Help for the Regional Economic Forecaster: Vector Autoregression

Paul A. Anderson

Economist
Research Department
Federal Reserve Bank of Minneapolis

- How fast will prices in the city or county rise in the next two years?
- How much will the state personal income tax base grow next year?
- What will the region's unemployment rate be next quarter?
- Would a state tax cut spur new business growth?
- What would be the effects of city-wide rent control?

These questions represent two general types of questions which public and private decision makers face every day: What will happen to the regional economy if current economic policies don't change? And what will happen if policies do change? These are not easy questions to answer, and increasingly people are turning to regional econometric models for help. But they shouldn't expect much help—at least not from models designed in the usual way.

The technique used in most regional and national models is designed mainly to answer the second kind of question, to analyze the effects of alternative economic policies. This technique has, however, proven untrustworthy for that task, and the theory it is based on has been seriously challenged. Since the technique is not designed primarily to produce accurate forecasts under current policies, it is probably not the best way to answer the first kind of question either. And for regional analysts it's almost surely not the best. Because of its heavy reliance on economic theory, the standard technique has trouble forecasting well or often with the fragmentary economic data typically available at the regional level.

Regional decision makers, therefore, should look for other techniques to help answer regional economic questions. Unfortunately, no good alternative is available yet for policy analysis, but there is one for simple forecasting: vector autoregression (VAR). The VAR method overcomes many of the standard technique's defects. It is designed especially to forecast, and it's based on statistical regularities, not economic theories. So, as our model of the Ninth Federal Reserve District demonstrates, VAR seems capable of producing regional forecasts which are, compared to the standard kind, more accurate, more frequent, and cheaper.

The standard way has its problems . . .

Regional econometric modeling is fairly new in the United States, but it has grown quickly, largely by imitating the development of national models. Before 1970, there were very few regional models; most people were more interested in building models of the U.S. economy as a whole. But since 1970, computer models of regional economies have sprung up all over. Today, there are operating models of cities (for example, Los Angeles, New York, and Philadelphia), counties (Luzerne County, Pennsylvania), states (Alaska, California, Colorado, Ohio, Oklahoma, Massachusetts, and Wisconsin), and multistate regions (the northeastern corridor of the United States).[1]

The technique which regional models have imitated is usually called the *structural* approach. In order to project the future course of the economy, this approach attempts to use economic theory and historical data to recreate the structure of the econ-

[1]For a recent survey of regional models, see Lawrence R. Klein and Norman J. Glickman's Econometric model-building at regional level in *Regional Science and Urban Economics*, vol. 7, no. 1 (March 1977), pp. 3-24.

omy as a system of equations. Which economic variables affect which and how is determined primarily by theories of how individuals, firms, and governments behave. The relationships among the variables are estimated to fit the historical data but restricted by what economic theories imply about what those relationships should be.

Building the restrictions of economic theory into the structure of an econometric model is very hard, but it is necessary for the model to be able to project the course of the economy under alternative government policies. This kind of prediction can't be done by just examining the historical data for what happened when the policies were used before; there are too many possible policies, and the economy is just too complicated. To clearly and accurately isolate and compare the effects of many different policies on many different economic variables, one needs to model the economy with detailed restrictions from economic theories, as the structural method does.

This structural modeling strategy was developed by the builders of the large national econometric models in the 1950s and 1960s. Since those national efforts appeared to be quite successful, it is not surprising that the regional modeling efforts of the 1970s proceeded along similar lines. But today there are good reasons to believe that imitating national models is not the most effective regional modeling strategy.

No one should ignore the national structural models' recent poor performance: they did not predict and could not explain the simultaneous high inflation and unemployment rates of this decade. In fact, a growing number of economists think this is evidence that current structural models are useless as policy analysis tools.[2] These economists argue that today's models have failed because they do not incorporate sufficiently sophisticated theories of economic behavior. Current models analyze the effects of policy changes under the assumption that the structure of the models (and of the economy) will not be affected. A more complete theory predicts, however, that people's behavior, and hence the economic mechanism, *will* change in response to changes in economic policies. (The criticisms of this group have received most public attention as the *rational expectations* criticisms.)

. . . especially with regional forecasting

In its current form, the structural technique thus seems unreliable to do the job it was mainly designed for, but it may still be useful for simple forecasting. Yet it is quite possible that a modeling technique designed specifically for forecasting would be able to do that job better. And regional model builders have another very good reason to look for such a new technique: The structural method is hard to apply at the regional level.

The structural method, with its basis in economic theory, needs complete data to forecast well. That is, it needs enough data to satisfy its theory. If a model's theory says a certain kind of economic indicator helps determine another, for example, and the determining indicator is not measured, forecasting accurately will obviously be difficult.

National models have few problems of this type because a large, detailed national economic data set is available. But regions of all sizes face the situation often; compared to the U.S. data, their data are incomplete. Some economic concepts are not measured at all in cities, counties, or states. Either no one has bothered to measure them, or the concepts are not measurable at that level. Some concepts may be measured in only parts of a region—in some of a state's counties but not others, for example. Many concepts are measured only infrequently, usually only once a year instead of quarterly or monthly. And data may be reported with a long lag: the most current year available may be not last year but three years ago. Finally, since regional economics is a relatively new field, many of the regional data series are very short: they have only been measured for a few years.[3]

Structural models for regions can and have been developed and used despite this fragmentary data, of course, but the forecasts they produce can suffer quite a bit. Most regional model builders choose to build annual models because annual data sets are the most complete available. This means, however, that they can only forecast once a year, not the most useful frequency for anyone who has to make decisions more often than that. This also means that the forecasts from most regional models can lose a lot of accuracy. Many times the annual data set, though the

---

[2]For a good summary of this viewpoint, see Robert E. Lucas, Jr., and Thomas J. Sargent's After Keynesian macroeconomics in the *Federal Reserve Bank of Minneapolis Quarterly Review*, Spring 1979, pp. 1-16.

[3]For another discussion of special regional data problems, see Klein and Glickman.

most complete available, is still not complete enough. For example, it still may not have some data for parts of the region or up-to-date data for some concepts.

The annual data set and its incompleteness force regional model builders to simplify their efforts and so can reduce their accuracy further. They must reduce the detail in their theories to fit the sketchier data, thus making the model less able to recreate the structure of the economy and predict accurately how economic agents will behave. They must throw out the partial monthly and quarterly data which could be used to produce more precise annual numbers if the theories behind the structural technique could handle different frequencies easily. And regional model builders using annual data must necessarily use all short series—typically, at most, only 25 observations. This severely limits how complex the interactions of the variables can be and so how close the model can come to an accurate forecast.

To get more complicated interactions, structural modelers generally interrelate regional and national variables, but the way many do this is not a good way to improve forecasting accuracy. Because directly incorporating these interrelationships is difficult, regional models are simply linked to large, detailed national models, so that the forecasts of national economic variables become input for the regional models. Regional economic variables are then calculated as simple functions of their national counterparts. Unless a region is really a scaled-down version of the nation, however, this procedure will distort regional forecasting; the interactions among variables will reflect too much of the nation's economic relationships and not enough of the region's.

For regional models, there's a better way

Regional model builders do not have to accept such unsatisfactory situations. It is true that so far no good alternatives to the structural technique for analyzing the effects of alternative government policies have been developed. But there is a good alternative for forecasting under current policies: *vector autoregression* (*VAR*).[4] This technique is designed especially to forecast, so it has a good chance of forecasting better than the structural method. And because of its design, VAR should definitely be better at regional forecasting.

VAR is a straightforward, powerful, statistical forecasting technique which can be applied to any set of historical data. Like the structural technique, it produces a system of equations which can project the future paths of economic variables using their historical data. Unlike the structural technique, however, the VAR method can be used to construct equations based entirely on regularities in the data themselves, not at all on economic theory.

Why such a technique can improve regional forecasting may be obvious. Very simply, because VAR doesn't depend on economic theories, it doesn't require complete data to forecast. All it needs is a collection of numbers (a *vector*) to relate to each other and to their pasts (*autoregression*).

This means VAR models can forecast much more often than structural models. Incomplete regional data do not restrict VAR models to annual forecasts. These regional models can forecast annually or quarterly or monthly, using whatever data are available for each frequency.

And these VAR forecasts should be more accurate than any structural models can produce. Regional VAR model builders will still be stuck with short, lagging, and missing data series, but they can use whatever they have more effectively. With VAR, regional model builders can easily use monthly and quarterly numbers to complicate the interactions of otherwise short series and so make their annual forecasts more precise. They can also increase the complexity of the models so that forecasts of any frequency will be more precise. This they can do by interrelating regional and national variables—and without the distortion structural modelers encounter. Regional VAR models need not become simple functions of U.S. models; the national variables can easily be built right into the system, just as additional influences on the regional variables.

An extra feature of VAR which regional model builders should appreciate is its cost: compared to structural models, VAR models should be much cheaper.

Generally, the costs of building and running an econometric model come from the people and com-

---

[4]Though the term *vector autoregression* can be applied to a broad class of techniques, we shall use it to represent a specific economic forecasting strategy developed by Christopher A. Sims and Robert B. Litterman. For a technical description of this approach and the key references, see the article by Thomas J. Sargent in this *Quarterly Review*.

puter time it needs. This is determined mainly by the size of the model and how complicated its computations must be. The smaller and simpler, the less time and money demanded.

Not being tied to economic theory is what should give VAR models the cost advantage over structural models. VAR models don't have to have all the variables which theory requires for structural models to work well, so VAR models should be relatively smaller. They don't need to relate the variables in the complex way economic theory prescribes, so they don't have to use as expensive procedures to estimate or solve the equations as structural models do. VAR models also don't require the repeated experiments to construct those relationships, which vague theories force on structural model builders. And regional VAR model builders don't have to spend any extra time and money trying to complete their data sets.

A Demonstration of Regional VAR Forecasting

We have constructed and used a VAR forecasting model for the Ninth Federal Reserve District,[5] and it seems to corroborate our claims for the VAR technique as a regional forecaster.

In its construction, our Ninth District model[6] is quite unlike a structural model. Our VAR model forecasts only five key measures of economic activity in the district—total employment, the labor force, personal income, retail sales, and the consumer price index—quite a small data set compared to what a structural model would require. As Table 1 shows, none of these indicators are available for all parts of the district, a point which would also handicap the structural technique. Our model is not confined to annual data, either. While VAR models can easily be built to use even more frequent data, for this demonstration we have chosen to build one which uses quarterly data. Our VAR model is not attached to a national model for complexity, as regional structural models often are, but it does not ignore national influences. To improve its accuracy, our model includes the influences on the region of these national forces: total employment, the labor force, the consumer price index, and the gross national product, adjusted for inflation.[7]

We have tested the forecasting accuracy of our model by having it predict regional conditions during a time period for which we have actual data to check its forecasts against. First the model's equations were estimated using data from 1960 through 1971. Then

[5]The Ninth Federal Reserve District consists of Minnesota, Montana, North and South Dakota, northwestern Wisconsin, and the Upper Peninsula of Michigan.

[6]Technical descriptions of the Ninth District VAR model's equations and restrictions are available on request to the author.

[7]However, regional variables have little effect on national conditions in our VAR model. This specification is based on formal tests described by Paul A. Anderson in A test of the exogeneity of national variables in a regional econometric model, Research Department Working Paper 124, Federal Reserve Bank of Minneapolis, Minnesota, June 1979.

Table 1

Geographical Coverage of Regional Data Used in VAR Model of the Ninth District

	Minnesota	Montana	North Dakota	South Dakota	NW Wisconsin	Upper Michigan
Civilian Employment	●	●	●	●		●
Civilian Labor Force	●	●	●	●		●
Personal Income	●	●	●	●		
Retail Sales	●					
Consumer Price Index	*					

*Minneapolis-St. Paul metropolitan area only.

the model was used to forecast. In each quarter from the first quarter of 1972 through the fourth quarter of 1976, the model forecasted the next eight quarters using only the economic information that would have been available at that time. (Such forecasts are called *ex ante* or before the fact forecasts.) Finally, we compared the model's forecasts to the actual reported levels.

Table 2 shows how far from the actuals our VAR's forecasts were on average. Since in each quarter the model used only data that would have been known in that quarter, these averages represent the size of the forecast errors that could be expected if the Ninth District VAR model were used today to forecast regional economic trends. Not surprisingly, this VAR model finds it harder to predict farther into the future. The average forecast errors for all of the regional variables are larger the longer the forecast horizon. The model seems to be most accurate at projecting labor market conditions and inflation.

How well does our VAR model do compared to structural regional models? This is somewhat hard to tell because structural model errors can't be computed as realistically as ours were. In two different ways, their errors are underestimated.

One way structural errors are underestimated is by calculating them for the same time period used to build the models. This is unavoidable because of the shortness of the annual data sets structural models use. There isn't enough data to use part of it to build the models and the rest to test the accuracy, as we did: all the data must contribute to model construction. The only way to test these models' accuracy, therefore, is to have them forecast some of the same years the models were designed explicitly to forecast well. The errors detected in such forecasts clearly will be smaller than can be expected in forecasts of other periods.

The other way structural errors are underestimated is by calculating them as *ex post* or after the fact errors. Rather than using only data which would have been known at the time of each forecast, as we did, most regional structural modelers use some data that could not have been known. In order to forecast, structural models require future values of national and regional policy variables. In order to realistically estimate forecast errors with historical data, therefore, modelers would have to figure out what forecasters really knew about future policy variables years ago. Since this is virtually impossible, modelers substitute actual levels for the future levels of these variables, as though forecasters had perfect foresight. Obviously, then, the errors in the resulting forecasts will typically be smaller than those of forecasts like ours, made with more realistic data.

To minimize this second distortion, we shall compare structural models' annual average errors to our VAR model's average errors for the time period it knew the most about: one quarter ahead. Table 3 shows this kind of comparison with errors of a structural model of the Philadelphia area and a group of seven regional structural models. Despite the bias against it, our Ninth District VAR looks much more accurate. Its one-quarter-ahead forecast errors are distinctly lower than both structural errors for all the regional variables the models have in common. In fact, the structural errors are about the same size as our four-quarter-ahead forecast errors, a comparison which puts our VAR model at a greater informational disadvantage.

The costs of regional econometric models are much easier to compare than their accuracy, and our Ninth District VAR appears to have a considerable advantage here. Since our model is so small, the cost of data collection was negligible. Constructing the VAR took roughly 50 worker hours, 2 weeks of calendar time, and $200 of computer time. And running a quarterly forecast on the computer costs less than $5.

Table 2

Average Forecast Errors*
of Ninth District VAR Model

	1 quarter ahead	4 quarters ahead	8 quarters ahead
Consumer Price Index	0.6%	1.9%	4.5%
Civilian Labor Force	0.7	1.5	3.6
Civilian Employment	0.8	1.7	2.4
Personal Income	1.4	3.6	5.0
Retail Sales	4.8	5.3	7.5

*Mean absolute value of the difference between the forecasted and actual levels as a percentage of the actual level. In each quarter of 1972 through 1976, the next eight quarters were forecasted.

Table 3

Comparing Average Forecast Errors*
of Regional Structural and VAR Models

	Ex post, inside sample errors of annual structural models		Ex ante, outside sample errors of Ninth District VAR model	
	Philadelphia region	Average of 7 regions	1 quarter ahead	4 quarters ahead
Consumer Price Index	3.2%	—	0.6%	1.9%
Civilian Labor Force	1.1	—	0.7	1.5
Civilian Employment	1.7	1.3%	0.8	1.7
Personal Income	6.7	3.4	1.4	3.6

*Mean absolute value of the difference between the forecasted and actual levels as a percentage of the actual level.

Sources: Two papers by Norman J. Glickman:
An econometric forecasting model for the Philadelphia region. *Journal of Regional Science*, vol. 11, no. 1 (April 1971), p. 25 (Table 2);
Son of 'The specification of regional economic models.' *Papers of the Regional Science Association*, vol. 32 (1974), p. 165 (Table 2).

It's an understatement to say that most of today's regional structural models cost much more than this. A recent example is the large New York Econometric Model which reportedly took about 2 years to develop at a cost of about $300,000.[8]

An Attractive Technique—for Everyone?

Our regional model demonstrates how VAR models may be useful for more than simple forecasting, too, just as structural models have been. Comparing incoming data with our model's forecasts has given us cues to economic developments in the area; when the forecasts have been off substantially, we've found that something worthy of detailed investigation is likely to be happening. Through simulation experiments, the model has helped assess the likely regional impacts of particular national developments—for example, what effect a U.S. recession would have locally. And, as reliable regional economic theories appear, we plan to build their restrictions into our VAR model so that the impacts of various policies can be evaluated.

Regional economic analysts with questions about the future, therefore, should consider vector autoregression a serious alternative to the structural modeling technique. Despite incomplete regional data, VAR models seem to be able to give them more frequent, more accurate, and cheaper answers.

But model builders without data problems should not ignore VAR. U.S. decision makers can't trust the structural method for help with policy analysis any more than regional decision makers can. And using the structural method for simple prediction is also extremely costly at the national level compared to VAR. Tests have shown, furthermore, that a small national VAR model could have done as well as or even better than the large national structural models in predicting recent U.S. economic activity.[9] Any model builders may thus find VAR an economical and accurate alternative.

[8]See Slow job growth for 1980s seen in New York econometric study, *New York Times*, April 3, 1979, p. A1.

[9]See Robert B. Litterman's Techniques of forecasting using vector autoregressions, Research Department Working Paper 115, Federal Reserve Bank of Minneapolis, Minnesota, February 1979.

The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.