AGEC 662 Notes
Summer 1996
B.A. McCarl

Forming Probability Distributions

An important part of any analysis of decision making under stochastic conditions is a probability distribution. These notes talk through some procedures for developing probability distributions as they apply to agricultural economics.

Probability distributions state the relative frequency of occurrence of a set of mutually exclusive events. Probability distributions can be univariate or multi variate. They give the relative frequency of observing a particular event. For example, a univariate distribution could describe the frequency of alternative yields for a crop at a particular point in time. In that case, the mutual exclusivity requirement would indicate that it was impossible to set more than one yield at one particular point in time. Multivariate distributions are needed where the outcomes involve more than one item wherein the level of one item does not predetermine another. For example, a multivariate distribution would be needed to describe the yield on two different plots of ground where the yields across plots are not independent or perfectly correlated. Mutual exclusively still must hold in that one cannot get more than one yield at one point in time on the individual plots of land. Thus, in the example case the probability distribution would express the relative frequency the simultaneous levels of a pair of a yield outcomes on the plots of ground.

Some Alternatives for Probability Distribution Description

There are a number of important alternative paths that can be followed in defining probability distributions. These are univariate-multivariate, continuous-discrete, full-moment based, known versus unknown distribution form, conditional versus unconditional and conditional on discrete or continuous events. Each merits discussion.

Univariate versus Multivariate

The distinction between univariate and multivariate probability distributions depends upon the events space. If the events are entirely described in terms of the value of one outcome or can be transformed to be described in terms of one outcome then a univariate probability distribution is present. If not the probability distribution is multivariate. Notice that having multiple items dependent upon the outcome of an event does not necessary imply a multivariate probability distribution. For example, one may have found that the yields of a number of crops are perfectly correlated across weather states. Thus, one might have multiple yields dependent upon the historical weather year. This is not a multivariate probability distribution, but is really a set of univariate probability distributions which are perfectly correlated across the single event weather.

Continuous-Discrete

An important distinction in terms of probability distributions is whether they are defined with respect to continuous or discrete events. Most theoretical probability distributions are generally developed with respect to a continuous event space. Most empirical probability distributions are developed from data on discrete events. For example, analysts often assume normality and a continuous set of yields. Most historical data is discrete with finite discrete observations. Continuous probability distributions only arise when one assumes or estimates a distributional form.

Full-Distribution Versus Moment Based

Probability distributions may be described either in terms of the moments of the distributions or in terms of the full probability distribution. Also, this often is combined with distributional assumptions. For example, one may use a technique which allows estimation of the mean and variance of the probability distribution, but does not estimate the full probability distribution. In cases, the mean and variance can fully describe the probability distribution for the normal distribution. Common techniques like regression estimate the mean of the probability distribution. There are techniques available such as the Just and Pope technique which estimate higher moments.

Known versus Unknown Distributional Form

There are a lot of applied analyses where discrete "empiricial" distributions are used without assumptions about the form of that distribution. One can use a flexible functional form distribution to approximate without finding the true distribution. Finally, one can fit a particular distributional form or assume for example that everything is normally distributed.

Conditional-Unconditional

An important distinction in applied work is the conditional-unconditional nature of probability distributions. Probability distributions are conditional if one assumes that an external event shifts the probability distribution. Consider the following examples. The current high prices in agriculture are in part a function of low government stocks that in turn make more likely higher profit price distributions. Thus, we could have a distribution of prices with and without government stocks being low. In turn, the distribution of prices would be conditional upon the level of stocks. Similarly, one could develop a distribution of yields for an irrigated crop with the yield distribution a function of water applied.

Ultimately, almost all probability distributions used in applied work are conditional in some form or fashion. For example, almost all distributions are conditional on time giving the distribution now. However, one does not usually expect probability distributions to be temporally stable.

Conditional and Discrete Versus Continuous Conditional Events

A very important distinction in some agricultural economics work is whether the conditional probability distribution is based on continuous or discrete exogenous events. For example, one can develop a probability distribution for prices with and without a NAFTA agreement being signed - a distribution conditional on discrete events. On the other hand, one may need to know how the probability distribution shifts when fertilizer usage changes and thus might want a probability distribution that relates yields and the distribution of yields to the amount of fertilizer that is used. Thus, if fertilizers increases by 5% one might want to know how the parameters of the probability distribution shift.

Formation Approaches for Probability Distributions

Fundamentally probability distributions formation can be approached in one of four different ways. Here we call these four ways: deductive assumption objective, subjective and simulated. These notes will discuss everything concentrate on objective formation of probability distributions. First, however we will overview these methods.

Distribution by Deductive Assumptions

A common methodology in agricultural economics is to assume a probability distribution form and then parameterize in a simple fashion using some data. Fundamentally, one usually assumes normality and then given a mean and variance estimate goes forward with the study. This certainly is common in the vast variety of studies done with expected value variance models. Other distribution assumptions can also be made.

Objective

The second formation method for probability distributions (and the item on which these notes will primarily concentrate) is the development of "objective" probability distributions from historical data. What one does is infer future probabilities of events based on their past frequency of occurance. In this case, one collects data on the variables of interest and attempts to manipulate these data into a probability distribution. One has to be careful when using historical data to make sure the data are all brought up-to-date. For example, using prices from a period of 20 years without adjusting for inflation is usually an improper procedure. Similarly, using data with important trends over time without adjusting for those trends is ordinarily improper.

Subjective Distributions

Distributions can be formed through subjective questioning of the decision makers. This embodies contacts with decision makers asking them to reveal their anticipated probability distribution. There are a number of problems with this, one of which one is the difficulty of accurately asking individuals for the probability distribution and a second is to avoid the probability distribution of being biased by currently observed factors (anchored). For example, during this year where prices are substantially higher than they have been in the past, individuals might give substantially higher probability distribution than the long-run distribution which might exist. There are also substantial problems in getting conditional distributions when doing subjective probability solicitation and in getting accurate distributions that the individuals believe in.

Simulated Probability Distributions

An increasingly more popular way of developing probability distributions is through the use of a simulation model. Such models involve, for example, the EPIC crop simulation model where one can develop a yield and fertilizer use or Richardson's FLIPSIM where one puts in some raw information on distributions of prices and yields and then simulates a net revenue probability distribution. The real difficulty with thte simulation approach is assuring the simulatoin validity generates applicable data.

Finding Probability Distributions Based on Objective Data

The most fundamental topic of these notes involves finding probability distributions based on historical or cross sectional data. Before launching into a discussion of techniques. First, let us talk about some desirable characteristics of a probability distribution.

Desirable Characteristics of Probability Distributions

There are several properties that characterize distributions which must be obtained in order to have a usable probability distribution: 1) each of the states of nature must be mutually exclusive; 2) a probability of occurrence of each of the states of nature must be an unbiased measure of the current probability of that state of nature occurring; 3) the sum of the probabilities across the states of nature must equal one

The second property is the most troubling in when using historical data. Suppose there is a trend in historical data. What if that trend is upward, then what one must do in the phase of the historical data is somehow remove the trend then extrapolate it to the time period of interest and form the probability distribution based on its historical observations in that time period. A very simple example.

The first questions regarding probability distributions is how do we form one from a set of historical data assuming that data comes from a single stationary probability distribution. In such a case, one would employ the following procedure to form the probability distribution.

Step 1 - take the historical data and array it from high to low.

Step 2 - decide if one wishes the probability distribution to just have the individual observations in it or have ranges if individual items assign each observation to a cell and go to Step 4.

Step 3 - if ranges are desiredfind the range size. Take the the high and low observations subtract their difference divide by the number of points to be in the distribution i.e., to find the lateral size if you want 10 points in the composite distribution and the low number was 90 while high number was 180 one could use an range interval of 9. Onee also might want to use an interval of 10 for convience.

Step 4 - go through the data and count the frequency for which the items fall in each range.

Step 5 - Take the range count and compute the relative frequency by divide the frequency in a cell by the total number of observations, this is the probability. Assign the mid point of the interval to be typical of the value.

An example, this procedure is presented in Table 1. The first column gives the observation number, the second column gives the raw item, the third column gives the sorted items and if we look this sort of strange we group this into groups of five starting at 2.5, so our first range is 2.5 to 7.5. Our second range is 7.5 to 12.5 etc. The resultant frequency table up here is in Table 2.

Making a Probability Distribution from a Set of Non-Stationary Data

One key word in the above discussion is stationarity. When one finds that the probability distribution is not stationary than a stationry series needs to be developed from the non-stationary set of data. The most common way of doing it is it utilize a regression to remove the systematic trends and then use the residuals to form the probability distribution. Let us consider a relatively simple ten point example of this. Suppose we use national average cotton yield data from Agricultural Statistics from 1977-1986. The years of data appear on Table 3. Note, the yields are higher at the end than the begging. Thus, a trend exists and the data are likely not stationry. In turn, suppose we run a simple time regression where we predict the data as a simple linear function of time. The regression output is given in panel B of Table 3. In turn, suppose we use the equation to develop a forecast for each of the years as in the third column of Panel A in Table 3. Note, that the year for 1977 the regression forecast 466 whereas the data is really 520. We then calculate the residual error. In the 1977 case the forecast equation has an error where it is 53 units underneath the actual observation. Finally, to form a stationary probability distribution we forecast to a period we don't have data for, in this case 1987. The equation shows a forecast of 610.7, then we add the residuals to each of the observations. So under 1977 residuals the forecast for 1987 would be 664. This final column in Panel A is a stationary series with assuming that the distribution of forecast errors in the past are relevant to the current time period. In turn, given these data we would form the probability distribution as in the last section.

A More Extensive Example of Forming Probability Distributions

One way of looking at the problems encountered in formation of objective probability distribution is to form one. Suppose that we wish to form a total revenue distribution for producers in an area. In this particular case, let us draw strawberry data on prices and yields for California from Agricultural Statistics for the period 1977-93. The data for prices and yields as well as the resultant multiplication of revenue are given in Table 4. The prices are in nominal dollars. Thus, the table also includes the implicit GNP deflator.

Now suppose we are going to construct a probability distribution for revenue but we choose to do this by first constructing separate stationary price and yield distributions. This will be done under the questionable assumption that the California producers are price takers in the strawberry market. (In fact, this may not be the case and this leads to some further discussion later).

Let's start by examining the probability distribution of price. Figure 1 shows the distribution of prices over time and the mean price. We could use these data to form the probability distribution. Here are seventeen years there so we could say there 1/17 chance of each price occurring. Thus, we would get the distribution in Table 5. The reason we should not use this for a probability distribution is there is an obvious bias in the data introduced by the fact that all of our low observations are in the first ten years of the data and all the high observations are in the last ten years of the data. We would try to remove this bias by converting the data from nominal to real prices (Table 6). Figure 2 gives the real price distribution and its mean. Again, is clearly not a data set we should want to use in forming the price probability distribution as now the prices get lower over time.

We might posit an individual looking at the strawberry market would mentally extrapolate the series with a linear time series model where price is equal to a constant plus b times the year. In this case, we estimate the price forecasting the equation using a simple linear model. The regression results appear in Table 7. The results show that price is trendling significantly down and that the equation explains 80% of the variation. A graph comparing the real price observations and the projections appears in Figure 3. We can now use this figure to provide a definition of risk. Mainly, suppose we posit that the decision maker's expectations follow the regression line and the deviations from those expectations is the observed risk. Thus, in 1982 the forecasted price was 72.8 whereas the actual real price is 73.4. Thus, there was an error where the observation was $.60 higher then the expectation, i.e., there was a deviation between the assumed expectation and the observation. In turn, we can develop more stationary series for prices by projecting the forecast out to 1994 for the price distribution and adding the residuals. These calculations appear in Table 6. Note that in 1994 the forecast equation predicts a price of $47.7. We then take the deviations which the real price and forecasting equation which in 1982 is $6.70 and then we add that $6.70 to the $47.7 projection comes up with an even of $54.4. Repeating this calculation for each year results in a stationary distribution of prices. Note, the effect this has on the probability distribution. The nominal prices had a standard error of 5.7. The real prices had a standard error of 8.0 and the stationary price distribution has a standard error of 3.5. Thus, we have gotten a narrower and more applicable probability distribution. This distribution is also on the graph in Figure 5. Note that the stationary price distribution has the smallest variance of all three of the probability distributions.

We can follow the same procedure for yields. Table 6 gives the historical yields, Table 8 the results from time dependent regression and Table 9 the projected yields the forecast errors and stationary probability distribution. Figure 6 gives the historical yields and Figure 7 gives the historical yield relationship with the trend line through it. Figure 7 again shows the difficulty of using the probability distribution derived straight from historical yields, i.e., certainly nobody in 1993 would expect yields that were observed in 1977 rather some yields in the neighborhood of 1993 would be more likely. Figure 8 shows the stationary probability distribution of yields.

Finally, now we can turn our attention to the probability distribution for revenue. Note, that if we take the observed deal times the historical price we get the data in Table 10, column 2. We take the real price times the observed yield we get the data in column 3 and if we take the detrended price distribution and detrended yield distribution we get the data in column 4. In this particular case, notice that the probability distribution of real revenue and stationary revenue are very similar. In fact the real revenue distribution has a smaller variance. This is most likely because of a strong correlation between prices and yields in the California case and when the yields are down prices are up. This leads to another challenge if one was really forming this probability distribution one would probably have to go through some sort of demand model formulation where one took yield and total acreage to form total production and then looked at the relationship between total production and price but also over a 20 year period one would also probably have to factor in the availability of imports from foreign countries, growth in population and some other demand factors before one could say develop a price probability conditional on acreage, yield, population, etc. yielding pure random.

General Lessons Learned on Objective Probabilities from the Above Exercises

The above exercises show several things. First, one must use objective data with caution when describing a probability distribution. The presence of trends and other systematic effects could bias the resultant probability distributions. What one needs to do is use a procedure to simulate expectations formation like regression or moving averages to develop a series of values that were expected for each of the points in the data. In turn, the deviations between the actual observations and values expected are a stationary risk measure. One can then add that stationary risk measure to the extrapolated expectation and form a stationary data set from which the probability distribution can be generated. Expectation models can be substantially more complicated than those used above. For example, if one is trying to estimate a probability distribution for crop yields with data where different amounts of water and fertilizer has been applied, by different people in different places one would have to try to systematically remove the management, location, weather, water and fertilization effects before one could get to the pure random probability distribution part of the model. One also may find that the residual terms are hetereoskedastic as will be the discussion of a later part of this set of notes.

Moving Onward Toward Conditional Distributions

Now, how do we form probability distributions for conditional situations. Fundamentally, the distribution above is conditional on the point in time. What we discuss now is the development of distributions which are conditional on either discrete or continuous events. In order to do this we need a more complex data set.

Suppose we use the data set from Tice. That data set contains corn yields from the Purdue Agronomy farm from a ten year fertilizer experiment. This data contains observations on year amount of nitrogen fertilizer applied, timing of fertilization (fall versus spring) and whether an denitrification inhibitor was used. The data for this amount to some 283 observations and are given in Table 11. Note in these data the first column gives the year, the second the amount of nitrogen used, the third a zero one indicator of whether the ampliation was done in the fall or spring (with a one indicating it was done in the fall) and a zero one variable indicating whether or not the denitrification inhibitor was used. The question now is how do we form a conditional probability distribution for yield dependent upon on the amount of nitrogen used, the application time and the inhibitor. This is just a simple extension of the procedure above.

The first step in forming such a probability distribution will again to make the data stationary. We choose to do this again through the use of regression by adding additional independent variables. In this regression we assume that yield is a quadratic function of nitrogen and has slope shifter for each of the years, nitrogen timing and the inhibitor. (Note in Tice's real analysis a set of weather data were used instead of the years but for class purposes this is certainly adequate).

The regression results are given in Table 12. Notice here the equation explains 78% of variation. Also nitrogen applied has a positive and significant effect while nitrogen squared has a negative effect indicating the diminishing returns to application of the variable input.

Also, fall applications of nitrogen reduces yields relative to spring application but the inhibiter increases yield.

The question now is how can we use this forecast equation to develop a conditional probability distribution. Suppose we wish a distribution for 30 pounds and 120 of nitrogen with spring application without the inhibitor. First, let us develop the residuals.. This is done by evaluating the function at each an every observation and forming a forecast then computing residuals. Thus, if we take this function and plug in for 1968 conditions and zero nitrogen applied and any other items which correspond to the first observation of the base set we get in a forecast of 51.7 bushels of corn whereas, the actual observed yield is 67 bushels, as a consequence, we have a forecast error of 15.3 bushels. We then repeat this procedure for all observations then take these residuals we could add them to a mean forecast equation to come up with a probability distribution. We form this mean forecast by evaluating the formula residuals plus the dummy variables. Under these circumstances the expected yield is 59.7 bushels for 30 pounds and 117 bushels for 120 pounds without the resultant probability distribution is graphed in Figure 12. In this distribution notice that the line on the left is for 30 units of nitrogen and the line on the right is for 120 units.

The formation of this forecast is a little bit more difficult than those above because some of the independent variables we use are in fact stochastic. Mainly, we assume a distribution of the dummy variables across the years also occurs there is a frequency for which the years occur. For the base forecast what we will do is, we will take the simple average across the dummy variables which is .06 I believe and add that to the forecast. But for forming the probability distributions what we will do is take the forecast of just the part of the model involving the non-stochastic independent variables, i.e. dropping the dummies. When we are forming a probability distribution observation that corresponds to a particular year and residual, we will add both the year dummy effect and the residual. Thus, for the 1968 case, we take valuation of the terms and squared term plus the 1968 dummy variable plus the residual for each of the replications in the data set and this gives us a 283 point probability distribution which is graphed in Figure 12.

Higher Order Moments - Just Pope

Note, the estimates of the distributions above are exactly parallel and shaped identically. This is a consequence of our assumptions. Namely, the mathematical formula for the regression written in arbitrary functional form is:

$$Y = f(x) + \epsilon$$

In this function f(x) is non-stochastic while ϵ is an error term with an expected value of zero and variance of σ_ϵ^2 . Now suppose we find the mean and variance of Y using mathematical statistics. To find the mean we use the formulas:

$$E(Y) = E(F(x) + \epsilon) = F(x) + E(\epsilon) = f(x)$$

The gives the expectation or average for Y and since f(x) is a nonstochastic we can simple Y factor it out of the expectation. Then since the expected value of the error term is zero this means that the expected value of Y is simply the f(x) function that we have estimated. In turn if we look for the variance of Y, the mathematical formula for it is:

$$V(Y) = E(Y - E(Y))^2$$
$$= E(f(x) + \epsilon - E(Y))^2 = E(f(x) + \epsilon - f(x))^2$$
$$= E(\epsilon)^2 = E(\epsilon - E(\epsilon))^2 = \sigma_\epsilon^2$$

In this case, we plug in our value for the expected value for y and simplify we get an expression which is simply the expected value of the error term the quantity squared. We then add in the expected value of ϵ without consequence because it is zero. The resultant expression is the definition of the variance of the error term. Thus the variance of Y is just the variance of the error term. Notice what this means is that the variance given any level of nitrogen fertilizer is always the same. This is potentially a limitation and leads us into our next form of probability distribution estimation. That is estimation controlling the higher order moments through regression.

Fundamentally the technique that we used above is one in which we estimate an expression for the mean which depends on X and then use this to make the distribution stationary. But as we show above with our linear functional form this leads to probability distributions with variance independent of the level of x. There may be cases where we wish to examine alternative specifications.

Let us do a little bit more mathematical statistics to show the implications of two additional functional forms we might use. If we take a traditional multiplicative approach to this problem, we would estimate the equation $y = (x)e^{\epsilon}$. This particular case we would assume that the expected value of the exponentiated error term was one. Thus, when we estimate the model in logs the expected value of the error term is zero and the variance of that term is σ_{ϵ}^{2}. In this particular case, if we go through our mathematical statistics estimation given:

$$Y = h(x)e^{\epsilon}$$

$$
\begin{aligned}
E(Y) &= E(h(x)e^{\epsilon}) \\
&= h(x)E(e^{\epsilon}) \\
&= h(x)
\end{aligned}
$$

$$V(Y) = E\,(h(x)e^{\epsilon} - E(y))^2 = h^2(x)E(e^{\epsilon} - E(e^{\epsilon}))^2 = h^2(x)e^{\sigma_{\epsilon}^2}$$

this shows that the expected value of h(x) again is the function estimated but that the variance now is expected to be the square of the h(x) function times the variance of the error term. Note in this case and in the one derived above we have made restrictive assumptions about how the variance behaves as x changes. Those restrictive assumptions arise because of the functional forms that we are using in our regression analysis. Just and Pope recognized this and proposed yet a third functional form

$$Y = f(x) + h^{1/2}(x)\epsilon$$

Here a mathematical statistics analysis leads to the conclusion that the expected value of Y is equal to f(x) and the variance of y is equal to h(x) times the variance of the error term.

$$
\begin{aligned}
E(Y) &= f(x) \\
Var(Y) &= h(x)\sigma_{\epsilon}^2
\end{aligned}
$$

What this means is that if we could estimate this function that f(x) would give the mean response to the usage of the inputs whereas $h^{½}(x)$ would tell how the variance of output changes as we change inputs. Note, this is an interpretation of the heteroskedasticity correction from econometrics. The procedure Just and Pope proposed for estimating this is as follows:

Step 1 - estimate the equation $Y = f(x) + \epsilon$

Step 2 - Compute the residuals from the equation given the estimated form of the equation

$$\hat{\epsilon} = Y - \tilde{F}(x).$$

Step 3, take the absolute values of Ê and find $h^{½}(x)$ using the functional form as follows.

$$|\hat{\epsilon}| = aX_1^{b_1}X_2^{b_2} \ldots . X_n^{b_n}\epsilon$$

Where a is an intercept term and b's are multiplicative powers on the various independent variables. This is a log linear regression so one can simple take the logarithms of both sides and do an OLS estimation. In turn, then one can go back to the original equation

$$Y = f(x) + h^{1/2}(x)\ \epsilon$$

and divide the equation through by our estimate of the $h^{1/2}(x)$ term as follows.

$$\frac{Y}{h^{1/2}(x)} = \frac{f(x)}{h^{1/2}(x)} + \epsilon$$

Estimate to find a new set of parameters for f(x). In turn observe the parameters for $h^{1/2}(x)$ which show how the variances change as the levels of x change.

We implement this procedure in terms of the above data set with three regressions the results of which are reported in Table 9. The first stage in Table 9 is identical to the regression we had earlier. However, in the second stage we take the residuals from the first stage and take the absolute value and log them for estimation. In this particular case then we get the estimates in the stage 2 column which show that constant a in the above function would be the antilog of 2.091. (Just and Pope mention while Buccola and McCarl discuss the fact that

0.6532 needs to be added to this estimate to remove bias). The stage 2 results also show nitrogen applied has a negative effect, (i.e., it is variance decreasing) but is not significant while fall application is variance increasing and significant and the inhibitor is variance decreasing but insignificant. These data show that one could conclude that fall versus spring application that the variances of fall applications are higher and that the variance of the spring application is lower. (This isn't surprising given that what that means is your applying the nitrogen closer to the time that the crop is being planted).

In turn, suppose we go back and form a new distribution. In this case what happens is we have to use the predicted error terms from the state 3 regression. In this particular case, the probability distribution is formed by evaluating the forecasted value for the particular outcome desired, i.e., 30 or 120 lbs. of nitrogen plus the error terms from stage 3 multiplied by the $h^{1/2}(x)$ term with the corrected intercept. The resultant probability distribution is then given in Figure **Final Thoughts on Conditional Distributions**

The basic approach estimating for conditional distribution is that followed above. Namely, what one does is estimate some sort of a function which has conditional item as an independent variable to find how the data are statistically influenced by the conditional item. That function may involve a function where one only has controlled for the first moment, one where the first and second moments are controlled for using the Just Pope procedures or one where even higher order moments are treated as in Antle. In turn one then one obtains predictions from the model on the mean Y values then adjusts the residuals by the estimated moments and forms the probability distribution. One could also simply use the moments from the regressions without forming a probability distribution.

Multi-variate Distributions

The next obvious step in the progression is to multi-variate distributions. Suppose one wishes to form a distribution on multicrop yields in different places including how they covary over time. What is commonly done in this case is that we estimate a set of regression equations not a single regression equation then form the distributions. For example, one might estimate a set of models with time as an independent variable of yields with one done for each crop in the 50 US states for each crop. Given those equations we could form univariate probability distributions by detrending (see Thaysen for an example). One can also preserve cross state correlations for example associating the 1977 observations in all the states for all the crops together. Thus, if 1977 tended to be unusually good weather year what would happen is the yields for corn, soybean, and wheat would all be above the mean. The residuals would reflect this now by maintaining the integrity of the 1977 operations across all crops and places one preserves correlation and therefore has a multi-variate distribution. This distribution gives not only the effects on the individual variables of changing inputs but also gives some idea of the interrelationship between the variables.

Continuous Distributions

An item missing from the above discussion which we will really not add very much on is how to estimate a continuous known form probability distribution. What one does in that case is apply some sort of regression based technique to develop a maximum likelihood estimate of the probability distribution by fitting the parameters according to its known functional form. Taylor presents some information on this. The only thing we have is that this is very difficult to do. We favor using empirical discrete distributions.

Table 1. Data for Probability Distribution Formation.

Obs Num	Raw Data	Sorted Data		Obs Num	Raw	Sorted	
1	7	7	5	39	35	31	30
2	22	8	10	40	21	31	30
3	31	10	10	41	28	31	30
4	20	12	10	42	31	32	30
5	24	13	15	43	15	32	30
6	41	14	15	44	34	34	35
7	13	15	15	45	23	35	35
8	24	16	15	46	38	36	35
9	20	17	15	47	24	36	35
10	23	18	20	48	36	38	40
11	18	18	20	49	32	41	40
12	12	19	20	50	26	45	45
13	26	20	20				
14	18	20	20	Mean	24.62	24.5	
15	25	20	20	Std er	8.158	8.078	
16	16	21	20				
17	32	22	20				
18	26	22	20				
19	27	23	25				
20	45	23	25				
21	17	23	25				
22	27	23	25				
23	26	24	25				
24	29	24	25				
25	8	24	25				
26	23	25	25				
27	36	25	25				
28	14	26	25				
29	31	26	25				
30	19	26	25				
31	20	26	25				
32	10	27	25				
33	30	27	25				
34	23	27	25				
35	22	28	30				
36	25	29	30				
37	27	30	30				
38	31	31	30				

Table 2. Probability Distribution Data For Forecast.

Interval	Mid Point	Frequency	Probability
2.5-7.5	5	1	.02
7.5-12.5	10	3	.06
12.5-17.5	15	5	.15
17.5-22.5	20	9	.18
22.5-27.5	25	16	.32
27.5-32.5	30	9	.18
32.5-37.5	35	4	.08
37.5-42.5	40	2	.04
42.5-47.5	45	1	.02
		50	1.00

Table 3. Dealing With Non Stationary Distributions.

Year	Data	Forecast	Residual	Stationary Series
1977	520	466.3	53.7	664
1978	420	480.8	-61	550
1979	547	495.2	51.8	663
1980	404	509.6	-106	505
1981	542	524.1	17.9	629
1982	590	538.5	51.5	662
1983	508	553.0	-45	566
1984	600	567.4	32.6	643
1985	630	581.8	48.2	659
1986	552	596.3	-44	566
1987		610.7		

Panel B Regression

Output	
R Squared	0.3588013
No. Of Observation	10
Degrees of Freedom	8
Constant	61.999951
Year	6.826

Table 4. Raw Data

Year	Strawberry Price/CWT	Strawberry Yield/Acre	Revenue/Acre	Implicit GNP Deflator
1977	33.2	186	6175	55.9
1978	31.7	175	5548	60.3
1979	38.9	175	6808	65.5
1980	41.2	195	8034	71.7
1981	42	200	8400	78.9
1982	48.9	219	10709	83.8
1983	45.5	207	9419	87.2
1984	41.7	229	9549	91
1985	44.3	230	10189	94.4
1986	49.5	229	11336	96.9
1987	49.4	242	11955	100
1988	46.2	250	11550	103.9
1989	47.1	248	11681	108.5
1990	46.3	272	12594	113.3
1991	46.3	293	13566	117.6
1992	52.1	265	13807	120.9
1993	52.4	276	14462	123.5
Mean	44.51176	228.8824	10339.98	92.54706
Standard Error	5.706434	34.88111	2599.357	20.47148

*Source Agricultural Statistics 1992 and 1994 Editions.

Table 5. Probability Distribution Nominal Prices.

Price Band	Relative Frequency
30-35	2/17
35-40	1/17
40-45	4/17
45-50	8/17
50-55	2/17

Table 6. Price Data - Manipulated

Year	Strawberry Price/CWT	Real Prices	Projected Price	Forecast Error	Stationary Price
1977	33.2	73.3	72.8	0.6	48.3
1978	31.7	64.9	71.3	-6.4	41.4
1979	38.9	73.3	69.8	3.5	51.3
1980	41.2	71.0	68.3	2.6	50.4
1981	42	65.7	66.9	-1.1	46.6
1982	48.9	72.1	65.4	6.7	54.4
1983	45.5	64.4	63.9	0.5	48.2
1984	41.7	56.6	62.5	-5.9	41.9
1985	44.3	58.0	61.0	-3.0	44.7
1986	49.5	63.1	59.5	3.6	51.3
1987	49.4	61.0	58.0	3.0	50.7
1988	46.2	54.9	56.6	-1.7	46.1
1989	47.1	53.6	55.1	-1.5	46.3
1990	46.3	50.5	53.6	-3.2	44.6
1991	46.3	48.6	52.1	-3.5	44.2
1992	52.1	53.2	50.7	2.5	50.3
1993	52.4	52.4	49.2	3.2	50.9
1994			47.7		
Mean	44.51176	60.98338	60.98338	4.0E-15	47.73303
Standard Error	5.706434	8.042726	7.212577	3.558674	3.558674

Table 7. Real Price Forecast Regression Real Price on Time

Regression Output:	
Constant	2983.422
Standard Error of Y Estimate	3.788498
R Squared	0.804219
Number of Observations	17
Degrees of Freedom	15
X Coefficient(s) (Time)	-1.47226
Standard Error of Coefficient	0.187559

Table 8. Yield Regression Yield on Time

Regression Output:	
Constant	-13383.9
Standard Error of Y Estimate	9.985054
R Squared	0.927696
Number of Observations	17
Degress of Freedom	15
X Coefficient(s) (Time)	6.857843
Standard Error of Coefficient	0.494334

Table 9. Yield Data

Year	Historical Yields	Projected Yield	Yield Projection Error	Stationary Yields
1977	186	174	12	303
1978	175	181	-6	285
1979	175	188	-13	278
1980	195	195	0	291
1981	200	201	-1	289
1982	219	208	11	301
1983	207	215	-8	282
1984	229	222	7	298
1985	230	229	1	292
1986	229	236	-7	284
1987	242	243	-1	290
1988	250	249	1	291
1989	248	256	-8	282
1990	272	263	9	299
1991	293	270	23	314
1992	265	277	-12	279
1993	276	284	-8	283
1994		291		
Mean	228.8824	228.8824	3.3E-12	290.6029
Standard Error	34.88111	33.59643	9.379325	9.379325

Table 10. Revenue Data

Year	Historical Revenue	Real Revenue	Stationary Revenue
1977	6175	13643	14621
1978	5548	11362	11779
1979	6808	12836	14244
1980	8034	13838	14653
1981	8400	13148	13475
1982	10709	15783	16390
1983	9419	13339	13626
1984	9549	12960	12460
1985	10189	13330	13042
1986	11336	14447	14565
1987	11955	14764	14704
1988	11550	13729	13417
1989	11681	13296	13056
1990	12594	13727	13348
1991	13566	14247	13862
1992	13807	14103	14013
1993	14462	14462	14405
Mean	10339.98	13707	13862
Standard Error	2599.357	926	1016

Table 11. Tices Raw Data

Year	Yield	N	Fall	Nserv	Year	Yield	N	Fall	Nserv	Year	Yield	N	Fall	Nserv
1968	67	0	1	0	1969	51	0	1	0	1970	36	0	1	0
1968	63	0	1	0	1969	50	0	1	0	1970	50	0	1	0
1968	34	0	1	0	1969	55	0	1	0	1970	47	0	1	0
1968	115	60	1	0	1969	81	60	1	0	1970	75	60	1	0
1968	107	60	1	0	1969	92	60	1	0	1970	86	60	1	0
1968	118	60	1	0	1969	94	60	1	0	1970	95	60	1	0
1968	145	120	1	0	1969	94	60	1	0	1970	114	120	1	0
1968	152	120	1	0	1969	92	60	1	0	1970	132	120	1	0
1968	141	120	1	0	1969	102	60	1	0	1970	126	120	1	0
1968	132	180	1	0	1969	113	120	1	0	1970	66	60	1	0
1968	152	180	1	0	1969	110	120	1	0	1970	81	60	1	0
1968	153	180	1	0	1969	106	120	1	0	1970	83	60	1	0
1968	116	60	0	0	1969	132	120	1	0	1970	82	120	1	0
1968	126	60	0	0	1969	131	120	1	0	1970	101	120	1	0
1968	121	60	0	0	1969	141	120	1	0	1970	116	120	1	0
1968	139	120	0	0	1969	119	180	1	0	1970	142	180	1	0
1968	147	120	0	0	1969	137	180	1	0	1970	146	180	1	0
1968	148	120	0	0	1969	140	180	1	0	1970	156	180	1	0
1968	154	180	0	0	1969	95	60	0	0	1970	58	60	0	0
1968	149	180	0	0	1969	105	60	0	0	1970	77	60	0	0
1968	151	180	0	0	1969	105	60	0	0	1970	77	60	0	0
					1969	144	120	0	0	1970	115	120	0	0
					1969	133	120	0	0	1970	114	120	0	0
					1969	141	120	0	0	1970	124	120	0	0
					1969	125	180	0	0	1970	143	180	0	0
					1969	118	180	0	0	1970	147	180	0	0
					1969	130	180	0	0	1970	152	180	0	0

Table 11. Continued.

Year	Yield	N	Fall	Nserv	Year	Yield	N	Fall	Nserv	Year	Yield	N	Fall	Nserv
1971	15	0	1	0	1972	21	0	1	0	1973	22	0	1	0
1971	18	0	1	0	1972	27	0	1	0	1973	20	0	1	0
1971	15	0	1	0	1972	19	0	1	0	1973	20	0	1	0
1971	120	60	1	0	1972	32	60	1	0	1973	100	60	1	0
1971	118	60	1	0	1972	33	60	1	0	1973	102	60	1	0
1971	128	60	1	0	1972	30	60	1	0	1973	114	60	1	0
1971	54	60	1	0	1972	46	60	1	0	1973	60	60	1	0
1971	81	60	1	0	1972	66	60	1	0	1973	42	60	1	0
1971	69	60	1	0	1972	79	60	1	0	1973	58	60	1	0
1971	177	120	1	0	1972	45	120	1	0	1973	146	120	1	0
1971	166	120	1	0	1972	43	120	1	0	1973	142	120	1	0
1971	158	120	1	0	1972	40	120	1	0	1973	135	120	1	0
1971	108	120	1	0	1972	102	120	1	0	1973	42	120	1	0
1971	110	120	1	0	1972	116	120	1	0	1973	86	120	1	0
1971	123	120	1	0	1972	113	120	1	0	1973	101	120	1	0
1971	179	180	1	0	1972	67	180	1	0	1973	149	180	1	0
1971	187	180	1	0	1972	67	180	1	0	1973	149	180	1	0
1971	183	180	1	0	1972	106	180	1	0	1973	150	180	1	0
1971	109	60	0	0	1972	68	60	0	0	1973	53	60	0	0
1971	114	60	0	0	1972	72	60	0	0	1973	80	60	0	0
1971	96	60	0	0	1972	78	60	0	0	1973	76	60	0	0
1971	158	120	0	0	1972	112	120	0	0	1973	124	120	0	0
1971	157	120	0	0	1972	115	120	0	0	1973	115	120	0	0
1971	170	120	0	0	1972	127	120	0	0	1973	139	120	0	0
1971	183	180	0	0	1972	145	180	0	0	1973	143	180	0	0
1971	176	180	0	0	1972	135	180	0	0	1973	144	180	0	0
1971	198	180	0	0	1972	142	180	0	0	1973	139	180	0	0

Table 11. Continued.

Year	Yield	N	Fall	Nserv	Year	Yield	N	Fall	Nserv	Year	Yield	N	Fall	Nserv
1974	27	0	1	0	1975	2	0	1	0	1976	15	0	1	0
1974	30	0	1	0	1975	1	0	1	0	1976	5	0	1	0
1974	32	0	1	0	1975	4	0	1	0	1976	17	0	1	0
1974	83	60	1	0	1975	88	60	1	0	1976	90	60	1	0
1974	103	60	1	0	1975	108	60	1	0	1976	112	60	1	0
1974	79	60	1	0	1975	115	60	1	0	1976	107	60	1	0
1974	76	60	1	0	1975	82	60	1	0	1976	51	60	1	0
1974	54	60	1	0	1975	74	60	1	0	1976	54	60	1	0
1974	71	60	1	0	1975	94	60	1	0	1976	68	60	1	0
1974	88	120	1	0	1975	149	120	1	0	1976	126	120	1	0
1974	103	120	1	0	1975	154	120	1	0	1976	133	120	1	0
1974	106	120	1	0	1975	159	120	1	0	1976	109	120	1	0
1974	97	120	1	0	1975	128	120	1	0	1976	91	120	1	0
1974	90	120	1	0	1975	122	120	1	0	1976	8	120	1	0
1974	92	120	1	0	1975	134	120	1	0	1976	104	120	1	0
1974	85	180	1	0	1975	165	180	1	0	1976	113	180	1	0
1974	101	180	1	0	1975	170	180	1	0	1976	140	180	1	0
1974	93	180	1	0	1975	169	180	1	0	1976	122	180	1	0
1974	86	60	0	0	1975	53	0	0	0	1976	64	60	0	0
1974	94	60	0	0	1975	60	0	0	0	1976	63	60	0	0
1974	88	60	0	0	1975	66	0	0	0	1976	64	60	0	0
1974	134	120	0	0	1975	108	120	0	0	1976	104	120	0	0
1974	126	120	0	0	1975	127	120	0	0	1976	107	120	0	0
1974	131	120	0	0	1975	141	120	0	0	1976	118	120	0	0
1974	134	120	0	0	1975	150	180	0	0	1976	124	180	0	0
1974	136	120	0	0	1975	158	180	0	0	1976	133	180	0	0
1974	150	120	0	0	1975	165	180	0	0	1976	136	180	0	0
					1975	86	0	1	0	1976	77	0	1	1
					1975	134	60	1	0	1976	103	60	0	0
					1975	153	60	1	1	1976	117	60	0	1
					1975	161	120	1	0	1976	133	120	0	0
					1975	171	120	1	1	1976	150	120	0	1
					1975	154	60	1	0	1976	151	180	0	0
					1975	151	60	0	1	1976	171	180	0	1
					1975	174	120	0	0					
					1975	174	120	0	1					
					1975	185	180	0	1					

Table 11. Continued.

Year	Yield	N	Fall	Nserv
1977	46	0	1	0
1977	48	0	1	0
1977	51	0	1	0
1977	85	60	1	0
1977	98	60	1	0
1977	112	60	1	0
1977	78	60	1	0
1977	73	60	1	0
1977	85	60	1	0
1977	114	120	1	0
1977	135	120	1	0
1977	119	120	1	0
1977	105	120	1	0
1977	110	120	1	0
1977	114	180	1	0
1977	117	180	1	0
1977	130	180	1	0
1977	116	180	1	0
1977	67	60	0	0
1977	74	60	0	0
1977	77	60	0	0
1977	95	60	0	0
1977	98	60	0	0
1977	107	60	0	0
1977	118	60	0	0
1977	113	60	0	0
1977	125	60	0	0
1977	123	150	1	0
1977	128	150	1	1

Table 12. Regression Model of Tice Data.

R Squared	0.7777254		
No. Of Observations	283		
Degrees of Freedom	269		
	Coef	Std Err	T
Constant	43.123313	20.268213	2.13
N applied	1.0174352	0.07368	13.8
N timing (1=Fall)	-11.64241	2.590484	-4.5
Inhib use	33.144929	6.8449267	4.84
N squared	-0.002508	0.0003669	-6.8
1968	20.227745	5.8412515	3.46
1969	4.7369295	5.4344857	0.87
1970	-2.485293	5.4344857	-0.5
1971	20.811004	5.4344857	3.83
1972	-28.22603	5.4344857	-5.2
1973	-5.818626	5.4344857	-1.1
1974	-10.05109	5.4296	-1.9
1975	17.728637	5.0823424	3.49
1976	-10.43183	5.17711059	-2
1977 (in intercept)	0		

Table 13. Just Pope Approach Applied to Tice Data.

	Stage 1			Stage 2			Stage 3		
R Squared	0.778			0.045			0.886		
No. of Observations	283			283			283		
Degrees of Freedom	269			279			279		
	Coef	Std err	T						
Constant	43.1233131	20.268213	2.12763272	2.09145598	1.14007739	1.83448598	43.0194002	5.0813523	8.4661322
N applied	1.01743522	0.0736877	13.8073975	-0.00046	0.001212	-0.3762994	1.03046209	0.076826	13.412958
N timing	-11.642413	2.59048396	-4.4943005	0.45767084 8	0.14473997	3.16202117	-11.745355	2.3158456	-5.071735
Inhib use	33.14	6.85492674	4.83519812	-0.4335093	0.36888404	-1.1751913	33.3990791	14.323646	2.3317442
N squared	-0.00251	0.00037	-6.8342305				-0.00249	0.0004	-6.831568
1968	20.2277445	5.84125152	3.46291278				16.5049886	5.3237476	3.1002576
1969	4.73692954	5.43448567	0.87164266				0.63315146	5.0961629	0.1242408
1970	-2.4852927	5.43448567	-0.4573188				-6.7963661	5.0961629	-1.333624
1971	20.8110036	5.43448567	3.8294339				22.453066	5.0961629	4.4058769
1972	-28.226033	5.43448567	-5.1938739				-24.587191	5.0961629	-4.824648
1973	-5.818626	5.43448567	-1.0706857				-8.6643557	5.0961629	-1.700172
1974	-10.05109	5.42959995	-1.8511659				-7.5560914	5.0876895	-1.485172
1975	17.728637	5.08234241	3.48828071				14.5007899	4.9183972	2.9482755
1976	-10.431833	5.17110587	-2.0173311				-13.941123	4.8832529	-2.854884
1977	0								

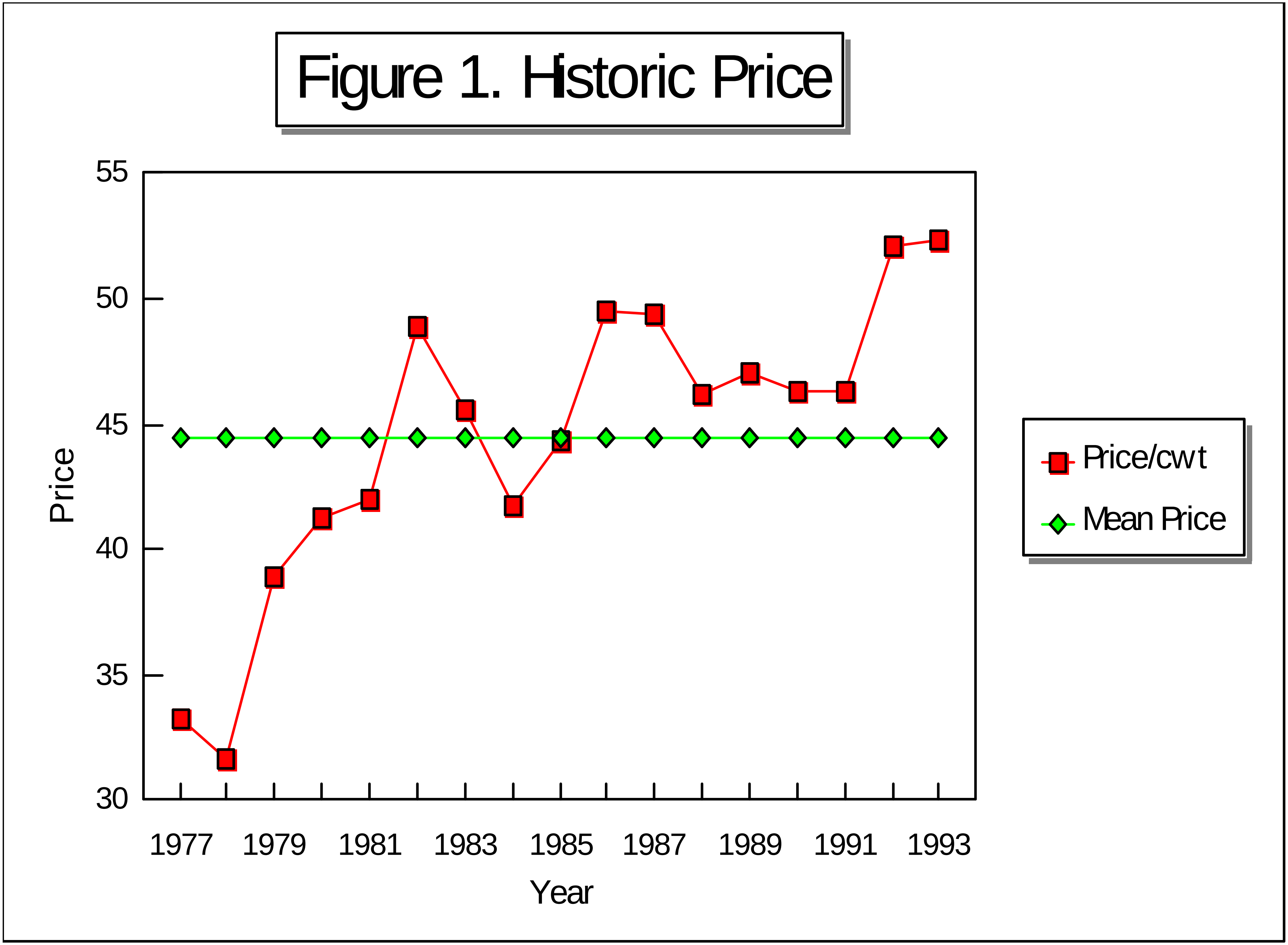
Figure 1. Historic Price
Price
55
50
45
40
35
30
1977
1979
1981
1983
1985
1987
1989
1991
1993
Year
Price/cwt
Mean Price

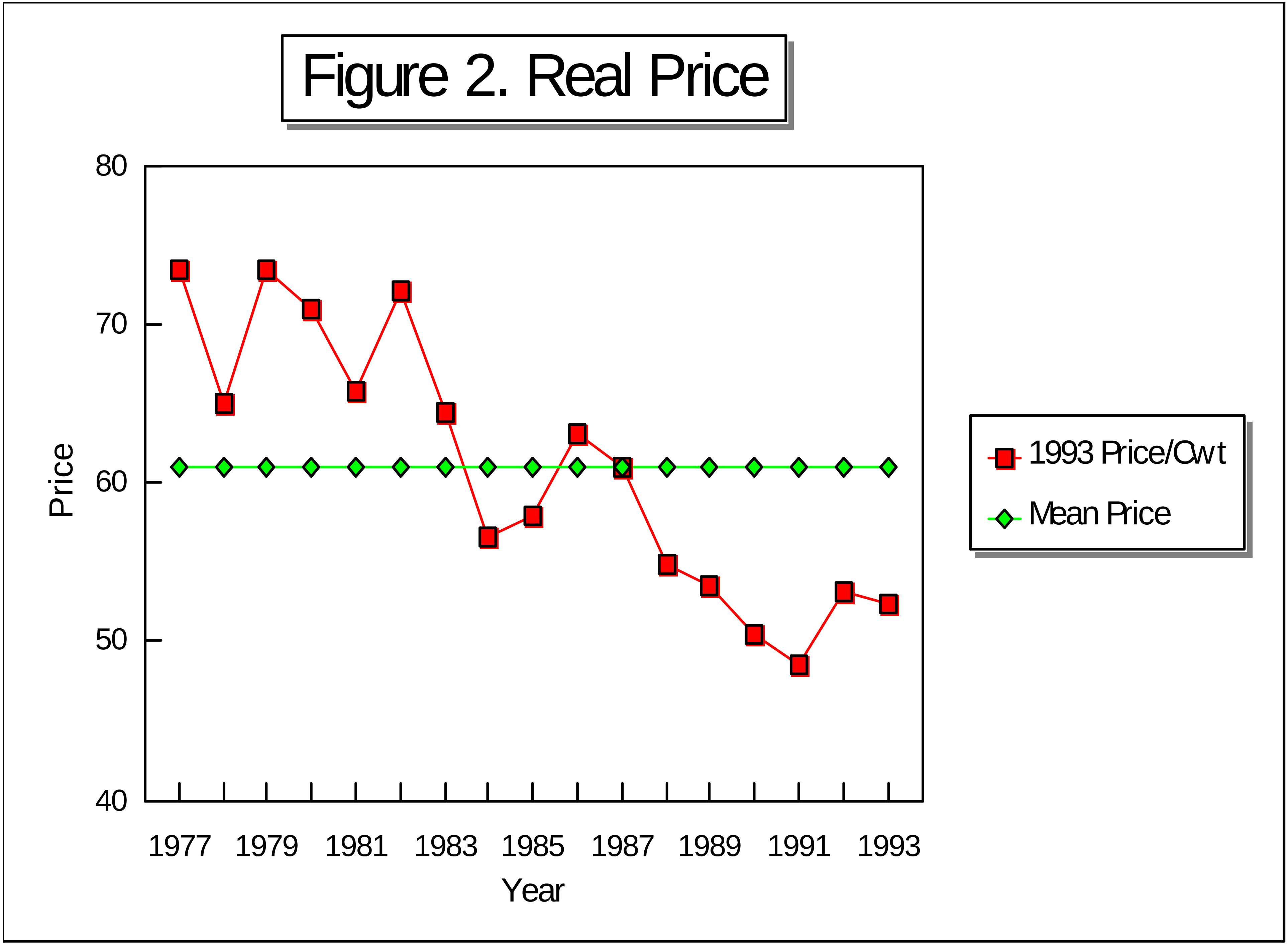
Figure 2. Real Price
80
70
60
50
40
Price
1977
1979
1981
1983
1985
1987
1989
1991
1993
Year
1993 Price/Cwt
Mean Price

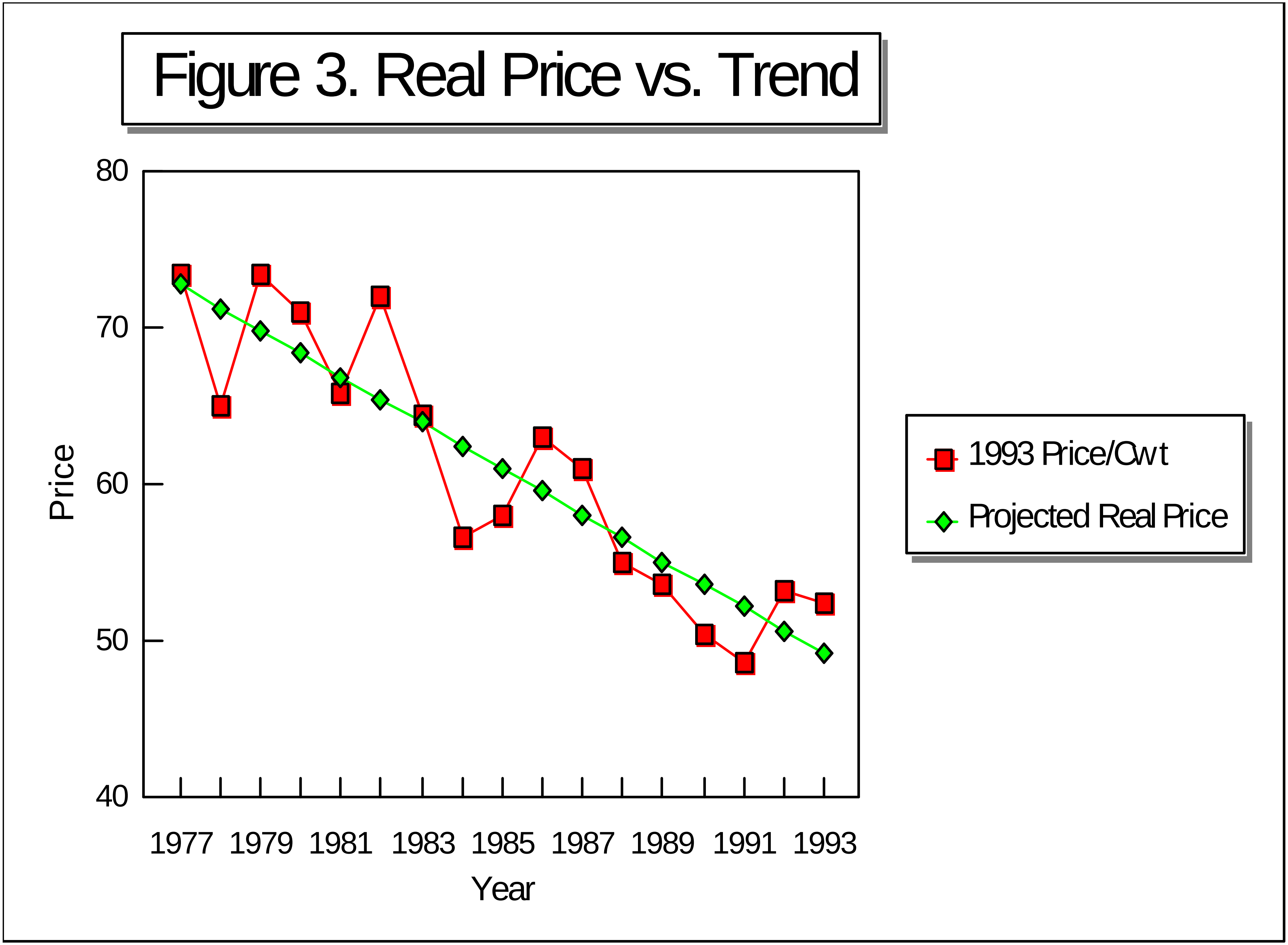
Figure 3. Real Price vs. Trend
Price
80
70
60
50
40
1977
1979
1981
1983
1985
1987
1989
1991
1993
Year
1993 Price/Cwt
Projected Real Price

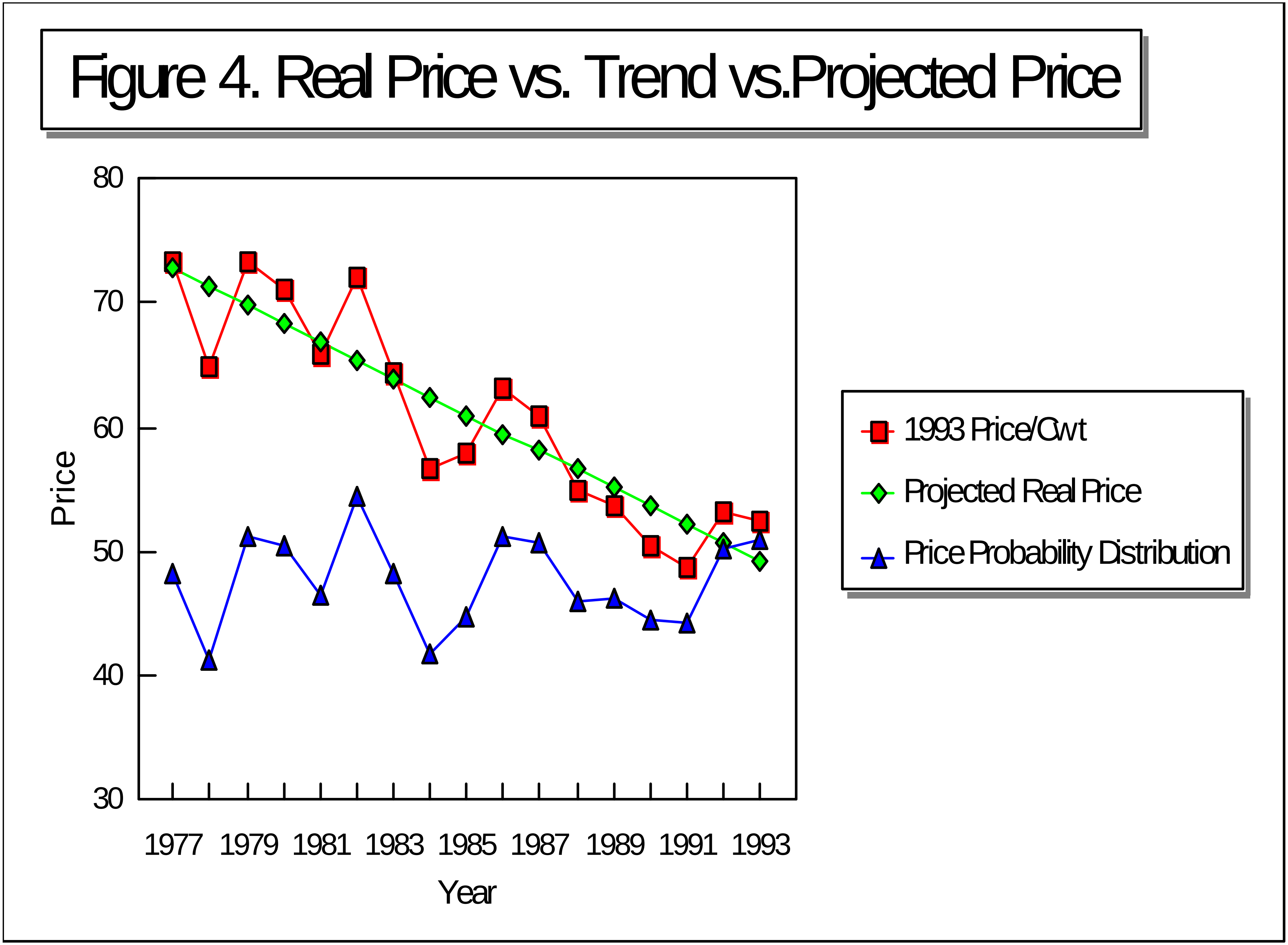

Figure 4. Real Price vs. Trend vs.Projected Price
80
70
60
50
40
30
Price
1977
1979
1981
1983
1985
1987
1989
1991
1993
Year
1993 Price/Cwt
Projected Real Price
Price Probability Distribution

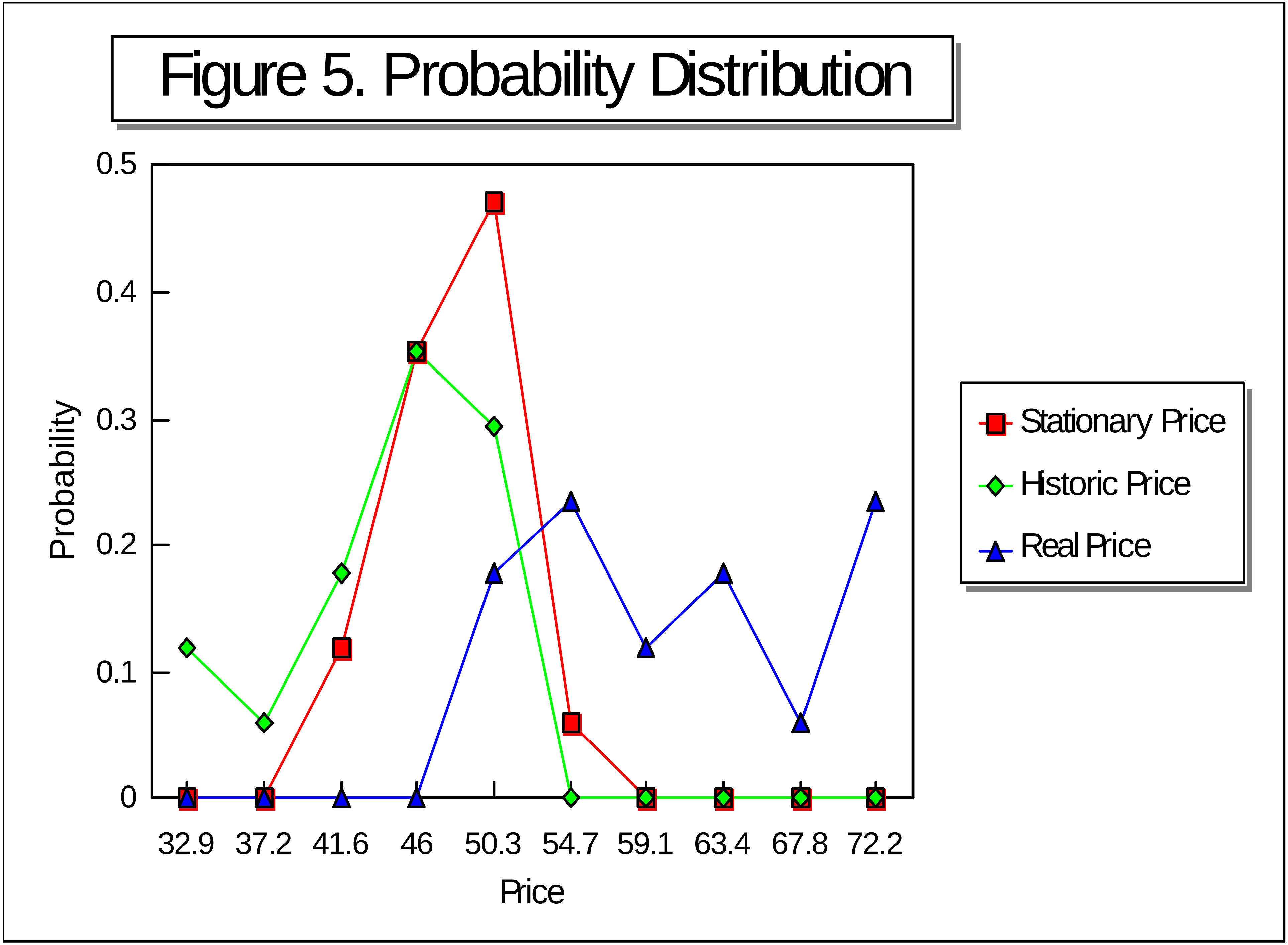
Figure 5. Probability Distribution
Probability
0.5
0.4
0.3
0.2
0.1
0
32.9
37.2
41.6
46
50.3
54.7
59.1
63.4
67.8
72.2
Price
Stationary Price
Historic Price
Real Price

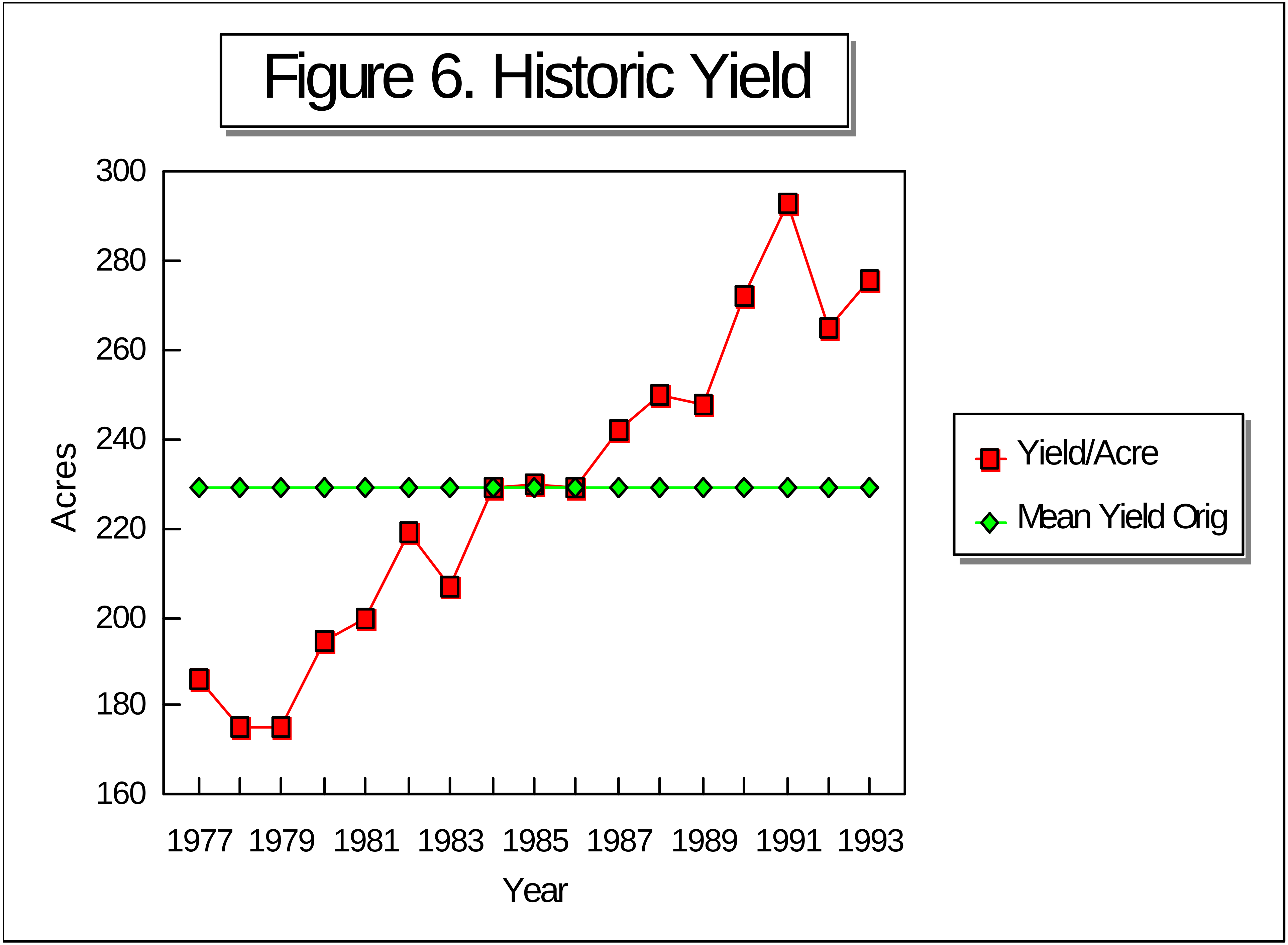
Figure 6. Historic Yield
300
280
260
240
220
200
180
160
Acres
1977
1979
1981
1983
1985
1987
1989
1991
1993
Year
Yield/Acre
Mean Yield Orig

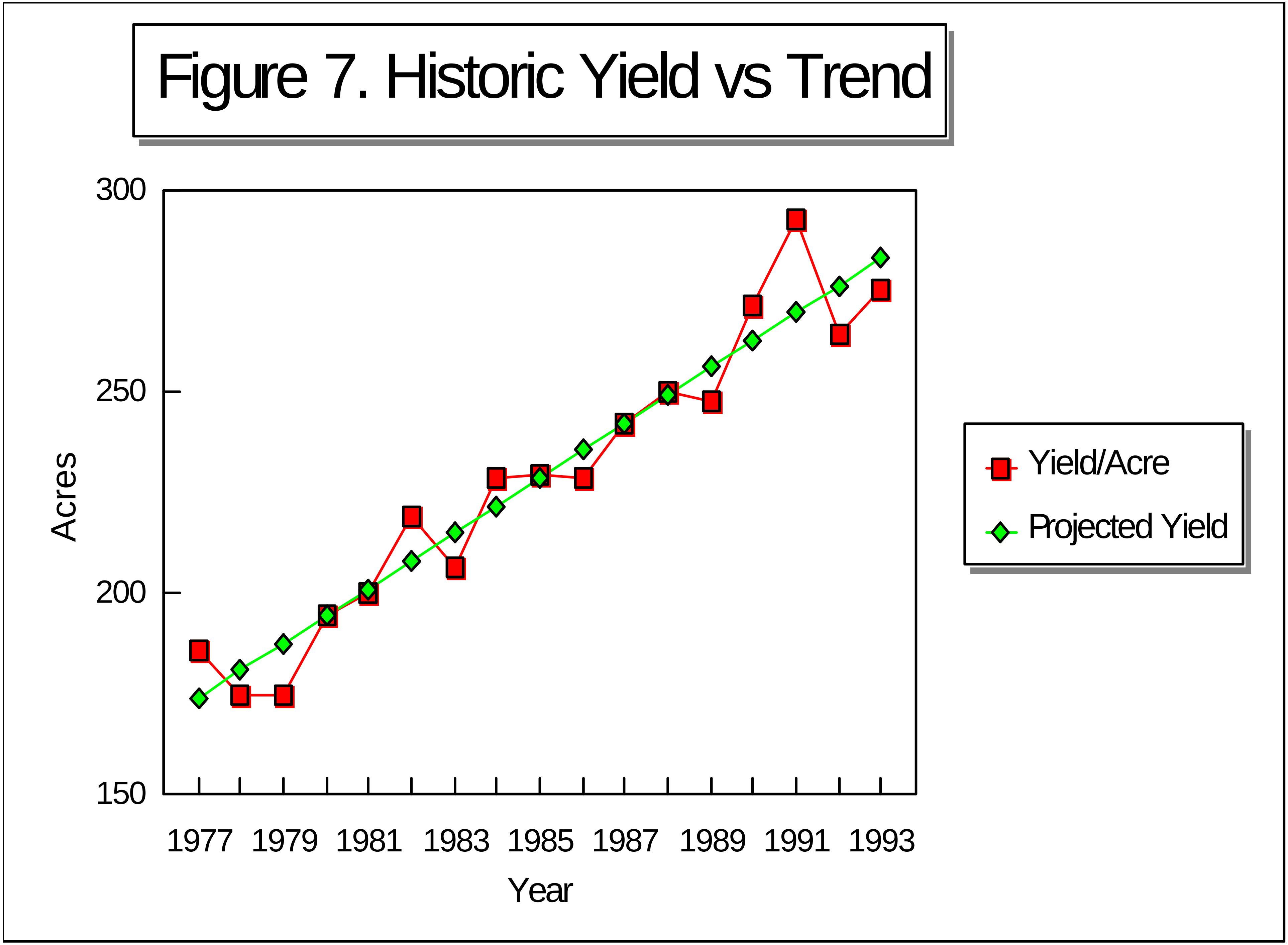
Figure 7. Historic Yield vs Trend
300
250
200
150
Acres
1977
1979
1981
1983
1985
1987
1989
1991
1993
Year
Yield/Acre
Projected Yield

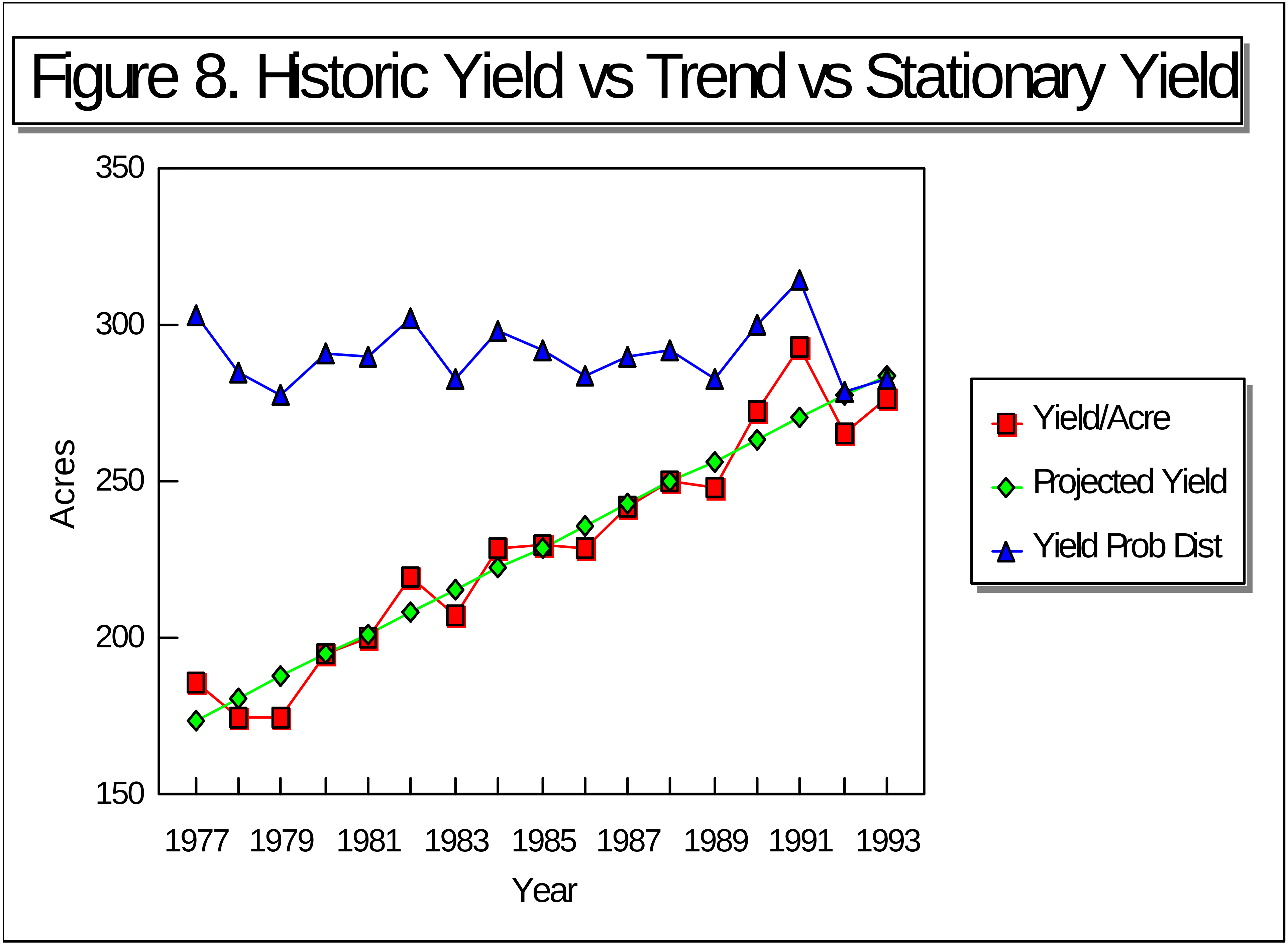
Figure 8. Historic Yield vs Trend vs Stationary Yield
350
300
250
200
150
Acres
1977
1979
1981
1983
1985
1987
1989
1991
1993
Year
Yield/Acre
Projected Yield
Yield Prob Dist

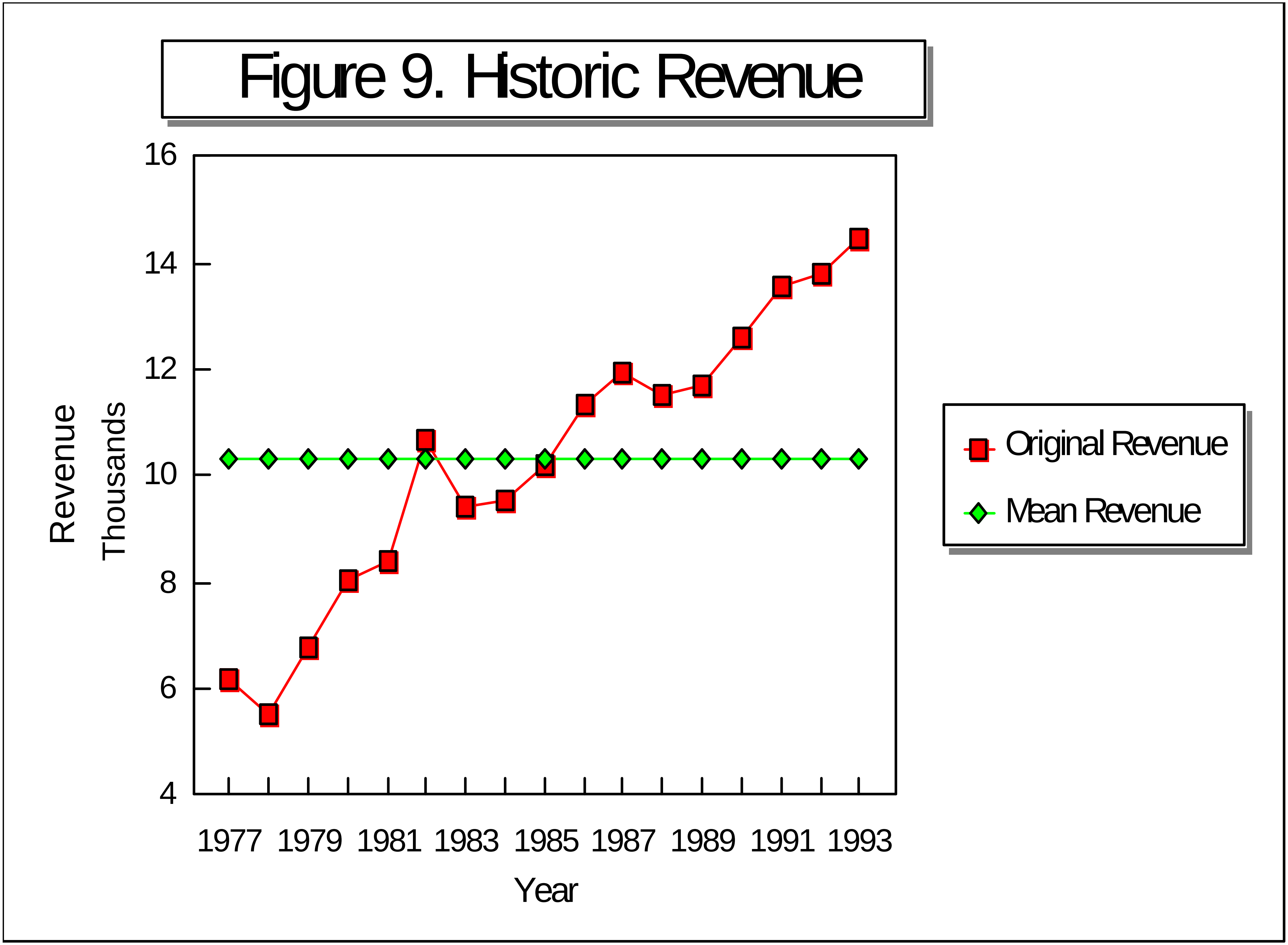
Figure 9. Historic Revenue
16
14
12
10
8
6
4
Revenue
Thousands
1977 1979 1981 1983 1985 1987 1989 1991 1993
Year
Original Revenue
Mean Revenue

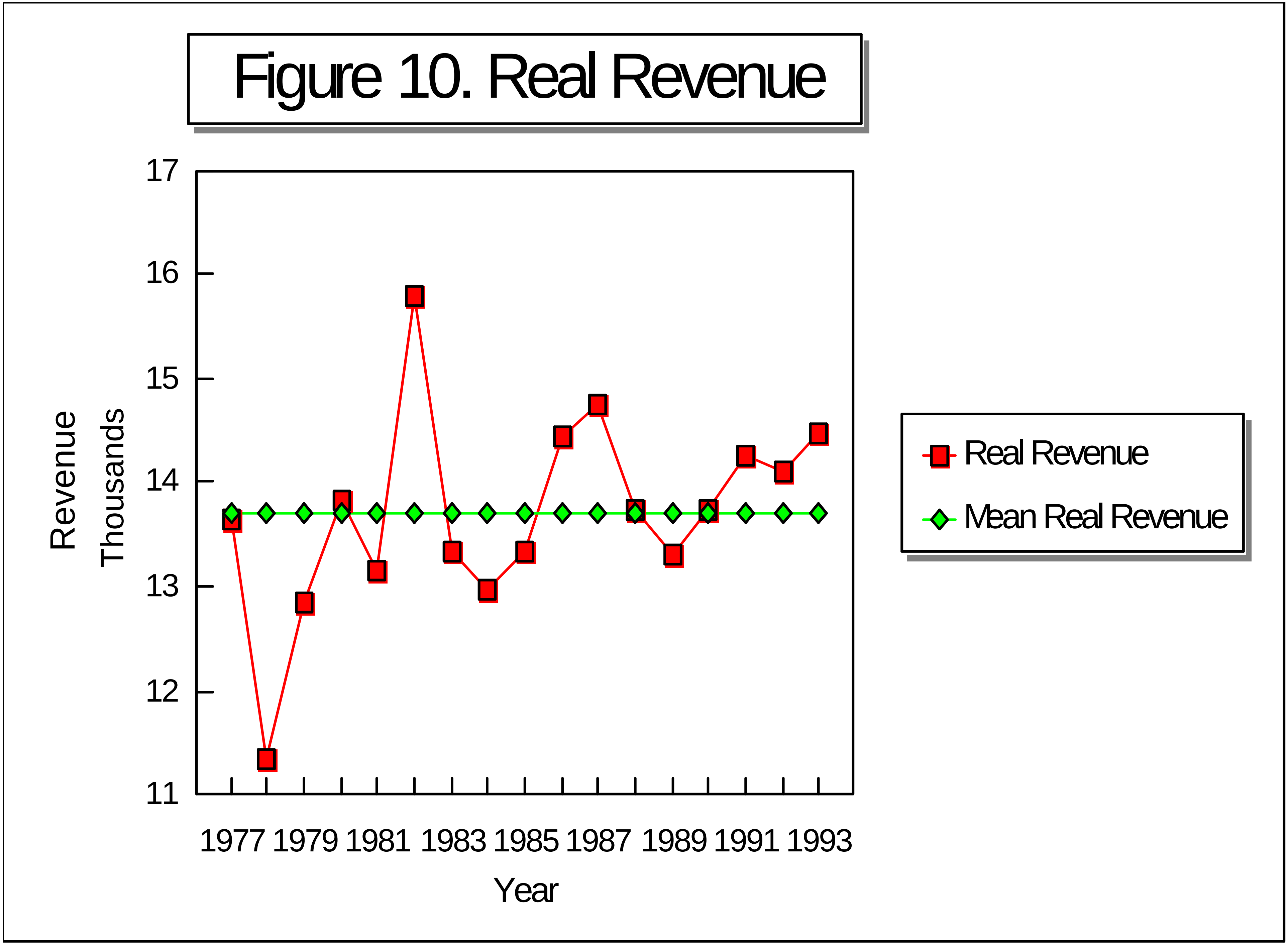

Figure 10. Real Revenue
17
16
15
14
13
12
11
Revenue
Thousands
1977 1979 1981 1983 1985 1987 1989 1991 1993
Year
Real Revenue
Mean Real Revenue

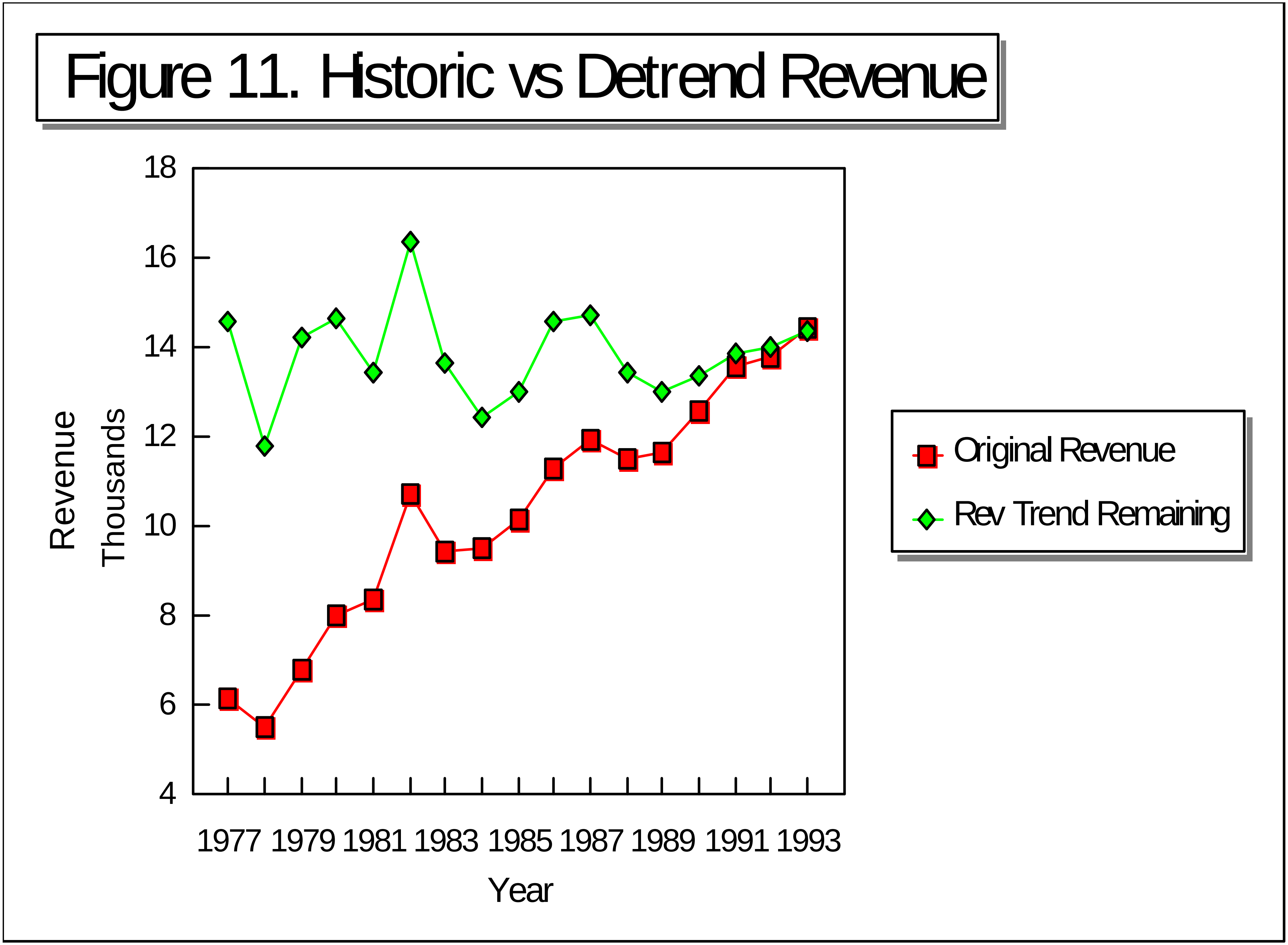
Figure 11. Historic vs Detrend Revenue
Revenue
Thousands
18
16
14
12
10
8
6
4
1977 1979 1981 1983 1985 1987 1989 1991 1993
Year
Original Revenue
Rev Trend Remaining

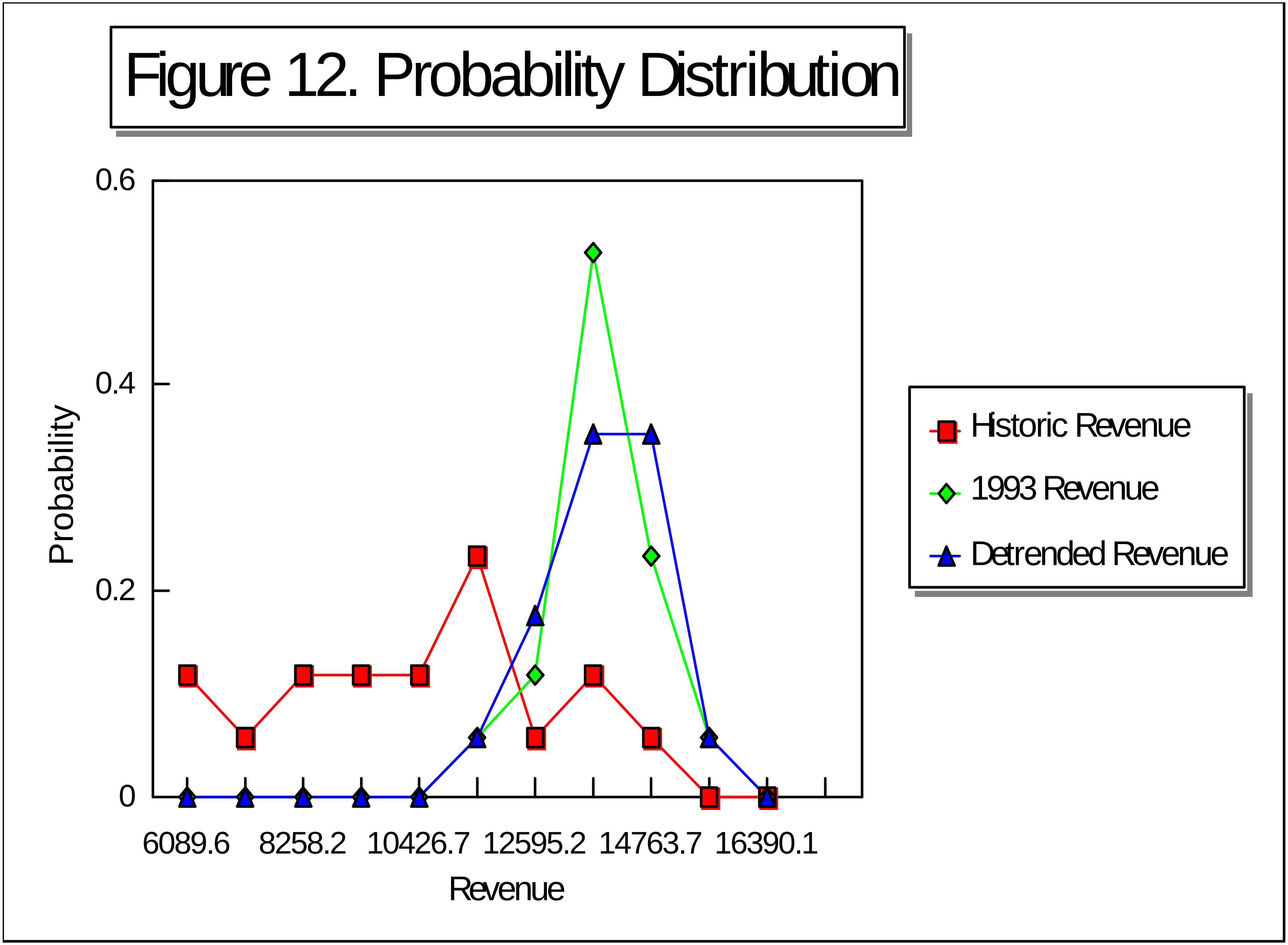
Figure 12. Probability Distribution
0.6
0.4
0.2
0
Probability
6089.6
8258.2
10426.7
12595.2
14763.7
16390.1
Revenue
Historic Revenue
1993 Revenue
Detrended Revenue