an option written on the geometric average of an asset's price provides a lower bound for the value of an option written on the arithmetic average of the price of the same asset.

Chart 4 shows the bounds imposed on an Asian option's value written on the arithmetic average. The value of a basic European call option forms the upper bound, and the value of a call option written on the geometric average delineates the lower bound, both having the same underlying parameters ($S, X, r, \sigma, \text{ and } t$).

The price of the arithmetic Asian option for a given spot price, computed using the Monte Carlo method, lies within the shaded area. The following relationships hold: \(^3\)

$$\text{Call}_{\text{European}} > \text{Asian Call}_{\text{arithmetic}} > \text{Asian Call}_{\text{geometric}}.$$  

The sections that follow illustrate the application of the valuation models discussed above to investment and hedging decisions in the foreign exchange markets. A Monte Carlo analysis like that used in Box 3 is employed to value the Asian option written on the arithmetic average with the exception that in the example the terminal price of the underlying asset is taken to be the arithmetic average of the $n$ daily stock prices, where $n$ is the number of days from time $t$ until expiration. \(^4\)
Numerical Examples. Suppose that a U.S. investor firmly believes that the Japanese economy will improve dramatically relative to other industrialized countries’ economies during the upcoming year. To act on this expectation, the investor purchases a one-year call option on the Japanese yen with a strike price of $0.0073. In this context, the investor is speculating or predicting that the cost of the yen in U.S. dollar terms will rise by the time the option expires—that is, the yen will appreciate against the U.S. dollar. Assume that as of the current date it costs the investor $0.0073 to buy one unit of Japanese yen (1/0.0073 translates to a current exchange rate of 136.9 yen per one U.S. dollar) and that this cost rises to $0.0079 (126.6 yen/$) over the one-year period. At the end of the year it costs more U.S. dollars ($0.0079 - $0.0073) to purchase one unit of Japanese yen (the dollar depreciated). Because the option is very valuable at expiration, the investor’s decision to purchase the one-year call option with a $0.0073 strike price turns out to have been a profitable one.

The example illustrates how an investor can use a basic currency call option in speculating on foreign currency movements. On the other hand, if the investor knew that dollars would have to be exchanged for yen at the end of the year—as might be the case for a business that imported goods from Japan with a commitment to make payment in yen—the purchase of the yen call option would constitute a foreign exchange risk hedging transaction.5

The following discussion illustrates more extensively how the path-dependent call option valuation models presented in this article are implemented using actual foreign exchange data on the Japanese yen and the German deutsche mark. The performance of these options is compared with that of a basic European call option written on the same currencies, and some of the risks inherent in using these instruments are pointed out.

The Data and Assumptions. Consider that on January 8, 1991, an investor is aware that one year from this date it will be necessary to purchase one million units of each of two foreign currencies: Japanese yen and deutsche marks. The investor’s objective is to eliminate the upside exchange rate risk in this purchase requirement. That is, he or she wishes to avoid being adversely affected by a rise in the foreign currencies’ value over the next year.

The investor is assumed to have four alternative options on the foreign currencies from which to choose: a basic call option, a lookback call option, a fixed-strike Asian option, and a floating-strike Asian option.6 All four options are European style, and the size of each option contract is for the purchase of one million units of the underlying foreign currency. To simplify matters, all options will be initially at-the-money. In other words, on the date they were purchased (time \( t_0 \)), the spot foreign exchange rates were equal to the options’ strike prices.

As will be discussed further, the choice of instruments can be difficult. In terms of the cost or premium required to purchase these options, the average-rate options are inexpensive relative to the lookback option. However, depending on the behavior of the exchange rates, the investor’s hedging needs, and his or her knowledge of exchange rate volatility, the more expensive lookback could be preferred.

Table 1 provides a hypothetical example of a buy-and-hold strategy on each of the four types of call options available to this investor. Each option’s performance as a hedging or investment vehicle is analyzed using the Japanese yen and the deutsche mark, given the net changes in each currency’s exchange rate over the one-year period.

Option transactions are initiated on January 8, 1991, to expire January 10, 1992 (367 calendar days). The basic data for this analysis are given in the first panel. So that all options will be priced at-the-money, the strike price is set equal to the initial spot exchange rate for the yen—$0.0073 per one unit of yen (the inverse of 133.3 yen per one U.S. dollar) and $0.6536 per one deutsche mark. The table also includes the domestic and foreign risk-free interest rates and estimates of asset price volatility for this time period.

The estimated initial cost of each option is shown in the second panel. The calculated premium for each option was multiplied by one million units of each currency to obtain the cost of the option contract. For both the fixed- and floating-strike average rate or Asian options, the Monte Carlo method was used in estimating value. Equations (3) and (1) were used to value the lookback and basic call options. Note the high cost of the at-the-money lookback call option relative to the basic call, as would be expected in light of the earlier discussion of the value of the strike-bonus option component of the lookback option.

The third panel shows the values of the four options at expiration. At this time the options are worth the greater of their intrinsic values, \( S - X \), or zero since no time premium remains. The profit/loss on each position reported in the fourth panel represents the options’ terminal values minus their initial costs.

Summary of the Results. As can be seen in the fourth panel, for the yen options all but the fixed-strike Asian call ended the period profitable. The fixed-strike
Table 1
A Hedging/Investment Application

<table>
<thead>
<tr>
<th></th>
<th>Yen</th>
<th>Mark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated volatility</td>
<td>14.14%</td>
<td>13.06%</td>
</tr>
<tr>
<td>Foreign interest rate</td>
<td>7.64%</td>
<td>9.13%</td>
</tr>
<tr>
<td>U.S. interest rate</td>
<td>6.91%</td>
<td>6.91%</td>
</tr>
<tr>
<td>Strike price</td>
<td>$0.0073</td>
<td>$0.6536</td>
</tr>
<tr>
<td>Average exchange rate</td>
<td>$0.0075</td>
<td>$0.6047</td>
</tr>
<tr>
<td>Minimum exchange rate</td>
<td>$0.0070</td>
<td>$0.5435</td>
</tr>
<tr>
<td>Spot (terminal) exchange rate</td>
<td>$0.0079</td>
<td>$0.6329</td>
</tr>
<tr>
<td>The U.S. dollar's net change</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

|                        | Depreciated | Appreciated |

Estimated Initial Cost of Call Option Contract
(at the money)\(^b\)

<table>
<thead>
<tr>
<th></th>
<th>Yen</th>
<th>Mark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asian (floating-strike)</td>
<td>$208</td>
<td>$14,930</td>
</tr>
<tr>
<td>Asian (fixed-strike)</td>
<td>$209</td>
<td>$15,032</td>
</tr>
<tr>
<td>Lookback call</td>
<td>$710</td>
<td>$54,000</td>
</tr>
<tr>
<td>Basic European call</td>
<td>$359</td>
<td>$25,200</td>
</tr>
</tbody>
</table>

Option Value at Expiration
(01/10/92)

<table>
<thead>
<tr>
<th></th>
<th>Yen</th>
<th>Mark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asian (floating-strike) max (Spot[T]—Average, 0)</td>
<td>$400</td>
<td>$28,200</td>
</tr>
<tr>
<td>Asian (fixed-strike)   max (Average—Strike, 0)</td>
<td>$123</td>
<td>$0</td>
</tr>
<tr>
<td>Lookback call          max (Spot[T]—Minimum, 0)</td>
<td>$823</td>
<td>$89,433</td>
</tr>
<tr>
<td>Basic European call    max (Spot[T]—Strike, 0)</td>
<td>$542</td>
<td>$0</td>
</tr>
</tbody>
</table>

Profit/(Loss) on Option\(^c\)

<table>
<thead>
<tr>
<th></th>
<th>Yen</th>
<th>Mark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asian (floating-strike)</td>
<td>$192</td>
<td>$13,270</td>
</tr>
<tr>
<td>Asian (fixed-strike)</td>
<td>($86)</td>
<td>($15,032)</td>
</tr>
<tr>
<td>Lookback call</td>
<td>$113</td>
<td>$35,433</td>
</tr>
<tr>
<td>Basic European call</td>
<td>$183</td>
<td>($25,200)</td>
</tr>
</tbody>
</table>

\(^a\) Option contract size is one million units of foreign currency. Time remaining until expiration (01/08/91–01/10/92) is 367 days.

\(^b\) Premium times one million units of currency.

\(^c\) Terminal values minus initial costs.
Asian option cost $86.00 more than its value at expiration (computed as the difference between the average of the daily exchange rates over 367 days and the strike price). Over the period analyzed the yen appreciated, net, from $0.0073 to $0.0079. Thus it cost more in U.S. dollar terms to purchase one yen at the maturity date than at origination. As can be seen, the basic call option struck initially at-the-money expired in-the-money, generating a profit of $183.00. The lookback call option was also profitable even though its initial cost was nearly twice that of the basic call. This value reflects the path taken by the yen over the period. As illustrated in Chart 5, the yen initially declined through June 1991 and then began to trend upward thereafter.

In the case of the deutsche mark, it is interesting that only the lookback options were profitable at expiration. The lookback call and the floating-strike average rate option generated profits of $35,433.00 and $13,270.00, respectively. The basic call option position resulted in a loss of $25,200.00 and the fixed-rate Asian option produced a loss of $15,032.00. As Chart 6 shows, over this one-year period the mark depreciated, net, against the U.S. dollar. If the dollar/mark exchange rate had followed a downward path over the period, all of the option positions would have expired worthless. However, the mark depreciated significantly in the first half of the year and then appreciated thereafter, ending just below its beginning rate.

As in the case with the Japanese yen, the profit associated with the lookback option can be attributed to the range achieved by the exchange rate during the period. In initially pricing the lookback call on the mark an annualized volatility estimate of 13.06 percent was used. The actual volatility over this period was a much higher 17.4 percent, and the unexpected increase in volatility positively affected the lookback option’s value.

An investor who accurately forecasts the range of the dollar/mark exchange rate over the assumed investment or hedging horizon given in this example would have found the lookback call or floating Asian options attractive investment vehicles. It should be emphasized, however, that investors must understand the risks inherent in these instruments to use them effectively (as is the case with most derivative products). This point takes on added importance given that these contracts are used mostly in foreign exchange transactions. For a variety of reasons it is extremely difficult
to forecast exchange rates, even when using large-scale structural econometric models. It is likely that such forecasting difficulties will also be encountered when attempting to forecast parameters of the price processes that determine the value of path-dependent foreign exchange options. In any number of cases other, more traditional derivative instruments such as futures and forwards or various types of swap products may represent a more cost-effective way for investors or hedgers to achieve their stated objectives. Thus, while many may find contracts including path-dependent options attractive, they will not satisfy the needs of every investor or hedger.

**Conclusion**

This article describes the way in which two popular path-dependent options—the lookback option and the average-rate or Asian option—are priced using modern options pricing techniques. The hedge portfolio and risk-neutral pricing approaches associated with the modern options pricing paradigm are used to price the lookback option, and Monte Carlo simulation is used in pricing the floating-strike and the fixed-strike average-rate or Asian options. A simple numerical example involving these options demonstrates the possible benefits, costs, and risks associated with their use.

Since their formal introduction in 1979, these options have experienced rapid growth. In one form or another they are now traded on exchanges, are used extensively in foreign exchange markets to hedge foreign exchange risks, and have found use in corporate mergers and acquisitions. While these options are likely to continue to find new applications and uses as hedging and/or investment vehicles, investors are advised to develop a full understanding of the risks inherent in these instruments before adding them to their investment or hedging portfolios.
Notes

1. Research published in 1979 by Goldman, Sosin, and Gatto paved the way for the modern treatment of path-dependent options. The authors had derived an explicit valuation formula for a hypothetical option with an exercise price to be set at the contract’s expiration rather than at origination, as it is for standard options. Payoff on such “lookback” options would depend on the price path followed by the contract’s underlying asset over the life of the option as well as on the option’s expiration date. Goldman, Sosin, and Gatto suggested that such options could be of value to investors, and they were proven correct when Macotta Metals Corporation of New York began trading lookback options on gold, silver, and platinum on March 16, 1982 (Hunter and Stowe 1992).

2. The Black-Scholes environment assumes that the natural logarithms of stock price returns, $\ln(S_t/S_{t-1})$, are normally distributed. In valuing an option written on the geometric average of an underlying asset’s value over time, this crucial assumption still holds because the product of the logarithms of stock price returns is normally distributed. However, this assumption fails to hold in the case of an Asian option written on the arithmetic average because the sum of the logarithms of the stock price returns over time is no longer normally distributed. Therefore, there is no closed-form analytical pricing equation for the average-rate option written on the arithmetic average (Ritchken, Sankarnsubramanian, and Vijh 1990).

3. The closed-form equation for pricing the Asian option written on the geometric average is derived in Ritchken, Sankarnsubramanian, and Vijh (1990).

4. In valuing the arithmetic average-rate option, knowledge of the valuation principles for the geometric average-rate option can be used to lower the standard error of the estimated value and therefore reduce the number of iterations required in the simulation. This procedure is known as the variance reduction technique and is described in Hull (1989) and Kemna and Vorst (1990).

5. Although the discussion to this point has focused only on the lookback call option, in the case of foreign currency options a call on one currency can be thought of as a put on another currency. That is, a call option to buy yen for U.S. dollars can be thought of as a put to sell U.S. dollars for yen. Thus, the same methodology discussed for valuing a lookback call option on a foreign currency can be used to value a lookback put option on a foreign currency.

6. Although this example considers only the use of the lookback call, average-rate, and basic European call option, in practice other hedging vehicles would also be considered or evaluated by the investor, depending on the investor’s objective.

References


Very few scientific breakthroughs emerge whole cloth outside of a well-established discipline, which typically includes both a theoretical context and an empirical tradition. Most progress builds on accepted and widely scrutinized work, either extending or debunking the findings of previous research. Consequently, most research begins with a review of previous relevant studies. Economic research is no different. The economics profession regularly takes stock of the current state of the discipline by evaluating the record since the last "big" breakthrough. These benchmarks constitute a map by which the discipline's development can be charted. A recent publication in this tradition, The State of Macroeconomics, edited by Seppo Honkapohja, is subtitled Proceedings of a Symposium: Whither Macroeconomics? While the title promises a wider assessment than the contents deliver, the collection contains several valuable essays that critically survey developments in macroeconomics since the mid-1970s.

The book consists of nine essays with commentaries and is divided into four parts. The first contains broad overviews of modern macroeconomics and ends with an assessment of the discipline's current state from the practical perspective of empirical work and policy analysis. The second and third parts focus, respectively, on two topics pertinent to policymakers—labor markets and exchange rates. The final section looks forward by discussing open issues in macroeconomics.

Macroeconomics is a young field. Many economic historians credit John Maynard Keynes with its emergence, in the 1930s and 1940s, as an approach...
distinct from microeconomics. Generally, microeconomics focuses on the composition of economic activity—that is, the manner in which individual units (consumers, households, firms) interact to determine the relative prices of goods and the quantities to be bought and sold—and on the allocation of all the resources in an economy among competing uses. In contrast, macroeconomics focuses on the level of economic activity, using highly aggregated data and summary measures of activity, such as employment, gross domestic product, and inflation. Macroeconomics is concerned with economic fluctuations—business cycles—and what can be done about them.

In asserting the existence of an unemployment equilibrium, Keynes defied the accepted wisdom of his day, which contended that unemployment exerted enough downward pressure on wages to eventually alleviate unemployment; hence, there could be no equilibrium with unemployment. Keynes’s assertion ignited a controversy that continued for years. A.C. Pigou, one of Keynes’s colleagues and a tenacious defender of the orthodox Keynes attacked, wrote in 1950 acknowledging and defining the unique macroeconomic perspective of Keynes’s *General Theory*: “Nobody before him, so far as I know, had brought all the relevant factors, real and monetary at once, together in a single formal scheme, through which their interplay could be coherently investigated.”

Forty years later Robert Solow echoed this idea, arguing that, before Keynes, “there was no determinant model lurking behind the story telling.” The *State of Macroeconomics* reflects this viewpoint in that it evaluates macroeconomics as a discipline that requires a general unified theory but is still in search of one. The book focuses mainly on one manifestation of this search—the continued struggle with unresolved issues on the proper role of microeconomic principles in macroeconomic theory and policy-making.

Between the 1940s and the early 1970s there was little explicit micro content to macroeconomics. That is, macroeconomics lacked well-specified assumptions about the rules that guided individual decision-making. One of the major breakthroughs of the 1970s demonstrated that the behavioral assumptions implicit in large-scale models of the economy were completely inconsistent with rational behavior at the individual level. Thus exposed and chastened, macroeconomists sought respectability by attempting to impose consistent microfoundations on their macroeconomic models.

After fifteen years of pursuing this laudable goal, the profession is humbled and at odds with itself. This book reflects that underlying dilemma—each of the essays discusses an approach to establishing acceptable foundations for some facet of macroeconomic research. In fact, this search for a general theory resonates throughout the entire macroeconomic literature of the last two decades. Macroeconomic models that are theoretically coherent, in that they have solid microfoundations, tend to be extraordinarily difficult to solve and are thus empirically irrelevant. On the other hand, models that deliver empirical implications have inadequate economic foundations, rendering them inappropriate for policy-making and forecasting even though they may describe data very well. Without a careful specification of individual behavior (or, equivalently, a “representative” individual), models that mimic past economic conditions are useless for forecasting because they are based on the outcomes of events rather than the decision-making process that determines the response to events. For example, it is not difficult to find a simple relationship between the unemployment rate and wage growth when prices are stable. However, a model that excludes the way in which inflation expectations are incorporated into decisions about wages and employment will fail to predict the unemployment rate/wage growth relationship once inflation expectations become a factor. Models that are invariant to the policy-setting (that is, that capture agents’ responses to underlying changes induced by new policy regimes) should be based on such fundamental factors as preferences, technology, endowments, and information, but such factors are difficult to observe and measure.

In response to these difficulties some economists have taken the high road of exquisitely correct microeconomic foundations, which rarely yield insights useful to policymakers; others have taken the low road of intense knowledge about local events (such as economic data), which, though useful within the context of an unchanged setting, cannot be relied upon otherwise—a severe limitation. The profession continues to struggle with this dilemma, and more than one economist has, in despair, dismissed the empirical relevance of more than a decade’s work on incorporating micro-into macroeconomics. Despite discouraging progress in coming to terms with the practical implications of microeconomic foundations, a large body of solid research has accumulated over the last fifteen years, and it continues to be pursued. *The State of Macroeconomics* examines only selected issues from this literature, but the selection is broad enough to reflect many of the contentious issues.

In the first article, “New Classical Macroeconomics, A Sympathetic Account,” Bennett McCallum reviews...
the 1970s' and 1980s' debates on macroeconomics and lays the groundwork for issues discussed in the remainder of the book. He distinguishes two main approaches, referring to them as the new classical and Keynesian—the new guard and the old guard, respectively. Many will argue that each of these two rubrics represents beliefs so diverse that they cannot be grouped under the same umbrella. Still, McCallum's dichotomy allows for a clear contrast, and the distinction is followed throughout the volume.

*The State of Macroeconomics* could be retitled *The New Classical Economics: Fact and Fiction* because most of its content focuses on the role of the new classical approach in contemporary macroeconomics. The new classical approach, so called because it harks back to the era when all economic analysis focused on interactions among individuals, is identified as attempting to deal explicitly with the microeconomic setting.

McCallum identifies several traits at the heart of the new classical economics. The first is the equilibrium approach to economic modeling, “in which private agents optimize in light of their own objectives and constraints and markets clear.” The phrase “optimize in light of their own objectives and constraints” describes rational behavior, which is taken for granted at the individual level. This assumption appears to be uncontroversial, and it is. However, its implications for economic modeling are profound in that this assumption, while necessary for constructing a framework suitable for policy analysis, at the same time complicates the task. In contrast, the equilibrium approach’s market-clearing assumption is quite controversial because it can be interpreted as referring to instantaneously clearing markets, which would require flexible prices; not only is this assumption unrealistic, but it also complicates modeling the imbalances that lead to cyclical fluctuations.

The second feature of the new classical approach that McCallum cites is the natural rate hypothesis, which states that the rate of unemployment can be only more or less inflation than anticipated. The acceptance of this hypothesis requires policymakers to acknowledge limits, for example, to countercyclical policy and limits on effective counter-cyclical policy (that is, random or capricious actions) had real effects, then the only role for effective counter-cyclical policy was in well-recognized rules that would limit price expectations. Later research showed that this result was a product of the extremely limiting assumptions of the models used to explore policy effects and of the structure of expectations formation in those models. Later models did not exclude systematic effects on economic performance but did not necessarily deliver them either. Still, the dominance of rules over discretion demonstrated in the early models had an intuitive appeal and a pertinence to then-current policy issues that made a lasting impression on the debate.

McCallum’s most controversial claim for the new classical approach is his dismissal of bubbles (sunspots) as important for understanding fluctuations in economic activity. Bubbles are self-fulfilling expectations, such as speculative booms in the stock market, that are not based on underlying changes in true value. In theory, sunspots can play an important role in generating cyclical fluctuations. The role of bubbles continues to be explored but is not one of the basic characteristics of the new classical approach.

Both of the commentaries on McCallum’s essay stress that the factors chosen by McCallum are neither necessary nor sufficient to define new classical theory. In general, the commenters object more to the exclusion of important elements than to the inclusion of the four chosen by McCallum.

In an essay titled “Keynesian Issues and Economic Theory,” Jean-Michel Grandmont disagrees with part of McCallum’s list, especially the discounting of bubble phenomena as important to explaining fluctuations. Grandmont identifies other factors as more definitive of the new classical approach to macroeconomics. He sees the new classical approach as encompassing the view that fluctuations are evidence of stability because the self-adjusting mechanisms of free markets generate them.

One contributor less sympathetic to the new classical macroeconomics, Michael Bordo, still identifies valuable contributions from it. He acknowledges theoretical developments from the new classical literature as justifying a good deal of eclecticism in macroeconomic policy-making. He cites two factors characteristic
of new classical models—the importance placed on forward-looking behavior and the recognition that price and wage decisions are made at different times—as both changing forever the way decisions are modeled regardless of the theoretical setting. Each of these characteristics, he argues, can be incorporated into nonclassical models without transforming them into classical models with classical policy implications, insofar as classical policy implications are generally characterized as strongly limiting the role of countercyclical monetary policy. Bordo also claims that the empirical relevance of many of the new classical characteristics is quite limited compared to the theoretical contribution. For example, while forward-looking expectations and flexible prices are “reasonable” assumptions, it may be necessary to make unreasonable assumptions about just how forward-looking and how flexible they must be to achieve the results claimed in the new classical models. Bordo also stresses that the debate opened by new classical theoretical developments about expectations has led to a recognition of the important role of political factors as well as policymakers’ credibility and commitment in the economic policy debate. Again, these elements have proved widely applicable outside the new classical setting.

The question of proper microfoundations, which dominates the broad issues discussed in the book’s first section, also plays a central role in the other sections, which deal with labor markets, open-economy macroeconomics, and methodology and future research. These questions are important especially in the issues of wage and price stickiness and exchange rate determination, but they appear even in the discussion of methodology.

The persistence of high unemployment in many economies for almost twenty years (since the first oil price shock in 1973) has stimulated a good deal of research into labor markets. This persistence is difficult to explain without the assumption of extreme inflexibility in wages. That is, to make empirical fact consistent with the market-clearing assumption of the new classical approach requires introducing some other impediment. Dale Mortensen reviews research on labor market hysteresis (the lack of a unique natural rate of unemployment) as an explanation for unemployment. This line of research provides several explanations of multiple labor market equilibria, and Mortensen concludes that they have some of the same policy implications—that government action may be necessary to move to a higher employment equilibrium. Interestingly, in these models the market-clearing assumption leads to a policy prescription that may be inconsistent with the new classical conclusions about policy’s irrelevance.

In international or open-economy macroeconomics the key issue has been the inability to explain, even in a theoretical setting, the movements in exchange rates since the early 1970s. Because both the exchange rate system and macroeconomic performance in general changed so dramatically at that time, issues of exchange rate determination and macroeconomic policy are closely linked. In his essay Rudiger Dornbusch reviews attempts to explain exchange rate determination in industrialized countries and contends that real exchange rates are determined over the long run primarily by the fundamental issue of resource allocation (a microeconomic phenomenon); he argues that this explanation is largely self-evident. Over the short run, Dornbusch contends, a number of competing approaches, distinguished by how binding their microfoundations are, can explain some episodes of exchange rate variability, but none has proved robust to generalization.

The era reviewed by most of the essays in The State of Macroeconomics is one during which new theory developed rapidly and virtuoso theory-making far outshone the empirical work produced to support it. Consequently, the work did not support dogmatic positions on policy issues, and policymakers continue to struggle with a multiplicity of intellectual foundations on which to base policy analysis. Not surprisingly, divergent views of the world support widely different policy prescriptions even when there is no dispute about the facts. However, the rigorous approach taken by the book’s various contributors and the willingness of even those with strong vested interests in the old or new regime to recognize the value of the opposition suggest that some macroeconomic issues will be resolved even as new ones emerge.

Notes
