An Evaluation of Alternative Strategies for Integrating Input-Output Information into Industry Employment Forecasting Equations

by

Dan S. Rickman

and

Steven R. Miller

Department of Economics
Oklahoma State University

Preliminary: Please do not quote without permission

Abstract

Alternative strategies for incorporating input-output information into industry employment forecasting equations are evaluated. The strategies differ along several dimensions. First, strategies that are based on using input-output information to select which industry employment variables to include as independent variables in each employment forecast equation are compared to those based on using input-output linkages to create aggregate demand variables for inclusion. Second, strategies differ according to the degree of endogeneity in the regional economy that they reflect. The strategies range from those that reflect only intermediate demand linkages, to those that also reflect induced consumption, investment and state and local government spending. Third, the strategies also differ according to whether model selection procedures are used to potentially reduce the set of independent variables in each equation. The particular model selection procedure used is that of Bayesian Model Averaging. Finally, the strategies differ according to whether restrictions are placed upon the coefficients during estimation. Relative forecast performance of the alternative models serve as one criterion in the evaluation of the models. Employment multipliers of the successful forecasting strategies also are examined to assess the usefulness of the different strategies for impact analysis. A primary finding of the study is that a strategy that imposed restrictions on inter-industry relationships through the use of input-output linkages in aggregation, produced comparable forecasts to less restrictive specifications, and proved more useful in impact analysis.

INTRODUCTION

Alternative strategies for incorporating input-output information into industry employment forecasting equations for the Oklahoma economy are evaluated. The strategies differ along several dimensions.

First, strategies that are based upon using input-output information to select which industry employment variables to include as independent variables in each employment forecast equation are compared to those based on using input-output linkages to create aggregate demand variables for inclusion. For the individual-industry strategy, the industries with the largest input-output row coefficients are selected as the independent variables for the row equation. Second, strategies differ according to the degree of endogeneity in the regional economy that they reflect. The strategies range from those that reflect only intermediate demand linkages, to those that also reflect induced consumption, investment and state and local government spending. These differing degrees of endogeneity reflect the differences in Type I, Type II and extended input-output multipliers. Thus, the comparison of the strategies indicates which endogeneities most likely exist in the Oklahoma economy. Third, the strategies also differ according to whether model selection procedures are used to potentially reduce the set of independent variables in each equation. The particular model selection procedure used is that of Bayesian Model Averaging. Bayesian Model Averaging is used to produce a model that contains Bayesian averaged estimates and a model with the highest posterior probability. Finally, the strategies differ according to whether restrictions are placed upon the coefficients during estimation. The restrictions are imposed using Bayesian mixed estimation akin to that used in Bayesian Vector Autoregression. In addition, all strategies are compared to model specification by stepwise regression.

Relative forecast performance of the alternative models serve as one criterion in the evaluation of the models. Employment multipliers of the successful forecasting strategies also are examined to assess the usefulness of the different strategies for impact analysis. A primary finding of the study is that a strategy that imposed restrictions on inter-industry relationships through the use of input-output linkages in aggregation, produced comparable forecasts to less restrictive specifications, and multipliers more useful in impact analysis. Differences in forecast accuracy only become significant at forecast horizons of 2-quarters ahead or more. Inter-industry linkages also become more important for forecast accuracy in the longer forecast horizons. Another finding is that, for Oklahoma, induced investment and state and local government spending responses are not apparent, and induced consumption responses are less than that predicted by Type II input-output multipliers. The Bayesian model selection procedure did not generally improve forecast performance. Finally, the strategies differed in their relative accuracy of industry employment versus that of total nonagricultural employment, owing to model differences in error cancellation.

METHOD

The models developed are based on two over-arching strategies: (1) including industry employment levels as right-hand-side variables in sector-specific employment forecasting equations; and (2) collapsing industry employment variables into aggregate demand variables before inclusion as independent variables. Within each strategy, however, additional model selection criteria are used in specifying the equations.

The first general strategy involves collapsing industry employment levels into aggregate demand variables using a regional input-output model: intermediate demand and local final demand. In addition, variables are included that reflect domestic and foreign export market linkages. Models are specified according to different conceptualizations of input-output multiplier responses, and also using Bayesian model selection. Finally, Bayesian mixed estimation is used to place parameter restrictions on the coefficients of the estimated equations.

The second general strategy is that of specifying employment growth in each sector as a function of growth rates in other sectors. Alternative methods are tried for specifying which sectors are included in each equation though. The model selection methods include stepwise regression, Bayesian model selection, and variable selection based on input-output linkages.

Aggregate Demand Variable Strategy

The general approach consists of calculating IO-based demand variables and including them in sector-specific employment forecasting equations. The approach is akin to that of Moghadam and Ballard (1988), which collapsed all possible intermediate demand linkages into a single time-series variable and included it in a forecasting equation. However, the current approach derives additional time-series variables based on IO linkages between sector employment and local final demand, domestic export demand, and foreign export demand.

Aggregation has the potential advantage of reducing problems of over-parameterization. Yet parameter restrictions are implicit in the aggregation. The linkages contained in the aggregated variables are uniformly adjusted during econometric estimation, limiting the within-sample fit.

Following the presentation in Rickman (forthcoming), I begin with the row equation of an input-output model: Equation (1). Equation (1) linearly relates output in private nonfarm industry i (Q_i) to output in other sectors (Q_j) through intermediate demands. Output also is linearly dependent upon final demands in Equation (1): consumption (C); residential investment (I_R); nonresidential investment (I_N); state and local government expenditures (G_{SL}); federal government expenditures (G_{FD}); and domestic and foreign exports (X). Mathematically expressed, Equation (1) is

$$(1)\ Q_i = \sum_{j=1}^{n}(\alpha_{ji} Q_j) + \alpha_{Ci} C + \alpha_{IRi} I_R + \alpha_{INi} I_N + \alpha_{SLi} G_{SL} + \alpha_{FDi} G_{FD} + \alpha_{Xi} X,$$

where n includes all nonfarm sectors including the state & local and federal government sectors; α_{ji} is the technical coefficient that relates output of i to output of j; α_C, α_{IR}, α_{IN}, α_{SL}, α_{FD} and α_X are the proportions of the corresponding final demands that are comprised by industry i output in the region.

Obtaining reduced-form employment relationships, however, requires converting industry output to employment. Following input-output models, I assume fixed proportions: $Q_i = E_i/\beta_i$, for all i, where β_i is the fixed ratio of employment to output in industry i. Therefore, substituting for output in terms of employment, and for the moment ignoring the federal government and export terms, I obtain:

$$(2)\ E_i \ = \ \beta_i\, [\sum_{j=1}^{n} (\alpha_{ji} / \beta_j) E_j \ + \ \alpha_{Ci} C \ + \ \alpha_{IRi} I_R \ + \ \alpha_{INi} I_N \ + \ \alpha_{SLi} G_{SL}].$$

Next, the local final demands need to be expressed in employment terms. To do so, I assume linear relationships between local final demand and labor income generated in the regional economy:

$$(3) C \ = \ \gamma_C \sum_{h=1}^{n} E_h W_h ,$$

$$(4)\ I_R \ = \ \gamma_{IR} \sum_{h=1}^{n} E_h W_h ,$$

$$(5)\ I_N \ = \ \gamma_{IN} \sum_{h=1}^{n} E_h W_h ,$$

$$(6)\ G_{SL} \ = \ \gamma_{SL} \sum_{h=1}^{n} E_h W_h ,$$

where W_h is the average annual wage rate obtained from the input-output table for industry h; and γ_C, γ_{IR}, γ_{IN} and γ_{SL} are the respective ratios of final demand to labor income in the input-output table. Equations (3)-(6) reflect final demand endogeneities found in extended input-output models (Batey and Rose, 1990). Federal government final demand spending is assumed to be exogenous to the region; thus, no expression is specified for federal government spending responses to changes in local economic activity.

Substituting Equations (3)-(6) into Equation (2), Equation (7) is obtained:

$$(7)\ E_i = \beta_i\, [\sum_{j=1}^{n} (\alpha_{ji} / \beta_j) E_j + (\alpha_{Ci}\gamma_C + \alpha_{IRi}\gamma_{IR} + \alpha_{INi}\gamma_{IN} + \alpha_{SLi}\gamma_{SL}) \sum_{h=1}^{n} E_h W_h].$$

Next, for illustration of econometric estimation I substitute the following:

$$(8)\, Y_i = E_i\ ,$$

and adding back exports, decomposed into domestic (including federal government spending) and foreign exports,

$$(11)\, X3_i = \beta_i(\alpha_{DXi} DX + \alpha_{FDi} G_{FD})\ ,$$

$$(12)\, X4_i = \beta_i \alpha_{FXi} FX\ ,$$

where DX and FX denote domestic exports and foreign exports, respectively; α_{DX} and α_{FX} represent the shares of domestic and foreign exports comprised of output in industry i, respectively; and subscripting for time t yields:

(13) $Y_{it} = X1_{it} + X2_{it} + X3_{it} + X4_{it}$.

Because of potential problems associated with non-stationarity (which is tested for later and found), Equation (13) is differentiated and converted to rates of change. Expressing the differentiated equation in a form suitable for econometric estimation, Equation (14) is obtained:

(14) $Y_{it}^{*} = \delta_1(\alpha_{X1i} X1_{it}^{*}) + \delta_2(\alpha_{X2i} X2_{it}^{*}) + \delta_3(\alpha_{X3i} X3_{it}^{*}) + \delta_4(\alpha_{X4i} X4_{it}^{*})^{+}\varepsilon_{it}$,

where α_{X1i}, α_{X2i}, α_{X3i}, α_{X4i} are the employment shares attributable to intermediate demand, local final demand, domestic exports (including the federal government) and foreign exports, respectively; the δ's are the parameters to be econometrically estimated, which equal unity if the IO-based variables replicate the historical data precisely. Equation (14) relates the percent change in industry employment to the percent changes in the demand components, weighted by the corresponding shares of industry employment attributable to the demand components in the base year of the input-output table. The variable, X1, is the intermediate demand shifter found in many econometric models that embed IO interindustry linkages, while, X2 contains linkages that are often omitted or proxied by macro-regional variables. The export variables, X3 and X4, also contain IO linkages that are omitted or proxied by similar variables across equations such as a U.S. macroeconomic variable. To allow for transmission and adjustment lags, Equation (14) is extended by adding lags of Y^* and X_k^* for all k to the right-hand-side. Equation (14) can be estimated in unrestricted form using ordinary least squares, or in restricted form using Bayesian estimation.

Bayesian Estimation

Estimation of Bayesian-restricted forms of Equation (14) can be accomplished using Litterman's (1980; 1986) extension of Theil's (1963) mixed estimation framework. To begin, for each industry i, let Y denote the vector of percent changes in industry employment in Equation (14), let X denote the corresponding matrix containing the independent variables, and let β represent the true parameter vector (the δ's). Mixed estimation consists of imposing stochastic restrictions on the estimation of β in the ordinary least squares model, resulting in a prior distribution of β with mean b and variance-covariance matrix ψ. To compute the mixed estimator, the data vectors are augmented using dummy observations:

(15) $$\begin{pmatrix} Y \\ r \end{pmatrix} = \begin{pmatrix} X \\ R \end{pmatrix} \beta + \begin{pmatrix} u \\ v \end{pmatrix},$$

where r is a vector of prior means (b), R is an identity matrix, E(u)=E(v)=0, and E(vv')=ψ. Note that the stochastic restrictions take the form: r = Rβ + v.

Then applying Aitken's generalized least-squares procedure to Equation (15), the mixed estimates for equation i are obtained:

(16) $\beta_{Theil} = (X'X + \sigma^2_u R'\psi^{-1}R)^{-1}(X'Y + \sigma^2_u R'\psi^{-1}r)$.

In calculating β_{Theil}, σ^2_u is obtained from the standard error of ordinary least squares estimation of the unrestricted form. Because of the infinite number of choices that could be made in specifying the uncertainty surrounding the prior means, the Litterman method of hyperparameters is used to specify ψ. To illustrate, let $\lambda^2(i,j)$ denote the variance of the prior for the coefficient on variable j in equation i:

(17) $\lambda^2(i,j) = \theta^2$, where $0 \leq \theta \leq \infty$.

The parameter θ reflects "overall tightness," with smaller values of θ reflecting less uncertainty around the prior means. Note that letting θ approach ∞ leads back to ordinary least squares. Therefore, ψ becomes a diagonal matrix with elements $\lambda^2(i,j)$.

Individual Industry Selection

Using employment growth of individual industries as explanatory variables avoids imposing restrictions on the estimates of inter-industry linkages. However, as the level of industry detail increases, over-parameterization can become a problem and sub-optimal forecast performance may result. One solution to over-parameterization of the VAR model is Bayesian estimation, producing the BVAR (Litterman, 1980; 1986). Industry-specific BVARs were successfully implemented by LeSage and Magura (1991) and Partridge and Rickman (1998). Yet even in BVARs, the number of time-series observations limits industry detail. An alternative is to reduce the number of variables to be included through model selection procedures.

Selection based on Input-Output Linkages

Input-output linkages have been used as a guide in specifying the inter-industry linkages in a VAR framework (Magura, 1987; Fawson and Criddle, 1994). The advantage of the approach is the reduction in parameters to be estimated. The disadvantage is that the variables omitted implicitly have parameters restricted to equal zero, which may cause bias. There also is the issue of how many industries to include, and whether to account for final demand relationships between sectors as well as intermediate demand relationships.

Stepwise Regression

The first procedure that comes to mind when considering model selection is a stepwise regression. Judging the quality of the model fit by the adjusted r-square, the process first finds the best single-variable model then seeks the best two-variable model from the remaining predictors. The process is continued adding the third variable, but now the procedure checks to see if any of the prior selected variables should be excluded. Whereas the combination of the first and second chosen variables is the superior two-variable model given that the first variable is included, the combination of the second and third variables may perform better as a two variable-model than the prior in the absence of the first chosen variable. The process of looking forward for new variables to add and looking backward for existing variables to remove continues until all variables contributing .05 to the adjusted r-square measure of goodness of fit have been exhausted.

Two noteworthy problems exist for the use of stepwise regression. First, stepwise regressions cannot guarantee a best model. The algorithm is often fooled by data irregularities that contribute to high adjusted r-squares such us nonstationarity. Furthermore, because stepwise regression uses single variable steps, it is vulnerable to finding local peaks and not global peaks in the adjusted r-squared. Second, measuring goodness of fit by the adjusted r-square, leads to bias toward including too many predictors in the model.

Bayesian Model Selection

An alternative to stepwise regression for model selection is Bayesian Model Averaging. Instead of maximizing the adjusted r-squared as in stepwise regression, in the Bayesian approach we maximize the posterior probability of Y_{t+j} given the data D. The Bayesian model selection approach has been shown to out-perform stepwise regression by Raftery, Madigan and Hoeting (1997) and has been used in regional employment forecasting by LeSage and Rey (2000).

The approach used here follows that of LeSage and Rey (2000). In their approach, the model selected is that which maximizes the posterior probability, in which the posterior probability of model M_k is given as:

$$(18)\, pr(Mk \mid D) = \frac{pr(D \mid M_k)\, pr(M_k)}{\sum_{l=1}^{K} pr(D \mid M_l)\, pr(M_l)}$$

where pr(D | Mk) represents the marginal likelihood of model Mk and pr(Mk) denotes the prior probability associated with model k. Following LeSage and Rey (2000), uniform prior probabilities were assigned to all models and Bayesian priors suggested by Raftery, Madigan and Hoeting were assigned to the parameters. The priors are relatively diffuse, allowing the sample data to determine the best model.[1]

We also employ the averaging solution of Leamer (1978):

$$(19)\; pr(y_{t+j} | D) = \sum_{k=1}^{K} pr(y_{t+j} | M_k, D) pr(M_k | D) .$$

This is an average of the posterior distributions of the k models, each weighted by the corresponding posterior model probability.[2]

IMPLEMENTATION

The equations are specified for thirty industries that comprise total private nonfarm employment in Oklahoma (shown in Table 1). Quarterly, non-seasonally adjusted, employment data from the second quarter of 1983 to the third quarter of 1998 are used (*Bureau of Labor Statistics*). The equations were estimated in rates-of-change, since Dickey-Fuller tests (Dickey and Fuller 1979; 1981) revealed that the employment levels were non-stationary, while their first-differences were stationary.[3] Quarterly seasonal dummy variables, minus one, also were added to each equation.

Models

Individual Industry

Specification of which industries' employment growth to include in each equation is based on econometric model selection procedures, input-output linkages, and a combination of the two strategies. First, both Bayesian model selection and stepwise regression are used to select among the thirty industries and seasonal dummy variables for inclusion in each equation: Bayes(30) and Stepwise(30). Next, intermediate demand linkages alone are used to select five industries to include in each equation (IO(5)).[4] These equations reflect linkages that underlie Type I multipliers. Final demand linkages are then used to determine an additional five industries to include in equations classified as dependent upon local final demand (IO(10)).[5] Finally, Bayesian model selection is used to select among the industries in IO(5) and IO(10) to produce Bayes(5) and Bayes(10), respectively.

Aggregate Demand

Alternative aggregate demand models are constructed that correspond to different input-output multiplier conceptualizations. One version is akin to the extended input-output model in that it fully implements the endogeneities of Equation (14), and is estimated using ordinary least squares (Extended IO). Applying the Bayesian model selection procedure to Equation (14) produces a version that only incorporates some of the aggregated demand variables in each equation (Baye_Ext_IO). Yet a third version is specified that utilizes the Bayesian model averaged estimates as prior means in Bayesian mixed estimation of the Extended IO model to produce Baye_Ave_IO.[6] Type I (Un_Type I) and Type II (Un_Type II) multiplier versions also are derived and estimated in unrestricted form. Un_Type I is obtained by excluding the final demand variable in Equation (14); whereas, Un_Type II model is obtained by including the final demand variable, but omitting the investment and state and local government demand responses; i.e., set γ_{IN},γ_{IR} and γ_{SL} equal to zero. Two final versions are created by separating the final demand variable into two variables: a consumption variable; and an investment/state and local government demand variable. Including these final demand variables in each equation produces Un_Ext_IO, and applying the Bayesian model selection procedure to the demand variables produces Baye_Un_Ext_IO.

Data

The IMPLAN data (Minnesota IMPLAN Group, 1996) are used as the source of input-output information in the alternative strategies. For the aggregate demand strategy, IMPLAN is used to derive the X_i and to calculate β, α_{X1} ,α_{X2} α_{X3},α_{X4}, γ_C, γ_{IR}, γ_{IN} and γ_{SL}. IMPLAN also is used for variable selection in IO(5) and IO(10).

Although IMPLAN uses national technical coefficients, implicitly assuming a uniform national production technology, IMPLAN regionalizes the national coefficients by accounting for differing regional trade dependence. The regionalization method that IMPLAN uses is the regional purchase coefficient (RPC) approach. Wage rates used to derive $X2^*$ are assumed exogenous.[7] Industry wage rates are calculated from the IMPLAN model as total wages and salaries plus proprietor income divided by corresponding employment. Domestic exports were assumed to grow proportionate to U.S. employment; thus, U.S. employment growth (*Bureau of Labor Statistics*) is substituted for $X3^*$ in Equation (14). Regarding foreign exports, $X4^*$ is proxied by GDP growth of Oklahoma's major sources of export demand, weighted by the share of exports going to each country in 1997 based on data from the Massachusetts Institute of Social and Economic Research (MISER) country of export destination of Oklahoma's manufactured exports.[8]

For individual industry models, only current period variables were considered because of the lack of degrees of freedom associated with including lags of 30 industries. Yet, for the aggregate demand strategy, we tested for the optimal lag length. The models were estimated using data from the first quarter of 1985 through the fourth quarter of 1994. A maximum lag length of six quarters for all variables in each equation was tried. The optimal lag length was allowed to vary by industry and was determined based on the Schwarz Criterion (SC).[9] With the exception of the durable goods and engineering services sectors, in which the optimal lag length was estimated to be one, the optimal lag was estimated to be zero in all sectors.

Out-of-sample forecasts needed to assess the historical fit of the models were generated for 1, 2, 4, 6 and 8 quarters ahead. To start with, the equations were estimated using data from the first quarter of 1985 through the fourth quarter of 1994. They were then used to produce a forecast (1, 2, 4, 6 or 8 quarters ahead), with the first quarter of 1995 being the first forecast-quarter. Next, rolling forecasts were calculated by re-estimating the equations after each one-quarter forecast step, with the third quarter of 1998 being the last forecast period. Therefore, fifteen 1-quarter, fourteen 2-quarter, twelve 4-quarter, ten 6-quarter and eight 8-quarter forecasts were calculated.

To calculate multipliers the equations were first estimated using data from the third quarter of 1983 through the fourth quarter of 1994. The estimated model is then used to produce a baseline 8-quarter-ahead forecast. Then separately for each sector, employment is increased by one thousand, and the model resolved for the total change in employment. The ratio of the total change in employment to the exogenous change of one thousand becomes the multiplier for that sector.

4. RESULTS

The forecast accuracy results for integration strategies based on selecting appropriate industries to include in each equation appear in Table 2. The results for integration strategies based on selecting appropriate aggregate demand variables appear in Table 3. Panel A in each table presents the weighted Mean Absolute Percent Error (MAPE) for each strategy, with each industry's MAPE weighted by its employment share of total nonagricultural employment. Panel B contains the MAPE for total nonagricultural employment, which is based on a comparison of the sum of the individual industry forecasts and actual total nonagricultural employment each quarter. The relative performance of integration strategies in terms of forecasting total nonagricultural employment can differ from that of the weighted industry forecasts because of differences in cancellation properties.[10] Tables 4 and 5 report weighted MAPE's for the manufacturing industries for both the individual industry and aggregate demand strategies, respectively.

Individual Industry Strategy Results

From panel A of Table 2 we see that the IO(5) strategy produced the most accurate forecasts for forecast horizons of 2, 4, 6, and 8 quarters ahead, while Bayes(5) was most accurate for 1-quarter ahead forecasts. IO(10) was second most accurate for forecast horizons of 2,4, and 8 quarters ahead. Stepwise(30) was uniformly least accurate, followed by Bayes(30) as the next least accurate (except for 8 quarters ahead). These results suggest that simply using input-output tables to select which variables enter each equation is more effective than using econometric-based model selection procedures. It also suggests that just using intermediate demand links in the IO table in equation specification is more accurate than also including variables based on induced-consumption responses.

From panel B we see that the econometric-based integration strategies also are least accurate in forecasting total nonagricultural employment. However, IO(10) is now most accurate in three forecast horizons (2, 4 and 6 quarters ahead) and second most accurate for 8-quarter ahead forecasts. Bayes(5) is most accurate for 1-quarter ahead forecasts, while Bayes(10) is most accurate for 8-quarter ahead forecasts. IO(5) is not most accurate, or even second most accurate, for any forecast horizon. Thus, better cancellation of errors appears associated with accounting for induced-consumption responses.

Regarding the relative performance of the econometric-based model selection procedures, they appear unable to compete with the use of input-output information. However, when combined with input-output information, the Bayesian model selection procedure shows some promise. For example, Bayes(5) produces the most accurate forecasts in both panels A and B for the 1-quarter ahead horizon.[11] This likely occurs because Bayes(5) contains the fewest inter-industry linkages and linkages to the national economy, which is likely accurate in the very short run. Yet, as the forecast horizon lengthens strategies that account for the linkages become more accurate, and the parsimonious Bayesian models become relatively less accurate.

Aggregate Demand Variable Strategy Results

From panel A of Table 3, the most accurate forecast strategy was that based on including an intermediate demand variable in each employment equation, but excluding induced demand variables: Un_Type I. Un_Type I had the most accurate forecasts for the 1, 2-, 4- and 6-quarter ahead forecasts, and was second most accurate for the 8-quarter-ahead forecasts. The most accurate 8-quarter ahead forecast was the strategy that included separate induced demand variables for consumption and investment/state and local government spending (B_Un_Ex_IO), in which it also was second most accurate for the 4- and 6-quarter ahead forecasts. The least accurate strategy for three of the forecast horizons (4, 6 and 8 quarters ahead) was that which also included an induced-demand variable for endogenous consumption, investment, and state and local government spending (Extended IO). The Bayesian Model Averaged strategy (Baye_Av_IO) was least accurate for the 1- and 2-quarter ahead forecasts. In comparing Extended IO to Baye_Ex_IO, and Unres_Ex_IO to B_Un_Ex_IO, the Bayesian selected models were more accurate in 8 of the 10 cases.

The Type I model also produced the most accurate 2-, 4-, 6- and 8-quarter ahead total nonagricultural employment forecasts and second most accurate 1-quarter forecasts. For the 1-quarter-ahead forecasts, the Bayesian selected extended IO model (Baye_Ex_IO) was most accurate. The Bayesian selected models in general were less accurate for total nonagricultural employment than their counterparts. Given their relative accuracy for individual industries, the Bayesian selected models produced relatively less error cancellation across industries.

Forecasting Manufacturing

To assess the comparative accuracy of the integration strategies across industries, the industry MAPEs for manufacturing were aggregated using each industry's share of total manufacturing employment (Tables 4 and 5). As shown in Tables 4 and 5, the MAPEs for manufacturing are larger than those obtained for all industries. This may be due to greater volatility in manufacturing overall, or greater volatility in the components because they reflect a greater level of sector disaggregation.

In contrast to the results from panel A in Table 2, in which IO(5) was most accurate, Table 4 shows that IO(10) was the most accurate strategy for selecting variables to enter each manufacturing employment equation. IO(5) was only second most accurate for three forecast horizons (2, 4, 6). The stepwise regression strategy (Stepwise(30)) was again least accurate, and was the only strategy that produced dramatically different MAPEs from all other models. Bayes(5) and Bayes(10) had identical specifications for the manufacturing equations as the Bayesian selection procedure eliminated all variables in IO(10) that were not in IO(5).

In Table 5, the Type I multiplier strategy is most accurate for the 4-, 6- and 8-quarter ahead forecast horizon, and second most accurate for the 2-quarter ahead forecasts. This confirms the findings for accuracy over all industries. The least accurate models were those that included one aggregate variable for all induced responses (Extended IO and Unres_Ex_IO).

Individual Industry vs. Aggregate Demand Strategies

A comparison of Table 3 with Table 2 shows that generally the more accurate strategies are those which are based on including individual industries in each equation, not aggregate variables. Only for the 2- and 8-quarter ahead forecasts of nonagricultural employment are the Type I multiplier forecasts more accurate than the best forecasts in Table 2. And only the Type I multiplier model consistently comes close to the accuracy of the best individual industry model. The IO restrictions imposed in the aggregation of industry employment generally cause forecast accuracy to deteriorate, particularly those that include linkages based on induced spending responses. For industry-specific forecasts, Type I multiplier strategies (IO(5) in Table 2 and Type I in Table 3) are more accurate. Only for total nonagricultural employment forecasts in Table 2, does accounting for induced consumption responses (IO(10)) improve accuracy.

Nevertheless, models in which IO linkages are integrated with econometric equations often are used for both forecasting and impact analysis. Strategies that consist of simply adding key industry employment variables to each equation may produce unsatisfactory multipliers. For one, many of the variables left out may be important. Second, while not necessarily being a problem for forecasting, collinearity makes the econometric estimates inefficient, which affects the calculated multipliers.[12] Therefore, the multipliers of IO(5), IO(10) and Type I are compared to each other and the restricted versions of Type I and Type II strategies (Res_TypeI and Res_TypeII).[13]

As shown in Table 6, the average multipliers in IO(5) and IO(10) are close to that of the restricted Type I multiplier model. However, the variation appears much greater in IO(5) and IO(10), in which some are even negative. Theoretically, negative values are highly improbable.

The unrestricted Type I multipliers also are slightly closer to those of the restricted Type II model than the restricted Type I model. Thus, it may be that the econometric estimates in the Type I model are picking up some endogenous consumption responses, not just to the degree predicted by the restricted Type II model. To further compare the multipliers, matched-t tests and Pearson Correlations are calculated and displayed in Table 7.

From Panels A and B in Table 7, we see that the unrestricted Type I multipliers do not significantly differ from those of IO(5) and IO(10) at the 0.05 level, but they also are not very positively correlated either. To be sure, IO(5) and IO(10) are not very correlated with any of the aggregate demand variable models. The two models only incorporating intermediate demand responses are barely correlated at 0.105, while those incorporating endogenous consumption responses are negatively correlated with each other at –0.308. In addition, the multipliers of the unrestricted Type I multiplier model are more correlated with the restricted Type II model (r=0.890) than the restricted Type I model (r=0.839), which suggests that the econometric estimates captured endogenous consumption responses. The fact that the unrestricted Type I multipliers are significantly smaller than the restricted Type II multipliers suggests that the assumption of constant marginal propensity to consume overstates the consumption responses to changes in income.

5. CONCLUDING REMARKS AND SUGGESTED RESEARCH DIRECTIONS

This paper evaluated alternative strategies for incorporating input-output information into industry employment econometric equations for the state of Oklahoma. The evaluation focused on both the relative forecast performance of the alternative models, and the employment multipliers of the successful forecasting strategies. A primary finding of the study is that a strategy that imposed restrictions on inter-industry relationships through the use of input-output linkages in aggregation, produced comparable forecasts to less restrictive specifications, and multipliers more useful in impact analysis.

Differences in forecast accuracy generally only become dramatic at forecast horizons of 2-quarters ahead or more. Inter-industry linkages also become more important for forecast accuracy in the longer forecast horizons. Another finding is that, for Oklahoma, induced investment and state and local government spending responses are not apparent, and induced consumption responses are less than that predicted by Type II input-output multipliers. A Bayesian model selection procedure did not generally improve forecast performance. Finally, the strategies differed in their relative accuracy of industry employment versus that of total nonagricultural employment, owing to model differences in error cancellation.

This study could be extended several ways. For one, in the Bayesian model selection procedure, incorporating input-output information into the priors in some fashion may improve the model selection procedure. In addition, incorporating changes in productivity and factor intensities may improve model performance. The result would be a model that could possibly capture neoclassical linkages found in structural econometric models, but were not imposed as exact restrictions, and would be supported by historical data.

ENDNOTES

[1]The procedure for implementing the Bayesian approach is described in LeSage and Rey (2000), where the corresponding MATLAB program is available at http://www.spatial-econometrics.com/.

[2]The procedure of LeSage and Rey (2000) also produces Bayesian model averaged estimates, in which k includes all models with posterior probabilities of at least one percent.

[3] Differencing has been shown to be particularly important for the estimation of unconstrained models (Cromwell and Hannan, 1993) since non-stationarity causes the least squares estimates to be inconsistent (Granger and Newbold, 1974).

[4]The five industries chosen were those with the largest $(\beta_i/\beta_j)\alpha_{ij}$.

[5]The additional five industries selected were those classified as exogenous that had the largest wage bill in the input-output model: Mining, Food and Kindred Products, Rubber and Misc. Plastics, Fabricated Metals, and Transportation Equipment. They were included in equations in which local consumption demand was estimated to be at least one-third of final demand: Construction, Textiles and Apparel, Other Nondurable Goods, and non-manufacturing industries (except Transportation, Business and Engineering Services).

[6]For the two industries that contained one-period lags, prior means equal to zero were specified for the lagged variables.

[7]Assuming exogenous wage rates allows the employment equations to be a self-contained model.

[8]The sources of export demand for Oklahoma products used in the weighting and construction of $X4^*$ are: Canada, Mexico, Japan, the Netherlands, the United Kingdom and France. The shares were normalized to sum to unity.

[9]Own-lags also were included.

[10]For a discussion of how forecast errors may cancel across time and industries see Rickman (forthcoming).

[11]Use of Stepwise regression to select among the five industries in each equation uniformly produced less accurate forecasts than Bayes(5).

[12] Forecast performance does not deteriorate in the presence of multicollinearity if it is stable into the forecast period.

[13]For the restricted versions θ=0.01, while for Bay_Av_IO θ=0.2.

REFERENCES

Batey, Peter J. and Adam Z. Rose. "Extended Input-Output Models: Progress and Potential," *International Regional Science Review* 13 (1990): 27-49.

Brucker, Sharon M., Steven E. Hastings, and William R. Latham III. "A Comparison of Five Ready-Made Model Systems," *Review of Regional Studies* 17 (1987): 1-16.

_____________. "The Variation of Estimated Impacts from Five Regional Input-Output Models," *International Regional Science Review* 13 (1990): 119-139.

Cromwell, Jeff B., and Michael J. Hannan. "The Utility of Impulse Response Functions in Regional Analysis: Some Critical Issues," *International Regional Science Review* 15 (1993): 199-222.

Dickey, David A., and Wayne A. Fuller. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association* 74 (1979): 427-431.

_____________. "Likelihood Ratio Tests for Autoregressive Time Series with a Unit Root," *Econometrica* 49 (1981): 1057-1072.

Fawson, Chris, and Keith R. Criddle. "A Comparative Analysis of Time Series Approaches to Modeling Intersectoral and Intercounty Employment Linkages in Rural Regional Labor Markets," *Journal of Regional Science* 34 (1994), 57-74.

Granger, Clive and P. Newbold, "Spurious Regressions in Econometrics," *Journal of Econometrics* 2 (1974): 111-120.

LeSage, James P. and Michael Magura. "Using Bayesian Techniques for Data Pooling in Regional Payroll Forecasting," *Journal of Business & Economic Statistics* 8 (1990): 127-135.

_____________. "Using Interindustry Input-Output Relations as a Bayesian Prior in Employment Forecasting Models," *International Journal of Forecasting* 7 (1991): 231-238.

LeSage, James P. and Sergio J. Rey. "Restrictions in Integrated Econometric+Input-Output Modeling," unpublished manuscript (2000).

Litterman, Robert B. "A Bayesian Procedure for Forecasting With Vector Autoregression," Working Paper, Massachusetts Institute of Technology, Department of Economics (1980).

______________"Forecasting with Bayesian Vector Autoregressions-Five Years of Experience," *Journal of Business & Economic Statistics* 4 (1986): 25-38.

Magura, Michael. "The Use of Input-Output Tables in Specifying Interindustry and Interregional Labor Market Linkages," *Papers of the Regional Science Association*, 62, 117-23.

Minnesota IMPLAN Group, Inc. *IMPLAN Professional*, Stillwater, MN (1996).

Moghadam, K. and K. Ballard. “Small area modeling of the industry section (SAMIS): An Integrated Econometric-Interindustry Approach,” *Environment and Planning A* 20 (1987): 665-669.

Partridge, Mark D. and Dan S. Rickman. “Generalizing the Bayesian Vector Autoregression Approach for Regional Interindustry Employment Forecasting,” *Journal of Business and Economic Statistics* 16 (1998a): 62-72.

Partridge, Mark D. and Dan S. Rickman. “Regional Computable General Equilibrium Modeling: A Survey and Critical Appraisal,” *International Regional Science Review 2*1 (1998b): 205-248.

Raynauld, Jacques, and Jean-Guy Simonato. "Seasonal BVAR Models: A Search Along Some Time Domain Priors," *Journal of Econometrics* 55 (1993): 203-219.

Rey, Sergio J. “The Performance of Alternative Integration Strategies for Combining Regional Econometric and Input-Output Models,” *International Regional Science Review* 21 (1998): 1-36.

Rickman, Dan S. “Using Input-Output Information for Bayesian Forecasting of Industry Employment in a Regional Econometric Model, *International Regional Science Review* (forthcoming).

Rickman, Dan S. and R. Keith Schwer. “A Systematic Comparison of the REMI and IMPLAN Models: The Case of Southern Nevada,” *Review of Regional Studies* 23 (1993): 143-161.

_____________. “A Comparison of the Multipliers of IMPLAN, REMI, and RIMS II: Benchmarking Ready-Made Models for Comparison,” *Annals of Regional Science* 29 (1995): 363-374.

Theil, Henri. “Use of Incomplete Prior Information in Regression Analysis,” *Journal of the American Statistical Association* 58 (1963): 401-414.

Treyz, George I, Dan S. Rickman and Gang Shao. “The REMI Economic-Demographic Forecasting and Simulation Model,” *International Regional Science Review* 14 (1992): 221-254.

Table 1. Industry Classification

1987 SIC code	Industry	%Oklahoma Private Nonag Employment (1995:1-1998:3 Average)
10	Mining	2.90
15	Construction	4.55
20	Food and Kindred Products	1.59
28,31	Other Nondurable Goods	0.40
22,23	Textiles and Apparel	0.73
24	Lumber and Wood Products	0.40
25,39	Other Durable Goods	0.55
26	Paper and Allied Products	0.40
27	Printing and Publishing	1.06
29	Petroleum Products	0.44
30	Rubber and Misc. Plastics	1.25
32	Stone, Clay and Glass Products	0.84
33	Primary Metals	0.48
34	Fabricated Metals	1.99
35	Industrial Machinery	2.89
36	Electrical Machinery	0.96
37	Transportation Equipment	1.70
38	Instruments	0.42
40-45,47	Transportation	4.29
46,48,49	Public Utilities	2.76

50,51	Wholesale Trade	6.00
58	Eating and Drinking Establishments	8.38
52-57,59	Rest of Retail	14.73
60-69	Finance, Insurance & Real Estate (FIRE)	6.17
70	Hotel and Motel Services	0.99
73	Business Services	6.87
80	Health Services	10.49
82	Education Services	1.12
87	Engineering Services	2.15
72,75,76,78,79,81,83, 84,86,88,89	Other Services	12.53

Table 2. Forecast Accuracy Comparison: Individual Industry Variables

	Panel A: Sector-Weighted Mean Absolute Percent Errors				
Model	1-qtr forecast	2-qtr forecast	4-qtr forecast	6-qtr forecast	8-qtr forecast
Bayes(30)	1.2035	1.7329	3.0106	3.7324	3.9298
Stepwise(30)	1.3294	2.102	3.8096	5.7938	7.8676
IO(5)	1.0245	1.4501*	2.033*	2.9804*	3.2433*
IO(10)	1.061	1.4835	2.092	3.272	3.3574
Bayes(5)	0.9908*	1.5005	2.2193	3.1194	3.7664
Bayes(10)	1.0279	1.564	2.413	3.4345	4.2199
	Panel B: Mean Absolute Percent Errors for Nonagricultural Employment				
	1-qtr forecast	2-qtr forecast	4-qtr forecast	6-qtr forecast	8-qtr forecast
Bayes(30)	0.2412	0.4775	0.8522	1.2804	1.6895
Stepwise(30)	0.2623	0.4879	0.7636	1.2879	2.1518
IO(5)	0.2660	0.4992	0.7367	1.1456	1.5534
IO(10)	0.2546	0.4279*	0.5560*	0.828*	1.0951
Bayes(5)	0.2390*	0.4330	0.7434	1.0528	1.4304
Bayes(10)	0.2419	0.4505	0.6892	0.8834	0.7765*

*Denotes most accurate strategy for that period.

Table 3. Forecast Accuracy Comparison: Aggregate Demand Variables

	Panel A: Sector-Weighted Mean Absolute Percent Errors				
Model	1-qtr forecast	2-qtr forecast	4-qtr forecast	6-qtr forecast	8-qtr forecast
Extended IO	1.0521	1.8295	3.0962	4.1697	4.9574
Baye_Ex_IO	1.0545	1.8633	2.7578	3.8364	4.4058
Baye_Av_IO	1.1828	2.7466	2.9869	3.7324	4.228
Un_Type II	1.0328	1.7001	2.793	3.8172	4.4913
Un_Type I	1.0154*	1.4851*	2.2444*	3.1646*	3.7543
Un_Ext_IO	1.0893	1.7609	2.7279	3.7637	4.4493
B_Un_Ext_IO	1.0771	1.7243	2.4526	3.2911	3.7038*
	Panel B: Mean Absolute Percent Errors for Nonagricultural Employment				
	1-qtr forecast	2-qtr forecast	4-qtr forecast	6-qtr forecast	8-qtr forecast
Extended IO	0.2722	0.9254	1.7518	2.5233	3.6458
Baye_Ex_IO	0.256*	1.0415	1.8782	2.7373	3.5680
Baye_Av_IO	0.4574	2.3174	2.3061	2.9048	3.2681
Un_Type II	0.2704	0.7254	1.443	2.0638	2.857
Un_Type I	0.2663	0.4175*	0.662*	0.9143*	0.7699*
Un_Ext_IO	0.3644	0.6363	1.2071	1.4448	1.5981
B_Un_Ext_IO	0.3200	0.8296	1.3468	1.7532	2.1985

*Denotes most accurate strategy for that period.

Table 4. Forecast Accuracy Comparison for Manufacturing: Individual Industry Variables

	Sector-Weighted Mean Absolute Percent Errors				
Model	1-qtr forecast	2-qtr forecast	4-qtr forecast	6-qtr forecast	8-qtr forecast
Bayes(30)	1.6848	2.2262	4.185	5.6119	5.8769
Stepwise(30)	2.3502	3.5458	6.458	10.3349	14.3927
IO(5)	1.6099	2.2184	3.4643	4.7393	5.8518
IO(10)	1.5951	2.2016*	3.441*	4.6234*	5.4896*
Bayes(5)	1.5942*	2.248	3.5737	4.6665	5.8677
Bayes(10)	1.5942*	2.248	3.5737	4.6665	5.8677

*Denotes most accurate strategy for that period.

Table 5. Forecast Accuracy Comparison for Manufacturing: Aggregate Demand Variables

	Sector-Weighted Mean Absolute Percent Errors				
Model	1-qtr forecast	2-qtr forecast	4-qtr forecast	6-qtr forecast	8-qtr forecast
Extended IO	1.7638	2.6651	4.6326	6.3982	8.02
Baye_Ex_IO	1.6576*	2.7488	4.3628	5.8293	7.4804
Baye_Av_IO	1.7283	3.2318	4.3306	5.6119	6.9337
Type II	1.786	2.4194	4.1447	5.7255	7.1149
Type I	1.7229	2.3361	3.8387*	5.1981*	6.3277*
Unres_Ex_IO	1.7617	2.3074*	3.8496	5.4846	6.8473
B_Un_Ex_IO	1.6896	2.5317	3.9759	5.3217	6.8453

*Denotes most accurate strategy for that period.

Table 6. Multipliers

	Multiplier Definition					
Sector	Restricted TypeII	Restricted Type I	Unrest. Type I	Un_TypeII	IO(5)	IO(10)
Mining	2.302	1.347	1.866	2.253	3.826	7.359
Const	2.169	1.429	1.899	1.939	2.487	3.103
Food	3.476	2.012	2.816	2.844	1.823	4.1
Oth Non	3.890	2.043	3.150	3.731	0.457	-0.029
Text&Appl	1.911	1.286	1.527	1.728	1.004	0.999
Lumber	2.407	1.575	2.133	2.319	1.164	0.91
Oth Dur	2.225	1.322	1.609	2.210	-0.457	1.004
Paper	4.517	2.090	3.601	3.830	9.043	7.39
Printing	2.268	1.333	1.739	2.255	0.086	0.253
Petro	11.321	5.756	5.634	6.522	2.168	1.398
Rubber	3.220	1.656	2.220	3.028	0.156	-1.03
Stone	3.080	1.587	1.951	2.845	0.584	0.372
Prim Metal	3.890	2.309	2.969	3.632	1.316	2.052
Fab Metal	2.656	1.384	2.038	2.772	0.74	4.628
Ind Mch	3.046	1.515	2.121	3.102	2.443	2.187
Elec Eq	2.864	1.016	2.958	3.064	0.719	1.154
Tran Eq	4.502	2.241	3.180	3.664	0.356	0.654
Instruments	3.478	1.815	2.918	3.734	1.311	1.325
Tran Srv	2.491	1.433	2.152	2.889	1.53	1.504
Comm	3.985	2.103	4.521	5.457	-0.622	-10.003
Wholesale	2.458	1.390	1.927	2.579	3.537	4.021
Eat Ret	1.517	1.157	1.480	1.560	1.136	1.303
Rest of Ret	1.444	1.084	1.176	1.506	1.531	1.227
FIRE	2.033	1.322	2.312	2.768	3.869	2.521
Hotel	1.728	1.241	1.551	1.707	1	1
Oth Srv	1.611	1.165	1.539	1.452	1.677	2.16
Bus Srv	1.613	1.137	1.208	1.618	1.011	0.985
Health Srv	1.925	1.225	1.488	2.112	3.82	5.005
Educ Srv	1.717	1.220	1.743	1.950	0.977	0.105
Eng Srv	1.823	1.210	1.569	1.981	1	1
Average	2.919	1.647	2.300	2.768	1.656	1.6219

Table 7. Statistical Comparisons of Multipliers

Panel A: P-Values for Matched t-tests					
	Un_TypeI	Res_TypeI	Res_TypeII	Un_TypeII	IO(5)
Un_TypeI					
Res_TypeI	0.000				
Res_TypeII	0.003	0.000			
Un_TypeII	0.000	0.000	0.403		
IO(5)	0.087	0.980	0.008	0.007	
IO(10)	0.281	0.966	0.058	0.083	0.931
Panel B: Pearson Correlation					
	Un_TypeI	Res_TypeI	Res_TypeII	Un_TypeII	IO(5)
Un_TypeI					
Res_TypeI	0.839				
Res_TypeII	0.890	0.980			
Un_TypeII	0.969	0.829	0.885		
IO(5)	0.105	0.080	0.107	0.046	
IO(10)	-0.263	-0.080	-0.070	-0.308	0.697