Economics Letters 26 (1988) 335–340
North-Holland

AN EXAMPLE OF CORRELATION AMONG RESIDUALS IN DIRECTLY ORDERED DATA *

Marc J.I. GAUDRY

Université de Montréal, Montréal, Canada H3C 3J7

Ulrich C.H. BLUM

Universität Karlsruhe, D-7500 Karlsruhe, FRG

Received 19 January 1988

We distinguish between *naturally* ordered data and data that are ordered *directly* by the researcher. After ordering observations according to a behavioral criterion, we estimate a single-order correlation parameter of the residuals simultaneously with the other parameters and the functional form of a savings equation for higher-income white-collar German households.

1. Introduction

In this note, we draw attention to a particular new orientation of the analysis of correlation among residuals that may prove as fruitful as established orientations. We do this by distinguishing data directly ordered (DO) by the researcher from naturally ordered (NO) data. Consider the very simple case of a standard linear regression model to which we add a simple error process:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + v, \tag{1}$$

$$v = \rho R v + w, \tag{2}$$

where y, X_1 and X_2 are $T \times 1$ vectors of observations, v and w are $T \times 1$ vectors of residuals, the β_k and ρ are parameters and R is the $T \times T$ residue impact matrix where element r_{ti}, for instance for the tth row

$$v_t = \rho \sum_i r_{ti} v_i + w_t, \qquad t = d_1 + 1, \ldots, n - d_2, \tag{3}$$

is defined as

$$\begin{aligned} r_{ti} &= 0, && t = i, \\ &= \text{value established by an impact criterion}, && t \neq i. \end{aligned} \tag{4}$$

* The research for this paper was supported by the F.C.A.R. program of Québec, the S.S.H.R.C. and the N.S.E.R.C. of Canada and the D.F.G. of Germany. Parts of Gaudry's and Blum's contributions were made during their tenures respectively as Alexander von Humboldt Research Scholar in 1985 at the University of Karlsruhe and as Alexander von Humboldt Scholar in 1986/87 at the Université de Montréal. The authors thank Anne Mouls for computing assistance. A full version of this paper is available from the authors.

The r_{ti} elements of the residue impact matrix describe the interaction among residuals. Here we comment on various specifications of residue impact criteria (RIC) used to define (4) and give an empirical example with directly ordered data, using maximum-likelihood estimation to establish the usefulness of the additional single-order correlation parameter.

2. Residue impact criteria

We assume throughout that the sample on y and the X_k contains no missing observations. This requirement is crucial if the eventual ordering of values is to yield correct relative positions or distances among observations.

A. Naturally ordered data

Of course, time series and spatial data are 'naturally' ordered. It is convenient to associate the natural order of time series observations with their running index. Then, each row vector r_t of the residue impact matrix has the familiar appearance

$$\begin{array}{lcccccccc} & 1 & ,\ldots, & t-1, & t, & t+1 & ,\ldots, & T, \\ r_t & =(0 & \ldots & 1 & 0 & 0 & \ldots & 0), \\ r_{t+1} & =(0 & \ldots & 0 & 1 & 0 & \ldots & 0), \end{array} \tag{5}$$

if the residue impact criterion is

[Each residual is correlated with its previous value]. (NORIC-1)

Because NORIC-1 assumes that later observations have no influence on the present, there can only be zeroes on the right-hand-side of diagonal elements of eq. (5). If spatial data are used, one would naturally wish to take into account the likely correlation between neighboring zones. As a zone can have a variable number of contiguous neighbours, a residue impact criterion such as

[Each residue is correlated with its contiguous neighbors]. (NORIC-2)

might well yield rows of the impact matrix such as

$$\begin{array}{lcccccccccccc} & & & & t & & & & & & & & T \\ r_t & =(0 & \ldots & 0 & 1 & 0 & 1 & 0 & 0 & \ldots & 0 & 0 & 1), \\ r_{t+1} & =(0 & \ldots & 0 & 0 & 1 & 0 & 1 & 0 & \ldots & 0 & 1 & 1), \end{array} \tag{6}$$

if the tth zone is a common neighbor of adjacent zones $t-1$, $t+1$, and T but zone $t+1$ has more borders than zone t (namely t, $t+2$, $T-1$, and T). One could also make impact a continuous function of an economic interaction variable, such as travel distance or cost, as in Bolduc (1985) who also summarizes this literature, and greatly improves the realism of the model. Although the data are ordered in space, there is in general, for eq. (6), no simply way to use the running index to order the observations in a way that fits the ordering required by a residue impact criterion such as NORIC-2.

B. Directly ordered data

As soon as the residue impact criterion requires more than searching for statistically significant leads and lags among the running index values, the modeling dimension involved in defining residue impact criteria has to be made explicit: making correlation a function of explicit variables, as just mentioned, is an example. But if the chosen impact criterion is independent from spatial location, one can hardly say that the ordering is *natural*. In effect, the ordering becomes *directly* defined by the impact criterion used to test a specific economic hypothesis. This can be seen most easily with a cross-section of individuals. Assume that we are interested in explaining a particular dimension of behavior, say individual (or group) savings. A residue impact criterion might be

[To test leadership, assume that all residuals are positively correlated with that of individual t^*]. (DORIC-1)

This implies a residue impact matrix where the (non-diagonal) elements of the column t^* are 1 and all others are 0. A negative value of the correlation coefficient could be interpreted as opposition to individual t^*. More generally, if one is interested in phenomena of imitation (looking up to) or distanciation (looking down on) with respect to individuals (or groups) differing in some way or other, ordering criteria might be defined with respect to any variable, including factors not used in the regression model. For instance,

[Rank individuals in decreasing order of net income and assume that each residual is positively correlated with that of the closest individual having 1000 units more of annual net income]. (DORIC-2)

As a rule, directly ordered data resulting from behavioral residue impact criteria will not permit easy use of the running index. The analysis of correlation is transformed into a primarily economic modeling problem and the 'economic' distance between observations becomes in principle more complicated than the simple distance of time series analysis: one might define a correlation that depends on a function of many variables!

3. Example: Higher-income white-collar German household savings

To illustrate the usefulness of directly ordering data, consider the following household savings function:

$$SAV_t = f(SPI_t,\ PER_t,\ AGE_t,\ INT_t,\ TAX_t,\ GIN_t) + u_t, \tag{7}$$

where, neglecting subscripts (units in parentheses):

SAV = increase in net wealth (includes housing but not durables; DM),
SPI = compulsory (old age) Social Security contributions (DM),
PER = number of persons permanently living in the household,
AGE = age of household head (years),
INT = rate of return on household portfolio (% ÷ 100)
TAX = income, capital and wealth taxes paid (DM),
GIN = gross household income including transfers (DM).

The available data, pertaining to German households in 1978, are published by the German Federal Statistical Office (1982) for various types of households, notably white-collar households. As explained in Blum (1986), the two parts of the stratified sample were merged and observations repeated to match their known share in the population. For reasons related to income tax brackets and social security contribution limits, we retained households with a gross monthly income of 4000 DM or more.

As we have no strong a priori views on the functional form of this function, we use the following econometric specification, outlined in Gaudry and Dagenais (1979):

$$y_t^{(\lambda_y)} = \beta_0 + \sum_k \beta_k X_{k_t}^{(\lambda_{xk})} + u_t, \tag{8}$$

$$u_t = \left[\exp\left(\delta_0 + \delta_m Z_{m_t}^{(\lambda_{zm})}\right)\right]^{1/2} v_t, \tag{9}$$

$$v_t = \rho v_{t-26} + w_t, \tag{10}$$

where (λ_y), (λ_{xk}) and (λ_{zm}) denote Box–Cox transformations applied to each variable. In this case, *gross income* is used to obtain homoskedastic errors and the lag $t-26$ is chosen according to the following residue impact criterion stressing the existence of a minimum economic distance, below which households do not notice behavioral differences:

[Assume that every household 'looks up' to households 7 percentiles higher in the distribution of disposable income].

(DORIC-3)

The observations were ordered according to *disposable income* and 7 percentile points (corresponding to 26 consecutive observations) was the minimum number of observations that never involved looking up to households in the same sample stratum. Because of the presence of repeated observations, the Durbin–Watson statistic, sometimes usable [King and Evans (1985)], is not helpful. We will in any case use the likehood ratio test.

The first part of table 1 contains elasticities of E(y), evaluated at the mean value of the explanatory variables, t-statistics conditional upon the optimal values of the Box–Cox transformations found in the second part of the table and, under the parentheses of the t-statistics, a number identifying the associated Box–Cox transformation. The second part of the table contains autocorrelation and heteroskedasticity parameters and t-statistics, as well as values of the maximized-likehood obtained with the L-1.1 algorithm [Liem et al. (1983)]. For our model without correlation, the values of the log likelihood for homoskedastic linear and logarithmic specifications were -364.138 and -285.497 respectively. Columns 3 and 4 respectively show that the use of one or two Box–Cox transformations – one on the endogenous and one on all exogenous variables – did not yield significant gains over the logarithmic case shown in Column 2. However, taking heteroskedasticity (according to eq. 9) into account in Column 5 shows both a large gain in the log likehood and significant differences between the two Box–Cox estimates. Column 6 shows that correcting for autocorrelation under the assumption of homoskedasticity also improves on the result of Column 4. Column 7, the most general model, shows the effet of simultaneously correcting for heteroskedasticity and autocorrelation. Compared to the results of Column 5, we note that the additional autocorrelation parameter is significant even when heteroskedasticity is taken into account. This suggests that households 'keep up with the Joneses', i.e. look up to higher income households in their savings behavior, a reasonable result for this category of households.

Table 2 contains elasticities and partial derivatives of savings with respect to all variables for the most interesting specifications of table 1. It is clear that the marginal effects are at least as reasonable in the most general model with autocorrelation as in other specifications. For instance, the linear model implies an unplausible but statistically significant complementarity between taxes and savings.

Table 1
Savings function of higher-income white-collar German households (1978): Elasticities (t = statistics) and other results.

Elasticities (T-statistics)	Code no. = Dep.var =	1 *WHSAV*	2 *WHSAV*	3 *WHSAV*	4 *WHSAV*	5 *WHSAV*	6 *WHSAV*	7 *WHSAV*
WH = White collar workers (high)								
Payments to soc. pens. insurance	*WHSPI*	0.042	−0.011	−0.013	−0.018	0.023	−0.024	0.025
		(3.04)	(−0.58)	(−0.45)	(−0.65)	(0.94)	(−0.80)	(0.99)
		LAM 1	LAM 1	LAM 1	LAM 1	LAM 1	LAM 1	LAM 1
Household size	*WHPER*	0.103	−0.166	−0.165	−0.151	−0.106	−0.149	−0.100
		(2.35)	(−3.50)	(−2.87)	(−2.50)	(−3.86)	(−2.25)	(−3.62)
		LAM 1	LAM 1	LAM 1	LAM 1	LAM 1	LAM 1	LAM 1
Age of household's head	*WHAGE*	−0.270	−0.399	−0.400	−0.391	−0.364	−0.383	−0.356
		(−6.57)	(−10.13)	(−9.57)	(−9.55)	(−8.65)	(−8.86)	(−8.95)
		LAM 1	LAM 1	LAM 1	LAM 1	LAM 1	LAM 1	LAM 1
Avg. return on household portfolio	*WHINT*	−0.649	−0.054	−0.051	−0.100	0.260	−0.055	−0.329
		(−3.74)	(−0.34)	(−0.45)	(−0.91)	(1.93)	−0.50)	(2.50)
		LAM 1	LAM 1	LAM 1	LAM 1	LAM 1	LAM 1	LAM 1
Income and capital taxes	*WHTAX*	0.655	−0.359	−0.356	−0.309	−0.229	−0.324	−0.221
		(5.45)	(−2.50)	(−2.24)	(−1.80)	(−3.91)	(−1.75)	(−4.07)
		LAM 1	LAM 1	LAM 1	LAM 1	LAM 1	LAM 1	LAM 1
Income (before taxes)	*WHGIN*	1.080	2.640	2.630	2.560	2.370	2.600	2.340
		(5.37)	(11.85)	(9.95)	(9.14)	(22.94)	(8.69)	(23.82)
		LAM 1	LAM 1	LAM 1	LAM 1	LAM 1	LAM 1	LAM 1
Regression constant	*Constant*	−0.732	−2.000	−2.030	−1.680	−7.230	−1.720	−8.690
		(−2.76)	(−6.35)	(−5.85)	(−4.94)	(−17.19)	(−4.48)	(−18.10)
Parameters (T-statistics)	Code no. = Dep.var. =	1 *WHSAV*	2 *WHSAV*	3 *WHSAV*	4 *WHSAV*	5 *WHSAV*	6 *WHSAV*	7 *WHSAV*
Autorcorrelation								
							0.132	0.158
Order 26	RHO 26						(1.79)	(1.98)
Box–Cox transformations								
xa							(1.79)	(1.98)
Lambda(Y)	*LAM Y*	1.000	0.000	−0.024	−0.004	−0.401	−0.002	−0.437
		FIXED	FIXED	(−0.23)	(−0.03)	(−2.40)	(−0.02)	(−2.70)
				(−9.63)	(−8.93)	(−8.40)	(−9.11)	(−8.88)
Lambda(X)–Group 1	*LAM 1*	1.000	0.000	−0.024	0.077	−1.028	0.070	−1.156
		FIXED	FIXED	(−0.23)	(0.25)	(−2.70)	(0.21)	(−3.09)
				(−9.63)	(−2.97)	(−5.32)	(−2.84)	(−5.76)
Heteroskedasticity								
Delta coefficients								
Income (before taxes)	*WHGIN*					−0.035		−0.074
						(−7.67)		(−8.28)
Box–Cox transformations								
Income (before taxes)	*WHGIN*					2.519		2.165
						(2.26)		(2.10)
						(1.36)		(1.13)
Log-likelihood		−364.138	−285.497	−285.438	−285.353	−266.103	−282.537	−262.111
Pseudo-R^2		0.962	0.976	0.976	0.976	0.979	0.977	0.979
Number of observations		341	341	341	341	341	341	341
Sample – First observation		27	27	27	27	27	27	27
– Last observation		367	367	367	367	367	367	367
Number of independent variables		7	7	7	7	7	7	7

Table 2
Savings function of higher-income white-collar German households (1978): Elasticities and [partial] derivatives.

		1	2	4	7
		Linear	Log	Box–Cox	Box–Cox + het. + correlation
Payments to the social pens. Insur.	*WHSPI*	0.042 [0.046]	−0.011 [−0.011]	−0.018 [−0.018]	0.025 [0.025]
Household size	*WHPER*	0.103 [20.500]	−0.166 [−30.200]	−0.151 [−27.400]	−0.100 [−18.500]
Age of Household's head	*WHAGE*	−0.270 [−4.070]	−0.399 [−5.520]	−0.391 [−5.400]	−0.356 [−5.010]
Average return on household portfolio	*WHINT*	−0.649 [−9741.300]	−0.054 [−748.000]	−0.100 [−1370.000]	0.329 [4608.500]
Income and capital taxes	*WHTAX*	0.655 [0.448]	−0.359 [−0.226]	−0.309 [−0.194]	−0.221 [−0.141]
Income (before taxes)	*WHGIN*	1.080 [0.128]	2.640 [0.286]	2.560 [0.278]	2.340 [0.258]

In general, we note that payments to the Social Pension Insurance System do not affect savings – clearly because the average income of our group lies above the income contribution limit and consequently contributions vary very little. From the influence of household size we conclude that one additional person reduces savings by about DM 20 – a reasonable result because we know that this additional person does not earn. A reduction of savings with the aging of the household's head is compatible with the life cycle model. We also note the convincing result that income taxes reduce savings. The value of the marginal propensity to save is credible.

4. Conclusion

We have pointed out that the structure of correlation among residuals could be exploited to test behavioral assumptions and noted that the formulation of required 'residue impact criteria' generally implied a direct ordering of observations that only exceptionally made use of whatever degree of natural ordering the data might have possessed. We provided an example of directly ordered data where information in the residuals made behavioral sense and improved the regression results of a Box–Cox specification controlled for heteroskedasticity of a very general form.

References

Blum, U.C., 1986, *Raumwirkungen des Gesetzlichen Rentenversicherungsbudget*, Karlsruher Beiträge zur Wirtschaftspolitik und Wirtschaftsforschung no. 13 (Loeper Verlag, Karlsruhe).

Bolduc, D., 1985, Deux procédures d'estimation en présence d'autocorrélation spatiale dans les résidus, Publication no. 443 (Centre de recherche sur les transports, Université de Montréal, Montréal).

Gaudry, M.I.J. and M.G. Dagenais, 1979, Heteroskedasticity and the use of Box–Cox transformations, Economics Letters 2/3, 225–229.

King, M.L. and M.A. Evans, 1985, The Durbin–Watson test and cross-sectional data, Economics Letters 18, 31–34.

Liem, T.C., M.G. Dagenais and M.J.I. Gaudry, 1983, L-1.1, A program for Box–Cox transformations with heteroskedastic and autoregressive residuals, Publication no. 301 (Centre de recherche sur les transports, Université de Montréal, Montréal).

Statistisches Bundesamt, 1982, Einkommens- und Verbrauchsstichprobe 1978, no. 4 (Stuttgart-Mainz).