Emilia Tomczyk[‡]

Quantification of expectations for general economic situation on the basis of industrial production index with data revisions

Abstract

Use of appropriate data vintages and taking data revisions into account have only recently became a staple of applied econometric analysis. In this paper, the topic of data vintage in regression quantification procedures is readdressed for survey data on general economic situation. From empirical analysis it follows that quantification of survey data on general economic situation on the basis of industrial production index does not present a significant improvement over the use of response balance. Additionally, results obtained for real-time and end-of-sample data are very similar and do not suggest superiority of any of these two data vintages as far as quantification of survey data on general economic situation is concerned.

Keywords: end-of-sample (EOS) data, real time (RTV) data, data revisions, survey data, general economic situation, expectations, quantification, regression method

JEL classification: C82, C83, D84

[‡] Institute of Econometrics, Warsaw School of Economics.

1. Introduction

Every researcher attempting to perform an aggregate analysis of qualitative survey data faces the problem of quantification of qualitative responses into time series data. In this paper, quantification of business survey data on general economic situation is accompanied by evaluating whether data vintage influences results of quantification procedures used to convert categorical questionnaire data into qualitative time series. Data revisions introduce an additional degree of uncertainty to exploration of business tendency survey data, and hence seem to offer an interesting topic of analysis for the 2015 CIRET seminar with a special focus on Economic Cycles and Uncertainty.

Questionnaires on general economic situation, or general business conditions, are a great challenge with respect to quantification of survey data. In both widely used approaches to quantifying survey data, that is, probabilistic and regression methods, it is necessary to define an *objective* economic variable to scale *subjective* survey data. While (more or less) neutral equivalents for series such as production and employment levels, prices, or volume of exports are available in official statistics, there is no *objective* or *unique* measure of general economic conditions. Business tendency indicators available in statistical publications (for example, in Statistical Bulletins published by the Polish Central Statistical Office – CSO) are themselves based on survey data and therefore inappropriate as a source of scaling factors for other surveys.

Two approaches to this problem have been proposed in literature. One of them offers GDP as a proxy for general business conditions and therefore a suitable *objective* equivalent of survey data on business situation. Unfortunately, GDP data are not available with monthly frequency; this drawback limits their usefulness for the purpose of modeling and quantifying monthly business survey data. Also, GDP values production in terms of purchasers' and other final demand sectors' prices, and therefore may not be an optimal measure of business conditions observed in manufacturing sectors.

The other approach recommends the use of industrial production as a proxy for general economic situation. This line of inquiry enjoys a long history: in its *Statistical Releases and Historical Data* section, the Federal Reserve Board points to the industrial production index as a measure of current business conditions dating back to the founding of the Fed system. Since then, indicators of industrial production served as proxies for business conditions in numerous applied economics papers (for an example, from securities markets, see Chen *et al*., 2007). In this paper, I continue this line of research on measures of general economic situation, and employ the volume

index of industrial production sold to scale business survey data on general economic situation.

The analysis of impact of data vintage on quantification procedures constitutes the second dimension of this paper. Various definitions of real-time (RTV) and end-of-sample (EoS) data, along with discussion of advantages of including data revision in quantification models and review of literature, have been presented in my previous publications (see Tomczyk 2013, 2014). In this paper, the topic of data vintage in quantification procedures is readdressed for survey data on general economic situation.

2. Description of data

Reported and expected changes in general economic situation (abbreviated from "General situation of the economy regardless of situation in your sector and enterprise"; see Appendix, question number 8) are taken from the monthly business tendency survey administered by the Research Institute for Economic Development (RIED) at the Warsaw School of Economics. Each survey question requires respondents to evaluate both current situation (as compared to the last month) and expectations for the next 3-4 months by assigning them to one of three categories: increase/improvement, no change, or decrease/decline. Previous studies based on RIED survey data (see Tomczyk, 2008) demonstrate that expectations series defined for three- and four-month forecast horizons do not differ significantly, and the former is used in this paper.

Aggregated survey results are regularly published and commented on in RIED bulletins: each month, a number of respondents is given, along with a percentage of respondents who observed increase/no change/decline and who expect increase/no change/decline in a given area of economic activity, along with a response balance (also called balance statistic) calculated as a difference between the percentage of 'optimists' (those who judge current situation favorably or predict improvement) and 'pessimists' (those who evaluate present situation unfavorably or predict decline).

Let us define the following:

A_t^1 – percentage of respondents who report improvement in general economic situation between t and $t-1$,

A_t^2 – percentage of respondents who report no change in general economic situation between t and $t-1$,

A_t^3 – percentage of respondents who report a decline in general economic situation between t and $t-1$,

P_t^1 – percentage of respondents who expect improvement in general economic situation between t and $t + 3$,

P_t^2 – percentage of respondents who expect no change in general economic situation between t and $t + 3$,

P_t^3 – percentage of respondents who expect a decline in general economic situation between t and $t + 3$.

The response balances calculated for the observed changes:

$$BA_t = A_t^1 - A_t^3, \tag{1}$$

and for the expectations:

$$BP_t = P_t^1 - P_t^3, \tag{2}$$

offer the simplest method of quantification – that is, converting qualitative business survey data into quantitative time series. More sophisticated procedures can be grouped into probabilistic and regressive quantification methods. In this paper, I focus on the regression method which is recommended for quantifying variables over which survey respondents exercise at least limited control (see Nardo, 2003) and which previously had been successfully used to quantify RIED survey data (see Tomczyk, 2008).

Quantification models are not designed to reflect a causal relationship, however, both probabilistic and regression quantification procedures require an *objective* variable to be defined to provide a scaling factor for the *subjective* assessments offered by survey respondents. The extent and frequency of revisions in the volume index of production sold in manufacturing, published by CSO, have been described in Tomczyk (2013, 2014). To summarize, the only regular data revisions in the past two decades were due to changes of the base period in 2004, 2009 and 2013. Apart from these systematic revisions, the index is (occasionally) revised one month after the initial release, and there are no further updates. An illustrative example of the structure of data revisions in the volume index of production for the last six months is shown in Table 1. Each column represents vintage of data and contains data that would have been available at a given moment. The last cell in each column (shaded grey) is the initial release of a value corresponding to a given date. The history of data revisions are represented by rows; data revisions are marked in bold.

Revisions seem regular but small, however, the month-to-month changes in expectations concerning general economic situation and expressed in business tendency surveys also tend to be minute. It is plausible, therefore, that quantification procedures exhibit dependency on even minor updates in

the input data – an effect that has been confirmed in analyses of RIED survey responses concerning changes in production (see Tomczyk, 2014).

Table 1. Revisions of the volume index of industrial production sold (in manufacturing) for November 2014 – April 2015.

	November 2014	December 2014	January 2015	Februar y 2015	March 2015	April 2015
November 2014	119.6	119.6	119.6	119.6	119.6	119.6
December 2014		113.3	113.3	113.3	113.3	113.3
January 2015			112.3	**112.2**	112.2	112.2
February 2015				114.5	**114.7**	114.7
March 2015					131.1	**131.2**
April 2015						120.9

Source: CSO Bulletins.

In Section 3, the results of quantification procedures are reported for quantification models with explanatory variables derived from the RIED business survey, and dependent variables defined as changes in the volume index of production sold, for two data vintages:

- RTV (real time data): initial release available in a given month,
- EoS (end-of-sample): final data which became available one month after the initial announcement.

The sample covers the period of January 2005 to April 2015 (124 observations). Basic statistics for both data vintages are summarized in Table 2.

Table 2. Summary statistics of revisions in the volume index of industrial production sold.

	Initial release (RTV)	Final release (EoS)
mean	97.46	97.47
standard deviation	16.13	16.12
minimum	61.20	61.40
maximum	131.10	131.20

Source: own calculations on the basis of Central Statistical Office data.

Summary characteristics of data vintages exhibit close similarity and suggest that the use of either RTV or EoS data will provide identical empirical results. Situation changes, however, when we consider direction of data

revisions. In Table 3, the structure of revisions in the volume index of industrial production sold is presented.

Table 3. Direction of revisions in the volume index of industrial production.

Direction of revision	Percentage in sample
Initial value larger than final value	26%
Initial value smaller than final value	41%
No revision	33%

Source: own calculations on the basis of Central Statistical Office data.

From Table 3 it follows that upward revisions (that is, corrections from a smaller initial value to a higher final number) are significantly more frequent than downward revisions and also more frequent than no revisions at all. These results suggest that revision process is not unbiased; formal analysis of unbiasedness would be necessary, however, to confirm this initial conclusion.

To conclude description of data, it is worth noting that CSO publishes business tendency indicators (BTIs) in three time series: indicator of the general business tendency climate, BTI diagnosis and BTI forecast, all presented in seasonally adjusted and unadjusted versions and across subsectors. Unfortunately, the full set of data on these indicators is available only from February 2009. As far as data revisions are concerned, a few minor corrections have been introduced between 2005 and 2014 in the indicator of the general business tendency climate; two remaining business conditions series, BTI diagnosis and BTI forecast, have not been revised in the past two decades.

3. Results of quantification procedure

For the purpose of quantifying RIED data on general economic situation, two versions of the regression method are used. In the Anderson model (1952), the following equation is estimated:

$$ {}_t x_{t+1} = \alpha \cdot A_t^1 + \beta \cdot A_t^3 + \nu_t , \tag{3} $$

where ${}_t x_{t+1}$ describes relative changes in value of variable x – in this case, the volume index of industrial production published in CSO Statistical Bulletins – between t and $t - 1$. Assuming that the same relationship holds for expectations reported in surveys, and that the error term in equation (3) meets

standard OLS assumptions, parameters α and β are estimated, and quantitative measure of expectations is constructed on the basis of the following equation:

$$ {}_t\hat{x}_{t+1} = \hat{\alpha} \cdot P_t^1 + \hat{\beta} \cdot P_t^3 , \tag{4} $$

where $\hat{\alpha}$ and $\hat{\beta}$ are OLS-estimators of (3) and reflect an average change in variable ${}_t x_{t+1}$ for respondents expecting, respectively, an increase or a decrease of the dependent variable. The HAC standard errors are usually used to account for possible serial correlation of the error term in (3) due to inertia often observed in expectations series, and heteroskedasticity likely to result from learning patterns imbedded in expectations formation processes.

A modification of the Anderson model was proposed by Thomas (1995) to allow for a special case in which normal or typical situation that respondents compare their current situation to includes a growth rate:

$$ {}_t x_{t+1} = \gamma + \delta \cdot A_t^3 + \xi_t , \tag{5} $$

where δ < 0. The Thomas quantitative measure of expectations is given by the formula:

$$ {}_t\hat{x}_{t+1} = \hat{\gamma} + \hat{\delta} \cdot P_t^3 , \tag{6} $$

where $\hat{\gamma}$ and $\hat{\delta}$ are estimates obtained on the basis of equation (5). The Thomas model, often used for quantifying data on variables like production or prices, offers an additional advantage of limiting the degree of multicollinearity between percentages of ‘optimistic’ and ‘pessimistic’ respondents which typically occurs in the Anderson model.

The quantification models described above are commonly used in converting survey data into time series needed for further analysis. However, vintage of data on the basis of which the models are estimated is rarely addressed. In the case of real time data (RTV), the dependent variable in the regression quantification models (that is, changes in volume of industrial production) is defined on the basis of the volume index of industrial production sold available in real time, IP_t^{RTV}:

$$ P_t^{RTV} = \frac{IP_t^{RTV}}{IP_{t-1}^{RTV}} - 1, \quad t = 1, ..., 124. \tag{7} $$

Variable ($P_t^{RTV} \cdot 100$) is interpreted as a percentage change in volume of industrial production as compared to the last month.

For final end-of-sample (EoS) data, the dependent variable in regression quantification models is defined on the basis of the final announcement of the volume index of industrial production sold, IP_t^{EOS}:

$$P_t^{EOS} = \frac{IP_t^{EOS}}{IP_{t-1}^{EOS}} - 1, \quad t = 1, ..., 124. \tag{8}$$

Nonetheless, equations (7) and (8) do not necessarily reflect economic processes that business tendency survey respondents aim to assess or forecast. Another plausible possibility may be offered: that respondents evaluate current changes in production against recent – let us say, observed during the last quarter – averages. Let us define:

$$P_t^{RTV-AV} = \frac{IP_t^{RTV}}{\frac{1}{3}\sum_{s=1}^{3} IP_{t-s}^{RTV}} - 1, \tag{9}$$

for real-time data and:

$$P_t^{EoS-AV} = \frac{IP_t^{EoS}}{\frac{1}{3}\sum_{s=1}^{3} IP_{t-s}^{EoS}} - 1, \tag{10}$$

for end-of-sample data. Variables ($P_t^{RTV-AV} \cdot 100$) and ($P_t^{EoS-AV} \cdot 100$) reflect percentage changes in volume of industrial production as compared to the average calculated on the basis of last three months, for real-time and end-of-sample data, respectively.

In line with the discussion presented above, the Anderson and Thomas quantification models have been estimated for both RTV and EoS data and for both definitions of the dependent variable: relative to the last month, and relative to the average of the last quarter. Two general conclusions emerge from the initial estimates of equations (3) and (5). First, none of the Anderson quantification models exhibit an appropriate sign of the estimated coefficient for explanatory variable A_t^3. In the models for RTV and EoS data, and for dependent variables defined with respect to the last month's or average values, estimated coefficients of A_t^3 are positive instead of negative. On this

basis, the Anderson model must be rejected as a method of quantifying RIED survey data on general economic situation. Second, the only models which remain in accordance with the quantification assumptions are the Thomas models estimated for a dependent variable defined with respect to the average of the last quarter. For both RTV and EoS data, estimated coefficients of δ are negative, as initially expected. Table 4 presents the results of the Thomas quantification model estimated with dependent variables P_t^{RTV-AV} (real-time data) and P_t^{EoS-AV} (end-of-sample data).

Table 4. The Thomas model (3) with HAC standard errors.

dependent variable	P_t^{RTV-AV}	P_t^{EoS-AV}
$\hat{\gamma}$	2.1785	2.1730
$\hat{\delta}$	-0.0328	-0.0326

Source: own calculations.

Results presented in Table 4 have the following interpretation: in enterprises that within the last month noted deterioration in general economic situation, an average decline was equal to a little more than 3%. This conclusion holds for both data vintages: there are no perceptible differences between results obtained on the basis of RTV and EoS data. Sizes of coefficient estimates are comparable with those obtained in other quantification models published in literature; however, they are not statistically different from zero.

The final question remains: do expectations time series constructed on the basis of the estimates shown in Table 4 present an improvement over the easily available response balance, expressed by equation (2)? It does not seem so. The correlation coefficient of the Thomas expectations series with simple balance statistics is equal to 0.7235 – a high correlation in the world of quantified survey data. What is more, additional assumptions are required for the use of quantification methods and accurate interpretation of their results. For example, from the pairs of equations describing the Anderson and Thomas quantification procedures, (3)-(4) and (5)-(6) respectively, it is clear that expectations for the next three months are calculated on the basis of estimates obtained on the basis of one-month observed changes. This simplification constitutes a significant weakness of the regression method, shared by all commonly used quantification procedures. Also, there exists no empirical confirmation for the assumption that a relationship between *objective* time series and *subjective* assessments can be described by the same

equation as a relationship between survey expectations and the quantified measure of expectations. To summarize, the expectations series obtained on the basis of the Thomas quantification model do not clearly overpower balance statistics as a measure of general economic situation reported by respondents of the RIED business tendency survey.

4. Conclusions and directions for future research

This paper compares results of regression quantification procedures of general economic situation survey data for two data vintages: real-time and end-of-sample, and for two definitions of a dependent variable in quantification models: relative to the last month, and relative to the average of the last quarter. The conclusions may be summarized as follows: quantification of responses to question 8 in the RIED business tendency survey with CSO data on industrial production does not present a significant improvement over the use of response balances as far as construction of expectations series is concerned. For most of the quantification models considered, survey data on general business situation do not fulfill the basic assumptions as to the signs of estimated coefficients. Only the Thomas model constructed with the dependent variable defined with respect to the last quarter's average exhibits a correct sign of the estimated coefficient. Still, the correlation coefficient of the expectations series obtained on this basis with simple response balance is relatively high (0.7235) and therefore suggests that the use of balance statistics may be of similar research value – and, as an additional advantage, unburdened by supplementary assumptions. Also, the results obtained for RTV and EoS data are very similar and do not suggest superiority of any of these two data vintages as far as quantification of survey data on general economic situation is concerned.

There are several directions of future research worth pursuing. First, economic categories other than industrial production or GDP – for example, changes in levels of orders or financial standing of manufacturing companies – may be considered as possible dependent variables in quantifications models. Second, since upward revisions in the production index are observed more often than downward revisions, and also more frequently than no revisions at all, tests of unbiasedness of CSO data revisions offer an interesting line of research. Third, based on results published in Arnold (2013) it seems worthwhile to test whether differences in empirical results with respect to data vintage, not discernible in this paper, depend on the phase of the business cycle in Poland.

References

Anderson, O. (1952). The business test of the IFO-Institute for Economic Research, Munich, and its theoretical model, *Revue de l'Institut International de Statistique*, 20 (1), 1-17.

Arnold, E. A. (2013). *The role of revisions and disagreement in professional forecasts*, National Bank of Poland Working Paper No. 153.

Chen, M.-H., Chen, S.-J., Kuo, Y.-C. (2007). The Critical Role of Monetary Policy in the Link between Business Conditions and Security Returns in Taiwan, *Asia Pacific Management Review*, 12 (1), 1-12.

Nardo, M. (2003). The quantification of qualitative survey data: A critical assessment, *Journal of Economic Surveys*, 17 (5), 645-668.

Thomas, D. G. (1995). Output expectations within manufacturing industry, *Applied Economics*, 27 (5), 403–408.

Tomczyk, E. (2008). Racjonalność oczekiwań polskich przedsiębiorców: reguła czy wyjątek? Wnioski z testu koniunktury IRG SGH, *Prace i Materiały Instytutu Rozwoju Gospodarczego SGH*, 80, 375-388.

Tomczyk, E. (2013). End-of-sample vs. real time data: perspectives for analysis of expectations, *Prace i Materiały Inastytutu Rozwoju Gospodarczego SGH*, 93, 45-69.

Tomczyk, E. (2014). *Influence of data vintage on quantification of expectations*, Department of Applied Econometrics Working Paper 14-04.

Appendix. Monthly RIED questionnaire in industry

		Observed within the last month	Expected for the next 3-4 months
01	Level of production (value or physical units)	up unchanged down	will increase will remain unchanged will decrease
02	Level of orders	up normal down	will increase will remain normal will decrease
03	Level of export orders	up normal down not applicable	will increase will remain normal will decrease not applicable
04	Stocks of finished goods	up unchanged down	will increase will remain unchanged will decrease
05	Prices of goods produced	up unchanged down	will increase will remain unchanged will decrease
06	Level of employment	up unchanged down	will increase will remain unchanged will decrease
07	Financial standing	improved unchanged deteriorated	will improve will remain unchanged will deteriorate
08	General situation of the economy regardless of situation in your sector and enterprise	improved unchanged deteriorated	will improve will remain unchanged will deteriorate

Source: the RIED database.