Chapter 1
Regression with a Two-Dimensional Dependent Variable

Lester D. Taylor

1.1 Introduction

This chapter focuses on how one might estimate a model in which the dependent variable is a point in the plane rather than a point on the real line. A situation that comes to mind is a market in which there are just two suppliers and the desire is to estimate the market shares of the two. An example would be determination of the respective shares of AT&T and MCI in the early days of competition in the long-distance telephone market. The standard approach in this situation (when such would have still been relevant) would be to specify a two-equation model, in which one equation explains calling activity in the aggregate long-distance market and a second equation that determines the two carriers' relative shares. An equation for aggregate residential calling activity might, for example, relate total long-distance minutes to aggregate household income, a measure of market size, and an index of long-distance prices; while the allocation equation might then specify MCI's share of total minutes as a function of MCI's average price per minute relative to the same for AT&T, plus other quantities thought to be important.

The purpose of these notes is to suggest an approach that can be applied in situations of this type in which the variable to be explained is defined in terms of polar coordinates on a two-dimensional plane. Again, two equations will be involved, but the approach allows for generalization to higher dimensions, and, even more interestingly, can be applied in circumstances in which the quantity to be explained represents the logarithm of a negative number. The latter, as will be seen, involves regression in the complex plane.

L. D. Taylor (✉)
University of Arizona, Tucson, AZ 86718, USA
e-mail: ltaylor@email.arizona.edu

J. Alleman et al. (eds.), *Demand for Communications Services - Insights and Perspectives*,
The Economics of Information, Communication, and Entertainment,
DOI: 10.1007/978-1-4614-7993-2_1,

1.2 Regression in Polar Coordinates

Assume that one has two firms selling in the same market, with sales of y_1 and y_2, respectively. Total sales will then be given by $y = y_1 + y_2$. The situation can be depicted as the vector (y_1, y_2) in the y_1y_2 plane, with y_1 and y_2 measured along their respective axes. In polar coordinates, the point (y_1, y_2) can be expressed as:

$$y_1 = r\cos\theta \tag{1.1}$$

$$y_2 = r\sin\theta \tag{1.2}$$

where

$$r = (y_1^2 + y_2^2)^{1/2} \tag{1.3}$$

$$\cos\theta = \frac{y_1}{\left(y_1^2 + y_2^2\right)^{1/2}} \tag{1.4}$$

$$\sin\theta = \frac{y_2}{\left(y_1^2 + y_2^2\right)^{1/2}} \tag{1.5}$$

One can now specify a two-equation model for determining y_1, y_2, and y in terms of $\cos\theta$ and r (or equivalently in $\sin\theta$ and r):

$$\cos\theta = f(X, \varepsilon) \tag{1.6}$$

and

$$\mathrm{r} = g(Z, \eta), \tag{1.7}$$

for some functions f and g, X and Z relevant predictors, and ε and η unobserved error terms.

At this point, the two-equation model in expressions (1.6) and (1.7) differs from the standard approach in that the market "budget constraint" ($y = y_1 + y_2$) is not estimated directly, but rather indirectly through the equation for the radius vector r. This being the case, one can legitimately ask, why take the trouble to work with polar coordinates? The answer is that this framework easily allows for the analysis of a market with three sellers and can probably be extended to markets in which n firms for $n \geq 4$ compete. Adding a third supplier to the market, with sales equal to y_3, the polar coordinates for the point (y_1, y_2, y_3) in 3-space will be given by:

$$y_1 = r\cos\alpha \tag{1.8}$$

$$y_2 = r\cos\beta \tag{1.9}$$

$$y_3 = r\cos\gamma \tag{1.10}$$

$$r = (y_1^2 + y_2^2 + y_3^2)^{1/2}, \tag{1.11}$$

where cos α, cos β, and cos γ are the direction cosines associated with (y_1, y_2, y_3) (now viewed as a vector from the origin). From expressions (1.8)–(1.10), one then has:

$$\cos\alpha = \frac{y_1}{(y_1^2 + y_2^2 + y_3^2)^{1/2}} \tag{1.12}$$

$$\cos\beta = \frac{y_2}{(y_1^2 + y_2^2 + y_3^2)^{1/2}} \tag{1.13}$$

$$\cos\gamma = \frac{y_3}{(y_1^2 + y_2^2 + y_3^2)^{1/2}} \tag{1.14}$$

A three-equation model for estimating the sales vector (y_1, y_2, y_3) can then be obtained by specifying explanatory equations for r in expression (1.11) and for any two of the cosine expressions in (1.12)–(1.14).

1.3 Regression in the Complex Plane

An alternative way of expressing a two-dimensional variable (y_1, y_2) is as

$$z = y_1 + iy_2 \tag{1.15}$$

in the complex plane, where y_1 and y_2 are real and i = $\sqrt{-1}$. The question that is now explored is whether there is any way of dealing with complex variables in a regression model. The answer appears to be yes, but before showing this to be the case, let me describe the circumstance that motivated the question to begin with. As is well-known, the double-logarithmic function has long been a workhorse in empirical econometrics, especially in applied demand analysis. However, a serious drawback of the double-logarithmic function is that it cannot accommodate variables that take on negative values, for the simple reason that the logarithm of a negative number is not defined as a real number, but rather as a complex number. Thus, if a way can be found for regression models to accommodate complex numbers, logarithms of negative numbers could be accommodated as well.

The place to begin, obviously, is with the derivation of the logarithm of a negative number. To this end, let v be a positive number, so that $-v$ is negative. The question, then, is what is $\ln(-v)$, which one can write as

$$\begin{aligned} \ln(-v) &= \ln(-1v) \\ &= \ln(-1) + \ln(v), \end{aligned} \tag{1.16}$$

which means that problem becomes to find an expression for ln(−1). However, from the famous equation of Euler,[1]

$$e^{i\pi} + 1 = 0 \tag{1.17}$$

one has, after rearranging and taking logarithms,

$$\ln(-1) = i\pi \tag{1.18}$$

Consequently,

$$\ln(-v) = i\pi + \ln(v). \tag{1.19}$$

To proceed, one now writes $\ln(-v)$ as the complex number,

$$\mathrm{z} = \ln(v) + i\pi, \tag{1.20}$$

so that (in polar coordinates):

$$\ln(v) = r\cos\theta \tag{1.21}$$

$$i\pi = i\, r\sin\theta, \tag{1.22}$$

where r, which represents the "length" of z—obtained by multiplying z by its complex conjugate, $\ln(v)$–$i\pi$, is equal to

$$r = \left[\pi^2 + \left(\ln(v)^2\right)\right]^{1/2}. \tag{1.23}$$

This is the important expression for the issue in question.

To apply this result, suppose that one has a sample of N observations on variables y and x that one assumes are related according to

$$f(y, x) = 0, \tag{1.24}$$

for some function f. Assume that both y and x have values that are negative, as well as positive, and suppose that (for whatever reason) one feels that f should be double-logarithmic, that is, one posits:

$$\ln(y_i) = \alpha + \beta\ln(x_i) + \varepsilon_i, i = 1, \ldots, N. \tag{1.25}$$

From the foregoing, the model to be estimated can then be written as:

$$z_i = \alpha + \beta w_i + \varepsilon_i, \tag{1.26}$$

where

$$\ln(y) \text{ if } y > 0$$

---

[1] See Nahin (1998, p. 67).

$$z = \left[\pi^2 + \left(\ln(-y)^2\right]^{1/2} \quad \text{if } y \le 0 \tag{1.27}$$

and

$$\ln(x) \text{ if } x > 0$$

$$w = \left[\pi^2 + \left(\ln(-x)^2\right]^{1/2} \quad \text{if } x \le 0. \tag{1.28}$$

1.4 An Example

In the Third Edition of *Consumer Demand in The United States*, the structure and stability of consumption expenditures in the United States was undertaken using a principal component analysis of 14 exhaustive categories of consumption expenditure using 16 quarters of data for 1996–1999 from the quarterly Consumer Expenditure Surveys conducted by the Bureau of Labor Statistics (BLS).[2] Among other things, the first two principal components (i.e., those associated with the two largest latent roots) were found to account for about 85 % of the variation in total consumption expenditures across households in the samples. Without going into details, one of the things pursued in the analysis was an attempt to explain these two principal components, in linear regressions, as functions of total expenditure and an array of socio-demographical predictors such as family size, age, and education. The estimated equations for these two principal components using data from the fourth quarter of 1999 are given in Table 1.1.[3] For comparison, estimates from an equation for the first principal component in which the dependent variable and total expenditure are expressed in logarithms are presented as well. As is evident, the double-log specification gives the better results. Any idea, however, of estimating a double-logarithmic equation for the second principal component was thwarted by the fact that about 10 % of its values are negative.

The results from applying the procedure described above to the principal component just referred to are given in Table 1.2. As mentioned, the underlying data are from the BLS Consumer Expenditure Survey for the fourth quarter of 1999, and consist of a sample of 5,649 U. S. households. The dependent variable in the model estimated was prepared according to expression (1.27), with the logarithm of y, for $y \le 0$, calculated for the absolute value of y. The dependent variable is therefore, $z = \ln(y)$ for the observations for which y is positive and $[\pi^2 + \ln(-y)^2]^{1/2}$ for the

[2] See Taylor and Houthakker (2010), Chap. 5).

[3] Households with after-tax income less than \$5,000 are excluded from the analysis.

Table 1.1 Principal component regressions BLS consumer expenditure survey 1999 Q4

Variables	PC1 linear		PC2 linear		PC1 double-log	
	Estimated coefficient	*t*-value	Estimated coefficient	*t*-value	Estimated coefficient	*t*-value
intercept	123.89	0.34	936.53	0.94	−1.5521	−20.85
totexp	0.48	209.39	−0.08	−14.04	1.0723	215.23
NO_EARNR	−108.59	−5.28	64.41	1.03	−0.0088	−2.20
AGE_REF	−5.27	−4.67	6.46	2.10	−0.0007	−3.52
FAM_SIZE	−37.14	−1.90	43.88	0.83	−0.0060	−1.80
dsinglehh	177.78	3.48	−297.13	−2.14	0.0527	5.95
drural	77.15	1.32	−549.14	−3.47	−0.0140	−1.41
dnochild	40.45	0.77	−299.72	−2.10	−0.0220	−2.46
dchild1	303.05	4.28	143.66	0.75	0.0612	5.06
dchild4	−63.40	−0.81	479.38	2.25	−0.0021	−0.16
ded10	−6.50	−0.08	92.49	0.40	−0.0022	−0.15
dedless12	183.89	0.53	−207.06	−0.22	0.1193	2.01
ded12	39.82	0.12	288.75	0.31	0.1000	1.69
dsomecoll	−4.09	−0.01	292.48	0.31	0.0810	1.37
ded15	−279.17	−0.80	1419.24	1.50	0.0727	1.22
dgradschool	−358.85	−1.02	2022.19	2.12	0.0675	1.13
dnortheast	−63.66	−1.24	96.39	0.69	-0.0232	-2.65
dmidwest	−91.50	−1.92	−266.89	−2.06	−0.0395	−4.85
dsouth	−26.85	−0.60	−424.42	−3.51	−0.0266	−3.49
dwhite	−202.39	−2.64	208.95	1.00	−0.0328	−2.51
dblack	8.01	−0.09	129.24	0.52	0.0107	0.69
dmale	−39.06	−1.13	59.61	0.64	−0.0088	−1.50
dfdstmps	138.25	1.55	−525.22	−2.17	0.0439	2.86
d4	−12.15	−0.20	141.63	0.87	0.0076	0.74

observations for which y is negative. All values of total expenditure are positive and accordingly require no special treatment.

From Table 1.2, one sees that, not only is total expenditure an extremely important predictor, but also that the R^2 of the logarithmic equation is considerably higher than the R^2 for the linear model in Table 1.1: 0.5204 versus 0.0598. However, as the dependent variable in logarithmic equation is obviously measured in different units than the dependent variable in the linear equation, a more meaningful comparison is to compute an R^2 for this equation with the predicted values in original (i.e., arithmetic) units. To do this, one defines two dummy variables:

$$\delta_1 = 1 \text{ if } y > 0 \tag{1.29}$$

$$\delta_2 = -1 \text{ if } y \leq 0$$

$$\delta_2 = -1 \quad \text{if } y \leq 0, \tag{1.30}$$

Table 1.2 Double logarithmic estimation of second principal component using expression (1.27)

Variables	Estimated coefficient	t-value
intercept	−3.1450	−11.54
Lntotexp	1.2045	66.03
NO_EARNR	−0.1161	−7.97
AGE_REF	−0.0014	−1.92
FAM_SIZE	−0.0017	−0.14
dsinglehh	0.2594	8.00
drural	−0.1269	−3.48
dnochild	−0.0294	−0.89
dchild1	0.2519	5.69
dchild4	0.1586	3.24
ded10	0.0410	0.76
dedless12	−0.0839	−0.39
ded12	−0.2072	−0.96
dsomecoll	−0.2101	−0.97
ded15	−0.2290	−1.05
dgradschool	−0.1416	−0.65
dnortheast	0.0522	1.63
dmidwest	0.0084	0.28
dsouth	−0.0118	−0.42
dwhite	−0.0117	−0.24
dblack	0.1248	2.18
dmale	−0.0693	−3.21
dfdstmps	0.2198	3.91
d4	−0.0266	−0.71
$R^2 = 0.5204$		

and then:

$$p = \left(\hat{z}^2 - \boldsymbol{\delta}_1 \boldsymbol{\pi}^2\right)^{1/2}. \tag{1.31}$$

A predicted value in arithmetic units, $\hat{p}$ follows from multiplying the exponential of p by δ_2:

$$\hat{p} = \delta_2 e^p \tag{1.32}$$

An R^2 in arithmetic units can now be obtained from the simple regression of y on $\hat{p}$[4]:

$$\begin{aligned} \hat{y} = &- 587.30 + 1.3105\,\hat{p} \qquad R^2 = 0.6714 \\ &(-19.39) \quad (107.42). \end{aligned} \tag{1.33}$$

[4] t-ratios are in parentheses. All calculations are done in SAS.

However, before concluding that the nonlinear model is really much better than the linear model, it must be noted that the double-log model contains information that the linear model does not, namely, that certain of the observations on the dependent variable take on negative values. Formally, this can be viewed as an econometric exercise in "switching regimes," in which (again, for whatever reason) one regime gives rise to positive values for the dependent variable while a second regime provides for negative values. Thus, one sees that when R^2s are calculated in comparable units, the value of 0.0598 of the linear model is a rather pale shadow of the value of 0.6714 of the "double-logarithmic" model. Consequently, a more appropriate test of the linear model *vis-à-vis* the double-logarithmic one would be to include such a "regime change" in its estimation. The standard way of this doing this would be to re-estimate the linear model with all the independent variables interacted with the dummy variable defined in expression (1.30). However, a much easier, cleaner and essentially equivalent procedure is to estimate the model as follows:

$$y = a_0 + a_1\delta_1 + (b_0 + b_1\delta_1)\hat{y}_p + \varepsilon, \tag{1.34}$$

where $\hat{y}_p$ is the predicted value of y in the original linear model and δ_1 is the dummy variable defined in expression (1.29). The resulting equation is:

$$\begin{aligned} \hat{y} &= 2407.26 \; - \; 8813.13\delta_1 \; - \; (0.5952 - 4.7517\delta_1)\hat{y}_p \\ &\quad (43.47) \quad (-109.94) \qquad (-14.41)(51.39) \\ R^2 &= 0.5728. \end{aligned} \tag{1.35}$$

However, "fairness" now requires that one does a comparable estimation for the nonlinear model:

$$\begin{aligned} \hat{y} &= 315.41 \; - \; 80.88\delta_1 \; + \; (1.7235 \; + \; 0.8878\delta_1)\hat{p} \\ &\quad (11.04) \quad (-0.90) \quad (-52.58) \quad (75.95) \\ R^2 &= 0.8085. \end{aligned} \tag{1.36}$$

As heads may be starting to swim at this point, it will be useful to spell out exactly what has been found:

To begin with, one has a quantity, y, that can take negative as well as positive values, whose relationship with another variable one has reason to think may be logarithmic.

As the logarithm of a negative number is a complex number, the model is estimated with a "logarithmic" dependent variable as defined in expression (1.27). The results, for the example considered, show that the nonlinear model provides a much better fit (as measured by the R^2 between the actual and predicted values measured in arithmetic units) than the linear model.

Since the nonlinear model treats negative values of the dependent variable differently than positive values, the nonlinear model can accordingly be viewed as allowing for "regime change." When this is allowed for in the linear model (by allowing negative and positive y to have different structures), the fit of the linear model (per Eq. (1.35)) is greatly improved. However, the same is also seen to be true (*cf.*, Eq. (1.35)) for the nonlinear model.

The conclusion, accordingly, is that, for the data in this example, a nonlinear model allowing for logarithms of negative numbers gives better results than a linear model: an R^2 of 0.81 versus 0.58 (from Eqs. (1.35) and (1.36)).

On the other hand, there is still some work to be done, for the fact that knowledge that negative values of the variable being explained are to be treated differently as arising from a different "regime" means that a model for explaining "regime" needs to be specified as well. Since "positive–negative" is clearly of a "yes–no" variety, one can view this as a need to specify a model for explaining the dummy variable δ_1 in expression (1.29). As an illustration (but no more than that), results from the estimation of a simple linear "discriminant" function, with δ_1 as the dependent variable and the predictors from the original models (total expenditure, age, family, education, etc.) as independent variables are given in Eq. (1.37)[5]:

$$
\begin{aligned}
\hat{\delta}_1 &= \underset{(0.83)}{0.0725} + \underset{(27.04)}{0.00001473\text{totexp}} + \text{other variables} \\
R^2 &= 0.1265
\end{aligned}
\tag{1.37}
$$

1.4.1 An Additional Example

A second example of the framework described above will now be presented using data from the Bill Harvesting II Survey that was conducted by PNR & Associates in the mid-1990s. Among other things, information in this survey was collected on households that made long-distance toll calls (both intra-LATA and inter-LATA) using both their local exchange carrier and another long-distance company.[6] While data from that era are obviously ancient history in relation to the questions and problems of today's information environment, they nevertheless provide a useful data set for illustrating the analysis of markets in which households face twin suppliers of a service.

For notation, let v and w denote toll minutes carried by the local exchange company (LEC) and long-distance carrier (OC), respectively, at prices p_{lec} and p_{oc}. In view of expressions (1.6) and (1.7) from earlier, the models for both intra-LATA and inter-LATA toll calling will be assumed as follows:

[5] Interestingly, a much improved fit is obtained in a model with total expenditure and the thirteen other principal components (which, by construction, are orthogonal to the principal component that is being explained) as predictors. The R^2 of this model is 0.46.

[6] Other studies involving the analysis of these data include Taylor and Rappoport (1997) and Kridel et al. (2002).

Table 1.3 IntraLATA toll-calling regression estimates bill harvesting data

Variables	cos θ		v/w	
	Estimated coefficient	*t*-ratio	Estimated coefficient	*t*-ratio
Models				
constant	0.5802	6.90	32.0549	2.48
income	−0.0033	−0.83	−0.0921	−0.15
age	0.0006	0.11	−1.8350	−2.13
hhcomp	0.0158	1.48	0.5623	0.34
hhsize	0.0257	2.06	−1.6065	−0.84
educ	0.0093	0.84	0.4702	0.28
lecplan	0.1887	4.02	34.5264	4.78
relpricelec/oc	−0.0950	−6.13	−6.5848	−2.77
	$R^2 = 0.1391$	df = 653	$R^2 = 0.0579$	df = 653
Variables	z		v + w	
	Estimated coefficient	*t*-ratio	Estimated coefficient	*t*-ratio
Models				
constant	160.3564	4.69	175.1315	4.75
income	2.6379	1.74	2.7803	1.70
age	2.1106	−2.58	−5.3457	−2.35
hhcomp	4.0298	1.31	6.0187	1.39
hhsize	4.6863	0.30	2.9896	0.59
educ	4.1600	−0.25	−0.5218	−0.12
lecplan	17.6718	6.36	129.1676	6.78
pricelec	−345.1262	−4.85	−393.7434	−5.13
priceoc	−74.9757	−1.44	−98.7487	−1.75
	$R^2 = 0.1277$	df = 652	$R^2 = 0.1414$	df = 652

$$\cos\theta = a + b \text{ income} + c(p_{\mathrm{lec}}/p_{\mathrm{oc}}) + \text{socio-demographic variables} + \mathrm{e} \tag{1.38}$$

$$\mathrm{r} = \alpha + \beta \text{ income} + \gamma p_{\mathrm{lec}} + \lambda p_{\mathrm{oc}} + \text{socio-demographic variables} + \varepsilon, \tag{1.39}$$

where

$$\cos\theta = \frac{v}{r} \tag{1.40}$$

$$z = \left(v^2 + w^2\right)^{1/2}. \tag{1.41}$$

The estimated equations for intra-LATA and inter-LATA toll calling are tabulated in Tables 1.3 and 1.4. As the concern with the exercise is primarily with procedure, only a few remarks are in order about the results as such. In the "shares" equations (i.e., with cos θ as the dependent variable), the relative price is the most important predictor (as is to be expected), while income is of little

Table 1.4 InterLATA toll-calling regression estimates bill harvesting data

Variables	Cos θ		v/w	
	Estimated coefficient	*t*-ratio	Estimated coefficient	*t*-ratio
Models				
constant	0.2955	2.74	1.2294	0.20
income	−0.0048	−0.92	0.2326	0.78
age	0.0156	2.23	−0.1116	−0.28
hhcomp	0.0157	1.21	1.1785	1.59
hhsize	0.0062	−0.36	0.0616	0.06
educ	0.0239	1.56	0.3775	0.43
lecplan	−0.0135	−0.16	−3.6824	−0.75
relpricelec/oc	−0.0855	−3.41	−2.2135	−1.54
	$R^2 = 0.0626$	df = 387	$R^2 = 0.0184$	df = 387
Variables	z		v + w	
	Estimated coefficient	*t*-ratio	Estimated coefficient	*t*-ratio
Models				
constant	217.1338	3.41	234.9010	3.50
income	2.3795	0.94	2.1877	0.82
age	−7.7823	−2.32	−7.8776	−2.23
hhcomp	−0.5843	−0.09	−1.7198	−0.26
hhsize	−3.8697	−0.48	−4.2879	−0.50
educ	20.5839	2.79	24.1515	3.11
lecplan	−3.9209	−0.09	−2.8309	−0.06
pricelec	−91.1808	−0.82	−124.9212	−1.07
priceoc	−576.5074	−2.98	−599.6088	−2.94
	$R^2 = 0.0838$	df = 386	$R^2 = 0.0888$	df = 386

consequence. In the "aggregate" equations (i.e., with z as the dependent variable), of the two prices, the LEC price is the more important for intra-LATA calling and the OC price for inter-LATA. Once again, income is of little consequence in either market. R^2*s*, though modest, are respectable for cross-sectional data. For comparison, models are also estimated in which the dependent variables are the ratio (v/w) and sum ($v + w$) of LEC and OC minutes.

Elasticities of interest that can be calculated from these four models include the elasticities of the LEC and OC intra-LATA and inter-LATA minutes with respect to the LEC price relative to the OC price and the elasticities of aggregate intra-LATA and inter-LATA minutes with respect to the each of the carrier's absolute price.[7] The resulting elasticities, calculated at sample mean values, are tabulated in Table 1.5. The elasticities in the "comparison" models are seen to be quite a bit

[7] The elasticity for LEC minutes in the "cos θ" equation is calculated as $\hat{c}\bar{h}\bar{z}/\bar{v}$, where h denotes the ratio of the LEC price to the OC price. The "aggregate" elasticities are calculated, not for the sum of LEC and OC minutes, but for the radius vector z (the positive square root of the sum of squares of LEC and OC minutes). The OC share elasticities are calculated from equations in which the dependent variable is sin θ.

Table 1.5 Price elasticities Models in Tables 1.3 and 1.4

cos θ, Z		V/W, V + W	
Elasticity	Value	Elasticity	Value
IntraLATA toll			
Share		Share	
LEC (own)	−0.18	LEC (own)	−0.52
LEC (cross)	0.12	LEC (cross)	0.55
OC (own)	−0.24	OC (own)	−0.29
OC (cross)	0.34	OC (cross)	0.59
Aggregate		Aggregate	
LEC price	−0.45	LEC price	−0.45
OC price	−0.11	OC price	−0.12
InterLATA toll			
Share		Share	
LEC (own)	−0.24	LEC (own)	−0.52
LEC (cross)	0.12	LEC (cross)	0.15
OC (own)	−0.05	OC (own)	−0.20
OC (cross)	0.09	OC (cross)	0.40
Aggregate		Aggregate	
LEC price	−0.11	LEC price	−0.13
OC price	−0.70	OC price	−0.66

larger than in the "polar-coordinate" models for the LEC and OC shares but are virtually the same in the two models for aggregate minutes.

As the dependent variables in the "polar-coordinates" and "comparison" models are in different units, comparable measures of fit are calculated, as earlier, as R^2s between actual and predicted values for the ratio of LEC to OC minutes for the share models and sum of LEC and OC minutes for the aggregate models. For the "polar-coordinate" equations, estimates of LEC and OC minutes (i.e., v and w) are derived from the estimates of cos θ to form estimates of v/w and $v + w$. R^2s are then obtained from simple regressions of actual values on these quantities. The resulting R^2s are presented in Table 1.6. Neither model does a good job of predicting minutes of non-LEC carriers.

Table 1.6 Comparable R^2s for share and aggregate models in Tables 1.3 and 1.4

Toll market	cos θ	z	v/w	v + w
Models				
IntraLATA	0.0428	0.1432	0.0579	0.1414
InterLATA	0.0058	0.0892	0.0184	0.0888

1.5 Final Words

The purpose of these notes has been to suggest procedures for dealing with dependent variables in regression models that can be represented as points in the plane. The "trick," if it should be seen as such, is to represent dependent variables in polar coordinates, in which case two-equation models can be specified in which estimation proceeds in terms of functions involving cosines, sines, and radius-vectors. Situations for which this procedure is relevant include analyses of markets in which there are duopoly suppliers. The approach allows for generalization to higher dimensions, and, perhaps most interestingly, can be applied in circumstances in which values of the dependent variable can be points in the complex plane. The procedures are illustrated using cross-sectional data on household toll calling from a PNR & Associates Bill Harvesting survey of the mid-1990s and data from the BLS Survey of Consumer Expenditures for the fourth quarter of 1999.

References

Kridel DJ, Rappoport PN, Taylor LD (2002) IntraLATA long-distance demand: carrier choice, usage demand, and price elasticities. Int J Forecast 18(4):545–559

Nahin P (1998) An imaginary tale: the story of the square root of −1. Princeton University Press, New Jersey

Taylor LD, Houthakker HS (2010) Consumer demand in the United States: prices, income, and consumer behavior, 3rd edn. Springer, Berlin

Taylor LD, Rappoport PN (1997) Toll price elasticities from a sample of 6,500 residential telephone bills. Inf Econ Policy 9(1):51–70