WORKING PAPER No. 30a-b
May 1993

Reinhold Bergström

Side 1

THE FULL TRICHOTOMOUS SCALE COMPARED WITH NET BALANCES IN QUALITATIVE BUSINESS SURVEY DATA - Experiences from the Swedish Business Tendency Surveys

Svensk resumé på sid 17.

Side 19

QUANTITATIVE PRODUCTION SERIES COMPARED WITH QUALITATIVE BUSINESS SURVEY SERIES FOR FIVE SECTORS OF THE SWEDISH MANUFACTURING INDUSTRY

Svensk resumé på sid 65.

The full trichotomous scale compared with net balances in qualitative business survey data - experiences from the Swedish Business Tendency Surveys

by

Reinhold Bergström

May 1993

Barjanej.doc

Abstract

The full trichotomous scale compared with net balances in qualitative business survey data - experiences from the Swedish Business Tendency Surveys

by

Reinhold Bergström
Department of Statistics
Uppsala University

Many qualitative business tendency survey variables have three scale steps of the type "increase", "unchanged" and "decrease". The information content in the full trichotomous scale is compared with that of balances - the difference between the two extreme alternatives - in models of the production volume measured as ordinary quantitative series. Data from the Swedish Business Tendency Survey (BTS) are used. Total manufacturing and three sectors of the manufacturing industry are studied and several BTS series are considered. In general it is found that use of the full trichotomous scale does not lead to improved models compared with use of balances only. The effect of transformations of BTS variables is also investigated. A logit-transformation does not lead to improved results.

1. Introduction.[1]

Business survey (barometer) data are typically qualitative with very few scale steps. The most common situation is questions with three possible alternatives of the type "increase", "no change" and "decrease". Questions with two alternatives of the type yes/no also occur frequently. As used in practice, the information from the trichotomous variables is often further condensed by the introduction of balances (net tendencies) which are computed as the difference between the "increase" and "decrease" alternatives. In Sweden, The National Institute of Economic Research (NIER) has conducted the Business Tendency Survey (BTS) since the 1950s.

Bergström (1992a) and (1992b) recently analyzed the relationship between production measured as an ordinary quantitative variable and different BTS variables using models of the Autoregressive Distributed Lag (ADL) type. The BTS variables were used in the form of balances only. A strong connection between the annual change in the quantitative production series and different barometer series, particularly production, was found in many cases. Christoffersson et al (1992) and Rahiala & Teräsvirta (1992) also recently analyzed the connection between quantitative and qualitative series using other methods and in the case of Rahiala & Teräsvirta the individual "increase" and "decrease" alternatives in addition to balances.

When BTS variables are used, they are in most cases used in their original form, i.e. as the proportion of firms with a certain alternative. This may not always be the optimal representation and Öller (1989) and (1990) recently studied whether transformations of BTS variables might yield more efficient results.

The purpose of the present paper is to see whether the full trichotomous set of observations contains more information than just the balances, when the purpose is an explanation of the production volume measured as an ordinary quantitative series. The starting-point for the analyses will be the models obtained in Bergström (1992a) and (1992b). In addition to data for total manufacturing (M) data for three sectors of the manufacturing industry will also be considered, namely food, beverages and tobacco (F, sector 31) in the following "Food", wood and wood products (W, sector 33), in the following "Wood" and engineering (sector 38)

[1]I am grateful to the NIER and particularly Roger Knudsen and Pekka Kääntä for putting data at my disposal. I also wish to thank Ann-Sofi Lönnquist and Anders Sundström for help with the collection of further data and preliminary analyses. The research reported in the present paper has been supported by the NIER and the Economic Council of Sweden.

excluding shipyards (E), in the following "Engineering". In addition, the effects of transformations of the basic trichotomous variables will also be analyzed.

2. The Basic Model

Our basic dependent variable is the annual change in the logarithm of the production volume. Denoting this variable, which is almost identical to the relative change in production, by $y = \Delta 4m$ we consider models of the following types

$$y_t = \beta_0 + \beta_1 x_t + u_t \tag{1}$$

and

$$y_t = \beta_0 + \beta_1 y_{t-1} + .. + \beta_k y_{t-k} + \gamma_0 x_t + \gamma_1 x_{t-1} + .. + \gamma_r x_{t-r} + u_t \tag{2}$$

where x_t is a BTS variable in the form of balances.

If we denote the share of firms reporting "increase" and "decrease" by z and w respectively we can formulate a more general model than (1) in the following way

$$y_t = \alpha_0 + \alpha_1 z_t + \alpha_2 w_t + u_t \tag{3}$$

If $\alpha_2 = -\alpha_1$ we obtain (1) as a special case of (3). In such a case the full trichotomous variable yields no extra information relative to use of the balance $x = z-w$ only. It is consequently of great interest to test the hypothesis $H_0 : \alpha_1 + \alpha_2 = 0$. This can be done either by a t-test or by an equivalent F-test of the restriction of model (3) that is implied by model (1). The F-statistic can be computed as

$$F = \frac{R^2_{(3)} - R^2_{(1)}}{(1-R^2_{(3)})/(n-p)} \tag{4}$$

where p is the number of parameters included in model (3). Under the null hypothesis F is distributed as F(1; n-p). Corresponding tests can be performed for the more general model (2).

3. Total Manufacturing (M)

Bergström (1992a) found that the BTS variable B101 which is defined as the production volume in the present quarter as compared with the previous quarter was strongly related to both the annual and quarterly change in production volume. It was also found that inclusion of lagged values of the dependent variable improved the explanatory power of the basic model. A final result obtained was that other BTS variables did not provide additional information to models that included B101.

Descriptive statistics for some of the BTS variables including the important variable B101 are shown in Tables 1 and 2. The variables are defined in Appendix B. For B101 the alternative "Unchanged" is the most common one. "Larger" occurs more often than "Smaller", which is consistent with the positive trend in the quantitative production variable. The variability in the balances is about twice that in the individual extreme alternatives as measured by the standard deviation or the range. There is a high negative correlation between the alternatives "Larger" and "Smaller", while each of these alternatives are only weakly correlated to the "Unchanged" category. The balance is very strongly correlated to both "Larger" and "Smaller".

Table 3 shows results for the model obtained in Bergström (1992a) and a model with B101 as the only explanatory variable. In the latter case we obtain an R^2 of 0.625, which is not much higher than with either "Larger" (0.607) or "Smaller" (0.558). With both these alternatives, we obtain only a slight increase in the R^2 (0.631), implying non-rejection of the hypothesis $\alpha_1 + \alpha_2 = 0$. The alternative "Larger" is more strongly significant and has a numerically larger parameter, 0.0034 compared with -0.0017. In this model addition of each of the alternatives "Larger" and "Smaller" to a model that includes only the other alternative improves the model.

In a model with lagged values of the dependent variable also included, we are again unable to reject the hypothesis of no additional information in the separate variables compared with the balance variable. The alternative "Larger" again has a parameter that is about twice that of the alternative "Smaller" and the latter variable is now not quite significant.

In Bergström (1992a), it was found that although no other variable contained significant extra information in the presence of B101, a number of variables were strongly related to the change in production volume. In Tables 1 and 2 some information on the most important of these variables is given. In these tables we also include results for the ex ante variable B301, which corresponds to the ex post variable B101 but is measured one quarter earlier. Thus B301 is a genuine prediction variable in contrast to B101, which only has predictive value due to its early availability.

The results for the additional variables are mainly similar to those already reported for B101. Thus the correlation between the two extreme alternatives is always negative and of considerable size. Each of the "Increase" and "Decrease" alternatives is strongly correlated with the balance. In general, we find a low correlation between the middle alternative and the balance. There are exceptions to this, however. Thus B203 which is an as-of-now judgement of orderstocks and B213 which is a similar judgement of stocks of finished products show high correlations between the respective balances and the middle alternative "adequate". For the variable B203 the middle alternative is less frequent than for the other variables and has a larger variability.

From Table 4 we find that for five variables out of six, we cannot reject the hypothesis of no extra information in the full set of variables. The exception is the variable B203, where R^2 increases from 0.436 to 0.499. In this case, the parameter of the "Comparatively Large" alternative is even of the wrong sign. It is noticeable that in general we do not obtain extra information in spite of the fact that the absolute values of the estimated parameters may differ considerably. The reason for this is the high correlation between the two extreme alternatives

4. Three Sectors of the Manufacturing Industry

We have analyzed three sectors of the manufacturing industry in detail. Table 5 gives simple descriptive results. To some extent they are similar to what we observed for total manufacturing, but some differences are noteworthy. For the food industry, we notice a lower correlation between the balances and the extreme alternatives and also a weak correlation between these two alternatives. For the wood industry there is a high correlation between the balance and the unchanged category.

In Bergström (1992b), it was noticed that the balance variable B101 is very weakly related to changes in production for the food industry. This is confirmed in Table 6, which also shows that there is no extra information value in the trichotomous scale compared with the balance. A similar conclusion is obtained for the best Autoregressive-Distributed Lag (ADL) model of the food industry, $F(1;59)=1.17$.

For the wood industry there is no extra information in the trichotomous scale in a model based on the BTS variable in current form only (Table 6). The preferred model includes three different lags of B101, which results in an inclusion of six variables when the trichotomous scale is used. This obviously leads to problems with multicollinearity. The extension of the sample period by three years compared with Bergström (1992b) renders $B101_{-2}$ insignificant. The model with the trichotomous scale results in a significant improvement, $F(3;54)=4.16$. The L, S_{-1} and S_{-2} terms are significant with expected signs. There is no obvious reason why the current L and the lagged S alternatives should be significant, while current S and lagged L is insignificant.

For the engineering industry we obtain the same result as for the other two sectors in a model based on B101 in current form only, i. e. no extra information in the full trichotomous scale. The selected ADL model includes B101 in current form and lagged one period. Starting with an ADL model with only the current form of B101 and lags of the dependent variable we find that inclusion of the L and S alternatives does not lead to an improvement, $F(1;74)=0.82$. The estimated parameters 0.00150 and -0.00156 are very close in absolute value to the parameter of the balance which is 0.00152. In the selected ADL model use of the full trichotomous scale does not lead to an improvement either, $F(2;72)=0.88$. Two of the four parameters are significant. All four parameters have the expected sign.

5. Transformed Explanatory Variables

The BTS variables have limited variability, something which might be a drawback, when we compare them with volume variables which in principle can take on any positive value. The upper limit of 100 should not cause much problem as the observations in general are well below this limit. The lower limit 0 is more problematical as the observed values for the "decrease" alternative often are close to 0.

One way of increasing the variability used e.g. by Öller (1989) is to employ the logit-transformation logit z = log (z/(100-z)) where z is a frequency in the trichotomous scale. The frequency 0 occurs in a few cases. To make it possible to use the formula in such cases we have added 0.5 to all observed values.

Table 10 shows the results of the estimation based on logit-transformed variables. These results should be compared with the corresponding results for the variables in original form in Tables 2 and 6. Overall, there is very little difference between the two approaches. The linear alternative works slightly better in terms of R^2 except for the food industry, where all alternatives are very unsatisfactory.

As regards the parameter estimates there is a general tendency. While the estimates of the Larger and Smaller parameters are numerically very similar in the linear alternative, the former parameter in general is larger when the variables are logit-transformed.

6. Conclusions

Many qualitative business tendency survey variables have three scale steps of the type "increase", "unchanged" and "decrease". The information content in the full trichotomous scale is compared with that of balances - the difference between the two extreme alternatives - in models of the production volume measured as ordinary quantitative series. Data from the Swedish Business Tendency Survey (BTS) are used. Total manufacturing and three sectors of the manufacturing industry are studied and several BTS series are considered. In general it is found that use of the full trichotomous scale does not lead to improved models compared with use of balances only. The effect of transformations of BTS variables is also investigated. A logit-transformation does not lead to improved results.

References

Bergström, R., (1992a). The Relationship between Manufacturing Production and Different Business Survey Series in Sweden. Konjunkturinstitutet/ Ekonomiska Rådet. Working Paper No. 12. Feb 1992.

Bergström, R., (1992b). Quantitative production series compared with qualitative business survey series for five sectors of the Swedish manufacturing industry. Research Report 92:x, Department of Statistics, Uppsala.

Christoffersson, A, Roberts, R and Eriksson, U (1992). The Relationship Between Manufacturing and Various BTS series in Sweden Illuminated by Frequency and Complex Demodulate Methods, Konjunkturinstitutet/Ekonomiska Rådet Working Paper No 15, April 1992.

Öller, L.E., (1989). Aggregating problems when forecasting industrial production using business survey data.Ministry of Finance Discussion Paper no 20, Helsinki.

Öller, L.E., (1990). Forecasting the business cycle using survey data. International Journal of Forecasting, 6, 453-461.

Rahiala, M and Teräsvirta, T (1992), Business Survey Data in Forecasting the Output of Swedish and Finnish Metal and Engineering Industries: A Kalman Filter Approach. Journal of Forecasting, 12, 255-271.

APPENDIX A

The Data

For the production volume we use the production index computed by Statistics, Sweden (Statistiska Centralbyrån, SCB). We use monthly index figures adjusted for calendar irregularities (normalmånadskorrigerade indextal) as available in the databank at the NIER at the end of 1990. From the monthly figures given as integer values, quarterly index values are computed by a direct averaging of the monthly figures. These are used with the accuracy that is obtained in the averaging process.

The BTS variables are available both as balances and as original increase/decrease, yes/no or corresponding individual categories. For the balances we have data from the databank at the NIER for the period 1968:1-90:3 for total manufacturing for most of the variables, while for sectors 31 and 33, data are only available for the period 1974:1-90:3. For the engineering industry data are available even before 1968 in many cases. These data do not always exactly coincide with the originally published figures.

For the basic trichotomous series, data are available only from 1978:1 in the data base at the NIER. These figures are not always exactly consistent with the net balances, but the differences are small. To increase the length of the series, we have collected data from the original publications back to 1968 for the individual sectors. For total manufacturing the first available such quarter in the publications is 1973:2.

From these series, balances have been constructed. These have been compared with the balances in the databank at the NIER. The differences are very small, which is illustrated by the following correlations between the two sets of balances: M 0.9999 (73:2-90:3), F 0.9906 (74:1-90:3), W 0.9966 (74:1-90:3) and E 0.9986 (68:1-90:3). In view of these results, we have used the longer published trichotomous series.

Appendix B

BTS variables included

	Ex post variables	Alternatives		
Present quarter compared with previous quarter				
B101	Production volume	Larger (L)	Unchanged (U)	Smaller (S)
B107	Purchases of raw materials	Larger (L)	Unchanged (U)	Smaller (S)
Situation 15 days before the end of the quarter				
B203	As-of-now judgement of orderbooks	Comparatively large (CL)	Adequate (A)	Comparatively small (CS)
B204	Number of workers employed	Larger (L)	Unchanged (U)	Smaller (S)
B213	As-of-now judgement of stocks of finished goods	Too large (TL)	Adequate (A)	Too small (TS)
	Ex ante variable			
Next quarter compared with present quarter				
B301	Production Volume	Larger (L)	Unchanged (U)	Smaller (S)

TABLE 1 Descriptive statistics for some BTS variables in trichotomous form. Total Manufacturing. B denotes balance and the other alternatives are defined in Appendix B. Sample period 1973:2-90:3. (B301 1973:3-90:3).

Variable	Alternative	Mean	St dev	Min	Max	Range
B101	L	19.7	6.6	7	32	25
	U	66.4	3.9	57	74	17
	S	13.9	7.6	2	28	26
	B	5.7	13.6	-20	28	48
B107	L	20.8	9.2	8	52	44
	U	65.5	6.1	43	79	26
	S	13.7	7.9	2	34	32
	B	7.2	16.3	-24	53	77
B203	CL	14.2	9.7	2	38	36
	A	44.8	9.5	23	58	35
	CS	41.0	17.5	11	74	63
	B	-26.7	26.7	-71	27	98
B204	L	11.4	5.5	3	23	20
	U	74.0	4.5	63	84	21
	S	14.6	7.7	4	33	29
	B	-3.2	12.6	-30	18	48
B213	TL	28.5	11.4	10	53	43
	A	64.4	8.7	44	80	36
	TS	7.1	5.0	1	23	22
	B	21.4	15.3	-10	50	60
B301	L	20.1	5.3	10	35	25
	U	67.6	3.5	59	76	17
	S	12.3	6.6	3	29	26
	B	7.8	11.5	-17	32	49

TABLE 2 Correlation between balances, individual components and Δ4m. Total manufacturing 1973:2-90:3. (B301 1973:3-90:3)

	Balances vs				Δ4m vs		
Variable	L	S	U	L vs S	L	S	U
B101	0.958	-0.968	0.268	-0.855	0.779	-0.747	0.141
B107	0.950	-0.919	-0.232	-0.752	0.722	-0.761	-0.096
B203	0.966	-0.989	0.842	-0.917	0.572	-0.690	0.692
B204	0.936	-0.967	0.509	-0.816	0.712	-0.683	0.299
B213	0.974	-0.843	-0.800	-0.704	-0.679	0.490	0.614
B301	0.952	-0.969	0.373	-0.847	0.705	-0.667	0.182

TABLE 3 Models for Δ4m. Total manufacturing. 1973:2 - 90:3.
B=Balance, L=Larger, S=Smaller. Constant terms not shown.

B101 B	0.00249 (0.00023)	0.00167 (0.00025)	B101 L	0.00340 (0.00093)	0.00223 (0.00082)
			S	-0.00171 (0.00081)	-0.00120 (0.00069)
$\Delta 4m_{-1}$		0.287 (0.095)			0.275 (0.097)
$\Delta 4m_{-2}$		0.293 (0.097)			0.300 (0.098)
$\Delta 4m_{-4}$		-0.384 (0.096)			-0.385 (0.097)
$\Delta 4m_{-5}$		0.206 (0.086)			0.197 (0.087)
R^2	0.625	0.761		0.631	0.763
F				1.06	0.49

TABLE 4 Balances vs trichotomous variables for total manufacturing. Dependent variable Δ4m. 1973:2-90:3. Constant terms not shown.

	Balances		Larger	Smaller		
	$\hat{\beta}$	R^2	$\hat{\beta}_1$	$\hat{\beta}_2$	R^2	F
B101	0.00249 (0.00023)	0.625	0.00340 (0.00093)	-0.00171 (0.00081)	0.631	1.07
B107	0.00207 (0.00020)	0.618	0.00161 (0.00053)	-0.00271 (0.00061)	0.631	2.28
B203	0.00106 (0.00015)	0.436	-0.00169 (0.00096)	-0.00255 (0.00053)	0.499	8.40
B204	0.00246 (0.00028)	0.526	0.00361 (0.00113)	-0.00171 (0.00080)	0.538	1.72
B213	-0.00186 (0.00026)	0.439	-0.00248 (0.00047)	0.00020 (0.00109)	0.461	2.81
B301	0.00264 (0.00032)	0.507	0.00395 (0.00129)	-0.00162 (0.00105)	0.514	1.01

TABLE 5 Correlation between balances and individual components of the BTS variable B101 for three sectors of the manufacturing industry.

Sector	Period	Balance vs			L vs
		L	S	U	S
F (31)	74:1-90:3	0.835	-0.799	-0.166	-0.360
W (33)	74:1-90:3	0.910	-0.969	0.684	-0.791
E (38)	68:1-90:3	0.945	-0.924	-0.265	-0.753

TABLE 6 Models for $\Delta 4m$ with the BTS variable B101 in current form as the only explanatory variable. Constant terms not shown.

		Sector		
		F	W	E
Balances only	B101	0.00040 (0.00027)	0.00281 (0.00032)	0.00209 (0.00020)
	R^2	0.0350	0.5541	0.5682
Trichotomous scale	B101: L	0.00038 (0.00043)	0.00313 (0.00137)	0.00189 (0.00053)
	S	-0.00042 (0.00051)	-0.00263 (0.00082)	-0.00235 (0.00064)
	R^2	0.0351	0.5546	0.5691
	Sample period	75:1-90:3	75:1-90:3	69:1-90:3

TABLE 7 ADL models for the food industry. Constant term not shown. Dependent variable: $\Delta 4m$. Sample period 75:1-90:3.

$\Delta 4m_{-4}$	-0.327 (0.116)	$\Delta 4m_{-4}$	-0.332 (0.116)		
$B101_{-2}$	0.00060 (0.00025)	B101: L_{-2}	0.00027 (0.00039)	S_{-2}	-0.00103 (0.00047)
R^2	0.195		0.210		

TABLE 8 ADL models for the wood industry. Constant terms not shown. Dependent variable: $\Delta 4m$. Sample period 75:1-90:3.

$\Delta 4m_{-3}$	0.251 (0.098)	$\Delta 4m_{-3}$		0.284 (0.092)		
$\Delta 4m_{-4}$	-0.255 (0.092)	$\Delta 4m_{-4}$		-0.244 (0.085)		
B101	0.00148 (0.00042)	B101:	L	0.00348 (0.00120)	S	-0.00056 (0.00076)
$B101_{-1}$	0.00100 (0.00041)		L_{-1}	-0.00156 (0.00132)	S_{-1}	-0.00236 (0.00073)
$B101_{-2}$	0.00061 (0.00044)		L_{-2}	-0.00167 (0.00126)	S_{-2}	-0.00173 (0.00071)
R^2	0.6996			0.7560		

TABLE 9 ADL models for the engineering industry. Constant term and parameters of D802 and D812 not shown. Dependent variable $\Delta 4m$. Sample period 70:2-90:3.

Δm_{-1}	0.097 (0.097)	Δm_{-1}	0.090 (0.098)		
Δm_{-2}	0.285 (0.083)	Δm_{-2}	0.282 (0.084)		
Δm_{-4}	-0.235 (0.073)	Δm_{-4}	-0.236 (0.074)		
B101	0.00118 (0.00027)	B101:L	0.00085 (0.00061)	S	-0.00142 (0.00054)
$B101_{-1}$	0.00093 (0.00032)	L_{-1}	0.00121 (0.00058)	S_{-1}	-0.00076 (0.00060)
R^2	0.7472		0.7532		

TABLE 10 Models for $\Delta 4m$ with the "L" and "S" alternatives of B101 in logit-transformed form as the only explanatory variables.

	F	W	E	M
Constant	0.0051 (0.0145)	0.0290 (0.0456)	0.0640 (0.0275)	0.0312 (0.0319)
L	0.0048 (0.0050)	0.0531 (0.0173)	0.0488 (0.0113)	0.0432 (0.0123)
S	-0.0036 (0.0035)	-0.0219 (0.0110)	-0.0112 (0.0069)	-0.0220 (0.0076)
R^2	0.0426	0.5205	0.5558	0.6274
Period	69:1-90:3	69:1-90:3	69:1-90:3	73:2-90:3

Användning av den fullständiga trikotoma skalan i konjunkturbarometerserier jämfört med användning av enbart nettotal - erfarenheter från den svenska konjunkturbarometern.

Barometerdata föreligger ofta i form av variabler som har tre skalsteg med benämningar av typen "större än", "oförändrad" och "mindre än". Vid användning i praktiken komprimeras informationen i denna trikotoma skala mycket ofta genom att s k nettotal beräknas. Nettotalet är differensen mellan andelen svarande som uppger alternativen "större än" och "mindre än". En intressant fråga är om någon väsentlig information går förlorad vid denna förenkling.

Som inledning till den vidare analysen presenteras först deskriptiva resultat som visar hur de tre alternativen i ett antal barometerserier utvecklats var för sig, hur de samvarierar inbördes samt hur de samvarierar med motsvarande nettotal och förändringstakten i den kvantitativa produktionsvolymserien.

Huvudfrågeställningen studeras genom analys av modeller där produktionsvolymen mätt som en vanlig kvantitativ serie är beroende variabel. De modeller som används är av s k ADL typ och innehåller laggade värden av den beroende variabeln och barometerserier som förklarande variabler. Dessa modeller, som utvecklats i ett tidigare arbete, baserades då enbart på barometerdata i form av nettotal. Genom att estimera motsvarande modeller med fullständigt utnyttjande av informationen i den trikotoma variabeln kan vi undersöka om förklaringsvärdet hos den senare modellen är större. Detta kan ske genom direkt testning av vissa parameterrestriktioner.

Ett genomgående resultat är att förklaringsvärdet inte är signifikant större i de modeller som innehåller variabler baserade på den trikotoma skalan. Det är vidare uppenbart att om barometerdata från flera olika tidpunkter används är multikollinearitetsproblemen större vid dessa modeller.

Slutligen undersöks också om man kan förbättra förklaringsvärdet hos barometerserier, genom att transformera dem. Speciellt undersöks logit-transformation av de enskilda beståndsdelarna i de trikotoma serierna. Det visar sig att sådana transformationer inte ökar förklaringsvärdet hos serierna.