The Review *of* Economics *and* Statistics

VOL. LXXIII | MAY 1991 | NUMBER 2

A TEST OF THE HABIT FORMATION HYPOTHESIS USING HOUSEHOLD DATA

Dale Heien and Cathy Durham*

Abstract—Numerous studies have confirmed the importance of habit formation, as represented by a lagged dependent variable, in demand analysis. Although all work to date has been based on aggregate time series data, this study uses household level BLS Interview Panel data to test the habit hypothesis. An interrelated demand system for seventeen goods is estimated from cross-section data and compared with a similar system based on time series data. The results show that the habit component is significantly different between the two data sets and much smaller in the cross-section data. Habit effects, while not as large in cross-section data as time series, are still highly significant. Several explanations are offered concerning why the two sets of estimates differ.

NUMEROUS studies have confirmed the importance of habit in consumer demand behavior. Even though the effects of habit can be modeled in several different ways, many of them result in specifications which contain a lagged dependent variable. Models of this type permit the researcher to draw a distinction between short- and long-run impacts. Although these models present estimation problems and are sometimes criticized as ad hoc, they are widely used in empirical research and policy analysis. Since the estimates of these models often show long-run effects which are substantially greater than the immediate impacts, their validity is of considerable interest.

Virtually all previous empirical work incorporating a lagged dependent variable has been based on aggregate time series data. For reasons discussed below, we conjecture that due to the use of time series data these habit effects are overstated. This paper attempts to shed light on this issue by testing a particular version of the linear habit formation hypothesis utilizing household microdata. This particular hypothesis says that the habitual component of consumption is proportionate to past consumption. To the authors' knowledge, this is the first comprehensive study which uses household data to estimate a demand model incorporating habit effects.[1]

Received for publication September 1, 1989. Revision accepted for publication July 18, 1990.

* University of California, Davis.

The authors are indebted to Oscar Burt, Richard Sexton and two anonymous referees whose comments and suggestions improved the manuscript considerably. Giannini Foundation Paper No. 953.

Background and Focus

Considerable effort has been devoted to the specification and estimation of habit effects. Although originally developed for investment and supply analysis, the models of Koyck (1954) and Nerlove (1958) were quickly adapted to consumer demand problems. These models have in common a lagged dependent variable, which implies either a geometric distributed lag, a partial adjustment mechanism, or an adaptive expectations scheme. Other conceptualizations of consumer demand models which utilize a lagged dependent variable are due to Stone and Rowe (1957) and Houthakker and Taylor (1970). Habit effects were introduced into complete demand systems by Pollak and Wales (1969).[2] Their term for this effect was the habit formation hypothesis. Some researchers have relied on that hypothesis, while others have introduced habit, or dynamic, effects into complete systems in other ways. Examples of these other dynamic systems models include Anderson and Blundell (1983), Houthakker and Taylor (1970), Phlips (1972), Rossi (1987), Spin-

[1] Benus, Kmenta and Shapiro (1976) estimate a partial adjustment model for total food expenditure using cross section data with prices. Darrough, Pollak and Wales (1983) and Tsujimura and Sato (1964) use grouped means from household data to estimate models which contain lagged dependent variables.

[2] Stone (1954) was the first to suggest using the lagged dependent variable as a measure of habit in a demand system.

newyn (1981), and Wissenberger (1986). Habit formation has also played an important role in theoretical analyses of consumer behavior.[3]

The empirical results from both single equation models and complete demand systems indicate that habit plays a very important role in consumer behavior.[4] Virtually all of the testing of the habit formation hypothesis has been done with market level time series data. The use of time series data was unavoidable since no longitudinal household data with lagged dependent variable information existed on detailed commodities.

For habit formation models, the use of time series data presents various estimation problems. First, the lagged dependent variable is often highly correlated with the dependent variable, resulting in estimates of the habit effect which are quite large. While this high correlation may reflect actual consumer behavior, statistical scenarios can be envisioned wherein it arises in a spurious manner. For example, assume that demand curves do not contain any habitual component, that relative prices change very little and that income displays a smooth pattern over time. In this case consumption will change mainly as a function of income. As a result, current consumption will be highly correlated with lagged consumption even though habit is not present. Also, constraints on industry capacity can generate a high correlation between total output, which is approximately equal to aggregate consumption for many industries, and lagged total output. For food items, where output fluctuates considerably, weaker habit effects are observed. Also, the manner in which some time series data are gathered, i.e., benchmarking off of other, more aggregate (and hence smoother) time series, may also explain the high correlation.

Second, there is considerable multicollinearity in the data used to estimate these time series models. The problems inherent in this situation are well known. Also, many of the studies were done using highly aggregated commodity groups, e.g. food, housing, services. As the groups become larger, they will tend to be more highly correlated with lagged consumption, since total expenditure is highly correlated with lagged total expenditure. Common, of course, to both time series and cross-section data is the estimation bias imparted due to presence of a lagged dependent variable.[5] Omission of important variables which are correlated with lagged consumption will, of course, tend to overstate the effects of lagged consumption and, as a result, overstate habit effects. Also, simultaneous equation bias is more likely to occur with market aggregates than with cross-section data.

Model Specification

The purpose of this analysis is to test the linear habit formation hypothesis on individual household data. To date, tests of the habit formation hypothesis have been conducted within the framework of a complete system of demand equations.[6] The data base used for this study does not contain information on prices paid by households. Hence, it is not possible to estimate a complete demand system of sufficient generality

[3] See, for example, Gorman (1967) and Pollak (1970).

[4] In the Pollak and Wales (1969) study, the coefficients for the lagged dependent variables were 0.749 (food), 0.797 (clothing), 0.928 (shelter), and 0.939 (miscellaneous). However, using an error specification which permitted serial correlation, Howe, Pollak and Wales (1979) found that the Linear Expenditure System had large and significant habit effects while the Quadratic Expenditure System had habit effects which were nearly zero and not significant. In that study, Howe, Pollak and Wales cautioned against strong conclusions regarding the effect of habit due to the possibility of a bimodal likelihood function.

It should also be noted that the presence of a lagged dependent variable is consistent with explanations other than habit. Examples of other explanations include technological, legal, institutional restraints, and/or expectations. These conceptualizations give rise to models such as the partial adjustment model, the geometric lag, or adaptive expectations. To the extent that the reduced forms of these models are the same as one would obtain under the linear habit formation hypothesis, the results given in this paper can be extrapolated to these other models also, although the reasons for the presence of lags are different.

[5] However, OLS estimates will be consistent and asymptotically normal, see Johnston (1984) pp. 360–362. Since, the cross section data base used here is quite large, the bias problem may be ameliorated. When autocorrelation is present the problem is more complicated. However, the reduction in efficiency is of the order of $(1-\alpha\phi)^{-1}$ where α is the autocorrelation coefficient and ϕ is the coefficient on the lagged dependent variable, see Harvey (1981) pp. 268–270. For most of the cross section estimates here both α and ϕ are small.

[6] A complete system of demand relations is one which satisfies the properties of homogeneity, symmetry and adding up.

to realistically represent consumer behavior.[7] However, in order to provide comparability with past studies an interrelated demand system based on the Quadratic Expenditure System (QES) was chosen as the vehicle to test this hypothesis.[8] As mentioned above, the data base used in this study does not contain information on prices. First, consider the QES in expenditure form with all prices normalized at unity,[9] or,

$$e_{ih} = \gamma_{ih} + \alpha_i\left(m_h - \sum_{j=1}^{n}\gamma_{jh}\right) + \beta_i\left(m_h - \sum_{j=1}^{n}\gamma_{jh}\right)^2 \tag{1}$$

where e_{ih} is the expenditure on the i^{th} good by the h^{th} household and m_h is total expenditure on all goods by the household. Since there is no point in preserving the parameter structure of the QES, β_i represents a combination of several QES parameters.

Demographic variables can be incorporated into models of this type through a technique known as translation.[10] Applying this concept to the above model gives

$$\gamma_{ih} = \rho_{io} + \sum_{k=1}^{s}\rho_{ik}\, d_{kh} \qquad i = 1,\ldots,n, \tag{2}$$

where ρ_{io} and the ρ_{ik}'s are parameters to be estimated and the d_{kh} are the demographic variables, of which there are s, for the h^{th} household. In order to specify the linear habit formation hypothesis we write the model given by (1), with time subscripts, as

$$e_{iht} = \gamma_{iht} + \alpha_i\left(m_{ht} - \sum_{j=1}^{n}\gamma_{jht}\right) + \beta_i\left(m_{ht} - \sum_{j=1}^{n}\gamma_{jht}\right)^2. \tag{3}$$

The habit formation hypothesis is now added to (2) so that,

$$\gamma_{iht} = \rho_{io} + \sum_{k=1}^{s}\rho_{ik}\, d_{kht} + \phi_i e_{iht-1} \qquad i = 1,\ldots,n. \tag{4}$$

The relations given by (3) present a nonlinear estimation problem due to the presence of supernumerary expenditure—$(m_{ht} - \Sigma_{j=1}^{n}\gamma_{jht})$. In order to avoid nonlinear simultaneous equation estimation[11] the system given by (3) was simplified to

$$e_{iht} = \gamma_{iht} + \alpha_i m_{ht} + \beta_i m_t^2 \tag{5}$$

which is simply a demographically augmented quadratic Engel curve.

This rather indirect derivation of the Engel curve given by (5) was presented for two reasons. First, the derivation shows the relation between (5) and a complete system of demand relations which employs the linear habit formation hypothesis. Second, we will present an iterated estimation of (3) for comparison with (5). The system given by (5) is not a complete demand system, but it is an approximation to a complete system and should suffice reasonably well for the purpose at hand.

Estimation

The use of household data is desirable since it avoids the problem of aggregation over consumers, provides a large sample, and limits the collinearity problems inherent in time series.[12]

[7] It is possible, for example, to estimate the demand system based on a Cobb-Douglas utility function without price data. Pollak and Wales (1978) show how the Linear Expenditure System (LES) can be estimated with data from two different price situations even though prices are not observed in either situation. Unfortunately, the LES has linear Engel curves, which for household data is not realistic.

[8] For further development of the QES see Howe, Pollak and Wales (1979). An interrelated demand system is defined as one which is estimated by a technique which takes cross equation error covariances into account.

[9] In cross-section studies where price data are not available, it is often customary to assume that all consumers face the same prices. This assumption is equivalent to normalizing all prices at unity.

[10] For a rigorous treatment of demographic effects in demand systems, see Pollak and Wales (1981).

[11] Some authors, see Howe, Pollak and Wales (1979), have indicated the possibility of bimodal likelihood functions. Informal discussions by one of the present authors with users of multiple equation nonlinear maximum likelihood algorithms indicates similar concerns. Also, size and cost constraints on available software and hardware prevented the estimation of a large (320 parameters and 2067 observations) nonlinear system.

[12] As an example of the collinearity in time series data, for data used in this study the average correlation coefficient between q_t and q_{t-1} is 0.981, for q_t and m_t it is 0.972, and for m_t and m_{t-1} it is 0.994. The respective statistics for the cross section model are 0.434, 0.393, and 0.659. The data are discussed later.

However, the use of household data for detailed commodities presents a major estimation problem. In surveys as large as the one used in this study, not every household will consume something in each of the categories unless the categories are broadly aggregated.

The fact that the observed expenditures on particular items sometimes take on zero values means that the dependent variable is truncated. The problem of truncated dependent variables, first recognized by Tobin (1958), results in biased and inconsistent estimates. For single equation models, works by Olsen (1980) and Greene (1981) indicate that the bias is proportional to the probability of a limit observation. However, the problem here is more complex, involving a set of demand relations. It is possible to estimate models of this type by maximum likelihood, but such procedures generally are not computationally feasible or reliable. Lee (1978) extended the single equation two-step Amemiya (1974) censored regression estimator to the multiple equation case.[13] Lee shows that the two-stage estimator resulting from this procedure is asymptotically more efficient than other two-stage estimators.

This particular application involves the special case of a simultaneous model in which the dependent variables are censored by a subset of unobservable latent variables. The dependent variables, which are the expenditures for the sixteen nondurables and services under consideration, are either zero or some positive amount for each household. Those cases where expenditure is zero are considered to be censored by an unobservable latent variable which induces the decision not to purchase that particular item in the survey period. The decision to buy or not to buy can be indicated by a binary indicator variable, which is a function of the latent variable and estimated as a probit model.[14] The procedure involves two steps. First, a probit regression is computed which determines the probability that a given household will consume the good in question. This regression is then used to compute the Mills' ratio for each household. The Mills' ratio is in turn used as an instrument in the second stage estimation of the demand relations.

The decision to consume is modeled as a dichotomous choice problem

$$C_{iht} = f(m_{ht}, d_{1ht}, \ldots, d_{sht}, C_{iht-1}) \tag{6}$$

where C_{iht} is 1 if the h^{th} household consumes the i^{th} item (i.e., if $e_{iht} > 0$), in the t^{th} period and 0 if the household does not consume the item in question. The other variables are as defined above. Little, if any, theoretical work has been done regarding the specification of (6). However, prices and demographic effects should play roles similar to those expected in traditional demand analysis. In addition, expenditure is included in the specification since Jackson (1984) has shown that variety is an increasing function of income, here proxied by expenditure.

Next, for the i^{th} good for the h^{th} household which consumes the item, the Mills' ratio

$$R_{iht} = \phi(p_{iht}, d_{ht}, m_{ht}, C_{iht-1}) / \Phi(p_{iht}, d_{ht}, m_{ht}, C_{iht-1}) \tag{7}$$

was computed where ϕ and Φ are the density and cumulative probability functions, respectively. For those households who do not consume the item in question the Mills' ratio is given by

$$R_{iht} = \phi(p_{iht}, d_{ht}, m_{ht}, C_{iht-1}) / (1 - \Phi(p_{iht}, d_{ht}, m_{ht}, C_{iht-1})). \tag{8}$$

The Mills' ratio for each item is then used as an instrument in the estimation of the demand system by specifying

$$\gamma_{iht} = \rho_{io} + \sum_{k=1}^{s} \rho_{ik} d_{kht} + \phi_i e_{iht-1} + \delta_i R_{iht} \qquad i = 1, \ldots, n. \tag{9}$$

The data base used to estimate the model was taken from the Bureau of Labor Statistics, Consumer Expenditure Survey: Interview Survey. This survey contains information on expenditures on over five hundred items as well as various demographic variables and income information. In the Interview Panel, approximately 5000 Households are visited by an interviewer every three months for fifteen months on an ongoing basis, recording expenditure data four times and basic household

[13] This procedure treats the zero consumption problem as a regression which is censored by latent variables. Other researchers have modelled the problem from different perspectives. Lee and Pitt (1986) consider the problem from the point of view of the dual, while Deaton and Irish (1984) present a model which considers measurement errors.

[14] See Lee (1978), p. 358. The assumptions underlying this model (and its proofs) are that the error terms from the model are approximately normal with zero means and finite variance–covariance matrix which is constant over all observations, i.e., i.i.d. For a more detailed description of the procedure used here, see Heien and Wessells (1990).

and demographic information on the first visit. Although recorded on a monthly basis, expenditures were added up to obtain quarterly totals. This was done for two reasons. First, it was assumed that three months provided a better planning horizon for consumer behavior. Second, for many items, the monthly expenditures were equal, i.e., they were a quarterly total divided by three. Next, renters were dropped from the data base.[15] Hence the demand system estimated here pertains to homeowners only.

As mentioned above, expenditures are recorded for over five hundred separate categories. These categories were aggregated in sixteen expenditure groups for purposes of this study. These groups along with the sample average expenditure and their budget shares are given in appendix A, columns (1) and (2). For purposes of this study only nondurable goods and services were considered.[16] The demographic variables used in this study were the number of males in the household over fifteen years of age, the number of females over fifteen, the number of males two to fifteen, the number of females two to fifteen, and the number of children under two. Dummy variables were created for the regions of the country (North Central, South, and West) and for the quarters. The default variables were East for region and Fall for season. A dummy variable was also created for location (city), with suburban/rural as the default.

The data contained in this sample cover the 1980-III to 1981-IV period. Although the data are available in panel form, constraints on the computer software and the data itself prevented treating the model as a moving panel of cross-section data over the entire time period. However, it was possible for any given quarter to match a household's expenditures for each item with the previous quarter's expenditure for as long as the household was in the survey. Hence, for each household in the data set, there exists information on current and lagged expenditures for all items in the budget. As a result of this it was also possible to test for first order autocorrelation by utilizing those households with a matching observation from the previous quarter. However, it was not possible to test for fourth order autocorrelation, which is perhaps equally likely in quarterly data, because a household's expenditure data are recorded for four quarters only. The same consideration prevented the use of an error components model.

The data base described above was still too large for existing software to accommodate. Hence, approximately 60% of the observations were dropped randomly using a pseudo random number generator. This deletion resulted in a final data base used for estimation of 2067 observations. Several OLS regressions were run on the data before and after the random deletion to check on the effects of reducing the sample size. There were virtually no differences in the estimated coefficients.

As noted above, the main drawback to the Interview Survey is that price data are not recorded. It is possible to attempt to remedy this situation by using area price indexes for each household in a particular area for each category of expenditure. However, such an approach invariably captures effects other than the intended ones. Also, since the price for any given good will be the same for all households in a given region, the price variable will be linearly dependent with the regional dummy variables. For these reasons no attempt was made to add prices to the system. This specification assumes that the relative price structure remained constant over the 1980-II to 1981-IV time period.

The probit model was estimated first for each of the sixteen categories. The specification for the expenditure choice equations is

$$C_{iht} = \beta_{io} + \sum_{k=1}^{s} \beta_{ik} d_{kht} + \beta_{is+1} C_{iht-1} + \beta_{is+2} m_{iht} + \beta_{is+3} m_{iht}^2 \tag{10}$$

where C_{iht} is as defined in (6). Equation (10) was then estimated by the probit technique and the Mills' ratios were computed for each item for each household. The equations given by (5), with (4) used as the specification for γ_{iht}, were then estimated using iterative Three Stage Least Squares, which is equivalent to maximum likeli-

[15] Barnes and Gillingham (1984), using cross-section data found that there were statistically significant differences between renters and homeowners in estimated demand relations.

[16] Although lagged dependent variables have long played a role in durable demand analysis the underlying models are conceptually different than the habit formation hypothesis. The authors plan to investigate the demand for durables in a future study.

hood. The estimates are presented in appendix A.

One of the major problems plaguing time series estimates has, of course, been autocorrelation. Hence, testing of relations given in appendix A for autocorrelation is advisable. Since the estimated model given by (5) and (4) contains a lagged dependent variable, it is necessary to use the Durbin h-test. Because of constraints on the data set mentioned above, it is possible to test a first order autocorrelation scheme between the second and third quarters and between the third and fourth quarters only. The results of these tests are given in columns (5) and (6) of table 1. Using an h-value of 2.0 (93% significance), the hypothesis of first order autocorrelation cannot be rejected in 23 of the 32 cases. However, the values of the first order autocorrelation coefficients are quite low. The highest coefficient is 0.31 (Miscellaneous) and the mean (of the abso-

TABLE 1.—SELECTED STATISTICS FOR 16 GOOD QUADRATIC ENGEL CURVE SYSTEM

	Cross Section Data Estimates								Time Series Data Estimates					
	Non-zero			AC Stats		Super	Exp Elasticity		LDV Coefficients		AC Stats	Exp Elasticity		
Category	w	LDV	R^2	r/h	r/h	LDV	Impact	LR	Original	Corrected	r/t	Impact	LR	t-ratio
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Food at Home	98.7	0.222	0.598	0.11	0.19	0.369	0.391	0.503	0.973	0.949	0.34	−0.022	−0.701	9.52
	0.169			3.22	5.67						1.77			
Alcohol at Home	89.0	0.199	0.251	0.05	0.13	0.295	0.501	0.622	0.941	0.964	−0.11	0.338	5.636	19.45
	0.038			1.81	3.99						0.48			
Food Away	84.7	0.121	0.401	0.09	0.03	0.246	1.287	1.462	0.917	0.969	−0.04	0.029	0.331	21.97
	0.192			2.62	0.75						0.21			
Alcohol Away	37.1	0.361	0.402	0.01	−0.12	0.384	0.848	1.325	0.931	0.234	0.95	0.237	3.384	0.77
	0.006			0.11	3.64						14.86			
Tobacco	11.1	0.259	0.289	0.21	0.23	0.274	0.579	0.782	0.905	1.041	−0.45	−0.114	−1.135	8.09
	0.002			5.66	6.53						2.45			
Clothing	92.3	0.192	0.488	0.01	0.02	0.353	1.263	1.561	0.912	0.963	0.06	0.289	3.215	10.62
	0.061			0.05	0.44						0.31			
Entertainment	96.3	0.011	0.839	0.25	0.11	0.029	1.381	1.391	0.901	0.948	−0.13	−0.169	−1.693	11.03
	0.087			6.75	2.74						0.64			
Personal Care	81.3	0.528	0.508	0.02	−0.26	0.584	0.329	0.686	0.487	0.328	0.58	0.551	1.057	1.38
	0.011			0.66	7.71						3.51			
Medical Care	89.2	0.185	0.209	0.05	0.17	0.314	0.788	0.961	1.042	1.028	−0.06	0.006	−0.138	17.16
	0.052			1.54	5.07						0.31			
Local Transport	96.8	0.061	0.443	0.14	0.16	0.295	0.669	0.806	0.265	0.393	−0.28	1.532	2.071	1.65
	0.145			4.02	4.53						1.41			
Travel Transport	10.6	0.275	0.446	0.09	0.07	0.132	2.041	2.172	0.641	0.746	−0.13	0.355	0.985	3.69
	0.011			2.47	1.96						0.63			
Miscellaneous	83.6	0.542	0.291	0.31	0.15	0.075	2.091	2.152	1.052	1.061	−0.57	−0.025	0.483	20.26
	0.049			8.67	4.22						3.44			
Home Heating	80.4	0.328	0.381	0.04	−0.08	0.353	0.272	0.372	1.051	1.089	−0.17	−0.057	1.105	14.89
	0.029			1.26	2.42						0.84			
Electricity	99.1	0.418	0.467	−0.18	−0.14	0.639	0.242	0.526	1.035	1.052	0.05	−0.001	0.039	26.08
	0.034			5.54	4.26						0.23			
Household Services	98.5	0.031	0.415	0.18	−0.12	0.448	0.412	0.611	0.942	0.896	−0.24	−0.093	−1.555	12.37
	0.031			5.26	3.46						1.21			
Homeownership	71.8	0.166	0.627	0.12	0.17	0.643	0.566	0.976	0.916	0.921	0.21	0.334	3.929	16.77
	0.085			3.29	4.62						1.04			

(1) Expenditure category.
(2) Percentage of Households who consume the item/Budget share.
(3) Coefficient for lagged dependent variable.
(4) R^2 from regression.
(5) Autocorrelation coefficient for second to third quarter/ Durbin h-statistic.
(6) Autocorrelation coefficient for third to fourth quarter/ Durbin h-statistic.
(7) Coefficient for LDV for specification with supernumeray expenditure, equation (8).
(8) Impact (first period) expenditure elasticity
(9) Long-run expenditure elasticity.
(10) Coefficient for LDV in regressions which are uncorrected for autocorrelation.
(11) Coefficient for LDV in regressions which are corrected for autocorrelation.
(12) Autocorrelation coefficient for corrected regressions/t-ratio for coefficient.
(13) Impact (one period) expenditure elasticity.
(14) Long-run expenditure elasticity.
(15) t-ratio for test of equality of (3) and (11).

lute value, since some are negative) over all goods is 0.12. Hence, we conclude, due to the low values of the autocorrelation coefficients, that autocorrelation is not a problem for this data set and specification. The model was re-run using an autocorrelation correction procedure. The estimates of the coefficients and their standard errors changed very little under this estimation procedure.

In order to provide a comparison of the habit formation effect for the household data, a demand system similar to (5) based on time series data was also estimated. The specification of the time series interrelated demand model is given by

$$q_{it} = \gamma_{it} + \alpha_i m_t / p_{it} + \beta_i p_{it} \Pi_t m_t^2 \qquad i = 1, \ldots, n, \quad (11)$$

where

$$\gamma_{it} = \rho_{io} + \phi_i q_{it-1} \quad (12)$$

and

$$\Pi_t = \prod_{j=1}^{n} p_{jt}^{-w_j}, \quad (13)$$

the q_{it}'s are the quantities of the i^{th} good, the p_{it}'s are their respective prices, the w_j's are the sample average budget shares, and the other variables are as defined above, only they now pertain to market aggregates. Hence, the household subscript is no longer appropriate. The time series data cover the same sixteen categories of goods and services as the household data, and were obtained from the Personal Consumption Expenditures, U.S. National Income and Products Accounts. The data used are annual, they are on a per capita basis, and cover the period 1960–86.[17] Unlike the household data, this series includes data on the prices of the sixteen goods. In order to account for the impact of prices over time we have included the price terms in (11). The demand system given by (11), (12) and (13) was estimated in the same manner as the system which is based on the household data. Since the time series data pertain to market aggregates, the Mills' ratio approach was not required. The estimates of the time series model and the probit relations used in the household model are not given in order to conserve space. Both are available on request from the authors.

The impact on the habit formation parameters of specifying an Engel curve in terms of supernumerary expenditure was also examined by estimating (3). This was accomplished by taking the estimates of γ_{iht}, from the original estimation of (5) and computing supernumerary expenditure. This then permitted the estimation of (3), given supernumerary expenditure. This estimation results in a new set of estimates of (4) which can then be used to recompute supernumerary expenditure and reestimate (3). This iterative procedure was continued until the parameter estimates converged. This estimation procedure was used on both sets of data.

Hence, to summarize, we estimated four models: two based on household data and two based on time series data. For the household data the two models are given by (3) and (5), both using (9) for γ_{iht}. For the time series data the two models are given by (11) and

$$q_{it} = \gamma_{it} + \alpha_i \left(m_t - \sum_{j=1}^{n} p_{jt}\gamma_{jt} \right) / p_{it} + \beta_i p_{it} \Pi_t \left(m_t - \sum_{j=1}^{n} p_{jt}\gamma_{jt} \right)^2 \quad (14)$$

with γ_{it} given by (12) for both models. A household data version of (5) was also estimated without the demographic effects and without the Mills' ratios. This was done to further test for differences between the lagged dependent variable coefficients in the time series model and in household models when demographic effects are excluded.

Results

The information needed for comparing the two systems is given in table 1. Column (3) gives the slope coefficients for the habit formation hypothesis in the household model with demographic effects included. Only two of the sixteen coefficients are greater than 0.5 and the simple average is 0.244. These magnitudes are considerably below what has been estimated in time series studies, c.f. Pollak and Wales (1969) cited in footnote 2. The estimates of the slope coefficients for the

[17] It would have been preferable to use quarterly data for comparability purposes. However, the quarterly time series data pertain to only ten nondurable categories and do not correspond well to the household data in terms of the categories covered. Some additional evidence concerning this problem is discussed later.

habit formation hypothesis in the time series model corrected for autocorrelation are given in column (11).[18] The simple average of these estimates is 0.849, well over three times that of the household average. It should be noted that the t-ratios on the slope coefficients for the lagged dependent variables in the household model are quite high (see appendix A), indicating that habit is a very significant variable in determining demand. The t-ratios were also high in the time series model, although generally less than their counterparts in the cross-section model.

In order to formalize these results, a test was performed on the hypothesis that the estimated values of the slope coefficients for the time series habit formation model (the ϕ_i's) are equal to the slope coefficients found in the household model. This was accomplished using a likelihood ratio test. The value of the likelihood function for the time series model mentioned above was compared with the value of the likelihood function for the same time series model except that the values of the ϕ_i's were restricted to be those found in the model estimated with household data. The test statistic is

$$R = -2\ln(L_1 - L_2) \tag{15}$$

where L_1 and L_2 are the values of the likelihood functions for the restricted and the unrestricted models, respectively. The test statistic is distributed as χ^2 with 16 degrees of freedom. The value of the test statistic was 220.42 which greatly exceeds any plausible theoretical value of a χ^2 variate. Therefore, we reject the hypothesis that the slope coefficients from the household model are equal to their counterparts from the time series model.

The same test was also performed using the habit formation slope coefficients from the household model without demographics. The lagged dependent variable coefficients from this model are slightly larger than those from the model which included demographic effects. The test statistic again exceeded 200 and hence the hypothesis is also rejected for this set of coefficients. Although there is a bias toward rejecting the null hypothesis, the exceedingly large values found here for the χ^2 statistic should remove any concern about this bias. The same test was also made using the supernumerary time series model given by (14). In both the time series and household supernumerary models, the habit coefficients were somewhat higher than their counterparts. However, the test results were again similar to those given above showing a strong rejection of the equality of the habit coefficients.

To further confirm the above findings, a coefficient-by-coefficient test between the household model with demographic effects and the time series model was performed using the t-test for the equality of means from two different populations. In fourteen out of the sixteen cases the time series lagged dependent variable coefficient is larger.[19] The two categories where it is smaller are Alcohol Away from Home and Personal Care. The hypothesis is that the difference between the time series coefficient and its cross-section counterpart is zero. The results, presented in column (15) of table 1, strongly reject the hypothesis in all cases but two—Alcohol Away from Home and Personal Care.

The presence of habit formation effects makes it possible for researchers to draw a distinction between impact, or immediate, and long-run effects. In table 1 these effects are presented in the form of expenditure elasticities for both systems. For the household system, the estimates of the impact and long-run expenditure elasticity are given in columns (8) and (9), while the corresponding estimates for the time series model are found in columns (13) and (14). For both models, the long-run elasticities are larger in all cases.[20] The differences between impact and long-run elasticities for the household based system are much less than those computed for the time

[18] The time series model was tested for autocorrelation and the results are given in table 1, column (12). The first entry (r) is the autocorrelation coefficient and the second entry (t) is the t-ratio from the significance test on r. With the exception of Alcohol Away from Home (and possibly Personal Care and Local Transport), the coefficients on the lagged dependent variable changed very little after the correction. The results discussed above pertain to the corrected version.

[19] The comparison is between columns (3) and (13) in table 1.

[20] For several categories (Tobacco, Medical Care, Miscellaneous, Home Heating, and Electricity) the coefficient on the lagged dependent variable exceeds unity. For a Koyck type interpretation, this implies explosive behavior. As mentioned above, lagged dependent variable models may overstate habit effects with time series data due to the omission of relevant variables such as demographics. Pooling techniques such as those developed by Jorgenson and others (see references in Jorgenson and Slesnick (1987)) can be used to ameliorate this problem.

series based system. This difference is a direct reflection of the differences between the slope coefficients in the two systems.[21]

Conclusions

In the opening section it was conjectured that the large habit effects found in previous studies may be overstated. The main reason for this conjecture is that virtually all of the previous studies were conducted using aggregate time series data. Reasons given for the overstatement of habit effects in time series data are collinearity problems arising from mutual trends and simultaneity, autocorrelation, aggregation and omitted variables. In order to examine this conjecture, we tested the habit formation hypothesis within a system of interrelated demand equations for sixteen commodities. The data base used for this study is a cross-section of households over time. The novel feature of this data base is that for each household for each commodity, data exist on the lagged expenditures. As a result, it was possible to test the habit formation hypothesis with cross-section data. The linear habit formation hypothesis is one of the most commonly used dynamic models, although by no means the only one. A habit formation model was also estimated using time series data on the same categories of goods and services to provide a basis of comparison with the household model.

On the basis of the evidence offered above, we conclude that for the data sets used in this study, habit effects from time series data differ substantially from those based on cross-section data.[22] The habit effects (as measured by the magnitude of the coefficient of the lagged dependent variable) based on time series data are also consistently much larger. We conclude that the differences between these two sets of estimates are statistically significant. Lastly, it should be emphasized that while habit effects are not as large in cross-section data as in time series data, they are still highly significant and play an important role in consumer behavior.

[21] A related point is the impact of the time interval under consideration. Sexauer (1977) analyzed this problem in the Houthakker-Taylor model and found that as the interval becomes shorter (e.g., from annual to quarterly) the effect of habit (which is positive) is outweighed by the stock effect (which is negative). However, the stock effect is present only for durables, which are not considered here, and nondurables such as clothing and food at home. Most of the budget (over 75%) in this study is for services and nondurables which cannot be stored, such as food away from home and alcohol away from home. For those items with no stock effect the average coefficient on the lagged variable was 0.311 for the household data and 0.653 for the time series data.

[22] One reviewer raised the interesting possibility that the time series estimates may capture interdependent preferences of the type described by Pollak (1976). If that is the case then the cross-section models are misspecified because they have omitted them. Hence, our results are consistent with the situation where both habit formation and interdependent preferences are present, i.e., the interdependent preferences account for the differences between the time series and the cross-section estimates. We note, however, that previous studies have ascribed their findings to habit effects, not interdependent preferences. A reviewer also noted that the cross-section estimates may be biased downward due to our treatment of seasonality. The lagged dependent variable does not have the seasonal component removed, while the current dependent variable has seasonality as an explanatory variable.

REFERENCES

Amemiya, T., "Multivariate Regression and Simultaneous Equation Models When the Dependent Variables Are Truncated Normal," *Econometrica* 42 (Nov. 1974), 999–1012.

Anderson, Gordon, and Richard Blundell, "Testing Restrictions in a Flexible Demand System: An Application to Consumers' Expenditure in Canada," *The Review of Economic Studies* 50 (July 1983), 397–410.

Barnes, Roberta, and Robert Gillingham, "Demographic Effects in Demand Analysis: Estimation of the Quadratic Expenditure System Using Microdata," this REVIEW 66 (Nov. 1984), 591–601.

Benus, J., J. Kmenta and H. Shapiro, "The Dynamics of Budget Allocation to Food Expenditures," this REVIEW 58 (May 1976), 129–138.

Darrough, Masako N., Robert A. Pollak and Terence J. Wales, "Dynamic and Stochastic Structure: An Analysis of Three Time Series of Household Budget Studies," this REVIEW 65 (May 1983), 274–281.

Deaton, Angus, and Margaret Irish, "Statistical Models for Zero Expenditures in Households Budgets," *Journal of Public Economics* 23 (Feb./Mar. 1984).

Gorman, W. H., "Tastes, Habits and Choices," *International Economic Review* 8 (June 1967), 218–222.

Greene, William H., "On the Asymptotic Bias of the Ordinary Least-Squares Estimator of the Tobit Model," *Econometrica* 49 (Mar. 1981), 505–514.

Harvey, A. C., *The Econometric Analysis of Time Series* (Deddington, Oxford: Philip Allan, 1981).

Heien, Dale, and Cathy Wessells, "Demand Systems Estimation with Microdata: A Censored Regression Approach," *Journal of Business and Economic Statistics* 8 (July 1990), 365–371.

Houthakker, Hendrik S., and Lester D. Taylor, *Consumer Demand in the United States: Analyses and Projections*, second and enlarged edition (Cambridge: Harvard University Press, 1970).

Howe, Howard, Robert A. Pollak, and Terence J. Wales, "Theory and Time Series Estimation of the Quadratic Expenditure System," *Econometrica* 47 (Sept. 1979), 1231–1248.

Jackson, Laurence Fraser, "Hierarchic Demand and the Engel Curve for Variety," this REVIEW 66 (Feb. 1984), 8–15.

Johnston, J., *Econometric Methods* (New York: McGraw-Hill, 1984).

Jorgenson, Dale W., and Daniel T. Slesnick, "Aggregate Consumer Behavior and Household Equivalence Scales," *Journal of Business and Economic Statistics* 5 (Apr. 1987), 219–232.

Koyck, L. M., *Distributed Lags and Investment Analysis* (Amsterdam: North-Holland, 1954).

Lee, Lung-Fei, "Simultaneous Equations Models with Discrete and Censored Dependent Variables," in P. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications* (Cambridge: MIT Press, 1978), 346–364.

Lee, Lung-Fei, and Mark M. Pitt, "Microeconometric Demand Systems with Binding Nonnegativity Constraints: The Dual Approach," *Econometrica* 54 (Sept. 1986), 1237–1242.

Nerlove, Marc, *Distributed Lags and Demand Analysis*, Agricultural Handbook No. 141 (Washington, D.C.: USDA, 1958).

Olsen, V., "Approximating a Truncated Normal Regression with the Method of Moments," *Econometrica* 48 (July 1980), 1099–1105.

Phlips, Louis, "A Dynamic Version of the Linear Expenditure Model," this REVIEW 54 (Nov. 1972), 450–458.

Pollak, Robert A., "Habit Formation and Dynamic Demand Functions," *Journal of Political Economy* 78 (July/Aug. 1970), 745–763.

______, "Interdependent Preferences," *The American Economic Review* 66 (June 1976), 309–320.

Pollak, Robert A., and Terence J. Wales, "Estimation of the Linear Expenditure System," *Econometrica* 37 (Oct. 1969), 611–628.

______, "Estimation of Complete Demand Systems from Household Budget Data: The Linear and Quadratic Expenditure Systems," *American Economic Review* 68 (June 1978), 348–359.

______, "Demographic Variables in Demand Analysis, *Econometrica* 49 (Nov. 1981), 1533–1558.

Rossi, Nicola, "An Intertemporally Quasi Separable Demand System," this REVIEW 69 (Aug. 1987), 449–457.

Sexauer, Benjamin, "The Role of Habits and Stocks in Consumer Expenditures," *The Quarterly Journal of Economics* 91 (Feb. 1977), 127–142.

Spinnewyn, Frans, "Rational Habit Formation," *European Economic Review* 15 (Jan. 1981), 91–109.

Stone, Richard, "Linear Expenditure Systems and Demand Analysis: An Application to the British Pattern of Demand," *Economic Journal* 64 (Sept. 1954), 511–527.

Stone, Richard, and D. A. Rowe, "The Market Demand for Durable Goods," *Econometrica* 25 (July 1957), 423–443.

Tobin, James, "Estimation of Relationships for Limited Dependent Variables," *Econometrica* 26 (Jan. 1958), 24–36.

Tsujimura, Kotaro, and Tamotsu Sato, "Irreversibility of Consumer Behavior in Terms of Numerical Preference Fields," this REVIEW 46 (Aug. 1964), 305–319.

Wissenberger, Edgar, "An Intertemporal System of Dynamic Consumer Demand Functions," *European Economic Review* 20 (Aug. 1986), 859–891.

APPENDIX A—ECONOMETRIC ESTIMATES OF PARAMETERS OF 16 GOOD QUADRATIC ENGEL CURVE SYSTEM FOR HOMEOWNERS

Category	Exp/w	Int	LDV	$M > 15$	$F > 15$	$M < 15$	$F < 15$	$C < 2$	City	Single	Span
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Food at Home	707.55	21.70	0.22	87.88	87.98	74.69	65.71	70.98	2.16	−36.82	31.12
	0.17	(0.76)	(13.96)	(8.08)	(7.39)	(7.32)	(6.18)	(2.95)	(0.87)	(1.07)	(0.78)
Alcohol at Home	158.14	−20.79	0.2	20.3	1.97	1.92	17.01	−15.13	32.1	6.59	−39.52
	0.04	(1.36)	(10.63)	(3.55)	(0.32)	(0.36)	(3.01)	(1.17)	(2.41)	(0.36)	(1.84)
Food Away	807.03	−61.74	0.12	−12.99	−15.41	−25.54	−28.77	−70.34	16.43	13.06	−5.45
	0.19	(2.97)	(7.04)	(1.68)	(1.82)	(3.50)	(3.76)	(4.03)	(0.91)	(0.52)	(0.19)
Alcohol Away	23.17	−5.19	0.36	−0.83	−4.35	−1.92	−3.43	−7.42	2.74	−7.91	0.71
	0.01	(1.23)	(21.36)	(0.52)	(2.51)	(1.29)	(2.20)	(2.09)	(0.75)	(1.56)	(0.12)
Tobacco	9.31	−0.42	0.26	−0.34	1.78	−0.81	−0.47	0.17	−3.40	0.49	4.01
	0.0	(0.12)	(19.38)	(0.26)	(1.24)	(0.65)	(0.36)	(0.59)	(1.11)	(0.12)	(0.81)
Clothing	253.07	−86.98	0.19	−38.90	−0.44	−12.97	−0.30	−35.64	−3.22	12.59	9.39
	0.06	(3.51)	(13.62)	(4.22)	(0.44)	(1.50)	(0.33)	(1.72)	(0.15)	(0.42)	(0.27)
Entertainment	366.39	269.85	0.11	−2.04	−26.56	41.92	37.02	6.73	28.52	−69.33	−105.12
	0.09	(7.92)	(0.98)	(0.16)	(1.91)	(3.50)	(2.95)	(0.24)	(0.96)	(1.69)	(2.20)
Personal Care	46.29	−3.28	0.53	2.09	3.93	−2.10	−0.95	−5.68	5.55	0.47	−7.48
	0.01	(0.92)	(30.16)	(1.56)	(2.67)	(1.67)	(0.72)	(1.88)	(1.78)	(0.11)	(1.49)
Medical Care	216.39	111.72	0.18	−14.94	−12.27	−26.09	−31.0	−8.05	−84.39	−11.39	−4.34
	0.05	(3.89)	(10.46)	(1.40)	(1.05)	(2.59)	(2.94)	(0.34)	(3.39)	(0.33)	(0.11)
Local Trans	608.41	−33.75	0.17	43.53	54.19	−0.86	−32.32	−31.53	−83.89	−1.99	−19.73
	0.15	(1.01)	(10.82)	(3.45)	(3.94)	(0.73)	(2.62)	(1.12)	(2.88)	(0.49)	(0.42)
Travel Trans	41.52	−4.44	0.60	−23.49	−9.81	−11.68	−17.45	−12.37	−6.14	−11.36	7.52
	0.01	(0.33)	(3.93)	(4.60)	(1.76)	(2.44)	(3.47)	(1.08)	(0.52)	(0.69)	(0.39)

APPENDIX A—(*Continued*)

Miscellaneous	205.43	−233.01	0.31	20.58	−34.39	−30.40	−20.84	−24.64	−51.90	38.92	−48.34
	0.05	(5.16)	(1.95)	(1.22)	(1.87)	(1.92)	(1.26)	(0.65)	(1.32)	(0.72)	(0.77)
Home Heating	125.42	3.27	0.28	−5.68	−2.38	0.28	−0.58	−12.76	113.49	38.26	−9.01
	0.03	(0.27)	(14.03)	(1.24)	(0.47)	(0.65)	(0.13)	(1.24)	(10.25)	(2.59)	(0.52)
Electricity	142.34	43.10	0.54	4.57	6.83	1.21	−0.36	7.70	−21.58	7.08	0.30
	0.03	(6.13)	(28.66)	(1.75)	(2.39)	(0.50)	(0.14)	(1.32)	(3.56)	(0.85)	(0.31)
Household Services	129.62	7.0	0.33	−6.66	10.03	0.46	−1.15	7.66	25.49	−12.82	21.79
	0.03	(0.92)	(18.43)	(2.35)	(3.21)	(0.17)	(0.41)	(1.20)	(3.83)	(1.40)	(2.04)
Homeownership	354.91	44.44	0.42	−53.27	−55.03	7.87	18.76	101.56	−18.58	2.22	59.52
	0.09	(1.83)	(36.83)	(5.87)	(5.52)	(0.92)	(2.09)	(4.96)	(0.88)	(0.76)	(1.75)
Total	4195.0										
	1.0										

Note: The numbers in parentheses are the t-ratios.
(1) Expenditure category
(2) Sample mean expenditure/Budget share
(3) Intercept
(4) Lagged Dependent Variable Coefficient
(5) Number of Males Fifteen or Older Living in the Household
(6) Number of Females Fifteen or Older Living in the Household
(7) Number of Males Fifteen or Younger Living in the Household
(8) Number of Females Fifteen or Younger Living in the Household
(9) Number of Children Less than Two
(10) Dummy Variable for City Dweller
(11) Single
(12) Spanish Speaking Ethnic Background

Category	Black	NC	S	W	Win	Spr	Sum	MR	α	β	R^2
(1)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)
Food at Home	−19.93	−43.94	−35.23	−33.18	−6.72	−12.08	18.74	140.24	0.90	−0.20E-05	0.59
	(0.70)	(2.22)	(1.70)	(1.47)	(0.35)	(0.63)	(1.0)	(6.29)	(16.52)	(8.70)	
Alcohol at Home	−10.22	−15.08	−11.55	10.94	−0.7	7.03	16.25	63.68	0.29	−0.65E-05	0.22
	(0.67)	(1.43)	(1.04)	(0.90)	(0.68)	(0.69)	(1.63)	(10.36)	(10.08)	(5.38)	
Food Away	11.63	−5.61	−8.45	−6.15	−40.0	−8.42	−1.14	43.48	0.96	−0.18E-05	0.61
	(0.56)	(0.39)	(0.56)	(0.37)	(2.90)	(0.61)	(0.84)	(5.49)	(23.31)	(11.0)	
Alcohol Away	−2.16	0.99	−3.37	4.72	0.90	5.98	9.55	26.33	0.65	−0.11E-05	0.35
	(0.52)	(0.34)	(1.10)	(1.42)	(0.32)	(2.13)	(3.46)	(18.87)	(8.03)	(3.18)	
Tobacco	4.63	6.54	1.43	2.26	−0.75	−2.14	−1.48	38.31	0.16	−0.46E-05	0.22
	(1.33)	(2.69)	(0.56)	(0.81)	(0.32)	(0.91)	(0.64)	(21.52)	(2.49)	(1.66)	
Clothing	25.69	16.38	−6.90	−63.71	68.84	−61.93	−31.59	32.07	0.11	−0.15E-05	0.52
	(1.05)	(0.96)	(0.39)	(3.27)	(4.19)	(3.73)	(1.96)	(2.94)	(22.87)	(7.47)	
Entertainment	−102.77	9.54	−13.25	27.18	40.16	−13.72	3.83	5.13	−0.12	0.21E-05	0.45
	(3.04)	(0.41)	(0.54)	(1.01)	(1.77)	(0.60)	(0.17)	(0.29)	(19.37)	(78.12)	
Personal Care	0.64	−2.37	1.67	−4.46	0.59	2.50	−4.70	21.94	0.48	−0.59 E-05	0.52
	(0.18)	(0.96)	(0.65)	(1.57)	(0.25)	(1.05)	(2.01)	(15.80)	(6.83)	(2.05)	
Medical Care	3.22	20.28	39.24	17.39	−19.86	−15.02	−19.83	94.29	0.61	−0.10E-05	0.41
	(0.11)	(1.03)	(1.89)	(0.77)	(1.04)	(0.79)	(1.06)	(8.90)	(11.34)	(4.59)	
Local Trans	−32.09	56.79	51.31	52.50	−21.80	26.48	79.02	46.69	0.15	−0.32E-05	0.45
	(0.96)	(2.45)	(2.11)	(1.98)	(0.98)	(1.18)	(3.61)	(1.91)	(23.14)	(12.09)	
Travel Trans	6.46	2.74	−1.32	63.61	−14.04	−5.94	−5.78	167.76	0.28	−0.26E-05	0.40
	(0.48)	(0.29)	(0.13)	(5.88)	(1.55)	(0.65)	(0.65)	(31.25)	(10.73)	(2.45)	
Miscellaneous	65.12	11.07	68.09	−48.39	−6.61	−15.05	−25.68	41.98	0.17	−0.21E-05	0.27
	(1.46)	(0.36)	(2.09)	(1.36)	(0.22)	(0.50)	(0.87)	(3.03)	(19.55)	(6.02)	
Home Heating	30.23	−77.13	−122.42	−119.76	77.59	71.37	−34.76	82.63	0.13	−0.30E-05	0.42
	(2.48)	(8.87)	(13.01)	(11.77)	(9.50)	(8.47)	(4.23)	(15.39)	(5.57)	(3.15)	
Electricity	−6.53	−0.88	24.0	−10.16	−29.40	−13.45	−12.49	51.58	0.13	−0.32E-05	0.45
	(0.95)	(0.18)	(4.73)	(1.84)	(6.34)	(2.91)	(2.75)	(6.43)	(9.66)	(5.78)	
Household Services	17.40	−0.52	−0.69	−1.77	−10.75	−15.63	−7.26	193.11	0.20	−0.51E-05	0.52
	(2.30)	(0.99)	(0.12)	(0.29)	(2.12)	(3.09)	(1.46)	(18.60)	(14.0)	(8.44)	
Homeownership	−12.76	36.38	49.28	134.10	−29.04	−44.48	−16.76	122.50	0.68	−0.16E-05	0.60
	(0.53)	(2.17)	(2.81)	(6.95)	(1.80)	(2.75)	(1.06)	(14.97)	(14.36)	(8.23)	

Note: The numbers in parentheses are the t-ratios.
(13) Afro-American Ethnic Background
(14) Household Lives in the North Central Region
(15) Household Lives in the Southern Region
(16) Household Lives in the Western Region
(17) Survey Conducted during the Winter Quarter
(18) Survey Conducted during the Spring Quarter
(19) Survey Conducted during the Summer Quarter
(20) Coefficient for the Mills Ratio
(21) Coefficient for Total Expenditure
(22) Coefficient for Total Expenditure Squared
(23) Coefficient of Determination