David A. Dickey

David A. Dickey is a professor of statistics at North Carolina State University.

Commentary

FIRST, LET ME EXPRESS my appreciation for the invitation to participate in this conference. I have made several visits to the Federal Reserve Bank of St. Louis and have enjoyed the hospitality of Ted and his associates. Carl Christ's paper was interesting and thoughtful, prompting us to look again at some philosophical issues in econometric modeling.

Trying to describe an ideal econometric model makes sense to me. When I am in the market for a car, camera or other piece of technological equipment, I often look at the top-of-the-line item to see what it can do and then decide which features I can give up to make my purchase affordable. Carl Christ has done the same sort of shopping for an econometric model, searching for the best of all possible models. We likely cannot afford it, in the sense that we cannot really afford to formulate a model now and go fishing for several years while test data accumulates. Nevertheless, looking at the top-of-the-line type of model will let us see an upper bound on what we can expect models to give us and will give us a target point to move toward even though we have no hope of actually hitting the target.

Researchers see some of the same statistical strengths and weaknesses of econometrics when they apply statistics to the biological and physical sciences. In both sciences you must decide which independent variables are of interest. Often these are control variables like fertilizer, water, insecticides or in our case, interest rates. In actual agricultural practice, insecticide and water are often applied in response to observations on the state of the growing plants. Similarly in economics, it is often hard to tell if a control variable, the Aaa bond rate, for example, is a response to observations on the economy or a driver of them. Agronomists perform greenhouse experiments in which they fertilize plants in amounts long and short of the perceived optimum to map out a response curve for yield as a function of fertilizer. In contrast, economists are reluctant to experiment by (knowingly) setting control variables at nonoptimal values.

It is well known in agriculture that greenhouse results often do not translate directly to the field, so agronomists, like econometricians, distinguish micro from macro environments. Biologists also typically know the lag relationships, if any, involved in their experiments. Yield in August may be related to fertilizer application in June, but when do we finish harvesting the effects of a bank closure or a tax increase? Biological organisms in the field and the economy respond to a great number of inputs and a big decision is which to put into the model and which to leave as part of the error term.

An aspect of model choice that Christ does not particularly stress is the choice of model form. This is sometimes chosen to fit the data at hand well and so can be part of a data mining operation. Many physical models, as well as econometric models, are not linear. Einstein's famous $E=MC^2$ is an example. In economics, the well-known $MV=PY$ can be made linear by taking logarithms. Such transformations have implications for variance on the original scale—a point I think is not often appreciated. If $\log(M) = \log(P) + \log(Y) - \log(V) + e$ with e normal, then $MV = PY[\exp(e)]$ and therefore the error is multiplicative, causing heterogeneity of variance in the untransformed data. Further, nonlinear models like $MV=PY$ pose problems of aggregation. For example, suppose such a relationship holds in all segments of an economy. Will it then hold in the aggregate? Not necessarily. To illustrate, note that

$$(20)(2) = (4)(10)$$

and

$$(12)(6) = (18)(4).$$

However, if we average 20 and 12, average 2 and 6, average 4 and 18, and average 10 and 4, we find that $(16)(4)=64$, but that $(11)(7)=77$. So apart from any estimation errors, even exact relationships can hold on some scales but not on others.

Despite all these potential problems, people have an inherent tendency to observe their environment and draw inferences. There seems to be an optimism that with enough information we can solve any of our problems, regardless of whether they are economic problems, medical problems or other kinds of problems. Attempts at problem solving will certainly persist, and analysis and criticism of these attempts, such as Christ's, are worthwhile activities. In fact, I think one of his main points is that we are all statisticians, observing our world and modifying our models based on the data. This may be done with or without numerical calculation. Model selection is influenced by our previous observations in a way that is hard to quantify.

I found the Mitchell quote from 1927 somewhat offensive. The idea that with enough calculations, any two series can be found correlated at 90 percent surely cannot be true of informed and careful statisticians and econometricians. Nevertheless, I can agree with the nature, if not the extent, of the problem. I can imagine someone noticing how a black cat had crossed his path on several occasions before a misfortune, thus giving birth to a superstition.

Surely, however, the past must be somewhat like the future. Living in North Carolina, for example, I do not carry earthquake insurance, but I might if I lived in San Francisco. I suspect that early mankind anticipated being cold in winter even without a good understanding of meteorology. I do not think we should dismiss modeling as a whole based on Lucas-critique types of considerations. Christ gives an example of a simple model that seems to have held up over a fairly long period. This is good news and I would go further to suggest that we not give up on statistical modeling even if we can't get quite such good results every time. Along these lines, I agree with Christ that ARIMA and VAR are not as informative as a good econometric model, but they may do less damage to our understanding of the economy than a mediocre econometric model.

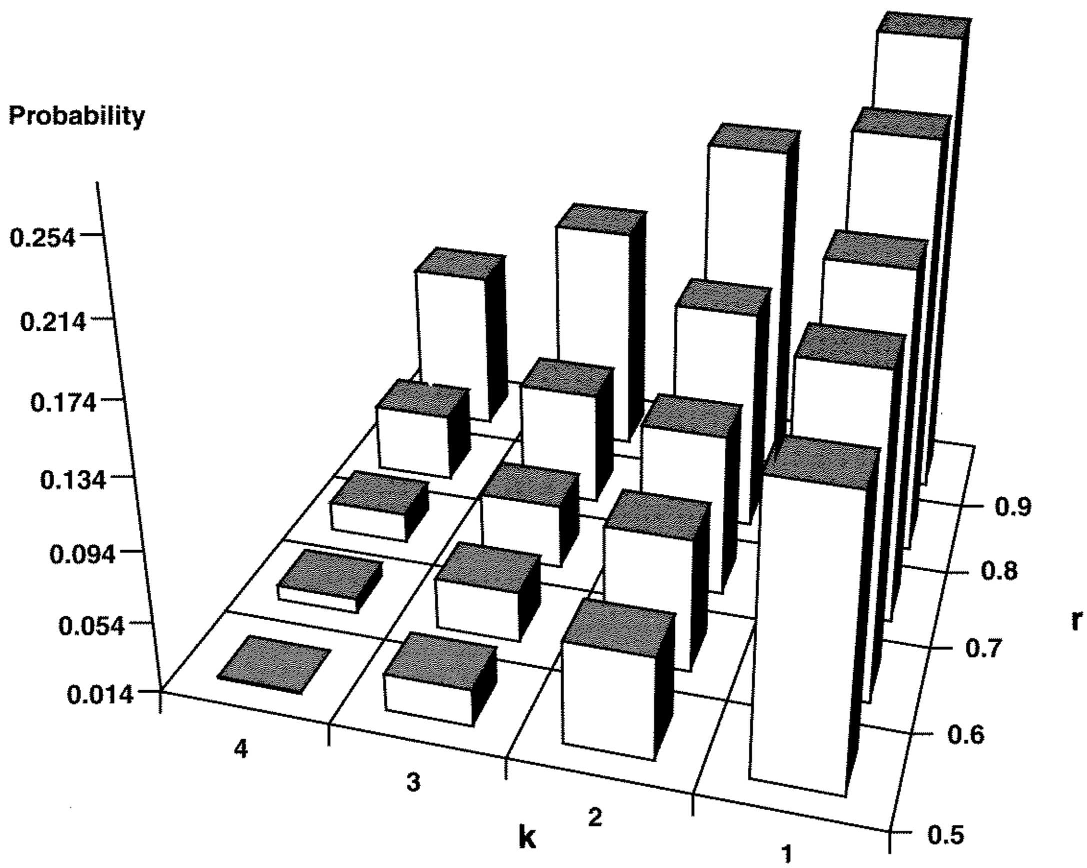
As a technical person, I feel obliged to address at least one or two technical points. I note that in the paper, some time was spent trying to decide whether a quadrupling RMSE would be reasonable in a good forecasting model. When we look at the theoretical forecast error variances, we can argue that this variance could not increase by more than a factor of eight in going from a one-step-ahead to an eight-step-ahead forecast. To compensate for the difference between estimated and theoretical MSEs, it is then concluded that if the estimated error mean square goes up by a factor of 16 (RMSE up by a factor of four) our model would be suspect. The probability of this quadrupling of sample RMSE would depend on the autocorrelation and the number of forecasts used to estimate RMSE; for example, if we just look at a single one-step-ahead residual and a single eight-step-ahead residual, the estimated RMSEs will simply be the ratio of the absolute errors and hence will vary a lot around the true values.

Suppose MSE is calculated by averaging the squares of k independent one-step-ahead errors $e(n+1)$ and the squares of the k corresponding eight-step-ahead errors

$$z(n+8) = e(n+8) + r\,e(n+7) + r^2\,e(n+6) + \dots + r^7\,e(n+1)$$

from an AR(1) with autoregressive parameter r. I estimated the probability that the sum of k values $z(n+8)^2$ is more than 16 times the sum of the k corresponding values $e(n+1)^2$ by a Monte Carlo experiment with 10,000 replicates at each r and k. Figure 1 summarizes the results with r = 0.5, 0.6, 0.7, 0.8 and 0.9. The number of forecasts

Figure 1
Probability of Quadrupling
(versus number of squares in RMSE and first order autocorrelations)



from which RMSE is calculated is k = 1, 2, 3 or 4. It is seen that, because of the variation in RMSE around its theoretical expected value, the probability of the eight-step-ahead RMSE exceeding four times the one-step-ahead RMSE can be reasonably large (greater than 0.2 in the case that k = 1) even with a perfect model and relatively mild autocorrelation. As k gets large, and hence as RMSE converges to the theoretical value discussed in the paper, the probability declines. Figure 1 shows the empirical frequency of RMSE quadrupling.

As another minor technical point, I would like to say that a lot of new ideas are the same old vanilla ones with a few sprinkles thrown on. In his figure 3, Christ plots the inverse velocity against the inverse Aaa bond rate data with connecting lines indicating the time order of the data and with the regression line overlaid. We

Figure 2
Aaa Bond Data

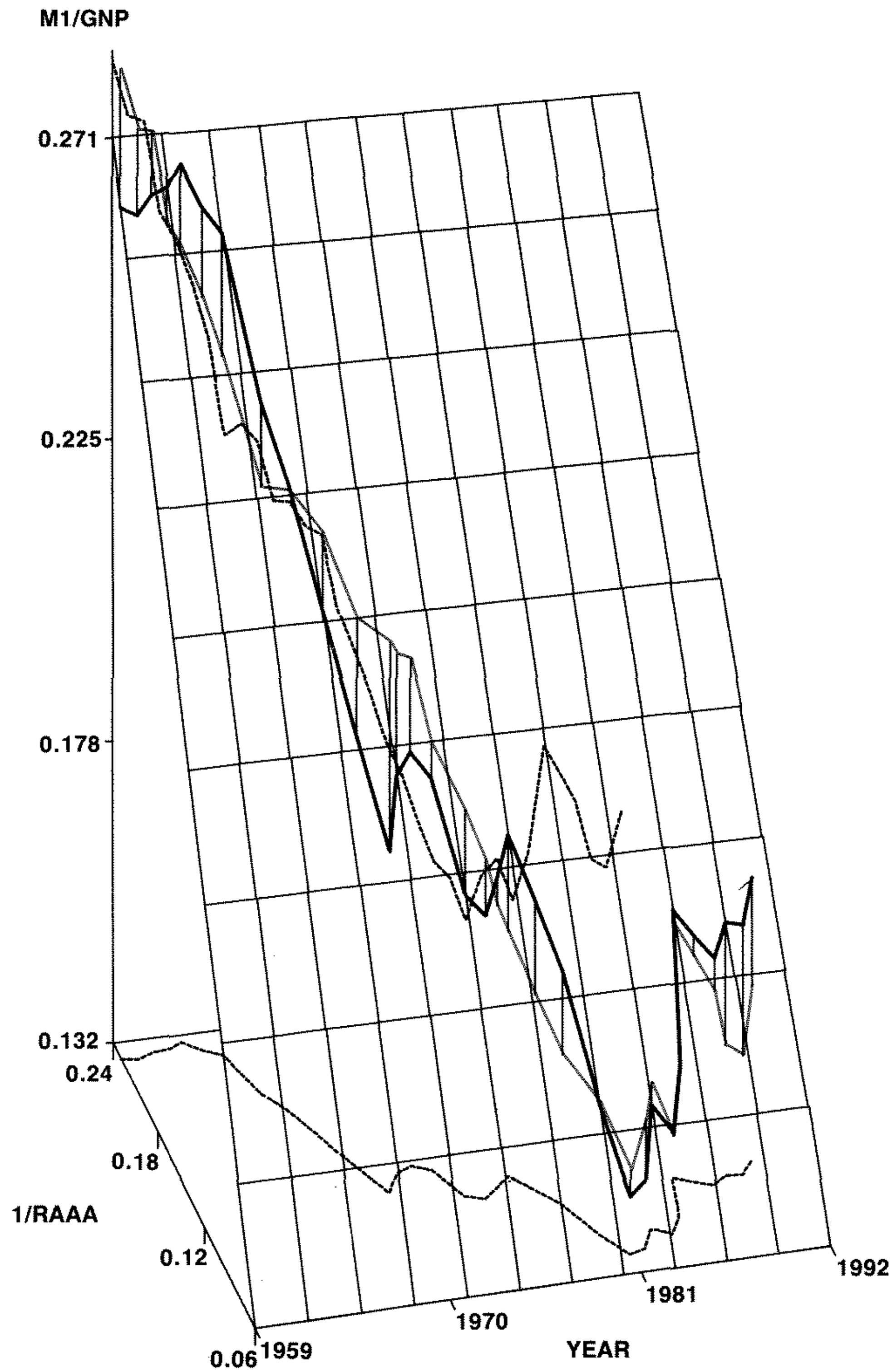
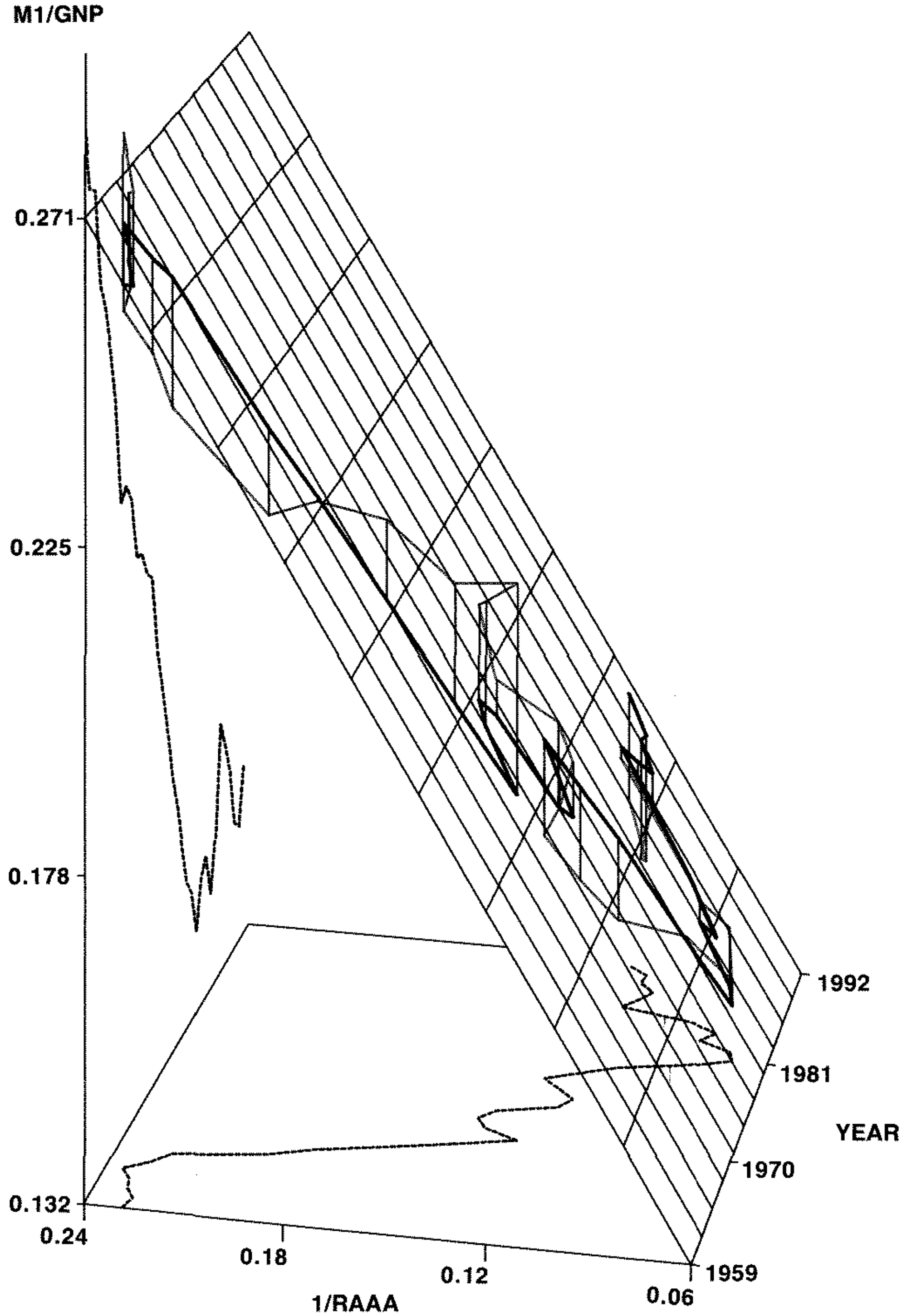


Figure 3
Aaa Bond Data



could think of this as the end view of a three-dimensional picture, which we rotate and tilt a bit in figures 2 and 3. The line is seen as the end of a three-dimensional plane. The data wander pretty far up and down and right and left but never get too far from the plane. Projections into the wall and floor of the plot show the two nonstationary looking series, also depicted in Christ's figure 2. The tightness of the data about the plane shows that a linear combination of the two series looks fairly stationary. This is the idea of cointegration. Regression is one way of finding cointegration in bivariate series. Other methods may give slightly different planes, but we can see that the main ideas of this currently popular econometric method are quite close to simpler time-tested ones.

In closing, I think we are at an exciting time for econometrics. Some of the computational burdens have been lifted, and we can concentrate more on proper model forms and formulation methods. Philosophical guidance such as that offered by Christ is important to keep in mind in our search.