BINGHAM McCUTCHEN LLP
DONN P. PICKETT (SBN 72257)
GEOFFREY M. HOWARD (SBN 157468)
HOLLY A. HOUSE (SBN 136045)
ZACHARY J. ALINDER (SBN 209009)
BREE HANN (SBN 215695)
Three Embarcadero Center
San Francisco, CA 94111-4067
Telephone: (415) 393-2000
Facsimile: (415) 393-2286
donn.pickett@bingham.com
geoff.howard@bingham.com
holly.house@bingham.com
zachary.alinder@bingham.com
bree.hann@bingham.com

BOIES, SCHILLER & FLEXNER LLP
DAVID BOIES (Admitted *Pro Hac Vice*)
333 Main Street
Armonk, NY 10504
Telephone: (914) 749-8200
Facsimile: (914) 749-8300
dboies@bsfllp.com
STEVEN C. HOLTZMAN (SBN 144177)
FRED NORTON (SBN 224725)
1999 Harrison St., Suite 900
Oakland, CA 94612
Telephone: (510) 874-1000
Facsimile: (510) 874-1460
sholtzman@bsfllp.com
fnorton@bsfllp.com

DORIAN DALEY (SBN 129049)
JENNIFER GLOSS (SBN 154227)
500 Oracle Parkway, M/S 5op7
Redwood City, CA 94070
Telephone: (650) 506-4846
Facsimile: (650) 506-7114
dorian.daley@oracle.com
jennifer.gloss@oracle.com

Attorneys for Plaintiffs Oracle USA, Inc., *et al.*

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
OAKLAND DIVISION

ORACLE USA, INC., *et al.*, Plaintiffs, v. SAP AG, *et al.*, Defendants.	CASE NO. 07-CV-01658 PJH (EDL) **EXHIBIT 2 TO THE DECLARATION OF DANIEL S. LEVY, PH.D. IN SUPPORT OF MOTION NO. 1: TO EXCLUDE TESTIMONY OF DEFENDANTS' EXPERT STEPHEN CLARKE** **FILED PURSUANT TO DKT. NO. 915**

EXHIBIT 2

NT

OMIC GROWTH

BRIUM THEORY

ITATIVE

GROWTH
AND FINANCE
GOTIATIONS

ECONOMETRICS

G. S. Maddala
University of Florida

McGRAW-HILL BOOK COMPANY
New York St. Louis San Francisco Auckland Bogotá
Düsseldorf Johannesburg London Madrid
Mexico Montreal New Delhi Panama
Paris São Paulo Singapore Sydney Tokyo Toronto

ECONOMETRICS

89101112 HDHD 89876543

This book was set in Times Roman by Computype, Inc. The editors were J. S. Dietrich and Michael Gardner; the production supervisor was Charles Hess. The drawings were done by Vantage Art, Inc.

Library of Congress Cataloging in Publication Data

Maddala, G S date
Econometrics.

(Economics handbook series)
Includes index.
1. Econometrics. I. Title.
HB139.M35 330'.01'82 76-26042
ISBN 0-07-039412-1

term. Often we find a constant term also included in regression equations with first differences. This procedure is valid only if there is a linear-trend term in the original equation. If the regression equation is

$$y_t = \alpha + \delta t + \beta x_t + u_t$$

then
$$y_{t-1} = \alpha + \delta(t-1) + \beta x_{t-1} + u_{t-1}$$

and on subtraction we get

$$(y_t - y_{t-1}) = \delta + \beta(x_t - x_{t-1}) + (u_t - u_{t-1})$$

which is an equation with the constant term δ.

Another important thing to note is that usually with time-series data one gets good R^2's when the regressions are estimated with the levels y_t and x_t but one gets poor R^2's if the regressions are estimated in first differences $(y_t - y_{t-1})$ and $(x_t - x_{t-1})$. Since usually a high R^2 is considered as proof of a strong relationship between the variables under investigation, there is a strong tendency to estimate the regression in terms of the levels rather than the first differences. This is sometimes called the "R^2 syndrome." However, if the Durbin-Watson statistic is very low, it often implies a misspecified equation, no matter what the value of the R^2 is. In such cases one should estimate the regression equation in first differences; and if the R^2 now is low, it is merely an indication that the variables y and x are not indeed related to each other, as the high R^2's obtained from the regressions of the levels might imply. Granger and Newbold[1] present some examples with artificially generated data where y, x, and the residual u are each highly autocorrelated series each generated independently so that there is no relationship between y and x, but the regression of y on x gives a high R^2 and a low Durbin-Watson statistic. When the regression is run in first differences, the R^2 is close to zero and the Durbin-Watson statistic is close to 2.0, thus demonstrating that the R^2 obtained earlier is spurious and that there is indeed no relationship between y and x. Thus regressions in first differences might often reveal the true nature of the relationship between y and x.

It is, of course, not always true that one should be estimating regression equations in first differences. In fact, if the Durbin-Watson statistic is greater than 1.2, which roughly implies that the correlation between u_t and u_{t-1} is less than $\frac{1}{2}(2.0 - 1.2)$, or 0.4, using first differences might actually increase the correlation between the resulting residuals $(u_t - u_{t-1})$ and $(u_{t-1} - u_{t-2})$. In such cases one should be using *quasi first differences* rather than first differences. For instance, if the Durbin-Watson statistic is 0.8, since this implies the correlation between u_t and u_{t-1} to be roughly $\frac{1}{2}(2.0 - 0.8)$, or 0.6, we should regress $(y_t - 0.6y_{t-1})$ on $(x_t - 0.6x_{t-1})$.

Finally, it should be emphasized that all this discussion of the Durbin-Watson statistic, first differences, and quasi first differences is relevant only if we believe that the residuals show first-order autocorrelation, that is, u_t and u_{t-1}

[1] C. W. J. Granger and P. Newbold, Spurious Regressions in Econometrics, *Journal of Econometrics*, vol. 2, no. 2, pp. 111–120, July 1974.

other qualities constant. For example, the coefficient of H indicates that an increase in 10 units of horsepower, ceteris paribus, results in a 1.2 percent increase in price. However, some of the coefficients have to be interpreted with caution. For example, the coefficient of P in the equation for 1960 says that the presence of power steering as "standard equipment" led to a 22.5 percent higher price in 1960. In this case the variable P is obviously not measuring the effect of power steering alone but is measuring the effect of "luxuriousness" of the car. It is also picking up the effects of A and B. This explains why the coefficient of A is so low in 1960. In fact, A, P, and B together can perhaps be replaced by a single dummy that measures "luxuriousness." These variables appear to be highly intercorrelated. Another coefficient, at first sight puzzling, is the coefficient of V, which, though not significant, is consistently negative. Though a V-8 costs more than a six-cylinder engine on a "comparable" car, what this coefficient says is that, holding horsepower and other variables constant, a V-8 is cheaper by about 4 percent. Since the V-8's have higher horsepower, what this coefficient is saying is that higher horsepower can be achieved more cheaply if one shifts to V-8 than by using the six-cylinder engine. It measures the decline in price per horsepower as one shifts to V-8's even though the total expenditure on horsepower goes up. This example illustrates the use of dummy variable and the interpretation of seemingly wrong coefficients.

As another example consider the estimates of liquid-asset demand by manufacturing corporations. Vogel and Maddala[1] computed regressions of the form $\log C = \alpha + \beta \log S$ where C = cash and S = sales, on the basis of data from the Internal Revenue Service, "Statistics of Income," for the year 1960–1961. The data consisted of 16 industry subgroups and 14 size classes, size being measured by total assets. When the regression equations were estimated separately for each industry, the estimates of β ranged from .929 to 1.077. The R^2's were uniformly high, ranging from .985 to .998. Thus one might conclude sales elasticity of demand for cash is close to 1. Also, when the data were pooled and a single equation estimated for the entire set of 224 observations, the estimate of β was .992 and $R^2 = .987$. When industry dummies were added, the estimate of β was .995 and $R^2 = .992$. From the high R^2's and relatively constant estimate of β one might be reassured that the sales elasticity is very close to 1. However, when asset-size dummies were introduced, the estimate of β fell to .334 with R^2 of .996. Also, all asset-size dummies were highly significant. The situation is described in Fig. 9-2. That the sales elasticity is significantly less than 1 is also confirmed by other evidence. This example illustrates how one can be very easily misled by high R^2's and apparent constancy of the coefficients. It also illustrates how one can get misleading results from grouped data, as mentioned in Chap. 6. When grouping is only by one variable, as in this case, more meaningful results will be obtained by considering a rectangular array of data consisting of several cross sections and analyzing it by pooled regressions

[1] R. C. Vogel and G. S. Maddala, Cross-Section Estimates of Liquid Asset Demand by Manufacturing Corporations, *The Journal of Finance*, December 1967.

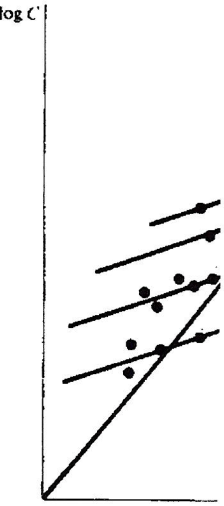

Figure 9-2 Bias due to om

and dummy variables
be given later.

As mentioned ea
As an illustration, cc
and chicken on the b
of demand functions

P_1

P_2

P_3

where P_1 = retail pric
P_2 = retail pric
P_3 = retail pric
x_1 = consump
x_2 = consump
x_3 = consump
y = disposabl
x_1, x_2, x_3 can be obt
retail divided by a
consumer price index
y are as follows:[1]

[1] There appears to be
basis of other information

H indicates that an
ılts in a 1.2 percent
o be interpreted with
or 1960 says that the
a 22.5 percent higher
easuring the effect of
usness" of the car. It
the coefficient of A
ıps be replaced by a
iables appear to be
ht puzzling, is the
tly negative. Though
rable" car, what this
ıbles constant, a V-8
ıorsepower, what this
əved more cheaply if
easures the decline in
total expenditure on
ımy variable and the

id-asset demand by
ed regressions of the
on the basis of data
for the year 1960–
ize classes, size being
ons were estimated
m .929 to 1.077. The
one might conclude
the data were pooled
24 observations, the
nies were added, the
ıd relatively constant
ty is very close to 1.
estimate of β fell to
ghly significant. The
significantly less than
ates how one can be
f the coefficients. It
n grouped data, as
able, as in this case,
rectangular array of
y pooled regressions

iquid Asset Demand by

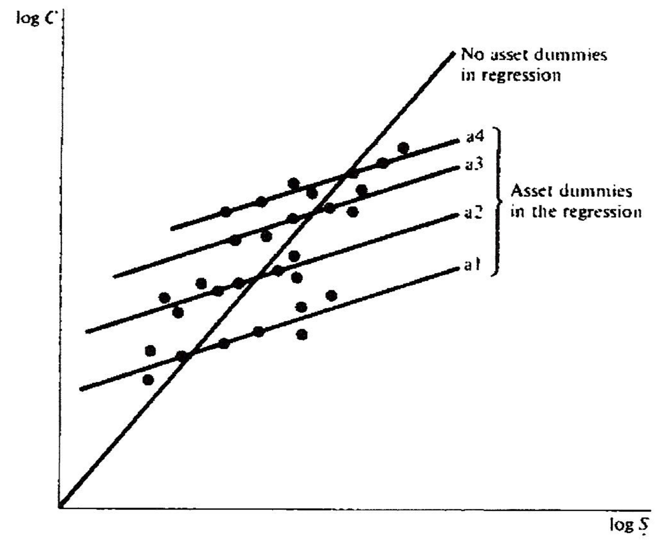

Figure 9-2 Bias due to omission of dummy variables.

and dummy variables. Some further examples of analysis from grouped data will be given later.

As mentioned earlier, dummy variables are not necessarily (0,1) variables. As an illustration, consider the joint estimation of the demand for beef, pork, and chicken on the basis of data presented in Table 7-5. Waugh estimates a set of demand functions of the form

$$\begin{aligned} P_1 &= \alpha_1 + \beta_{11}x_1 + \beta_{12}x_2 + \beta_{13}x_3 + \gamma_1 y + u_1 \\ P_2 &= \alpha_2 + \beta_{12}x_1 + \beta_{22}x_2 + \beta_{23}x_3 + \gamma_2 y + u_2 \\ P_3 &= \alpha_3 + \beta_{13}x_1 + \beta_{23}x_2 + \beta_{33}x_3 + \gamma_3 y + u_3 \end{aligned} \qquad (9\text{-}6)$$

where P_1 = retail price of beef
P_2 = retail price of pork
P_3 = retail price of chicken
x_1 = consumption of beef per capita
x_2 = consumption of pork per capita
x_3 = consumption of chicken per capita
y = disposable income per capita

x_1, x_2, x_3 can be obtained from Table 7-5. The prices in Table 7-5 are, however, retail divided by a consumer price index. Hence we multiplied them by the consumer price index p to get p_1, p_2, and p_3. This index p and disposable income y are as follows:[1]

[1] There appears to be a misprint in the price of beef given in Table 7-5 for the year 1950 (on the basis of other information given in Waugh). We corrected this to 83.3 from 88.3.