An Unobserved Components Forecasting Model of Non-Farm Employment for the Nashville MSA

Joachim Zietz and David A. Penn
Middle Tennessee State University - USA

Abstract. The study demonstrates how unobserved component modeling, also known as structural time series modeling, can be usefully applied to forecast non-farm employment for the Nashville MSA. Short-term out-of-sample forecasts are provided for total employment and its three components: services, construction, and manufacturing. The forecasts are compared to those of a simple vector autoregression. It is shown that the suggested methodology provides very accurate short-term forecasts even in the absence of a full set of independent regressors. In addition, it makes it possible to back out long-term trends, which aid the forecaster in making long-term projections of sectoral employment.

1. Introduction

Local decision-makers must make choices about the future based on the information at hand. For example, a local school system must decide how many teachers and staff to hire based on enrollment expected for the coming school year; a city government may wish to hire additional solid waste management workers if it believes housing construction will increase; and a local contractor may hire additional plumbers and cabinet makers if it anticipates higher demand for housing. In each of these cases, decision-makers must make judgments about economic conditions expected for their communities, but they typically find that information regarding future local economic conditions is difficult to obtain.

This difficulty is not a result of a dearth of literature on what drives local economic growth. The problem is that most of the literature is primarily interested in testing hypotheses about the contributions of national, sectoral, and local economic shocks to growth and in finding useful policy levers to influence the rate of growth.[1] There is much less academic interest in forecasting local economic activity. This preference for hypothesis testing and policy analysis applies even

more to the rapidly growing literature on regional convergence and on regional endogenous growth.[2]

Apart from the many studies focused on structural and policy analysis, there is also some work that is more directly useful for forecasting. Perhaps the most well-known is that of Treyz (1993) on a large-scale input-output based regional economic model.[3] The comparative advantage of large-scale regional modeling systems may not lie in short-term local forecasting but in the evaluation of policy changes at the level of the region and over the medium run. In addition, large systems tend to be costly to maintain and adapt, and they require significant data input without being significantly more accurate in the short run than simpler forecasting methods.[4] As a consequence, decision-makers often prefer forecasts that rely on easy-to-understand ad hoc forecasting methods based on a few select variables that are well understood.

The typical determinants used in simple ad hoc modeling exercises include some regional indicators, such as population growth, and some national indica-

---

[1] Carlino (2003) is a brief overview of the structural literature.

[2] Compare, for example, Doring and Schnellenbach (2006), Johansson et al. (2001), and Martin and Sunley (1998).

[3] The model is known under the name REMI. See http://www.remi.com.

[4] This is one of the reasons why so-called "black-box" models, such as those of the Box-Jenkins type or those of the vector autoregression type, gained so much popularity among forecasters relative to structural regression models.

tors, such as industrial production. One may doubt, however, that these variables, which tend to be readily available, are sufficient for a reliable forecast. More detailed local data may improve forecasting quality. Two problems, however, typically arise in this context: (i) more detailed data may not be available; (ii) there is little theoretical guidance to identify what these missing local data may have to be. Adding deterministic time trends or other deterministic components, such as seasonal or cyclical indicators, may capture some of this ignorance and lead to better within-sample fit. Unfortunately, they may also cause large errors in out-of-sample forecasting applications.

The purpose of this study is to show by way of an example how one may be able to utilize underlying but unobserved components of a time series to improve the forecasting accuracy of simple ad hoc models of local activity. The empirical methodology relies on structural time series modeling (e.g., Harvey 1989, Durbin and Koopman 2001). This technique is also known as unobserved components modeling. As this alternative name suggests, it allows the specification and estimation of model components that cannot be made explicit for lack of data or a viable theory. The methodology is illustrated for a multivariate forecasting model of non-farm employment of the Nashville MSA. Its forecasting performance is contrasted to that of a simple vector autoregression, which is another type of ad hoc modeling approach.

The study is organized as follows. The next section summarizes how the forecasting exercise is embedded in economic theory. Next, the empirical methodology is briefly introduced. This is followed by a description of the data and a discussion of the empirical results. The key points are summarized in a concluding section.

2. Determinants of Non-farm Employment

To forecast employment, the following simple theoretical framework is utilized. Assume that regional output (y) is given by a production function of the type

$$y = y(a, k, n), \tag{1}$$

where a is a measure of total factor productivity, k stands for capital, and n for labor input or employment. Total output can be subdivided by sector, for example into manufacturing (m), services (s), and construction (c). This may be helpful because employment in these three sectors tends to move in different directions in Middle Tennessee, with construction and

manufacturing following a declining trend, and services following a strongly upward trend,

$$y = y_m + y_s + y_c. \tag{2}$$

For each sector specified in (2), a production function along the lines of equation (1) is assumed to exist.

Over time, production technology is affected by relative factor prices. In particular, as total factor productivity and the real wage rise in manufacturing, the employment-output ratio should decline in manufacturing as production is becoming more capital intensive. Low-skill employment will likely shift out of manufacturing into the services. Whether the trend decline in the employment-output ratio for manufacturing will translate into a decline in absolute employment depends on the growth rate of manufacturing. High growth rates, as experienced in Middle Tennessee during the past two decades, may counteract the trend decline in the employment-output ratio and raise absolute employment in manufacturing. The high growth rates experienced in manufacturing in Middle Tennessee are difficult to explain endogenously. To a large extent they are the result of an exogenous event: the decision of domestic and foreign car manufacturers and their suppliers to settle in the region to take advantage of relatively low real wages, low unionization, and a central location.

Changes in the employment level of manufacturing relative to that of the services and construction can be brought about also by a change in the real exchange rate, which is another key relative price that affects the production structure among the sectors identified in equation (2). In particular, a rising real exchange rate acts like a tax on tradable goods and like a subsidy on non-tradables. As most of manufacturing is tradable and most services and construction are not, employment in manufacturing should decline but employment in construction and the services should rise with a rise in the real exchange rate (e.g., Zietz 1996).

Observed output is assumed constrained by the production technology but driven by demand factors. Demand is assumed to be determined by local (y^l) and national factors (y^n),

$$y^d = y^l + y^n \tag{3}$$

The key local factor that is driving demand is assumed to be population. A rise in population should raise employment in the non-tradable sectors services and construction. All other demand factors are national in origin. The following national demand factors are assumed of potential value in forecasting local employ-

ment: industrial production, gasoline prices, import surges from developing countries, such as China, and the yield spread.

A rise in industrial production is taken to increase the demand for products made locally and for employment for a given inter-industry and inter-regional input-output structure. The increase in employment is likely to be most visible in manufacturing. As a derivative to a rise in employment in manufacturing, however, employment in the other two sectors may also rise. Higher gasoline prices are assumed to reduce consumer spending and employment across the board. An import surge from a country such as China may significantly increase the availability of inexpensive foreign substitute products. This can be expected to reduce the demand for locally produced manufacturing products and for employment in manufacturing. The yield spread reflects the impact of monetary policy. It is defined as the difference between a long-term rate, such as the rate for 10-year bonds, and the 3-month t-bill rate. The yield spread is used as a leading indicator by the Conference Board and has been shown to be valuable for forecasting employment for numerous states, including Tennessee (Shoesmith 2003) and also nationally (Estrella and Mishkin 1998, Carlino and DeFina 1999). A rise in the yield spread is assumed to trigger a downturn in economic activity and employment within about three quarters.

3. Empirical Methodology

Following Harvey (1989), a univariate structural time series model can be expressed as

$$y_t = \mu_t + \sum_i \sum_j \alpha_{ij} x_{i,t-j} + \varepsilon_t \qquad \text{for } t = 1,\ldots,T \qquad (4)$$

where y_t is the dependent variable, $x_{i,t-j}$ regressor variable i subject to time lag j, α_{ij} a coefficient associated with variable $x_{i,t-j}$ and ε_t a zero mean constant variance error term. The term μ_t is a time-dependent intercept, which differentiates the model from a simple regression model. The intercept term in equation (4) is specified to follow a random walk process with drift as

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t \qquad \eta \sim \text{NID}(0, \sigma_\eta^2) \quad (5)$$
$$\beta_t = \beta_{t-1} + \xi_t \qquad \xi \sim \text{NID}(0, \sigma_\xi^2) \quad (6)$$

In the context of equations (5) and (6), μ_t can be interpreted as the "level" of a stochastic trend and the drift parameter β_t as its "slope." Both "level" and "slope"

are assumed to follow random walks, with their respective white-noise disturbances η_t and ζ_t independent of each other and of ε_t. This general trend model can be tested down to simpler form, such as a "level" only model, which would be written as

$$\mu_t = \mu_{t-1} + \eta_t \qquad \eta \sim \text{NID}(0, \sigma_\eta^2) \qquad (7)$$

The stochastic trend incorporated by μ_t can be made more flexible (equation 5) or less flexible (equation 7) depending on the complexity of the unobserved trend movements in the dependent variable. Which complexity is needed for a particular case is testable. One would typically start with an over-specified model and test whether a simplified model structure is not rejected by the data. The purpose of the stochastic trend is to capture those trend movements that are not explainable by the regressor variables, which make up the observed components part of the model.

For completeness, it should be mentioned that the model can be extended to include components other than a trend. Cyclical components are feasible, as are seasonal ones. The main purpose of including them is, as in any specification problem, to approximate the data generating process as closely as possible even in the absence of observable variables. All unobserved components can in principle be stochastic or deterministic. If they are stochastic, they are allowed to change over time. If they are deterministic, they have a fixed impact.

The modeling of unobserved components is a second best. Having variables in the model to capture the data generating process is a preferable alternative because it converts the modeling from a semi "black-box" exercise reminiscent of Box-Jenkins time series modeling to one where the driving forces are made explicit and can be given economic content. It should be apparent that identifying an unobservable component may help in this respect. For example, the plot of an unobserved component over time may provide enough clues to identify variables that can capture or at least approximate the time series behavior of the unobservable component.

The three components of employment are estimated with a multivariate model. That means, although three individual equations are specified, one for each employment sector, they are estimated jointly to account for common shocks, covariances, and interactions. These may come about because a decline in one employment segment, for example layoffs in construction, translates into an increase in employment in another, for example in the services, or vice versa.

Because there are relatively few explicit variables, some care needs to be exercised to let the estimates not be overly influenced by outliers. Outliers are points in time that create large residuals. This may happen not only at the level of equation (4) but also at the level of equations (5) or (6). In other words, outliers are possible for the stochastic trend components that are labeled "level" and "slope" in structural time series modeling.

4. Data and Estimation Results

The data are summarized in Table 1. Total non-farm employment is split up into three components: services, construction, and manufacturing. A number of variables are collected according to the discussion of the theoretical background. Not all of them turn out to be statistically significant. For the sake of brevity, the specification search is not detailed here. Rather, the reported results are limited to the preferred model. Independent variables that do not show in the preferred model of Table 2 can be assumed to be statisti-

cally immaterial for the prediction of non-farm employment or its three components. One example for that is the gasoline price.

The results are generated with the software package STAMP 5.0 (Koopman et al. 1995). Similar results can be obtained using the freely available SSF PACK (Koopman et al. 1999) or the UCM procedure of the statistical package SAS (SAS 2004) .

The first set of right-hand side variables in Table 2 [*construction*(-1), Δ*population*] relates to regional variables. The second set [*industrial production*(-3), *interest rate spread*(-8)] consists of national variables. The third set [Δ*Chinese imports*, *major currencies ex-rate*(-1)] contains international variables. Finally, there are observation specific dummy variables. The latter are of two types, those that are large one-month shocks to equation (4) and those that are large one-month shocks to the stochastic level of the equation, which is specified in equation (5). The former are denoted *Irr* in Table 2, the latter are identified as *Lvl*.

Table 1. Data Definitions and Sources

Variable	Definition	Source
services	employment in service sector of Nashville MSA, including public sector, in 1,000s, seasonally adjusted	Current Employment Survey (CES) monthly survey of employers, Bureau of Labor Statistics (www.bls.gov)
construction	employment in construction sector of Nashville MSA, in 1,000s, seasonally adjusted	
manufacturing	employment in manufacturing sector of Nashville MSA, in 1,000s, seasonally adjusted	
population	population in Nashville MSA, in 1,000s; population on July 1 each year interpolated to a monthly figure	U.S. Census Bureau (www.census.gov)
industrial production	index of industrial production	Fred, series INDPRO
interest rate spread	10 year treasury rate constant maturity – 3 month t-bill rate (secondary market)	Fred, series GS10 – series TB3MS
Chinese imports	U.S. Imports from China, Mainland, Customs Basis	Fred, series IMPCH
major currencies ex-rate	Trade Weighted Exchange Index: Major Currencies	Fred, series TWEXMMTH
gasoline price	Spot Oil Price: West Texas Intermediate	Fred, series OILPRICE

Notes: Fred stands for Federal Reserve Bank of St. Louis Economic Data Base (http://research.stlouisfed.org/fred2/).

Table 2. Estimation Results for Multivariate Model of Employment by Sector

Variables	Services		Construction		Manufacturing	
	coefficient	p-value	coefficient	p-value	coefficient	p-value
construction(-1)	-0.599	0.043			0.333	0.007
Δ*population*			0.805	0.003		
industrial production(-3)	1.012	0.000			0.215	0.008
interest rate spread(-8)					0.332	0.037
Δ*Chinese imports*	-0.0003	0.040				
major currencies ex-rate(-1)			0.030	0.020		
Irr 1994.10	-5.730	0.000				
Irr 1999.1			-0.876	0.000		
Lvl 1995.4					-2.536	0.000
Lvl 1999.4			1.111	0.000		
Lvl 2001. 1					3.486	0.000
R^2	0.9994		0.9973		0.9914	
p-values for:						
Durbin-Watson	0.7087		0.4962		0.3895	
Autocorrelation at lag (1)	0.7393		0.5009		0.4011	
Autocorrelation at lag (12)	0.7503		0.3689		0.9785	
Box-Ljung Q statistic	0.4758		0.0125*		0.8839	
Bowman-Shenton normality	0.9942		0.0001*		0.0037*	
Heteroskedasticity	0.8695		0.9985		0.9169	
One-step ahead out-of-sample predictive test						
12-period χ^2	0.9949		0.9989		0.9924	

Notes: Monthly data cover period 1990. 9 - 2005. 6. (T = 178); (-*t*) identifies a lag of *t* months; Δ is the first-difference operator; *Irr* stands for an observation specific dummy variable for the given date; *Lvl* is an observation specific dummy variable for the stochastic trend; * indicates rejection of the null of no problem at 5 percent level or better.

The three employment equations are not equally well determined. The equation for construction reveals some evidence for autocorrelation at the 5 percent level, although not at the 1 percent level. The equations for construction and manufacturing do not pass the test for normality of residuals at standard levels of significance. More observation specific dummy va-riables or the inclusion of additional variables may be able to remove this problem although at the expense of making the model more complicated or more unst-able.

Local population growth is a key direct determi-nant for employment in construction. This is plausible as population growth in the region leads to new hous-

ing construction. Indirectly, via increased employment in construction, population growth also feeds into more output and employment in manufacturing. More employment in construction, by contrast, reduces employment in services. The negative sign of construction in the service equation suggests that workers are likely moving back and forth between construction and services. As demand for construction workers increases, workers are attracted to it from the services due to the generally higher wages in construction, and vice versa.

The two national variables *industrial production* and *interest rate spread* affect employment only for the services and manufacturing. Only employment construction is affected as predicted by a change in the exchange rate. Somewhat surprisingly, a rise in imports from China reduces employment only in the services, but not in manufacturing. Experiments with a longer lag length (up to 10) for imports from China do not change this result.

Figures 1 through 3 contain multi-step out-of-sample extrapolations of the three components of nonfarm employment.[5] The extrapolations of construction
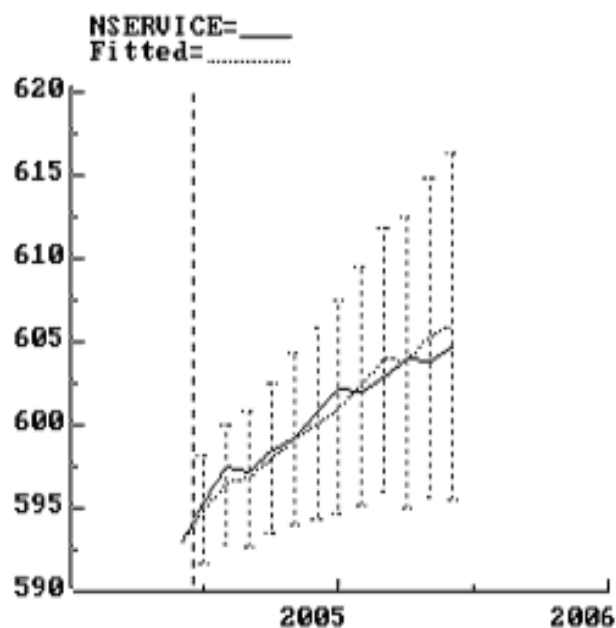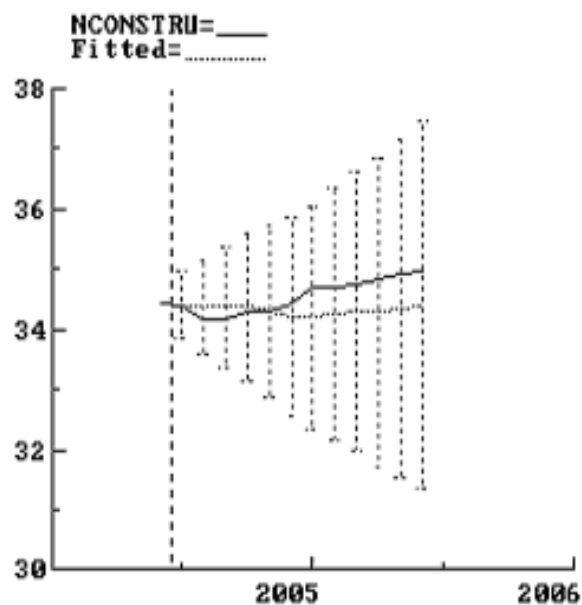

Figure 2. 12-period-ahead out-of-sample extrapolation of construction employment

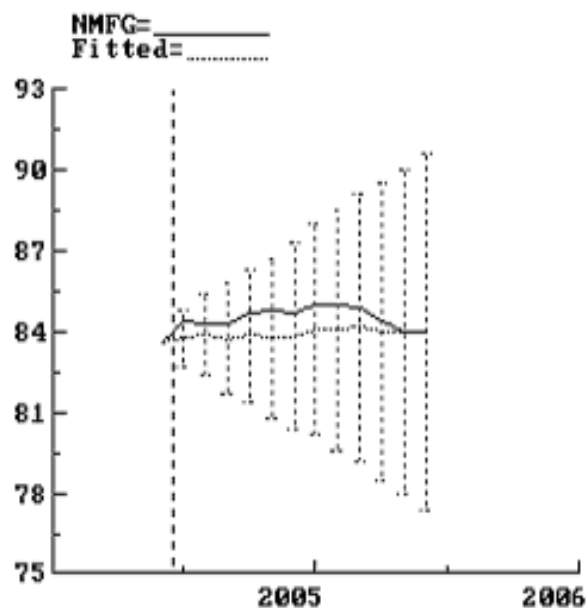**Figure 1.** 12-period-ahead out-of-sample extrapolation of service employment

Figure 3. 12-period-ahead out-of-sample extrapolation of manufacturing employment

---

[5] An augmented Kalman filter is used to optimally predict the variables out of sample (Durbin and Koopman 2001). The forecast is conditional on information at time 2004:6.

(Figure 2) consistently underpredict after about half a year. The figure also reveals that construction is a small sector relative to the other two sectors. More importantly, the underprediction of construction is matched by an overprediction of the service sector by

a similar amount toward the end of the one year fore-cast horizon. Since the forecast of manufacturing is very close to the actual value at the end of the forecast horizon, the out-of-sample prediction error for the sum of all three employment figures is rather small for a forecast horizon of one year (12 periods). This is documented in Figure 4. The out-of-sample forecast of total non-farm employment in this figure is derived as the sum of the predictions for each of the three employment components. It is apparent that the predictions tend to underestimate the true employment level for most months. This reflects the underprediction of manufacturing employment over most of the forecast horizon. The error does not cumulate for total non-farm employment: the prediction for a year in advance matches the actual employment level almost exactly.
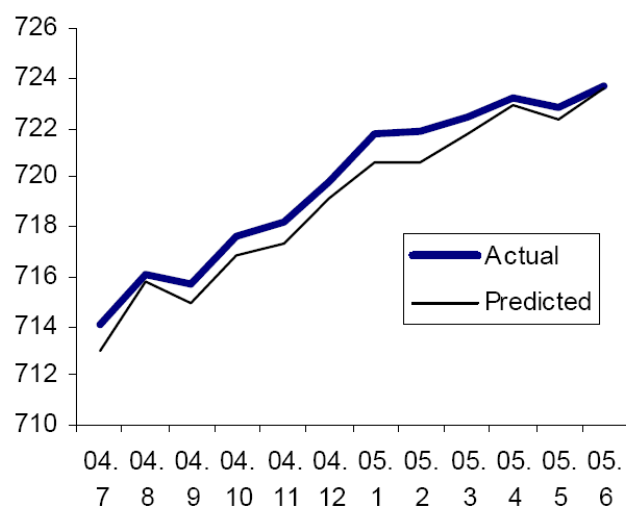

Figure 4. Total Non-farm Employment - Actual and 12-period Out-of-Sample Extrapolation

Another way to assess the out-of-sample extrapolations of the model detailed in Table 2 is to compare them to those of an alternative standard forecasting method, such as a vector autoregression (VAR). The results of such a comparison are provided in Table 3. It is apparent that the forecasting performance of a simple VAR with six lags on each employment component of total non-farm employment is significantly worse than that of the unobserved component model of Table 2. This applies not only to each of the three sectoral equations but also to the sum of the three sectoral forecasts.[6] It confirms the hesitations about VARs

---

[6] The performance of the VAR is even worse when observation specific dummy variables are included similar to those for the unobserved component model.

or its close relative Vector Error Correction Models (VECMs) expressed by Harvey (1997).[7]

Table 3. Sum of Squared Errors for a 12-Period Out-of-Sample Extrapolation, 2004:7 to 2005:6

	Model of Table 2	VAR with 6 lags
Services	0.94	44.58
Construction	1.65	4.24
Manufacturing	3.70	37.60
Total employment	7.52	125.95

Notes: VAR denotes a vector autoregression with a constant and six lags for each variable. Both models are estimated over the same time period, 1990. 9 - 2004. 6.

Figures 5 to 7 are of special interest for long-range forecasting. They depict the underlying unobserved components of the model identified in Table 2. There are two of these components for each sectoral employment equation, a level and a slope. These are defined by equations (5) and (6), respectively. Together they comprise the stochastic trend of the sectoral models. It is apparent from the level component of the stochastic trend that the job growth in the service sector was significant in the 1990s, leveled off perceptively around 2000, and has since then been again on an upward trend. The recent upward trend in the level, however, appears to be moderate compared to the one observed in the 1990s. Trend employment in construction, as identified by the level component, almost doubled from 1993 to 2000, then dropped off dramatically, and recovered after the end of the recession. But there appears to be little to suggest a continuation of the surge seen in the latter part of the 1990s. It rather looks like trend employment in construction will be steady or possibly slightly declining. Decline is also the key word that describes trend employment in manufacturing. Ever since 1995, trend employment has been going down in manufacturing. Only the early 1990s saw a positive trend growth, which was largely fueled by the movement of the automobile industry into the region. The graph shows that the trend decline in employment seems to have leveled off as of late.

---

[7] Compare Crane and Nourzad (1998) for an application of VECMs to regional employment forecasting.
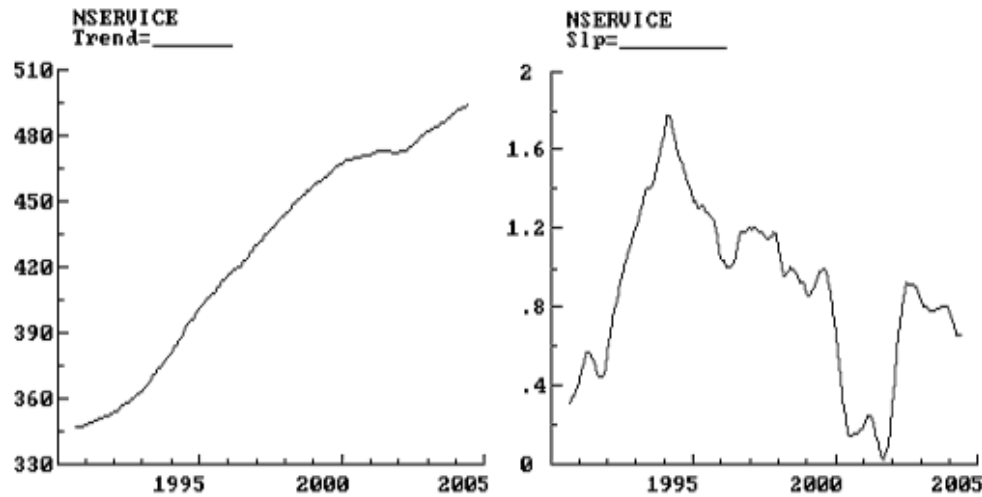
Figure 5. Unobserved Trend Components of Service Employment
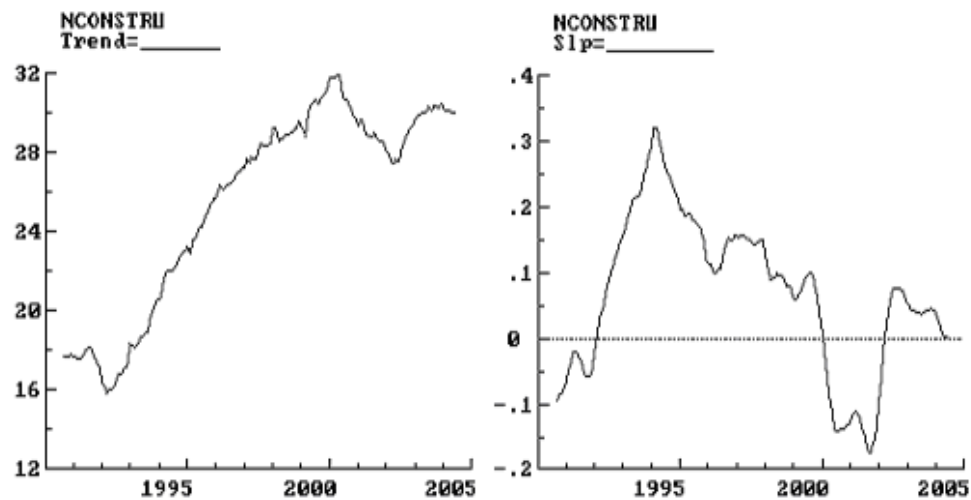

Figure 6. Unobserved Trend Components of Construction Employment

The slope components of the unobserved stochastic trends behave rather similarly for the three employment sectors. This is in apparent contrast to the diverging movements of the three level components. The slopes appear to be driven largely by the strong growth of the automobile industry in the early 1990s and the recession around the year 2000. The slope, therefore, captures the effect of special events on employment that are not adequately represented by any of the included variables.

5.   Conclusions

The purpose of this study has been to demonstrate the usefulness of structural time series (unobserved components) models for forecasting local (regional) employment. The proposed methodology has been applied to the forecasting of non-farm employment for the Nashville MSA. Forecasts of local economic conditions are in high demand by decision-makers in both government and the private sector. Our approach offers one alternative for regional economists interested in forecasting local economic activity.
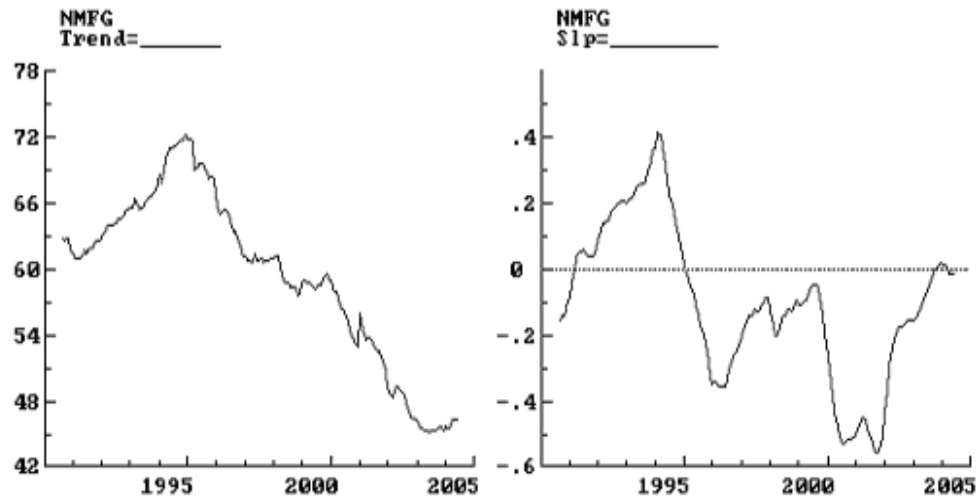
Figure 7. Unobserved Trend Components of Manufacturing Employment

The results suggest that unobserved component models can enhance a forecasting project in several respects. They tend to capture the data generating process even in the absence of a sufficient number of explicit regressor variables. In fact, the unobserved components serve as a quasi substitute for key missing regressors. This is desirable in all applications where the necessary data are not available or where there is little theory to help in the selection of independent variables. By capturing the data generating process, unobserved component models tend to compare favorably in terms of forecasting performance to other popular forecasting methods, such as VARs. This has been clearly brought out by the example discussed in this study. As in any regression analysis, unobserved component models allow one to explicitly control for as many variables as one wants. This makes it possible to control for policy effects or other changes in the economic environment, including unspecified effects, also known as outliers.

For the purpose of forecasting, unobserved component models do not only have the advantage that they can provide rather accurate short-run out-of-sample forecasts. Since they identify the underlying trend of a series, they quite naturally also provide the basis for longer range forecasts of the trend of a series. Depending on the type of data series, it would also be possible to identify the cyclical or seasonal pattern of a data series. Both can be modeled as unobserved components. This makes it possible in principle to remove seasonal or cyclical components from a series based on an explicit estimated model. In other words, there is

no need then to rely on one-size-fits-all filters, such as the X-12 seasonal adjustment program of the BLS or the Hodrick-Prescott filter.

We believe that unobserved component models should be considered by regional economists who are searching for alternative forecasting methods. These models allow the forecaster to predict local economic actvity with relative accuracy without the commitment of a large amount of time and resources. Forecasts that are accurate and timely should be of substantial benefit for local school systems, economic developers, and businesses that need information about local economic activity expected for the near future.

References

Carlino, G. and R. DeFina. 1999. The differential regional effects of monetary policy: Evidence from the U.S. states. *Journal of Regional Science* 39(2): 339–358.

Carlino, G. 2003. A confluence of events? Explaining fluctuations in local employment. Federal Reserve Bank of Philadelphia *Business Review*. First quarter, pp. 6-12.

Crane, S.E. and F. Nourzad. 1998. Improving local manufacturing employment forecasts using cointegration analysis. *Growth and Change* 29(2): 175-195.

Doring, T. and J. Schnellenbach. 2006. What do we know about geographical knowledge spillovers and regional growth?: A survey of the literature. *Regional Studies* 40(3), 375-395.

Durbin, J. and S.J. Koopman. 2001. *Time Series Analysis by State Space Methods*. Oxford and New York: Oxford University Press.

Estrella, A. and F.S. Mishkin. 1998. Predicting U.S. recessions: Financial variables as leading indicators. *Review of Economics and Statistics* 80(1): 45–61.

Harvey, A.C. 1997. Trends, cycles and autoregressions. *Economic Journal* 107(440): 192-201.

Harvey, A.C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge and New York: Cambridge University Press.

Johansson, B., C. Karlsson, and R.R. Stough. 2001. *Theories of Endogenous Regional Growth: Lessons for Regional Policies*. Heidelberg and New York: Springer.

Koopman, S.J., A.C. Harvey, J.A. Doornik, and N. Shephard. 1995. *Stamp 5.0: Structural Time Series Analyser, Modeller, and Predictor*. London: Chapman and Hall.

Koopman, S.J., N. Shephard, and J.A. Doornik. 1999. Statistical algorithms for models in state space using SsfPack 2.2. *Econometrics Journal* 2(1): 107–160.

Martin, R. and P. Sunley. 1998. Slow convergence? The new endogenous growth theory and regional development. *Economic Geography* 74(3), 201-227.

SAS. 2004. *SAS/ETS 9 User's Guide*. Carey, NC: SAS Publishing.

Shoesmith, G.L. 2003. Predicting national and regional recessions using probit modeling and interest rate spreads. *Journal of Regional Science* 43(2): 373–392.

Treyz, G.I. 1993. *Regional Economic Modeling: A Systematic Approach to Economic Forecasting and Policy Analysis*. Norwell, Mass. and Dordrecht: Kluwer Academic.

Zietz, J. 1996. The relative price of tradables and nontradables and the U.S. trade balance. *Open Economies Review* 7(2): 147-160.