ESTIMATING DEMAND CURVES
THE CASE OF MAIZE IN NEPAL

*Neal P. Cohen

Objectives

Since there are different reasons for estimating demand curves this paper will attempt to make clear the difference in the reasons and the implications of those differences. Further since different approaches are necessary depending on whether we are working with time series or cross sectional data this paper will illustrate the differences. This last part of the paper will deal with techniques that have been used to estimate demand curves. This paper is not designed for those with extensive mathematical or statistical background.

To make the information more readily understandable and to provide illustrations of the methods used, we shall illustrate the techniques using time series data for maize production over the past eight years in the Kingdom of Nepal.

---

* Dr. Neal P. Cohen is a fullbright lecturer in Economics, Economics Instruction Committee, Tribhuvan University, Kirtipur.

The major problem with estimating demand curves

Usually people who begin to estimate demand curves use data on the total amount of a product which is purchased and the prices over a period of time. However, since in economics the amount demanded and the price are jointly determined by the interaction of both supply and demand curves, this technique will not allow us to correctly isolate the demand curve. Simply, take the situation in diagram I. Price and quantity in period one are determined by the intersection of supply and demand curves in that period (SS_1 and DD_1). If both curves shift to SS_2 and DD_2 (possibly due to a decrease in the price of inputs and an increase in incomes), then the new price and quantity that are observed are P_2 and Q_2. This would continue if supply and demand continue to shift to SS_3 and DD_3, and SS_4 and DD_4. What we observe when we collect data is the intersection points, that is, P_1 and Q_1, P_2 and Q_2, P_3 and Q_3 and finally P_4 and Q_4. A line joining these four points is neither the demand curve nor the supply curve, but rather it is a sales curve over time. If however the situation is as in diagram II where only the demand curve is shifting over time, then the curve that is actually illustrated by the price and quantity information that is collected is the supply curve. Finally in diagram III we have that the demand curve stays still and the supply curve shifts; in this case the data actually will illustrate the demand curve.
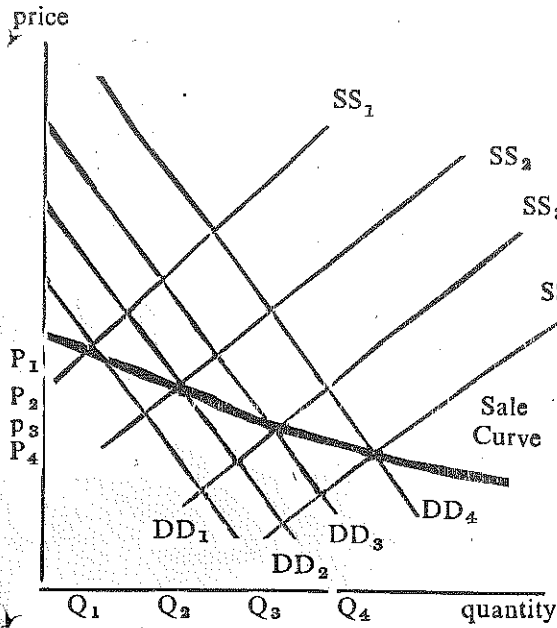
Supply and Demand Diagrams

Diagram I

price
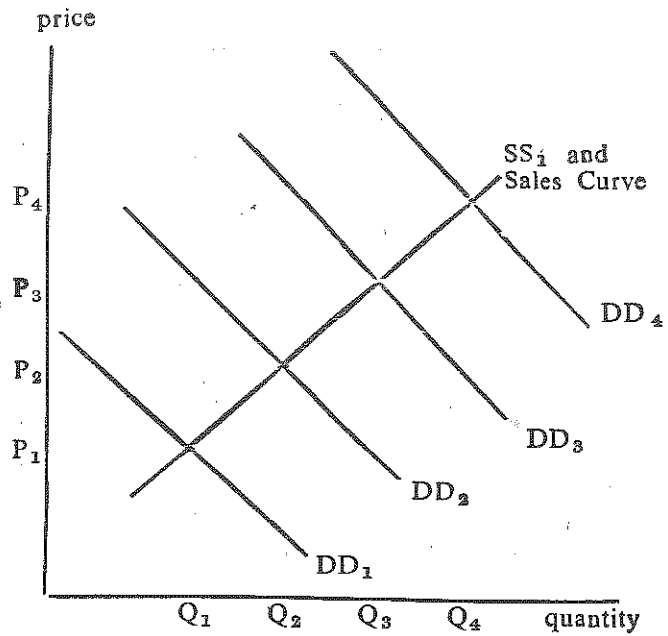

SS₁
SS₂
SS₃
SS₄
Sale Curve
DD₁
DD₂
DD₃
DD₄

P₁ P₂ p₃ P₄

Q₁ Q₂ Q₃ Q₄   quantity

Diagram II

price

SS₁ and Sales Curve
DD₄
DD₃
DD₂
DD₁

P₄ P₃ P₂ P₁

Q₁ Q₂ Q₃ Q₄   quantity

Diagram III

price

SS₁
SS₂
SS₃
SS₄
DD₁ and Sales Curve

P₁ P₂ P₃ P₄

Q₁ Q₂ Q₃ Q₄   quantity

Diagram IV

price

SS₁
SS₂
SS₃
SS₄
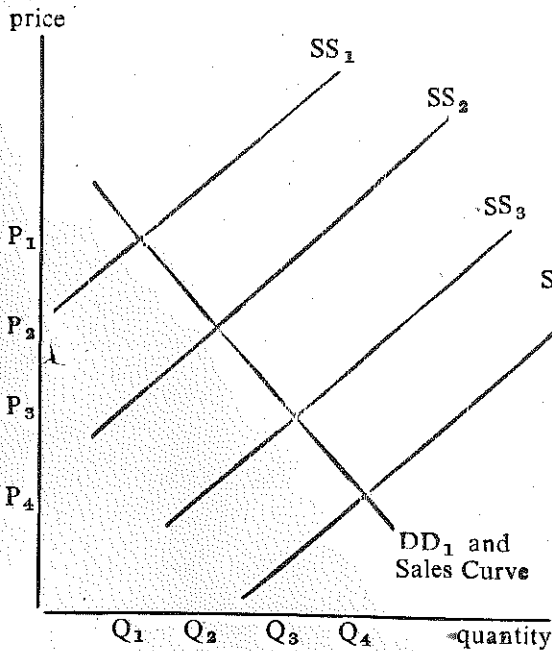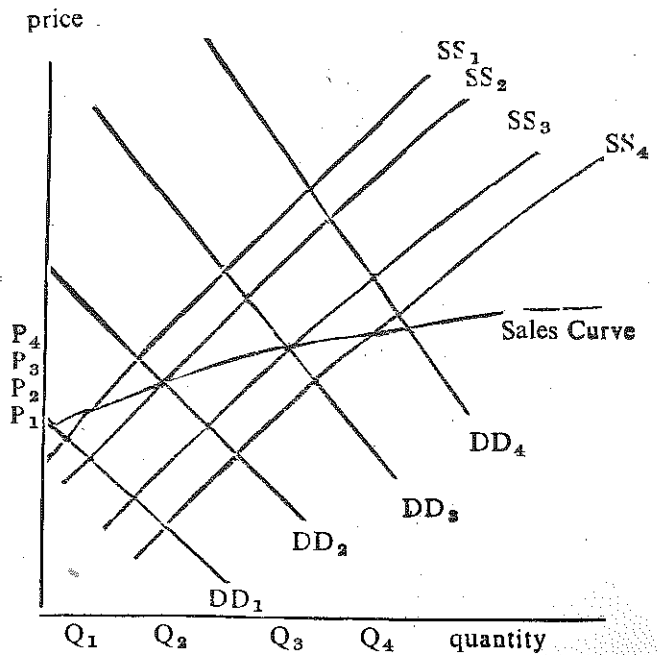Sales Curve
DD₄
DD₃
DD₂
DD₁

P₄ P₃ P₂ P₁

Q₁ Q₂ Q₃ Q₄   quantity

Thus, simply using price and quantity sold (or bought) will not necessarily yield a demand curve, unless we can assume that the demand curve has not shifted over time while the supply curve has. This assumption is so frequently violated that we shall have to find some techniques that will allow us to estimate a demand curve when the assumption of its being stationary does not hold. Further, note that if both the supply and demand curves are shifting rightwards than the slope of the sales curve is less negative than the slope of the demand curve, and in fact it can be positive (this will happen if the demand shifts are greater than the supply shifts, see diagram IV). If we use the data presented in table I for maize in Nepal we find that a simple regression such that $Q=f(P)$ yields a slope of $+0.000485$. This indicates that as the price increases the quantity sold increases. This is unacceptable for a demand curve. (Except possibly for status goods where the desire for a good is contingent on a high price, or for a "Giffin" good, and inferior good whose consumption is basic to daily life, neither is the case for maize in Nepal.) What has happened in the case of maize is that the situation illustrated by diagram IV is presently true, demand appears to be increasing faster than is supply.

Table I

Year	Price maize (indexed)	General price Index	Production maize in 1000 metric tons	Population in millions	GDP in constant rupees (indexed)
Col. 1	Col. 2	Col. 3	Col. 4	Col. 5	Col. 6
2026	100.0	100.0	7.65	10.98	100.0
2027	108.6	110.6	7.95	11.23	102.6
2028	107.5	116.0	8.33	11.56	101.4
2029	119.4	117.5	7.59	11.81	104 5
2030	152.7	130.7	8.22	12.06	103.9
2031	152.7	151.9	8.14	12.32	110.5
2032	175.3	177.5	8.27	12.57	114.3
2033	183.9	184.1	7.48	12 86	118.1

Sources: maize prices (Department of Food and Agricultural Marketing Services), General price index (Rastra Bank), maize production (Agricultural Marketing

Services Department), Population (International Financial Statistics), constant GDP (Central Bureau of Statistics), price of fertilizers–not reported here ("Agricultural Marketing Information".)

Is there anything wrong with estimating sales curves

Strictly speaking there is nothing wrong with estimating sales curves, however, they do not provide the information we want when we estimate demand curves. Usually we are trying to find out whether there will be sufficient supply available to meet the demand. If there is not sufficient supply then (1) there will be pressure on prices to increase, or some non-price method of allocation will have to be used, or some otherway of relieving the pressure of demand will have to be found; (2) we shall have to import the item to make up for the deficit or (3) we shall have to try to increase production of the item by increasing investment, capacity utilization or efficiency. Thus we essentially want the demand curve to find out whether there shall be a planning problem or not, to find out whethere will be pressure on domestic production, prices or the balance of payments.

We often engage in demand curve estimation to enhance our comprehension of the operation of the market and our understanding of economics. sales curve estimation does not help.

Purposes of demand curve estimation

There are two approaches to estimation. The first simply wants to obtain the most accurate possible forecast of demand for a product in a specific year, or in a specific location. In statistical terms, it wants to be able to axplain as much of the variation (or variance) in quantity as is possible using whatever variables will assist in reducing the amount of variation in the dependent variable (quantity) that is unexplained. Thus we decide to include an additional explanatory variable simply because it works, that is, it assists in increasing the proportion of the variation explained. We do not ask whether the coefficients derived make sense, but simply does the variable contribute to our goal of a high r^2 (or $\bar{r}^2$),

With this approach we do not necessarily even need to use any statistical techniques. Anything that appears to yield acceptable forecasts can be used. This could be astrology, seers, folk methods. All we are interested in with this approach is that the forecasts have been correct in the past, and we have confidence that they will be correct in the future.

The second approach to forecasting wants that the forecasts are not only reasonably accurate but that the method used to get the forecasts is "scientifically acceptable, that is, it is methodologically sound. Further we want that the individual elements to the forecast make economic sense, and not just the final total forecast of demand. Not only is the proportion of variation explained reasonably high, but the estimates of price, income and cross elasticity make economic sense, they are of the correct sign and magnitude. In the example presented earlier the sign is not acceptable, since we expect for the demand curve, that as the price of maize increases that the demand for maize will decrease. If the sign were negative than it might be acceptable in that it would indicate that consumers of the product are very slightly responsive to changes in price. Basically it would indicate that consumers want so many kilos of maize and within the observed price range, price had very little impact on their decision. For a necessity like maize we would be surprised to find that consumers were very price responsive.

Similarly, when we are interested in the coefficients we may feel that income plays a significant role in quantity demanded. In general we would expect, that as incomes increase the quantity demanded will increase. Thus if we notice a negative sign, then we would have to try to explain it (possibly we are dealing with an inferior good, that is, one whose demand is inversely related to income). We would not expect that people are very income responsive to maize. If incomes increase by 50% we would expect that the amount of maize purchase to increase by some percentage less than 50%. We do not expect that for most food items that a 50% increase in income will increase the amount purchased by, say, 75%. Thus, this second approach is willing to accept a trade off whereby we have a lower proportion of the variation being explained but in return there are acceptable coefficients.

With the second approach we can better forecast the effect on quantity of just a change in price or income. The first approach may yield a better overall forecast

allowing all the variables to change; but it will usually not yeild as good a forecast if only one variable changes.

Constant Marginal propensity or Constant elasticity models

Since we are usually interested in the values of the coefficients and the techniques that are necessary to obtain acceptable coefficients are more difficult, we shall almost exclusively discuss this approach in the remaining portion of this article.

Usually the information that we want from such a model is the price, income or cross elasticity of demand. These are the percentage change in the quantity demanded of maize divided by the percentage change in price (or income or the price of some other good).

$$\frac{\% \text{ change in Q}}{\% \text{ change in P}} = \frac{\text{the change in Q}}{Q} \Big/ \frac{\text{the change in P}}{P} = \text{elasticity}$$

We usually make the assumption that this elasticity is constant, that is, whether a 10% price increase occurs when the price is low or when the price is high the result will still be a, say, 3% decrease in quantity demanded. Thus if the price increase from Rs. 1.00 to Rs. 1.10 (an increase of 10%), or from Rs. 5.00 to Rs. 5.50 (an increase of 10%), that quantity will decrease by 3%. Thus in the first instance this could be from 300 units to 291, and in the second from 200 to 194 units (both are decreases of 3%). In this case the elasticity is $-3.0\%/+10.0\% = -0.3$ Necessities usually have demand elasticities between -1.0 and 0.0. The more general the nature of the good the closer will be the elasticity to 0.0. Thus the demand for wheat will be closer to zero than the demand for bread, Which in turn will be closer to zero than the demand for a specific brand of bread. Basic industrial production items, such as cement, steel or bricks will also exhibit a demand elasticity close to zero and similarly the more specific the item the less will be its elasticity, the more negative it will be.

Income elasticity can be either positive or negative and depending on whether its value, without a sign, is greater or less than 1.0 it will be elastic or inelastic (respectively). This is used to determine normal and inferior goods.

The cross elasticity is used to determine whether another good is competitive or a substitute (cross elasticity is positive), or whether it is a compliment (cross elasticity negative).

This first approach assumes that these elasticities are constant over the relevant range of values being studied.

The alternative approach is to assume that there is a constant marginal propensity. This would say that a Rs. 1.00 increase in price will always decrease the quantity demanded by 5 units (as an example). Thus if the price changes from Rs. 1.00 to Rs 2.00 then quantity will change from 300 to 295 units, if price changes from Rs. 5.00 to Rs. 6.00 then quantity demandend may change from 200 to 195. In the first case there was $1\frac{1}{3}\%$ decrease in quantity due to a 100% increase in price, or an elasticity of $-1.666/100 = -0.0166$. In the second case a 2 5% decrease in quantity due to a 20% increase in price or an elasticity of -0 125.

Usually the first model which assumed constant elasticities is preferred for supply and demand curve estimation. While it is a little more difficult to estimate, it is sufficiently simply to warrent the additional work.

For the second model, constant marginal propensities, we would regress quantity as a function of price. For the First model, constant elasticities, we convert both the quantity figures and the price figures to logs and estimate the regression as the log of quantity is a function of the log of price. The coefficient on price is the elasticity.

Earlier we estimated the sales curve for maize using the untransformed data, i. e. Q and P and derived

$$Q = 7.887 + 0.000485 \ (p)$$

To estimate price elasticity using this approach we simply take the coefficient in front of the price variable times P/Q. The price elasticity for 2033 would be $0.000485 \times (183.9/7.48) = 0$ 0119. The elasticity for such an equation is usually estimated at the mid points, that is at the average values of P and Q, thus the average price elasticity is

0.000485 × (137.5125/7.95375) = 0.00839.   Evaluating the elasticity for each year we derive the following:

Yearly elasticities using constant marginal propensity model

Year	Elasticity
2026	0.0063
2027	0.0066
2028	0.0063
2029	0.0076
2030	0.0090
2031	0.0091
2032	0.0103
2033	0.0119

Thus this model shows that elasticity is gradually becoming more positive. But it assumes that the marginal propensity is constant at 0.000485 and for each increase in the maize price index of one, that quantity will increase by 0.000485 thousand metric tons.

If we transform all the Q's and P's into logs and then do the regression we derive

$$\log Q = 0.8731 + 0.01275 \; (\log P)$$

The price elasticity of demand is a constant 0.01275.

Unless there is a good reason to assume the constant marginal propensity model it is usual within economics to assume the constant elasticity model and to work fully in logs. (It makes no difference to the elasticity figure whether we work in common or natural logs, in this paper we use the common, or base 10 logs)

What variables ought to be included in the model ?

We can only use as variables items whose value varies, we could not include as a variable the price of a product whose price has not changed. It is not a variable. Thus the first rule is to include items whose value varies; very little variation in a variable frequently results in insignificant coefficients, that is, coefficients where we are not sure whether the sign is correct.

We ought to try to include whatever variable we think may effect demand. The demand for maize is probably a function of the price of maize, incomes in Nepal, price of substitutes, and population, amongst other variables. The demand for petrol would be a function of price, income, number of vehicles within the Kingdom. When dealing with time series data we frequently find it advisable to include "time" as a variable (it is included as an unlogged variable). The reason for including this variable is that for some items there is a learning process which increases over time, thus regardless of any change in population, price or income, the demand for certain products is increasing simply because people are becoming more aware of the items and its uses.

Dummy variables are frequently found to be necessary is demand analysis. A dummy variable takes on a value of zero or one (there is no need to log this variable, especially since the log of zero is undefined). This is saying that there are some other variables which cannot be quantified but which are felt to be important. Thus tourism Nepal would be a function of a number of variables and possibly also a dummy variable which would take on a value of 1 only during the years of Indo-Pakistani wars. Regardless of changes in the cost of coming to Nepal or incomes elsewhere, when there is a conflict in the neighbouring countries this will effect Nepali tourism. If we are dealing with cross sectional deta then we may feel that a variable that will effect amount of biscuits consumed within any one household will be whether the household lives near a regular distribution point or not. Thus those households which are near a regular distribution point would have a dummy variable set at one: and the others would have it set at zero. The rationale is that if an item is regularly available you are more likely to consume it. Two families with the same income, same number of members and facing the same prices would consume different amounts simply

because the first family lives closer to a regular source of the item. The dummy variable measures this closeness. We should only use dummy variables when there are no other variables available which can takes on values wider and more diverse than zero and one.

When we are forecasting using our model we can only forecast within a limited range: if the price has ranged from Rs. 1 to Rs. 7, then we cannot, with much accuracy, forecast what will happen if the price rises to Rs. 20. The further away the price is from the midpoint of the observed prices, the less accurate is the forecast. Similarly, if we have observed data during a period when there were no export controls, then we will predict incorrectly when there are export controls. If one of our variables is a dummy indicating a war period, then we can predict the effect of a comparable war fought when incomes and prices are different. If the war is markedly different then earlier wars then the forecast will still be incorrect since the dummy could only take on a value of zero and one, and was unable to take into consideration the intensity of the war.

Time series or cross sectional data

Time series data are data that have been collected over a period of time (days, months or pears). Thus the data we presented in table I were time series since they reported quantity and prices over the period from 2026 to 2033. These data can be aggregated for an entire country (as ours), or be for one zone, district, panchayat or even for one family. That does not change the essential time series nature of the data.

For cross-sectional data we hold time constant and allow other items to vary. Thus when we do a survey, a questionnaire, we try to do it in as short a period of time as possible so that time is not a variable. Similarly, data on the number of tourists into Nepal from different countries in one year are cross sectional data. Time is the same for all the data.

A major problem we often have with cross sectional demand curves is that price is frequently constant. Thus the price of maize is reasonably constant throughout

Kathmandu today (although it may be different from what it was last week, at another time). Thus price cannot be included as a variable. It is sometimes possible to use cross-sectional data collected at different points in time in one regression. This combination of cross sectional and time series data may enable us to estimate the price elasticity.

In a country such as Nepal which is undergoing reasonably rapid changes, the use of time series data from too many years back might invalidate the results. Tourism in Nepal has changed so much that to use data of the past twenty years will probably distort the findings as to the relationship between the variables currently existant.

Data sources:

Time series data are usually the easier to gather because we have to use secondary sources. We utilize data collected by some government agency over the past years, quarters of years, or months. While these data are easy to use there are a number of possible problems with its use.

If one data source gives us data for the last five years and another gives us data for the previous five years we are usually unable to combine the two sources and have data for ten years. The reason is simply that the defination of the variables, the methods of collection might have changed. What is frequently possible, if there is at least one overlapping year, is to index the variable, thus not using the raw figures.

In general, we need to reassure ourselves that the definitions used by the source collecting the data agrees with the defination that we need in the variable we want to use; and that they have been reasonably consistant in their methods of collection. If we use price data of maize from one source and price of millet from another source the problem of comparability of the data arises, especially if we want to use the two prices as reflecting substitutes in the mind of the consumer or producer.

Quite often we are unable to gather the exact data that we would like; in this case we are often able to find a variable that is a reasonably good proxy for what we

would like to have. A proxy variable is defined as a variable that we believe has a very high correlation with a variable we want to use, but are unable to get; or we do not feel that the variable we are able to get is of sufficient accuracy. If we would like to include gross domestic product as a variable and distrust the figures reported we may feel that figures on the production of rice are more accurate, and that the figures are highly correlated with the true GDP. Similarly a general price index may not be available but prices of a few products may exist; we can construct a proxy index that may be a reasonably good approxition of the missing general price index. We may feel that distance from a foreign country to Nepal is a good proxy for the cost of travelling from that country to Nepal. Econometric work has many examples of proxy variables. We need to know when to use and how to select proxy variables.

Cross - sectional data for demand curves can be collected from official sources in which case the same problems listed earlier are applicable, or we can attempt to collect data through a questionnaire. Before we collect the data from a questionnaire, it would be wise to think through all the possible variables that may effect the demand or supply curve and to be sure that we have collected data which will allow all relevant variables to be tested. Thus in an example of demand for maize we would want to include some general variables to occupation, age, education, size of household and income of the household, which do the purchases and coooking in the household. There are possibilities that any of these variables may significantly explain the quantity purchased, and we may want to include them. We collect the data to allow us to have a choice later in the analysis including that variable. Other variables we may want include are: approximate amount purchased in the past week (or month), price paid, what they consider to be the positive and negative characteristics of the type of maize they are purchasing. For some products we could ask people to rank order the product and its substitutes as to overall quality, value, contributions to good health. For supply curves we would also need to collect information on the cost of some major inputs, for maize this would be fertilizers and/or irrigation Other variables will be needed depending on the specific product in which we are interested.

For all quistionnaries we need to be aware of the problems of people misinterpreting the questions we are asking, of being told what people would like us to

think is true, what they wish were true, or what a faulty memory remembers.

In addition to questionnaires we need to insure unbiased results. This is usually done by the process of random (or stratified random) sampling, where every individual has an equal liklihood of being sampled. In this case the results can be generalized to the community being surveyed, without sampling the entire coummunity.

All the problems enumerated in this brief section are solvable, and frequently even if the data reported or collected are not precisely what we would have liked, with some thought we are usually able to retrive some useful information.

Interpretation of results

Before getting into the techniques of doing the regression work, it will be useful to take a result and see how to analyze it. Using one of the techniques discussed later we derived maize demand in Nepal:

$$\log \; Q_d = 1.225 - 0.861 \; (\log \; GDP_c) + 1.674 \; (\log \; Pop.) - 0.178 \; (\log \; P)$$

Where Q_d is the quantity demanded (production minus exports plus imports, in thousands of metric tons), GDP_c is an index of gross domestic product deflated by the general price index, Pop. is population in millions of people and P is the price of maize, indexed.

When all the variables are loged then the coefficients are elasticities. The first thing we notice is that the signs on population and price are correct. The sign on GDP_c presents some problems. It is indicating that as income increases people shift out of maize into some other product. This would tend to indicate that it is an inferior good. However, if we combine this coefficient with the population one then we are able to say that unless real GDP increases more than $1.96(1.674/0.861)$times as fast as the percentage change in population that total demand will still increase. That is, if real income increases by 3%and population increases by 2%, then total demand will still increase by 0.765% (derived by taking 3% times the coefficient on GDP_c and adding it to 2% times coefficient on population, $(3 \times -0.861) + (2 \times 1.764) = 0.765$). Currently population is increasing by 2.16% per year

and real income by 2.65% per year, thus total demand is increasing by 1.33% per year (exclusive of price changes). It appears that total quantity demanded will be increasing in the near future. From this result we can see that there will be pressure for increased production for domestic uses simply with increase in population. For every 1% increase in population there is a 1.674% increase in the demand for maize If Nepal wants to generate any surplus maize for export purposes then it will be necessary to increase production by more than 3.62% (2.16×1.674) in order to generate some surplus for export purposes. (The 3.62% holds unless we also want to assume changes in read GDP and/or price of maize.)

The price variable has a very low coefficient. This is an example of an inelastic good, a doubling of the price of maize will lead to a reduction in the amount purchased by 17.8%. Thus the people who buy maize at present are not particularly concerned about price. They decide on quantity and find that price does not substantially change their decisions. This not because prices have not varied, they have increased by as much as 30% in a year. People must view the price as still sufficiently low that its increase does not effect quantity very substantially, yet.

We could use the results to forecast demand under the conditions where the government wants real income by 3% per year, lower population growth to 2%, and would like to keep prices reasonably stable, not increasing by more than 5% per year. We find that during a five years period the amount demanded will decrease by 0.6%. If supply analysis indicates that supply is constant, then we will not generate much surplus for export purpose.

Techniques of deriving the results

There are many different techniques that have been derived to estimate demand curves since H. L. Moore (Economic Cycles: Their Law and use, Macmillan, New York) estimated in 1914 the demand for pig iron. Using ordinary least squares he found that the price elasticity was positive, indicating that the higher the price, the higher the demand. His problem was the one we analyzed in the second section of this paper on the major problem in estimating demand curves, that is, he estimated a sales

curve where the demand curve was shifting more than the supply curve. Thus the use of ordinary least squares is fraught with danger signs and can only be used in very limited situations. Had we used it for quantity demanded of maize we would have got

$$\log Q_d = 2.055 - 0.934 \ (\log \ GDP_c) + 0.507 \ (\log \ Pop.) + 0.115 \ (\log \ P)$$

which shows population much less elastic than the result shown earlier, but, more importantly, the coefficient on price is the wrong sign. Like Moore we find that the higher the price the more people will purchase maize. This result is so seldom observed in actuality that we can dismiss it.

Method of Henry schultz

Henry Schultz was one of the first to do extensive research on the properties of demand cnrves, and his approach to estimating the demand for agricultural products is still being used. (His book was The Theory and Measurement of Demand, University of Chicago Press, Chicago, 1938.) His method can be summarized in a number of steps.

1) Prices are defined as relative prices, that is, he uses the price of the product divided by a general price index, or by the price of the most important substitute.

2) Quantity is defined as quantity per capita, quantity is divided by population.

3) Take logs of the relative prices and the quantity per capita.

4) The results in step (3) are regressed against time separately, i. e. log (P/GPI) = f (time) and log (Q/Pop) = f (time).

5) Compute the errors from the step (4) regressions, that is, the forecasted values each year less the actual observed values.

6) The errors in the quantity per capita regression are themselves regressed against the errors in the relative price regrression.

Schultz further divided his analysis into different historical periods and did all the work separately for each distinct period.

What this method essentially does is to remove as many of the usually observed shiftors of the demand curve in an attempt to estimate it.

Using our data we shall follow Schultz's method step-by-step and then look at the results.

1) We divide the index of price of maize by the index of the Kathmandu consummer prices, i. e. 100/100=1.00, 108.6/110.6=0.983; 107.5/116.0=0.927, etc. and then multiply the results by 100 to develop a new index number (col. 2 divided by col. 3 of table 1 equals col. 2 of table 2).

2) We divide quantity by population and get 7.65/10.98=0.697; 7.95/11.23=0.708; 8 33/11.56=0.721, etc. (col. 4 divided by col. 5 of table 1 equals col. 3 of table 2).

3) We take logs of the results and thus get for relative prices that the log of 100 is 2.000, the log of 98.3 is 1.9921, the log of 92.7 is 1.9969, etc. The logs of quantity per capita are: of 0.697 it is -0.1569, of 0.708 the log is -0.1500, etc. (the results are in col. 4 and 5 of table 2).

4) We regress col. 4 as a function of col 2; and col. 5 as a function of col 1 of table 2. The results are:

log (P/GPI)=1.9964+0.0021 (time)

log (Q/Pop)=-0.1425-0.0095 (time)

5) We find the forecasted value for relative prices by finding estimated relative price when t=0, t=1, t=2 etc. and substract from this figure the actual relative price log of 2.000. Thus, the estimated relative price log is $1.9964+0.0021\times0=1.9964$. From this we substract 2.000 and derive-0.0036 which is the error and is reported in col. 6. We do the same for quantity per capita. The forecast for 2026 would be $-.1425 - 0.0095\times0 = -.1425$, and we substract $-.1569$ and get$+0.145$ which is

reported in col. 7 of table 2. For 2027 we change the times 0 ($\times$0) in the above two examples to times one ($\times$1) and continue the work.

6) We regress the results in col. 7 against the results in col. 6 and derive.

$$Qpc_{errors} = 0 + 0.0152 \,(\, P/GPI_{errors} \,)$$

Since the sum of the errors equals zero there is no constant term. The 0.0152 is Schultz's estimate the maize elasticity of demand.

Table – 2

Data donversions for demand estimation Schultz's method

Year	coded year col. 1	P/GPI col. 2	Q/Pop col. 3	log P/GPI col. 4	log Q/Pop col. 5	errors P/GPI col. 6	errors Q/Pop col. 7
2026	0	100	.6967	2.0000	−.1569	−.0036	.0145
2027	1	98.19	.7079	1.9921	−.1500	.0064	−.0019
2028	2	92.67	.7206	1.9669	−.1423	.0337	−.0191
2029	3	101.62	.6427	2.0070	−.1920	−.0043	.0211
2030	4	116.83	.6816	2.0676	−.1665	−.0628	−.0139
2031	5	100.53	.6607	2.0023	−.1800	.0046	−.0098
2032	6	98.76	.6579	1.9946	−.1818	.0144	−.0175
2033	7	99.89	.5816	1.9995	−.2353	.0115	.0266

While the mathematics of Schultz's method is extensive, it is also reasonably simple. Its assumptions tend to be very easily violated when dealing with non–farm output and thus it ought not be used in those situations. In situations where the supply of the product is somewhat less than perfectly inelastic, that is, there is some

farmer responsiveness to price, then this model begins to yield spurious results. In fact, the greater is the price responsiveness of farmers the more likely the estimated demand responsiveness to price to be positive, or less negative than it actually is. This is the case here. The estimated coefficient is too high. It does not take a very high price elasticity of supply to make the demand elasticity of supply truly negative, in our expample.

In our case we find that the elasticity is a positive one, but very inelastic. Since the r^2 (proportion of the variation explained) was a dismal 0.0234, we have very little confidence in the sign, or in its magnitude. What the result indicates is that Schultz's method finds that there is very little price responsiveness of buyers of maize. The demand is very unresponsive to price. A positive response to price is highly unlikely, thus leading the way for us to conclude that there are probably other variables that ought to have been included in the analysis.

Schultz's method ought to be used for those agricultural goods where there is is a strong trend in production and consumption and where the major factor controlling output is not price, bus availability of fertilizers, seeds, or climatic conditions.

Fox's method of estimating price elasticity

This method, which has found favour within the U. S. Department of Agriculture, begins by assuming that quantity is a predetermined variable; that it does not respond to price. The causation runs almost exclusively from changes in quantity to changes in price, and not vice-versa. Nor does this model allow for two—way causation, where price effects and is effected by quantity. Thus this method re-arranges the regression to make price the dependent variable, and quantity (along with whatever other variables are considered important) as the independent variable.

All the variables are expressed in logs as he is working with the costant elasticity form of the demand curve.

We estimate the price responsive model as:

$$\log P = f (\log Q, \log Gdp_c, \log Pop)$$

Taking logs of the data presented in table I we get as our result

$$\log P = -3.4883 + 0.3586(\log Q + 0.8616)(\log GDP_c) + 3.2950(\log Pop)$$

$$(1.82)^{**} \quad (0.45) \qquad (0.58) \qquad\qquad (2.01)^{**}$$

$$R^2 = .938 \quad F = 20.01^{***}$$

(The t – values are in brackets under the coefficients, the stars refer to the significance of the relevant statistics, one star indicates significant at the 20%, two at the 10% and three is significant at the 5% level; we do not perform and significance test for the R^2.)

Since the demand curve needs to have log Q on the left hand side we need to transform the above result, and get

$$\log Q = 9.7276 + 2.7886 (\log P) - 2.4027 (\log GDP_c) - 9.1885 (\log Pop)$$

From the first equation we note that both increases in real income and increases in population put pressure on prices to rise; a realistic assumtion. In fact, the yearly 2.16% insrease in population is putting pressure on maize prices to rise by 7.12% per year (3.295 times 2.16). The positive sign for real income indicates that as incomes increase there will be further pressure on prices to rise and the faster Nepali real income increases the greater will be that pressure. The positive quantity elasticity indicates that as production increases price will increase. As this is highly unlikely we have to conclude that either we face a situation of missing variable or that this method does not work well for maize consumption in Nepal. The letter appears to be true since we ought to assume that there is some degree of responsiveness by the farmers to price, we cannot make Fox' assumption that quantity is predetermined or that other variables than price are paramount for the farmers, As price rises the farmers will respond somewhat by planting more maize.

Two stage least squares, instrumental Estimates

It is best to approach the estimation of a demand curve from the viewpoint of an entire economic model, that is

$$Q_s = f(P) \qquad\qquad Q_d = f(P) \qquad\qquad Q_s = Q_d$$

Further we ought not assume that anything is predetermined; that farmers will respond to price and consumers will respond to price, and that price itself will respond to the quantity available. As there is more production there ought to be pressure on prices to fall, if there is a shortage there will be pressure on prices to rise.

Regression theory assumes that there is no correlation between the independent variable (P in the above equations) and the error term. However, from economic theory we know that if the price is too high supply will be excessive and demand too small, or there will be a positive correlation and a negative correlation respectively in the two situations. Thus there is a need for some way to purge the price variable of the correlation with the error term.

The method used by two stage least squares (2SLS) is not to use price as an independent variable directly, but to use estimates of price. Simply, in the first stage we regress price against the exogenous or predetermined variables in the model and derive the forecasted values of price given the values of the predetermined variables. This price is then used in the second stage.

The model we will try to estimate is

$$Q_s = f(P, P_{-1} \text{ and } P_{fert})$$

$$Q_d = f(P, GDP_c, \text{ and } Pop)$$

$$Q_s = Q_d$$

Where P_{-1} is last year's maize price, P_{fert} is the price of fertilizers, indexed). In the model we do not try to estimate GDP_c, population or the price of fertilizers. They

are exogenous and are not substantially effected by the price and quantity of maize.

First stage,

We do the regression,

$$P = f(P_{fert}, GDP_c \text{ and Pop})$$

and derive,

$$P = 98.565 + 0.6084(P_{fert}) - 2.1800(GDP_c) + 32.3064(pop)$$

Using the value of the price of fertilizers, the real income and the population in 026 we derive the estimated, forecasted, price in that year. This is reported below as P_{est}.

Table – 3

Estimates of price derived in the first stage of 2SLS

year	Past estimated price	Past actual price
026	99.00	100.0
027	101.41	108.6
028	115.41	107.5
029	131.21	119.4
030	140.60	152.7
031	154.26	152.7
032	174.49	175.3
033	183.73	183.9

If the model does not include exogenous variables then we need to choose instruments which are correlated with price, but are not correlated with the error term. The closer our instruments are to meeting these conditions the better they will be and the truer will be our final coefficients.

Second stage

We now do the regressions we are interested in but substitute price estimates, P_{est} for actual price, that is

$$Q_s = f(P_{est}, P_{-1}, P_{fert}) \text{ Supply}$$

$$Q_d = (P_{est}, GDP_c, Pop) \quad \text{Demand}$$

As before we shall use the constant elasticity model and convert all the raw data into logs.

Our results:

$$\log Q = .761 + .360 (\log P_{est}) + .003 (\log P_{-1}) - .298 (\log P_{fert})$$
$$(2.47)^{***} \qquad\qquad (.02) \qquad\qquad (.88)$$
$$(1.02)^{*}$$

$$R^2 = .256, \qquad F = .458$$

Since the coefficient on lagged prices is so insignificant it is wise to drop it. Our results thus become

$$\log Q = .765 + 0.358 (\log P_{est}) - .295 (\log P_{fert})$$
$$(4.30)^{***} \qquad\qquad (1.31)^{*}$$
$$(1.28)^{*}$$

$$R^2 = .256 \qquad F = .859$$

For the demamd curve

$$\log Q = 1.225 - .178 (\log P_{est}) - .861 (\log GDP_c) + 1.674 (\log Pop)$$
$$(.58) \qquad (.24) \qquad (1.03)^{*} \qquad\qquad (.55)$$

$$R^2 = .245 \qquad F = .433$$

All of the signs in the above equations make economic sense and the magnitudes are fully acceptable. For the supply curve we find that the farmers are responsive to price, but there is still an inelastic response, that is, a 10% increase in price yields a less that 10% (3.5% to be exact) response in production. Further as the prices of fertilizers increase the farmers will decrease the amount of maize produced, presumably switching to products that require less fertilizers, or different types. Note that in the first stage we did find that an increase in the price of fertilizers will have a positive effect on the price of maize. The equation also indicates that farmers are not particularly responsive to last year's price, thus the posibility of a cobweb is lessened and the role for reasonably quick government action increased. Had the farmers strongly reacted this year to last year's price than it may take two years for a change in government policy to become effective.

In the demand curve we also note that demand is price inelastic, but for the first time the sign is the correct one, negative. Thus a 10% increase in price will result in a 1.78% decrease in demand. Maize is the type of good where price is not the prime consideration. Rather we note that as real income per capita increases slightly demand still rises (the coefficient on log Pop is greater than the coefficient on log GDP_c), but if the rise in real income is more than 1.94 times as great as the percentage change in population, then total demand for maize will drop. If real income alone increases total demand for maize will drop, thus indicating that people will switch out of maize into other products. It would be viewed as an inferior good. However, there will still be increasing total demand because of the elasticity with respect to population.

Ordinary multiple linear regression

This method is statistically unsound, except under rare circumstances presented in diagram III. Had we done it using the above variables, we would have got.

$$\log Q = .836 + .271 (\log P) - .037 (\log_{-1}) - 205 (\log P_{fert}) \text{ supply}$$

$$\log Q = 2.005 + .115 (\log P) - .934 (\log Gdp_c) + .507 (\log Pop) \text{ demand}$$

This formulation shows that there is less pressure on demand due to population, but the sign in the demand curve on prices is unacceptable.

Serial correlation, Autocorrelation

This is a problem that we frequently confront when working with time series data. It will not be a problem when we work with cross sectional data. A problem of serial correlation exists when the forecasting errors are correlated with each other. If knowing there was a positive error last year, I am able to accurately predict a positive error this year then we have a problem of positive serial correlation. A positive error last year leading to a prediction of a negative error this year is an example of negative serial correlation (but this is very seldom observed). The Durbin–Watson is the most popular statistic used for checking for serial correlation. We first find all the errors.

Table IV

Forecasting errors 2SLS

Year	Supply errors	Demand errors
026	.0068	.0061
027	−.0077	−.0057
028	−.0089	−.0105
029	.0244	.0242
030	.0005	.0017
031	−.0103	−.0088
032	−.0228	−.0232
033	.0201	.0207

The Durbin–Watson formula is

$$DW = \frac{\Sigma(e_t - e_{-1})^2}{\Sigma e_t^2}$$

For supply the DW is 2.35 for demand it is 2.27. Tables for significant points for DW indicate that this does not show any cignificant positive or negative serial correlation.

If we had found significant serial correlation the procedure for correcting it can be tedious. The simplest method is to perform a regression

$$E_t = f(e_{-1})$$

on which we would get for the supply curve

$$e'_t = -.003 - .244 (e._1)$$

(the larger the number of observations the closer will the constant term be to zero)

We now have to transform ALL the variables in the second stage thusly

$$\overline{(\log Q)} = (\log Q)_t - (-.244)(\log Q)_{-1}$$

The regression we actually estimate becomes

$$\overline{\log Q} = f(\overline{\log P_{est}}, \overline{\log P_{-1}}, \overline{\log P_{fert}})$$

After the regression coefficients are derived we can easily convert back to log Q, log P_{est}, etc, form by substituting what $\overline{\log Q}$, etc, are equal to, however, we shall now ha \ additi onal variables in the final regression, namely $Q_{-1}, P_{est-1}, P_{-2}, P_{fert-1}$.

Summary

While 2SLS is the most difficult method it is also statistically the most sound and is most likely to yield coefficients that are unbiased and consistent (that is, as the sample increases, and remaining problems will tend to disappear). Thus, when we are interested in the coefficients and want to draw information from them, 2SLS ought to be used. Ordinary least squares will yield spurious results. In some situations the

the method of Schultz or Fox can be used. Generally these techniques work well when there is little response of the farmers to price.

It ought to be obvious that we can only forecast when we have included all the relevant variables. Thus we cannot forecast the effect of price changes unless we include the price variable. We cannot forecast the effect of changing word income unless that variable is included. Further, the greater ther distance a forecasted independent variable is from the mean value in the data set, the less accurate is the resulting dependent variable forecast. In our exmple forecasting quantity when the price of maize index stands at 400 would be very hazardous and probably inaccurate. If the price of fertilizers should drop to 50 then we would find our forecasts also inaccurate.

We are unable to forecast the effects ef major changes in the society. Thus we could not predict what would happen to maize consumption in Nepal where India to annex the country. We cannot forecast forward many years. However, this technique will enable us to predict supply and demand forward a few years (at least enough for a five year plan), IF there are no drastic changes in the operation of the economy.

If the sole interest in forecasting is accurancy and the coefficients and the information they yeild is of no inportance than the niceties that this article hae discussed become less important. In this situation ordinary least squares with many explanatory variables (and there is probably no need to log the variables) will work sufficiently well. We find from Fox's method and the first stage of ordinary least squares that we can forecast price well (the proportion of the variation that we can explain is over 90% in both cases). However, in no case we are able to forecast quantity as well. Additional thought needs to be given to determine additional variables that ought to be tried in an attempt to improve our forecasting ability.