Inference For Multivariate Time Series Models With Trend

Christopher A. Sims

Dept. of Economics

Yale University

37 Hillhouse Ave.

New Haven CT 06520-1962

Originally presented at the August 1992 ASA meetings.

Revised 7/96

Revised 10/96[*]

[*] Though much of the text discusses the figures, in this version of the paper they are not present, because the originals were not preserved in digital form. The truly diligent reader could reconstruct the figures from the coefficient estimates given. The plan is to update the figures before too long, in any case.

Large-sample asymptotic distribution theory is likely to be a poor approximation when samples are not large relative to the time scale of the estimated model's dynamics. As economists have become increasingly aware of this point, they have moved initially toward using asymptotic theory that assumes model dynamics separate cleanly into a non-stationary component (associated with the unit roots) and a stationary component. I have argued elsewhere that the result may be worse than simply using the standard distribution theory associated with stationary asymptotics. This is because the standard stationary asymptotics preserve the usual correspondence between classical confidence statements and p-values and characteristics of the shape of the likelihood function. This correspondence is distorted by use of the "correct" asymptotic theory.

On the other hand, as the Bayesian approach to inference in models with strong low-frequency components has gained the attention of more researchers and its advantages have been recognized, disputes have broken out over the appropriate choice of a prior distribution. Classical criteria for a prior representing "ignorance" or "objectivity" seem to suggest strongly non-flat priors for these models. The flat prior on dynamic model coefficients in most instances becomes strongly non-flat when we transform it consistently with a change in the time interval at which data is measured. The likelihood function itself, conditional on initial observations, is Gaussian in shape in linear Gaussian time series models, and in large samples it can be shown to have approximately this shape even for non-Gaussian disturbances and for unit-root model dynamics.[1] There seems to be no leading proposal for a non-flat prior that retains the transparency and computational simplicity of reporting the likelihood function itself in these models. Economists trying to draw prescriptions for practice from, e.g., the articles on these issues assembled in the 1991 special issue of the *Journal of Applied Econometrics*[2] might well find the material abstract and puzzling.

1. Initial Transients as Trends

In my contribution to the *Journal of Applied Econometrics* symposium, I pointed out that estimation methods that condition on initial values, treating them as carrying no information about model dynamics, tend to imply that in the first part of the sample the behavior of the data is dominated by a large “transient”. That is, the estimates imply that the initial data points are very far from the deterministic trend line or steady state, in the sense that the estimated model implies that future deviations as great as the initial deviation will be extremely rare. In that paper I displayed a graph, roughly reproduced here as Figure 1, of the actual data, the trend line, and the transient generated by initial conditions, for a linear AR model of the log of U.S. real GNP, including a trend term. One can see from the graph that over the first third of the sample the data stay below the trend line, with the initial deviation from trend much larger than occurs anywhere later in the sample. I argued that this type of result was probably typical in time series data, especially when polynomial trend terms are included in the model.

In this paper I expand this point and discuss its implications. Section 1 documents that the phenomenon is widespread. It shows that my earlier suggestion that it is more important the higher the order of polynomial trend terms in the model is incorrect -- it is in a certain sense

---

[1] See Kim, Jae-Young [1992]

[2] My own contribution to the discussion is Sims [1991].

stronger when such terms are not included. It also shows a tendency for the phenomenon to be stronger in multivariate models.

Section 2 discusses methodological implications of these patterns. It argues that these estimated models with large initial transients are seldom appealing models of the data. Though considerable attention has been devoted to distinguishing models conditional on initial data values with roots near the unit circle and no polynomial trend from models with polynomial trend (which typically are estimated to have roots farther from the unit circle), these models in fact tend to have similar implications and are usually unappealing for the same reasons. Section 3 discusses methods for focusing attention on a more plausible region of the parameter space, in which initial conditions and model dynamics are not in such sharp conflict. A Bayesian perspective seems to be the best way to approach this problem, and it remains difficult. However, when we recognize that there is no way to avoid taking a stance on how initial conditions are related to model dynamics, and that this is the most important dimension of indeterminacy in choosing a reasonable prior, the argument over how to choose a prior takes on a more concrete form. We must decide how to use the initial conditions, and there are several leading approaches that are relatively easily interpreted from this stance.

2. Graphical Examples of Large Estimated Initial Transients

If we fit a 5th-order linear univariate AR model to the log of U.S. real GDP[3], including a constant term, but no higher-order deterministic components, the estimates are

$$(1-1.3343L+.1845L^2+.2495L^3+.0075L^4-.1032L^5)Y(t)=.03764+e(t) \tag{1}$$

This equation has one real root of .9943 and two complex pairs of absolute value .68 and .48, respectively. The model with linear trend term underlying Figure 1 has maximal root .943. Thus it appears that by suppressing the linear trend term we find stronger support for stochastic trend, a non-deterministic drift in GDP rather than a pre-ordained monotone movement along a smooth path. But this impression is misleading, as shown in Figure 2.

With a root as close to one as .9943, the model implies that the variance of GDP is much larger than the variance of a one-period innovation, and indeed it implies a standard deviation for the level of GDP about its steady state of about .12, more than ten times larger than the estimated standard error of one-period forecast errors, which is about .01. But the actual data are many standard deviations away from the estimated steady-state value. The estimates model actual GNP as starting about 17 standard deviations from the steady-state in 1949, reaching a position only about 6 standard deviations away in 1991. The observed behavior of the data is dominated by the movement in the forecast path generated form the 1948 initial conditions, shown as the dotted line around which the actual path fluctuates. This deterministic forecast path is very smooth and close to linear. Apparently the main function of the estimated near-unit root in this model is to allow it to use a large initial transient to mimic the behavior of a monotonic deterministic trend.

The reason this happens can be seen in a simple first-order univariate autoregression with constant. If

[3] GDP numbers are used as far back as they are available. They are spliced to GNP numbers for earlier years.

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

and if $\rho \to 1$ while *a* remains fixed, both the implied unconditional standard deviation of *Y*, $\sqrt{\sigma_\varepsilon^2 \big/ \left(1-\rho^2\right)}$, and the implied steady-state, $\alpha/(1-\rho)$, go to infinity. But their ratio, a "*t*-statistic" for the distance of the steady-state from zero, goes to infinity. Thus stationary AR models approach the form of a random walk with trend by sending the estimated steady-state farther and farther from the location of the observed data, in standard-deviation units as well as in absolute units.

This kind of result appears in intensified form in highly multivariate models, as these have higher-dimensional vectors of initial conditions and allow for correspondingly more intricate initial transients. Figure 3 shows results for GDP from a 9-variable vector autoregression, using the same variables employed in the forecasting model from which I at one time generated forecasts each quarter. The steady state is slightly farther away from the data absolutely, and much farther away in standard deviation units, than in the univariate model. Most of the variables in the 9-variable model show this pattern of steady-state far above data, smooth positively sloped initial transient generating an upward trend line. Some, like the commodity price index shown in Figure 4, show more interesting curvature in the trend line. Three -- international value of the dollar, the unemployment rate, and the T-bill rate, (Figures 5-7) show data series entering the two-standard-error band about steady state. However, even for these variables, all or most of the data points in the first half of the sample lie outside the two-standard-deviation band about the steady state.

3. Modeling the Joint Behavior of Consumption and GNP

As an illustration, consider alternative approaches to dynamic modeling of the relation between quarterly aggregate real consumption and GNP. The most straightforward approach is to estimate a linear VAR by OLS, conditioning on initial observations. If we include two lags of each variable, in logs, and a constant, we obtain

$$C = \left(.9542L + .0456L^2\right)C + \left(.1427L - .1432L^2\right)Y + .01258 + \varepsilon_C \tag{3}$$

$$Y = (.2991L - .2300L)C + 1.2139L - .2908L)Y + .09799 + \varepsilon_Y \tag{4}$$

This system has roots of .9993, .8959, .4014, and -.1286. Forecasts with this model for 20 years after the end of the sample are displayed in Figure 8, together with one-standard-error bands. Both the forecasts and the standard-error bands are computed from the stochastic model for the data implied by taking the OLS point-estimates as fixed. The steady-state-values are not shown on the figure, as they are both between 17 and 19. A chi-squared statistic $x'\Sigma^{-1}x$, with x the deviation of the two-dimensional initial observation from steady state and Σ the unconditional covariance matrix of the data, is 2829.

This system is clearly for practical purposes a model with linear deterministic trend. The curvature in the forecast paths is difficult to see, and the error bands around the path expand at something like the linear rate we would expect for a random walk with drift.

Now consider what happens when we estimate the same system by maximum likelihood, using the full likelihood without conditioning on initial observations.[4] The estimated system is

$$C = (.9443L - .0364L)C + (.1571L - .1409L)Y + .01385 + \varepsilon_C \tag{5}$$

$$Y = (.2877L - .2228L)C + (1.2045L - .2773L\ )Y + .09635 + \varepsilon_Y \tag{6}$$

The coefficients of this system are close to those of the system estimated by OLS. So, apparently, are the roots: .9952, .8947, .3811 and -.1222. However for purposes of long-term forecasting, this estimated system is quite different from the first. The steady-state is now log$C$=7.35, log$Y$=7.88, values close to the initial values, and the χ^2 statistic on the initial deviations from steady state, though significant, is an order of magnitude smaller at 153. Chart 9 shows the forecast paths and standard error bands. There is still a more or less linear trend, but it is now negative. Standard error bands expand much as in the model estimated by OLS.

Which of these models is better? There can be no purely objective answer. If one uses OLS, one is implicitly treating as plausible the possibility of very persistent effects of initial conditions, even if they are of a kind implied by the model to be rare occurrences in the long run. While this sounds on the face of it unappealing, probably most economists would be more comfortable with projections like Figure 8 than with projections like Figure 9. If steady-state is far enough away, the model implies that approach to it is nearly linear and uniform within the sample and for a long period outside it. Over a certain fairly long span of time, there is a kind of temporal homogeneity to the model. Thus estimates like Figure 8, and also Figures 2 and 3, appear reasonable. On the other hand, Figures 4-7, in which a deterministic component of the model behaves differently in the first part of the sample from the way it behaves in the latter part of the sample, raise doubts. Whether they are reasonable depends on whether we believe that the initial conditions for the sample were in fact generated by a different mechanism than that which generated subsequent data. These data begin with the immediate post-World-War-II period, which was unusual. Whether it is unusual enough to justify the kind of temporal inhomogeneity in Figures 4-7 cannot be determined by looking at these data.

A pattern like Figure 9 raises different questions. Use of the unconditional likelihood treats the sample symmetrically from start to finish, and indeed in Figure 9 we find that the data at the end of the sample, rather than at the beginning, are the most deviant from the steady-state. If one believed that, over a long enough span, economic history will show symmetric expansions and contractions, the forecasts in Figure 9 would look good. Most economists do not believe this, at least at the time scale implied by Figure 9, but there is no way to make the data tell us objectively that this is an unreasonable interpretation.

Is there a way to specify a prior that favors temporal homogeneity, giving credence to stationary models that oscillate about a steady state as well as to models that oscillate about a temporally homogeneous trend, but not to models that in effect use different models for different

---

[4] See the Appendix for a discussion of computational methods used to formulate and maximize the likelihood in a way that remains numerically stable near the boundary of the stable region.

parts of the sample? Use of methods that condition on initial observations inevitably allow strong but disappearing initial transients. Use of full-sample likelihood avoids treating the first part of the sample asymmetrically from the rest, but may still produce unappealing estimates (like those in Figure 9) that imply some part of the sample is strongly deviant. A method I have been using for some time is to use Theil mixed estimation -- "dummy observations" to incorporate into a Bayesian prior belief that no-change forecasts should be "good" at the beginning of the sample. In a VAR, one can do this by creating observations in which the right-hand side lagged values of the vector X are all set at $X(t-s) = X(1),\ s = 1,\ldots,k$, where k is the number of lags in the VAR. The vector of left-hand side variables in the VAR is also set at $X(1)$. The observation can then be scaled to give it a reasonable weight relative to the sample (though a weight of one can be argued to be a good a priori choice). Such a dummy observation adds information to the prior favoring a no-change forecast. A no-change forecast may be good because the model is close to steady-state at $X(1)$, or because the presence of unit roots makes a no-change forecast always fairly good. Short-lived initial transients will show relatively rapid rates of change initially, and will therefore tend to be more strongly down-weighted by this type of prior than long-lived smooth trends.

My experience with using such priors in a forecasting model is summarized in Sims [1993]. Unfortunately, in the example studied for this paper the OLS estimates yield such smooth behavior that the effect of such priors on forecasts like those in Figure 8 are barely visible. Probably in a 9-variable, 5-lag model they would moderate the unsettling patterns in Figures 4-7, but results to show that are not available at this writing.

4. A Growth Model

Some of the debate and ambiguity about how to model time series like these may be exacerbated by the use of purely "statistical" models, in which economic interpretations of the model are vague and informal. When a model with a more complete economic interpretation is used, it becomes easier to discuss what might be reasonable beliefs about parameters, and indeed some of the apparent sensitivity of results to modest variations in assumptions may disappear.

Consider the neoclassical growth model in which a representative agent maximizes

$$E\int_{t=0}^{\infty} \frac{C_t^{1-\gamma}}{1-\gamma} e^{-\beta t} dt \tag{7}$$

subject to

$$C + \dot{K} + \delta K = \theta K^{\alpha} \ . \tag{8}$$

Assume *ß* and ***q*** evolve stochastically according to

$$d\beta = -\nu(\beta - \bar{\beta})dt + \sigma_{\beta} dW_{\beta} \tag{9}$$

$$\begin{aligned} d(\log\theta) &= \left(-\mu(\theta - \bar{\theta}) + \eta\right)dt + \sigma_{\theta} dW_{\theta} \\ d\eta &= -\phi\eta dt + \sigma_{\eta} dW_{\eta} \end{aligned} \tag{10}$$

The three "*W*" processes, the Wiener process disturbances in (9) and (10) are assumed to be mutually independent.

With modern computer technology, maximum likelihood estimation of a linearized version of this model (which will be accurate for "small" fluctuations around steady state) is not much more difficult than estimation of the full-sample ML linear VAR model. The 2-lag, 2-variable VAR model has 13 free parameters (counting disturbance covariance matrix elements), while the model in (7)-(9) has 11. I have in fact estimated the model using the same data on C and Y as for the VAR. (K is treated as an unobserved component.) The methods used are described in the Appendix. It should be noted that the model is explicitly aggregated over time to account for the fact that observed data are approximately time averages over quarters. The model achieves a better fit than the 2x2 VAR (a log likelihood higher by about 200) despite its slightly smaller number of parameters. It produces the forecasts shown in Figure 10, which appear to me to be the most reasonable in these figures -- they show some growth slowdown as likely, but not certain, and they show forecast error bands widening relatively rapidly as the time horizon lengthens.

In this model one can also experiment with full-sample likelihood approaches versus likelihood conditional on two initial observations. The differences are in the expected directions. The χ^2 statistics for deviation of initial observation from steady-state are 13.27 for the conditional model and 10.89 for the unconditional one. The differences are much smaller, however, and indeed the implied forecasts from parameters estimated with the two criteria are so little different that they cannot be distinguished by eye on the graphs (and therefore are not shown separately).

It is interesting to speculate on why this model, despite its freedom to generate complex deterministic trending behavior by choice of the parameters in (9) and (10), does not do so, even when estimated with conditional likelihood. One explanation may be the fact that *C/GDP* has been declining since the early 50's. In this model, the explanation for such a decline would be the approach of steady-state.[5] With steady-state very far away, it would be hard to explain a drift in the savings rate. This may also explain the small estimated value of d – if d is large, the proportional rate of decline in net savings is magnified, and forces the model to put steady-state nearer to explain the more rapid drop in accumulation.

Full ML Parameter Estimates

a: 0.469049	b: 0.047156	g: 1.87164	d: 0.000615183	q: 41.993	m: 8.71e-05
f: 20.1329	n: 0.49895	σ_β: 0.01596	σ_η: 0.55605	σ_θ: .00034356	

Note that some care is required in interpreting the estimated standard errors of disturbances. The m near zero and the corresponding σ_β of .016 do imply that there is a component of q that is close to a logarithmic random walk with a one-period prediction error of about 1.5%, before time averaging. But the large σ_η corresponds to a f of 20, meaning this component's dynamics are very rapidly damped. In time averages over a period of length .25 (one quarter), the contribution of this component to variance will be strongly attenuated. Indeed, this component shows so little

[5] This simple model has no government or trade balance, so the fact that in reality I/Y has mostly risen over the same period does not affect the fact that the model, which fits only C and *GDP* data, would interpret the declining ratio of C to *GDP* as reflecting convergence to steady state.

serial correlation that it probably contributes little to the model's explanatory power. At the quarterly level of time aggregation, $h\ dt$ is not behaving very differently from $\sigma_\theta dW_\theta$, so the model is close to what would have been produced with h suppressed and σ_θ increased to about .032. This implies a one-quarter-ahead forecast standard error in q of about 1.6%.

5. Conclusions

The examples studied here should make it clear that the fact that initial conditions are "asymptotically unimportant" cannot be an excuse for failing to be explicit about how we are (or are not) modeling them. Also, the fact that dealing with initial conditions explicitly is computationally difficult should be of diminishing importance. The methods applied in this paper, though non-standard and time-consuming to program, took modest computing time to execute. They could be standardized to allow wider use.

The issues discussed here bear potentially on a wide variety of applications. The literature on "convergence" in economic growth has looked through large cross-sections for a pattern of movement toward country or region-specific trend lines or steady states. The phenomena studied here probably play a role in that literature, a role that has not been fully examined.

The practice of imposing unit roots and cointegration on models as if it were exact *a priori* knowledge may also have an interpretation as an approximate Bayesian procedure reflecting a prior preferring temporal homogeneity to models that imply large initial transients. However just as flat priors, particularly combined with conditional likelihood, may in effect give too much credibility to models that show stationary fluctuations around large transients, they may well give too much credibility to models with strong “long run relationships”. The long run relationships may turn out to have been estimated as showing complex deterministic behavior.

APPENDIX

NUMERICAL METHODS

This paper takes an approach to estimating its non-linear dynamic stochastic general equilibrium (DSGE) model that is very similar to, but more primitive than, the approach taken by Leeper and Sims [1994 and Jinill Kim [1995] which were completed later. As the structural model of this paper is meant only to show the promise of structural modeling for better treatment of trends and long-term forecasts, the computational methods used for this paper's structural model are only sketched below.

For all of the models of C and Y considered in this paper, the likelihood is taken to be Gaussian. For the underlying structural model of C and Y, of course, the distribution of the data is implied not to be Gaussian, but we are using this model linearized about deterministic steady state, so with Gaussian driving disturbances the data are implied to be Gaussian. Let $X_t = \left[C_t \ Y_t \right]'$ be the data vector and $R_X(t)$ be its matrix-valued autocovariance function. Let $\bar{X} = \left[\bar{C} \ \bar{Y} \right]'$ be the deterministic steady-state values of C and Y. Define

$$\tilde{X}_T = \left[X_1', \ldots, X_T' \right]' ; \ \bar{X}_T = 1_T \otimes \bar{X} \tag{11}$$

where 1_T is a $T \times 1$ vector of 1's; and define

$$\Omega_T = \left[R_X(j-i) \right] \tag{12}$$

Then the log-likelihood for the sample $X_1, \ldots, X_T$ is

$$\mathsf{L}_T = -.5 \log \left| \Omega_T \right| -.5 \left(\tilde{X}_T - \bar{X}_T \right)' \Omega_T^{-1} \left(\tilde{X}_T - \bar{X}_T \right) . \tag{13}$$

To this point all we have said applies both to the case of a linear VAR model and to that of the linearized structural model. The differences are only in the way $\bar{X}$ and R_X are derived. Indeed the most computationally demanding part of the likelihood evaluation is the inversion of the large, ill-conditioned Ω_T matrix, so that the more complicated derivations for $\bar{X}$ and R_X for the structural model do not increase the time required for the calculation by very much. For this paper Ω_T remained well-conditioned enough to invert by a Cholesky factorization approach. For larger models, it would probably be worthwhile to use an approach that efficiently exploited the block-Toeplitz structure of Ω_T.

We obtain conditional likelihood by replacing $\bar{X}_T$ with the conditional mean $\bar{X}_T^C$ of $\tilde{X}_T$ given X_0 and replacing Ω_T with Ω_T^C , the covariance matrix of $\tilde{X}_T$ conditional on X_0 . Letting $V = \left[R_X(1)', \ldots R_X(T)' \right]'$, we have

$$\bar{X}_T^C - \bar{X}_T = V R_X(0)^{-1} \left(X_0 - \bar{X} \right) \tag{14}$$

$$\Omega_T^C = \Omega_T - V R_X(0)^{-1} V' \tag{15}$$

and

$$\mathsf{L}_T^C = -.5 \log \left| \Omega_T^C \right| -.5 \left(\tilde{X}_T - \bar{X}_T^C \right)' \left(\Omega_T^C \right)^{-1} \left(\tilde{X}_T - \bar{X}_T^C \right) \tag{16}$$

Of course for the linear VAR model the maximum likelihood estimator can be computed directly by the usual least-squares formulas, so there is no need to use (16), but we use (16) for estimating the growth model.

Very significant further computational problems arise in computing R_X from the autoregressive form of the model, in implementing the computations required to form (16), and in maximizing likelihood. It is likely that a recursive approach to forming the likelihood using the Kalman filter is more efficient and accurate than working directly with (16) as was done in this paper's computations. The reader is referred to Leeper and Sims [1994 and Jinill Kim [1995] for a more detailed discussion of these issues.

REFERENCES

Kim, Jae-Young [1992]. "", processed, Yale University.

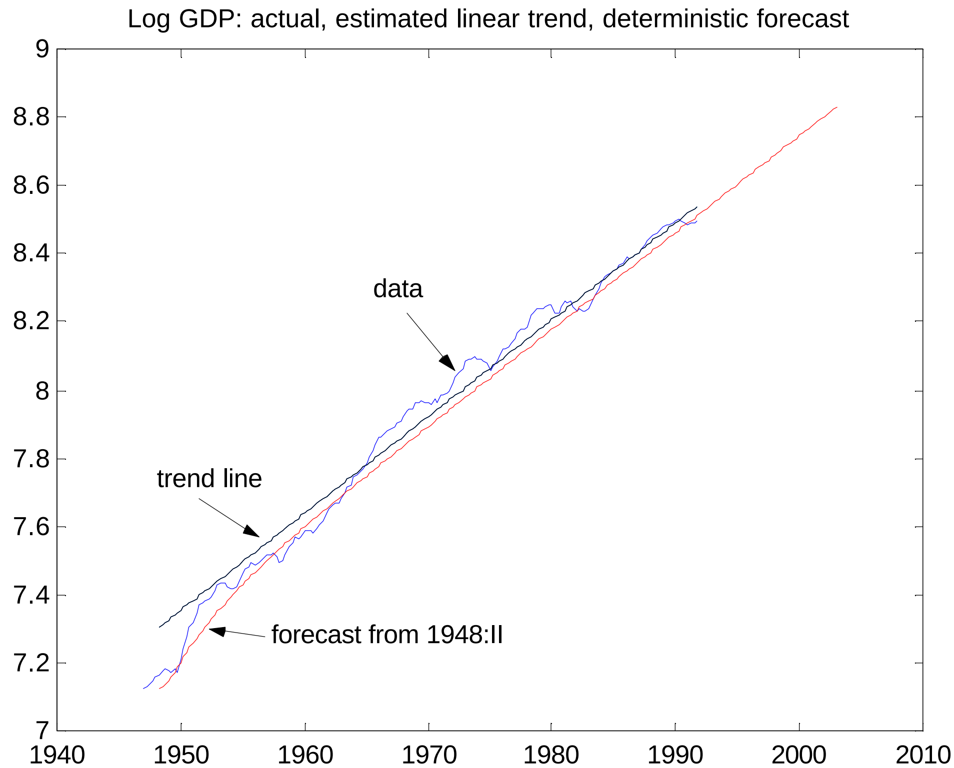
Kim, Jinill [1995]. “Effects of Monetary Policy in a Stochastic Equilibrium Model with Real and Nominal Rigidities,” processed, Yale University.

Leeper, Eric and C.A. Sims [1994]. “Toward a Modern Macro Model Usable for Policy Analysis,” *NBER Macroeconomics Annual*, 81-117.

Sims, C. A. [1993]. “A 9 Variable Probabilistic Macroeconomic Forecasting Model,” in *Business Cycles, Indicators, and Forecasting*, James H. Stock and Mark W. Watson, editors, NBER Studies in Business Cycles Volume 28, 1993, 179-214.

_________________ [1991]. “Comment on ‘To Criticize the Critics’ by Peter C.B. Phillips”, *Journal of Applied Econometrics*, 6, 1991, 423-434.

Figure 1



Log GDP: actual, estimated linear trend, deterministic forecast
9
8.8
8.6
8.4
8.2
8
7.8
7.6
7.4
7.2
7
1940
1950
1960
1970
1980
1990
2000
2010
data
trend line
forecast from 1948:II

Figure 2

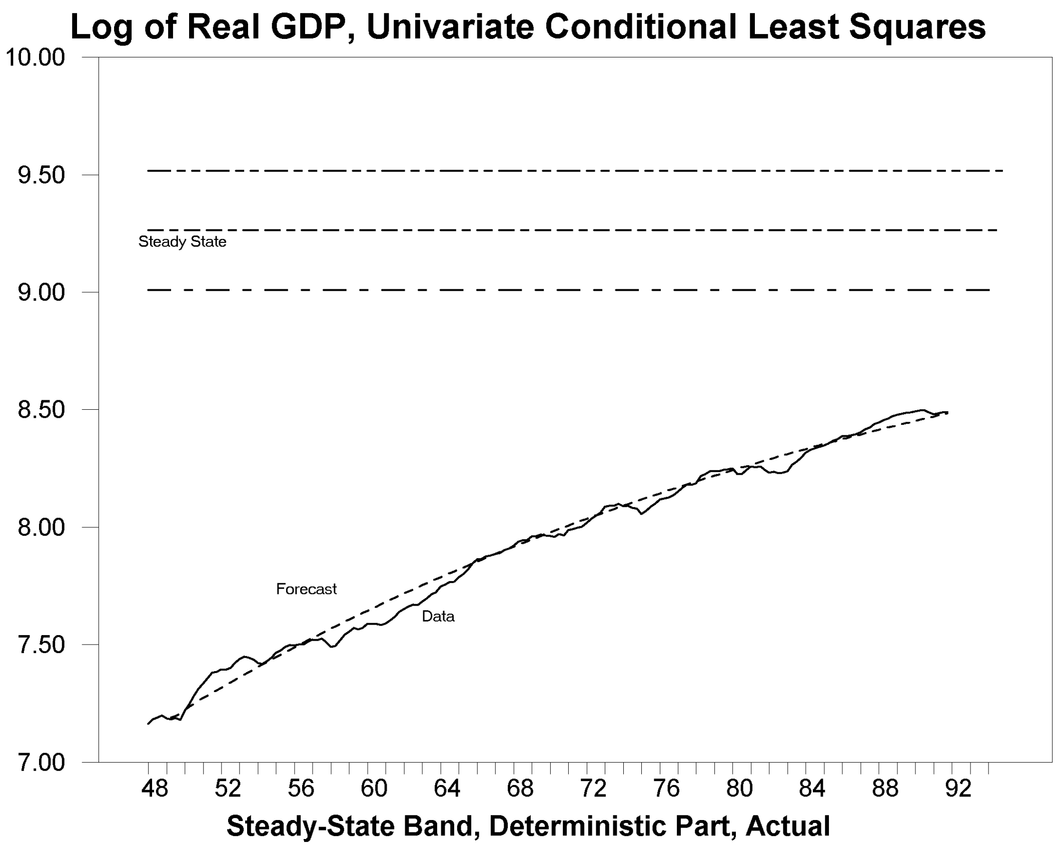

Log of Real GDP, Univariate Conditional Least Squares
10.00
9.50
9.00
8.50
8.00
7.50
7.00
Steady State
Forecast
Data
48
52
56
60
64
68
72
76
80
84
88
92
Steady-State Band, Deterministic Part, Actual

Figure 3

GDP from 9-variable model

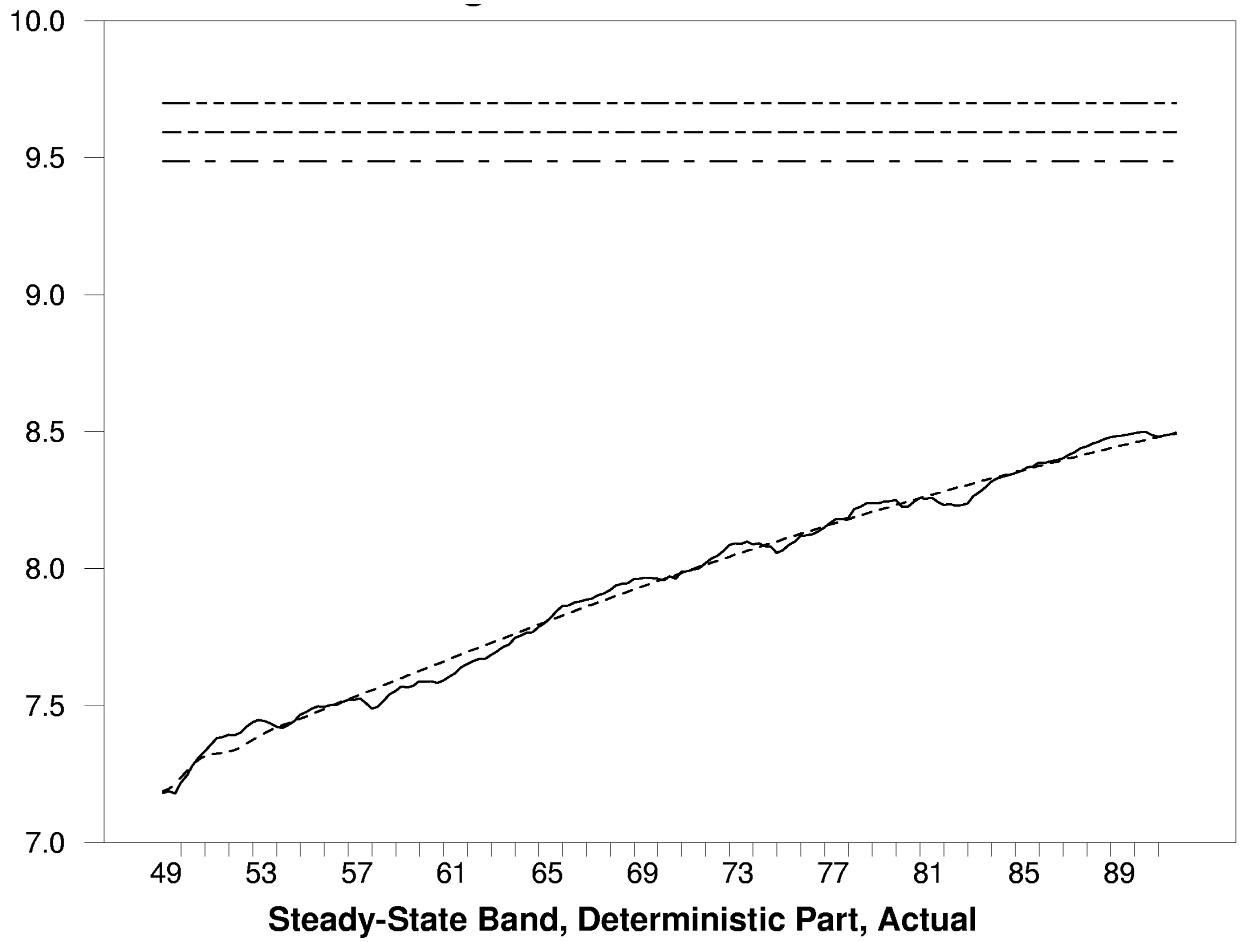


Figure 4

Commodity Prices from 9-variable model

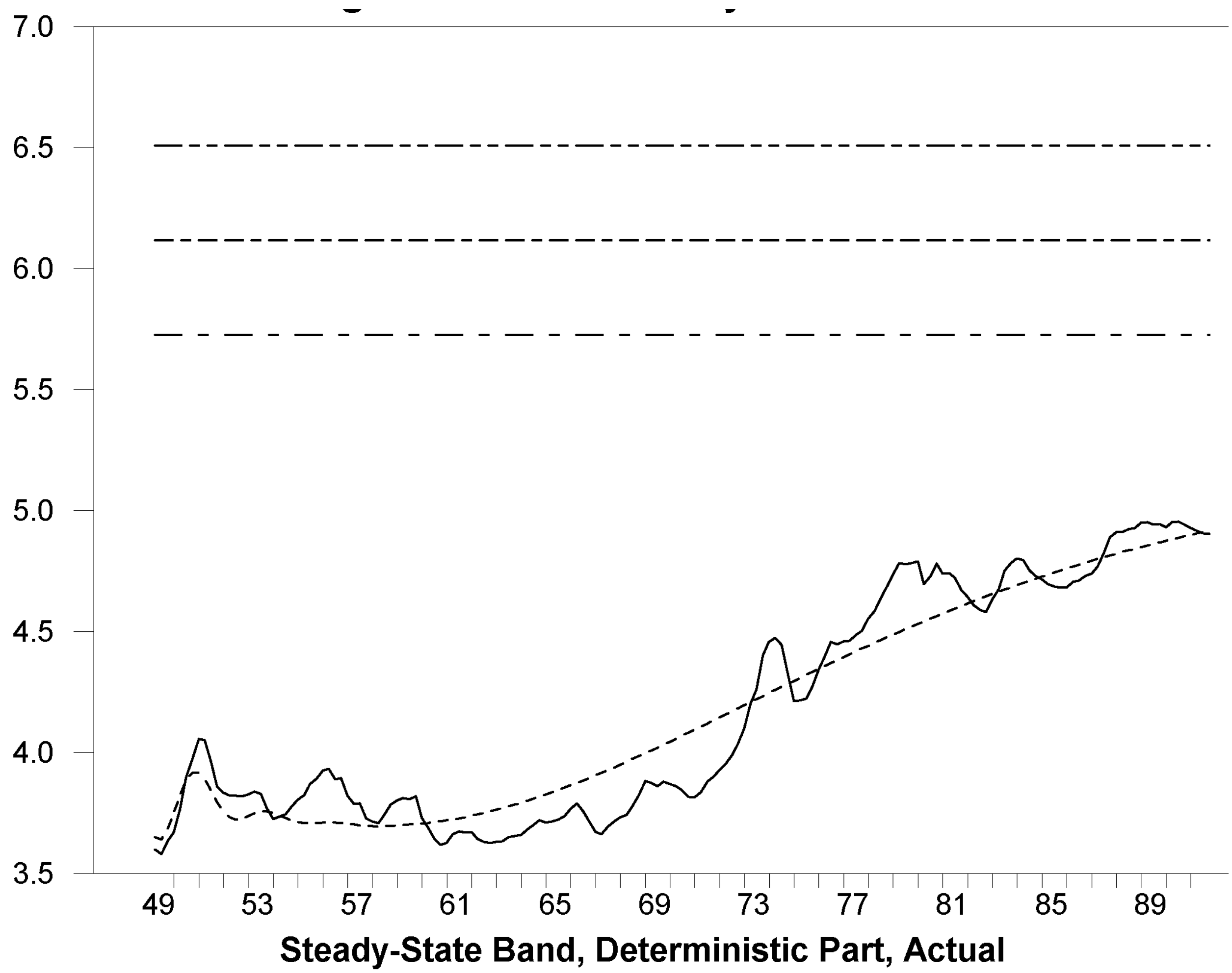


Figure 5

Value of the Dollar from 9-variable model

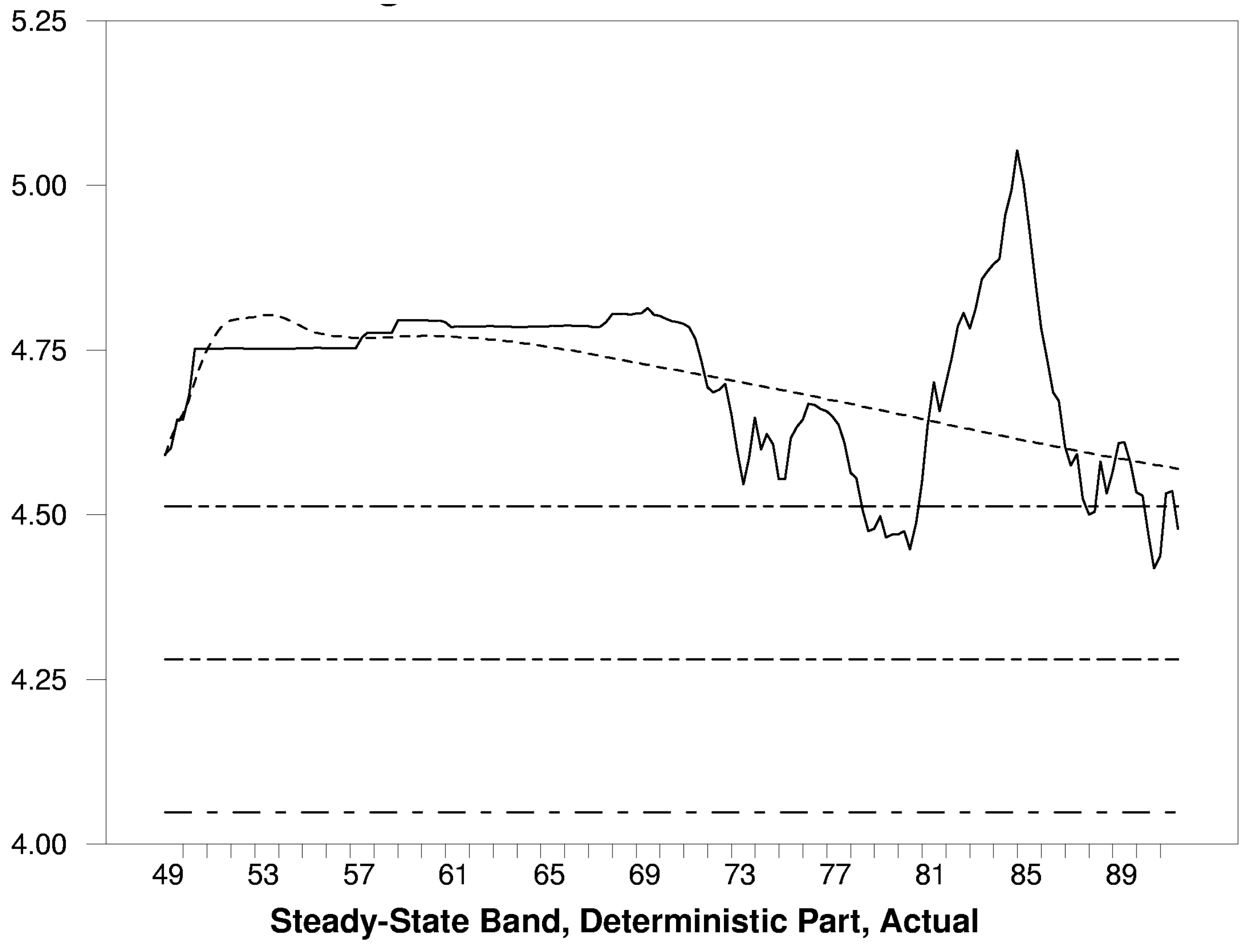


Figure 6

Unemployment from 9-variable model

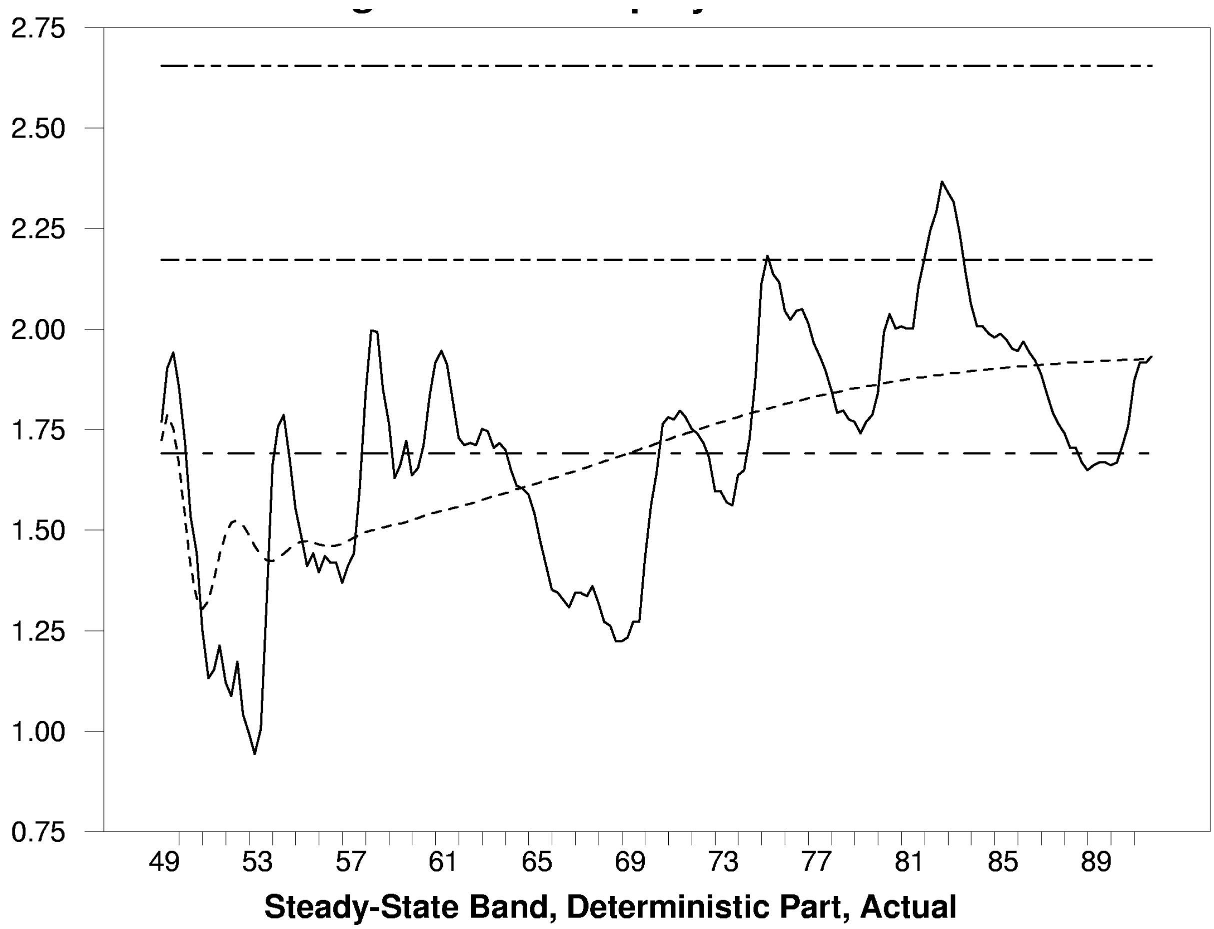


Figure 7

Tbill Rate from 9-variable model

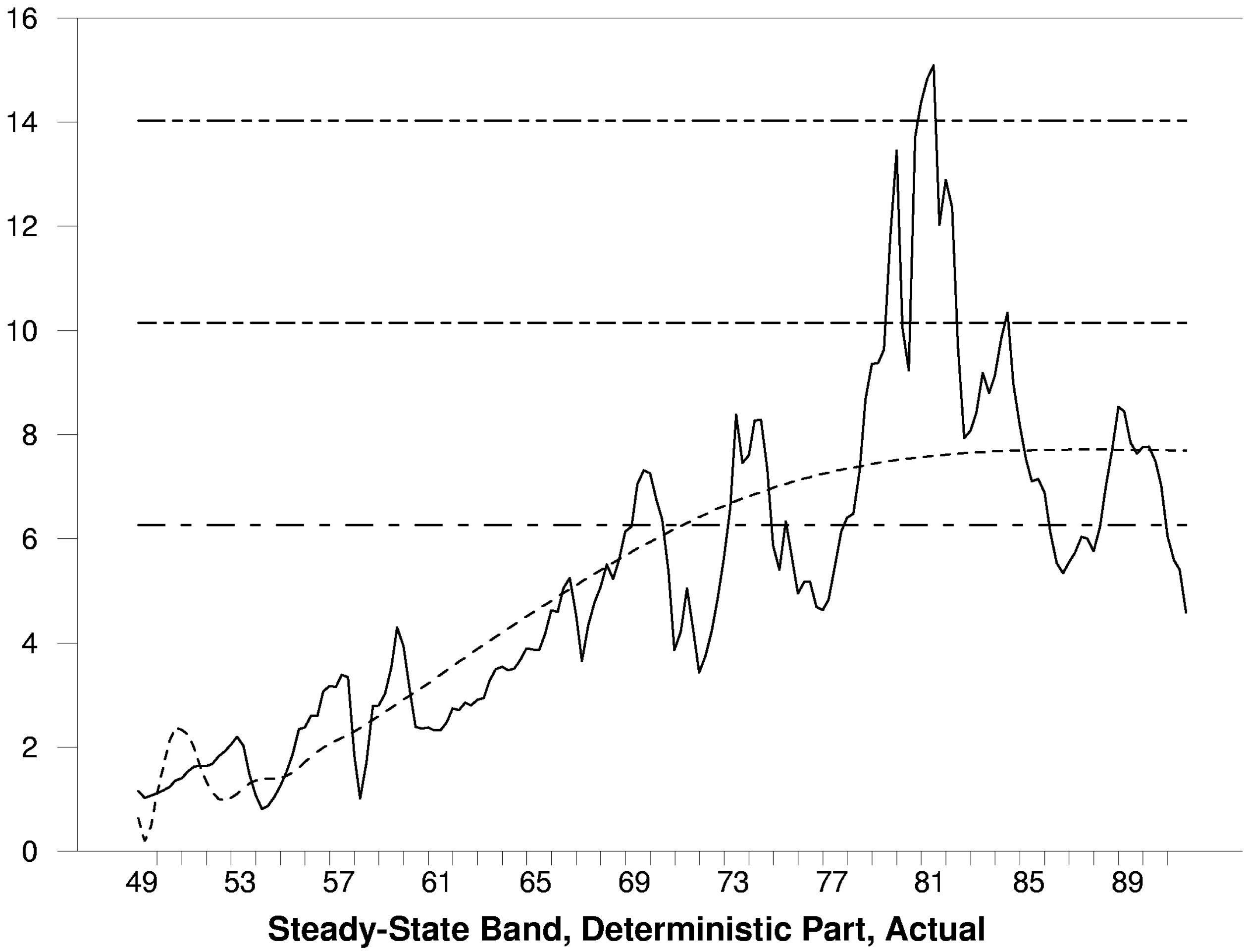
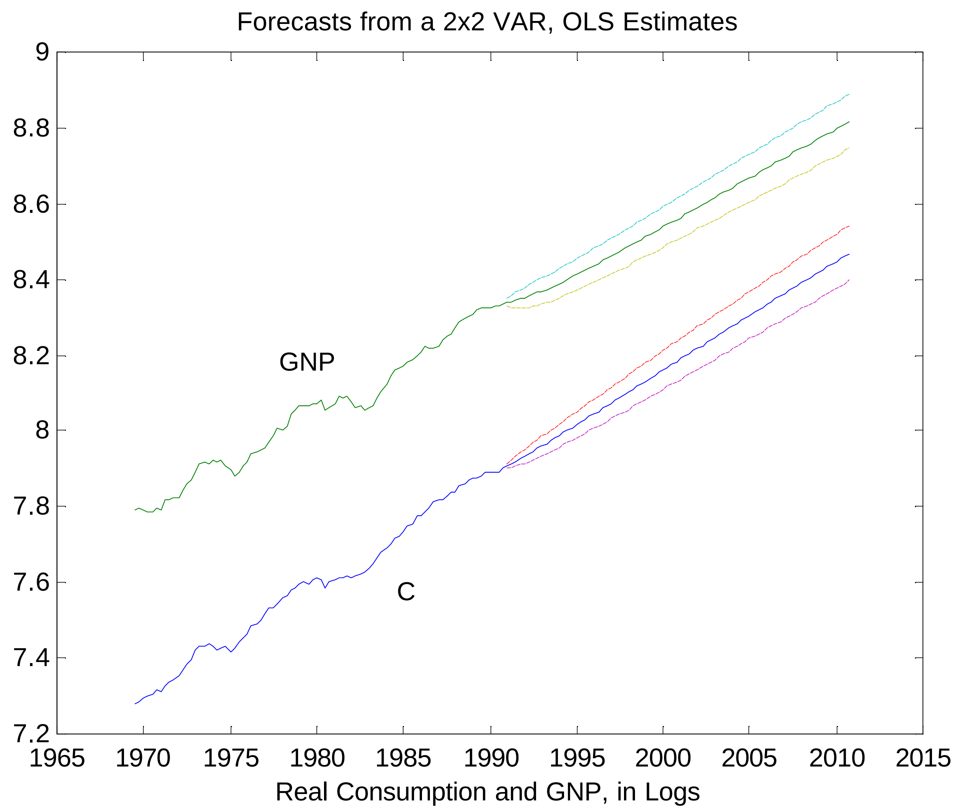


Figure 8



Forecasts from a 2x2 VAR, OLS Estimates
9
8.8
8.6
8.4
8.2
8
7.8
7.6
7.4
7.2
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
GNP
C
Real Consumption and GNP, in Logs

Figure 9

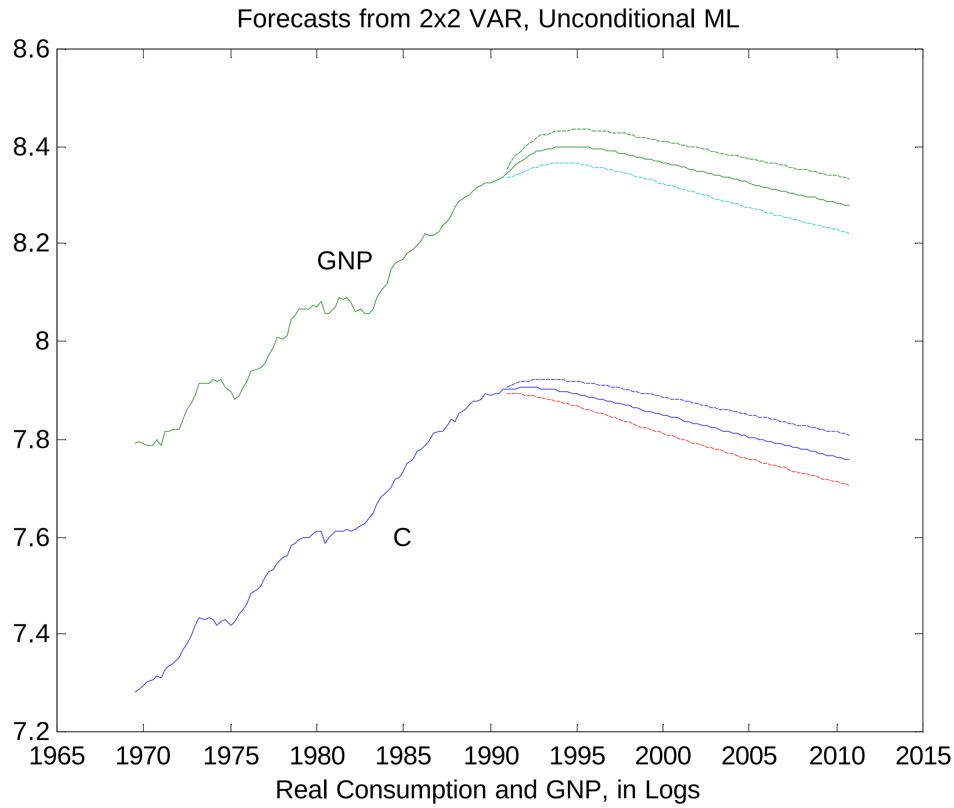
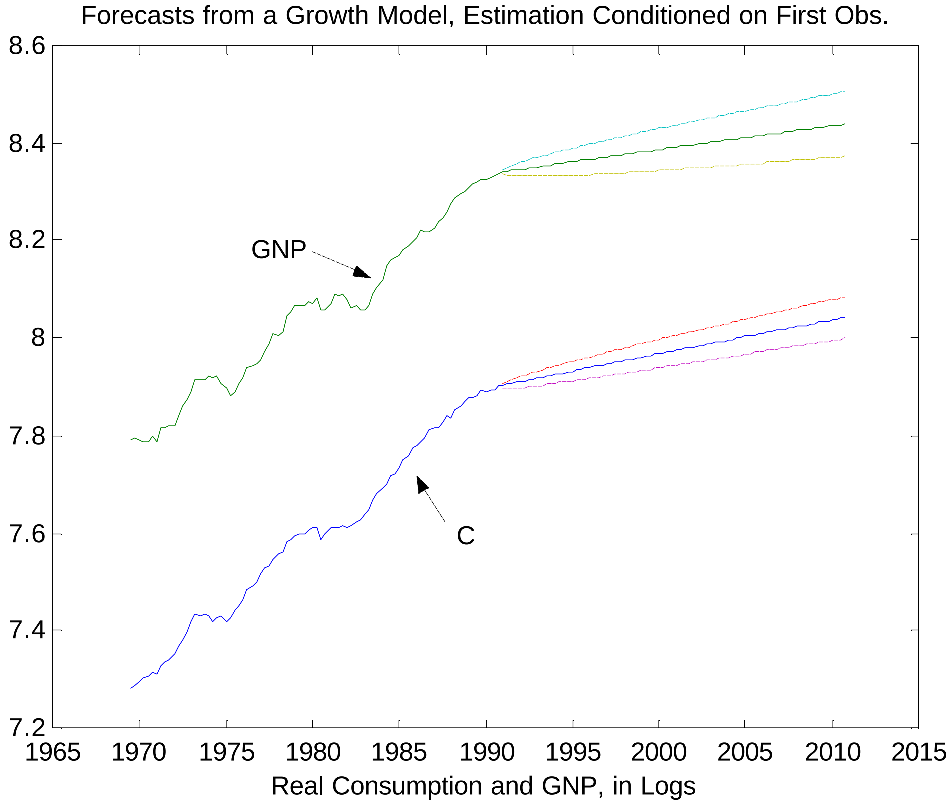

Forecasts from 2x2 VAR, Unconditional ML
8.6
8.4
8.2
8
7.8
7.6
7.4
7.2
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
GNP
C
Real Consumption and GNP, in Logs

Figure 10


Forecasts from a Growth Model, Estimation Conditioned on First Obs.
8.6
8.4
8.2
8
7.8
7.6
7.4
7.2
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
GNP
C
Real Consumption and GNP, in Logs