Discussion Paper 22

Institute for Empirical Macroeconomics
Federal Reserve Bank of Minneapolis
250 Marquette Avenue
Minneapolis, Minnesota 55480

December 1989

MODELING TRENDS

Christopher A. Sims
Institute for Empirical Macroeconomics
and University of Minnesota

ABSTRACT

Models of low-frequency behavior of time series may have strongly conflicting substantive implications while fitting the data nearly equally well. We should develop methods which display the resulting uncertainty rather than adopt modeling conventions which hide it. One step toward this goal may be to consider "overparameterized" stationary ARMA models.

KEY WORDS: Nonstationarity, unit roots, trends

This material is based on work supported by the National Science Foundation under Grant No. SES-8722451. The Government has certain rights to this material.

MODELING TRENDS

Macroeconomists have recently been exploring what the time series data can tell us about long run behavior of economic variables. There is, I think, a consensus emerging that statistical models which are similar in their ability to fit the long run characteristics of the data may differ substantially in their behavioral interpretations. (Recent examples exploring both these conclusions are Campbell and Mankiw [1987], Christiano and Eichenbaum [1989], and Cochrane [1988].) This paper develops the implications of this consensus for modeling strategies and also points out a class of models for trending time series which may deserve wider use.

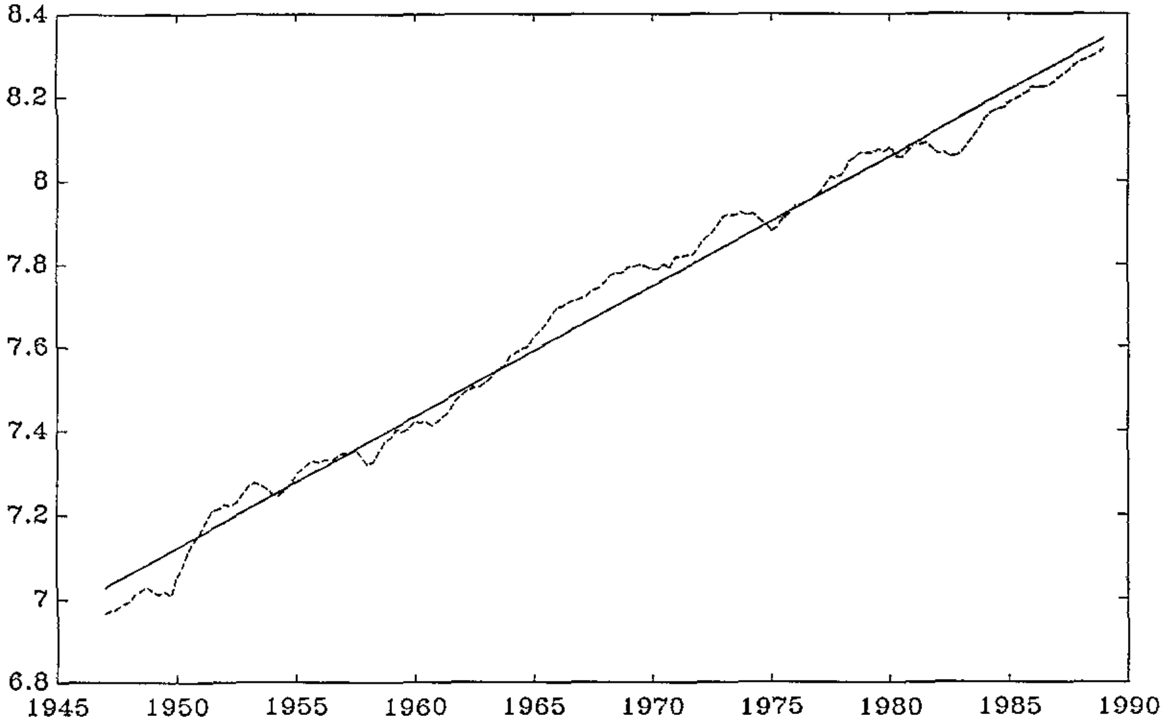
I. Common Sense

The most commonly used models for economic time series showing persistence are all special cases of general ARMA specifications with exogenous polynomial trends. The two leading special cases are the unit root and the trend-stationary models. A unit root model has its variance growing at a polynomial rate for large values of its time argument t and may also have a deterministic polynomial "drift" component in its mean. The trend-stationary model has a polynomial in time as its mean, but variance converging to a constant for large t. Most practical modelers have paid little attention to a third special case, the stationary subclass of ARMA models, for time series showing an "obvious" trendlike behavior. In any case, if the series at hand seems to have a trend when plotted, usually it is easy to construct a test statistic which rejects stationarity against an alternative with polynomial trend or accepts a unit root null hypothesis at conventional significance levels.

Postwar quarterly data on real GNP spans about 170 observations, over which its log has grown at a 3.2% annual rate. If we plot it against a linear trend, there is no obvious tendency for it to be flying away from the trend line at the ends of the sample. (See Figure 1.) Common sense tells us that it should be very hard to reject a model with a linear trend of 3.2% per year in the log of real GNP. The fluctuations of real GNP around its trend line change sign more often than would be very likely if they themselves represented a random walk, but a random walk does fluctuate around its initial level for a while with some nontrivial probability. Furthermore, the deviations of a random walk from a fitted linear trend line necessarily fluctuate in sign. Figure 2 shows the fluctuations of GNP around its trend line and the fluctuations of a simulated Gaussian random walk about its fitted linear trend line. Common sense tells us that it will be hard to be confident that GNP's deviations from a linear trend are not close to a random walk.

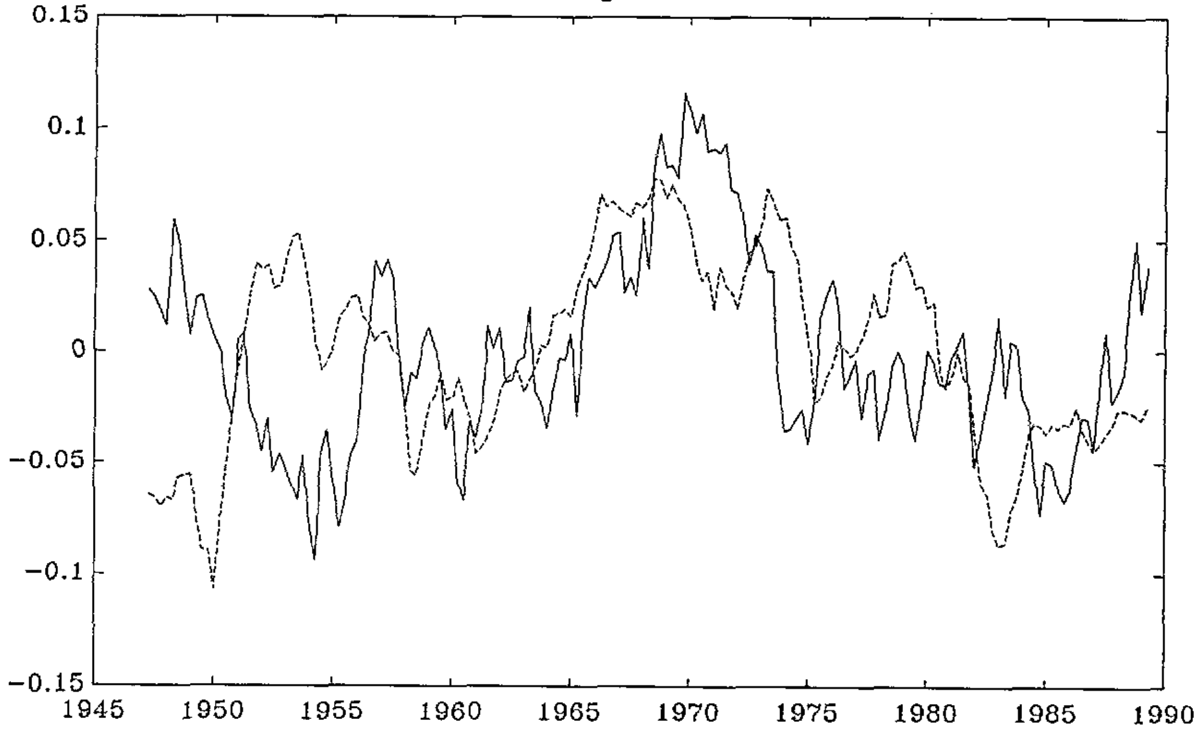
But take another look at Figure 1. GNP is below its trend line at the start and the end of the sample and stays above the trend line for a long stretch toward the middle of the sample. Isn't this what we would expect if there were a tendency of the growth rate to drop over the period? Indeed, in current discussions of prospects for the economy in the U.S. and the world generally, the apparent slacking off in the "trend" growth rate of output is a major concern. Most economists would agree that there is a definite possibility that the trend growth rate has dropped permanently, or at least for the next several decades. No ARMA model which includes trend (and this includes unit-root ARMA models which contain a positive constant term) is consistent with the possibility of a permanent drop in the growth rate.

Figure 1



Log of Real GNP

Figure 2



Deviations from Trend: GNP and Random Walk

We can fit a model with a bounded mean function to the GNP series and obtain a better fit than with a linear trend model. For example, one can fit an equation of the form

$$Y(t) = \alpha + \beta\cos(\gamma t+\phi) \tag{1}$$

with greek letters treated as free parameters and Y representing log of real GNP. My calculations show that a choice of α=2.11, β=6.68, γ=.00215, and ϕ=-.76 gives a trend line which fits substantially better than a linear trend, though when it is plotted over the sample its deviation from linearity is hard to see. Of course this model of trend implies that we are nearing a peak of the level of GNP, which will arrive about the year 2044 and be followed by over 700 years of mostly negative growth.

This latter model of deterministic trend is of course not to be taken seriously. However, in my view, it is no more ridiculous than the linear trend model, which ought also not to be taken seriously. The problem with both models is their deterministic treatment of low-frequency components of the process. It seems apparent that we could construct models in which there is a stochastic trend component whose sample paths look like those emerging from deterministic trend models, yet which cannot be extrapolated with certainty. Wouldn't this be more reasonable?

II. Distinguishing Among Models of Persistence

We can model persistence with ARMA processes with unit roots, ARMA processes with explosive roots, processes with deterministic polynomial or exponential trends, fractionally integrated processes, and covariance stationary ARMA processes with high power at low frequencies, among other possibilities. This section presents a number of results emphasizing the artificiality of most attempts to get finite spans of data to distinguish among these models.

Finite parameterizations of these models are matters of convenience; we ordinarily expect that the orders of ARMA processes and the degrees of polynomial trends will be adjusted in interaction with the data. Once we recognize this, so that the order of parameterization is treated as data-dependent, we have to recognize that we should not expect to distinguish the models by looking at the data.

It is known, but perhaps not widely appreciated, that in a certain well-defined sense ARMA models with explosive autoregressive roots are equivalent to models with deterministic exponential trends. To take the simplest example, suppose

$$Y(t) = \rho Y(t-1) + \varepsilon(t) \tag{2}$$

with $\rho>1$ and $E_t[\varepsilon(t+1)]=0$, where E_t means expectation conditional on all $Y(s)$ for $s\leq t$. This model implies that

$$Y(t) = -\rho^{-1}(1-\rho^{-1}L^{-1})^{-1}\varepsilon(t+1) \quad + \lim_{s\to\infty} Y(t+s)\rho^{-s} \ . \tag{3}$$

If the ε's all have finite variance, the limit appearing on the right of (3) exists, not as a constant but as a random variable. To see why, observe that

$$Y(t) = \rho^t Y(0) + \sum_{s=0}^{t-1} \varepsilon(t-s)\rho^s$$

$$= \rho^t\left(Y(0)+ \sum_{s=1}^{\infty}\varepsilon(s)\rho^{-s}\right) - \sum_{s=1}^{\infty}\varepsilon(t+s)\rho^{-s} \tag{4}$$

Assuming $Y(0)$ and the ε's are jointly normally distributed, the two terms on the right of (4) are jointly normally distributed. Furthermore, if we call the second term on the right of (4) $z(t)$, it is clear that $z(t)$, $t=1,\ldots,\infty$, has the distribution of a draw from a stationary AR process with parameter ρ^{-1}. Also, because $Y(0)$ is uncorrelated with $\varepsilon(t)$ for $t\geq 1$, the first term on the right of (4), which we will call Q, has a non-degenerate normal distribution conditional on $\{z_s$, all $s\geq 1\}$. Thus we can think of generating a particular realized path for $Y(t)$, $t=0,\ldots,\infty$, by first drawing a realized path of a Gaussian AR(1) process with parameter ρ^{-1}, treating that as $\{z(t), t=0,\ldots,\infty\}$, then drawing a realization of Q from the appropriate normal distribution $Q|\{z(t), t=0,\ldots,\infty\}$, then applying (4) to generate $Y(t)$, $t=0,\ldots,\infty$.

Suppose our model was instead

$$Y(t) = A\rho^{t} + \rho^{-1}Y(t-1) + \eta(t) \ . \tag{5}$$

Here $E_t[\eta(t+1)]=0$, all t, as usual. If we do not know A with certainty, then even if we know ρ with certainty, there is no way to use a single realized time path for Y to distinguish (5) from (2). Another way to put it is that if we treat A as a random variable jointly normal with the i.i.d. η's, the probability measure on the space of paths implied by (5) is equivalent to the probability measure implied by (2). "Equivalence" here is in the technical measure-theoretic sense that no class of paths with nonzero probability under one measure has zero probability under the other. This means that there is no way to define a decision procedure which picks the true data generating process with probability one using an observed time series of unbounded length.

It might be argued that the symmetry between the exponent of the trend in (5) and the inverse of the root of the AR component is special, so that it will be easy to distinguish this model from general models containing trends at rates not exactly matching inverses of AR roots. But this comforting conclusion is available only if we treat the order of parameterization as given a priori. Individual roots of stationary ARMA processes are not identified when we recognize that the order of the process is not given a priori. In particular, we have the result

Proposition 1: Given an arbitrary root outside the unit circle, an arbitrary covariance-stationary process can be approximated arbitrarily well, in the sense of one-step-ahead mean square linear prediction error, by finite-order AR processes whose autoregressive operator has the given root.

Proof: Let the process be Y and the prespecified root be R>1. We know that the minimum one-step-ahead mean square prediction error linear forecast for Y(t) can be approximated arbitrarily well as b(L)Y(t) for b(L) a finite order polynomial in positive powers of L. Let $c(L)=(1-R^{-1}L)^{-1}b(L)$. This will generally be a polynomial in L of infinite order, but we can approximate c(L) arbitrarily well (in ℓ_1 or ℓ_2 norm) by truncating c(L) at some finite power of L, calling the result $\hat{c}(L)$. Then since $c(L)(1-R^{-1}L)=b(L)$, $\hat{c}(L)(1-R^{-1}L)$ approximates b(L) arbitrarily well -- and of course it has R as a root. Approximation of b in ℓ_2 is sufficient to guarantee approximation of b(L)Y(t) in the variance norm for stationary Y. q.e.d.

Proposition 1 means that a model's list of roots is not identified if we do not know a priori the order of model. The proposition can be extended -- the process need not be stationary if it is difference-stationary of some order; a finite list of roots can be prespecified instead of a single root.

Proposition 1 together with our previous discussion about (2) and (5) implies that we should not take comfort in the requirement that roots in the trend and stationary components in (5) match up. In practice, we will be able to find finite order models in which the roots do match up and which fit arbitrarily well, even if in the true model they do not match up. Thus models with exponentially explosive, deterministic trend are essentially indistinguishable from models with no deterministic trend but with explosive autoregressive roots.

It is possible to distinguish a model with exponentially explosive deterministic trend (or, equivalently, with an explosive AR root) from the class of models with polynomial trends and roots on or outside the unit circle. There is no analogue to Proposition 1 if the prespecified root is strictly inside the unit circle. Thus expected prediction error from stationary ARMA models, no matter how high their orders, remain worse by an amount bounded away from zero than those from a true model with an explosive root. Another expression of essentially the same mathematical fact is that it is impossible to uniformly approximate e^t on the interval $[0,T]$ by polynomials of degree Q without letting the degree of Q grow linearly in T. Thus in a large enough sample exponential trend behavior will be clearly distinguished from polynomial trend behavior. On the other hand, an arbitrary stationary ARMA model can be approximated arbitrary well by models containing at least one explosive root, so long as that root can be chosen arbitrarily close to the unit circle. Correspondingly, polynomial trends can be approximated well by linear combinations of exponentials, so long as we are allowed to choose the exponentials to have arbitrarily slow rates of growth.

Proposition 2: Models with stationary fluctuations about deterministic polynomial trends can be approximated arbitrarily well by stationary finite order AR models without constant terms.

Proof: Suppose the stochastic process Y has polynomial component P(t) of order m, so

$$Y(t) = P(t) + Z(t) \tag{6}$$

with Z stationary and linearly regular (meaning that it contains no linearly deterministic component). If we difference Y m+1 times we extinguish the P component (including the constant term) and are left with $Z^*(t)=\Delta^m Y(t)=\Delta^m Z(t)$. This process is itself stationary, and its innovation (error from best one-step-ahead linear prediction) is the same as that of Y. Since it is linearly regular, Z^* has the property that $\hat{Z}_k(t)=E[Z^*(t)|Z^*(t-s),\ s=1,..,k]$ converges in mean square to $E[Z^*(t)|Z^*(t-s),\ \text{all } s\geq 1]$ as $k\to\infty$. But $\hat{Z}_k(t)$ is just the least squares estimate of a k'th order autoregressive predictor for $Z^*(t)$. It will be a finite linear combination of past Z^*'s, thus of past Y's, and will achieve arbitrarily close to optimal predictions of Z^*, and thus of Y. q.e.d.

Proposition 2 could be extended to cover the case of true models which are difference-stationary processes plus polynomial trend. All that is required is that in the first step we difference by the larger of the order of P or the order of integration in Z. The purely autoregressive predictors constructed in the proof of Proposition 2 will generally involve long averages of past data which in effect reconstruct the polynomial component of Y from Z^*.

III. An Interesting Class of Models

The results of the preceding section suggest that, if we are interested mainly in models without strongly exponentially explosive deterministic components, we could stay entirely within the class of stationary AR models, even stationary AR models without constant terms, without paying any permanent penalty in fit or predictive power. In practice, though, econometricians have tended to abandon such models when data seem to show trend-like behavior. It seems to me we have done so more out of a desire for computational convenience than out of careful consideration of the models' implications.

To stay in the stationary class of models in the face of trend-like behavior in the data requires that we model components which evolve very slowly relative to the span of our data. When we allow for such components, we have to recognize that the data are likely to be only weakly informative about their nature. It is important in applications of a model not to allow aspects of it about which the data has little to say to be set essentially arbitrarily by the data. Usually in non-Bayesian approaches, and often even in Bayesian approaches, this is accomplished by making arbitrary choices of certain parameters. In the context at hand, it is often done by differencing the data, preliminary regression on a polynomial in t, or initial prefiltering with a high-pass filter other than a simple difference operator. Nonetheless it is possible in principle to leave in the model parameters describing the low-frequency characteristics of the data about which the data are weakly informative. This will generate larger implied standard errors on estimates and numerical problems, just as in the similar case of near collinearity in regressors. But just as in that case, it may often be better scientific practice to expose the uncertainty actually present than to resolve it by an arbitrary restriction.

Let us consider two examples, one theoretical and one applied, to see the difficulties and possibilities of stationary modeling of trending series. Suppose the data contained a first-order AR component with parameter .8, innovation standard deviation .006, and a component, independent of the first, which moved slowly and smoothly enough to show a fairly steady growth rate over 169 periods (a postwar quarterly sample). If the latter had the form

$$(1-\lambda L)^2 Z(t) = \varepsilon(t) \tag{7}$$

we could get a standard error of about .002 on ε and a correlation of of .4 between growth rates of Z separated by 169 quarters by choosing a λ of about .998. Adding on the first-order AR component gives an overall MA representation for the process of

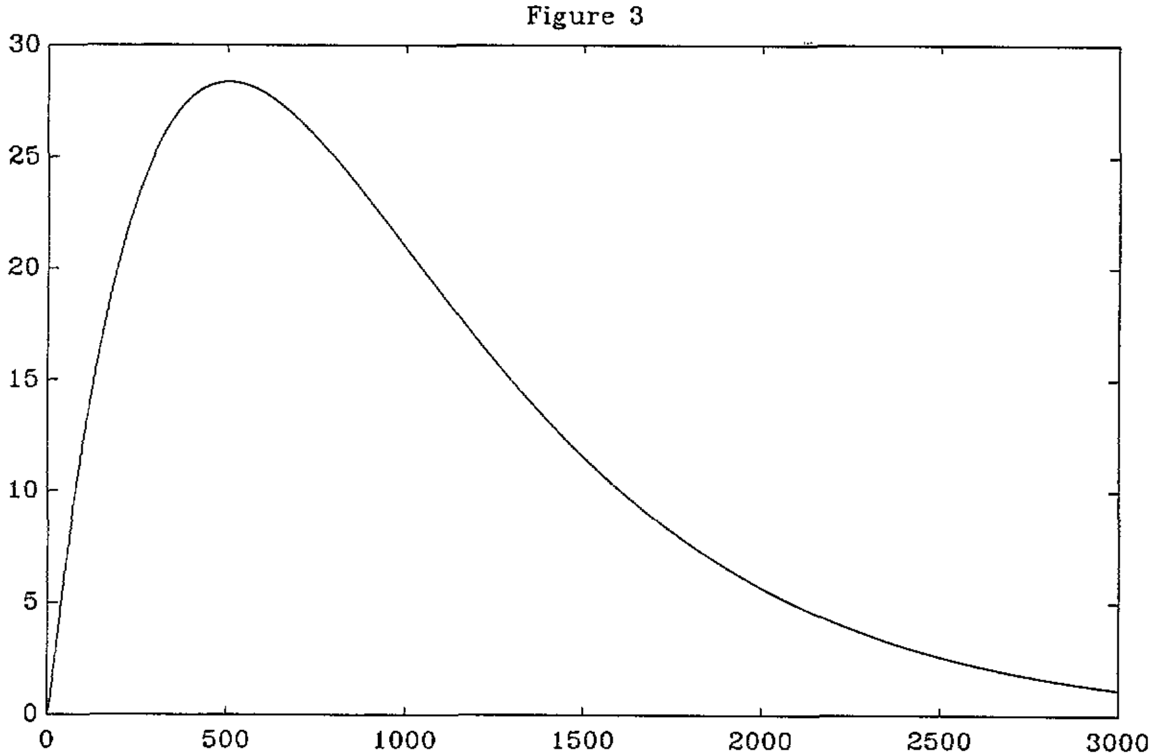
$$Y(t) = \frac{1-.5712L+.624L^2}{1-2.798L+2.593L^2-.797} \quad . \tag{8}$$

The numerator has imaginary roots of .786±.083i, which are close to the denominator root of .8. Usual good practice for constructing ARMA forecasting models would suggest reducing the order of numerator and denominator, therefore. However, this near-cancellation can be expected to occur when we allow low frequency components of the process to be freely parameterized.

Figure 3 shows the impulse response of the model (8). The response is very slowly decaying, and implies forecast errors increase sharply with increased forecast horizon.

IV. Log GNP, For Example

Since it has been so much studied, it is interesting to see the results from attempting a stationary ARMA model of log GNP. The results reported here all use the RATS program to estimate ARMA models, which means that a conditional likelihood is maximized rather than the the unconditional likelihood which would be usable in principle for the stationary models. At the edge of the unit



Impulse Response, Example Model

circle it is not obvious that using the marginal distribution of initial conditions implied by the model, as the unconditional likelihood does, is appropriate. Matrix singularity problems would create difficulties with these near-unit-circle models in an unconditional likelihood calculation. The RATS approach instead develops convergence problems. It is an interesting fact that generating forecasts from the ARMA difference equations conditional on initial zero values for disturbances often leads to convergence to global best fits with MA roots inside the unit circle. These anomalous fits are substantially better than can be obtained with models which stay on the proper side of the unit circle, and they do represent legitimate forecasting formulas for future data as functions of past data. However, they imply that one-step-ahead forecasts of Y(t) have a coefficient on Y(0) which increases with t. This makes it implausible that their in-sample performance will persist much outside the sample and justifies ignoring them. Avoiding them required frequent restarts of the RATS algorithm.

The Table reports the fitted versions of four equations: ARMA(3,3) with constant; ARMA(3,3) with trend; ARMA(3,3) with a $\cos(.002t+\theta)$ trend term; and ARMA(2,3) fit to differenced data. The last of these is reported with the AR operator multiplied by 1-L for comparability with the other three models. The best fit is the model with trigonometric trend, though its margin of advantage over the linear model is very small.

A *t* test for the restriction that the sum of coefficients on lagged Y in the first column is one, which tests the null hypothesis of a unit root, i.e. of the column 4 model, yields a value of 1.204 and a marginal significance level of .23. Thus a differenced model is very close in fit to the model on levels data. Nonetheless, as can be seen in the figures for the long term forecast growth rates, the implication of the models are substantively different. The differenced model (because it converts the constant into an implied linear trend) forecasts more persistent growth, averaging 3.3% over 8 years compared to 2.2% for the stationary model. From the forecast standard errors in the table one can see that the standard error on the annualized 31-quarter growth rate is approximately 1%. Thus the difference between these two models in forecast long-term growth rates is only about one standard error. Nonetheless for practical purposes a report that the growth rate will be about 3.3% for the next 8 years, with only about one sixth probability of being less than 2.3%, is quite different from one that it will be about 2.2%, with only about one sixth probability of being above 3.2%. To arrive at a reasonable conclusion from these two models one must assess their relative a priori plausibility and weight together the results appropriately.

The *t* test for the null hypothesis of 0 coefficient on t in column 2, which tests column 1 against column 2, is 2.26, for a marginal significance level of .024. While as usual conventional significance levels are an unreliable guide, here the usual Bayesian rule of thumb, the Schwarz criterion, is also unreliable because of the failure of all the coefficients to be converging at the usual $T^{-.5}$ rate. A Bayesian assessment of the evidence must consider how concentrated the likelihood function is relative to our prior on the coefficient for t. Furthermore, it is unreasonable to suppose our priors on the coefficient of t and on the coefficients on lagged Y's and lagged ε's are independent. My informal experiments with priors suggest that, unless prior uncertainty about the trend coefficients is very great (say an initial uniform prior on the range -6% to +12% for the trend rate of growth), the observed data makes the models with trend a posteriori more likely than they were a priori. On the other hand, if the prior probability on the stationary model is not small, the evidence is not strong enough to make its posterior probability negligible. Here the case that there are substantively important differences among the models is even stronger. The trend models differ among themselves on the forecast growth rate by almost as much as the stationary models and also make the implied standard error of the forecast much smaller, so that the differences in implied distributions for the growth rate are sharper.

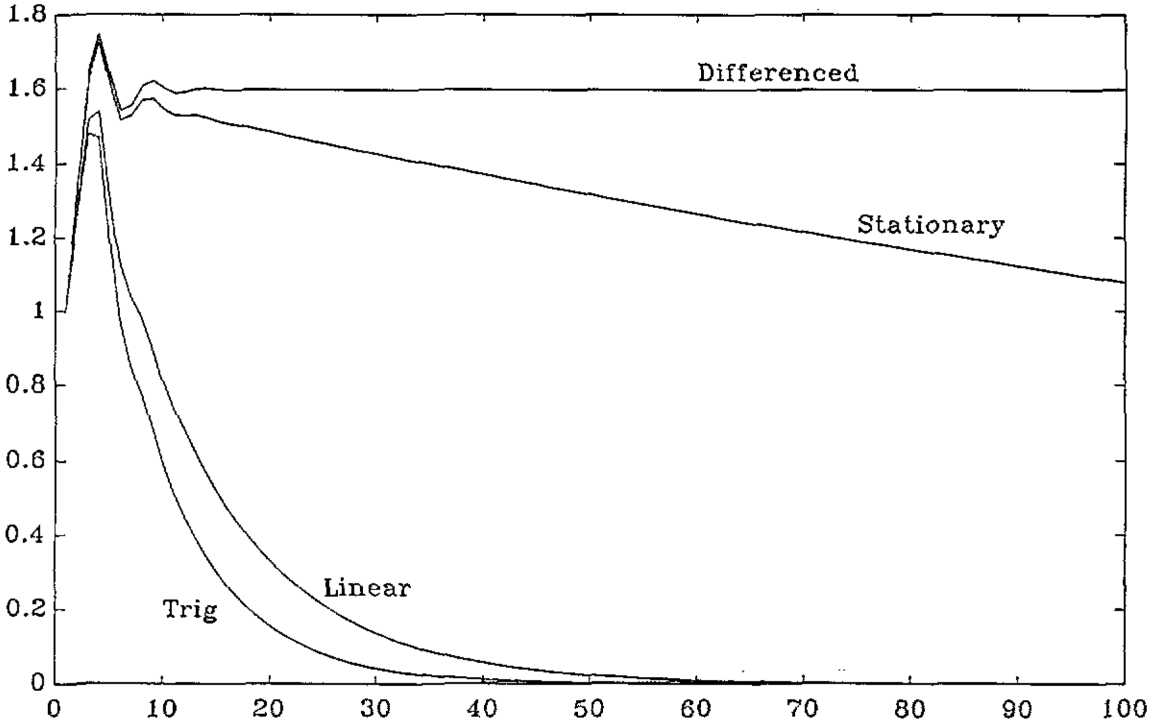
Figures 4-5 show the impulse responses for the three estimated models. Observe that the time scale for the stationary model is completely different from that for the others. The decay of the response to shock is fastest for the best-fitting model, the trigonometric trend model. The lower part of the Table reports implied forecast standard errors at horizons up to 4 years. At longer horizons, the stationary model implies much larger forecast errors than do the other two models. Also reported in the Table are forecast average annual growth rates over the 31 quarters following 1989:1. The trigonometric trend model and the stationary model forecast 2.1% and 2.2% rates, while the linear trend model forecasts a 3.1% rate.

V. Conclusions

The preceding section's applied example illustrates the main points of this paper. We can fit the data well with a variety of approaches to modeling low frequencies. Stationary models, which are not a priori implausible, can fit roughly as well as models with non-stationary trends. There are large differences in practical implications among models with similar fits.

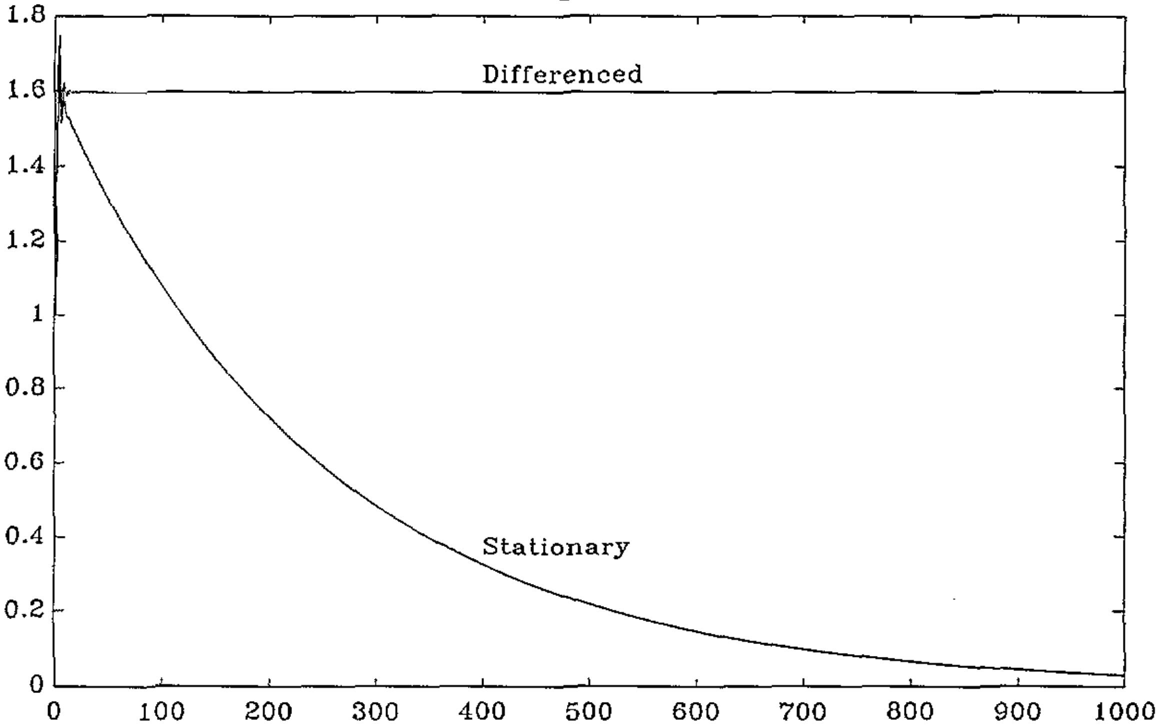
Putting these individual conclusions together suggests that

Figure 4



Impulse Responses, First 25 years

Figure 5



Impulse Responses, First 250 Years

econometricians should be paying more attention to giving an honest account of the uncertainty about long run implications of their statistical models.

REFERENCES

Campbell, J.Y. and N.G. Mankiw [1987], "Are Output Fluctuations Transitory?" Quarterly Journal of Economics 102, 857-80.

Christiano, L. J. and M. Eichenbaum [1989], "Unit Roots in Real GNP: Do We Know and Do We Care?" Discussion paper no. 18, Institute for Empirical Macroeconomics, Minneapolis.

Cochrane, J.H. [1988], "How Big is the Random Walk in GNP?" Journal of Political Economy 96, 893-920.

TABLE

Coeff. on Vbl. Lag		Pure ARMA	Lin. Trend	Trig. Trend	Diff.'d Data
const.		.0431	.684	.437	.00883
		(.0288)	(.287)	(.358)	(.00255)
Y	1	1.355	1.160	1.141	1.408
		(.309)	(.322)	(.365)	(.311)
	2	−.826	−.588	−.562	−.889
		(.422)	(.439)	(.491)	
	3	.466	.332	.287	.481
		(.174)	(.191)	(.210)	(.171)
ε	1	−.042	.107	.111	−.085
		(.316)	(.326)	(.367)	(.318)
	2	.669	.633	.611	.668
		(.118)	(.130)	(.135)	(.115)
	3	.144	.194	.199	.131
		(.144)	(.142)	(.155)	(.147)
t			.000722		
			(.000319)		
cos(.002t)				.503	
				(.468)	
sin(.002t)				.605	
				(.296)	
SEE		.00999	.00980	.00978	.01000
Forecast Annual Growth 89:1-97:4		2.2%	3.1%	2.1%	3.3%
Std. Err. of Fcst by Horizon in Qtrs.					
1		.010	.010	.010	.010
4		.029	.026	.026	.029
8		.042	.034	.032	.043
16		.061	.039	.034	.063
32		.084	.040	.035	.089
Roots (root,period)					
AR		.996,∞	.914,∞	.874,∞	1.000, ∞
		.684, 4.81	.602, 4.60	.573, 4.70	.694, 4.94
MA		.847, 4.40	.827, 4.29	.817, 4.32	.847, 4.45
		.200, 2.00	.284, 2.00	.298, 2.00	.183, 2.0