An Evaluation of Alternative Strategies for Incorporating Interindustry Relationships into a Regional Employment Forecasting Model

*Dan S. Rickman and Steven R. Miller**

Abstract: Alternative strategies for incorporating interindustry relationships into industry employment forecasting equations are evaluated. Strategies that are based on using input-output information to select which industry employment variables to include as independent variables in each equation are compared to those of including input-output-based aggregate demand variables. The strategies also differ according to the degree of endogeneity between industry employment levels, whether model selection procedures are used, and whether restrictions are placed upon the coefficients during estimation. Relative forecast performance of the alternative models and an examination of the employment multipliers serve as criteria in the evaluation of the models.

I. INTRODUCTION

A variety of strategies have been employed to incorporate interindustry relationships into regional economic forecasting models. A common strategy has been to use input-output (IO) tables to collapse interindustry relationships into aggregate demand shifters for inclusion in industry-specific econometric forecasting equations (e.g., Duobinis 1981; Moghadam and Ballard 1988; Rickman 2001). However, the restrictions imposed on interindustry relationships in the construction of the demand shift variables may be inaccurate, which limits forecast performance. Alternatively, linkages in an IO model have been used to determine which industry variables to include in each industry-specific equation (e.g., Magura 1987; Fawson and Criddle 1994). Although the variables included do not impose any restrictions on the interindustry relationships, variables are omitted that may be relevant, imposing exact restrictions of zero on some interindustry relationships. In using the Bayesian Model Averaging approach, LeSage and Rey (2000) selected variables to include in each equation based on the combination of variables with the highest Bayesian posterior probability, avoiding imposing a priori exclusion restrictions. Yet, in comparing forecasts from this model to one with an IO-based demand shift variable, they used national IO coefficients and national industry employment in place of regional coefficients and regional employment in the construction of the aggregate demand shifter. Using the Oklahoma economy as an example, this paper takes up the issue of how alternative strategies for incorporating interindustry relationships into industry employment equations compare.

*Department of Economics & Legal Studies, Oklahoma State University, Stillwater, OK.

The strategies differ along several dimensions. First, strategies that are based on selecting industry employment variables to include as independent variables in each employment forecast equation are compared to those based on using IO linkages to create aggregate demand variables for inclusion. Second, strategies differ according to whether IO information or econometric model selection is used to specify which industry employment variables to include as independent variables in each employment forecast equation. One particular econometric model selection procedure used is that of Bayesian Model Averaging (BMA). BMA is used to produce a model with the highest posterior probability for each industry. Specification by stepwise regression also is performed to provide a forecast performance benchmark. Third, 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 IO multipliers. Finally, the strategies differ according to whether restrictions are placed upon the coefficients during estimation. The restrictions are imposed using mixed estimation akin to that used in Bayesian Vector Autoregression.

The relative forecast performance of the alternative models serves as one criterion in the evaluation of the models. Employment multipliers of the more successful forecasting strategies also are examined to assess their usefulness for impact analysis. A primary finding of the study is that a strategy of imposing restrictions on interindustry relationships through the use of IO linkages in aggregation produced somewhat comparable forecasts to models with selected individual industries to include in each equation, and produced multipliers deemed more useful in impact analysis. Also, interindustry linkages are shown to be more important for forecast accuracy in the longer forecast horizons. The BMA selection procedure does not generally produce superior forecast performance. Finally, the strategies differ in their relative accuracy of industry employment versus that of total nonagricultural employment, owing to model differences in error cancellation.

II. METHOD

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

The first general strategy involves using a regional IO model to collapse industry employment levels into aggregate demand variables (i.e., intermediate demand and local final demand variables). In addition, variables are included that reflect domestic and foreign export market linkages. Models are specified according to different conceptualizations of IO multiplier responses, and also using

Bayesian model selection. Finally, for some models, mixed estimation is used to restrict the coefficients in their estimation.

The second general strategy is that of specifying employment in each sector as a function of employment 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 IO linkages.

Aggregate Demand Variable Strategy

The aggregate demand variable 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), in which all IO intermediate demand linkages were collapsed into a single time-series variable for inclusion in a forecasting equation. However, following Rickman (2001), 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 forecast problems associated with overparameterization. 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 that can be obtained.

Following Rickman (2001), employment in industry i, time t (Y_{it}), can be written as being attributable to intermediate demands of other industries ($X1_i$), local final demands ($X2_i$), domestic export demand ($X3_i$) and foreign export demand ($X4_i$):

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

Intermediate demands in employment terms are obtained by using the regional IO coefficients (converted into employment terms), and aggregating over all j industries for industry i. Local final demands include household income induced consumption effects, as well as induced investment and state and local government spending. Domestic export demand is assumed proportionate to U.S. employment in that industry, and accounts for federal government spending in the region. Foreign export demand is assumed proportionate to income of Oklahoma's major trading partners.[1]

Because of potential problems associated with nonstationarity, Equation 1 is differentiated and converted to rates of change. Expressing the differentiated equation in a form suitable for econometric estimation, Equation 2 is obtained:

(2) $Y_{it}^* = \delta_1\left(\alpha_{X1t}X1_{it}^*\right) + \delta_2\left(\alpha_{X2t}X2_{it}^*\right) + \delta_3\left(\alpha_{X3t}X3_{it}^*\right) + \delta_4\left(\alpha_{X4t}X4_{it}^*\right) + \varepsilon_{it}$,

where α_{X1i}, α_{X2i}, α_{X3i}, and α_{X4i} are the employment shares attributable to intermediate demand, local final demand, domestic exports (including the federal government),

[1]For details of the derivations, see Rickman (2001).

and foreign exports, respectively. The δ's are the parameters to be estimated econometrically, which equal unity if the IO-based variables replicate the historical data precisely. Equation 2 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 IO table. To allow for transmission and adjustment lags, Equation 2 is extended by adding lags of Y^* and X_k^* for all k to the right-hand side. Equation 2 can be estimated in unrestricted form using ordinary least squares (OLS), or in restricted form using mixed estimation.

Mixed Estimation

Restricted estimation of Equation 2 can be accomplished using Litterman's (1980, 1986) extension of Theil's (1963) mixed estimation framework. For each industry i, let Y denote the vector of percent changes in industry employment in Equation 2, 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 OLS model. The mixed estimates (β_{mix}) for equation i are obtained as:

$$(3) \qquad \beta_{mix} = \left(X'X + \sigma_u^2 R'\psi^{-1}R \right)^{-1} \left(X'Y + \sigma_u^2 R'\psi^{-1}r \right),$$

where R is an identity matrix, r contains the prior means, and ψ reflects the tightness of the prior.

In calculating β_{mix}, σ_u^2 is obtained from the standard error of OLS estimation of the unrestricted form. Because of the infinite number of choices that could be made in specifying the uncertainty surrounding the prior means, a simplified version of 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:

$$(4) \qquad \lambda^2(i,j) = \theta^2, \text{ where } 0 \le \theta \le \infty.$$

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 OLS. 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 interindustry linkages as in the aggregate demand variable strategy. However, as the level of industry detail increases, overparameterization can become a problem and suboptimal forecast performance may result. One solution to overparameterization 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 greatly limits industry detail. An alternative is to reduce the number of variables to be included through model selection procedures.

Selection Based on IO Linkages

IO linkages have been used as a guide in specifying the interindustry 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 econometric procedure that often comes to mind when considering model selection is stepwise regression. Judging the quality of the model fit by the adjusted r-squared, stepwise regression first finds the best single-variable model, and then seeks the best two-variable model from the remaining predictors. The process continues by adding a 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-squared measure of goodness of fit have been exhausted. Yet, the adjusted r-squared method of model specification has in general been suggested as leading to the inclusion of too many variables (Amemiya 1985, pp. 50-51).

Bayesian Model Selection

An alternative to stepwise regression for model selection is BMA. Instead of maximizing the adjusted r-squared as in stepwise regression, in BMA the posterior probability of Y_{t+j} given the data D is maximized. BMA has been shown to outperform 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:

(5) $pr(M_k|D) = \left(pr(D|M_k)pr(M_k)\right) \big/ \left(\sum_l pr(D|M_l)pr(M_l)\right),$

where $pr(D|M_k)$ represents the marginal likelihood of model M_k and $pr(M_k)$ 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 (1997) were assigned to the

parameters. The priors are relatively diffuse, allowing the sample data to determine the best model.[2]

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

$$(6) \qquad \text{pr}(Y_{t+j}|D) = \sum_k \text{pr}(Y_{t+j}|M_k, D)\text{pr}(M_k|D).$$

The result in Equation 6 is an average of the posterior distributions of the K models that have posterior probabilities of at least 1 percent, with each posterior distribution weighted by the corresponding posterior model probability.

III. IMPLEMENTATION

TABLE 1

Industry Classification

1987 SIC Code	Industry	%Oklahoma Private Nonag. Emp. (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

Equation 2 is specified for each of 30 industries that comprise total private nonagricultural employment in Oklahoma (shown in Table 1). Quarterly nonseasonally 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 nonstationary, while their first-differences

[2]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/.

were stationary.[3] Quarterly seasonal dummy variables, minus one, also were added to each equation.

Models

Aggregate Demand

Alternative aggregate demand models are constructed that correspond to different IO multiplier concepts. One version is akin to the extended IO model (Batey and Rose 1990), in that it fully implements the endogeneities of Equation 2, and is estimated using OLS (Extended IO). Applying the BMA procedure to Equation 2 produces a version that may only include some of the aggregate demand variables in each equation (Baye_Ext_IO), such that the posterior probability is maximized for each equation in model selection, yet estimated with OLS. Yet a third version is specified that utilizes the Bayesian model averaged estimates as prior means in mixed estimation of Extended IO to produce Baye_Ave_IO.[4] Type I (Un_Type I) and Type II (Un_Type II) multiplier versions also are derived and estimated in unrestricted form with OLS. Un_Type I is obtained by excluding the local final demand variable (X2) in Equation 2, whereas the Un_Type II model is obtained by including X2, but omitting the investment and state and local government demand responses in its construction.

Individual Industry

Specification of which industries' employment growth to include in each equation is based upon econometric model selection procedures, IO linkages, and a combination of the two strategies. First, both Bayesian model selection and stepwise regression are used to select among the 30 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)).[5] These equations solely 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)).[6] 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. Thus, for both Bayes(5) and Bayes (10), each equation not only reflects the use of IO information, but it also is the specification that produced the largest posterior probability among the alternatives provided by the IO model.

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

[4]For the two industries that contained one-period lags, prior means equal to zero were specified for the lagged variables. For Bay_Av_IO, θ=0.2.

[5]The five industries chosen for inclusion in each equation were those that contained the largest interindustry employment linkages in the IO table with the dependent variable industry.

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

Data

The IMPLAN data (Minnesota IMPLAN Group 1996) were used as the source of IO information in the alternative strategies. For the aggregate demand strategy, IMPLAN was used to derive X_i, α_{X1}, α_{X2}, α_{X3}, and α_{X4}. IMPLAN also was used for variable selection in IO(5) and IO(10). Wage rates used to derive X2* were assumed exogenous.[7] Industry wage rates were calculated from the IMPLAN model as total wages and salaries plus proprietor income divided by corresponding employment. U.S. employment growth used in X3* was from the Bureau of Labor Statistics. Regarding foreign exports, X4* was 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 on country of export destination of Oklahoma's manufactured exports from the Massachusetts Institute of Social and Economic Research.[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.[9] With the exception of the durable goods and engineering services sectors, in which the optimal lag length was estimated to be one quarter, the optimal lag was estimated to be zero in all sectors. Thus, for all but two industries, only contemporaneous inter-industry relationships are captured in the equations.

Out-of-sample forecasts used to assess the relative accuracy of the models were generated for one, two, four, six, and eight quarters ahead. To start with, the equations of each pre-specified model were estimated using data from the first quarter of 1985 through the fourth quarter of 1994. They were then used to produce a forecast (one, two, four, six, or eight 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. Yet, although this procedure produces out-of-sample forecasts, actual values for the data used to construct the exogenous variables (X3 and X4) were used, such that the forecasts are not strictly ex ante forecasts. Therefore, 15 one-quarter, 14 two-quarter, 12 four-quarter, 10 six-quarter, and 8 eight-quarter forecasts were calculated.

To calculate multipliers, the equations were first estimated using data from the third quarter of 1983 through the third quarter of 1996. The estimated model was then used to produce a baseline eight-quarter-ahead forecast. Then, separately for each sector, employment was increased by 1,000, and the model was resolved

---

[7]The assumption of exogenous wage rates allows the employment equations to be a self-contained block of a larger econometric model. Further discussion of how IMPLAN was used can be found in Rickman (2001).
[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.

for the total change in employment. The ratio of the total change in employment to the exogenous change of 1,000 became the multiplier for that sector.

IV. RESULTS

The forecast accuracy results for aggregate demand variable integration strategies appear in Table 2. The results for integration strategies based on selecting industry variables to include in each equation appear in Table 3. Panel A in each table presents the weighted Mean Absolute Percent Error (MAPE) for each strategy, with each MAPE weighted by the industry's 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 employment 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 error cancellation properties.[10]

Aggregate Demand Variable Strategy Results

TABLE 2

Forecast Accuracy Comparison: Aggregate Demand Variables

| Model | Panel A: Sector-Weighted Mean Absolute Percent Errors | | | | |
	1 qtr ahead	2 qtr ahead	4 qtr ahead	6 qtr ahead	8 qtr ahead
Extended IO	1.052	1.830	3.096	4.170	4.957
Baye_Ex_IO	1.055	1.863	2.758	3.836	4.406
Baye_Av_IO	1.183	2.747	2.987	3.732	4.228
Un_Type II	1.033	1.700	2.793	3.817	4.491
Un_Type I	1.015*	1.485*	2.244*	3.165*	3.754*
	Panel B: Mean Absolute Percent Errors for Nonfarm Employment				
	1 qtr ahead	2 qtr ahead	4 qtr ahead	6 qtr ahead	8 qtr ahead
Extended IO	0.272	0.925	1.752	2.523	3.646
Baye_Ex_IO	0.256*	1.042	1.878	2.737	3.568
Baye_Av_IO	0.457	2.317	2.306	2.905	3.268
Un_Type II	0.270	0.725	1.443	2.064	2.857
Un_Type I	0.266	0.418*	0.662*	0.914*	0.770*

*Denotes most accurate strategy for that period.

From panel A of Table 2, the most accurate forecast strategy was that which included an intermediate demand variable in each employment equation, while omitting the local final demand variable (Un_Type I).[11] The least accurate strategy for three of the forecast horizons (four, six, and eight quarters ahead) was that which also included the local final demand variable for endogenous consumption, investment, and state and local government spending (Extended IO). The BMA aggregate demand variable strategy (Baye_Av_IO) was least accurate for the one-

---

[10]For a discussion of how forecast errors might cancel across time and industries, see Rickman (2001).

[11]This result is consistent with Rickman (forthcoming). Yet, the focus of Rickman (forthcoming) was the relative accuracy of alternative IO multiplier concepts within an aggregate demand variable forecasting strategy. Forecasting was not the emphasis, no comparisons were made to the alternative individual industry strategies for incorporating interindustry relationships, and BMA was not used for model specification in the aggregate demand strategy (i.e., Baye_Ex_IO and Baye_Ave_IO were not constructed).

and two-quarter-ahead forecasts, but became second most accurate for the longer-term six- and eight-quarter-ahead forecasts.

Un_Type I also produced the most accurate two-, four-, six-, and eight-quarter-ahead total nonagricultural employment forecasts, and the second most accurate one-quarter forecasts. Where Un_Type I was most accurate, it was by a large margin. For the one-quarter-ahead forecasts, the Bayesian selected extended IO model (Baye_Ex_IO) was most accurate.

Individual Industry Strategy Results

TABLE 3

Forecast Accuracy Comparison: Individual Industry Variables

| Model | Panel A: Sector-Weighted Mean Absolute Percent Errors | | | | |
	1 qtr ahead	2 qtr ahead	4 qtr ahead	6 qtr ahead	8 qtr ahead
Bayes(30)	1.204	1.733	3.011	3.732	3.930
Stepwise(30)	1.329	2.102	3.810	5.794	7.868
IO(5)	1.025	1.451*	2.033*	2.980*	3.243*
IO(10)	1.061	1.484	2.092	3.272	3.357
Bayes(5)	0.991*	1.501	2.219	3.119	3.766
Bayes(10)	1.028	1.564	2.413	3.435	4.220
	Panel B: Mean Absolute Percent Errors for Nonfarm Employment				
	1 qtr ahead	2 qtr ahead	4 qtr ahead	6 qtr ahead	8 qtr ahead
Bayes(30)	0.241	0.478	0.852	1.280	1.690
Stepwise(30)	0.262	0.488	0.764	1.288	2.152
IO(5)	0.266	0.499	0.737	1.146	1.553
IO(10)	0.255	0.428*	0.556*	0.828*	1.095
Bayes(5)	0.239*	0.433	0.743	1.053	1.434
Bayes(10)	0.242	0.451	0.689	0.883	0.777*

*Denotes most accurate strategy for that period.

From panel A of Table 3, we see that IO(5) produced the most accurate forecasts for forecast horizons of two, four, six, and eight quarters ahead, while Bayes(5) was most accurate for one-quarter-ahead forecasts. IO(10) was second most accurate for forecast horizons of two, four, and eight quarters ahead. Stepwise(30) was uniformly least accurate, followed by Bayes(30) as the next least accurate (except for eight quarters ahead). Stepwise(30) averaged twenty variables (not counting seasonal dummy variables) in each equation, while Bayes(30) only averaged about two variables per equation. Thus, the poor forecast performance of Stepwise(30) is likely attributable to overparameterization. On the contrary, the Bayesian model selection procedure perhaps produced specifications that were too parsimonious. Consistent with the results in Table 2, the results in panel A of Table 3 suggest that just accounting for intermediate demand relationships in the IO table in equation specification is more accurate than also incorporating variables based on induced consumption responses.

From panel B we see that the econometric-based integration strategies also are generally least accurate in forecasting total nonagricultural employment. However, IO(10) is now most accurate in three forecast horizons (two, four, and six quarters ahead) and second most accurate for eight-quarter-ahead forecasts.

Bayes(5) is most accurate for one-quarter-ahead forecasts, while Bayes(10) is most accurate for eight-quarter-ahead forecasts. IO(5) is not most accurate, or even second most accurate, for any forecast horizon. Thus, greater 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 IO information. However, when combined with IO 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 one-quarter-ahead horizon.[12] This likely occurs because Bayes(5) contains the fewest interindustry linkages, and fewest 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 regional interindustry linkages become more accurate, and the parsimonious Bayesian models become relatively less accurate. In addition, one potential explanation for the relative accuracy of Bayes(5) and Bayes(10) compared to Bayes(30) is that potential overparameterization and multicollinearity limit the effectiveness of the BMA procedure, in which reducing the potential set of variables using the IO table mitigated these problems.

Individual Industry versus Aggregate Demand Strategies

A comparison of Table 3 with Table 2 shows that generally the more accurate strategies are those that are based upon including individual industries in each equation, not aggregate demand variables. Only for the two- and eight-quarter-ahead forecasts of nonagricultural employment are the unrestricted-Type I multiplier forecasts more accurate than the best forecasts in Table 3. And only the unrestricted-Type I multiplier model consistently comes close to the accuracy of the best individual industry model. In fact, the average MAPE across aggregate demand models (not shown) is only smaller than the average MAPE for the industry variable models at the one- and eight-quarter-ahead horizons, and this is solely due to the relative inaccuracy of Stepwise(30).

The IO restrictions imposed in the aggregation of industry employment generally caused 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 3 and Un_Type I in Table 2) are more accurate. Only for total nonagricultural employment forecasts in Table 3 does accounting for induced consumption responses (IO(10)) improve accuracy.

Nevertheless, in addition to their use in forecasting, econometric models are also often used for impact analysis. Hence, an additional criterion that can be used to evaluate the competing models is the reasonableness of the multipliers. 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

---

[12]The use of stepwise regression to select among the five industries in each equation uniformly produced less accurate forecasts than Bayes(5).

forecasting, collinearity makes the econometric estimates inefficient, which affects the calculated multipliers.[13] Therefore, the multipliers of IO(5), IO(10), and Un_Type I are compared to each other. Restricted versions of Type I and Type II strategies (Res_TypeI and Res_TypeII) are also used in the comparison for benchmarking purposes, in which the δ's were set equal to unity using mixed estimation.[14]

TABLE 4

Multiplier Comparison

Sector	Res_TypeII	Res_TypeI	Un_TypeI	Un_Type II	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	4.100
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	0.910
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	7.390
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	-1.030
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	0.740	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	1.530	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	1.000	1.000
Oth Srv	1.611	1.165	1.539	1.452	1.677	2.160
Bus Srv	1.613	1.137	1.208	1.618	1.011	0.985
Health Srv	1.925	1.225	1.488	2.112	3.820	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	1.000	1.000
Average	2.919	1.647	2.300	2.768	1.656	1.622

As shown in Table 4, the average multipliers in IO(5) and IO(10) are close to that of the restricted Type I multiplier model (Res_Type I). 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 OLS estimates in the Type I model are picking up some endogenous consumption responses, not just to the

---

[13]Forecast performance does not deteriorate in the presence of multicollinearity if it is stable into the forecast period, though if overparameterization occurs, forecast performance will be suboptimal.
[14]For the restricted versions θ=0.01.

degree predicted by the restricted Type II model (Rickman forthcoming). To compare the multipliers further, matched t-tests and Pearson Correlations are calculated and displayed in Table 5.

TABLE 5

Statistical Comparisons of Multipliers

	Un_TypeI	Res_TypeI	Res_TypeII	Un_TypeII	IO(5)
	Panel A: P-Values for Matched t-tests				
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

From panels A and B in Table 5, 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.

V. CONCLUDING REMARKS AND SUGGESTED RESEARCH DIRECTIONS

This paper evaluated alternative strategies for incorporating interindustry relationships 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 more successful forecasting strategies. A primary finding of the study is that a strategy that imposed restrictions on interindustry relationships through the use of IO linkages in aggregation produced somewhat comparable forecasts to less-restrictive specifications, and multipliers that appear more reasonable for use in impact analysis.

Differences in forecast accuracy generally only become dramatic at forecast horizons of two quarters ahead or more. Interindustry linkages also become more important for forecast accuracy in the longer forecast horizons. Econometric model selection procedures, including Bayesian model selection, generally produced less-accurate forecasts; whereas, using IO relationships in model specification

produced relatively more-accurate forecasts. Yet, when combined with other strategies for reducing the dimensionality, Bayesian model selection shows some promise. 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 type of comparison should be conducted for other regions to assess the generalizability of the results. For example, endogenous spending responses (e.g., investment and government) may be more prevalent in a larger and more developed economy, such that multipliers reflecting them more accurately represent the economy (Rey 1998). However, there is no a priori reason to expect that the relative accuracy of the aggregate-demand-variable versus individual-industry-variables approaches would change.

This study could be extended several ways. For one, in the Bayesian model selection procedure, incorporating IO information into the priors in some fashion may improve the model selection procedure. In addition, incorporating changes in productivity and factor intensities in the construction of the aggregate demand variables may improve its corresponding model performance. Other IO multiplier concepts also could be used in the construction of the aggregate demand variables. For example, in regions with slack labor markets, Type II multipliers produce too-large multipliers, suggesting the use of a variation of the Type IV multiplier (van Dijk and Oosterhaven 1986), which adjusts the estimated induced income gains downward to account for the likely reduction in transfer payments. The result of such exercises would be a relatively parsimonious econometric model that captures structural features of a regional economy, and yet was demonstrated to yield relatively accurate forecasts.

REFERENCES

Amemiya, T. *Advanced Econometrics*. Cambridge, Mass.: Harvard University Press, 1985.

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

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

Dickey, D.A., and W.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.

Duobinis, S.F. "An Econometric Model of the Chicago Standard Metropolitan Area." *Journal of Regional Science* 21(1981), 293-319.

Fawson, C., and K.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, C., and P. Newbold. "Spurious Regressions in Econometrics." *Journal of Econometrics* 2 (1974), 111-120.

Leamer, E.E. *Specification Searches.* New York: Wiley, 1978.

LeSage, J.P., and M. Magura. "Using Interindustry Input-Output Relations as a Bayesian Prior in Employment Forecasting Models." *International Journal of Forecasting* 7 (1991), 231-238.

LeSage, J.P., and S.J. Rey. "Restrictions in Integrated Econometric-Input-Output Modeling." Toledo, Oh.: University of Toledo Department of Economics, 2000. Available at: http://www.econ.utoledo.edu/faculty/lesage/lesage.html.

Litterman, R.B. "A Bayesian Procedure for Forecasting with Vector Autoregression." Working Paper, Massachusetts Institute of Technology, Department of Economics. Cambridge, Mass., 1980.

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

Magura, M. "The Use of Input-Output Tables in Specifying Interindustry and Interregional Labor Market Linkages." *Papers of the Regional Science Association* 62 (1987), 117-23.

Minnesota IMPLAN Group, Inc. *IMPLAN Professional.* Stillwater, Minn., 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, M.D., and D.S. Rickman. "Generalizing the Bayesian Vector Autoregression Approach for Regional Interindustry Employment Forecasting." *Journal of Business and Economic Statistics* 16 (1998), 62-72.

Raftery, A., D. Madigan, and J. Hoeting. "Bayesian Model Averaging for Linear Regression Models." *Journal of the American Statistical Association* 92 (1997), 179-191.

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

Rickman, D.S. "Using Input-Output Information for Bayesian Forecasting of Industry Employment in a Regional Econometric Model." *International Regional Science Review* 24 (2001), 226-244.

_____. "A Bayesian Forecasting Approach to Evaluating and Constructing Regional Input-Output Based Employment Multipliers." *Papers in Regional Science* (forthcoming).

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

van Dijk, J., and J. Oosterhaven. "Regional Impacts of Migrants' Expenditures: An Input-Output/Vacancy-Chain Approach." In P.W.J. Batey and M. Madden (eds.) *Integrated Analysis of Regional Systems.* London Papers in Regional Science 15. London: Pion Press, 1986.