FORECASTING STATE LEVEL ECONOMIC ACTIVITY: AN ERROR CORRECTION MODEL WITH EXOGENOUS NATIONAL STRUCTURAL FORECAST COMPONENTS

Michael Hicks, Ball State University*

INTRODUCTION

Projecting economic activity at the state and substate level is an area of considerable interest to policymakers. The use of economic forecasts for budget preparation, economic development, and workforce investment planning efforts constitute just a few of the applications of formal economic modeling in regions. While many state governments maintain economic forecasting capacity, few substate regions outside very large metropolitan areas retain a forecasting capacity. Further, there are a number of states where the forecasting of economic activity is outsourced from the executive branch to universities and private sector forecasting firms. The U.S. Department of Labor also provides some limited labor market forecasts through state agencies as a part of the Labor Market Information programs.[1] Despite the abundance of forecasting tools that have been developed over the past two decades, significant demand for low cost, state, and substate economic forecasts remains strong.

This paper develops a method for projecting state and substate economic activity by combining large-scale structural models with low cost, flexible but accurate vector error correction models. The emphasis of this work is in the practical nature of the forecasting model. I proceed with an explanation of the modeling process, with a brief description of the national forecast model I employ. I then outline the error correction model (ECM) with exogenous variables, its structure, and econometric considerations. I conclude with several forecast performance metrics and a summary of the paper.

THE NATIONAL MODEL

Several high-quality national forecasts currently exist. These include the well-known commercial models offered by Global Insight and Standard and Poor's, as well as academic models such as the well respected FAIRMODEL produced by Yale University's professor Ray Fair.

I employ the FAIRMODEL for my estimates. Researchers in states or regions with access to commercial models, or locally produced models may opt to use these models. The procedures I describe here are largely invariant to the choice of exogenous variables employed in the ECM model of states and regions.

The FAIRMODEL is a solid choice as a national forecast model for state and local predictions for three reasons. First, it is provided as a free model. The ease of downloading and application is superb. Second, it enjoys a long history of accurate national forecasts. It has been widely used in research and forecasting settings, and remains the most respected private forecast model. Finally, the model permits significant flexibility in assumptions. For example, there are five monetary policy assumptions which may be employed in the forecast.[2]

The FAIRMODEL then seems a logical approach under a common cost function criterion for prediction models. This is particularly true where the forecast for state and local public policy acts as an economic baseline rather than as a budgetary tool. The model results I display in a later section suggest it also performs well on other criterion as well.

The FAIRMODEL employs a 30-equation stochastic representation of the U.S. economy which is frequently recalibrated. The model uses a 2SLS estimation procedure, with over 100 identities, and thus a significant choice in potential variables for use at the state and local area. The model has provided 98 separate forecasts since its inception, with a mean absolute error of 1.07 percent across all predicted variables. The most recent published forecast error for gross domestic product (GDP) growth is displayed in Table 1.

THE VECM MODEL

The choice of a vector error correction model (VECM) for state level forecasts provides a

* I would like to thank participants of the 62nd annual conference of the *Association of University Business and Economic Research*, session on Forecast Model Performance, especially Gary Horvath, Martin Shields and David Keyser.

Table 1
Forecast Performance of FAIRMODEL, U.S. GDP Growth

	2006:Q2	*2006:Q3*	*2006:Q4*	*2007:Q1*	*2007:Q2*	*2007:Q3*	*2007:Q4*	*2008:Q1*	*2008:Q2*	*2008:Q3*
Actual	2.7	0.8	1.5	0.0	4.8	4.8	-0.2	0.9	2.8	-0.3
2006:Q2	3.6	2.9	2.7	2.5	2.4	2.4	2.4	2.5	2.6	2.6
2006:Q3		1.6	2.3	2.4	2.4	2.6	2.6	2.7	2.7	2.7
2006:Q4			2.1	2.6	2.7	2.7	2.7	2.7	2.7	2.7
2007:Q1				3.2	2.9	2.6	2.6	2.5	2.6	2.6
2007:Q2					3.2	3.0	2.9	2.8	2.8	2.7
2007:Q3						4.1	3.7	3.4	3.1	2.8
2007:Q4							3.4	3.2	3.1	2.8
2008:Q1								3.1	2.5	2.7

Source: Fair (2008, Table 4).

relatively straightforward platform. Suppose the following vector autoregression representation:

$$\boldsymbol{Y} = \boldsymbol{\Phi Y} + \boldsymbol{e},$$

where the lagged endogenous variables include an as yet unspecified number of lags, which are referred to predetermined variables. Obtaining first differences of this representation provides us with:

$$\boldsymbol{Y} - \boldsymbol{Y} = \boldsymbol{\Phi}\ (\boldsymbol{Y} - \boldsymbol{Y})\boldsymbol{e},$$

in which the as yet undetermined number of lags are of the first differences of the endogenous variables. This model is now integrated by an order, or I(1). This step is necessary to later construct the error correction model, but a common approach at this juncture is to test these series for non-stationarity. If the times series are not stationary with I(1), the series will require subsequent differencing to construct a stationary time series from which to forecast.[3] While it is not uncommon for time series to require higher order integration prior to obtaining a unit root, the types of series employed in state and local forecast models will typically be I(1).[4]

We seek for our forecast a cointegrating equation(s) for the variables *Y* in the VAR representation above. A cointegrating equation is a linear combination of two or more equations which may consist of a constant, trend, and error component. Before describing in some technical detail the cointegrating equation, it is helpful to review the economic explanation for the equation and its use.

The presence of nonstationary variables imposes a real constraint on the forecaster. The absence of a unit root suggests that results from a direct modeling effort will yield a spurious regression. Differencing these variables to I(1) or greater, while providing a series that is stationary (possesses a unit root) will result in significant losses in economic information. This is particularly true regarding a long-term equilibrium condition between the variables of interest (say inputs and outputs). The relationship may be reestablished through the use of another equation that outlines equilibrium relationships between the time series. This equilibrium equation can include a stochastic element.

For a forecaster, this offers an especially useful representation, since it offers both a tractable econometric model and an attractive economic explanation based on economic theory. What remains is in evaluating and testing the number and type of cointegrating relationships.

If there are z equations in the VAR representation, and the cointegrating rank of the VAR is p, there are up to z-p linear combinations of the variables, which are referred to as cointegrating equations. Also, $p \leq z$-1. A shortcut to estimating the cointegrating rank is possible through simple hypothesis test of z-1 cointegrating equations. Cointegrating equations which do not exhibit significance in a standard t-test should be discarded. This should be cautioned against in some settings, but in state or local forecasts of economic activity, the potential error would seem miniscule relative to the cost.

The specification choice for a cointegrating equation includes a constant, trend, and exogenous variables. Suppose two variables share a common trend and intercept then:

$$y_t = \alpha + \beta t + e_t^y$$
$$z_t = \lambda + \delta t + e_t^z.$$

Then the cointegrating relationship can be constructed as:

$$w_t = (\alpha + \Pi\lambda) + (\beta + \Pi\delta)t + e_t^y + \Pi e_t^z,$$

where Π is referred to as the cointegrating vector. The single equation extract of the error correction model takes the form:

$$\Delta y_{\mathbf{t}} = \mathbf{\Phi}(\Delta\mathbf{Y}) + (\alpha + \Pi\lambda) + (\beta + \Pi\delta)t + e_t,$$

which is the I(1) VAR representation with the cointegrating equation for variable z represented by the cointegrating relationship and the VAR construct.

The inclusion of exogenous variables into this relationship does not alter the estimation process unless they comprise a component of the cointegrating equation. In practice I have not observed exogenous variables as part of the cointegrating equation in regional forecasting models.

Touched briefly on, but left unanswered as yet is the determination of the optimal lag length selection for $\mathbf{\Phi}(\Delta\mathbf{Y})$. The most common approach is to employ an information criterion selection process. The Akaike Information Criterion or Schwarz-Bayesian methods are commonly employed.[5] Testing various lag lengths and choosing that which minimizes the information criterion is the usual technique. This leaves us with a single representative model.

The model we construct employs quarterly state data from 1990 through the second quarter 2008. The series we forecast are real personal income, earnings at place of work, and earnings in the manufacturing, wholesale, retail, construction, finance, information, and health care sectors. We employ national GDP as the exogenous variable, using history and forecasts from the FAIRMODEL's 2nd quarter 2008 forecast. All variables were differenced and we were unable to reject a unit root at the 1 percent level for the I(1) series using the augmented Dickey-Fuller test.

We assumed initially eight cointegrating equations, but rejected two, leaving us with eight cointegrating vectors. We included both an intercept and trend, but no exogenous variables in the cointegrating equation. The optimal lag lengths were generated for the entire VECM by minimizing the Akaike Information Criterion.

In Table 2 these diagnostics offer mixed results for this model, but are strikingly similar to results obtained in other similar modeling efforts employing ECM with exogenous variables. Importantly, sample model performance does not include the standard for forecasting models. Rather, sample predictive ability should have the objective of forecasting. To evaluate this, we now turn our attention to model performance on earlier estimates of this type of model.

MODEL PERFORMANCE

I have prepared four separate forecasts of state level economic activity using this approach—two in West Virginia, and two in Indiana. The West Virginia models are derived from the West Virginia Econometric Model (Hicks and Simpson, 1999) and consist of a draft with test diagnostics and a population forecast model. The out-of-sample performance of each are reported in Table 3.

Table 2
Selected Model Diagnostics

	PI	*Earnings*	*Manufac*	*Whole*	*Retail*	*Finance*	*Inform*	*Constr*	*Health*
Log likelihood	-247.61	-231.80	-208.27	-46.97	-70.24	-100.15	-2.49	-79.79	-61.43
Akaike AIC	8.48	8.02	7.34	2.67	3.34	4.21	1.38	3.62	3.08
Schwarz SC	9.94	9.48	8.80	4.12	4.80	5.66	2.83	5.07	4.54
Mean dependent	10.98	7.89	1.15	0.46	0.28	0.43	0.09	0.46	1.21
S.D. dependent	16.90	14.76	8.83	1.04	1.40	2.41	0.60	2.00	1.65
t-statistic on Exogenous variable (U.S. GDP)	14.88	11.83	8.41	0.81	1.14	1.76	0.43	1.31	1.00

Table 3
The West Virginia Econometric Model, out-of-sample performance

Beta Test Results: percentage error in out-of-sample forecasts of employment levels

	Non-Farm Employment	*Con*	*Man*	*Durable Goods*	*Non-Durable Goods*	*TCPU*	*Wholesale Trade*	*Retail Trade*	*FIRE*	*Unemployment Rate*
1998:1	-0.54%	-1.62%	-3.06%	-3.41%	-2.54%	-1.37%	-1.23%	-0.22%	0.51%	24%
1998:2	1.14%	3.99%	0.23%	0.67%	-0.36%	-0.23%	2.35%	0.26%	-0.67%	26.84%
1998:3	0.82%	-1.11%	0.64%	1.30%	-0.32%	0.20%	-0.07%	0.77%	0.12%	-2.11%
1998:4	-0.99%	0.52%	-4.68%	-5.82%	-2.97%	-1.34%	-0.47%	-0.83%	-1.42%	38.52%
1999:1	0.19%	-0.22%	0.21%	0.15%	0.29%	-0.38%	-0.48%	0.20%	-3.23%	27.40%
1999:2	1.70%	6.77%	-2.36%	-3.33%	-0.95%	2.56%	0.88%	-0.19%	0.14%	25.20%
1999:3	1.92%	-0.91%	-0.17%	-0.13%	-0.23%	0.77%	0.54%	0.77%	-0.10%	19.90%
Average	**0.61%**	**1.06%**	**-1.31%**	**-1.51%**	**-1.01%**	**0.03%**	**0.22%**	**0.11%**	**-0.67%**	**22.79%**

Table 4
West Virginia Econometric Model Population Forecast, 2001-2006

2001	2002	2003	2004	2005	2006
1.25%	**1.17%**	**1.01%**	**0.83%**	**0.88%**	**0.76%**

Actual population forecast performance is available from this model, as the result of a long-term forecast performed in 2000. These results appear in Table 4.

Our experience in Indiana includes the annual Indiana Labor Market Forecast©, which began in 2007. This forecast employs data from the Quarterly Workforce Indicators into a time-series model of employment dynamics. The four quarters of experience we have established yield robust model performance results in what are among the most volatile economic variables to forecast. See Table 5.

These results offer a much more hopeful prognosis of this approach. One additional estimate is an out-of-sample prediction of the Indiana model detailed in the section above. This model performance was compared to the results of an official state level forecast in Indiana. See Table 6.

Table 5
Indiana Labor Market Forecast Performance

	Employment	*New Hires*
2006:Q3	-0.04%	3.52%
2006:Q4	-0.11%	8.46%
2007:Q1	-0.57%	1.34%
2007:Q2	0.36%	4.86%

SUMMARY

This paper outlines a tractable, low-cost modeling hybrid for state and local economic forecasting. The model relies on an external forecast to develop exogenous variables for an error correction model. We then detail the development of the ECM, diagnostics for model specification, and finally an example model of Indiana. We also include model performance for four of these models (two in West Virginia and two in Indiana). We also compare these to a State Budget Agency forecast in Indiana.

The VECM models offered superb overall forecasting performance in a series of out-of-sample forecast comparisons. In personal income estimates, in West Virginia, over six quarters from 1999-2001 the average error was -0.31 percent. Population estimates from 2001 through 2006 at the aggregate state level yielded an average error of 0.98 percent. The model applied to the State of Indiana showed a 4-quarter error of .09 percent in employment and only 4.55 percent in the highly

Table 6
ECM vs. State Budget Agency 2007:1 to 2008:2

		ECM		*SBA*	
GDP	*Actual*	*Forecast*	*MAD*	*Forecast*	*MAD*
2007:Q1	13510.9	13543.104	0.24%	13,613.40	0.76%
2007:Q2	13737.5	13757.527	0.15%	13,769.10	0.23%
2007:Q3	13950.6	13958.6	0.06%	13,943.00	-0.05%
2007:Q4	14031.2	14073.465	0.30%	14,126.50	0.68%
2008:Q1	14150.8	14295.178	1.01%	14,314.30	1.16%
		Average Error	**0.62%**	**Average Error**	**0.75%**
		ECM		*SBA*	
Personal Income	*Actual*	*Forecast*	*MAD*	*Forecast*	*MAD*
2007:Q1	11,451,855	11,382,780	-0.60%	11,244.70	-1.81%
2007:Q2	11,568,700	11,543,620	-0.22%	11,380.20	-1.63%
2007:Q3	11,722,750	11,712,370	-0.09%	11,542.00	-1.54%
2007:Q4	11,867,043	11,854,420	-0.11%	11,696.70	-1.44%
2008:Q1	12,002,122	12,011,010	0.07%	11,862.30	-1.16%
		Average Error	**-0.16%**	**Average Error**	**-1.26%**

volatile new hire series from the Quarterly Workforce Indicators series. As a comparison of this model against official state forecasts, the VECM error was only 82.7 percent of the official state forecast, with an aggregate error over six quarters of 0.62 percent for State GDP. State Personal Income was also forecast during the same period, with an error only 12.7 percent of the official forecast.

The results of these models are promising for state and local economic forecasts. The techniques employed here are not new, but their application to the very real problem of timely and accurate low-cost forecasts for state and local policymakers is rare.

Notes

1 See for example the LMI data at Indiana's Division of Workforce Development.

2 The complete reference is available at Fair (2004). The entire model in electronic form, detailed appendices and data, an analysis of model performance and other reference material is available at http://fairmodel.econ.yale.edu/.

3 The stationarity test most commonly available in software packages is the Augmented Dickey-Fuller test. Some packages also include the Phillips-Peron test. See Greene (2003).

4 These variables are likely to include employment, personal income, state GDP, unemployment rates and earnings by industry. Many of these will be stationary in levels, especially in slower growing regions. However the first differencing of these variables, while not necessary, is a commonly applied approach, and consistent with the intent of the error correction model.

5 Akaike (1974) and Schwarz (1978).

References

Akaike, Hirotugu. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19 (1974): 716–723.

Fair, Ray.
Estimating How the Macroeconomy Works. Cambridge, MA: Harvard University Press, 2004.
The Forecasting Record of the U.S. Model. New Haven, CT: Yale University, 2008.

Greene, William. *Econometric Analysis,* 5rd ed. Upper Saddle River, NJ: Prentice Hall, 2003.

Hicks, Michael and Marc Simpson. The West Virginia Econometric Model. Huntington, WV: Marshall University, 1999.

Schwarz, Gideon. Estimating the Dimension of a Model. *Annals of Statistics* 6 (1978):461-464.