Carnegie-Rochester Conference Series on Public Policy 47 (1997) 281-290
North-Holland

Real business-cycle realizations, 1925-1995
A comment

Kevin D. Hoover*
University of California, Davis

In their paper, Gregor Smith and Stanley Zin address an important question: how well do real-business-cycle models account for the historical behavior of the economy? This empirical exercise is in the spirit of traditional econometrics. It would hardly be surprising to see such comparisons in other contexts, but it is surprising in the case of real-business-cycle models. The standard empirical methodology, following Kydland and Prescott's (1982) original article on real-business-cycle models, is calibration. Typically, calibrated real-business-cycle models have been evaluated through subjective comparisons between selected sample moments for the actual economy and the corresponding moments from simulated data generated using the model. Smith and Zin's paper is especially attractive because it seeks to bridge the gap between the calibration methodology and traditional econometrics.

The central empirical question about any model is, how well does the model reproduce characteristic features of the world? If a model does reproduce those features well—perchance, perfectly—one cannot logically deduce that the model is the true one that generated the data. What one can logically deduce is that the model belongs to a class—much narrower than the class of all possible models—that can reproduce the data to a given degree of accuracy. That is itself useful knowledge—especially if rival models do not fall within such a class. Calibrators have typically emphasized (e.g., Kydland and Prescott 1991 and 1996) that their models are highly stylized and are meant to capture certain dimensions of the data, while ignoring others. Kydland and Prescott, for example, argue that standard econometrics biases the investigator against the real-business-cycle model, because it makes the model responsible, as it were, for features of the data it was not designed to capture. The standards of traditional econometrics may undermine the usefulness of economic models for policy analysis, because policy analysis

*I thank Kevin Salyer for helpful comments on an earlier draft.

0167-2231/97/$17.00/
PII 0167-2231(98)00013-X

requires a structural understanding. The real-business-cycle model, despite the fact that it is a stylized or, as philosophers of science prefer to say, *idealized* model, captures essential core features of the economy. It is, therefore, useful for policy analysis, even though it explicitly applies only to a limited domain.[1]

Smith and Zin are to be congratulated for recognizing that, while the calibrators' complaint against traditional econometrics may have some merit, it is nevertheless an essential test of any model, even a calibrated one, that it carry nonredundant information about historical time series. As they put it succinctly: "Having a zero variance of the difference between two series is a much stricter criterion than having a zero difference of their variances" (p. 244). Matching sample moments narrows the class of acceptable models, but that class is still likely to be impossibly broad and not very discriminating. Matching historical realizations is likely to narrow the class in a far more telling way. The interesting question is how to make the comparison to historical statistics in a way that both provides a stringent and discriminating test and accounts for the calibrators' legitimate concern that real-business-cycle models not be penalized for not capturing aspects of the data that they were, in fact, designed to ignore. Smith and Zin have made a good start. There are a few issues that I believe need further attention.

1 The treatment of capital biases the model against success

The state variables of the Smith and Zin model are capital, the technology shock, and the government expenditure shock. The two shocks are supposed to be measured, but the capital stock is calibrated in the sense that only its initial value (in 1925 or 1955) is specified, and the capital variable is then allowed to evolve according to its law of motion within the model.

Conditioning on predicted capital biases the model away from fitting the actual data. The capital variable bears the burden of every failure of the model to fit the data, and the errors are compounded and passed on to the prediction of the capital variable in the next period. This is clear from the B panels of Figures 2-6 and 9-13, which show that the model is a terrible predictor of capital. Failure in this dimension, however, tells us very little about how good the model is. It just says that the model is not perfect, which we knew *a priori*, and that the modeling strategy piles the imperfections onto its predicted capital stock. Of course, since other predictions depend on the capital variable, they are likely to be inaccurate, too.

What should we conclude if the model fits better for, say, 1940 when it is

[1]Hoover (1995) attempts to relate the real-business-cycle model to the philosophical literature on idealization. Hoover (1994) raises some general philosophical questions about the empirical evaluation of idealized models.

started in 1937 than when it is started in 1925? I think only this: that the sins of the parents have had less time to be visited on the children. A fairer test of the model would be to use the alternative that they note (p. 251) and update the capital stock variable continually (i.e., to use the *actual* capital stock as the state variable), so that the model generated step-ahead forecasts for capital (equivalently predicted investment) and allowed bygones to be bygones. Clearly there are times when one is concerned about the ability of a model to forecast many periods ahead. The issue here is not forecasting ability; it is rather how closely the model captures the true underlying relationships in the economy, understanding of which is essential for policy analysis. After all, the model already conditions on *current* technology and government spending shocks, so many-period-ahead forecasts are ruled out.

Given the incredible demands placed on predicted capital in the model, the fact, as Smith and Zin (p. 258) observe, that the model fits the early 1990s better than the earlier recessions is probably more due to luck than to the nature of the model itself.

2 Success at modeling output is uninformative

Ideally, one would like to observe technology shocks directly, feed them into the model as an input, and check to see how well predicted output (Smith and Zin's Y_S) matches actual output (Y_t). Unfortunately, technology shocks are latent variables, not directly observable. Following the tradition of real-business-cycle modelers, Smith and Zin measure technology shocks through the Solow residual. Every difference between actual output and predicted output is ascribed to technology. If the world were precisely Cobb-Douglas in technology, and if we knew the true factor-share parameters, then the Solow residual would in fact capture the true technology shock (z_t).

As Smith and Zin observe, the Solow residual may well be a poor measure of true technology shocks. They write (p. 259) "... it is tenuous to view Solow residuals as pure measures of technological progress..." In addition to their references to Jorgenson and Griliches, we should note that Hall (1986, 1990) has pointed out that the existence of market power distorts the Solow residual as a measure of technology shocks. Hall, Jorgenson, and Griliches argue that the Solow residual picks up features of the economy that are not technological change. In an unpublished work, James Hartley (1994) investigates whether the Solow residual picks up technological change at all. He creates simulated data in the linear-quadratic real-business-cycle model of Hansen and Sargent (1990;1996) in which it is possible to introduce genuine shocks to the underlying production technology. He then finds that for a range of reasonable parameterizations, Solow residuals calculated on this simulated data miss the magnitude and often even the direction of the true technology shocks.

These doubts about the Solow residual to one side, I am willing to regard it as probably a pretty good approximation for the purposes to which Solow (1957) put it: estimating the secular importance of technological change. And I am even willing to entertain the possibility that it may be acceptable as a measure of higher frequency technology shocks when one is interested principally in characterizing its stochastic properties in order to guide the simulation of a calibrated model—this is the typical use in the real-business-cycle literature. Nonetheless, I do not believe that predicted output from a model driven by Solow residuals can be usefully compared to actual historical output.

To see the difficulty, consider a simpler version of a real-business-cycle model in which we abstract from time trends. Initially, let labor be supplied inelastically and condition as suggested in Section 2 above on the actual capital stock (K_t) rather than, as in Smith and Zin's paper, the predicted value (K_s). The Solow residual (z_t) can be calculated in log-linear form:

$$log(z_t) = log(Y_t) - \alpha log(K_t) - (1-\alpha) log(n_t). \tag{1}$$

The log-linear version of the production function is given by

$$log(Y_S) = log(z_s) + \alpha log(K_t) + (1-\alpha) log(n_s). \tag{2}$$

To generate a forecast for output, we notice that labor is inelastically supplied, so $n_s = n_t$, and then we substitute the measured Solow residual for z_s in equation (2), with the result that our forecast of output is

$$\begin{aligned} log(Y_s) &= log(Y_t) - \alpha log(K_t) - (1-\alpha) log(n_t) + \\ &+ \alpha log(K_t) + (1-\alpha) log(n_t) = log(Y_t) \end{aligned} \tag{3}$$

or

$$Y_S = Y_t \tag{3'}$$

Our model fits perfectly; the correlation between predicted and actual output is unity. Does anyone believe that this is a demonstration of its goodness?[2]

The procedure here reminds me of an article that I once saw that measured poverty as the difference between people's income and a poverty line.[3] It then ran a regression of poverty on income, which naturally had an R^2 of unity and *very* large t-statistics. The author then proclaimed a perfect understanding of the roots of poverty and a policy prescription: people are

[2]The essential point of this section was anticipated by King (1995, p. 84).

[3]My memory is that this was in the *Journal for Irreproducible Results*, but diligent search was unable to locate it. I would be grateful to anyone who can supply the correct reference.

poor because they do not have money; to end poverty, give the poor more money.

Smith and Zin, of course, do not get a perfect fit. The reason is that the inputs to their production function do not recapitulate the capital and labor measures used to calculate the Solow residual. In particular, even if my earlier point is granted about using the current capital stock, the labor input (n_s) is determined by other features of the model—in fact, by features that are considered the most characteristic of real-business-cycle models, such as intertemporal substitutability of labor and optimal investment and consumption planning.[4] So, it is natural to relax our assumption of an inelastic labor supply. Then equation (3) becomes

$$\begin{aligned} log(Y_s) &= log(Y_t) - \alpha log(K_t) - (1-\alpha)log(n_t) + \alpha log(K_t) + (1-\alpha)log(n_s) \\ &= log(Y_t) - (1-\alpha)[log(n_t) - log(n_s)]. \qquad (4) \end{aligned}$$

How well predicted output fits actual output is seen to depend on how well predicted labor fits actual labor. Still, there is an artifactual element to the correlation between predicted and actual output. Notice that the share parameter α is not given in nature, but is a modeling choice. If α is modeled to be close to one, then the predicted and actual output are again nearly perfectly correlated. Now, it might be objected that we *know* α is not close to one but close to 0.33 (Smith and Zin's assumption). I agree. But information about the true size of α comes from the calibrator's supply of exogenous information and has nothing to do with the fit of the model to historical data or with traditional econometrics. It underscores Kydland and Prescott's advocacy of external sources of information to pin down free parameters. We must not forget that whether α is one, 0.33, or zero, actual output shows up on the right-hand side of equation (4) only because we put it there in the construction of the Solow residual, not because the model generated it by closely matching the structure of the economy.[5]

Of course, it would be a marvelous testament to the success of the model if the right-hand side of equation (4) turned out to be very nearly Y_t. That would occur because the model's predicted labor was very nearly the actual labor. But why test that indirectly by comparing Y_S to Y_t? It is more direct and to the point to compare n_s to n_t.

Some of the same problems may arise in testing the historical accuracy of the model for labor since the Solow residual also contains current labor

[4]Additionally, independently detrending the Solow residual and the other inputs to the production function may introduce discrepancies between actual and model-generated data.

[5]Hoover and Salyer (1996) provide simulation evidence that the Solow residual does not convey useful information about technology shocks and that the apparent success of real-business-cycle models in matching historical data for output is wholly an artifact of the use of current output in the construction of the Solow residual.

information by construction. Smith and Zin's most revealing test is, therefore, the test of predicted consumption against actual consumption (see the D panels of their Figures 2-6 and 9-13).

To give up the comparison of historical and predicted output is not to give up the comparison of historical and predicted data altogether. As Smith and Zin (p. 259) note, there are other ways to back out combinations of the technology shock and the government expenditure shock from the model. If, for instance, one were to condition on actual output, inherited capital, and the government expenditure shock, then one could back out another measure of z_s. But, given that we have nothing independent to compare it to, the more interesting point is that we can back out measures of consumption and measures of labor uncorrupted by actual labor. Historical comparisons on these dimensions would be useful tests of the model.

3 Success should be judged comparatively

Even if we restrict our attention to consumption, how well does Smith and Zin's model perform? We do not really know. A perfect model would track actual consumption perfectly; it would show spectral coherence of unity and phase shift of zero. We do not have that, but are we close? The only standards are impressionistic. It is rarely an interesting question whether a particular model is a good characterization of the world. Far more important and informative is the question whether it is a better characterization than some other competing model. The great puzzle of calibrated models is how to evaluate them comparatively.

Suppose that, at the most basic level, we agree on the fundamental structure of the model but not on how it should be calibrated. How then should we resolve disputes over the parameterization? Should α be 0.33 or 0.4; should ρ be 0.99 or 0.95? A metric is needed to adjudicate such disputes. Kydland and Prescott rightly point out that not just any metric will do. It must be a metric that does not penalize the calibrated model for failures on dimension it is not designed to capture. There have been efforts such as Watson's (1993) asymmetrical measures of goodness of fit that address this question. It may even be possible to employ more traditional econometric estimation in a manner that accounts for Kydland and Prescott's concerns. Smith and Zin follow a mixed approach in which the decision rules are estimated by a generalized method of moments, while production and preference parameters are calibrated. Berkowitz (1996) shows how one can estimate models in the frequency domain in ways that emphasize frequencies of interest (say, two to eight years for real-business-cycle models) and deemphasize irrelevant frequencies (for business-cycle models, very low and very high frequencies).

Proponents of real-business-cycle models sometimes talk as if the essential elements of macroeconomics were accepted without dispute, so that all acceptable models would share common core features. Models would then differ only in relatively secondary matters: do we have only technology shocks or should there also be fiscal shocks? is labor divisible or indivisible? does the model acknowledge the institution of overtime pay? does the model account for household production? A variety of models can be built on fundamentally the same core assumptions, each distinguished by peripheral assumptions and parameterizations. Again, we are presented with the problem of how to discriminate among such models in a disciplined manner.

While I have characterized a variety of models as possessing a common core, some of the variants of the core model are sufficiently different that they might be regarded as challenging the core. A monetary model, for example, could be seen just as a core model with an additional shock. But the question of how to introduce monetary elements is a vexed one, and the monetary model may also be regarded as fundamentally different from the real-business-cycle model, rather than a variant of it. Even more radically opposed models are seriously entertained. It would be possible to generate an idealized, Keynesian model. Farmer (1993) presents a model the core of which features increasing returns and belief shocks.

Allegiance to any particular core is likely to insulate models from discriminating empirical tests. Any failure of the model to match the data on the usual calibrators' standards is likely to result in the addition of another epicycle. Confidence in the core inadvertently turns the most common modeling assumptions into, to use the terminology of the philosopher of science Imré Lakatos (1970), the *hardcore* of a research program, insulated from severe empirical test by the heuristics of the peripheral assumptions. Openmindedness and a decent respect for the opinions of other researchers should, however, make us feel uneasy about such insulation.

If we merely ask, how well does this model fit the data? we will never resolve disputes over the core. But different cores are not intrinsically incomparable: after all, competing models purport to explain the same actual data. We must, therefore, go beyond the exercise in Smith and Zin's paper and run horse-races between competing models. Smith and Zin's correct insight is that there is nothing in the calibration methodology that makes it inevitable that the models not be judged with reference to historical data. A good model necessarily carries nonredundant information about the historical time-series. More importantly, a better model carries more such information than its rivals.

In the published version of their paper, Smith and Zin (p. 260) concede the usefulness of such comparisons and mention the possibility of applying similar methods to the models of Taylor and of Leeper and Sims. Unfortunately, they do not suggest a method of systematic comparison or an appropriate metric. Where we need to go farther than Smith and Zin do in this paper is to isolate the particular dimensions along which some historical comparisons are useful and to compare models with the full range of peripheral and core assumptions along these dimensions.

A natural way to make such comparisons is in an encompassing framework (see Hendry 1995, ch. 14). Kydland and Prescott have pointed out that real-business-cycle or other calibrated models are unlikely ever to fit as well as econometric models with many free parameters. It is a fair question to ask, however, whether imposing the restrictions implicit in a calibrated model permits us to eliminate some of those free parameters. That provides some measure of the information carried by a calibrated model relative to an econometric alternative. It is then natural to compare competing models on the basis of this relative, incremental information. One calibrated model is superior to another (one model *encompasses* the other) if the incremental information it carries includes all of the incremental information of the rival model. Of course both models may also carry very little incremental information (i.e., not allow us to validly restrict the econometric alternative), or each might carry unique information (neither encompasses the other, rather, a satisfactory model must incorporate features of both models).[6]

There is still a question about what the econometric alternative should describe. On this score, Smith and Zin are, I believe, right: the historical time-series provides a stringent challenge for the modeler. Horse-races are necessary. Work such as Smith and Zin's takes a step toward making such horse-races possible.

[6] Hoover (1994) sketches the main lines of such a strategy.

References

Berkowitz, J., (1996). Generalized Spectral Estimation, Finance and Economics Discussion Series, no. 97-37, Division of Research ad Statistics, Division of Monetary Affairs, Board of Governors of the Federal Reserve System, August.

Farmer, R.E.A., (1993). *The Macroeconomics of Self-Fulfilling Prophecies.* Cambridge, MA: MIT Press.

Hall, R.E., (1990). Invariance Properties of Solow's Productivity Residual, Peter Diamond (ed.) *Growth/Productivity/Unemployment: Essays to Celebrate Bob Solow's Birthday.* Cambridge: MIT Press, 71-112.

Hall, R.E., (1986). Market Structure and Macroeconomics Fluctuations, *Brookings Papers on Economic Activity*, no. **2**: 265-338.

Hansen, L.P. and Sargent, T.J., (1996). Recursive Models of Dynamic Linear Economies, manuscript (July 23, 1996 draft) available as of 25 November 1997 on the World Wide Web at http://riffle.stanford.edu/hansen.html.

Hansen, L.P. and Sargent, T.J., (1990). Recursive Linear Models of Dynamic Economies, *National Bureau of Economic Research*, Working Paper, no. 3479.

Hartley, J.E., (1994). Technology in Macroeconomic Models, unpublished doctoral dissertation, Department of Economics, University of California, Davis.

Hendry, D.F., (1995). *Dynamic Econometrics.* Oxford: Oxford University Press.

Hoover, K.D., (1994). Six Queries about Idealization in an Empirical Context, *Poznzan Studies in the Philosophy of the Sciences and the Humanities*, **38**: 43-53.

Hoover, K.D., (1995). Facts and Artifacts: Calibration and the Empirical Assessment of Real-business-cycle Models, *Oxford Economic Papers*, **47(1)**: 24-44.

Hoover, K.D. and Salyer, K.D., (1996). Technology Shocks or Colored Noise?

Why Real-Business-Cycle Models Cannot Explain Actual Business Cycles, unpublished manuscript, Department of Economics, University of California, Davis, **25**: April.

King, R.G., (1995). Quantity Theory and Econometrics, *Federal Reserve Bank of Richmond Economic Quarterly*, **81(3)**: 53-105.

Kydland, F.E. and Prescott, E.C., (1982). Time to Build and Aggregate Fluctuations, *Econometrica*, **50(6)**: 1345-1369.

Kydland, F.E. and Prescott, E.C., (1991). The Econometrics of the General Equilibrium approach to Business Cycles, *Scandinavian Journal of Economics*, **93(2)**: 161-178.

Kydland, F.E. and Prescott, E.C., (1996). The Computational Experiment: An Econometric Tool, *Journal of Economic Perspectives*, **10(1)**: winter, 69-86.

Lakatos, I., (1970). Falsification and the Methodology of Scientific Research Programmes. Imré Lakatos and Alan Musgrave (eds.). *Criticism and the Growth of Knowledge*. Cambridge: University of Cambridge Press.

Solow, R.M., (1957). Technical Change and the Aggregate Production Function, *Review of Economics and Statistics*, **39**: 312-320.

Watson, M.W., (1993). Measures of Fit for Calibrated Models, *Journal of Political Economy*, **101(6)**: 1011-1041.