

Evaluating Alternative Index Designs For the Texas Manufacturing Outlook Survey

By Fiona Sigalla, Frank Berger, Tom Fomby, Keith Phillips and Mine Yücel¹
Federal Reserve Bank of Dallas

This paper has been prepared for the Third Joint European Commission-OECD Workshop on
International Development of Business and Consumer Tendency Surveys,
Brussels, 12-13 November 2007.

The views expressed are the authors and do not necessarily represent those of the Federal
Reserve Bank of Dallas or the Federal Reserve System

¹ Sigalla is an Economist, Berger is Director of Technical Support and Data Analysis, Phillips is Senior Economist and Policy Advisor and Yücel is Senior Economist and Vice President. All are employees of the Federal Reserve Bank of Dallas. Fomby is a Research Associate of the Federal Reserve Bank of Dallas and Professor of Economics, Southern Methodist University, Dallas, Texas.

The Texas Manufacturing Outlook Survey was created under the leadership of Harvey Rosenblum, Executive Vice President and Director of Research. The authors thank Richard McKenzie and his colleagues at the OECD for assistance with building our survey and this analysis. Jim Orr with the Federal Reserve Bank of New York contributed helpful comments. Mike Nicholson and Raghav Virmani provided excellent research assistance. This research could not have been conducted without the help and support of the Dallas Fed Statistics Department, especially Mario Hernandez and Stephen Douglass.

Introduction

Business tendency surveys are used around the world to monitor economic activity and expectations about future growth. The surveys provide immediate information about the health of the economy and the outlook for changes in economic activity.

Business tendency surveys rely on qualitative responses from business leaders to generate indexes explaining changes in current conditions. Qualitative responses—asking whether an indicator has “increased,” “decreased” or “remained the same”—can provide insights into changes in factors such as sales and orders that typically would not be available for weeks or months if precise values were obtained. For sub-national surveys, much of the information is not available from any other source. Information can also be collected beyond those available from quantitative series, such as business leaders’ judgment about changes in the level of general business activity.

Anecdotal reports like those from business tendency surveys have been gaining importance for monitoring economic activity because the information can be gathered quickly and are not subject to revision. Analysts have long been aware of the difficulties associated with obtaining quantitative economic information in a timely fashion, with most indicators available with a delay of weeks or months. In recent years, economists have become increasingly aware of the impact data revisions have on our understanding of current conditions. Survey responses are measured with little error and are not revised.

Several business tendency surveys are conducted in the United States, including well known surveys by the Institute for Supply Management and the Federal Reserve Bank of Philadelphia. These surveys have been found to be helpful for understanding current and future movements in economic conditions. U.S. business tendency surveys are very similar in design and operation. Without exception, U.S. survey data have been compiled as simple unweighted indexes, either of the balance or diffusion form.²

Outside the United States, business tendency surveys tend to use different survey designs and aggregation methods. The Organization for Economic Cooperation and Development (OECD) recommends weighting survey responses using industry type, firm size and response rate within firm size category.³ More than 30 countries currently conduct business tendency surveys and most follow OECD weighting recommendations in full or in part.

This paper will evaluate the impact of alternative weighting methodologies using data from the Texas Manufacturing Outlook Survey (TMOS). The costs and benefits of various types of index weighting will be discussed. Numerous index designs—including OECD recommended and others—will be tested against measures of economic activity to determine the impact and

² For a discussion of the difference between the balance form and diffusion form for index design, see section 4.

³ OECD “Business Tendency Surveys: A Handbook” Organisation for Economic Co-operation and Development, Statistics. 2003

value of weighting responses for producing indexes that provide the most information available about current economic conditions.

The Texas Manufacturing Outlook Survey

The Federal Reserve Bank of Dallas began a business tendency survey in May 2004 to assess economic activity in the state. TMOS is conducted monthly to monitor changes in the state's manufacturing sector.

Texas—one of the country's largest and fastest growing manufacturing states—produced \$139.7 billion worth of manufacturing value added in 2006, or 8.7 percent of the U.S. total. The state ranks second behind California in factory production and first as an exporter of manufactured products. Texas produces a wide variety of goods. According to 2005 estimates, Texas turns out 19 percent of U.S. petroleum and coal products and nearly 15 percent of chemical products. The state also produces just over 13 percent of the nation's value added in computer and electronics products, nearly 10 percent of its machinery, and 10 percent of its nonmetallic mineral products, such as brick, glass and cement.

The sample for the TMOS was drawn from *Reference USA*, a database that lists over 45,000 manufacturing operations in the state. This list is compiled using various sources, including phone records, financial records and phone calls directly to manufacturers. The sample focused on the headquarters for each firm, so as not to get duplicate responses from branch or affiliate operations. The sample was random but focused on larger firms with more than 100 employees. For some industry categories, it was necessary to sample firms with fewer than 100 employees.

Letters of invitation were sent to 2500 manufacturers in April 2004 and 130 business leaders agreed to participate. In the first year of the survey, between 65 and 80 firms responded each month. The survey panel was deepened in January 2007, with nearly 250 manufacturers receiving the survey questionnaire each month, boosting the response rate to between 115 and 126 respondents.

Table 1 details the breakdown of TMOS survey respondents by firm size as compared to the distribution of firm size in the population. Since the original sample intentionally excluded most firms with less than 100 employees, Table 1 also examines the relationship between the TMOS firm size proportions above 100 and the population firm size proportions above 100. Table 2 provides a similar breakdown of TMOS survey respondents by industry type.

With respect to Table 1 and the firm sizes > 100 , a Chi-square Goodness-of-Fit test indicates that the distribution of the respondents by firm size is statistically indistinguishable from the distribution of the firm sizes in the population. In a similar manner, in Table 2 the Chi-square Goodness-of-Fit test of the distribution of the respondents by industry affiliation is statistically indistinguishable from the distribution of manufacturing firms by industry in the population. Thus, given these two Goodness-of-Fit test results, we conclude that, with respect to

firm size and industry affiliation, our respondent panel is representative of our manufacturing population.⁴

Data collection for the survey begins mid-month, and results are released near the end of each month.⁵ Participants receive an e-mail with a unique token permitting one-time access to a secure internet server containing the questionnaire. Answers are electronically transmitted to the Federal Reserve Bank of Dallas. The survey form is available for seven consecutive business days. Respondents are reminded that the survey is closing, and late responses are not accepted. The process is strictly confidential; the names of individual firms participating in the survey are never disclosed.

Manufacturers are asked to answer questions about their Texas operations, such as changes in production, capacity utilization, orders, inventories, prices, employees and capital expenditures. Two questions solicit opinions about all of their operations, including those outside the state. Business leaders are asked to assess company outlook and the level of general business activity. A copy of the survey form is presented in Appendix 1.

For all questions, participants are asked whether the indicator has increased, decreased or remained unchanged. Answers cover changes over the previous month and expectations for activity six months into the future. Survey respondents are instructed that their answers should exclude the effects of normal seasonal changes.

Seasonal Adjustment

In most instances, business leaders are aware of the underlying seasonal pattern inherent in their industry and make decisions adjusting for these typical fluctuations. Participants in the Texas Manufacturing Outlook Survey are asked to respond to survey questions “other than the normal seasonal change”—effectively participants are asked to seasonally adjust their responses themselves.

Despite such qualifiers, most Business Tendency Surveys routinely adjust their survey variables for seasonal variation.⁶ Most use either the Census Bureau’s X11 or X12-ARIMA procedures or TRAMO/SEATS.⁷

⁴The Chi-square Goodness-of-Fit statistic for firm sizes greater than 100 in Table 1 is $\chi^2_5 = 2.27$ ($p = 0.81$). The critical value at the 5 percent confidence level is 11.07. Therefore the null hypothesis is not rejected. The Chi-Square Goodness-of-Fit statistic for industry type in Table 2 is $\chi^2_{20} = 19.29$ ($p = 0.50$). Critical value at the 5 percent level is 31.41. Once again, we are unable to reject the null hypothesis.

⁵ Survey data are released on the last Monday of the month being observed. For example, September data is released on the last Monday of September. Collection of survey data typically commences nine business days prior to public release. Depending on where the last Monday falls during the month, survey data are collected starting on the second or third Tuesday of the month—thus the start of the period can be anytime between roughly the 11th and the 18th.

⁶ With the exception of the Texas Manufacturing Outlook Survey, all Federal Reserves seasonally adjust their series.

It is hard to know the extent to which respondents are able to fully comprehend and adjust their own seasonal pattern. It may be easier for participants to be aware of the seasonal variation more in some aspects of their business than in others. Some variables may not have seasonal variation.

In this study we have chosen to pre-test all of our survey variables using the standard F-test for stable seasonality reported in X11 and X12 programs. Each question's indexes are independently tested for a statistically significant seasonal pattern and are adjusted only if necessary. See Appendix 2 for details about the seasonality test.

Index Design for Business Tendency Surveys

Business tendency surveys are useful because qualitative survey responses can be aggregated into quantifiable indexes. There are several business tendency surveys evaluating activity in the United States. A national survey is conducted by the Institute of Supply Management. The Federal Reserve System conducts regional surveys at some reserve banks. All of these surveys—including the Texas Manufacturing Outlook Survey—aggregate responses into simple balance indexes that give an equal weight to all respondents.

Business tendency surveys are also conducted in all OECD countries and many non-OECD countries. Through years of experience working with member and nonmember countries, the OECD has developed recommendations for collecting and evaluating data generated from business tendency surveys. The OECD publishes a handbook of recommendations for improving survey performance, including methodology for weighting business tendency surveys using firm size, industry and survey response weights.⁸

Unweighted Indexes

A balance index provides insight into the dispersion of change by measuring the proportion of responses that contribute positively and negatively to the index.⁹ For each unweighted balance index on a given question the i th respondent in the k th industry is registered

⁷ Unlike most series that can vary between negative infinity and positive infinity, responses from business tendency surveys are bounded. (Typically between -100 and 100 or between zero and 100). Rather than straightforwardly applying the X11 or X12 programs to the balance (diffusion) indexes, some researchers suggests that care must be used when seasonally adjusting bounded series. Analysis of special procedures for seasonally adjusting bounded series suggests the error introduced by using traditional seasonal adjustment techniques is trivial.

⁸ According to the Handbook, the European Union established a standard framework for business tendency surveys during the 1970s. The OECD worked with the EU during the 1990s to adapt this system for use by other countries. The European Commission is preparing a handbook of recommendations for business tendency surveys. Prior to that handbook, the OECD handbook provided the only international standards for best practice for the development of business tendency surveys.

⁹ Balance indexes are centered at zero and range from -100 to +100. Some business tendency surveys use a diffusion index that is centered around 50 (with index values from 0 to 100). Balance indexes can easily be converted to diffusion indexes by dividing by 2 and adding 50.

as -1 if the response is “decrease”, 0 if the response is “no change” or +1 if the response is increase. Let $R_{ik} = \{-1,0,+1\}$ denote the response of the i th respondent in the k th industry to a specific survey question.

$$R_{ik} = -1,0,+1$$

The unweighted industry balance index is:

$$b_k = \frac{\sum R_{ik}}{n_k} \cdot 100$$

where

$$k = 1,2,\dots,K$$

represents the index of the K industries and n_k represents the number of respondents in the k th industry.

Then an overall (total) unweighted balance index for a particular question can be calculated as

$$b_u = \frac{b_1 n_1 + b_2 n_2 + \dots + b_K n_K}{n_1 + n_2 + \dots + n_K} \cdot 100 = \frac{\sum_{i,k} R_{ik}}{N} \cdot 100$$

When all firms report that activity has increased, an index will register 100. An index will register -100 when all firms report a decrease. An index will be zero when the number of firms reporting an increase or a decrease is equal.

Weighted Indexes

A weighted index groups participants by size and industry-type and weights the responses according to their contribution to the economic activity of the area surveyed.

There are three layers of weighting:

- (1) By Industry
- (2) By Size of the Respondent
- (3) By Participation Rate as a Percent of Response

(1) By Industry (Industry Weights by Value Added)

$$b_{(1)} = \frac{b_{u1}VA_1 + b_{u2}VA_2 + \dots + b_{uK}VA_K}{VA_1 + VA_2 + \dots + VA_K}$$

where VA_k represents the value added produced by the k th industry and b_{uk} represents the unweighted balance index of the k th industry.

(2) By Industry and by Size of the Respondent

$$b_{(2)} = \frac{b_{e1}VA_1 + b_{e2}VA_2 + \dots + b_{eK}VA_K}{VA_1 + VA_2 + \dots + VA_K}$$

where the employment-weighted balance indices by industry are calculated by the formula

$$b_{ek} = \frac{\sum_i e_{ik}R_{ik}}{\sum_i e_{ik}}$$

and where e_{ik} is the employment size of the i th respondent's firm that belongs to the k th industry.¹⁰

(3) By Industry, Size of the Respondent, and Participation Rate (Industry Weights by VA, Respondent Weights by Firm Size, and Participation Rate as a Percent of Response)

$$b_{(3)} = \frac{b_{ep1}VA_1 + b_{ep2}VA_2 + \dots + b_{epK}VA_K}{VA_1 + VA_2 + \dots + VA_K}$$

where the firm-participation rate-weighted balance indices by industry are calculated by the following formula

¹⁰ TMOS indexes break firms into seven size categories as detailed in Table 1. Each respondent is assigned a size value which is equal to the midpoint of its size category. For example, category 1, which contains firms with 99 or fewer employees, is assigned a value of 50. Category 2, which contains firms with 100 to 249 workers, is assigned a value of 175. The largest category contains firms with 10,000 workers or more workers and is assigned the value 10,000. (There are no participating TMOS firms in this size category).

$$b_{epk} = \frac{\sum_i e_{ik} \left(\frac{1}{p_{ik}} \right) R_{ik}}{\sum_i e_{ik} \left(\frac{1}{p_{ik}} \right)}$$

where p_{ik} is the proportion of firms in the i th firm's size class (as detailed in Table 1) that are reporting in the k th industry.

Evaluating Various Weighting Methodologies

More than 50 countries conduct business tendency surveys to monitor and forecast changes in economic activity. As detailed in Appendix 2, most countries tend to follow OECD recommendations in full or in part, generating indexes by weighting survey responses by firm size and often also by industry type.

There are no observable trends associated with the choice of various weighting methodologies. For example, it does not appear that the size of the country or the size of the survey sample contributes to the decision to weight survey responses. It is notable, however, that surveys conducted in the United States use unweighted index construction.¹¹

As shown in Charts 1a-f, the choice of weighting scheme can significantly affect index results.¹² The pattern of movements between the two types of indexes is quite similar, but the standard deviations of unweighted indexes are smaller than for weighted indexes.

It is possible unweighted indexes provide some smoothing of survey responses that reduce noise in the data series. It is also possible that smoothing hides useful information about changes in the economy.

Conceptual differences between weighted and unweighted indexes

Unweighted indexes give equal weight to all responding firms. Changes in activity at large and small firms have an identical impact on the index. Responses also contribute equally regardless of the industry's contribution to the overall economy.

¹¹ A quick survey found very little explanation for the choice of index construction methodology. An exception is Austria, where the following note was listed: "The results of the Austrian business cycle surveys are not weighted, as we found out that weighting would not lead to remarkably improved results (this is due to the fact, that large companies are overbalanced in our sample, i.e. the sample is "weighted by itself")."

¹² The TMOS weighted indexes were calculated using slightly different methodology than the OECD recommends. As recommended, value added weights were used for most variables—the exception being employment, where employment weights were used. Statistical tests determined that employment provided a superior fit to target variables than value added for the employment question. With both employment and value-added weights, a three-year moving average was used. Value added estimates were drawn from the Census and Annual Surveys of Manufacturers. Industry value added weights for each year were formed by averaging the Census/Annual Survey figures from the three previous available years. Firm size weighting was achieved using the firm size breakdowns noted in table 1 and discussed in footnote 10.

An unweighted index may fail to capture important information about changes in the business cycle by giving less weight to large firms or industries. For example, changes in an economy's manufacturing output can be impacted heavily by the performance of large firms: an increase in production at a large firm can overwhelm the effect of many small firms reducing production. But in this example an unweighted index would aggregate large and small firms identically, registering a reduction in production when, in fact, production increased.

An unweighted index can be distorted because it gives the most impact to changes in the industry with the most reporting firms. For example, if many responses are received from a small industry but only a few responses are received from a large industry then the index would be weighted more heavily towards changes in the smaller industry. Similarly—even with full participation—an unweighted index could be distorted if a relatively small industry is comprised of numerous firms and a larger industry has a handful of large companies.

A weighted index requires that within each industry and firm class size the sample size is large enough to get a representative estimate of the change in that industry or firm class size. Insufficient sample sizes increase the likelihood that idiosyncratic firm issues, such as management difficulties or supply chain disruptions, will have a significant impact on the overall index. An unweighted index averages the responses of all firms, making it less likely that idiosyncratic firm developments impact the index. An unweighted index also reduces volatility stemming from changes in participation of responding firms since no responding firm is more important than another

If larger standard deviations are the result of important movements in the economy, then weighted indexes should have a tighter relationship with other measures of economic activity than an unweighted index.¹³

Econometrically evaluating alternative weighting schemes for the Texas Manufacturing Outlook Survey

As explained earlier, TMOS indexes can be calculated using various weighting schemes based on firm size, participation rate and output. Alternative indexes were tested for their ability to explain other variables of interest. In this paper, we limit our discussion to two weighting schemes: 1) an unweighted index that corresponds to the methodology commonly used in the United States and 2) the fully weighted scheme recommended by the OECD and used by many countries.

Regional business tendency surveys are valuable because there are limited data series available to evaluate economic conditions. However, this paucity of data also introduces a challenge for evaluating the effectiveness of survey results. A lack of knowledge about the true

¹³ It would be interesting to look into the source of the larger standard deviation – is it caused by firms moving in and out of the sample or by higher weights on more volatile firms and industries. This can be examined by creating changes in an index that from month-to-month has only the same firms reporting as the previous month. The difference between this hybrid series and our current series is due to firms coming in and out of the sample.

value of the variable in question hampers the method of determining which index weighting method is most appropriate.

For example, if there was an accurate measure of new orders in manufacturing then that series could be regressed on alternative designs of weighted survey indexes involving new order questions to see which best fit the true data.

So a number of dependent variables are collected for Texas and national manufacturing industries, offering a broad target of indicators that can be evaluated to determine how various weighting schemes help explain factory activity.

The alternative index designs are evaluated against two Texas variables. Texas manufacturing employment is the only analogous measures of the data asked about in the survey. A monthly Texas manufacturing coincident index (TXMCI) conforms well to real output in manufacturing and has the added advantage of being monthly.¹⁴

Survey variables are also tested against national manufacturing activity. While the survey is designed to measure Texas manufacturing activity, it is possible that short-run changes in Texas manufacturing may reflect national patterns. The alternatively weighted indexes are also evaluated related to changes in U.S. manufacturing employment, U.S. industrial production and levels of the Institute of Supply Management manufacturing composite, production and new orders indexes.

How Well Does the Survey Reflect the Economy?

To analyze the usefulness of the TMOS in tracking the economy, the Texas and national variables are tested to see if one or more of the balance indexes from the TMOS can explain and/or predict their movements. The variables of interest are referred to as “target” variables and the balance indices as the “survey” variables.

In what follows, we will evaluate the usefulness of the “current” TMOS balance indexes as they relate, first, to good in-sample fit and, second, to good out-of-sample forecasting performance vis-à-vis naïve forecasting methods.¹⁵

To determine if the survey variables are useful we examine their explanatory power in an autoregressive distributed lag model framework. If the survey variable in question does not

¹⁴ The TXMCI aggregates changes in quarterly real gross Texas product in manufacturing and Texas manufacturing employment using a dynamic Kalman filter due to Stock and Watson (1991.) Applying this procedure to the two component series results in an index in which 79 percent of its variance is due to movements in Texas manufacturing RGSP and 21 percent is due to movements in manufacturing employment. For a more complete description of applying the Stock and Watson procedure to construct regional coincident indexes and a description of how quarterly data is used with monthly data to produce a monthly index see Phillips (2005).

¹⁵ This analysis is only evaluating TMOS survey indexes based on questions involving changes in current conditions compared to the previous month as opposed to the questions asking about changes expected six months from now.

explain a significant amount of the variation in the target variable beyond what can be explained solely by the past history of the target variable, then the survey variable is not an effective instrument in explaining the target variable. In contrast, if the survey variable does explain a statistically significant amount of the variation in the target variable, the survey variable is judged to be useful. This judgment is in-sample. That is, in this section we evaluate the effectiveness of the proposed survey variables only on the basis of estimation and hypothesis testing applied to our entire data set.

Let y denote a target variable of interest that is in stationary form, e.g. monthly percentage change in Texas employment. Second, let x denote a survey variable whose relevance for explaining the target variable is in question. Here we are assuming that the survey variables (balance indexes) are all stationary even though our sample of balance index values is short and in certain instances have short term trends in them. In looking at balance indexes of longer history (like the Philadelphia Fed balance indexes) the fact that the balance indexes are bounded (they must be between -100 and +100) eventually ensures stationary.

The econometric model we entertain is a second-order autoregressive distributed lag model

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \beta x_{t-d} + \varepsilon_t \quad (1)$$

with a potentially second order autocorrelation of the errors

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + u_t \quad (2)$$

In each estimation, the two autoregressive terms in equation (1) were retained while the choices of the distributed lag delay were taken to be $d = 0$ or $d = 1$ and the choices of autoregressive order for the error equation (2) were taken to be either 2, 1, or 0 (the uncorrelated error case).¹⁶ For a fixed choice of the delay parameter d , the autocorrelation function of the least squares residuals of equation (1) and the t-statistics of the autocorrelation coefficients of equation (2) were examined and, depending on the significance of the t-statistics of the autocorrelation coefficients, an order of the autocorrelation error was chosen. In the case that no autocorrelation seemed present, Ordinary Least Squares was used to estimate equation (1); otherwise, Generalized Least Squares of a specified order of autocorrelation was used (in SAS Proc AUTOREG). Then, for choosing across the possible of orders of d , the model with the highest R^2 was chosen.¹⁷

¹⁶ There is uncertainty with respect to the timing of the relationship between the “current” situation survey questions we examine and the “current” target variables of interest. This is because of the difference in the timing of the collection period for the survey results and the collection period for the target variables the survey variables are trying to predict. In most cases we found that the optimal delay (d) is 1 in that this delay, more so than the delay $d = 0$, maximizes the explanatory power of the survey variable. Interesting the case of $d = 1$ gives a more advanced notice of the future movement in the target variable than the delay $d = 0$.

¹⁷ In estimating the autoregressive distributed lag model of equation (1), some econometricians might prefer to “lag the equation to white noise” by adding lags of the dependent variable and/or the explanatory variable until the

Once an “optimal” test equation was established for the given survey variable (x) and a given target variable (y), the t-statistic and corresponding one-tailed p-value was recorded and evaluated for statistical significance. The one-tailed (right-tailed) p-value was examined because, a priori, it is expected that the survey variables will be positively correlated with the target variable, given that the survey variables are balance indices and the larger (smaller) the balance index, the more positive (negative) the performance of the economy as measured by the target variables.

Of course, when considering a given target variable, there may be more than one survey variable that has a statistically significant degree of explanatory power vis-à-vis equation (1). In the case that several survey variables have statistically significant explanatory power (say, $p < 0.10$), then one might choose to use judgment and distinguish between them based on not only the p-value but also relevancy of the survey question to the particular target variable of interest.

Results

Selective results of the econometric evaluation are presented in Tables 3 through 8. We examine the efficacy of our survey indexes both in-sample and out-of-sample. These tables report results with the best overall fitting survey variables for each target variable and with two index calculation methods, designated as unweighted and weighted.¹⁸ The first index scheme represents the unweighted calculation of the balance index, where the responses are all given equal weight regardless of the size of the firm, participation rate or industry. In the weighted designation, the individual responses within an industry are weighted by firm size and participation rate in calculating each industry balance index. In addition, the industry balance indices are aggregated to the total manufacturing balance index using value-added or employment size industry weights.

For the out-of-sample forecasts, we start out by restricting the sample to the first 24 observations and forecast one period ahead using the regression results from that sample. Then

residuals of the model are white noise as compared to fixing the autoregressive and distributed lag orders of the model and then modeling the errors as potentially be autocorrelated in equation (2) as we have chosen to do. However, in our case we had so many variable combinations that we choose to set up an econometrically flexible and plausible “production run” model that could easily be pre-tested, estimated, and summarized as compared to developing the customized, lag-to-white noise results for the many possible variable combinations we were faced with. Given that our econometric model of equations (1) and (2) scarcely ever exhibited autocorrelation in the errors of the model and that often the second autoregressive term was not statistically significant, we feel that, if anything, the reported statistical significances of the survey variables are likely to be understated rather than overstated. Likewise, when considering the out-of-sample forecasting results reported below we also might be able to improve the forecasting performances of the survey index models by customizing them based on “lagging the model to white noise.”

¹⁸ By “best overall fitting” we mean that we looked at the autoregressive distributed lag models of each survey variable across the two weighting options we report and highlight the survey variable that had the most explanatory power.

we add one observation, rerun the regression and again forecast one period ahead. With this procedure we generate 13 out-of-sample forecasts.

In-sample results. The TMOS survey indexes evaluated are employment, production, general business activity and new orders. Also, a composite index is generated including new orders, production and employment using weights similar to the ISM composite (adjusted for the fact that TMOS does not have the supplier delivery variable).

Table 3 reports statistical results of our examination of two regional target variables: Texas manufacturing employment and the Texas Manufacturing Coincident Index (TXMCI). The p-value in the table represents the one-tailed significance of the survey variable. Overall, the index variables do a fairly good job explaining Texas manufacturing employment and TXMCI. The TMOS employment index, lagged one period, explains changes in Texas manufacturing employment well. The unweighted form of the index is more significant and the equation has a higher adjusted- R^2 . The weighted forms of all four indexes are significant in the TXMCI equations. The new orders and composite indexes are also significant in explaining changes in the TXMCI in their unweighted forms.

TMOS indexes also do a good job explaining national indicators. Table 4 reports statistical results for five national target variables: U.S. manufacturing employment, the manufacturing component of U.S. industrial production (IP), the Institute for Supply Management's production (ISMP), composite (ISMC) and new orders (ISMNO) indexes. The TMOS employment index is significant in explaining U.S. manufacturing employment. The unweighted form of the employment index is more significant than the weighted form, as is the case for Texas manufacturing employment. For U.S. industrial production, TMOS production, general business activity, new orders and the composite index are all significant in their respective equations. The weighted form of the indexes has a higher degree of significance (significant at the one percent level) except for the new orders index, where the unweighted form does somewhat better.

In the ISM equations, except for the general business activity index in the ISMC and ISMP equations, the weighted form of every TMOS index analyzed is significant. Table 4 shows that the unweighted form of the TMOS indexes are also significant except in two cases, the new orders and composite indexes in the ISMP equation.

In 12 of the 17 equations, the weighted form of the TMOS indexes does a better job explaining the targeted national variables, with both higher significance and better fit. Table 7 scores these results.

Out-of sample results. Tables 5 and 6 illustrate the out-of-sample results of the weighted and unweighted forms of the indexes. To judge the efficacy of the out-of-sample forecasts of our index variables, we use a naïve Box-Jenkins autoregressive model and compare the mean absolute errors (MAE) for employment and U.S. industrial production and the mean absolute percent error for the ISM for the naïve model and our models. For the targeted Texas variables, the unweighted form of the TMOS indexes performs better. The only index that has a lower MAE with the weighted form is the general business activity index in the TMCI equation (Table

5). When comparing these results with a pure autoregressive model, it is not clear whether the survey variables have much additional predictive power. Although we have not performed any formal tests, the MAEs from the naïve models are probably not significantly different than the MAEs from the models including the survey variables.

The out-of-sample forecasts of the national indicators follow a similar pattern to the in-sample estimates. Table 6 shows that the weighted form of the TMOS indexes have lower MAEs for a majority of the national indicators. The weighted forms of the indexes forecast better except for the general business activity index in all three of the ISM equations. Overall, in 14 out of 17 national forecasts, the weighted indexes have lower MAEs (see Table 8). The models with the survey variables seem to predict better than the naïve model in a majority of the equations, but again, the MAEs are very close and probably not significantly different from each other.

Conclusions

Using a large and diverse sample, weighting by firm, participation rate and industry size has strong theoretical justification. But in practical applications with smaller samples and samples that change from month to month, the choice can be unclear. In this paper we use the three years of data that we have for the TMOS to study whether weighted or unweighted indexes best fit selected Texas and U.S. manufacturing data.

Overall, our results show that several of the balance indexes from the TMOS have significant explanatory power for some important regional and national variables. For both the regional and national indicators, the weighted indexes of production and new orders seemed to perform better, while the unweighted seemed to work best for employment.

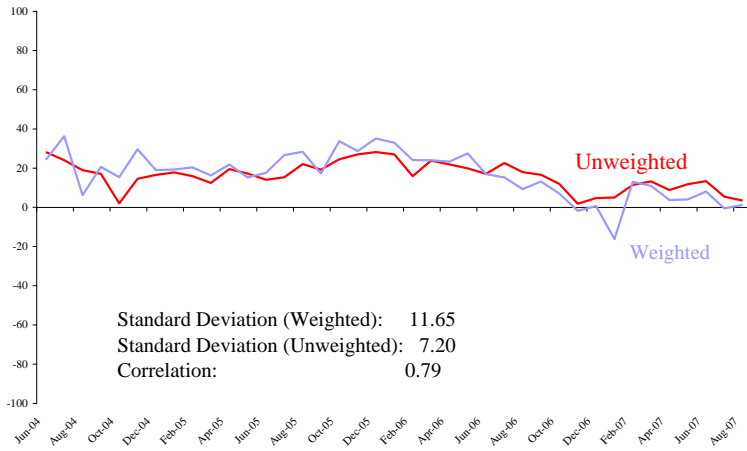
The unweighted form of the TMOS employment index fits both Texas and U.S. manufacturing employment very well. For the Texas Manufacturing Coincident Index, we get a better fit with the weighted TMOS indexes.

For the national target variables, the weighted TMOS indexes fit better in 12 out of 17 equations, although the differences between the weighted and unweighted results are marginal in some cases. Overall, the weighted indexes explain the target variables better 16 out of 22 times.

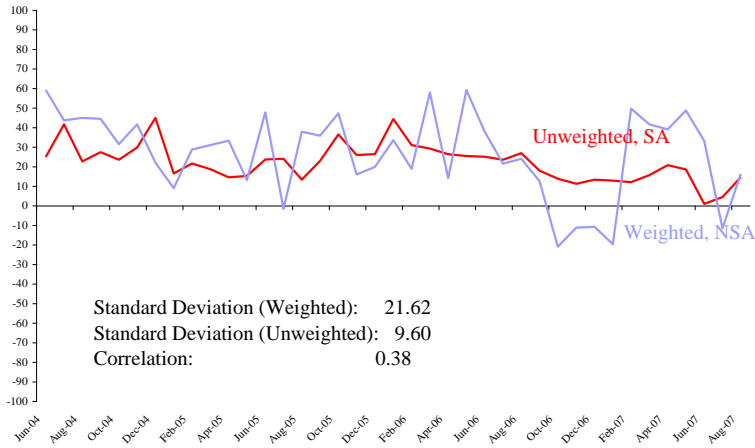
In the out-of-sample forecasts, the unweighted indexes forecast better than the weighted for the Texas economic indicators. The pattern is very similar to the in-sample results for the national target variables: the weighted indexes predict better in 13 out of the 17 cases. However, given our small sample size, it is not certain whether the models including our indexes perform better than a naïve, autoregressive model. With a larger data set, we hope to obtain more conclusive results.

Charts 1a-f Select Indexes from Texas Manufacturing Outlook Survey—Comparison of Weighted and Unweighted Indexes

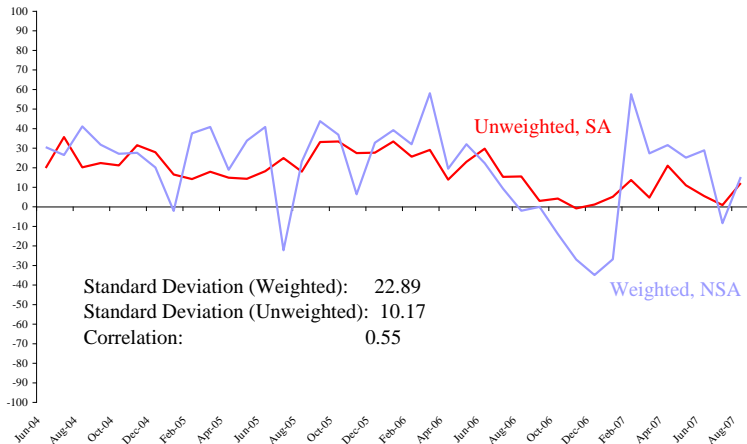
Employment, SA



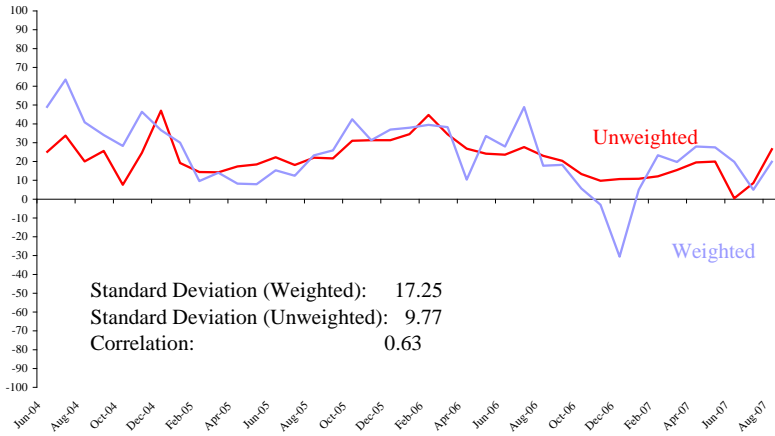
Production



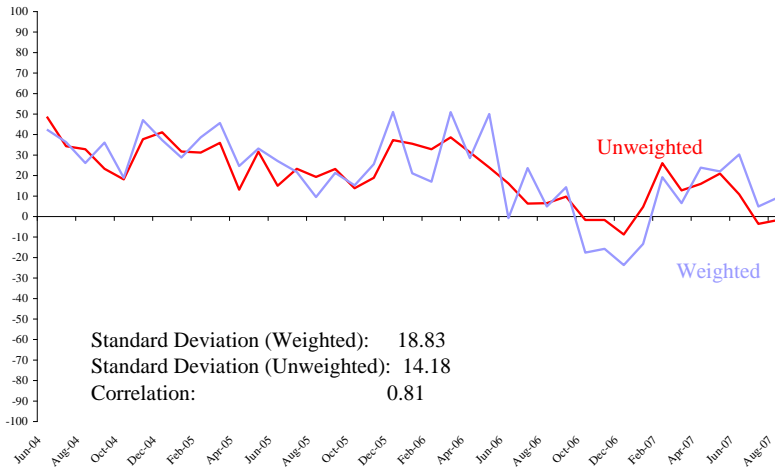
New Orders



Shipments, SA



General Business Activity, NSA



Composite, SA

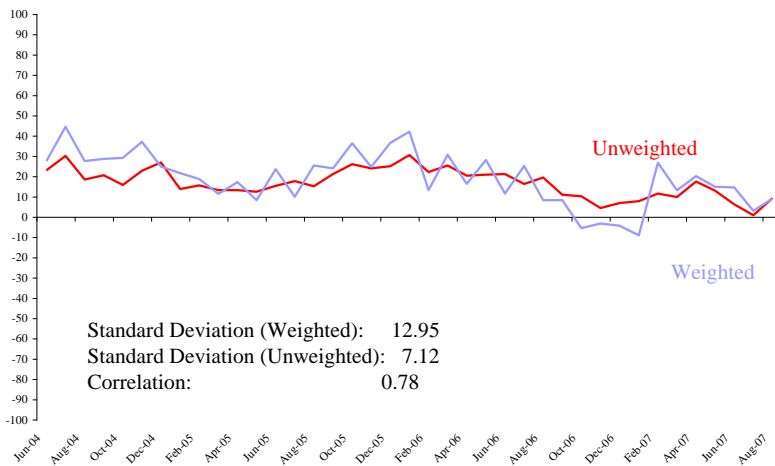


Table 1.
Comparison of Population Proportions with Survey Proportions by Firm Size

By size class	Total number of Texas Manufacturing Firms August 2007	Population percentage of firms by size category	Population percentage of firms by size category excluding firms with fewer than 100 employees	TMOS Sample in August 2007	TMOS Sample percentage of firms by size category	TMOS Sample percentage of firms by size category excluding firms with fewer than 100 employees	TMOS Sample as a Percent of Total Population of Firms
<100	5101	74.5%		29	26.1%		0.6%
100-249	1091	15.9%	62.5%	52	46.8%	63.4%	4.8%
250-499	360	5.3%	20.6%	19	17.1%	23.2%	5.3%
500-999	176	2.6%	10.1%	5	4.5%	6.1%	2.8%
1000-4999	107	1.6%	6.1%	6	5.4%	7.3%	5.6%
5000-9999	9	0.1%	0.5%	0	0.0%	0.0%	0.0%
>9999	2	0.0%	0.1%	0	0.0%	0.0%	0.0%
Total	6846	100.0%	100.0%	111	100.0%	100.0%	1.6%

Table 2.
Comparison of Population Proportions with Survey Proportions by NAICS Code

NAICS		Total number of Texas Manufacturing Firms August 2007	Population percentage of firms by NAICS Code	TMOS Sample in August 2007	TMOS Sample percentage of firms by size category	TMOS Sample as a Percent of Total Population of Firms
311	Food	420	6.1%	10	9.0%	2.4%
312	Beverage & Tobacco	46	0.7%	1	0.9%	2.2%
313	Textile Mills	48	0.7%	0	0.0%	0.0%
314	Textile Product Mills	101	1.5%	0	0.0%	0.0%
315	Apparel	74	1.1%	1	0.9%	1.4%
316	Leather and Allied Products	33	0.5%	0	0.0%	0.0%
321	Wood	266	3.9%	5	4.5%	1.9%
322	Paper	177	2.6%	7	6.3%	4.0%
323	Printing & Related	340	5.0%	3	2.7%	0.9%
324	Petroleum & Coal Products	116	1.7%	2	1.8%	1.7%
325	Chemical	473	6.9%	5	4.5%	1.1%
326	Plastics & Rubber Products	381	5.6%	7	6.3%	1.8%
327	Nonmetallic Mineral	347	5.1%	7	6.3%	2.0%
331	Primary Metal	201	2.9%	5	4.5%	2.5%
332	Fabricated Metal Product	1238	18.1%	23	20.7%	1.9%
333	Machinery	750	11.0%	15	13.5%	2.0%
334	Computer & Electronic	603	8.8%	7	6.3%	1.2%
335	Electrical Eqpt, Appliance & Comp	220	3.2%	2	1.8%	0.9%
336	Transportation Equipment	364	5.3%	4	3.6%	1.1%
337	Furniture & Related	236	3.4%	4	3.6%	1.7%
339	Miscellaneous	412	6.0%	3	2.7%	0.7%
		6846	100.0%	111	100.0%	1.6%

Table 3.
In-sample fits for Texas dependent variables
(Alternative weighting schemes)

Dependent variable	Unweighted		Weighted	
	R ²	p ¹	R ²	p ¹
1. Texas payroll employment				
a. Employment index	0.2199	0.0026	0.0844	0.0526
2. Texas Manufacturing Coincident Index				
a. Production index	0.3315	0.1687	0.4018	0.0178
b. General business activity index	0.3374	0.1368	0.4025	0.0174
c. New orders index	0.3727	0.0436	0.4283	0.0078
d. Composite index	0.3864	0.0285	0.3935	0.0229

¹ One-tail p-values on BOS index (lagged one period) included in second order autoregressive model.

Table 4.
In-sample fits for various U.S. dependent variables
(Alternative weighting schemes)

Dependent variable	Unweighted		Weighted	
	R2	p ¹	R2	p ¹
1. U.S. payroll employment				
a. Employment index	0.1275	0.00775	0.0725	0.01825
2. U.S. Manufacturing Industrial Production				
a. Production index	0.1028	0.0535	0.1627	0.0139
b. General business activity index	0.1540	0.0169	0.2844	0.0008
c. New orders index	0.2749	0.0010	0.1475	0.0196
d. Composite index	0.1531	0.0172	0.2439	0.0022
3. ISM (production)				
a. Production index	0.4993	0.4034	0.6448	0.0005
b. General business activity index	0.5281	0.0825	0.5079	0.2183
c. New orders index	0.5152	0.1875	0.6024	0.0034
d. Composite index	0.5043	0.2697	0.5155	0.1476
4. ISM (Composite)				
a. Production index	0.6329	0.1703	0.6965	0.0045
b. General business activity index	0.6541	0.0476	0.6267	0.2684
c. New orders index	0.6621	0.0303	0.6818	0.0100
d. Composite index	0.6515	0.0553	0.6585	0.0372
5. ISM (New orders)				
a. Production index	0.3637	0.2610	0.4855	0.0035
b. General business activity index	0.3910	0.0903	0.3920	0.0872
c. New orders index	0.4382	0.0187	0.4264	0.0275
d. Composite index	0.3933	0.0835	0.4873	0.0036

¹ One-tail p-values on BOS index (lagged one period) included in second order autoregressive model.

Table 5.
Out-of-sample forecast results for Texas variables
(Alternative weighting schemes)

Dependent variable	Unweighted ¹	Weighted ¹
1. Texas payroll employment		
a. Employment index (lags only, MAE = .00110)	0.00104	0.00124
2. Texas Manufacturing Coincident Index		
a. Production index	0.00288	0.00321
b. General business activity index	0.00363	0.00302
c. New orders index	0.00295	0.00332
d. Composite index (lags only, MAE = .00298)	0.00316	0.00322

¹Mean absolute error

Table 6.
Out-of-sample forecast results for U.S. variables
(Alternative weighting schemes)

Dependent variable	Unweighted ¹	Weighted ¹
1. U.S. payroll employment		
a. Employment index (lags only, MAE = .000789)	0.000863	0.000835
2. U.S. Manufacturing Industrial Production		
a. Production index	0.0050	0.0038
b. General business activity index	0.0040	0.0036
c. New orders index	0.0042	0.0042
d. Composite index (lags only, MAE = .00450)	0.0043	0.0039
3. ISM (production)		
a. Production index	0.0473	0.0441
b. General business activity index	0.0465	0.0503
c. New orders index	0.0456	0.0439
d. Composite index (lags only, MAE = .0469)	0.0476	0.0491
4. ISM (Composite)		
a. Production index	0.0271	0.0250
b. General business activity index	0.0259	0.0280
c. New orders index	0.0261	0.0260
d. Composite index (lags only, MAE = .0267)	0.0276	0.0276
5. ISM (New orders)		
a. Production index	0.0552	0.0485
b. General business activity index	0.0507	0.0517
c. New orders index	0.0509	0.0489
d. Composite index (lags only, MAE = .0543)	0.0557	0.0480

¹Mean absolute error for employment and industrial production, mean absolute percent error for ISM

Table 7.
Scoring table for various weighting schemes (in-sample)

<u>Texas Dependent Variables</u>	
Unweighted	Weighted
1/5	4/5

<u>U.S. Dependent Variables</u>	
Unweighted	Weighted
5/17	12/17

<u>All Variables</u>	
Unweighted	Weighted
6/22	16/22

Table 8.
Scoring table for various weighting schemes (out-of-sample)

<u>Texas Dependent Variables</u>	
Unweighted	Weighted
4/5	1/5

<u>U.S. Dependent Variables</u>	
Unweighted	Weighted
4/17	13/17

<u>All Variables</u>	
Unweighted	Weighted
8/22	14/22

Appendix 1

CONFIDENTIAL: Individual replies will not be published or quoted.

Instructions for Submitting the Survey

- Please complete this survey by checking the boxes that best describe your business's activity.

Name of Company:	ABI:
Contact:	Phone Number:
Title:	E-mail:
Address:	City/State/Zip

	January vs. December Other than the normal seasonal change	Six months from now Other than the normal seasonal change
--	--	---

Company Business Indicators Relating to Your Facilities and Products in Texas:

	Increase	No Change	Decrease	Increase	No Change	Decrease
Production						
Capacity utilization						
Volume of New orders						
Growth rate of orders						
Unfilled orders						
Volume of shipments						
Delivery time						
Materials inventories						
Finished goods inventories						
Prices paid for raw materials						
Prices received for finished goods						
Wages and benefits						
Number of employees						
Average employee workweek						
Capital expenditures						

General Business Conditions:

	Improved	Remained the Same	Worsened	Improved	Remained the Same	Worsened
How has the outlook for your company changed?						
What is your evaluation of the level of general business activity?						

Please feel free to comment on any issues that may be affecting your business.

Appendix 2

Seasonally Adjusting Indexes from Business Tendency Surveys

In this appendix, we detail the test used to evaluate seasonality.

To establish the notation of this test, assume the data is observed monthly and thus there are $k=12$ seasons (months) within the year. Let the number of observations for each of the k months be denoted by n_1, n_2, \dots, n_k and thus the total number of observations is $n = n_1 + n_2 + \dots + n_k$. Furthermore, let x_{ij} represent the seasonal/irregular component (SI) that has been previously extracted from the original data with the j -index denoting a particular month, $j = 1, 2, \dots, k$ and the i -index denoting the succession of years in which the j -th month's observations occur, $i = 1, 2, \dots, n_j$.

We use a one-way analysis of variance test referred to in the X11/X12 literature¹⁹ as the "test for stable seasonality." Essentially the test examines the extent to which the monthly sample means, $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$, differ from one another. Assuming that each month's observations follow a random variable X_j that is normally distributed with mean m_j and standard deviation, σ , the null and alternative hypotheses of the stable seasonality test are

$$H_0 : m_1 = m_2 = \dots = m_k$$

$$H_1 : m_p \neq m_q \text{ for at least one pair (p,q) .}$$

The null hypothesis represents the case of no seasonality in the data while the alternative hypothesis represents the case of seasonality in the data. The so-called analysis of variance equation is written as follows:

$$\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2 = \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2$$

where \bar{x} represents the overall (grand) mean of the data. Equivalently,

$$S^2 = S_A^2 + S_R^2$$

and the total variation in the x_{ij} , S^2 , can be can be partitioned into the variation due to the monthly averages, S_A^2 , and the residual variation, S_R^2 . If the null hypothesis is true, it follows that the statistic

¹⁹ For example, see D. Ladiray and B. Quenneville (2001, p.57-58).

$$F_s = \frac{S_A^2 / (k - 1)}{S_R^2 / (n - k)}$$

is distributed as an F random variable with $k - 1$ and $n - k$ degrees of freedom. Thus, a large p-value associated with this statistic would indicate the absence of seasonality in the data. On the other hand, a small p-value would indicate the presence of seasonality in the data. As pointed out by Ladiray and Quenneville (2001, p. 58), the assumptions of this test are likely to be violated because of the probable autocorrelation in the seasonal/irregular data, x_{ij} . Thus they recommend setting the level of significance of the test to be $p = 0.001$, one-tenth of one percent. We adopt this recommendation in all of our stable seasonal tests of the Texas Manufacturing Outlook survey variables.

Appendix 3

Weighting of Business Tendency Surveys (Organization operating survey in parenthesis)

Countries using no weighting—equal weighting

Austria (Austrian Institute for Economic Research--WIFO)
China P.R. (National Bureau of Statistics)
Hungary (GKI – Economic Research Company)
Korea (The Bank of Korea: Statistics Survey Team, Economic Statistics Department)
United States (Institute for Supply Management)
U.S. Regional Federal Reserve Banks (Atlanta, Dallas, Kansas City, New York,
Philadelphia, Richmond)

Countries using firm size weighting (either value added and/or employment)

Canada (Statistics Canada) (?)
Greece (Foundation of Economic & Industrial Research—FEIR)
Indonesia (BPS, Statistics Indonesia)
Ireland (Irish Business & Employers Confederation—IBEC)
Luxembourg (Service Central de la Statistique et des Etudes Economiques)
Mexico (Bank of Mexico)
New Zealand (New Zealand Institute of Economic Research)
Norway (Statistics Norway)
Poland (Central Statistical Office)
Russian Federation (Center of Economic Analysis)
Sweden (Konjunkturinstitutet—National Institute of Economic Research)
Turkey (State Institute of Statistics)

Countries using both firm size and industry weighting

Belgium (Banque Nationale de Belgique)
Chinese Taipei (Council for Economic Planning & Development—CEPD)
Czech Republic (Czech Statistical Office)
Denmark (Danmarks Statistik)
Finland (Confederation of Finnish Industry and Employers—CFIE)
France (Institut National de la Statistique et des Etudes Economicques—INSEE)
Germany (ifo Institut für Wirtschaftsforschung)
Hong Kong, China (Census and Statistics Department)
Italy (Istitute for Studies and Eeconomic Análisis—ISEA)
Netherlands (Centraal Bureau voor de Statistiek)
Portugal (Instituto Nacional de Estatística)
Slovakia (Statistical Office of the Slovak Republic)
South Africa (Bureau for Economic Research—Department of Trade and Industry)
Spain (Ministerio de Industria, Turismo y Comercio)
Switzerland (Swiss Institute for Business Cycle Research)
United Kingdom (Confederation of British Industry)

Unknown:

Argentina (Instituto Nacional de Estadística y Censos—INDEC)
Australia (Australian Chamber of Commerce & Industry and Westpac Banking Corporation)
Bolivia (Banco Central de Bolivia)
Brazil (Fundação Getúlio Vargas—FGV)
Bulgaria (National Statistical Institute)
Colombia (Fundación para la Educación Superior y el Desarrollo—FESD)
Dominican Republic (Banco Central de la República Dominicana)
Ecuador (Banco Central del Ecuador)
Estonia (Estonian Institute for Market Research—EIMR)
India (National Council of Applied Economic Research—NCAER)
Japan (Bank of Japan)
Kazakhstan (Economic Research Institute—ERI)
Laos (National Statistical Centre)
Lebanon (Bank of Lebanon)
Malaysia (Bank Negara Malaysia)
Peru (Banco Central de Reserva del Perú)
Philippines (Bangko Sentral ng Pilipinas)
Latvia (Central Statistical Bureau of Latvia)
Lithuania (Lithuanian Department of Statistics)
Romania (National Commission for Statistics)
Saudi Arabia (Ministry of Planning)
Singapore (Centre for Business Research & Development—CBRD)
Slovenia (Statistical Office of the Republic of Slovenia)
Thailand (Bank of Thailand)
Ukraine (Ministry of Statistics—Statistical Research Institute)
Uzbekistan (Institute of Economics—IE)
Venezuela (Banco Central de Venezuela)
Vietnam (General Statistical Office of Vietnam)

Source: Information for this table collected from the OECD Handbook of Methods and OECD website:
http://www.oecd.org/document/37/0,2340,en_2649_34353_33776933_1_1_1_1,00.html

References

OECD “Business Tendency Surveys: A Handbook” Organisation for Economic Co-operation and Development, Statistics. 2003

J.H. Stock and M.W. Watson, A Probability Model of the Coincident Economic Indicators, in: *Leading Economic Indicators: New Approaches and Forecasting Records*, K. Lahiri and G. Moore, eds, Cambridge University Press, 1991, pp. 63–89.

Phillips, Keith, “A New Monthly Index of the Texas Business Cycle,” Journal of Economic and Social Measurement, Volume 30, Number 4, 2005.

Berger, Frank, Tom Fomby, Fiona Sigalla, and Mine Yucel. 2007. “The Texas Manufacturing Outlook Survey: A Tool for Understanding the Economy.” Federal Reserve Bank of Dallas *Southwest Economy*, Issue 1 (January/February).

Deitz, Richard and Charles Steindel. 2005. “The Predictive Abilities of the New York Fed’s Empire State Manufacturing Survey.” Federal Reserve Bank of New York Current Issues in Economics and Finance Second District Highlights. 11, no. 1 (January).

Harris, Matthew, Raymond E. Owens and Pierre-Daniel G. Sarte. 2004. “Using Manufacturing Surveys to Assess Economic Conditions.” Federal Reserve Bank of Richmond *Economic Quarterly* 90, no. 4 (fall).

Keeton, William R. and Michael Verba. 2004. “What Can Regional Manufacturing Surveys Tell Us? Lessons from the Tenth District.” Federal Reserve Bank of Kansas City *Economic Review* 89, no. 3 (third quarter).

Schiller, Timothy and Michael Trebing. 2003. “Taking the Measure of Manufacturing.” Federal Reserve Bank of Philadelphia *Business Review*, fourth quarter.

http://www.oecd.org/document/37/0,2340,en_2649_34353_33776933_1_1_1_1,00.html