

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Anticipations and Purchases: An Analysis of Consumer Behavior

Volume Author/Editor: F. Thomas Juster

Volume Publisher: Princeton University Press

Volume ISBN: 0-87014-079-5

Volume URL: <http://www.nber.org/books/just64-1>

Publication Date: 1964

Chapter Title: Buying Intentions and Purchase Probability: I

Chapter Author: F. Thomas Juster

Chapter URL: <http://www.nber.org/chapters/c1031>

Chapter pages in book: (43 - 103)

CHAPTER 3

Buying Intentions and Purchase Probability: I

Introduction

THE empirical analysis in Chapter 2 demonstrated the existence of a very systematic relation between the proportions of households reporting intentions to buy a particular commodity, given alternative questions about buying intentions, and the subsequent purchase rates of households classified as intenders and nonintenders by the alternative questions. Using the terminology developed above, let:

- p_i be the fraction of the i th subsample reporting intentions to buy at the beginning of the forecast period, where the i th subsample was asked a particular question or set of questions about intentions to buy;
- x_i be the fraction of the i th sample purchasing during the forecast period, i.e., the period over which purchase behavior is to be analyzed;
- r_i be the fraction of intenders in the i th subsample who bought during the forecast period;
- s_i be the fraction of nonintenders in the i th subsample who bought during the forecast period.

The following empirical generalizations are observable in the data. First, the fraction of intenders who buy— r_i —is always greater than the fraction of nonintenders who buy— s_i —whatever the forecast period examined or the alternative definitions of intender and nonintender. Hence, the responses on a survey of buying intentions have behavioral significance; on the average, the subsequent purchase rates of yes and no responders do differ. Secondly, the proportion of intenders— p_i —is negatively correlated with the fraction of both intenders and nonintenders who purchase; intentions questions with relatively high values of p tend to have relatively low values of both r and s . Thirdly, the proportion of intenders is negatively correlated with $r_i - s_i$, that is, with the algebraic difference in purchase rates between intenders and nonintenders.¹ The first two generalizations have already been discussed at length in Chapter 2; the third is documented by Table 8.

The Probability Hypothesis

All three generalizations are consistent with the intuitively plausible hypothesis that responses to questions about buying intentions are essentially a reflection of the respondent's subjective estimate of his purchase

¹ The second and third generalizations apply only to a comparison among alternative intentions questions for a given commodity.

TABLE 8
DIFFERENCE IN PURCHASE RATES OF INTENDERS (τ) AND NONINTENDERS (ρ), SELECTED INTENTIONS QUESTIONS AND PURCHASE PERIODS

Buying Intentions Question	Auto	Furniture	Carpets and Rugs	Hi-Fi Components	Range	Refrigerator	Washing Machine	TV	Air Conditioner	Clothes Dryer	Dishwasher	Food Freezer	Garbage Disposal Unit
SIX MONTHS SUBSEQUENT TO SURVEY													
A ₁	40.8	35.4	31.8	25.0	43.7	46.4	34.6	14.8	41.8	24.4	37.3	29.5	25.6
B ₁	42.5	28.1	20.2	29.0	32.9	34.3	32.4	19.6	17.2	30.6	25.7	26.2	23.9
C ₁	33.2	29.8	20.8	18.1	28.5	26.9	30.4	22.7	23.1	19.3	21.3	22.0	24.0
D ₁	25.9	23.1	12.6	18.8	22.3	28.9	22.2	16.4	22.4	16.4	24.1	16.2	15.5
A ₁ or A ₂	22.8	21.0	15.1	9.6	22.4	21.6	17.9	7.5	20.0	14.7	15.8	13.2	13.3
C ₂	22.3	20.2	13.6	9.3	19.2	20.6	21.8	12.2	10.4	9.6	9.2	7.6	11.7
B ₁ or B ₂	13.9	16.7	9.9	9.0	12.7	13.0	19.5	12.5	8.6	9.7	8.1	5.9	6.5
TWELVE MONTHS SUBSEQUENT TO SURVEY													
A ₁	46.0	38.6	42.8	36.0	49.1	53.0	37.8	30.8	42.8	41.6	48.8	37.8	36.1
B ₁	44.5	34.1	33.7	41.9	50.1	45.7	40.9	40.7	17.7	38.8	41.0	29.1	46.2
C ₁	40.3	35.3	27.6	33.5	41.7	38.7	38.2	36.0	24.6	28.5	31.3	29.8	33.2
D ₁	33.9	30.7	23.7	28.8	38.1	41.4	34.1	27.1	23.8	26.6	38.4	25.7	28.8
A ₁ or A ₂	31.9	25.0	26.1	20.8	30.1	27.0	28.3	20.9	22.0	25.3	23.0	20.0	21.9
C ₂	32.9	27.2	20.3	17.5	30.1	29.3	28.5	24.9	11.1	18.4	15.0	11.5	18.0
B ₁ or B ₂	19.9	23.9	19.8	17.9	22.0	22.0	28.7	23.3	9.9	17.1	15.3	10.3	16.3

SOURCE: Derived from columns 9 through 12 of Table 2. For buying intentions questions, see Table 1.

probability, designated Q . More precisely, the statement "I intend [plan] to buy X within Y months" simply means that the respondent judges that his probability of purchasing X within Y months is high enough for him to consider yes a more accurate answer than no, given the particular question asked.² Moreover, I read the data as suggesting that households have a rather precise notion as to their purchase probability, given the commodity and the time period, and further suggesting that the distribution of purchase probabilities among households is a continuous function. The hypothesis that responses to questions about buying intentions reflect judgments about the respondent's subjective probability of purchase is essentially a priori—I can think of no other reasonable meaning that might be attached to the statement that "I intend to buy X within Y months." The other two parts of the probability hypothesis—that respondents have relatively precise quantitative notions of their purchase probability and that the probability function is continuous—are inferences from the data.

Some Alternative Hypotheses

Not only are the data broadly consistent with these hypotheses but they also clearly contradict some alternative hypotheses. It could be argued, for example, that households have only the very roughest of notions about purchase probability, and that any more refined classification than a simple division into "high-" or "low-probability" groups would be meaningless. But if that were so it would be observed that either (1) the proportion of intenders who purchase is about the same for all intentions variants and is independent of the proportion of the sample classified as intenders, or (2) the proportion of intenders in the sample is about the same for most or all of the intentions variants. The data, however, do not show either of these characteristics; rather, they indicate quite clearly that a change in the intentions question will alter both the proportion of

² A yes answer does not necessarily imply a purchase probability equal to 0.5 or more, although the data are not inconsistent with this proposition. All questions about buying intentions in this and other surveys use imprecise language that will be interpreted differently by different respondents. For example, the probabilities associated with the statement "I plan to purchase a car within a year" may range all the way from perfect subjective certainty ($Q = 1$) to something very much less ($Q = 0.25$). The reason is that the meaning of the words "plan" or "intend" is imprecise. To some people "plan" implies that they have thought about purchasing and might buy within the year; to others "plan" may convey a much more rigorous interpretation: that they have figured out next year's budget, they will purchase a car unless an unexpectedly adverse (and important) event happens, and they have already shopped around or will in the near future.

intenders and the purchase rates of intenders and nonintenders in a way consistent with the probability model.

Another possibility is that the universe distribution is not continuous but basically trichotomous, comprising those who *really* expect to buy, those who do not, and those for whom contingencies determine the answer.³ But if that were true the purchase rates for definite intenders and definite nonintenders ought to approach 1.0 and zero, respectively, substantially higher (lower) than those observed in the data.⁴

Analysis of the Probability Model

The probability hypothesis is worth elaborating before its consistency with the data is examined more thoroughly. Let me begin by assuming that all households know the subjective probability that they will purchase a particular product during some specified future period of time—the forecast period. These probabilities are distributed in some way between zero (perfect certainty of nonpurchase) and unity (perfect certainty of purchase). The mean of this *ex-ante* probability distribution—designated x' —is the expected value of the population purchase rate during the forecast period. Evidently, x' constitutes the best estimate of the population purchase rate that can be obtained from survey data of this sort.⁵

The distribution of purchase probabilities must be highly skewed for any product purchased by a small fraction of the population, since most households must have probabilities close to zero for mean probability in the population to be relatively small. For an item purchased by a large majority of the population during the forecast period, the distribution must be skewed in the opposite direction, since most households are bound in that case to have relatively high probabilities. The distribu-

³ Trichotomous distributions of this sort undoubtedly do exist. Suppose, for example, a survey were made of "expected births during the next six months." Everyone with a wife more than three months' pregnant would evidently say yes, everyone with a wife either two months pregnant or not pregnant at all would answer no, and a few people would say that "it depends." The birth rate would be close to 1.0 for the yes group, close to zero for the no group, and around 0.5 for those reporting that "it depends." And the responses would be the same whether the survey asked about "definite" or "probable" births. But if the survey had been concerned with expected births during the next two years, the data would begin to take on the character of a continuous distribution.

⁴ Wholly unanticipated developments might account for some departure from these expected purchase rates, as will be discussed below; but it seems implausible to me that such events are frequent enough to account for the observed results.

⁵ Although mean probability (x') in the sample is the best estimate of population purchases (x), it does not necessarily follow that x' will be an accurate predictor of x . For example, important and unforeseen events may result in an actual value of x greater or smaller than x' for any given forecast period; hence, the time series correlation between the two cannot be determined a priori.

tion is likely to be symmetrical for an item purchased by about half the population, although it could resemble either a normal (bell-shaped) function, a relatively peaked or relatively flat function, or possibly a U-shaped function.

An Illustrative Distribution Function

Given some distribution of purchase probabilities, what would be observed from an analysis of data obtained by a survey of intentions to buy and a follow-up survey of actual purchases? Given the characteristics (planning period and certainty dimension) of the buying-intentions question asked in the survey, all households with subjective purchase probabilities (Q) greater than the minimum probability associated with the question—designated C_i —would report that they intended to buy; all others would report the reverse. It will not necessarily be true that every household reporting an intention to buy has a higher purchase probability than every other household, since the minimum probability implied by the question may itself vary among households. It is obvious, however, that mean probability for intenders will exceed that for nonintenders except under extreme assumptions.⁶

Alteration of the certainty dimension (definite, probable, etc.) or time horizon (six months, twelve months, etc.) of the intentions question would result in a different distribution of intenders and nonintenders, since C_i must vary with the certainty dimension of the question and Q with time horizon. The lower the degree of certainty implied by the i th survey question, holding the time horizon constant, the lower the cut-off probability as judged by the typical respondent. Similarly, the longer the time horizon of the question, holding degree of certainty constant, the higher the Q of the typical respondent. For if a respondent thinks the chances are one in four that he will buy within six months, he will typically

⁶ If all households interpret the i th intentions question as having the same cut-off point, all intenders must have a higher probability than all nonintenders; hence mean probability— q —is necessarily greater for intenders than for nonintenders. If C_i varies among households, q for intenders will exceed that for nonintenders except in the extreme case where Q and C_i are perfectly correlated, that is, where households with high values of Q systematically interpret the question as having a high cut-off probability, and vice versa. Otherwise, it must be true that q is greater for intenders than for nonintenders even if C_i varies: In any subgroup of households that interpret the i th question as having the same C_i , mean probability for intenders will exceed that for nonintenders. The means for all such subgroups combined will evidently have the same property, since in each subgroup q for intenders, although different among the groups, must be greater than q for nonintenders. Thus it is only necessary to find some variation in Q among households with a common interpretation of C_i in order to demonstrate the above proposition.

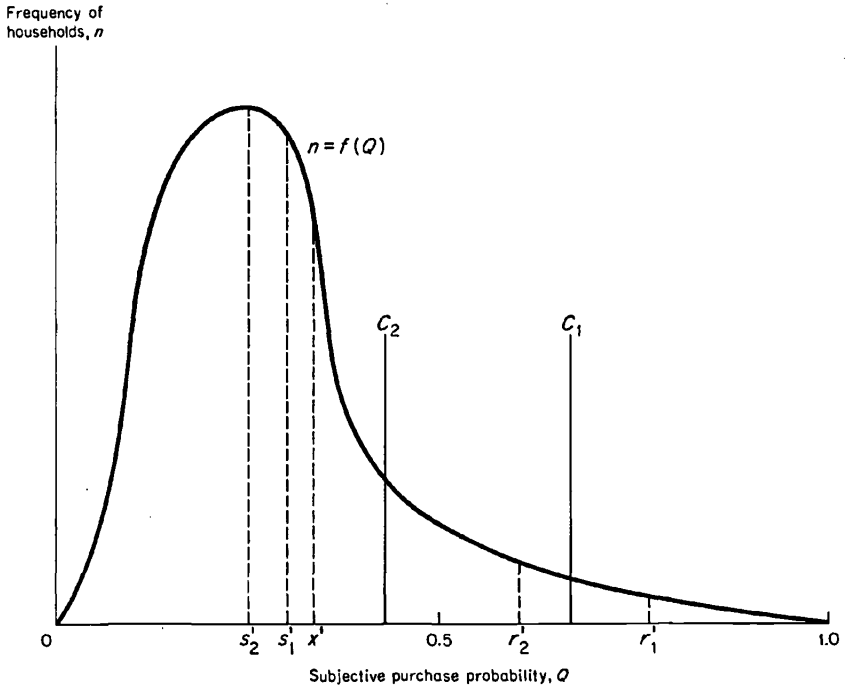
BUYING INTENTIONS AND PURCHASE PROBABILITY: I

judge the chances better than one in four that he will buy within a year, and better still that he will buy within two years.

Further, it can be shown that not only C_i but also the mean probabilities for intenders and nonintenders will be related to the specifications of the buying-intentions question, with the mean probability for both groups being higher the higher the cut-off probability. These relations are

CHART 1

Illustrative Distribution of Households
by Subjective Purchase Probability



illustrated in Chart 1. For ease of exposition I assume that C_i is the same for all households, although this assumption is not essential to most of the results. The probability distribution relating to a specified forecast period (say, one year) is shown as a skewed function, $n = f(Q)$, with most households having relatively low values of Q . If a sample of 100 households is asked whether they “definitely” intend to buy within a year, all those with purchase probabilities higher than the C implied by the question— C_1 in Chart 1—will answer yes; the rest of the sample will answer no. If asked whether they might “possibly” buy within a year, a

BUYING INTENTIONS AND PURCHASE PROBABILITY: I

larger number will respond affirmatively, since this question must be interpreted as having a lower cut-off probability, designated as C_2 .⁷ Thus, numbers of households (N) equal to, respectively,

$$N_1 = \int_{C_1}^1 f(Q) dq,$$

and

$$N_2 = \int_{C_2}^1 f(Q) dq$$

will report that they "definitely" or "probably" will buy within a year. Similarly, numbers of households equal to, respectively,

$$100 - N_1 = \int_0^{C_1} f(Q) dq,$$

and

$$100 - N_2 = \int_0^{C_2} f(Q) dq$$

will report that they do not intend to buy within a year when asked about definite or possible purchases. The relative frequencies of definite intenders and nonintenders correspond to the areas under the distribution function from C_1 to 1.0 and from zero to C_1 , respectively. When these households are asked about possible rather than definite purchases, the areas from C_2 to 1.0 and zero to C_2 represent the respective numbers of intenders and nonintenders.

The mean of the distribution function, x' , is roughly 0.25 on the scale in Chart 1. The mean probability among intenders and nonintenders depends on the cut-off probability. When C_1 is the cut-off, the magnitude designated r'_1 is the mean probability for intenders; that designated s'_1 , the mean probability for nonintenders. When C_2 is the cut-off point, both means are necessarily lower— r'_2 for intenders and s'_2 for nonintenders: lowering the cut-off probability shifts the highest probability segment of the original nonintender group into the intender classification; mean probability for the "shifter" is necessarily lower than for (original) intenders and higher than for (original) nonintenders.

A number of generalizations can be derived from Chart 1. First, mean probability for intenders is always higher than for nonintenders; hence, the algebraic difference between the two is necessarily positive. This generalization holds for any cut-off probability C_i and, therefore, for any

⁷ Anyone who responds affirmatively to the first question will obviously respond affirmatively to the second, but the reverse is not necessarily true. It follows that the maximum probability among "marginal" intenders—those who respond no to definite but yes to possible purchase, must be less than the minimum probability for those who respond yes to both questions.

survey question about intentions to buy. Secondly, both r'_i and s'_i must be positively related to C_i , although the relationship is generally nonlinear. Since the proportion of intenders in the sample— p_i —depends on and is negatively correlated with C_i , both r'_i and s'_i must be negatively related to p_i . Third, the algebraic quantity $(r' - s')_i$ also appears to be positively correlated with C_i , negatively, with p_i . This result is entirely due to the shape of the distribution function in Chart 1, and is neither true for all possible distribution functions nor even necessarily true for all functions similar to the one shown above.⁸ The relation between $r_i - s_i$ and p_i is discussed at greater length in the appendix to this chapter.⁹

Finally, $(r' - s')_i$ as a function of C_i generates a relatively large variance as C_i moves from zero to unity. This is readily seen by moving the cut-off point to extreme positions. When the cut-off probability is very high, p will approach zero, r' will approach unity, and s' will approach x' . When the cut-off probability is very low, p will approach unity, r' will approach x' , and s' will approach zero. Thus, the variance of $(r' - s')_i$ —as a function of C_i —depends largely on whether $1 - x'$ and x' are close together or far apart. Since the distribution function in Chart 1 has a mean x' of about 0.25, $(r' - s')_i$ will have a relatively large variance.

Some Alternative Distribution Functions

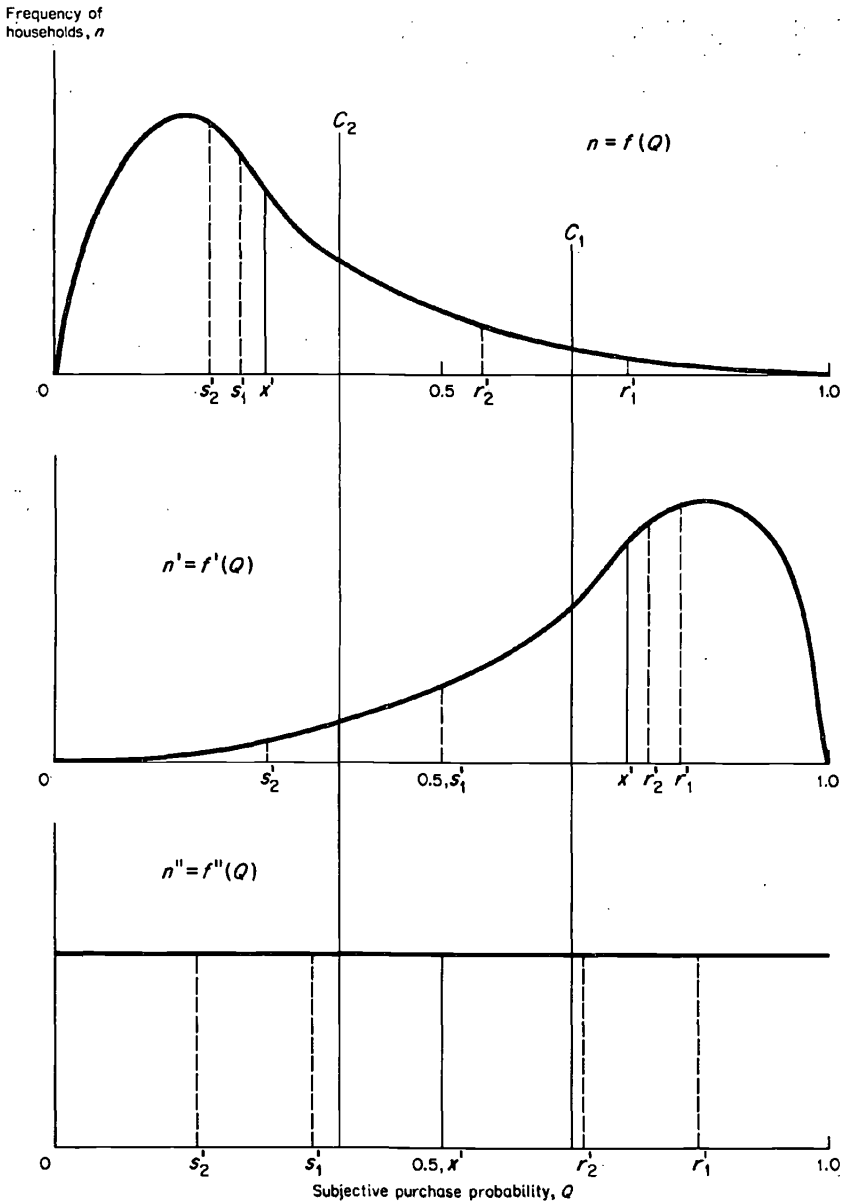
It is useful to observe how these relations would be altered by different assumptions about the shape of the distribution function. Chart 2 contains three types of functions: the top panel has a skewed function much like that in Chart 1. The function in the middle panel is also skewed, but the skew is in the opposite direction. The bottom panel shows a rectangular distribution function. If the same questions about intentions to buy,

⁸ For the function $n = f(Q)$ in Chart 1, variations in the cut-off probability that are confined to the region around C_1 and C_2 cause large movements in r' relative to s' . The distribution function in Chart 1 happens to have a relatively small density and a flat slope in the probability region around these cut-off points. Consequently, a change in C has a relatively large impact on the mean probability for households with probabilities higher than C_i , that is, on r'_i . The peak frequencies for the function in Chart 1, and the vast bulk of the total frequencies, are below both cut-off points; therefore a change in C has a relatively small influence on the mean probability for households with probabilities lower than C_i , that is, on s'_i .

⁹ The conclusions in the appendix are as follows: the relation between $(r' - s')_i$ and p_i is rather complicated, and appears to depend on the absolute size of both p_i and x' . Given the general shape of the distribution function, $(r' - s')_i$ seems to be negatively correlated with p_i throughout the entire range of p_i (from zero to 100 per cent) provided that x' is sufficiently small. The correlation between the two seems to be positive throughout the entire range of p_i provided that x' is sufficiently large. For intermediate values of x' the correlation seems to depend on the size of p_i , being negative when p_i is small and positive when p_i is large.

CHART 2

Alternative Distributions of Households
by Subjective Purchase Probability



involving the same cut-off points, were asked about commodities characterized by these three distributions, the following relations would be observed.

For all three functions r'_i is necessarily greater than s'_i ; hence, $(r' - s')_i$ is always positive. In the top panel, r'_i , s'_i , and $(r' - s')_i$ are all positively correlated with C_i and, therefore, negatively with p_i . Given a lower cut-off probability, p will increase; and the mean probabilities for intenders and nonintenders (necessarily) and the difference between the two (generally) will decline. For the middle panel— $n' = f'(Q) - r'_i$ and s'_i are also positively correlated with C_i and negatively with p_i , as before; but $(r' - s')_i$ appears to have exactly the reverse relationship—negative correlation with C_i , positive with p_i . This is clearly true when C_i is such that p_i is relatively large, but does not necessarily follow for all values of C_i and p_i . In the rectangular distribution— $n'' = f''(Q) - r'_i$ and s'_i are again positively correlated with C_i and negatively with p_i , but there is no correlation at all between $(r' - s')_i$ and p_i , because $(r' - s')_i$ is a constant, equal to 0.5. Finally, $(r' - s')_i$ —as a function of C_i or p_i —has a relatively large variance in the two skewed distributions but none in the rectangular one: the appropriate generalization here is that the more skewed the distribution function, the greater the variance in $(r' - s')_i$.¹⁰

A Mathematical Note

It is interesting to note that almost all the above analytical generalizations can be obtained with a minimum of assumptions about the distribution function. Let

x' = mean purchase probability in the sample, $0 \leq x' \leq 1$

p = fraction of the sample reporting intentions to buy, $0 \leq p \leq 1$

$1 - p$ = fraction not reporting intentions to buy

r' = mean purchase probability among intenders, $0 \leq r' \leq 1$

s' = mean purchase probability among nonintenders, $0 \leq s' \leq 1$

By definition,

$$1.0 \quad x' = r'p + s'(1 - p)$$

$$1.1 \quad x' = p(r' - s') + s'$$

Since x' is a weighted average of r' and s' and intenders must have

¹⁰ If the distribution function were a normal curve the relations in the rectangular function would generally hold, except that $(r' - s')_i$ would not be a constant. The quantity $(r' - s')_i$ would be (linearly) uncorrelated with p_i and would have a relatively small variance; there would be positive correlation between $(r' - s')_i$ and p_i in the region where p_i was relatively large; negative correlation between the two, where p_i was relatively small.

higher mean probabilities than nonintenders, it follows that:

$$\begin{aligned} r' &> x' > s' \\ r' &> s' \end{aligned}$$

Further, since x' is the mean of a frequency distribution comprising the points (r', p) and $[s', (1 - p)]$, the variance of the distribution— V —can be written in any one of the following ways:

$$2.0 \quad V = pr'^2 + (1 - p)s'^2 - x'^2$$

$$2.1 \quad V = p(1 - p)(r' - s')^2$$

$$2.2 \quad V = (r' - x')(x' - s')$$

If both x' (the mean probability) and V (the variance) are assumed to be constants, the following relations can be derived. Rewriting equation 2.2:

$$2.2 \quad r' = [V/(x' - s')] + x'$$

Hence,

$$2.21 \quad \frac{dr'}{ds'} = V/(x' - s')^2$$

and $\frac{dr'}{ds'} > 0$. Rewriting equation 1.1,

$$1.1 \quad p = \frac{x' - s'}{r' - s'}$$

Therefore,

$$1.11 \quad \frac{dp}{dr'} = \frac{-2p}{(r' - s')}$$

and $\frac{dp}{dr'} < 0$, since $r' > s'$. In addition,

$$1.12 \quad \frac{dp}{ds'} = 2[(x' - r')/(r' - s')^2]$$

Hence,

$$\frac{dp}{ds} < 0$$

Finally, rewriting equation 2.1

$$2.1 \quad (r' - s')^2 = V/p(1 - p)$$

Therefore,

$$2.11 \quad \frac{d(r' - s')}{dp} = \frac{-V(1 - 2p)}{2p^2(1 - p)^2(r' - s')}$$

Hence,

$$\frac{d(r' - s')}{dp} \begin{cases} < 0 \text{ when } p < 0.5 \\ = 0 \text{ when } p = 0.5 \\ > 0 \text{ when } p > 0.5 \end{cases}$$

Relation 2.21 says that r' and s' are positively correlated although not necessarily in a linear fashion. Relations 1.11 and 1.12 say that both r' and s' are negatively correlated with p . All these generalizations hold for any type of distribution function and for any value of p , and are identical to the comparable relations derived from Chart 2. Relation 2.11 contradicts one of the generalizations discussed above, however, since it specifies that the association between $r' - s'$ and p depends entirely on whether $p \gtrless 0.5$. In Charts 1 and 2, the illustrative negative correlation between $r' - s'$ and p is, in fact, associated with a value of p less than 0.5 [illustrative function $n = f(Q)$], while the illustrative positive correlation between the two is associated with a value of p greater than 0.5 [$n' = f'(Q)$, in Chart 2]. The contradiction is real, however, since it was argued above that $r' - s'$ is negatively correlated with p throughout the entire range of p provided that x' is sufficiently small (see note 9). The geometrical analysis is based on the assumption that the entire distribution is fixed and given, but relation 2.11 is based on the much weaker assumption that the mean and variance of the distribution are fixed. The difference in assumption is sufficient to account for the difference in conclusions.¹¹

Empirical Tests of the Model

The empirical data contained in Chapter 2 and in Table 8 are obviously related to these analytical constructs. The mean purchase probability in the i th sample (x'_i) is evidently a forecast of the proportion of the i th sample that will actually purchase (x_i). Similarly, the mean probabilities for intenders (r'_i) and nonintenders (s'_i) are forecasts of the purchase rates for intenders (r_i) and nonintenders (s_i). Alternatively, the *ex post* x_i is an

¹¹ The estimate of the variance underlying relation 2.11 is based on two points. If these two points are drawn from a "true" distribution that is actually continuous (for example, from any of the distributions shown in Charts 1 and 2), the variance computed from combinations of observed points (r', p); and [$s', (1 - p)$]; will not be the same for every i . An estimate of the variance computed in this way will be relatively large when p takes on intermediate values, relatively small when p is either very large or very small. At the extreme, i.e., when p is either zero or one, the variance estimated from two such points will be zero, since the mean will be the same as either r'_i or s'_i . For the variance to be constant, as specified in relation 2.11, the (implicit) distribution must itself shift as p changes. As a consequence, this relation will never hold exactly if the distribution function is continuous, and will hold approximately only within moderate ranges of p .

BUYING INTENTIONS AND PURCHASE PROBABILITY: I

estimate of what the *ex ante* x'_i must have been; r_i , of what r'_i must have been; and s_i , of what s'_i must have been. Interpreted in this way columns 9 and 11 of Table 2 contain alternative estimates of r'_i ; corresponding to different cut-off points and proportions of intenders, while columns 10 and 12 have a complementary set of estimates of s'_i . The data do not contain any direct empirical estimates for C_i , although column 3 of Table 2 shows the proportion of households with probabilities between (the unknown) C_i and 1.0, that is, p_i .

THE PROBLEM OF BIAS

It is important to note that reconstruction of the *ex ante* probability distribution may not be possible from the observed data, which consist of purchase rates (estimates of *ex ante* mean probability) for a number of complementary chunks drawn from the underlying distribution. Two kinds of bias must be reckoned with in assuming that the observed purchase rates for intenders or nonintenders are valid estimates of mean *ex ante* probability in the appropriate segment of the distribution. The first is essentially a regression bias between *ex ante* and *ex post* magnitudes; the second and probably more important bias arises because the cut-off probability for any given buying-intentions question is not likely to be invariant among households.

Unforeseen events will influence the relation between the observed r_i and s_i points in Table 2 and the corresponding *ex ante* values of r'_i and s'_i . Many households report a very low or zero *ex ante* purchase probability, given the commodity and the forecast period. But some of these will encounter conditions that are wholly unanticipated, e.g., a severe accident, fully covered by insurance, with a new automobile; others will disregard likely contingencies, e.g., the necessity of replacing an old washing machine should it break down. In a survey of buying intentions, both these types of households would report that they do not intend to purchase, but in a follow-up survey many would have actually purchased. On the other side, any number of unexpected contingencies could result in postponement of a prospective purchase that had an *ex ante* probability of close to 1.0. All households in this situation would have reported buying intentions, but many would not have been observed to purchase.

Unanticipated events of this sort (those whose probability of occurrence is not taken account of in the estimate of subjective purchase probability) will influence the relation between the *ex post* rate of purchase and the *ex ante* mean probability for a group of households even if such events are distributed at random among the population. Unfavorable surprises will

have a systematically stronger influence on the difference between expected and observed behavior the higher the mean *ex ante* probability for any group of households, while favorable surprises will have a systematically stronger influence on this difference the lower the mean *ex ante* probability. As a consequence r tends to underestimate r' while s tends to overestimate s' ; there seems to be no inherent reason for x to be a biased estimate of x' . And these considerations imply a progressively more serious upward bias in estimates of s'_i based on s_i as C_i decreases.¹² The bias in estimates of r'_i obtained from r_i is just the reverse; here, the empirical estimates are too low as C_i approaches 1.0, gradually becoming unbiased as C_i approaches zero.

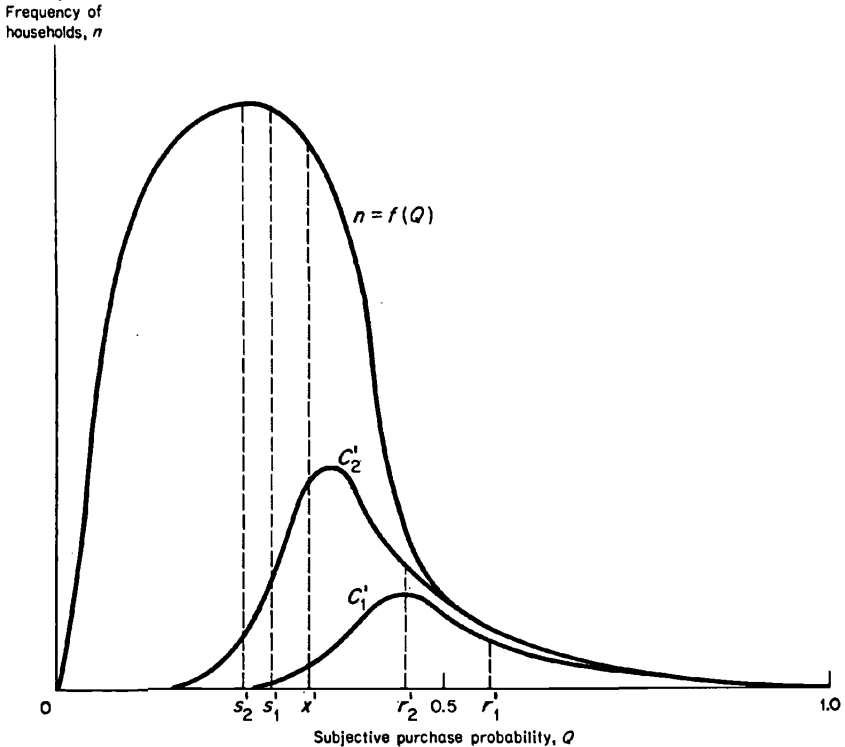
The second source of bias is illustrated by Chart 3, which shows the effect on r' and s' of variation in the cut-off probability among households. Assume that a distribution of purchase probabilities among households, $n = f(Q)$, is given. Whether the cut-off probability for the i th intentions question varies or not, every household with a purchase probability higher than (its estimate of) C_i will report an intention to buy, and every household with a purchase probability lower than (its estimate of) C_i will report no intention to buy. By definition, all intenders must be located to the right of their own estimated C_i ; all nonintenders, to the left. If C_i is invariant among households, every intender must have a higher purchase probability than every nonintender; and C_i can be represented by a vertical line, as in Charts 1 and 2. But if C_i varies among households, these cut-off points themselves constitute a distribution, and purchase probability will be lower for some intenders than for some nonintenders.

In Chart 3, C'_1 and C'_2 illustrate the distribution of *purchase* probabilities among households that reported buying intentions when asked questions 1 and 2, on the assumption that cut-off probabilities for these questions are not the same for all households in the sample. The distribution of *cut-off* probabilities for intenders and nonintenders cannot be wholly inferred from the shape of C'_1 or C'_2 . In Chart 3, for example, households with a purchase probability of 0.7 might or might not have reported buying intentions, since some intenders have purchase probabilities as high as 0.7 and so do some nonintenders. Some of these 0.7-probability households must have cut-off probabilities equal to or less than 0.7 (other-

¹² The bias in estimates of s' obtained from s is presumably more serious the greater the possibility of unexpected breakdowns or major repair bills, which in turn are related to the number of households that own the item. Therefore, the s function for widely owned items such as automobiles, ranges, washing machines, etc., ought to deviate more from the true s' function than the s function for less widely owned items such as dishwashers, clothes dryers, etc.

wise none of the group would have reported buying intentions) and some must have cut-off probabilities higher than 0.7 (otherwise all would have reported intentions). But the actual cut-off probabilities among either intenders or nonintenders are unknown.¹³

CHART 3
 Illustrative Distribution of Households
 Having Variable Cut-Off Probability



It is readily demonstrated that any variation in cut-off probabilities will necessarily result in values of r' , hence also r , that are smaller than

¹³ In general, the distribution of cut-off probabilities is likely to show less variance than the distribution of purchase probabilities among intenders. Cut-off probabilities for intenders at the low end of the probability range must be about the same as their purchase probabilities; such households would not have reported intentions to buy if their C_i were above Q , and their Q is so low that C_i cannot be very far below it. On the other hand, the cut-off probabilities for intenders at the high end of the range could be anywhere below their purchase probabilities. It seems intuitively plausible that the distribution of cut-off probabilities among both intenders and nonintenders approximates a normal distribution.

if the cut-off probability were invariant, given the same proportion of the sample reporting buying intentions. In Chart 3, for example, the area under C'_1 is the proportion of households reporting intentions on question 1, and r'_1 is the mean purchase probability for these households. If cut-off probabilities were invariant C'_1 would consist of a vertical line from the horizontal axis to the distribution function, and would then coincide with the function itself. If such were the case and if a given proportion of households reported intentions, r'_1 would necessarily be further to the right than shown. On similar reasoning, s'_1 would be further to the left than it is on Chart 3. It follows that r_i would be larger and s_i smaller if the cut-off probability were invariant, other things equal; consequently, $r_i - s_i$ and the correlation between intentions and purchases would also be larger. Finally, it is evident that the previous assumption of invariance in C_i among households is simply a limiting case of the general proposition that C_i constitutes a distribution—the limiting case being one in which the distribution is peaked to such an extent that it can be represented by a single vertical line with zero variance.¹⁴

If household cut-off probabilities for an intentions question do vary, it becomes much more difficult to make an accurate estimate of the distribution of *ex ante* purchase probabilities from data on the observed purchase rates for intenders and nonintenders. For intentions questions 1 and 2 in Chart 3, the mean probabilities for intenders are r'_1 and r'_2 ; for nonintenders, s'_1 and s'_2 . The mean probability for the entire sample is x' . Clearly, the relations among r'_i , s'_i , p_i , and x' depend not only on the distribution function— $n = f(Q)$ —but also on the distribution of cut-off points among intenders— $C_i = f(D)$ —where D represents whatever variables are associated with the distribution of cut-off probabilities for the i th intentions question. The latter function cannot be observed at all in the empirical data. Hence, it is necessary to make the extreme assump-

¹⁴ It is implicit in this analysis that the distribution of purchase probabilities among intenders—the C'_1 and C'_2 functions in Chart 3—will not overlap for a set of $i \dots j$ intentions questions if either the certainty or the time specifications of the question are altered, but may overlap if both are altered simultaneously in different directions. For example, the distribution of purchase probabilities among a group of households reporting intentions to buy within six months must be wholly contained within a similar distribution among a group reporting intentions to buy within twelve months. Any household reporting an intention to buy within six months must necessarily report that it intends to buy within twelve months unless the questions are misunderstood; and any household reporting that it did not intend to buy within twelve months can hardly report that it intends to buy within six months. Similar reasoning applies to intentions questions about definite plans to buy within X months and possible purchases within X months. On a priori grounds, however, it is not possible to tell whether the distributions would overlap for households "planning to buy within six months" or "definitely planning to buy within twelve months."

BUYING INTENTIONS AND PURCHASE PROBABILITY: I

tion of invariance in the cut-off probabilities in using the (*ex post*) r_i , s_i , p_i , and x variables to estimate the (*ex ante*) distribution function. And if the distribution of cut-off probabilities has greater variance for some intentions questions than for others, as seems likely, estimation of the distribution function from the observed points becomes even more troublesome.

A TEST OF PLAUSIBILITY

The probability model described above can be subjected to several kinds of tests. If (1) the distribution function can be specified and if it can be assumed that (2) regression bias is negligible and that (3) the cut-off probability for any given intentions question is invariant among households, an estimate of the cut-off probability (C_i), and the variance of the distribution (T_i) can be obtained from p_i , r_i , and x .¹⁶ Since the data provide alternative measures of buying intentions ($i \dots j$) and corresponding measures of r and p , it is possible to estimate the respective cut-off points $C_{i \dots j}$ and variances $T_{i \dots j}$. That is, given the distribution function, the observed data (r_1, p_1, x) , (r_2, p_2, x) , etc., will yield estimates of (C_1, T_1) , (C_2, T_2) , etc. If the correct distribution function has been specified, all the $T_{i \dots j}$ will be the same; and the $C_{i \dots j}$ will follow a consistent pattern related to the specifications of the buying intentions questions.

A test of this kind seems to me of limited usefulness. There is good reason to suppose that none of the necessary assumptions hold, and I would judge that the difference between assumption and reality is quantitatively important. A number of functions that fit the data reasonably well could be located. But the "best" function may well be different for one commodity than for another, and I do not see any practical way of deciding which, among several that provide reasonably good fits, is the "true" function.

The basic hypothesis is obviously not contradicted by the data. Thus, it is known from the start that the observed results, in a very general sense, could have been generated by responses that reflect a continuous distribution of purchase probabilities. In Tables 9 and 10 the data have been rearranged to show the distribution of households by two alternative classifications based on the variant questions about buying intentions. Table 9 summarizes the fraction of households (in parentheses), and the proportion of these purchasing, that reported definite, probable-possible, or no intentions to buy specific items. Panel A contains data for each of the three

¹⁶ The variable s_i is redundant to the calculation because it is determined completely by r_i , p_i , and x .

BUYING INTENTIONS AND PURCHASE PROBABILITY: I

available categories (definite, probable-possible, none). The fraction of households in each category corresponds to areas of the distribution function between the respective (but unknown) cut-off probabilities— C_d (minimum probability for definite intenders), C_p (minimum for probable-

TABLE 9
PERCENTAGE DISTRIBUTION OF HOUSEHOLDS BY SPECIFIED CERTAINTY OF THEIR BUYING INTENTIONS, AND THE PROPORTION OF EACH GROUP PURCHASING WITHIN SIX MONTHS OF INTENTIONS SURVEY

	A. All Households			B. Intenders			C. Nonintenders		
	Purchase Probabilities: $\geq C_d$ $< C_d$, $\geq C_p$ $< C_p$, ≥ 0			Purchase Probabilities: $\geq C_d$ $\geq C_p$ ≥ 0			Purchase Probabilities: $< C_d$ $< C_p$ < 0		
Automobile	56.6 (8.1)	27.2 (22.9)	12.1 (69.0)	56.6 (8.1)	34.9 (31.0)	19.1 (100.0)	15.8 (91.9)	12.1 (69.0)	0.0 (0.0)
Furniture	49.3 (13.5)	21.0 (18.4)	12.0 (68.1)	49.3 (13.5)	33.0 (31.9)	18.7 (100.0)	13.9 (86.5)	12.0 (68.1)	0.0 (0.0)
Carpets and rugs	37.6 (5.7)	11.4 (11.4)	5.0 (82.9)	37.6 (5.7)	20.1 (17.1)	7.6 (100.0)	5.8 (94.3)	5.0 (82.9)	0.0 (0.0)
Washing machine	40.9 (3.3)	16.4 (8.6)	5.4 (88.1)	40.9 (3.3)	23.3 (11.9)	7.5 (100.0)	6.3 (96.7)	5.4 (88.1)	0.0 (0.0)
High-fidelity equipment	30.4 (4.3)	9.0 (12.7)	4.8 (83.0)	30.4 (4.3)	14.4 (17.0)	6.5 (100.0)	5.4 (95.7)	4.8 (83.0)	0.0 (0.0)
Refrigerator	50.8 (2.5)	17.0 (8.1)	3.3 (89.4)	50.8 (2.5)	24.9 (10.6)	5.6 (100.0)	4.4 (97.5)	3.3 (89.4)	0.0 (0.0)
Range	47.8 (2.5)	16.1 (5.9)	3.3 (91.6)	47.8 (2.5)	25.7 (8.4)	5.2 (100.0)	4.1 (97.5)	3.3 (91.6)	0.0 (0.0)
Clothes dryer	28.4 (2.8)	14.2 (8.6)	3.0 (88.6)	28.4 (2.8)	17.7 (11.4)	4.6 (100.0)	4.0 (97.2)	3.0 (88.6)	0.0 (0.0)
Television set	19.7 (2.5)	9.7 (9.1)	4.3 (88.4)	19.7 (2.5)	11.8 (11.6)	5.3 (100.0)	4.9 (97.5)	4.3 (88.4)	0.0 (0.0)
Air conditioner	46.3 (2.1)	18.0 (9.7)	3.0 (88.2)	46.3 (2.1)	23.0 (11.8)	5.4 (100.0)	4.5 (97.9)	3.0 (88.2)	0.0 (0.0)
Food freezer	31.7 (1.6)	10.9 (6.7)	1.6 (91.7)	31.7 (1.6)	14.8 (8.3)	2.7 (100.0)	2.2 (98.4)	1.6 (97.7)	0.0 (0.0)
Dishwasher	39.2 (1.9)	9.5 (5.6)	1.4 (92.5)	39.2 (1.9)	17.2 (7.5)	2.6 (100.0)	1.9 (98.1)	1.4 (92.5)	0.0 (0.0)
Garbage dis- posal unit	27.5 (1.5)	9.6 (3.6)	1.6 (94.9)	27.5 (1.5)	14.9 (5.1)	2.3 (100.0)	1.9 (98.5)	1.6 (94.9)	0.0 (0.0)

SOURCE: All data obtained directly from Table 2, except for the middle column in Panel A, which was calculated from columns 3 and 4 of Table 2.

NOTE: Figures in parentheses are proportions of households with specified purchase probabilities; that is, reporting specified intentions. Upper figure in each cell is proportion of these who actually purchased. See accompanying text for explanation of symbols.

possible intenders), and C_n (minimum for nonintenders). C_n of course equals zero, but the other cut-off points are not observable. The proportions purchasing are estimates of mean probability in the distribution function *between* these cut-off points. Panel B shows the proportion of households *above* the various cut-off points and the respective proportions

purchasing; these correspond to estimates of areas under the distribution function and mean probabilities *above* the respective cut-off points, that is, of p_i and r'_i . Panel C contains companion figures for the proportion of households (areas), and the proportion purchasing (mean probabilities) *below* the same cut-off points. These are estimates of $(1 - p)_i$ and s'_i , respectively.

Table 10 follows the same format except that time, rather than explicit certainty specification, is used as the basis for classification. More observations are possible here because households were asked about intentions to buy within six months, within twelve months, "later," or not at all. Hence, panel A has estimates of the area under the distribution function and the corresponding mean probability based on the fraction of households (areas), and the proportions purchasing (mean probabilities), that intended to buy within six months, seven to twelve months, twelve months to later, and not at all. The corresponding estimates of r'_i and s'_i in panels B and C are based on the proportion purchasing among households intending or not intending to buy within six months, twelve months, later, or not at all. The cut-off probability for six-month intenders is designated as C_6 ; for twelve-month intenders, C_{12} ; for later intenders, C_{30} ; and for nonintenders, C_n .¹⁶

The last column of panel B in each table shows the proportion of households with probabilities higher than zero, i.e., the entire sample, and their estimated mean probability, i.e., the observed proportion of the sample

¹⁶ Some of the calculated mean probabilities in panel A of Table 10 are based on reponses from different samples of households. In such cases, sampling variation will occasionally result in logically implausible observations. In Table 10, for example, several of the calculated mean probabilities turn out to be negative.

The calculation of mean probability estimates based on purchase rates from different samples is as follows. Assume that 5 per cent of one sample report that they intend to buy within six months and that 10 per cent of a second sample report that they intend to buy within twelve months. Further assume that 50 per cent of the first sample and 20 per cent of the second actually purchase. The estimated proportion purchasing among those who plan to buy within the period from six to twelve months after the survey data is $[(0.10 \times 0.20) - (0.05 \times 0.50)] \div 0.05 = -0.10$. This negative proportion (minus 10 per cent) is logically absurd, since those who reported intentions to buy within twelve months must consist of those intending to buy within six months plus those intending to buy in the period from seven to twelve months after the survey date. Since 50 per cent of the six-month intenders were assumed to buy, and 5 per cent of this sample were assumed to be six-month intenders, at least 2.5 per cent of the entire sample will buy even if no one but six-month intenders actually purchases. But in the illustration only 20 per cent of the twelve-month intenders (based on a different sample) were assumed to purchase; since 10 per cent of this sample are assumed to be intenders, only 2.0 per cent of the entire sample would appear to have purchased if no one but intenders actually purchases. Thus the arithmetic implies that negative purchases are made by those who would have been seven-to-twelve-month intenders and that their mean probability must therefore have been negative.

TABLE 10
 PERCENTAGE DISTRIBUTION OF HOUSEHOLDS BY INTENDED TIME OF PURCHASE, AND THE PROPORTION OF EACH GROUP PURCHASING
 WITHIN SIX MONTHS OF INTENTIONS SURVEY

	A. All Households—Purchase Probabilities:			B. Intenders—Purchase Probabilities:			C. Nonintenders—Purchase Probabilities:				
	$\geq C_6$	$< C_6, \geq C_{12}$	$< C_{12}, \geq C_{30}$	$\geq C_6$	$\geq C_{12}$	$\geq C_{30}$	$\geq C_6$	$< C_6$	$< C_{12}$	$< C_{30}$	< 0
Automobile	57.7 (10.8)	24.2 (13.6)	15.1 (25.0)	13.0 (50.6)	39.0 (24.4)	26.9 (49.4)	19.6 ^a (100.0)	15.2 (89.2)	13.1 (75.6)	13.0 (50.6)	0.0 (0.0)
Furniture	43.1 (15.7)	25.6 (11.0)	15.0 (11.3)	13.0 (62.0)	35.9 (15.7)	29.7 (38.0)	19.2 (100.0)	15.0 (84.3)	12.8 (73.3)	13.0 (62.0)	0.0 (0.0)
Carpets and rugs	25.8 (7.7)	7.5 (5.6)	10.7 (12.7)	4.6 (74.0)	18.1 (7.7)	14.5 (26.0)	7.2 (100.0)	5.6 (92.3)	5.5 (86.7)	4.6 (74.0)	0.0 (0.0)
Washing machine	37.8 (5.0)	15.8 (4.9)	16.8 (4.8)	4.1 (85.3)	26.9 (9.9)	23.6 (14.7)	7.0 (100.0)	5.4 (95.0)	4.7 (85.3)	4.1 (85.3)	0.0 (0.0)
High-fidelity equipment	34.0 (6.9)	15.3 (7.1)	0.7 (11.6)	4.6 (74.4)	34.0 (6.9)	13.7 (25.6)	7.6 (100.0)	5.0 (93.1)	5.7 (86.0)	4.6 (74.4)	0.0 (0.0)
Refrigerator	38.2 (3.3)	25.2 (2.5)	5.8 (9.3)	3.1 (84.9)	32.6 (3.3)	16.1 (15.1)	5.2 (100.0)	3.9 (96.7)	3.7 (94.2)	3.1 (84.9)	0.0 (0.0)
Range	36.4 (3.2)	16.5 (3.1)	5.3 (6.7)	2.9 (87.0)	36.4 (6.3)	15.6 (13.0)	5.0 (100.0)	3.5 (96.8)	4.3 (93.7)	2.9 (87.0)	0.0 (0.0)
Clothes dryer	33.7 (3.6)	5.5 (3.6)	6.9 (10.0)	2.5 (82.8)	19.6 (7.2)	12.2 (17.2)	4.3 (100.0)	3.1 (96.4)	3.2 (92.8)	2.5 (82.8)	0.0 (0.0)
Television set	24.5 (4.1)	17.6 (5.1)	8.1 (4.6)	4.0 (86.2)	24.5 (4.1)	16.5 (13.8)	5.8 (100.0)	4.9 (95.9)	4.3 (90.8)	4.0 (86.2)	0.0 (0.0)
Air conditioner	20.0 (3.9)	32.4 (3.2)	-4.4 (6.9)	2.2 (86.0)	20.0 (3.9)	10.8 (14.0)	4.1 (100.0)	2.8 (96.1)	3.2 (92.9)	2.2 (86.0)	0.0 (0.0)
Food freezer	28.2 (1.4)	14.0 (3.6)	1.5 (8.3)	1.8 (86.7)	28.2 (1.4)	7.7 (13.3)	2.6 (100.0)	2.0 (98.6)	1.8 (95.0)	1.8 (86.7)	0.0 (0.0)
Dishwasher	27.4 (2.3)	25.2 (2.2)	-1.0 (7.3)	1.3 (88.2)	27.4 (2.3)	9.4 (11.8)	2.8 (100.0)	1.7 (97.7)	2.2 (95.5)	1.3 (88.2)	0.0 (0.0)
Garbage disposal unit	25.4 (2.3)	3.5 (1.4)	1.4 (5.2)	1.4 (91.1)	25.4 (2.3)	7.9 (8.9)	2.1 (100.0)	1.5 (97.7)	1.6 (96.3)	1.4 (91.1)	0.0 (0.0)

SOURCE: Same as Table 9, except that data in the two middle columns of Panel A have been estimated from columns 3 and 4 of Table 2.

NOTE: Figures in parentheses are proportions of households with specified purchase probabilities, that is, reporting intentions to buy within a specified number of months; upper figure in each cell is

proportion of these who actually purchased. See accompanying text for explanation of symbols.

^a This column differs from the comparable one in Table 9 because it was derived from a different sample. The proportion of the sample that purchased each item in this table is an average for samples B and D. The comparable column in Table 9 is based on sample A.

purchasing. The largest proportion purchasing any one item during the six-month period is for automobiles (19.1 per cent). Hence, all of the distribution functions must be skewed even more than the function $n = f(Q)$ shown in Chart 2, since the mean there was about 0.25. The proportion of households in each category of panel A also documents the proposition that all the distribution functions are very highly skewed, and that the vast majority of these households have very low probabilities.

In the discussion of the model, I specified the relation between a set of $i \dots j$ buying intentions questions and the corresponding purchase rates for intenders and nonintenders, given a probability function like the one at the top of Chart 2. To repeat: (1) r_i is always greater than s_i ; (2) r_i and s_i are both positively related to C_i and, hence, negatively related to p_i ; (3) $(r - s)_i$ should be positively related to C_i , negatively to p_i ; (4) the variance of $(r - s)_{i \dots j}$ as a function of $C_{i \dots j}$ or $p_{i \dots j}$ is greater the more highly skewed the distribution. The first two generalizations hold regardless of the shape of the distribution function, and are clearly true for the data in panels B and C of both Tables 9 and 10. The third is also clearly observable. The quantity r falls much more rapidly than s as C_i decreases (and p_i increases); hence, $(r - s)_i$ falls as C_i decreases (and p_i increases). The fourth essentially says that the variance in $(r - s)_{i \dots j}$ should be greater for commodities with relatively small values of x , since the smaller is x' the more highly skewed is the distribution. This relationship is difficult to judge from the data in Tables 9 and 10; but the rank correlations between commodity purchase rates and the variance of the $r_i - s_i$ values are -0.32 and -0.68 for the six- and twelve-month purchase periods, respectively, in Table 9. In Table 10, the comparable rank correlations are, respectively, -0.15 and -0.55 .¹⁷ Thus all of the empirical observations are consistent with predictions of the probability model, although the model is specified in very general terms.

An analysis complementary to the above is contained in an appendix to this chapter; here some cumulative functions based on Chart 1 are developed. These functions are the precise *ex ante* counterparts to the observed r_i , s_i , p_i , and x_i points; observed data for several commodities are plotted and illustrative functions are fitted to the observed points. The appendix tests are necessarily impressionistic, as are those just described; both essentially document the proposition that the existence of a probability distribution is empirically plausible.

¹⁷ Strictly speaking the comparison of variance in $r_i - s_i$ with x is valid only if households assign the same cut-off point, given the intentions question, to all commodities; data in Chapter 4 suggest that this is unlikely to be the case.

A TEST OF AN ALTERNATIVE MODEL

I turn now to some quite different (and rather indirect) tests of the hypothesis. Table 9 contains data for intentions variants that differ as to certainty dimension (definite intenders, probable-possible intenders, etc.), but each variant has the same planning period. Table 10 contains data for intentions variants that differ as to length of the planning period (six-month intenders, twelve-month intenders, etc.) but have the same certainty dimension. Each table shows the fraction of households reporting intentions to buy for the $i \dots j$ intentions variants and the fraction that purchased within six months of the survey date. What is the most fruitful interpretation of responses to survey questions that differ as to planning period but not as to certainty dimension? Are such statements essentially judgments about the probable *timing* of purchases or do they reflect a judgment about purchase probability relating to *any* future period of time? In the analysis above, these data were treated as if they were probability statements. It seems more natural to interpret these data as judgments about probable timing. The available data permit a test of these two interpretations. The results, in my view, provide strong support for the proposition that households do seem to make the kind of probability judgment that is the central theme of this chapter.¹⁸

The alternative interpretations can be illustrated by the intentions data for groups B and D. Households in group B were asked whether they "intended to buy within the next six months"; D, whether they "intended to buy within the next twelve months." More households reported buying intentions in group D than B; as must be the case. In fact, twelve-month intenders can usefully be thought of as comprising two subgroups—those who would, if asked, have reported that they intended to buy within six months and those who would not have been willing to make this judgment. Those reporting intentions to buy within six months are designated as p_6 ; those reporting intentions to buy within a year, p_{12} ; and $p_{12} - p_6$ is designated as p_{6-12} (obviously $p_6 + p_{6-12} \equiv p_{12}$).

What are the empirical implications of the two interpretations? If statements about intentions to buy within varying time periods reflect probable timing—*when* the household expects to make a particular

¹⁸ It should be noted that the probability model does not depend on whether or not statements about intentions to buy within different forward periods actually represent judgments about the probability of purchase, unrelated to the timing of purchases. Nothing in the model requires that a probability interpretation be placed on responses to these particular questions, although the model gains in empirical plausibility if a probability interpretation turns out to be consistent with the empirical results.

purchase—the p_6 group ought to show a relatively heavy purchase rate in the six months subsequent to the survey (Period I) and a relatively lower purchase rate during the period from six to twelve months after the survey (Period II). The Period II purchase rate for the p_6 group would be zero if these statements were interpreted as reflecting perfect certainty about timing, but this extreme assumption is not necessary. The timing interpretation requires only that the purchase rate be lower in Period II than in period I for the p_6 group. Households in the p_{6-12} group, on the other hand, are saying that they expect to buy in Period II rather than in Period I. Some may end up buying in Period I, but the purchase rate in II will surely be higher.

The probability interpretation essentially says that the p_6 group has a higher cut-off and mean probability than the p_{6-12} group. The p_{12} group must have the same cut-off point as the p_{6-12} group, and its mean probability must be a weighted average of the means for its two subgroups. The probability model can be interpreted as specifying that these relations be independent of the forecast period. In short, one (plausible) version of the probability interpretation suggests that the only difference between six- and twelve-month intenders is that the former are more likely to purchase during any arbitrarily specified period. But in that case purchase rates ought to be higher for the p_6 than for the p_{6-12} group in Period I and, also, in Periods II, III, . . . , n .

All told there are four sets of intentions questions that permit comparisons of the sort just described. The B sample was asked about plans to buy within six months. The A sample was asked about definite plans to buy within twelve months and also about probable or possible purchases within twelve months. Both the C and D samples were asked about plans to buy within twelve months. Hence, the purchase rate of six-month planners— r_6 —can be compared with rates for four groups of twelve-month planners. The timing interpretation predicts that the six-month intenders will have a higher purchase rate during Period I and a lower purchase rate during Period II than any of the four groups of twelve-month intenders. That is, for Period I

$$r_6 > r_{12} > r_{6-12};$$

and for Period II

$$r_6 < r_{12} < r_{6-12}$$

The probability interpretation, on the other hand, predicts that for intentions questions with the same certainty dimension purchase probability is inversely correlated with the planning period, i.e., those planning

BUYING INTENTIONS AND PURCHASE PROBABILITY: I

to buy within n months have a higher purchase probability than those planning to buy within $n + 1$ months, irrespective of the forecast period. That is, for both periods, I and II,

$$r_6 > r_{12} > r_{6-12}$$

For intentions questions with the same planning period, the probability interpretation argues that the more restrictive the certainty dimension the higher the purchase probability of the group reporting intentions. For intentions questions that vary both as to planning period and certainty dimension, the probability model can be interpreted as requiring that the lower the fraction of households reporting intentions, the higher

TABLE 11
PREDICTED DIFFERENCES IN INTENDERS' PURCHASE RATES FOR PERIODS I AND II

<i>Period I</i>		<i>Period II</i>	
Probability Interpretation	Timing Interpretation	Probability Interpretation	Timing Interpretation
$B_1 < A_1$	$B_1 > A_1$	$B_1 < A_1$	$B_1 < A_1$
$B_1 > A_1 + A_2$	$B_1 > A_1 + A_2$	$B_1 > A_1 + A_2$	$B_1 < A_1 + A_2$
$B_1 > C_1$	$B_1 > C_1$	$B_1 > C_1$	$B_1 < C_1$
$B_1 > D_1$	$B_1 > D_1$	$B_1 > D_1$	$B_1 < D_1$

NOTE: The intentions questions are as follows:

- A_1 Definite intentions to buy within twelve months
- $A_1 + A_2$ Definite, probable, or possible intentions to buy within twelve months
- C_1 Intentions to buy within twelve months (if income is as expected)
- D_1 Intentions to buy within twelve months
- B_1 Intentions to buy within six months

the purchase probability of intenders. Since all these relations hold for any forecast period, the intentions variants ought to show the same pattern for Periods I, II, III, etc.

Given the characteristics of the six-month intentions question and the four twelve-month questions, the probability interpretation predicts that definite twelve-month intenders will buy more than six-month intenders during all forecast periods because the fraction of intenders is lower for the definite twelve-month variant. The probability interpretation predicts a priori (compare the questions) that six-month intenders will purchase more than the other three groups of twelve-month intenders in both periods; in addition, the fraction of intenders is lower for the six-month group than for these three twelve-month groups.

The survey data yield four pairs of observed purchase rates, for each of thirteen commodities, in which the probability and timing interpretations

BUYING INTENTIONS AND PURCHASE PROBABILITY: I

predict a different relation between purchase rates for six- and twelve-month intenders. One of the pairs involves Period I purchase rates; the other three, Period II purchase rates. In the remaining four comparisons the two interpretations predict the same outcome. The possible

TABLE 12
TEST OF PROBABILITY AND TIMING INTERPRETATIONS OF CONSUMERS UNION
SURVEY DATA
(proportions of intenders who purchased)

	<i>Intentions Specified by Variant</i>				
	A ₁	B ₁	A ₁ or A ₂	C ₁	D ₁
PERIOD I: WITHIN SIX MONTHS OF SURVEY DATE					
Automobile	56.6	57.7	34.9	46.1	39.0
Air conditioner	46.3	20.1	23.0	25.8	25.6
Clothes dryer	28.4	33.7	17.7	22.2	19.6
Dishwasher	39.2	27.4	17.2	22.9	26.3
Food freezer	31.7	28.2	14.8	23.9	18.0
Range	47.8	36.4	25.7	31.1	26.6
Refrigerator	50.8	38.2	24.9	30.1	32.6
Television set	19.7	24.5	11.8	26.8	20.7
Washing machine	40.9	37.8	23.3	34.5	26.9
Carpets and rugs	37.6	25.8	20.1	25.5	18.1
Furniture	49.3	43.1	33.0	40.5	35.9
Garbage disposal unit	27.5	25.4	14.9	25.3	17.1
High-fidelity equipment	30.4	34.0	14.4	22.1	24.5
PERIOD II: WITHIN SEVEN TO TWELVE MONTHS OF SURVEY DATE					
Automobile	18.9	15.0	20.4	19.6	19.7
Air conditioner	1.9	1.0	2.6	2.3	1.9
Clothes dryer	20.3	11.2	13.0	12.2	12.7
Dishwasher	13.7	17.7	9.1	11.5	16.2
Food freezer	9.8	5.1	7.9	9.1	10.8
Range	9.0	20.4	10.8	15.2	18.7
Refrigerator	9.2	14.6	7.6	14.2	14.8
Television set	21.2	26.4	17.4	18.6	18.2
Washing machine	8.0	13.3	14.1	11.7	17.2
Carpets and rugs	16.8	19.6	15.6	11.7	16.4
Furniture	16.0	17.9	15.9	16.8	18.0
Garbage disposal unit	12.5	23.8	10.4	10.7	14.6
High-fidelity equipment	17.9	18.6	16.6	21.0	15.5

SOURCE: Period I data from column 9 of Table 2. Period II data are the respective differences between columns 11 and 9 of Table 2.

comparisons are summarized in Table 11, along with the outcomes predicted by the two interpretations; the observed purchase rates are summarized in Table 12.

The evidence generally supports the probability interpretation of responses to questions about intentions to buy within different time periods. Of the fifty-two differences that discriminate between the

alternative interpretations, seventeen are so small (two percentage points or less) that they can be ignored; of the remaining thirty-five, twenty-four are consistent with the probability interpretation, hence inconsistent with the timing interpretation. And of the differences that are statistically more reliable, ten of thirteen significant at the 0.10 level and six of the seven differences significant at the 0.05 level¹⁹ are consistent with the probability hypothesis.

Intentions Data and Purchase Predictions

The accuracy of purchase predictions based on surveys of buying intentions evidently depends on the time series correlation between p , the proportion of intenders in the population, and x , the proportion of purchasers in the population. It has been demonstrated by Arthur Okun that p and x will be positively correlated over time if the purchase rate of intenders exceeds that of nonintenders, that is, if r exceeds s ; more precisely, Okun shows that p and x will be positively correlated if \bar{R} exceeds \bar{S} , where the latter variables are the respective means of the random variables r and s . The time series correlation between x and p turns out to be positively associated with the algebraic difference between \bar{R} and \bar{S} and with the time series variance in p .²⁰ These relations hold completely provided that the time series movements in both r and s are independent of those in p ; otherwise, the relation may be weaker or stronger than indicated by the above analysis.²¹

¹⁹ All deviant cases where the differences are fairly reliable involve purchase rates for automobiles. The majority of these purchases (in this sample) consist of new cars. Since Period I extends from April 1958 to October 1958 and Period II from October 1958 to March 1959, April planners waiting for new models would tend to show up as twelve-month planners and Period II purchasers; hence, buying intentions for new automobiles may have a strong element of probable timing.

²⁰ Formally, Okun shows that

$$M_{xp} = (\bar{r} - \bar{s})M_{pp} + \bar{p}M_{rp} + (1 - \bar{p})M_{sp} + M_{rpp} - M_{spp},$$

where M designates the respective variances or covariances. All the terms to the right of $(\bar{r} - \bar{s})M_{pp}$ are zero if both r and s are independent of p . If such is the case the estimated slope of the time series regression of x on p (M_{xp}/M_{pp}) is $(\bar{r} - \bar{s})$. In addition, the squared time series correlation between x and p turns out to be:

$$r^2_{xp} = (\bar{r} - \bar{s})^2 \frac{M_{pp}}{M_{xx}}$$

Thus, the size of the time series correlation between x and p depends on the size of $\bar{r} - \bar{s}$ and of M_{pp} and is positively associated with both.

²¹ For example, if p were to be positively correlated with r and/or s , the time series correlation between p and x would be larger than indicated by the $r - s$ difference; if the time series relation of p with r and/or s happens to be negative, the correlation between p and x will be smaller than suggested by the $r - s$ difference. To illustrate

BUYING INTENTIONS AND PURCHASE PROBABILITY: I

It is intuitively obvious, and easily demonstrated, that if both r and s are completely invariant over time, p will be a perfect linear predictor of x . Start with the definitional equation,

$$x = p(r - s) + s$$

If both r and s are wholly invariant, they can be replaced by the constants k and K , respectively. Then the above equation can be written $x = K + kp$, and x is a completely determined linear function of p with intercept of $K (= s)$ and slope of $k (= r - s)$. To the extent that either purchase rate (r or s) varies, either randomly or systematically, p will be an imperfect predictor of x .

THE IMPORTANCE OF NONINTENDERS FOR PREDICTION

Given that r and s are not, and should not be expected to be, invariant over time, what can be inferred about the predictive ability of alternative intentions surveys? First, the proportion of total purchases accounted for by intenders appears to be highly relevant in determining which intentions surveys or questions are likely to be better time series predictors of purchases. Intentions questions characterized by small proportions of intenders relative to purchasers, hence, by low ratios of purchases by intenders to total purchases, will necessarily yield poor predictions of population purchases unless changes in the proportion of intenders (p) happen to be correlated with changes in the purchase rate of nonintenders (s). Given the definitional relation among these variables, intentions questions with relatively high cut-off probabilities will necessarily be associated with relatively small values of p , and it has been shown above that such questions will have relatively large values of r , s , and $r - s$, and values of s that are very close to x ; but in that event most of the time series variance in x is a consequence not of changes in $(r - s)p$, the first term on the right-hand side of the definitional equation, but of changes in s , the second term. Whether p is an accurate predictor of x thus will depend largely on the correlation between p and s ; and there is no a priori reason

the common sense interpretation of these statements, assume that p and s are negatively correlated over time. If so, when p increases the product $s(1 - p)$ will decrease. If the latter falls, it is quite possible that x will also fall, since by definition,

$$x = rp + s(1 - p),$$

and the second term on the right-hand side of this identity is generally much larger than the first term. But in that case, an increase in buying intentions (p) would be associated with a fall in purchases (x), and the time series correlation between p and x would be negative.

to suppose that these two variables (the proportion of intenders in the population and the purchase rate of nonintenders) will be strongly related. In effect, an intentions question with a relatively high probability cut-off is likely to be a very good time series predictor of purchases by *intenders* (because the purchase rate among intenders, r , is likely to be comparatively stable over time), but a poor predictor of total purchases: most purchases are made by nonintenders; therefore, most of the time series variance in x , the population purchase rate, is a consequence of changes in s , the purchase rate of nonintenders; and p , the proportion of intenders in the population, may be either not closely related or unrelated to s .

On the other hand, an intentions question with a relatively low probability cut-off will necessarily be characterized by a relatively high value of p . Such questions have been observed to have relatively low values of r , s , and $r - s$, and values of s that are substantially below x . In this case p may be both a worse predictor of purchases by intenders (because r is likely to be less stable over time than in the previous case) and a better predictor of total purchases. In sum, time series changes in s are likely to dominate time series changes in x whenever p is small relative to x , simply because a large part of the variation in x will then be a direct consequence of variation in s ; as a consequence, the time series correlation between p and x will depend largely on the size of the correlation between p and s . If p is not so small, less of the variation in x will be a direct consequence of variation in s , and the time series correlation between p and x is less heavily dependent on the p, s correlation.²²

PREDICTION OF DIFFERENCES IN COMMODITY PURCHASE RATES

Table 2 contained estimates of the proportion of total purchases made by intenders for each of the alternative intentions questions. The estimates,

²² Another way of looking at this proposition is to note that if (1) r is essentially invariant over time, and (2) r is so large relative to s that $r - s$ can also be taken as invariant, time series changes in s will tend to show up as forecast errors in predicting x from p . Since

$$x = p(r - s) + s,$$

a linear regression of x on p will yield

$$x = a + bp + u,$$

where a is an estimate of \bar{s} , b of $(\bar{r} - \bar{s})$, and u is the error term. If p tends to be quite small because of the characteristics of the intentions question and if s and p are independent, the nonintender purchase rate during any specific period— s_t —is approximately $\bar{s} + u_j$. Moreover, when p is small in absolute terms it will also tend to have little time series variance. Hence, the approximate variance of x will be u , and the correlation between x and p will be very small. On the same reasoning, the higher the level of p , the larger its variance, the smaller the error variance, and the larger the correlation between p and x , other things equal.

BUYING INTENTIONS AND PURCHASE PROBABILITY: I

TABLE 13

PROPORTION OF TOTAL PURCHASES BY INTENDERS RELATED TO PURCHASE RATES FOR INTENDERS, NONINTENDERS, AND TOTAL SAMPLE

	<i>Intentions Question A₁</i> (definitely intend to buy)					<i>Intentions Questions A₁ or A₂</i> (definitely, probably, possibly will buy)				
	Intenders' Purchases as Per Cent of Total	<i>p</i>	<i>x</i>	<i>r</i>	<i>s</i>	Intenders' Purchases as Per Cent of Total	<i>p</i>	<i>x</i>	<i>r</i>	<i>s</i>
PURCHASES WITHIN SIX MONTHS										
Automobile	24	.081	.191	.566	.158	57	.310	.191	.349	.121
Furniture	36	.135	.187	.493	.139	56	.319	.187	.330	.120
Carpets and rugs	28	.057	.076	.376	.058	45	.171	.076	.201	.050
High-fidelity equipment	20	.043	.065	.304	.054	38	.170	.065	.144	.048
Range	23	.025	.052	.478	.041	42	.084	.052	.257	.033
Refrigerator	23	.025	.056	.508	.044	47	.106	.056	.249	.033
Washing machine	18	.033	.075	.409	.063	37	.119	.075	.233	.054
Television set	9	.025	.053	.197	.049	26	.116	.053	.118	.043
Air conditioner	18	.021	.054	.463	.045	50	.118	.054	.230	.030
Clothes dryer	17	.028	.046	.284	.040	44	.114	.046	.177	.030
Dishwasher	29	.019	.026	.392	.019	50	.075	.026	.172	.014
Food freezer	19	.016	.027	.317	.022	42	.083	.027	.148	.016
Garbage disposal unit	18	.015	.023	.275	.019	33	.051	.023	.149	.016
Squared correlation coefficients		$r^2_{px} = .847, r^2_{sx} \sim 1.0$					$r^2_{px} = .949, r^2_{sx} \sim 1.0$			
PURCHASES WITHIN TWELVE MONTHS										
Automobile	18	.081	.332	.775	.295	52	.310	.332	.553	.234
Furniture	28	.135	.319	.653	.267	49	.319	.319	.489	.239
Carpets and rugs	22	.057	.141	.544	.116	43	.171	.141	.357	.096
High-fidelity equipment	15	.043	.139	.483	.123	38	.170	.139	.310	.102
Range	16	.025	.089	.568	.077	34	.084	.089	.365	.064
Refrigerator	18	.025	.083	.600	.070	42	.106	.083	.325	.055
Washing machine	13	.033	.124	.489	.111	36	.119	.124	.374	.091
Television set	10	.025	.109	.409	.101	31	.116	.109	.292	.083
Air conditioner	16	.021	.063	.482	.054	48	.118	.063	.256	.036
Clothes dryer	17	.028	.081	.487	.071	43	.114	.081	.307	.054
Dishwasher	20	.019	.050	.529	.041	39	.075	.050	.263	.033
Food freezer	15	.016	.043	.415	.037	44	.083	.043	.227	.027
Garbage disposal unit	13	.015	.045	.400	.039	29	.051	.045	.253	.034
Squared correlation coefficients		$r^2_{px} = .852, r^2_{sx} \sim 1.0$					$r^2_{px} = .955, r^2_{sx} \sim 1.0$			

SOURCE: Table 2. See accompanying text for explanation of symbols.

for two of these intentions questions and all thirteen commodities, are shown in Table 13, along with purchase rates for intenders, nonintenders, and the sample as a whole. Although time series data on *p* and *x* for the alternative intentions questions are not available in the CU data, comparisons among commodities with different purchase rates constitute an interesting proxy.

BUYING INTENTIONS AND PURCHASE PROBABILITY: I

For both six- and twelve-month purchase periods, the correlation between p and x is noticeably stronger for the intentions question characterized by a relatively high ratio of p to x —hence, also, a high ratio of purchases by intenders to total purchases; the correlation between s and x approximates unity.²³ Data for the other intentions questions are also consistent with the generalization that the (p, x) correlation is stronger for questions with relatively low cut-off probabilities, with the exception of the two questions that have extremely low cut-off points— C_2 and (B_1 or B_2). The tabulation, below, of r^2_{px} values computed from the same kind of data as in Table 13, documents this proposition.²⁴

<i>Intentions Question</i>	<i>r²_{px} Among Commodities for Purchases Within</i>	
	6 Months	12 Months
A ₁	.847	.852
B ₁	.876	.876
C ₁	.926	.932
D ₁	.956	.965
A ₁ or A ₂	.949	.955

The evidence in Table 13 illuminates one of the major difficulties in using the current buying intentions surveys to predict purchases of durables. Those who report that they “definitely will buy during the next twelve months” account for only between 20 and 30 per cent of total purchases during the six months following the survey, and for an even smaller proportion during the twelve months following. Even those who report that they “definitely, probably, or possibly will buy during the next twelve months” generally account for only 40 to 50 per cent of total purchases during the six months following the survey and, again, for a smaller proportion during the twelve months. Clearly, the majority of durable goods purchases are made by households that are generally classified as nonintenders by consumer surveys. As a consequence, the accuracy of purchase predictions based on intentions surveys depends

²³ All these correlations have a strong upward bias because s is a major component of x , and x has an extremely large variance among commodities. More precisely, the correlations have a strong upward bias as a measure of the success that would be enjoyed in making short-run predictions of x from p . An appropriate analogy is that, although the correlation between income and consumption over the period 1900 to 1960 must be close to 1.0, predictions of the change in consumption between two adjacent years are likely to be quite poor if based solely on income. Since consumption is a large part of income, any substantial variance in consumption must be strongly associated with a corresponding variance in income.

²⁴ This proposition does not depend on the comparative “efficiencies” of the intentions question, as the term is used in Chapter 2. It is clear, for example, the C_1 is a more efficient question than A_1 or A_2 combined in terms of its ability to discriminate between buyers and nonbuyers of particular commodities; but the (p, x) correlation among commodities is somewhat stronger for the combination of A_1 or A_2 .

largely on whether or not changes in the proportion of intenders can successfully predict changes in the purchase rate of nonintenders, i.e., on whether or not p and s are strongly correlated over time.

From the data in Table 13 it appears that the probability cut-off point for definite, probable, or possible intenders is quite low, ranging from somewhere around 0.25 for automobiles to about 0.10 for garbage disposal units.²⁵ One possible conclusion is that estimates of the frequency distribution in the low-probability region are at least as critical as estimates of the distribution in the high-probability region, since at least half of the total purchase probabilities (hence, expected purchases) are contained in the former region.²⁶ It is also possible, however, that the low-probability region really contains a very small part of the total *ex ante* probabilities (hence, of expected purchases) and that these data give a misleading impression.

If the cut-off probabilities for any given intentions question are highly variable among households, a substantial number of relatively high-probability households will be classified as nonintenders. Since an estimate of the relative importance of low-probability households depends on the proportions and observed purchase rates of intenders and nonintenders, misclassification of households due to the inefficiency of the survey question could account for the relatively high proportion of purchases made by nonintenders. Alternatively, contingencies may be quantitatively important enough to produce this result. As already noted, my estimates are based on the observed r_i and s_i points. But r_i is too low as an estimate of r'_i (the mean probability in the distribution above C_i), and s_i is too high as an estimate of s'_i (the mean probability in the region

²⁵ The cut-off probability, C_i , must be lower than r'_i , the mean probability in the segment of the distribution located above C_i ; how much lower depends on the shape of the distribution. No direct estimates of either C_i or r'_i are available, but r_i , the proportion of intenders who purchase, constitutes a downwardly biased estimate of r'_i . It does not seem unreasonable to assume that

$$C_i = r_i - 0.25r_i,$$

which is the basis for the statement in the text.

²⁶ Part of the data in Table 13 forms the empirical basis for estimating the cumulative expected purchases function—the middle panel of Chart 6—developed in the appendix to this chapter. The above data indicate that 19.1 per cent of the sample purchased an automobile, 8.1 per cent reported definite intentions to buy, and 31.0 reported either definite, probable, or possible intentions to buy. Further, 56.6 per cent of the definite intenders purchased, accounting for 24 per cent of total purchases; and 34.9 per cent of the definite-probable-possible intenders purchased, accounting for 57 per cent of total purchases. If the cut-off probabilities are assumed to be equal to $r_i - 0.25r_i$ —see note 25, above—observed points on the appendix function 2.2 are: $C_i = 0.42$, $Br/Nr = .081 \times .566 = .046$; $C_i = 0.26$; $Br/Nr = .310 \times .349 = .108$; and $C_i = 0$, $Br/Nr = .191$.

below C_i). Thus, if the bias between *ex ante* and *ex post* values is sufficiently powerful, a similar analysis based on r'_i and s'_i might yield substantially different results.

IMPLICATIONS OF THE PROBABILITY MODEL FOR PREDICTION

For predicting future purchases of durables the probability model simply says that the best estimate of x , the proportion of households that will purchase during the forecast period, is x' , the mean of the distribution function. All current surveys of buying intentions are designed to provide an estimate of the proportion of households in the high-probability segment of the distribution above some unknown but presumably stable cut-off point C_i . Even if the C implicit in the language of the intentions question is in fact invariant over time—and the evidence suggests that it is²⁷—the mean of the distribution (x'_i) need not have a simple or fixed relation to p_i , the proportion of the population with purchase probabilities above C_i . Certain kinds of distributions do have the property that p is a unique function of x' ; most do not.

These relations are worth examining more carefully, since they lie at the root of the forecasting problem. In Chart 4 two distribution functions are shown, one relating to period $t_0-n_0 = f_0(Q)$ —and the other relating to period $t_1-n_1 = f_1(Q)$. If a survey of buying intentions is taken in t_0 , a fraction p_0 will report intentions, a fraction $1 - p_0$ that they do not intend to buy. The buying intentions question is assumed to have an invariant (among households) cut-off probability of C_0 , so that

$$p_0 = \left[\int_{C_0}^1 f_0(Q) dq \right] \div \left[\int_0^1 f_0(Q) dq \right]$$

$$1 - p_0 = \left[\int_0^{C_0} f_0(Q) dq \right] \div \left[\int_0^1 f_0(Q) dq \right]$$

A similar survey taken in t_1 , using the same intentions question, will show different results if the distribution of households is assumed to have shifted. If the cut-off probability of the intentions question is also invariant over time so that $C_0 = C_1$, p_1 intenders and $1 - p_1$ nonintenders will be observed, where

$$p_1 = \left[\int_{C_0}^1 f_1(Q) dq \right] \div \left[\int_0^1 f_1(Q) dq \right]$$

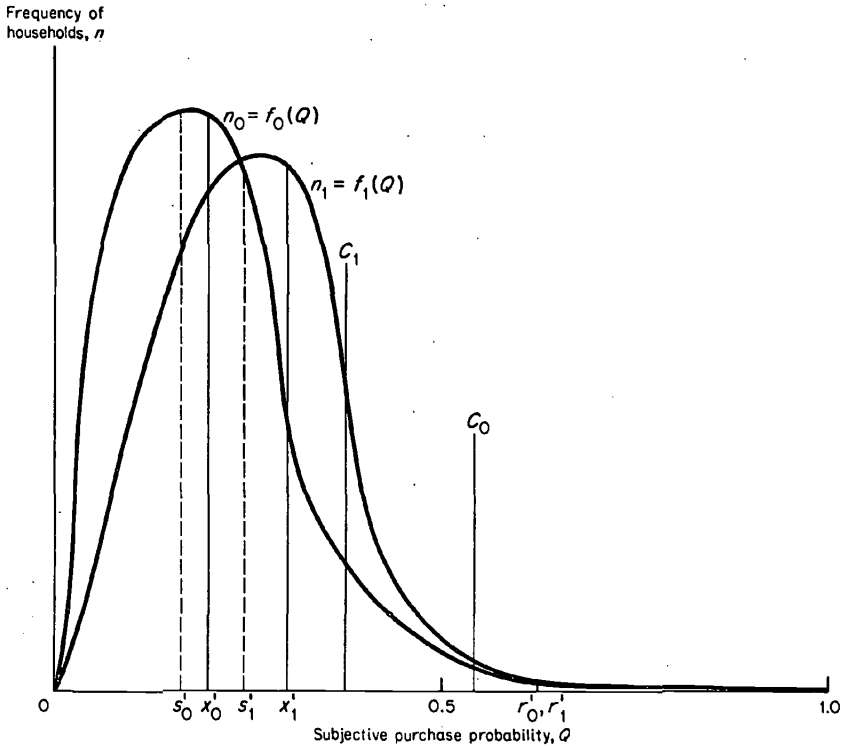
$$1 - p_1 = \left[\int_0^{C_0} f_1(Q) dq \right] \div \left[\int_0^1 f_1(Q) dq \right]$$

²⁷ Reinterview data from the Federal Reserve Board-Census Bureau survey indicates very little time series variance in the purchase rate of households reporting intentions to buy within six months. There is also no apparent correlation between the purchase rate for intenders and the population purchase rate.

BUYING INTENTIONS AND PURCHASE PROBABILITY: I

As the functions in Chart 4 are drawn, p_1 is somewhat higher than p_0 , although the percentage change p_1/p_0 appears to be quite small. However, the mean of the t_0 distribution (x'_0) appears to differ from the t_1 mean (x'_1) by something like 50 per cent, that is, $x'_1/x'_0 \sim 1.50$. In order to forecast the change in purchases it is necessary to predict the

CHART 4
Illustrative Time Series Changes
in Probability Distribution



ratio x_1/x_0 . The best estimate of this ratio is clearly x'_1/x'_0 ; p_1/p_0 may turn out to be the same or very close to x'_1/x'_0 ; but it may also turn out to be quite different.

The relationship between p_1/p_0 and x'_1/x'_0 is easily shown to depend on the level, as well as the change, of mean probabilities for intenders (r'_0, r'_1) and nonintenders (s'_0, s'_1). As I have noted before, mean probability in the sample as a whole is a weighted average of the means for

intenders and nonintenders, since

$$\begin{aligned}x'_0 &= r'_0 p_0 + s'_0(1 - p_0) \\x'_1 &= r'_1 p_1 + s'_1(1 - p_1)\end{aligned}$$

Hence the change in mean probability between t_0 and t_1 is

$$\frac{x'_1}{x'_0} = \frac{r'_1 p_1 + s'_1(1 - p_1)}{r'_0 p_0 + s'_0(1 - p_0)}$$

The ratio p_1/p_0 will equal x'_1/x'_0 only if $r'_1 = r'_0$ and if $s'_0 = s'_1 = 0$; otherwise, the two ratios will be the same only by chance. The point is not that the ratio p_1/p_0 will necessarily yield a bad forecast of x'_1/x'_0 , but that there is no obvious reason why it should yield a good forecast. From Chart 4 it is abundantly clear that one could construct distribution functions for periods t_0 and t_1 , with means of x'_0 and x'_1 , and either a very high ratio of p_1/p_0 or a very low one. Further, one might get quite different results simply by varying the intentions question; in Chart 4 a question with a lower cut-off probability than C_0 —say, C_1 —would be a better predictor because the new p_1/p_0 ratio would be higher than the one shown above. But a question with a higher cut-off probability—say, C_2 —would apparently give worse results.

The above argument can be stated more succinctly in general form. Prediction of durable goods purchases from consumer surveys requires that the survey yield an estimate of x' , the mean purchase probability in the population. If x' is a linear function of p_i , an estimate of p_i will be a perfect substitute for x' in linear regressions. If x' is a unique but non-linear function of p_i , an estimate of p_i will be a perfect substitute for x' provided that the functional relation between p_i and x' can be established. But if x' cannot be determined solely from p_i , the latter will be an inferior substitute, one that cannot reliably predict purchases. One final point is worth noting. All surveys are necessarily limited in sample size. The problem of sampling error is maximized for a survey that concentrates on measuring p , especially if, as is typically the case, intentions questions with relatively high cut-off probabilities are used.

In addition to these analytical considerations, the question of stability over time in the distribution function is of considerable importance. It may well be true that the proportion of households with probabilities above some relatively high value, say 0.50, is much less variable over time than the proportion with probabilities between 0.10 and 0.20.²⁸ Since a

²⁸ Very little empirical evidence is available on this point. One interesting scrap of data (unpublished, from a confidential market research study) suggests that, for the commodity under study, the entire shift in the probability distribution between two

large majority of households appear to have quite low purchase probabilities for most items, variations in the number of households in the probability region below 0.25 may be a more important source of changes in population mean probability than variations in the number of households in the region above 0.50 or 0.75.

An Exploratory Time Series Test

For the most part, the propositions just discussed cannot be tested directly with existing time series data on consumer intentions to buy. Empirical evidence does suggest that a relatively large share of the time series variance in purchases is due to variation over time in the purchases of nonintenders, rather than intenders. There is no survey evidence that bears directly on the predictability of these critically important movements in nonintenders' purchase rates. But the probability hypothesis suggests a possible proxy for nonintender purchase rates—more precisely, for the variation in mean *ex-ante* purchase probability among nonintenders. Empirically, the proxy variable performs extremely well; it explains a substantial part of the variance in over-all purchase rates that cannot be accounted for by changes in the proportion of intenders. I regard these results as suggestive but inconclusive, for two reasons. First, there are only eleven time series observations on which the relevant test can be run. Second, and more important, quantitative assumptions are necessary to conduct the test; while the precise assumptions used below are plausible and consistent with the probability hypothesis, other equally plausible and consistent quantitative assumptions could also have been made—with results considerably less striking than the ones presented below.

time periods took place in what must have been a very low probability range. During the same two periods the purchase rate changed by some 20 per cent. The pattern was roughly as follows:

<i>Proportion of Sample Reporting Intentions to Buy</i>	<i>Period 1</i>	<i>Period 2</i>
Within a year	2%	2%
Within a few years	4	4
Interested, no plans	10	14
Not interested	84	80

Here the only shift was between households "not interested," presumably with probabilities of zero or close to it, and those "interested, no plans," presumably with probabilities somewhat higher than zero but very low in absolute terms. A very crude calculation indicates that mean probability in the latter group is quite likely to have been less than 0.10, assuming that the observed change in purchases had been due solely to the increased frequency in that group.

BUYING INTENTIONS AND PURCHASE PROBABILITY: I

The basic data are the surveys of consumer intentions to buy automobiles, conducted since January 1959 by the U.S. Bureau of the Census. Other intentions surveys cover a longer time span but the detailed classifications of intenders needed to implement the test would require retabulation and possibly even recoding of the survey schedules. For data on automobile purchases, I rely on Ward's Automotive Reports. An alternative purchase series is available—the Census survey of buying intentions also asks about past purchases—but the Ward's series seems to be regarded as the more accurate estimate of automobile sales to consumer units; in addition, use of the Ward's data insures that the intentions and purchases series are completely independent.

Because the intentions data have a strong seasonal pattern, I rely exclusively on comparisons of change from year to year.²⁹ The variable to be predicted is the change in the proportion of households purchasing a new automobile during successive twelve-month purchase periods. The explanatory variables are the change in the (weighted) proportion of households reporting intentions to buy a car at the beginning of each purchase period, the corresponding change in the (weighted) proportion of households reporting that they "don't know" or are "uncertain" about car buying intentions, and the change in the (weighted) proportion of nonintenders adjusted by an index of change in the proportion of "don't know" households.

The weights are essentially estimates of mean purchase probability for the various classes of households. Those reporting that they definitely intend to buy a new car within six months are assigned a weight of 0.7—roughly, the observed purchase rate among such households during the twelve months subsequent to an intentions survey. Those reporting that they might buy a new car within six months are assigned a weight of 0.5, those reporting that they intend to (definitely will, probably will, or might) buy a new car within twelve months (but not within six months) are assigned a weight of 0.3, and those reporting that they intend to buy a *used* car within twelve months, a weight of 0.2. Summing up the weighted proportions of households reporting these various types of intentions yields

²⁹ The seasonal pattern in car buying intentions does not match the seasonal in consumer purchases of new automobiles. For the Census data, intentions to buy new cars are much higher in the October survey than in any of the other (January, April, and July) surveys; intentions to buy new cars may also be seasonally high in January. Purchases are at their peak in the spring. Of course, the seasonal in intentions should lead that in purchases.

Attempts to make a seasonal adjustment of intentions based on the limited number of available observations have convinced me that the risks of so doing are greater than those in comparisons of year-to-year change.

BUYING INTENTIONS AND PURCHASE PROBABILITY: I

a series designated as p' . Households reporting either that they "don't know" about automobile purchases within six months or within twelve months are each assigned a weight of 0.1; this series is designated as d . The two "don't know" categories have some overlap, but it is by no means complete. For example, a household might well report that they "don't know" about purchases within a six month period but "might" or "probably will" buy within twelve months. The probability weight is assumed to be invariant over time for all these groups; hence any change in expected purchases must be a consequence of changes in the proportion of households reporting intentions to buy or reporting "don't know."

For the nonintender series—designated as z —I use a probability weight of 0.05.³⁰ For nonintenders, however, I do not make the assumption that the probability weight is constant over time and that changes in the proportion of households measure changes in expected purchases. Rather, I assume that mean probability among nonintenders will vary over time, and that it varies directly with the change in the fraction of households reporting "don't know." In general terms, the argument is that an increase (decrease) in the proportion of "don't know" households is symptomatic of a general increase (decrease) in purchase probability among all households in the low probability regions of the distribution. One could, of course, use movements in other intender classes as being symptomatic of movements in purchase probability among nonintenders. It seems plausible to me, however, that movements in the intender class with the lowest purchase probability are more likely to reflect movements in average probability among nonintenders simply because these two classes are adjacent in the probability distribution. In effect, I am assuming that the p, s correlation is likely to be stronger the lower the probability interval defined by any given p . Concretely, I not only assume that the change in the proportion of "don't know" responses is symptomatic of a general increase (decrease) in mean probability among nonintenders, but is the best estimate of the amount of the increase (decrease). Other assumptions are equally plausible.

The basic data are shown in Tables 14 and 15. Table 14 has the propor-

³⁰ Nonintenders will not precisely equal the total sample minus all the groups discussed above. The Census data contain one classification comprising households reporting intentions to buy either new or used cars within twelve months. I have defined nonintenders as the total sample less the sum of this comprehensive category and of households reporting that they "don't know" about their twelve-month intentions.

Because I have included both six- and twelve-month "don't know" groups among the intenders, some households are counted twice; consequently, the sum of all intender groups, plus the nonintenders as defined in this note, slightly exceeds 100 per cent.

BUYING INTENTIONS AND PURCHASE PROBABILITY: I

tions of households in the buying categories of intention, don't know, and nonintention; Table 15 contains the weighted constructs designated above as p' , d , and z . As noted above, the existence of seasonal variation in the buying-intentions data requires either seasonal adjustment of the data or analysis based on differences between successive periods twelve months

TABLE 14
BASIC DATA FOR ANALYSIS OF TIME SERIES RELATIONS BETWEEN PURCHASES OF NEW
AUTOMOBILES AND INTENTIONS TO BUY AUTOMOBILES
(per cent of households reporting)

Survey Date	Definitely Intend to Buy a New Car Within 6 Mos. (1)	Might Buy a New Car Within 6 Mos. (2)	Intend to Buy a New Car Within 7-12 Mos. (3)	Intend to Buy a Used Car Within 12 Mos. (4)	<i>Don't Know About Buying Plans</i>		No Plans to Buy Within 12 Mos. (7)
					Within 6 Mos. (5)	Within 12 Mos. (6)	
Jan. 1959	1.2	1.9	3.6	8.5	3.4	6.9	76.0
April 1959	0.9	1.8	3.9	7.2	3.0	6.4	77.8
July 1959	1.2	2.0	3.8	7.8	3.1	6.5	76.4
Oct. 1959	1.8	2.9	4.9	9.3	2.8	5.6	73.2
Jan. 1960	1.3	2.2	3.9	8.4	3.6	7.8	73.5
April 1960	1.5	1.8	3.7	7.2	3.0	7.5	75.4
July 1960	1.2	1.9	3.8	7.2	2.8	7.1	76.1
Oct. 1960	1.5	2.2	4.2	8.0	2.8	7.1	74.3
Jan. 1961	1.4	2.1	3.9	8.3	2.7	7.4	74.7
April 1961	1.1	2.0	3.7	7.7	2.9	7.6	75.8
July 1961	1.4	2.0	4.2	7.9	2.9	7.3	75.3
Oct. 1961	1.6	2.1	4.4	8.2	3.1	8.0	73.5
Jan. 1962	1.5	2.2	4.1	8.2	3.0	7.7	74.2
April 1962	1.5	1.9	4.3	9.2	3.3	7.9	73.2
July 1962	1.4	2.0	4.0	8.1	2.9	7.6	75.0
Oct. 1962	1.8	2.3	5.0	8.3	3.2	8.0	72.5
Jan. 1963	1.5	2.4	3.9	8.7	3.2	8.1	73.3
April 1963	1.8	2.3	4.0	8.0	3.5	8.6	73.2

SOURCE: All data are from Table 1 in *Consumer Buying Intentions*, Department of Commerce, U.S. Bureau of Census, Current Population Reports, Series P-65, No. 2, May 22, 1963. The first six columns are shown directly in Table 1; the last column is equal to 100 per cent minus the sum of those with twelve-month intentions to buy new or used cars plus those who "don't know" about twelve-month intentions.

apart. I have selected the latter course; hence Δx is defined as the difference in the new automobile purchase rate between successive twelve-month periods, e.g., as the difference in the purchase rate for the periods January 1, 1960 to December 31, 1960 and January 1, 1959 to December 31, 1959; or between the periods April 1, 1960 to March 31, 1961, and April 1, 1959 to March 31, 1960, etc. The buying intentions series are treated similarly; hence, the intentions variables used in the empirical analysis are $\Delta p'$, Δd , and Δz ; these measure differences in the respective

BUYING INTENTIONS AND PURCHASE PROBABILITY: I

TABLE 15
WEIGHTED PERCENTAGES OF HOUSEHOLDS INTENDING TO BUY AND PURCHASING AUTOMOBILES

Time Period ^a	Intenders ^b (<i>p'</i>)	Don't Know ^c (<i>d</i>)	Nonintenders ^d (<i>z</i>)	Total (<i>p'</i> + <i>d</i> + <i>z</i>)	Purchasers ^e (<i>x</i>)
1 Jan. 1959-31 Dec. 1959	4.57	1.03	3.80	9.40	11.84
1 April 1959-31 Mar. 1960	4.14	.94	3.89	8.97	12.14
1 July 1959-30 June 1960	4.54	.96	3.82	9.32	12.27
1 Oct. 1959-30 Sept. 1960	6.04	.84	3.66	10.54	12.21
1 Jan. 1960-31 Dec. 1960	4.92	1.14	4.02	10.08	12.61
1 April 1960-31 Mar. 1961	4.50	1.05	4.06	9.61	11.91
1 July 1960-30 June 1961	4.37	.99	3.76	9.12	11.39
1 Oct. 1960-30 Sept. 1961	5.01	.99	4.16	10.16	11.03
1 Jan. 1961-31 Dec. 1961	4.86	1.01	3.44	9.31	11.07
1 April 1961-31 Mar. 1962	4.42	1.05	4.06	9.53	11.65
1 July 1961-30 June 1962	4.82	1.02	3.86	9.70	12.22
1 Oct. 1961-30 Sept. 1962	5.13	1.11	4.60	10.84	12.45
1 Jan. 1962-31 Dec. 1962	5.02	1.07	3.67	9.76	12.96
1 April 1962-31 Mar. 1963	5.13	1.12	4.26	10.51	13.20
1 July 1962-30 June 1963	4.80	1.05	3.93	9.78	13.30
1 Oct. 1962-30 Sept. 1963	5.57	1.12	4.60	11.29	n.a.
1 Jan. 1963-31 Dec. 1963	5.16	1.13	3.84	10.13	n.a.
1 April 1963-31 Mar. 1964	5.21	1.21	4.56	10.98	n.a.

^a Time period shown is period over which the purchase rate (*x*) is measured. The intentions surveys are taken during the first mentioned month of each period, usually in the third week.

^b Derived from data in Table 14: percentages in the first four columns multiplied by weights of 0.7, 0.5, 0.3 and 0.2, respectively, then summed. See above text for further description of procedures.

^c Derived from data in Table 14: percentages in the fifth and sixth columns multiplied by weight of 0.1, then summed. See above text for further description of procedures.

^d Derived from Table 14: percentages in the last column of Table 14 are first multiplied by weight of 0.05, then adjusted by the average of two link relative indexes obtained from the fifth and sixth columns in Table 14. To illustrate: for data in columns five and six the link relatives between January 1959 and January 1960 are 105.9 and 113.0 respectively; the average of these two indexes is 109.4. Thus, the January 1960 observation in the *z* column (4.02) is the January 1960 figure from the seventh column of Table 14 (73.5) \times 0.0547; (the 0.0547 figure is equal to 0.05 \times 109.4). Similarly, the January 1961 *z* observation is equal to the January 1961 figure from the last column of Table 14 (74.7) \times (0.0547 \times 84.3 = 0.0461); the 84.3 figure is the average of the link relatives between January 1961 and January 1960 for the fifth and sixth columns in Table 14. See above text for further description.

^e Derived from Table 1-B in *The Predictive Value of Consumer Buying Intentions: A Study of Aggregate Time Series*, April 1963, an unpublished manuscript by Theodore Flechsig of the Federal Reserve Board. Flechsig's data on sales come directly from *Ward's Automotive Reports*. He has converted total sales figures to proportions of households for a six-month time period, and I have simply added together his respective six-month figures to obtain a twelve-month series. The last few figures shown in this column represent figures calculated in the same manner as the rest except that they were not yet available at the time Flechsig's manuscript was circulated, hence are not contained in his Table 1-B.

intentions series based on surveys taken twelve months apart. As a consequence, only eleven (overlapping) observations are available for an investigation of the relation between purchases and buying intentions.

Table 15 provides the basic data for an examination of two related problems. First, is the time series variation in automobile purchase rates due mainly to variation in intenders', as opposed to nonintenders', purchases? If so, time series variation in purchases will be due mainly to changes in the proportion of intenders; if not, variation in purchases will be due mainly to changes in mean probability among nonintenders. If it is assumed that mean purchase probability for the various intender groups is as specified above (ranging from 0.7 for those with definite new car intentions within six months to 0.2 for those intending to buy used cars within twelve months) and is constant over time,³¹ the change in total purchases (Δx) can be separated into the change in intenders' purchases ($\Delta p'$) and a residual, $\Delta x - \Delta p'$. Δx is an *ex-post* magnitude, and $\Delta p'$ is *ex-ante* by construction; $\Delta x - \Delta p'$ can be thought of as the sum of purchases that nonintenders expected to make—an unobserved but *ex-ante* magnitude—and of purchases that reflect the influence of unforeseen events. Although the relative importance of unforeseen events cannot be measured directly, there is no reason to expect that it will be the dominant element in the variance of $\Delta x - \Delta p'$.

The data indicate that the variance of $\Delta p'$ —intenders purchases—is less than about one-fourth the size of the variance in $\Delta x - \Delta p'$. This empirical finding is not altered if intenders are defined to include households reporting that they "don't know" whether they will buy a car: the variance of $\Delta p' + \Delta d$ is almost exactly the same as that of $\Delta p'$ itself, and the variance of the new residual, $\Delta x - \Delta p' - \Delta d$, is still more than four times as large. I conclude that a major part of the time series variance in purchase rates consists of the variation in mean probability among nonintenders. It follows that the correlation between Δx and $\Delta p'$ will be relatively small unless changes in the proportion of intenders happen to be a good predictor of changes in nonintender mean probability, and hence nonintender purchase rates.

The second question concerns the correlation between intentions and purchases. The basic data are drawn from Table 15; the empirical results are summarized in Table 16.

³¹ What evidence I have seen suggests that mean probability among intenders is in fact quite stable over time. But it is not even necessary that such be the case: an exact estimate of the total purchases made by intenders will have the same variance over time as my constructed estimate provided that mean probability among intenders is uncorrelated with the proportion of households reporting intentions.

BUYING INTENTIONS AND PURCHASE PROBABILITY: I

Changes in weighted buying intentions ($\Delta p'$) explain about 40 per cent of the variance in purchase rate changes (eq. a in Table 16); the correlation is barely significant at the 0.05 level. By themselves, changes in the "don't know" (Δd) or nonintender (Δz) groups explain considerably less of the variance in purchase rate changes than does $\Delta p'$. However, the sum of $\Delta p'$ and Δd brings the explained variance up to almost 50 per cent (equation e), while summing all three expected purchases series $\Delta(p' + d + z)$ brings the explained variance up to roughly two-thirds of the total (equation f). The correlation between Δx and $\Delta(p' + d + z)$ is significantly different from zero at the 0.01 level, and is significantly *higher* than the correlation between Δx and $\Delta p'$ at the 0.05 level.

TABLE 16
TIME SERIES CORRELATIONS BETWEEN THE CHANGE IN NEW AUTOMOBILE
PURCHASE RATES AND THE CHANGE IN ALTERNATIVE MEASURES OF
AUTOMOBILE BUYING INTENTIONS

Equation Number	Equation	Squared Correlation Coefficient
(a)	$\Delta x = +.19 + 1.74\Delta p'$,	$r^2 = .39^a$
(b)	$\Delta x = +.03 + 5.53\Delta d$,	$r^2 = .13$
(c)	$\Delta x = +.07 + 1.97\Delta z$,	$r^2 = .23$
(d)	$\Delta x = +.09 + 1.80\Delta(p' + d)$,	$r^2 = .47^a$
(e)	$\Delta x = +.06 + 1.49\Delta(d + z)$,	$r^2 = .15$
(f)	$\Delta x = -.11 + 1.77\Delta(p' + d + z)$,	$r^2 = .67^b$

SOURCE: Estimated from basic data in Table 15. See the notes to Table 15 or the text for explanation of symbols.

^a *F* ratio significantly different from unity at 0.05 level.

^b *F* ratio significantly different from unity at 0.01 level.

In effect, a substantial part of what may be regarded as the residual variance in actual purchases is explained by adding a very crude proxy for nonintenders' expected purchases to expected purchases by intenders: changes in expected purchases by intenders explains 39 per cent of the variance in total purchases; accounting for changes in expected purchases by the "don't know" and nonintender households brings the explained variance up to 67 per cent. Thus almost half of the residual variance (28 of 61 percentage points) is explained by this crude approximation for nonintender purchases.

An alternate way of investigating the same proposition with somewhat different assumptions involves correlating Δd —the change in the proportion of "don't know" households—with the purchase rate of nonintenders. The latter can be estimated by the following procedure.

1. Using the probability weights assigned above to the various intender groups, estimate the actual purchases of nonintenders by calculating $x - p' - d$ from Table 15.
2. Estimate the purchase *rate* of nonintenders by dividing the proportion of nonintenders into estimated purchases by nonintenders; that is, divide $x - p' - d$ by column 7 of Table 14. The resulting values can be thought of as estimates of mean *ex-ante* purchase probability among nonintenders.

Designating the nonintender purchase rate as s , and its year-to-year change as Δs , we have

$$\Delta s = 0.04 + 2.67\Delta d, r^2 = 0.21$$

The correlation between Δs and Δd , while positive, is not significantly different from zero at conventional levels. This test thus implies a less strong association between changes in the purchase probability of nonintenders and changes in the proportion of "don't know" households than the set of regressions discussed earlier. It should be noted, however, that relative to the earlier regressions, the $\Delta s, \Delta d$ test has an added constraint as well as an extra degree of freedom.³²

Another interesting piece of time series evidence concerns the explanatory power of the above intender series net of other variables that are associated with purchase rates. In the final chapter of this monograph I present cross-section data in which durable goods purchases are regressed on some fifteen independent variables. Some of the independent variables represent circumstances that were known at the survey date (designated as initial-data variables), while others represent circumstances that were wholly or partly unforeseen at the survey date (designated as intervening variables). The probability hypothesis implies that initial-data variables should be more strongly related to buying intentions than to purchases, while the opposite relation should hold for intervening variables. Further, the hypothesis implies that intervening variables will make a larger contribution than initial-data variables to an explanation of the residual variance in purchases, buying intentions held constant. The argument is simply that the influence of initial-data variables will be accounted for by intentions because the values of such variables are known to the respondent

³² The two procedures would show substantially similar results if the regression of Δx on $\Delta(p' + d)$ yielded a slope coefficient close to unity. This is the general nature of the constraint; the data show that the regression coefficient of $\Delta(p' + d)$ is considerably higher than unity (see Table 16). On the other hand, the $\Delta s, \Delta d$ correlation permits the regression coefficient of Δd to vary, while the first procedure forces the coefficients of Δd and $\Delta p'$ to equality. There are other differences as well in the two procedures, but these are the most important ones.

BUYING INTENTIONS AND PURCHASE PROBABILITY: I

at the time intentions are reported; this is less true of intervening variables because they reflect wholly or partly unforeseen events that influence actual purchases but not buying intentions or purchase probabilities.³³

On this line of argument, changes in concurrent income, as compared with past income, should be *more* strongly associated with changes in purchases, *less* strongly associated with changes in buying intentions. Past income change is clearly an initial-data variable, while concurrent income is intervening to the extent that the change is not foreseen.

TABLE 17
TIME SERIES CORRELATIONS BETWEEN THE CHANGE IN NEW AUTOMOBILE PURCHASE RATES, THE CHANGE IN AUTOMOBILE BUYING INTENTIONS, AND THE CHANGE IN DISPOSABLE INCOME

Equation	Dependent Variable	Independent Variable(s)	Multiple R ² or Simple r ²	Partial r ² for		
				$\Delta(p' + d + z)$	Δy_{-1}	Δy
(g)	Δx	Δy_{-1}	.56 ^b			
(h)	Δx	Δy	.86 ^b			
(i)	$\Delta(p' + d + z)$	Δy_{-1}	.31			
(j)	$\Delta(p' + d + z)$	Δy	.43 ^a			
(k)	Δx	$\Delta y_{-1}, \Delta y$.90 ^b		.30	.78 ^b
(l)	$\Delta(p' + d + z)$	$\Delta y_{-1}, \Delta y$.46		.06	.22
(m)	Δx	$\Delta(p' + d + z), \Delta y_{-1}$.79 ^b	.53 ^a	.38	
(n)	Δx	$\Delta(p' + d + z), \Delta y$.94 ^b	.54 ^a		.81 ^b
(o)	Δx	$\Delta(p' + d + z), \Delta y_{-1}, \Delta y$.94 ^b	.38	.05	.71 ^b

SOURCE: Basic data for purchases and buying intentions are shown in Table 15; concurrent and past income changes calculated from the *Survey of Current Business*, U.S. Department of Commerce, various issues. See Table 15 and the above text for explanation of symbols.

^a F ratio significantly different from unity at 0.05 level.

^b F ratio significantly different from unity at 0.01 level.

Designating concurrent income change as Δy and past income change as Δy_{-1} ,³⁴ the results shown in Table 17 are obtained.

Concurrent income is more highly correlated with purchases than with buying intentions, it is more highly correlated with purchases than is past

³³ See Chapter 7, pp. 166-168, for a more detailed analysis.

³⁴ The symbol Δy_{-1} is defined as the difference in disposable income between the quarter immediately preceding the intentions survey and the comparable quarter one year earlier. The symbol Δy is defined as the average difference in disposable income between the first two quarters of a twelve-month purchase period and the comparable two quarters one year earlier; both differences are measured in billions of current dollars. Interestingly enough, when concurrent income change is measured further along in the purchase period, its correlation with the purchase rate is weakened, possibly because of a lag in the relation between the two. A similar pattern of results is shown in Flechsig's manuscript.

income, and it adds considerably more to the explanation of purchases net of buying intentions than does past income; all these results are in accord with predictions. Actually, a linear combination of change in buying intentions— $\Delta(p' + d + z)$ —and concurrent income— Δy —explains almost all of the variance in purchase rates, and an F test indicates that both partial correlations are highly significant. The explanation of purchases is hardly improved at all when past income is added to concurrent income and buying intentions (compare equations n and o in Table 17). However, many of the predicted relations involving income changes and buying intentions do not appear. Concurrent income is somewhat more strongly (rather than less strongly) associated with intentions than is past income and concurrent income adds more than past income to an explanation of buying intentions (compare equations i, j, and l in Table 17).

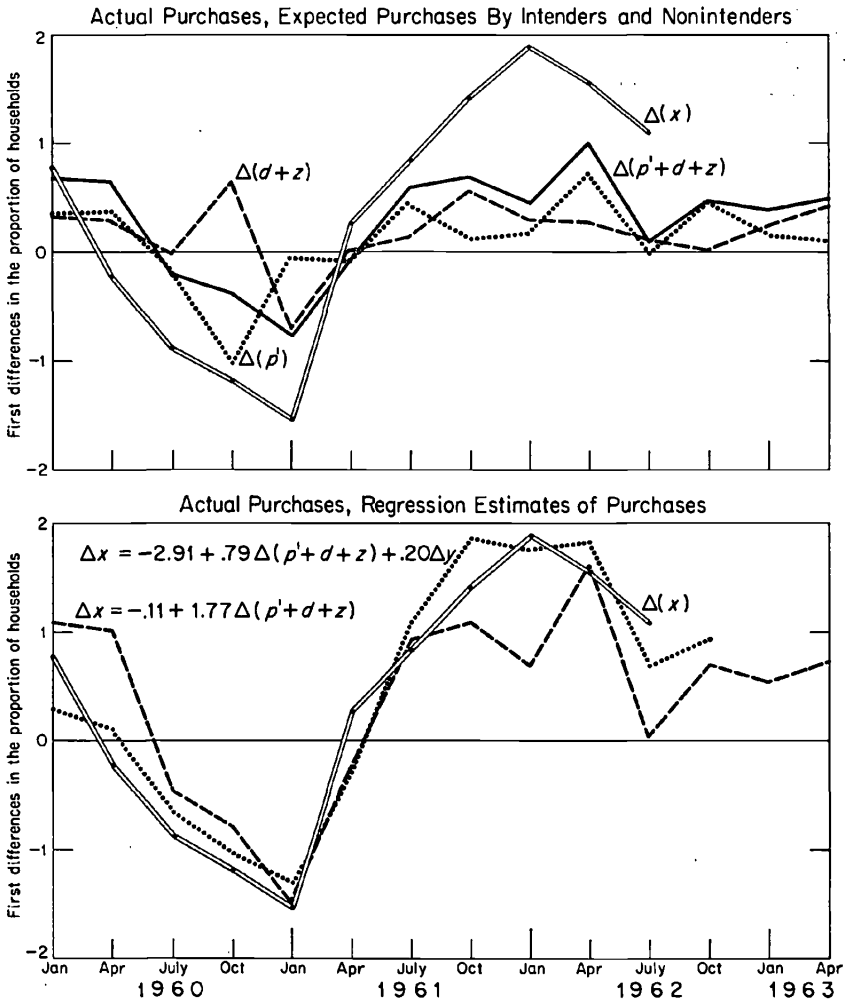
The findings of this exploratory time series investigation are summarized in Chart 5. The upper panel shows first differences in the actual purchase rate of new automobiles, and in the rate of expected purchases by intenders, nonintenders plus don't know households, and the sum of the two. In almost every time period, it is evident that adding nonintenders' expected purchases (estimated as described above) to those of intenders improves the fit with actual purchases. The lower panel shows first differences in actual purchases plotted against regression estimates based on expected purchases and on the combination of expected purchases and concurrent income. The latter combination of variables evidently explains almost all of the variance in actual purchases.

Let me note once again that the above results cannot be viewed as conclusive evidence in favor of either the analytical model developed in this chapter or of the empirical assumptions used above. The level of statistical significance attributed to these results is undoubtedly exaggerated; the particular constructs used in the correlation analysis were neither the first used nor the only reasonable ones that could have been used. Hence, more degrees of freedom have in fact been used up than appears to be the case, and there is no formally correct way to adjust the results to account for this. In my judgment, the chances are better than even that the general pattern of the empirical results is substantive; that is, some such assumption as made above about movements in mean purchase probability among nonintenders is likely to improve purchase predictions based on buying intentions surveys. I doubt that the particular assumption made above is the most appropriate one, and I suspect that direct measurement of purchase probability among nonintenders would turn out to be a superior predictor than any proxy variable.

CHART 5

Change from Preceding Year in Twelve-Month Purchase Rates of New Automobiles Compared with Corresponding Change in Rate of Expected Purchases

x = Actual purchase rate
 p' = Expected purchase rate of intenders
 $(d+z)$ = Expected purchase rate of nonintenders



Source: Basic data from Table 15.

Summary

On the whole, the evidence tends to support the hypothesis that questions about intentions to buy elicit responses drawn from a continuous distribution of purchase probabilities. No convincing evidence is available on the shape or the stability of the distribution functions, nor is it known whether the mean of these functions can be accurately estimated from the proportion of the sample in the upper tail. It also seems clear that the typical intentions questions in current use tend to be relatively inefficient in that substantial numbers of households reporting that they "intend to buy" have lower purchase probabilities than some households not so reporting. Given these results, further experimentation with survey design seems to be called for, and the objective is self-evident. If a continuous distribution of purchase probabilities characterizes the universe, a survey should attempt to estimate the mean of this distribution. The simplest procedure is to estimate the distribution and compute the mean directly. In this way no assumptions need be made about the distribution function, its stability over time, or the relation between the mean and the proportion of households in the upper tail. It may turn out that only parts of the distribution are of interest for prediction purposes, but I see no way of deciding this question on a-priori grounds.

There are, of course, many possible procedures by which such information might be obtained, including the direct approach. I know of no convincing evidence that survey respondents would experience more difficulty in answering the question "What are the odds that you will buy A next year?" than in answering the kinds of questions currently used in intentions surveys, e.g., "Do you intend to buy A next year?" and "How certain are you of these plans?" It is even possible that the direct approach is the best, particularly since differences among households in the interpretation of a survey question may well be greater when literary terms are used than when numerical scales are explicit in the question. That is to say, on a-priori grounds it could be argued that the variance in cut-off probability (and, therefore, in purchase probability) among those reporting that they "definitely intend to buy within n months" is greater than the variance among those reporting that "the odds on buying within n months are better than two to one."³⁵ These are questions of judgment, however, and nothing in the above data suggests that one or another way of measuring subjective probability is superior. It may be that a more complicated procedure, involving combined responses to several questions,

³⁵ The illustration assumes that average probability is the same in both cases.

would be more accurate than simpler and more direct methods. This question is entirely open; but the conclusion seems inescapable that more precise and comprehensive scaling of the entire probability distribution is an essential requirement for increasing the accuracy of predictions based on survey data.

Granting that the probability model constitutes the most plausible interpretation of intentions data, a number of problems remain about how such data are to be used for predicting purchases of durable goods. First, is the cut-off probability corresponding to a given intentions question the same for all commodities and for all households in the population? If not, what are the characteristics to which the C_i are related? These questions are discussed in the next chapter. Second, how do unanticipated events (contingencies) influence the translation of initial probabilities into purchases? The third is similar to the second. Do concurrent attitudes have any effect on the initial probabilities associated with responses to questions about buying intentions or about explicit probabilities? These problems are discussed in Chapter 6. Finally, are the *ex ante* purchase probabilities associated with responses to a specific set of questions about buying intentions (or explicit probabilities) invariant over time? That is, are time series differences in average probability for the population accurately reflected in the changing proportion of responses to a fixed scale, or does the probability associated with a particular response vary systematically with particular kinds of circumstances, e.g., the level of unemployment, the state of business activity, etc.? In principle, one needs a series of observations $C_i \dots_j$, relating to different time periods. My analysis implicitly assumes that a set of such observations will show only random variation due to sampling error. This problem is not discussed here; the necessary data cannot be readily obtained from any source, and cannot be obtained at all from the CU survey materials.

*Appendix I: A Geometrical Model of the
Subjective Probability Hypothesis*

By amplifying the distribution function in Chart 1, I shall elaborate the subjective probability hypothesis and present some crude tests of its empirical content. The upper panel of Chart 6 shows the same distribution function as in Chart 1, i.e.,

$$1.0 \qquad n = f(Q),$$

and the cumulative distribution function obtained by summing from $C = 0$

CHART 6

Illustrative Distributions of Households
by Subjective Purchase Probability, by Expected Purchases,
and by Average and Marginal Purchase Probability

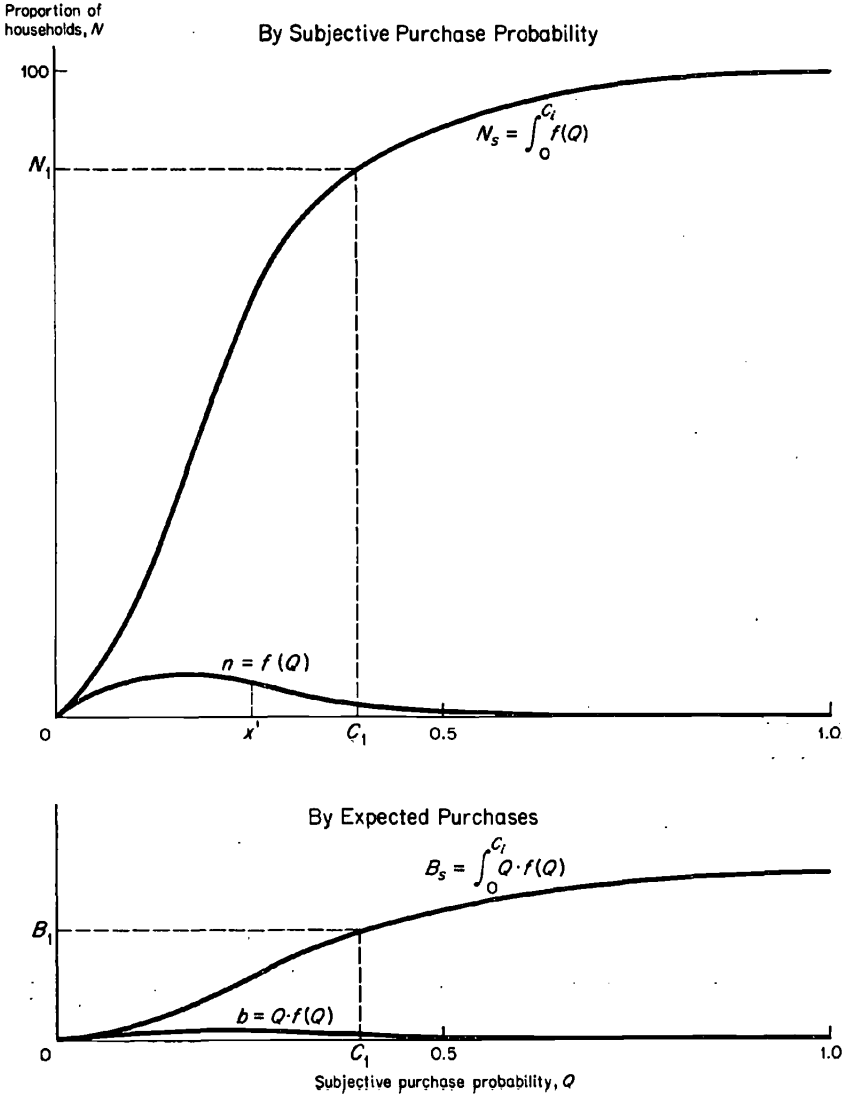
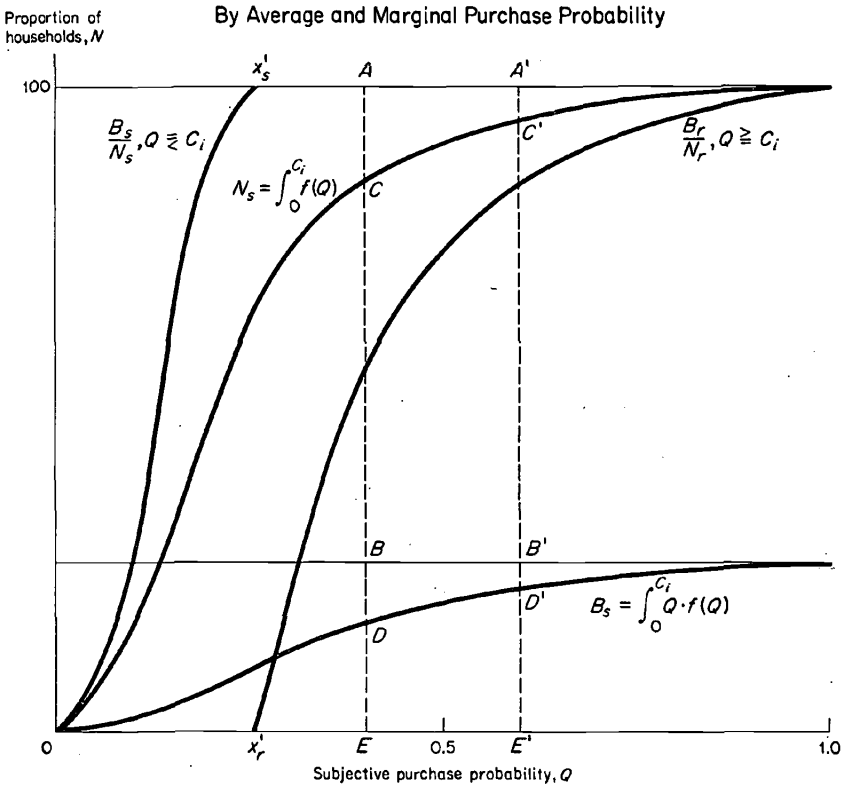


CHART 6 (concluded)



to $C = C_i$, that is,³⁶

$$1.1 \quad N_s = \int_0^{C_i} f(Q) dq.$$

The middle panel of Chart 6 shows an expected purchases (b) function obtained by multiplying the number of households at each probability level by the probability itself. If b households have purchase probabilities equal to 0.5, we expect these households, on the average, to make 0.5b purchases. The middle panel also contains a cumulative expected purchases function, obtained by summing from zero to C_i as above.

$$2.0 \quad b = Q \cdot f(Q)$$

$$2.1 \quad B_s = \int_0^{C_i} Q \cdot f(Q) dq$$

³⁶ I designate $\int_0^{C_i} f(Q)$, as N_s to indicate that it is the cumulative number of households with probabilities less than C_i , that is, the number that would not report intentions to buy if asked a question with a cut-off probability equal to C_i .

The two cumulative functions 1.1 and 2.1 have some useful properties. Equation 1.1 shows the number of households with probabilities equal to or less than C_i . Equation 2.1 shows the expected number of purchases by households with probabilities equal to or less than C_i . In the top panel of Chart 6, for example, N_1 households have subjective purchase probabilities equal to or less than C_1 . These N_1 households, on the average, will make purchases equal to B_1 , as indicated by the middle panel.

The ratio of these two cumulative functions— B_i/N_i —is thus the fraction of households with $Q \leq C_i$ expected to purchase during the forecast period. When $C_i = 1.0$, this ratio is evidently equal to mean probability for the sample as a whole, i.e., to x' ; when C_i approaches zero, this ratio will also approach zero. Two complementary cumulative functions, specifying the number of households and number of expected purchases for those with subjective probabilities equal to or *greater than* C_i , can also be constructed:

$$1.2 \quad N_r = \int_{C_i}^1 f(Q) dq$$

$$2.2 \quad B_r = \int_{C_i}^1 Q \cdot f(Q) dq$$

It follows that

$$1.3 \quad N_r + N_s = \int_0^1 f(Q) dq = N,$$

$$2.3 \quad B_r + B_s = \int_0^1 Q f(Q) dq = B,$$

and

$$3.0 \quad B/N = x'$$

Further, B_r/N_r is the fraction of households with $Q \geq C_i$ expected to purchase during the forecast period. When C_i is zero this fraction is equal to the proportion of the entire sample that is expected to purchase, i.e., to x' . As C_i approaches 1.0, this fraction will also approach 1.0.

The bottom panel of Chart 6 reproduces the cumulative functions 1.1 and 2.1 from the upper panels and, in addition, shows the relation between N and both B_r/N_r and B_s/N_s . The last two are, of course, average purchase probability for those with $Q \geq C_i$ and $Q \leq C_i$, respectively, since they relate numbers of expected purchases to numbers of people in the respective categories. These two ratios are plotted against the (horizontal) subjective probability scale. The ratio B_r/N_r varies from x' to 1.0 (and N varies from 0 to 100) as C_i varies from 0 to 1.0. The ratio B_s/N_s varies from 0 to x' as C_i varies from 0 to 1.0.

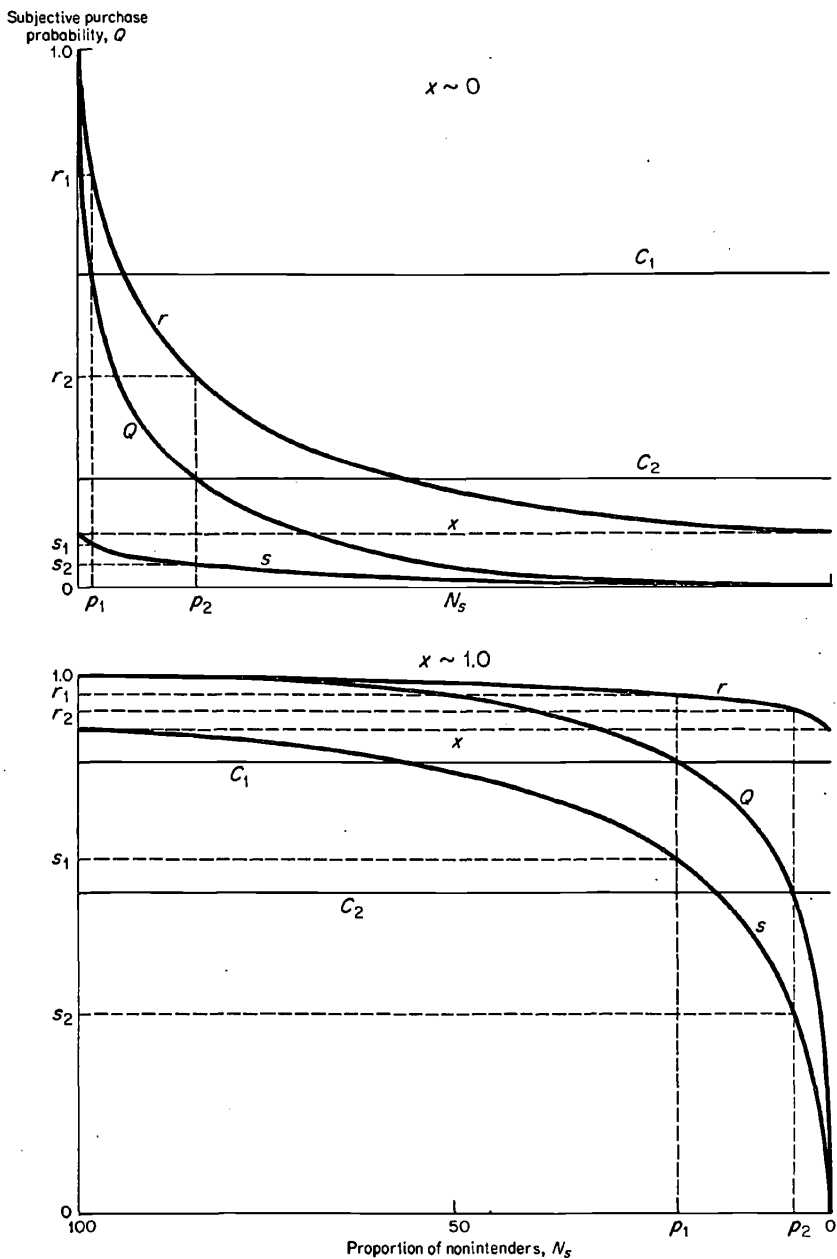
The construction of these ratios is illustrated in the bottom panel of Chart 6. When C_i is at E , N_s equals CE , and a proportion DE/CE is expected to purchase; when C_i increases to E' , N_s increases to $C'E'$; and a smaller proportion, equal to $D'E'/C'E'$, is expected to purchase. When C_i approaches zero the ratio also approaches zero, and when C_i approaches unity the ratio approaches x' . The resulting values of B_s/N_s will trace out a curve from the lower left-hand corner of the diagram to x'_s at the top. Similarly for B_r/N_r : when C_i is at E , N_r equals AC ; when C_i increases to E' , N_r falls to $A'C'$. In the former case, a proportion equal to BD/AC is expected to purchase; and in the latter, a larger proportion, equal to $B'D'/A'C'$. When C_i approaches 1.0, N_r approaches zero, and B_r/N_r approaches unity. When C_i approaches zero, N_r approaches 100, and B_r/N_r approaches x'_r . The resulting values of B_r/N_r trace out a curve from x'_r at the bottom of the diagram to the upper right-hand corner.

I have noted above that B_r/N_r and B_s/N_s are essentially average purchase probability functions for households with $Q \geq C_i$ and $\leq C_i$, respectively. The cumulative frequency function— $N_s = \int_0^{C_i} f(Q)$ —can also be interpreted as a probability function, marginal (in the sense used by economists in discussing cost or revenue functions) to both of the above average probability functions. The cumulative function 1.1 shows the number of households with probabilities equal to or less than some cut-off point, C_i . The marginal household in that function evidently has a probability of purchase, Q , equal to C_i . Consequently, the ratio of the slopes of the B and N functions at C_i must be the marginal probability of the C_i th household. Since the marginal probability is the same whether one is cumulating households from $C_i = 0$ to $C_i = 1.0$ or vice versa, the cumulative frequency distribution is marginal (in the above sense) to the average probability function with the origin in the lower left-hand corner ($B_s/N_s, Q \leq C_i$), as well as to the average probability function with the origin in the upper left-hand corner ($B_r/N_r, Q \geq C_i$).

Chart 7 is designed to make better use of the available data, and shows the lower panel of Chart 6 for alternative distributive functions. The axes have been rotated a quarter turn to the left, so that both marginal and average (subjective) probability are now on the vertical axes; the cumulative number (per cent) of households is on the horizontal. The horizontal axis represents the proportion of nonintenders when it is read as scaling 100 per cent to 0 per cent reading from left to right; it also represents the proportion of intenders when read as 100 per cent minus the proportion of nonintenders. The cumulative frequency function—

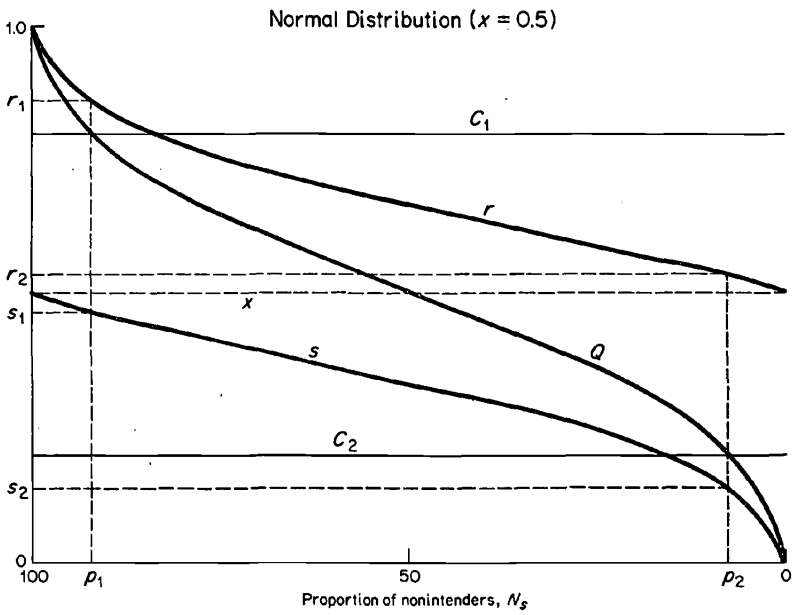
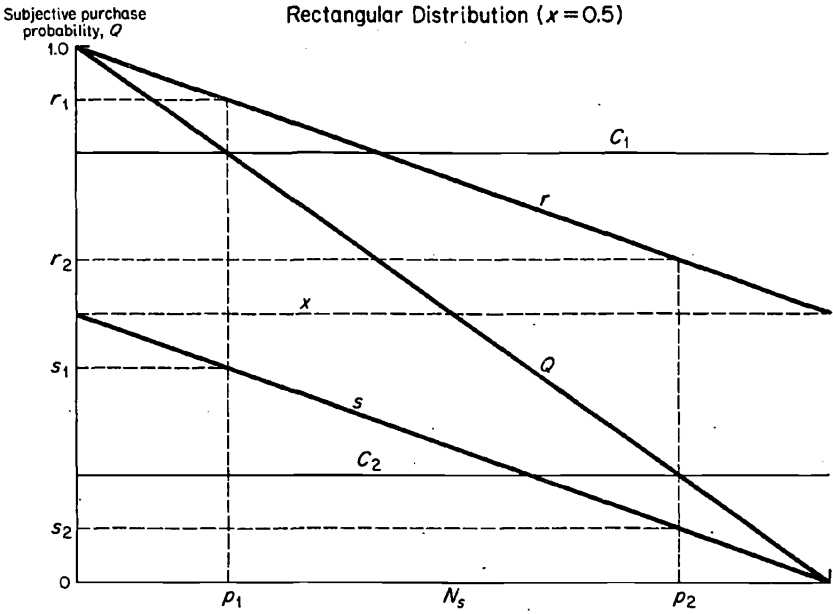
CHART 7

Illustrative Functional Relations Between
Alternative Intentions Questions and Purchase Rates
for Intenders and Nonintenders



BUYING INTENTIONS AND PURCHASE PROBABILITY: I

CHART 7 (concluded)



$N_s = \int_0^{C_i} f(Q)$ —is the marginal probability curve, designated as Q . One average probability curve is the ratio B_r/N_r ; the other is the ratio B_s/N_s . The first of these functions is designated r' , the second s' , in accordance with the nomenclature used above. A horizontal line has been drawn at x' , the mean probability for the sample as a whole. The top panel of Chart 7 shows these functions for a probability distribution with a very small value of x' ; second panel, the same functions for a distribution with a very large value of x' ; the third panel, the same functions for a rectangular distribution with $x' = 0.5$; and the fourth panel, the same functions for a normal distribution with x' also equal to 0.5.

Given a survey of intentions to buy, the survey question having a cut-off probability of C_i , and a subsequent survey of purchases, we will observe some fraction (p_i) of the sample reporting intentions, a fraction $(1 - p_i)$ not reporting intentions, and fractions r_i and s_i of p_i and $1 - p_i$, respectively, actually purchasing. A fraction (x_i) of the sample as a whole will actually purchase. I assume that *ex ante* mean purchase probability for any group is equal to the *ex post* fraction of the group observed to purchase, although there is some regression bias in the relation between these two measures as discussed above. Thus, x_i is an estimate of x'_i ; r_i , of r'_i ; and s_i , of s'_i . Several combinations of points $(r, p)_i$ and $(s, 1 - p)_i$ can be located from the empirical data, although there are not any directly observable measures of the corresponding cut-off probability, C_i . Two such combinations of points are illustrated in Chart 7, for cut-off points designated C_1 and C_2 .

From Chart 7 the same generalizations can be elicited as in the text about the relation among r , s , p , and x . First, r exceeds s for any C_i ; therefore, $r - s$ must always be positive. Second, both r_i and s_i are positively related to C_i , hence negatively related to p_i . These generalizations hold for the distribution functions in all four panels and are therefore independent of the value of x . Third, the relation between $r - s$ and p depends on x . When x is small (top panel), $r - s$ is positively related to C ; hence, negatively related to p . When x is large (second panel), $r - s$ is negatively related to C ; hence, positively related to p .⁸⁷ In the third

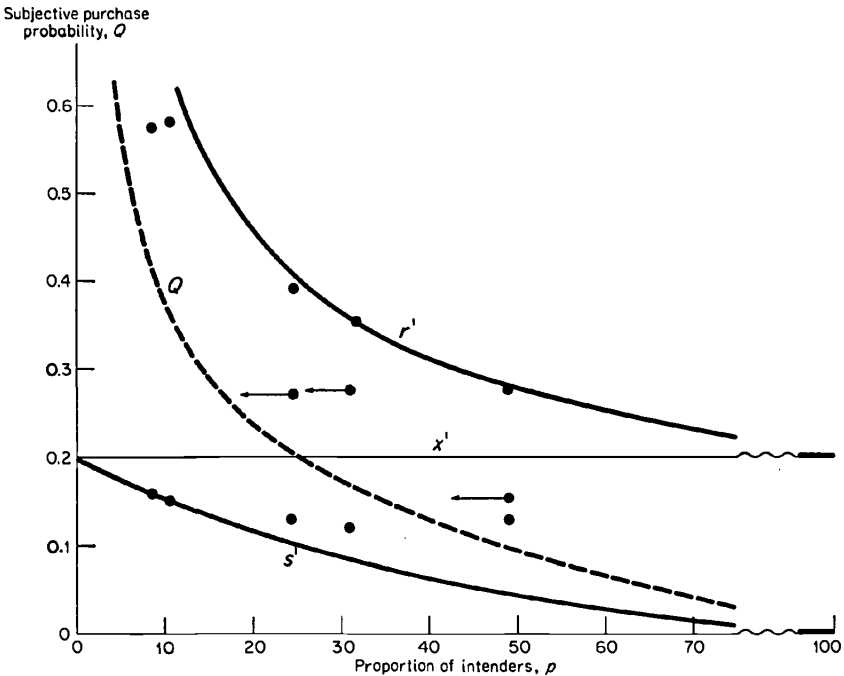
⁸⁷ The last two statements are clearly true in general for the functions shown in Chart 7: since $r - s$ is necessarily equal to x when p is 100 per cent—and $C = 0$ —and to $1 - x$ when p is zero per cent—and $C = 1.0$ —it follows that the correlation between $r - s$ and p will tend to be negative if $1 - x$ is larger than x , and to be positive if x is larger than $1 - x$. Thus, the correlation between $r - s$ and p must generally be negative for functions like that in the top panel, positive for functions like that in the second panel. It appears, however, that parts of the function may have the reverse relationship. For example, the correlation between $r - s$ and p for values of p close to 100 per cent

BUYING INTENTIONS AND PURCHASE PROBABILITY: I

panel, where $x' = 0.5$, there is evidently a zero correlation between $r - s$ and p because the former is a constant (0.5) throughout the entire range of p . The last panel also has a distribution function with an x' of 0.5, but with the characteristics of a normal (Gaussian) function rather than of a rectangular function. In this panel there is zero linear correlation

CHART 8

Proportion of Sample Reporting Intentions to Buy Automobiles, and Proportion of Intenders and Nonintenders Purchasing Within a Six-Month Period



Source: Basic data from Tables 9 and 10.

between $r - s$ and p ; however, there is a negative association between the two in the region where p is less than 0.5, a positive association where p is greater than 0.5. Finally, the variance in $r - s$, as a function of C or p ,

might well be positive for functions like those in the top panel; and for functions like those in the second panel, the correlation might well be negative for values of p close to zero, i.e., where nonintenders are close to 100 per cent. If functions were drawn for values of x slightly under 0.5 (top panel) or slightly over 0.5 (second panel) it would be impossible to tell by visual inspection whether the correlation between $r - s$ and p was positive or negative or zero in either panel.

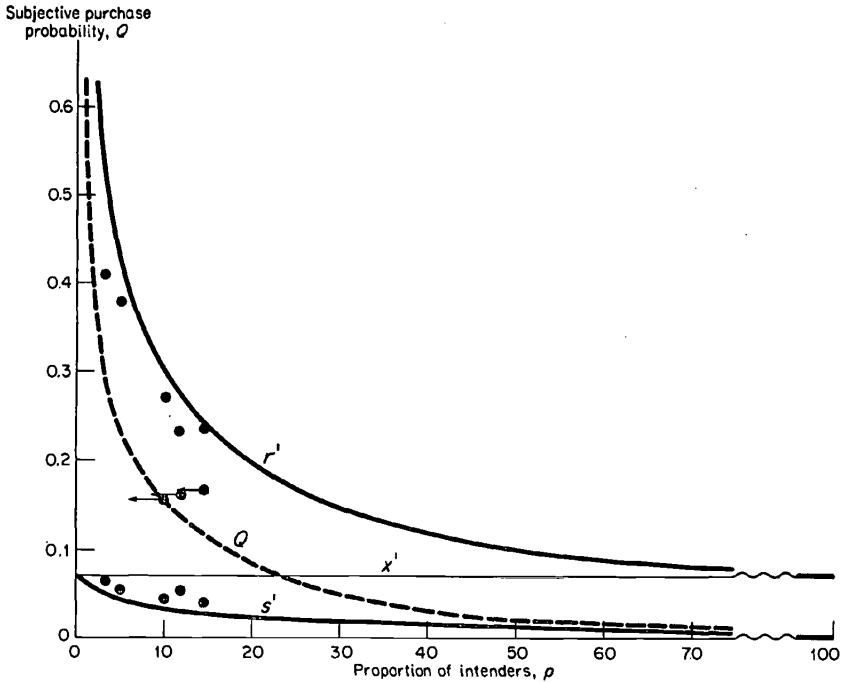
BUYING INTENTIONS AND PURCHASE PROBABILITY: I

depends on the value of x . The closer x comes to 0.5, the smaller the variance of $(r - s)$; the closer x comes to either 0 or 1, the larger the variance. All these generalizations can be observed in the data, as already discussed in the text.

It is an interesting exercise to plot some of the observed $(r, p)_i$ and $(s, 1 - p)_i$ points to get a clearer view of the probability distributions for

CHART 9

Proportion of Sample Reporting Intentions to Buy Washing Machines, and Proportion of Intenders and Nonintenders Purchasing Within a Six-Month Period



Source: Basic data from Tables 9 and 10.

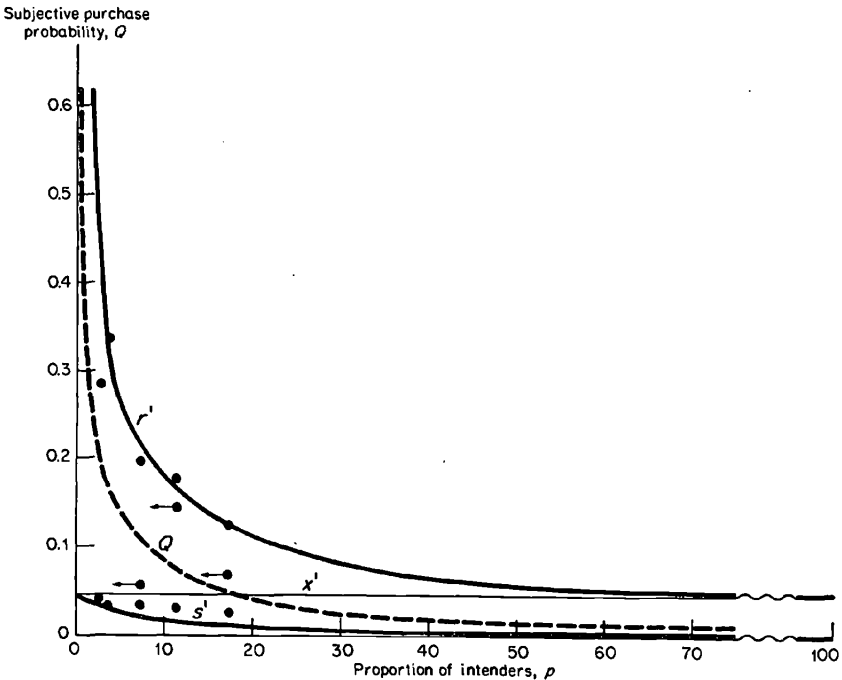
different items. All observed functions have values of x that are below 0.5; consequently, functions like those in the top panel of Chart 7 are appropriate. The data for automobiles, for washing machines, and for clothes dryers are plotted in Charts 8, 9 and 10. These data are representative of the least skewed function (automobiles), the most skewed (clothes dryers), and the "typical" function (washing machines). Points on the r and s functions can be located precisely from the basic data in

BUYING INTENTIONS AND PURCHASE PROBABILITY: I

Tables 9 and 10. Although some observations for the Q function are available, they all consist of averages for fairly large segments, that is, of mean probability between two unknown cut-off points. These are plotted at $(Q_i \dots_j, p_j)$; that is, the mean probability for households with probabilities lower than i but higher than j is located at p_j . Such points are necessarily too far to the right, and an arrow indicates this fact.³⁸

CHART 10

Proportion of Sample Reporting Intentions to Buy Clothes Dryers, and Proportion of Intenders and Nonintenders Purchasing Within a Six-Month Period



Source: Basic data from Tables 9 and 10.

The empirically observed points in these charts are plotted from data in Tables 9 and 10. Panel A in these tables contains points on the Q function; Panel B has points on the r' function; Panel C, points on the s' function. The r' and s' functions are drawn to correspond roughly to the observed points, account being taken of the regression bias noted above. Because of this bias the observed $(r, p)_i$ points will generally be below the

³⁸ It cannot be correct to plot these q points midway between p_i and p_j , since the Q function is not a straight line. The i, \dots, j points are so far apart that location at the midpoint is a poor estimate.

true $(r', p)_i$ point, since *ex ante* probability must be higher than the *ex post* purchase rate for groups of households in the upper tail of the distribution. Conversely, the observed $(s, 1 - p)_i$ points will generally be above the true $(s', 1 - p)_i$ points. In Charts 8, 9, and 10 the bias is clearly evident in the s' function; any reasonable estimate of this curve must lie below the observed s points, especially for relatively high values of p .

Appendix II: An Estimate of the Potential for Improvement in the Prediction of Purchases

One additional piece of empirical data can be brought to bear on the prediction problem. Granting the probability hypothesis, there are only two reasons that a survey of buying intentions will produce an erroneous forecast of the population purchase rate: first, the survey may not yield enough information for a reliable estimate of mean *ex ante* purchase probability, because the proportion reporting that they intend to buy is a less than perfect substitute for the mean; second, the *ex post* purchase rate may diverge from *ex ante* mean purchase probability because of unforeseen events. Evidently, only the first type of error can be reduced or eliminated by improvements in survey design.³⁹ But are errors of the first type large enough to merit a serious attempt to reduce or eliminate them?

The question cannot be answered with precision, but an estimate of the minimum size of type 1 errors can be obtained from the data. In Chapter 2 estimates are given of the proportion of the variance in actual purchases explained by responses to each of the intentions questions—that is, the cross-section r^2 (coefficient of determination) between X and P , where both X and P are dichotomous (1,0) variables. But $r^2_{X,P}$ is less than unity because of the effects of the two types of error just discussed.⁴⁰ And an

³⁹ Errors due to unforeseen events can be reduced only if both the functional relation between *ex ante* probability and unforeseen events, as well as the events themselves, can be specified. In that case perfectly accurate contingent forecasts would be possible.

⁴⁰ It may be objected that the cross-section correlation would be less than unity even if a perfectly accurate measure of *ex ante* purchase probability were available for each household and no unforeseen events took place during the period. The observed values of X are inherently dichotomous—either a household purchases the commodity or it does not—but *ex ante* probability is presumably a continuous function. Thus, of 100 households with probabilities of 0.1, on the average 10 will buy and 90 will not buy; and the cross-section correlation $r^2_{X,Q}$, where Q is purchase probability, will be less than unity.

The objection, while valid, does not affect the analysis. The observed values of $r^2_{X,P}$, where both X (purchases) and P (intentions to buy) are (1,0) variables, are unbiased estimates of $r^2_{x,p}$, where x and p are the proportions of purchasers and intenders drawn from a number of randomly selected subsamples. But if purchase probability were an available datum, the correlation between values of mean purchase probability and the proportion of purchasers drawn from the same randomly selected subsamples would be unity, assuming that no unforeseen events occur and that the subsamples are very large.

estimate of the maximum size of type 2 errors can be obtained from the data; hence, also, an estimate of the minimum size of type 1 errors.

The argument is as follows: errors due to the influence of unforeseen events essentially constitute a regression bias. As noted above in the section "The Problem of Bias," this is one of the reasons that r_i is a downwardly biased estimate of r'_i ; s_i , an upwardly biased estimate of s'_i . Although there is no way to estimate the true value of r'_i , there are several ways of estimating the maximum value of r'_i , given the commodity and the buying intentions question.

To begin with, it has previously been demonstrated that the following inequality must hold by definition:

$$1 \geq r', s' \geq 0;$$

that is, both r' and s' must be between zero and unity. Further, both r' and s' are also constrained by the definitional relation

$$x' = r'p + s'(1 - p)$$

Since r' cannot be greater than unity nor s' less than zero, it follows that

$$\begin{aligned} \text{m\AA}x. r' &= \frac{x'}{p}, \text{ and} \\ \text{m\AA}n. s' &= \frac{x' - p}{1 - p} \end{aligned}$$

Thus when p is greater than x' , r' will have a maximum value of less than unity and s' a minimum value of zero; where p is less than x' , r' will have a maximum value of unity and s' a minimum value greater than zero.⁴¹

The above analysis thus provides two independently derived upper limits to the value of r'_i , and two independently derived lower limits to the value of s'_i .⁴² These are

$$\begin{aligned} r'_i &\leq 1, \text{ and} \\ r'_i &\leq \frac{x'}{p}; \text{ in addition} \\ s'_i &\geq 0 \\ s'_i &\geq x' - p/1 - p \end{aligned}$$

⁴¹ Both the above constraints are ineffective when x is approximately equal to p . In such cases the value of (m\AA x. r' - m\AA n. s') will be approximately unity, and the value of m\AA x. $r^2_{x,p}$ will also be close to unity.

⁴² Estimates of m\AA x. r'_i and m\AA n. s'_i can also be obtained from Tchebycheff's Inequality,⁴³ but the procedure is more complicated and the results are much the same. Given the true mean and standard deviation of any distribution, Tchebycheff's Inequality states that no more than $1/k^2$ cases will be beyond k standard deviations from the mean, regardless of the shape of the distribution. Turning this theorem around, the pro-

An estimate of the maximum correlation between X and P can evidently be obtained from the above maximum or minimum estimates, since

$$\text{m}\hat{\text{a}}\text{x. } r^2_{X,P} = (\text{m}\hat{\text{a}}\text{x. } r'_i - \text{m}\hat{\text{i}}\text{n. } s'_i)^2 \frac{p(1-p)}{x(1-x)}$$

These estimates (of $\text{m}\hat{\text{a}}\text{x. } r^2_{X,P}$) constitute the proportion of variance in purchases that could conceivably be explained by buying intentions if unforeseen events are given the maximum possible weight; the residual variance $(1 - \text{m}\hat{\text{a}}\text{x. } r^2_{X,P})$ must therefore be an estimate of the minimum unexplained variation in purchases attributable to the use of a dichotomous (1,0) buying-intentions variable, rather than a continuous distribution of *ex ante* probabilities, in making predictions. Table 18 summarizes these estimates for each of the thirteen commodities and for several of the buying-intentions questions; the observed proportion of the variance in X explained by P is also shown.

The data suggest that, although unforeseen events may have served to reduce the purchases-intentions correlation very considerably, any reasonable estimate of their importance still leaves a good deal of variation in purchases that cannot be explained by a dichotomous intentions variable. In very few cases does it appear that much more than two-thirds of the total variance in purchases could have been explained by intentions, making an extremely generous allowance for the part of total unexplained variance due to unforeseen events. I conclude that there is a considerable potential for improvement if a survey can be designed that will yield more information about the probability distribution. To what degree this potential can be realized is, of course, a question on which the available data shed no light at all.

portion of cases equal to $1/k^2$ can be no more than k standard deviations from the mean. Defining p as $1/k^2$, and assuming that the mean of the *ex ante* distribution, x' , is equal to the observed purchase rate (x), it follows that p cases can be no more than the distance $x + \sigma'/\sqrt{p}$. But this is the same as saying that

$$\text{m}\hat{\text{a}}\text{x. } C_i = x + \sigma'/\sqrt{p}$$

Given that $\sigma' = (r' - s') \sqrt{p(1-p)}$ (see above, p. 53), and that $s' = \frac{x' - pr'}{1-p}$ (see above, p. 52), and assuming that the distribution above C_i is rectangular, it turns out that

$$\text{m}\hat{\text{a}}\text{x. } r'_i = \frac{1/2(1 + x' - x' \sqrt{1-p})}{1 - 1/2 \sqrt{1-p}}$$

Given $\text{m}\hat{\text{a}}\text{x. } r'_i$, and $\text{m}\hat{\text{i}}\text{n. } s'_i$, $\text{m}\hat{\text{a}}\text{x. } r^2_{X,P}$ can be readily estimated. The distribution free assumption in the Tchebycheff theorem, however, typically provides estimates of $\text{m}\hat{\text{a}}\text{x. } r'_i$ that are much the same as those shown in Table 18.

BUYING INTENTIONS AND PURCHASE PROBABILITY: I

TABLE 18
PROPORTION OF VARIANCE IN PURCHASES EXPLAINED BY
ALTERNATIVE BUYING INTENTIONS QUESTIONS

Commodity	Buying Intentions Questions ^a			
	A ₁	C ₁	D ₁	A ₁ + A ₂
	OBSERVED PROPORTION (r^2_{XP})			
Automobiles	.080	.120	.079	.072
Furniture	.096	.118	.067	.063
Carpets & rugs	.077	.076	.027	.046
High-fidelity equipment	.043	.059	.060	.021
Range	.095	.130	.055	.079
Refrigerator	.098	.084	.089	.083
Washing machine	.055	.114	.068	.049
Television set	.010	.068	.040	.012
Air conditioner	.066	.072	.084	.082
Clothes dryer	.036	.063	.044	.050
Dishwasher	.103	.070	.078	.067
Food freezer	.053	.065	.047	.049
Garbage disposal	.044	.082	.044	.031
	ESTIMATED MAXIMUM PROPORTION ($\text{m}\hat{\text{a}}\text{x. } r^2_{XP}$)			
Automobiles	.373	.854	.753	.523
Furniture	.681	.580	.652	.487
Carpets & rugs	.740	.493	.504	.399
High-fidelity equipment	.642	.491	.500	.307
Range	.486	.735	.781	.598
Refrigerator	.432	.679	.892	.522
Washing machine	.421	.792	.685	.601
Television set	.459	.710	.606	.424
Air conditioner	.376	.725	.560	.427
Clothes dryer	.596	.551	.578	.374
Dishwasher	.725	.615	.610	.328
Food freezer	.584	.722	.506	.309
Garbage disposal	.647	.665	.553	.436

SOURCE: Data in upper panel derived from Chapter 2; data in lower panel estimated according to procedures described in the text.

^a For a discussion of the intentions questions, see Chapter 2.

^b The procedures for estimating $\text{m}\hat{\text{a}}\text{x. } r^2_{XP}$ are described in the text. An illustration is as follows: For question A₁—definite intentions to buy automobiles within twelve months— $x = 0.191$ and $\hat{p} = 0.081$. Hence the appropriate estimate of $\text{m}\hat{\text{a}}\text{x. } r'_i$ is unity, since the alternative estimate— $\text{m}\hat{\text{a}}\text{x. } r'_i = \frac{x}{\hat{p}}$ —yields a value of r'_i greater than unity. The appropriate estimate of $\text{m}\hat{\text{i}}\text{n. } s'_i$ is

$$\text{m}\hat{\text{i}}\text{n. } s_i = x' - \hat{p}/1 - \hat{p} = 0.110/0.919 = 0.120.$$

Since

$$\begin{aligned} \text{m}\hat{\text{a}}\text{x. } r^2_{XP} &= (\text{m}\hat{\text{a}}\text{x. } r'_i - \text{m}\hat{\text{i}}\text{n. } s'_i)^2 \hat{p}(1 - \hat{p})/x(1 - x), \\ \text{m}\hat{\text{a}}\text{x. } r^2_{XP} &= (1.0 - .120)^2 .081(.919)/.191(.809); \\ \text{m}\hat{\text{a}}\text{x. } r^2_{XP} &= .373. \end{aligned}$$