PN-ABL-324

AGRICULTURAL POLICY ANALYSIS PROJECT, PHASE II

Under contract to the Agency for International Development, Bureau for Science and Technology, Office of Agriculture
Project Office Hampden Square, 4800 Montgomery Lane, Suite 500, Bethesda, MD 20814 • Telephone (301) 913-0500
Telex 312636 • Fax (301) 652-7530 • Fax (301) 652-7791

A QUARTERLY PALAY PRODUCTION FORECASTING MODEL

December 1991

APAP II Methods and Guidelines
No. 403

Prepared for

Agricultural Policy Analysis Project, Phase II (APAP II)
A.I.D. Contract No. DAN-4084-Z-00-8034-00 and the Government of the Philippines, Department of Agriculture

Author: David Dawe, Harvard Institute for International Development

Prime Contractor: **Abt Associates Inc.**, 55 Wheeler Street, Cambridge, MA 02138 • (617) 492-7100

Subcontractors. **Harvard Institute for International Development,** Harvard University, One Eliot Street, Cambridge, MA 02138 • (617) 495-2164
Food Research Institute, Stanford University, Stanford, CA 94305-6084 • (415) 723-3941
North Carolina State University, Department of Economics and Business, Box 7645, Raleigh, NC 27695-7645 • (919) 737-7187
Abel, Daft & Earley, 1410 King Street, Alexandria, VA 22314 • (703) 739-9090
International Science and Technology Institute, 1129 20th Street, NW, Suite 800, Washington, D C 20036 • (202) 785-0831
International Food Policy Research Institute, 1776 Massachusetts Avenue, NW, Washington, D C 20036 • (202) 862-5600

A QUARTERLY PALAY PRODUCTION FORECASTING MODEL

David Dawe
Harvard Institute for International Development
Department of Agriculture

Appendix F
Rayna Lyn Roxas

Cover Illustration
Marieju Riñon

UNITED STATES OF AMERICA

FUNDED BY:
UNITED STATES AGENCY
FOR INTERNATIONAL DEVELOPMENT

TABLE OF CONTENTS

PREFACE iv

NOTE vi

ABOUT THE DISKETTE vii

INTRODUCTION 1

A. Basic Description of the Model 1
B. What is a Regression, and How can it be used to make Forecasts? 2
C. Format of the Forecasts 4

I. USING REGRESSION ANALYSIS AND RAINFALL DATA TO FORECAST PALAY PRODUCTION 7

A. Choosing the Independent Variables 7
B. Some Simple TSP Commands and Notation 8
C. The R^2 Statistic 10
D. The Coefficient on the Independent Variable 12
E. Residuals 12
F. Changing the Sample Range 14
G. Using more than one Independent Variable 15
H. The t-statistic 18
I. The Size of the Most Recent Residual 20
J. Adjusted Rainfall Series and the "Right Sign" of the Independent Variable 24
K. The Use of Time Trends 29
L. Dummy Variables 32
M. Creation of other Independent Variables 36
N. Use of (-1) for Rainfall Variables 37
O. Regressions with only a Constant Term 37
P. Another Example 38

II. IMPLEMENTATION OF THE MODEL 46

A. Updating the Data 46
B. Updating the Regression Equations 49
C. Making the Forecast 50
D. Batch Files to Automate Forecast Generation 54
E. Printing of the "Quarterly Palay Production Outlook" 56
F. A Timetable for Generating and Issuing the Outlook Reports 58

III. APPENDICES

A. List of Restrictions on the Choice of Independent Variables and on the Observations Used. 60
B. List of Weather Stations and their Abbreviations; Conventions for Naming Variables. 62
C. List of Current Regression Equations for Forecasting. 64
D. Batch Files Used to Generate the Forecasts. 86
E. A Summary of the Steps Required to Update the Model Every Quarter. . . 101
F. How to Print the "Quarterly Palay Production Outlook" 102
G. List of Common TSP Commands. 108
H. Performance of the Model vs. BAS Forecasts. 109

PREFACE

This manual will describe the quarterly palay production forecasting model in use at the Policy Analysis Division (PAD) of the Planning and Monitoring Service (PMS) at the Department of Agriculture (DA) of the Republic of the Philippines. This model uses econometric techniques and rainfall data gathered by the Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) to make forecasts of palay production on a quarterly basis. The model was developed jointly by Yolanda Azarcon, David Dawe, Generoso deGuzman and Emeterio Solivas, all consultants to the DA. It is currently being implemented using the software package TSP® (Time Series Processor), version 6.5, although any software that performs simple regressions could conceivably be used. The text of this manual will explain the model in detail and how it can be used to make quarterly forecasts of palay production. Some understanding of simple statistics and econometrics is assumed, although a strong attempt has been made to limit the amount of such knowledge that is required, and it is probably possible to follow everything without any prior knowledge whatsoever.

Thanks are due to the Accelerated Agricultural Production Project (AAPP) of the United States Agency for International Development (USAID), which funded the development of this model as well as the writing and printing of this manual. The primary vehicle for provision of this funding was the Agricultural Policy Analysis Project (APAP) Phase II (Contract number DAN-4084-Z-00-8034-00), a worldwide project of USAID for which Abt Associates is the prime contractor and the Harvard Institute for International Development (HIID) is one of the subcontractors. Supplementary funding from the Research and Training Program for Agricultural Policy (RTPAP), another component of AAPP, was also provided. Without the financial assistance from USAID, this project would not have been possible.

Last, but not least, a big thank you is also due to Rayna Lyn Roxas, who besides writing Appendix F, also set the entire document in Word Perfect.® While Word Perfect is indeed a tremendous piece of software, a major effort is still required to prepare a

manual of this size, and it is certain that the manual would never have made it to the printer without that major effort on Rayna's part.

NOTE

The regression equations used in the model to forecast palay production need to be updated every year with new data in order to generate reasonable forecasts year in and year out. The equations given in Appendix C of this manual are current as of the fourth quarter of 1991 when the manual went to press. If you come across this manual a couple of years after its printing and would like to see the most recent updates of the regression equations used in forecasting, please contact the Policy Analysis Division of the Department of Agriculture for this information.

ABOUT THE DISKETTE

The diskette provided with some versions of this manual contains the TSP files for each zone (Zone1.h, Zone1.wf, and so on for each of Zones 2-5). These files were created in version 6.54 of Micro-TSP. In addition, a copy of each of the batch files is provided (qtr11991.bat, qtr21991.bat, qtr31991.bat and qtr41991.bat). If one wishes to use these files, however, one must have a copy of the TSP software. Finally, the diskette also contains a copy of the Word Perfect® file Outlook, which is the sample report shown in Appendix F without the graph. If no diskette was included in your copy of the manual, please contact the Policy Analysis Division of the DA for copies of the relevant files.

INTRODUCTION

A. Basic Description of the Model

The purpose of this model is to provide an independent estimate of palay production in addition to the forecasts currently being made by the Bureau of Agricultural Statistics (BAS) of the DA. The results of the model should be interpreted as a "second opinion" of sorts, and are best used as a complement to BAS standing crop estimates. In order to make this forecast, the model uses rainfall data from PAGASA. Economic variables (e.g. prices of palay, fertilizer and corn) were also tested, but consistently failed to improve the fit of the regressions. Thus, they have been excluded from the model altogether.

In order to utilize the rainfall data properly, production is estimated independently for each of five agro-climatic zones, with production in any one zone being a function of rainfall in that zone. The zones were determined in the following fashion. First, it was decided that only regional[1] production data would be used in the model (not provincial). This decision was made because administrative regions are typically fairly homogeneous in terms of climate and because the use of provincial data would substantially increase the amount of work that would have to go into the design and updating of the model. Next, cross-correlations of regional production data were computed to determine which regions' production moved together most closely. For example, it was found using such analysis that production in Region I (Ilocos) correlated most closely with production in Region III (Central Luzon). Thus, Regions I and III are both placed in Zone 1.

It was also decided to limit the total number of agro-climatic zones to five or six, again in the interests of simplicity and ease of implementation. If a region could plausibly

[1]In order to avoid confusion, the word "zone" will be used throughout this manual to refer to the agro-climatic zones in this model (for which production is estimated), while the word "region" will refer to the administrative regions that are part of the bureaucratic apparatus of the Philippine government. Each zone is thus composed of one or more regions.

be placed in one of two or more zones by virtue of the cross-correlation production analysis, the final choice was made by looking at a rainfall map of the Philippines that groups various areas in the country according to how many wet (rainfall greater than 200 mm) and dry months (rainfall less than 100 mm) occur in the course of the year. The final zones that emerged from this procedure are listed in Table 1.

Table 1

Agro-Climatic Zone	Regions Included
Zone 1	Ilocos, Cordilleras, Central Luzon (Regions I, III, CAR)
Zone 2	Cagayan Valley (Region II)
Zone 3	Southern Tagalog,Bicol, Central Visayas, Eastern Visayas, Northern Mindanao (Regions IV, V, VII, VIII, X)
Zone 4	Western Visayas (Region VI)
Zone 5	Western, Central and Southern Mindanao (Region IX, XI, XII)

B. What is a Regression, and How can it be used to make Forecasts?

Production was estimated for each zone using ordinary least squares (OLS) regression analysis. OLS analysis attempts to explain movements in one variable (called the dependent variable) with movements in other variables (independent variables). For example, one might want to regress production of palay in zone 2 in October (the dependent variable) on rainfall in Tuguegarao in the preceding June (the independent variable). The idea here is that the amount of rainfall in June will influence production in October, and regression analysis attempts to determine the exact quantitative nature of the influence that rainfall has on production. Thus, palay production is termed the dependent variable because it depends on rainfall, while rainfall is termed the independent variable because it is independent of

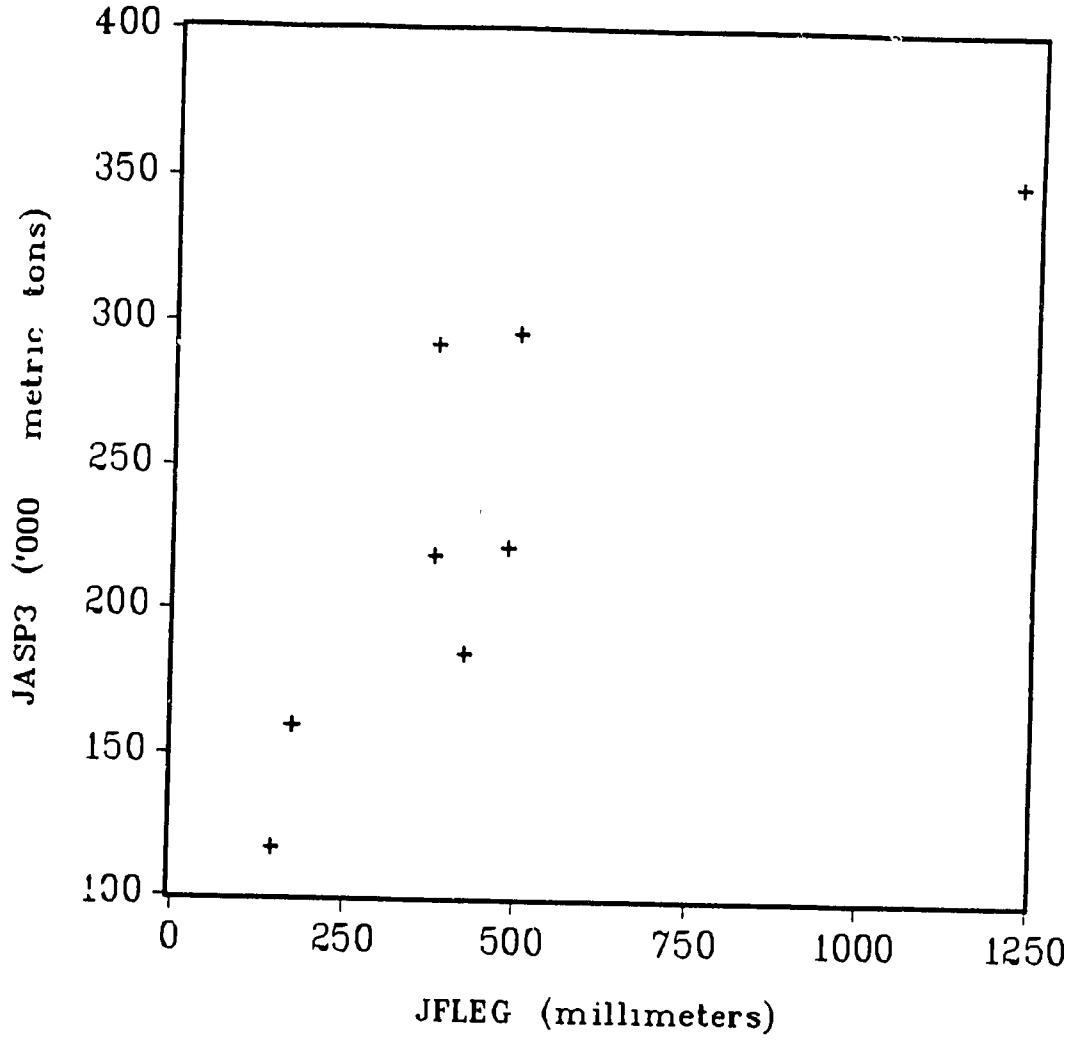
palay production (i.e. present palay production does not determine how much rain fell in the past).

Intuitively, a regression simply draws a straight line (the "drawing" is done with a mathematical formula that need not concern us at this stage) through a collection of points so that the points are as close as possible to the line. As an example, let us examine the following regression. The dependent variable (i.e. the variable we are trying to explain) will be third quarter palay production in zone 3 (which includes Bicol), and the independent variable will be total rainfall in Legaspi City in January and February. If we plot production versus rainfall on a graph for the years 1983-1990, we obtain the scatter plot shown in Figure 1. Clearly the points do not lie exactly on one straight line, but there is some trend to the points (i.e. they are not scattered randomly on the graph). A regression draws a single line that comes as close as possible to going through all of the points. The regression line for this particular set of data is shown in Figure 2.

In order to develop the model, independent variables must be found that explain production well. In the context of regression analysis, choosing a suitable independent variable to explain palay production involves finding a variable that, when plotted on a scatter diagram with production on the vertical axis, results in as perfect of a straight line as possible.

Once the regression line is drawn, it can be used to make predictions about the future in the following way. The first step is to take the new value of the independent variable and plot it on the horizontal axis. In this example, it is the total rainfall in Legaspi City in January and February of 1991. Next, one reads straight up to the regression line from that point on the horizontal axis. Then, one reads from the regression line straight over to the vertical axis and records the value on that axis. This number (203,384 metric tons) is the predicted value of production in the third quarter of 1991 given the new value of rainfall in Legaspi City for January and February of 1991 (which is 324.6 mm). This procedure is shown on the graph in Figure 2.

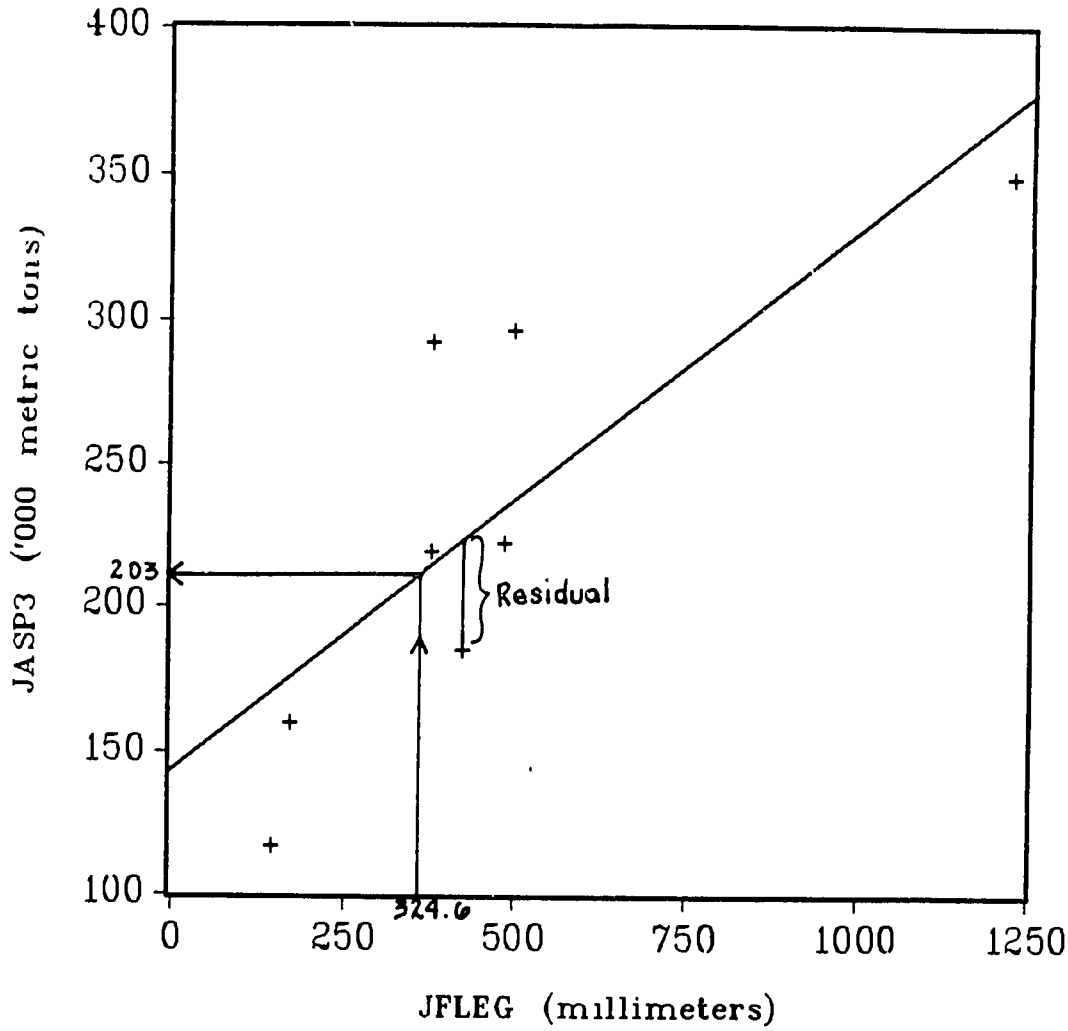
Figure 1



C. Format of the Forecasts

The production and rainfall data used to calculate the initial regressions covered the years 1975-1989, and were kindly provided by BAS and PAGASA. In order to implement the model, new data will continually be needed from these two sources so that regression parameters can be updated. Without continual updating, the predictions of the model are likely to be useless. Thus, the updating procedure is a critical part of the implementation.

Figure 2



Production estimates were made by calendar quarter or by month for each zone, whichever proved to be most effective in any given case. The final forecasts are always presented by calendar quarter because this is the way BAS presents its forecasts. In some cases, this is perhaps unfortunate because the cropping season cuts across quarters. For example, the dry season crop comes in March and April, but these months can not be combined in the model forecasts if comparability with BAS forecasts is desired. As another example, the wet season crop in Western Visayas begins in September (third quarter) and

continues into the fourth quarter. It was decided, however, that comparability with existing BAS forecasts was of primary importance, and so forecasts correspond exactly with calendar quarters.

With the preceding as background, we will proceed to a more detailed description of the model and its implementation.

I. USING REGRESSION ANALYSIS AND RAINFALL DATA TO FORECAST PALAY PRODUCTION

A. Choosing the Independent Variables

The easiest way to grasp the essence of the model is to go through an example. As our first example, we will attempt to forecast September palay production in zone 4 (Region VI) using rainfall data. It seems reasonable to expect that September production will be influenced by rainfall in the months of March, April, May and June. It is also true that production is probably influenced by rainfall in July and August, but here the model runs up against the constraint of being useful to the DA. Since September is part of the third quarter, a forecast of September production needs to be made at the beginning of July, which is the beginning of the third quarter and the time when BAS releases its standing crop estimates for that quarter. Thus, it is not possible to use July or August rainfall data to predict September production if the prediction is to be made by the third week of July. In fact, the situation is even worse than that. Due to the time required to process PAGASA field data, even June rainfall data is sometimes not available until very late in July or early in August. Thus, it is not feasible to use even June rainfall data in predicting September production.

We thus need to confine ourselves for the moment to the use of March, April and May rainfall data for Iloilo and Roxas City, which are the two main weather stations in the Western Visayas. The first step is to run some exploratory regressions to explain the dependent variable (September palay production in zone 4) with the aid of some independent variables. Before we run the exploratory regressions, however, we first need to discuss in more detail how to choose possible independent variables.

In the model, there are literally hundreds (if not thousands or millions) of possible variables to choose in order to explain production. Thus, we need to come up with some a priori restrictions on the variables that we can use for this purpose (over the course of this manual, there will be many such restrictions. For the convenience of users implementing the model, all of these restrictions are summarized in one place in Appendix A). The first

such restriction is an obvious one, namely that we will not attempt to explain production in one zone with rainfall data from another zone. Thus, rainfall in Dagupan (zone 1) can not be used to explain production in the Cagayan Valley (zone 2). This is obvious and needs no further explanation.

The second restriction is that rainfall data four, five or six months in advance of the month whose production is trying to be explained is the most logical choice, and this data should be tried before any other rainfall data is used. It does not make much sense to use rainfall data seven months or more prior to the production month because the crop is not yet in the ground at that point (nor is land preparation likely to have started). Rainfall three months or less prior to production, on the other hand, although it may influence production, is likely to run into the problem discussed above, namely that the data will not be available at the time the forecast needs to be made. This latter restriction, however, will depend on the particular month whose forecast is being made. For example, September production forecasts can not be made with June rainfall data (three months before) because of the time required to process PAGASA data. When predicting July production, however, it may be feasible to use rainfall data from two months before (i.e. May) because the May data is likely to be available by the first week in July. Thus, more data will effectively be at the disposal of the forecaster when making predictions for months at the beginning of calendar quarters.

B. Some Simple TSP Commands and Notation

Taking into account the above restrictions, we see that we are constrained to using rainfall data from Iloilo and Roxas City from March, April and May in order to explain September production in the Western Visayas. Although this set consists of just six data points, there are many different ways in which to use the available data. Before explaining how to do this, however, we will begin to introduce some notation for the variables. This notation corresponds to how variables are named in TSP. In TSP, all variable names must be eight letters or less. We will adopt the convention of beginning variable names with the first three letters of the month from which the data comes, followed by three letters that

designate the weather station from which the data comes (all eight letters will not be used at this point because we will later need the other two letters for naming more sophisticated variables). For example, rainfall in APRil in ILOilo will be designated by APRILO.[2] Our choice set for the independent variables to explain September palay production thus consists of MARILO, APRILO, MAYILO, MARROX, APRROX, and MAYROX.

In order to run some regressions that attempt to explain September production on the basis of rainfall data, we first need to access the TSP software.[3] From an IBM compatible computer using MS-DOS®, this can be done by typing the command TSP while in whatever subdirectory contains the TSP software. For example, if the software is in the subdirectory TSP, type TSP at the C:\TSP> prompt. The next step is to load the file ZONE4, which contains all the relevant data for zone 4. If the file is on a diskette in the A drive, this is done by typing LOAD A:\ZONE4 (followed by pressing the ENTER key). We are now ready to run some regressions.

Ordinary least squares (OLS) regressions are run using the command LS. The command LS is followed by the name of the dependent variable, which in this case is SEPP, where the second P stands for production. After the dependent variable comes the letter C (for constant[4]), followed by the names of the independent variables. As an example, the command LS SEPP C MAYILO (followed by pressing the ENTER key) would run a

---

[2]A complete listing of all weather stations and their three letter abbreviations is given in Appendix B.

[3]For readers who are not already familiar with TSP, it will probably help to have at hand a computer with TSP installed while going through the text. The commands can then be entered as they are discussed. TSP printouts are provided in the tables accompanying the text, however, for those readers who do not have access to TSP while reading the text.

[4]The constant in a regression signifies the point where the regression line intersects the vertical axis, or in other words what the value of production would be if all the independent variables had the value zero. If the reader desires a more detailed discussion of the constant, he or she should consult an elementary econometrics text. REGRESSIONS FOR THIS MODEL SHOULD NOT BE RUN WITHOUT A CONSTANT FOR ANY REASON WHATSOEVER, since the traditional interpretation of many of the regression statistics (including the R^2) changes radically when constants are omitted from the regression.

regression of September production on May rainfall in Iloilo (and a constant). Running this regression gives the results shown in Table 2. In deciding whether or not this regression will be useful to us in explaining past production and trying to predict future production, one needs to examine the regression statistics generated by TSP. In this manual, we will only discuss a few of the most important statistics, and we will not discuss them in textbook detail. The text of this manual is not intended to substitute for a course in econometrics, and the reader should consult an econometrics textbook if a more in-depth presentation on any particular topic is desired.

Table 2

```
LS // Dependent Variable is SEPP
Date: 11-14-1991 / Time: 14:57
SMPL range: 1975  -  1990
Number of observations: 16
=================================================================
     VARIABLE   COEFFICIENT   STD. ERROR      T-STAT.    2-TAIL SIG.
=================================================================
        C        145314.79     27120.579     5.3581007     0.000
     MAYILO      158.07338     183.27594     0.8624884     0.403
=================================================================
R-squared                0.050454    Mean of dependent var   163450.8
Adjusted R-squared      -0.017371    S.D. of dependent var   67924.89
S.E. of regression       68512.30    Sum of squared resid    6.57E+10
Durbin-Watson stat       1.597622    F-statistic             0.743886
Log likelihood          -199.7911
=================================================================
```

C. The R^2 Statistic

The first statistic we will discuss is the R^2 statistic. The R^2 more or less summarizes how well the independent variables explain the dependent variable. The R^2 varies from 0 to 1, with high values indicating that a high proportion of the variance of the dependent variable is explained by the independent variables as a group. A value of 0 indicates that none of the variance is explained, while a value of 1 indicates that all of the variance is explained (in practice, values of exactly 0 or 1 are never obtained - intermediate values are the rule). For example, an R^2 of 0.69 would imply that 69% of the variation in the

dependent variable is being explained by the independent variables. Thus, all other things being equal, a high R^2 is better than a low R^2.

From Table 2, one can see that the R^2 for this regression is very low, 0.05. For the palay production forecasting model, values for R^2 typically run from 0.45 to 0.80. An R^2 below 0.45 can be acceptable in some circumstances, but 0.05 is clearly very, very low and is simply not acceptable for the purposes of this model. Thus, we need to find an alternative regression that explains September production better.

Before running any more regressions, however, we need to stop and explain that the process of finding a suitable regression can be fairly time consuming. IT IS IMPORTANT TO EMPHASIZE, HOWEVER, THAT IT IS NOT SIMPLY A HIT OR MISS PROCEDURE. THE PROCESS OF FINDING A SUITABLE REGRESSION REQUIRES ONE TO EXAMINE CAREFULLY THE REGRESSION STATISTICS PROVIDED BY TSP. THESE STATISTICS WILL PROVIDE CLUES ABOUT WHAT REGRESSIONS TO TRY NEXT SO THAT THE MOST USEFUL REGRESSION CAN BE FOUND. Adopting such a strategy also makes the work of finding a suitable regression more interesting because it becomes more like solving a mystery. Simply running regression after regression in a random fashion quickly becomes boring, besides the fact that it will result in poor model performance.

There is some trial and error involved, however, so the next step might be to run the following regressions (remember, all of these commands must be followed by pressing the ENTER key):

```
LS SEPP C MARILO
LS SEPP C MARROX
LS SEPP C APRILO
LS SEPP C APRROX
LS SEPP C MAYROX
```

In other words, just experiment with simple rainfall variables (chosen subject to the restrictions noted above) at first and see what happens. Most of the above regressions fail

miserably (i.e. very low R^2), but a couple of them show some promise, namely the ones with APRILO and MAYROX. The R^2 for these regressions are still fairly low (about 0.10), but since these are the best, it may pay us to look more closely at them.

D. The Coefficient on the Independent Variable

Another important number produced by running the regression in TSP is the coefficient on the independent variable(s). The magnitude of this coefficient indicates the quantitative impact on production of a change in rainfall. Let us look ahead for a moment to Table 8, where we can see that the coefficient on APRILO is 452. This means that for every increase in rainfall of one millimeter (the units of rainfall used in the model), there will be, on average, an increase in palay production in September of 452 metric tons (which are the units of palay production used in the model). This coefficient is the basis for making forecasts of palay production, as will be seen more clearly in Section C of Chapter II.

E. Residuals

An important feature of TSP is the residual plot that it generates for each regression. Residuals are the difference between past actual values of production and the value of production that would have been predicted by the model given the values of the independent variables (recall Figure 2 and the discussion in Section B of the Introduction on using regressions to make predictions). Referring back to Figure 2, residuals are the vertical distance between the actual plotted point and the regression line. Obviously, a regression "fits" the points better if the residuals are small, and this is what the R^2 statistic roughly measures. A high value of R^2 implies that the residuals are small, and a low R^2 implies that the residuals are large.

The R^2 statistic is a summary statistic, and it roughly indicates what the sum of all the residuals is. It does not allow one to examine the residuals for individual points, however. This is precisely what the residual plot in TSP allows one to do, though, so let us

see what information is contained in it. To see the residual plot after the regression has been run (with the command LS SEPP C APRILO followed by ENTER), press ENTER (once again) followed by S (for screen) at the "Display the Residual, Actual and Fitted values?" prompt. As shown in Table 3, this command produces a list of all actual values, predicted (fitted) values and residuals (the difference between actual and predicted values). In addition, it also shows graphically all of the residuals. Each residual is represented by an asterisk, and the closer the asterisk is to the center line, the smaller the residual is in absolute value. In other words, the closer the asterisk is to the center line, the more closely the predicted value matches the actual for that particular year.

The first thing one should notice about the residual plot in Table 3 is that many of the large residuals (represented by asterisks far away from the center line) occur in the years 1975-1981. What this indicates is that the regression line fits the data reasonably well for the years 1982-1990, but not so well for years prior to 1982. The pattern for the regression LS SEPP C MAYROX is similar, and this may indicate that for some reason or other the period 1975-1981 in the Western Visayas was different than the way things are now.

Table 3

```
LS // Dependent Variable is SEPP
Date: 11-14-1991 / Time: 14:59
SMPL range: 1975  -  1990
Number of observations: 16
==================================================================
    VARIABLE    COEFFICIENT    STD. ERROR      T-STAT.    2-TAIL SIG.
==================================================================
       C         142571.64     23865.428     5.9739821      0.000
    APRILO       296.84180     242.32222     1.2249880      0.241
==================================================================
R-squared                0.096809    Mean of dependent var    163450.8
Adjusted R-squared       0.032295    S.D. of dependent var    67924.89
S.E. of regression       66819.06    Sum of squared resid     6.25E+10
Durbin-Watson stat       1.434039    F-statistic              1.500596
Log likelihood          -199.3907
==================================================================
```

```
==================================================================
            Residual Plot                 obs RESIDUAL  ACTUAL   FITTED
==================================================================
|     *     :           |         :       | 1975 -102013.  83571.0  185584.
|           :    *      |         :       | 1976 -32524.6  116162.  148687.
|           *           |         :       | 1977 -64914.6  77657.0  142572.
|           :*          |         :       | 1978 -51509.7  130067.  181577.
|           :           |   *     :       | 1979  13806.1  193572.  179766.
|           :           |      *  :       | 1980  32108.4  176847.  144739.
|           :           |         :    *  | 1981  103957.  270395.  166438.
|           :      *    |         :       | 1982 -17344.6  143275.  160620.
|           :*          |         :       | 1983 -55649.1  87635.0  143284.
|           :           |   *     :       | 1984  13744.3  176145.  162401.
|           :           |  *      :       | 1985  11398.5  233405.  222007.
|           :       *   |         :       | 1986 -8591.26  148585.  157176.
|           *           |         :       | 1987 -65862.6  78401.0  144264.
|           :           |         :    *  | 1988  99521.4  267384.  167863.
|           :           |  *      :       | 1989  5184.33  168149.  162965.
|           :           |         :     * | 1990  118689.  263962.  145273.
==================================================================
```

F. Changing the Sample Range

If such a pattern is observed, a good idea is to run the regression again, but this time exclude the early years that do not seem to fit so well. To do this in TSP, one must use the SMPL command, which stands for sample. To change the years for which the regression is run, one types the command SMPL 1982 1990, followed by pressing the ENTER key.[5] Now type the regression command once again (LS SEPP C APRILO), and let us examine the new results, which are presented in Table 4. The value of the R^2 has increased substantially (to 0.21), which means we are moving in the right direction. With the new sample range, let us rerun some of the other regressions above that failed to produce good results the first time. If we do this, we find in particular that LS SEPP C MAYILO provides a good fit to the data (see Table 5 and contrast it with Table 2, which is the same regression but over a different sample range. Note in particular the big increase in the R^2 to 0.42).

[5]Almost all commands in TSP need to be followed by ENTER in order to execute them. From now on, unless stated otherwise, it is to be assumed that all commands should be followed by pressing the ENTER key, and the manual will not state this explicitly anymore.

Table 4

```
LS // Dependent Variable is SEPP
Date: 11-14-1991 / Time: 15:02
SMPL range: 1982  -  1990
Number of observations: 9
==================================================================
     VARIABLE   COEFFICIENT   STD. ERROR      T-STAT.   2-TAIL SIG.
==================================================================
        C        147273.43     29589.928     4.9771472     0.002
     APRILO      392.33164     289.09094     1.3571219     0.217
==================================================================
R-squared                0.208304    Mean of dependent var   174104.6
Adjusted R-squared       0.095205    S.D. of dependent var   69434.73
S.E. of regression       66046.82    Sum of squared resid    3.05E+10
Durbin-Watson stat       1.836143    F-statistic             1.841780
Log likelihood          -111.5226
==================================================================
```

Table 5

```
LS // Dependent Variable is SEPP
Date: 11-14-1991 / Time: 15:04
SMPL range: 1982  -  1990
Number of observations: 9
==================================================================
     VARIABLE   COEFFICIENT   STD. ERROR      T-STAT.   2-TAIL SIG.
==================================================================
        C        111406.06     33831.053     3.2930118     0.013
     MAYILO      489.91705     219.23758     2.2346399     0.061
==================================================================
R-squared                0.416356    Mean of dependent var   174104.6
Adjusted R-squared       0.332978    S.D. of dependent var   69434.73
S.E. of regression       56708.31    Sum of squared resid    2.25E+10
Durbin-Watson stat       2.439002    F-statistic             4.993616
Log likelihood          -110.1506
==================================================================
```

G. Using more than one Independent Variable

At this point, it is time to run some regressions with more than one independent variable. The intuition behind a regression with more than one independent variable (this is called multiple regression) is similar to that with only one, except that now the scatter diagram of Figure 1 must be visualized in three (or more) dimensions. Regression analysis still draws a line through all of the points as best it can, but now the line must be visualized

in n-dimensional space instead of on a piece of paper. Such visualization is exceedingly difficult for most people, and there is no need to worry about it if it poses problems for you. The two dimensional case of Figure 1 is still useful as an analogy, however, and the reader should keep this figure in mind even when running regressions with many independent variables.

In principle, it is possible to run regressions with as many independent variables as one likes. If the number of independent variables (including the constant C, which should ALWAYS be included) exceeds the number of observations, however, the error message "Insufficient number of observations availabie" will result. On the other hand, if a regression is run with exactly the same number of independent variables as observations, the R^2 statistic will always be equal to one. Given that high R^2 values are good, one might think that one should always include as many independent variables as there are observations. This involves a serious misunderstanding of regression analysis, however.

If one includes as many independent variables as observations, the R^2 will always be one, regardless of what the independent variables are. For example, if the dependent variable is SEPP, and the sample range is 1982 - 1990 (nine observations), choosing nine independent variables (including the constant) such as the budget deficit in Chile lagged 83 years, the number of times your family ate chicken adobo, how many days you were sick with a cold and how many times your cousin visited Disneyland will result in a perfect R^2 of 1.00 as long as there are nine of them. Clearly these variables are not affecting palay production in the Western Visayas, but they are giving high values of R^2 nonetheless.

What is happening? A return to the two dimensional diagram of Figure 1 will help here. If we were to plot only two points on this diagram, it would always be possible to draw a straight line that goes through the two points perfectly (i.e. an R^2 of 1.00). This will be true no matter what the independent variable on the horizontal axis is. For example, let us plot (in our imagination) the budget deficit in Chile lagged 83 years and palay production in the Western Visayas in September for the sample period 1980-1981. This will result in two points, and it will be possible to draw a straight line that goes right through both of them. The situation here is one of two observations and two independent variables (the budget deficit in Chile lagged 83 years and the ubiquitous constant C), which as stated

above always leads to a perfect R^2 of 1.00 (i.e. the line goes exactly through all the points - two in this case). The situation of using nine independent variables to explain nine observations is entirely analogous in nine dimensional space (whatever that is), and it is hopefully obvious that we are not explaining anything with such methods.

Using six, seven or eight independent variables in conjunction with nine observations is not necessarily much different. Although perfect R^2's of 1.00 are not likely to result, very high values will nevertheless be obtained, and this will be the case regardless of what independent variables are chosen. In essence, then, using lots of independent variables relative to the number of observations is cheating. Regressions obtained using such methods, while yielding high values of R^2, will be of no help whatsoever in making predictions about future palay production. A useful rule of thumb is to never use more than 3 (or 4 at the absolute most) independent variables in addition to the constant C. This applies if the sample range is from 1975-1990. If the sample range has substantially fewer observations (e.g. 8, 9, or 10), then one should use no more than 2 independent variables in addition to the constant.

A related rule of thumb is to not reduce the number of observations below 8 under any circumstances. Thus, although this discussion initially focused on discarding certain observations at the beginning of the sample if they do not appear to fit too well, there is a limit to such an approach. In addition, the discarded observations should always be consecutive and should come from the beginning of the sample (e.g. do not discard the observations for 1988-1990).

Bearing the above restrictions on the number of independent variables in mind, let us return to the topic of multiple regression. Since both APRILO and MAYILO appear to result in high values of R^2, let us use both of them in a regression at the same time. Remember that we are still in the sample range 1982-1991 (9 observations), so the maximum number of independent variables that we can use is two. The command to run this regression is LS SEPP C APRILO MAYILO, and if we run this regression the value of R^2 is 0.70, as can be seen in Table 6. This is very good, but before we can conclude that the high R^2 is indicative of a good regression, we must discuss another of the statistics generated by TSP, the t-statistic.

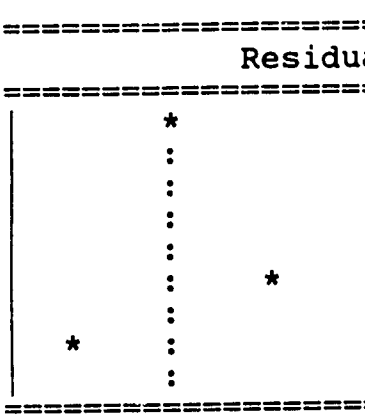
Table 6

```
LS // Dependent Variable is SEPP
Date: 11-14-1991 / Time: 15:06
SMPL range: 1982  -  1990
Number of observations: 9
```

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	74637.381	30692.217	2.4318016	0.051
MAYILO	533.07496	172.00007	3.0992717	0.021
APRILO	456.87868	194.73450	2.3461620	0.057

R-squared	0.695609	Mean of dependent var	174104.6
Adjusted R-squared	0.594145	S.D. of dependent var	69434.73
S.E. of regression	44234.62	Sum of squared resid	1.17E+10
Durbin-Watson stat	2.985256	F-statistic	6.855733
Log likelihood	-107.2212		

Residual Plot



obs	RESIDUAL	ACTUAL	FITTED
1982	-43366.4	143275.	186641.
1983	6890.21	87635.0	80744.8
1984	12456.5	176145.	163689.
1985	4948.84	233405.	228456.
1986	13940.7	148585.	134644.
1987	-21176.4	78401.0	99577.4
1988	48484.9	267384.	218899.
1989	-67626.4	168149.	235775.
1990	45448.1	263962.	218514.

H. The t-statistic

The t-statistic, or "t-stat" for short, is calculated by dividing the coefficient on any independent variable by its standard error (which is also given in the set of statistics generated by TSP). Basically, the t-statistic is a measure of how certain we are that the given independent variable is "truly" explaining some of the variation in the dependent variable, as opposed to being correlated with the dependent variable in a purely coincidental fashion. In order to understand this more clearly, look at the column in Table 6 labelled "2-TAIL SIG." The number in that column is, roughly speaking, the probability that the

correlation between the dependent variable and the given independent variable that actually exists would arise due to random chance. A very low number in that column indicates that it is highly unlikely (low probability) that the relationship between the dependent variable and the independent variable would arise as a result of random chance, and thus, that it is highly likely that the correlation between the two is due to some nonrandom, structural relationship. For example, we see in Table 6 that the two tail significance for MAYILO is 0.021. This means that there is only a 2.1% chance that the observed relationship between SEPP and MAYILO would arise due to random chance. This is a fairly low number, so the presumption is that MAYILO actually has an influence on SEPP.

A low number in the two tail significance column corresponds to a high t-statistic in absolute value. As the two tail significance approaches zero (one), the t-statistic approaches plus or minus infinity (zero). Thus, t-statistics can take on any value whatsoever. As a general rule of thumb, we will say that a t-statistic of greater than 2 indicates the given relationship is probably not arising due to random chance, and that there is indeed a structural relationship between the dependent and independent variable. In such a case, the variable should probably be included in the regression to predict palay production (provided it is of the right sign, which we will discuss in section J below). For our purposes, a t-statistic of 2 corresponds to a two tail significance of about 7%.[6]

For variables with t-statistics of less than 1.5 in absolute value (corresponding to a two tail significance of roughly 17%), we will say that there is a good chance that the correlation between the two variables is due to random chance alone. The given independent variable should then be excluded from the final regression. Variables with intermediate t-statistics (between 1.5 and 2.0 in absolute value) are borderline and can either be excluded or included depending on how much their presence improves the regression.

---

[6]The exact relationship depends on something known as the degrees of freedom, which will vary from regression to regression. This is beyond the scope of this discussion, and interested readers should consult an elementary statistics textbook.

As noted above, the R^2 is a summary statistic that explains how well the independent variables as a group explain movements in the dependent variable. The t-statistic, however, measures (in a sense) the contribution of each independent variable by itself. The higher the t-statistic is in absolute value, the more certain we are that that particular variable is truly explaining part of the variance in the dependent variable.

For the purposes of the above mentioned rule of thumb, the constant C is not to be considered an independent variable. In other words, even if the t-statistic for the constant is very low, it MUST still be included in the regression.

If an independent variable has a t-statistic that is high in absolute value, that independent variable is often called "significant" in econometric jargon. Thus, from now on, if an independent variable has a high t-statistic in absolute value, we will term it significant.

It should be noted that the t-statistic is not some immutable characteristic of an independent variable. Not only does it depend on which dependent variable we are considering (which should be obvious), it also depends on the sample range, and more importantly, on what other independent variables are included in the regression. When deciding whether or not to include an independent variable in a regression, the t-statistic that matters is the one in the final regression (and not one that was the result of some earlier regression).

Let us now examine the t-statistics for the independent variables in the above regression. Referring back to Table 6, we see that the t-statistics for APRILO and MAYILO are 2.3 and 3.1 respectively. Both of these satisfy the rule of thumb that the t-statistic be greater than two in absolute value, so it is a good idea to leave them in the regression. We are now at the point where we can begin to do some final "tuning-up" of the regression for predicting palay production in the Western Visayas in September.

I. The Size of the Most Recent Residual

If you do a little experimentation on your own, you will notice that running the regression LS SEPP C MAYILO APRROX gives an R^2 of 0.72 (see Table 7), which is even higher than the 0.70 above. This is a good thing to note, and argues that it should be used

instead of the former LS SEPP C MAYILO APRILO. If one looks at the residual plot for both regressions, though (Tables 6 and 7), one will see that the residual for 1990 is rather large in either case. This is in general not good, because it implies that the regression is not doing so well at predicting in the later years of the sample, which is what we really care about.

There are a couple of rules that can guide you in deciding whether or not such a residual is unacceptably large. In order to make this decision properly, the most important thing to note is how the scaling on the residual graph in TSP is done. In these residual graphs, the asterisks representing the residuals are always spread out over the graph so that the whole graph is more or less occupied. Thus, even if the largest residual was only 0.01% of the actual value of production (which is a very small number), it would still be far away from the center line. This is because asterisks are placed far away from the center line not if the residual is large relative to production (which is what we are interested in), but if the residual is large relative to other residuals. Since some residuals will always be large relative to other residuals (by definition), there will always be some asterisks far away from the center line. This is nothing to worry about if the size of the residuals relative to production is small. Thus, what we should do in assessing the 1990 residual is to first calculate how big it is relative to actual production (as a percentage). In this example (LS SEPP C APRILO MAYILO), we would divide 45,448.1 (the residual) by actual production of 263,962 to come up with an answer of 17%.

Table 7

```
LS // Dependent Variable is SEPP
Date: 11-14-1991 / Time: 15:08
SMPL range: 1982  -  1990
Number of observations: 9
==================================================================
    VARIABLE   COEFFICIENT   STD. ERROR      T-STAT.    2-TAIL SIG.
==================================================================
       C        71869.978     29784.362     2.4130105     0.052
    MAYILO      544.32807     165.87633     3.2815296     0.017
    APRROX      515.39051     203.15519     2.5369301     0.044
==================================================================
R-squared                0.718410    Mean of dependent var    174106.6
Adjusted R-squared       0.624546    S.D. of dependent var    69434.73
S.E. of regression       42545.64    Sum of squared resid     1.09E+10
Durbin-Watson stat       3.072690    F-statistic              7.653770
Log likelihood          -106.8708
==================================================================
```

```
==================================================================
            Residual Plot                 obs RESIDUAL  ACTUAL   FITTED
==================================================================
|      :      *      |            :      | 1982 -20010.4  143275.  163285.
|      :             |     *      :      | 1983  9875.25  87635.0  77759.7
|    * :             |            :      | 1984 -49654.0  176145.  225799.
|      :             |          * :      | 1985  35767.4  233405.  197638.
|      :             |     *      :      | 1986  13397.9  148585.  135187.
|      :      *      |            :      | 1987 -17822.5  78401.0  96223.5
|      :             |          * :      | 1988  34611.9  267384.  232772.
|    * :             |            :      | 1989 -52599.8  168149.  220749.
|      :             |            :*     | 1990  46434.4  263962.  217528.
==================================================================
```

A good rule of thumb to use here is that residuals of 10% or less are very small and need not be worried about at all. Residuals of much more than 25% for the most recent observation are rather large and may be grounds for abandoning the regression. Something in between these numbers is something to be noted, but if the regression is otherwise a good one and other regressions can not improve upon the last residual, then it is not worth worrying about. Keeping this in mind, we will try to improve a bit on the 1990 residual. If we can not improve the regression, however, we will consider what we have done already to be acceptable (high R^2, high t-statistics, and 1990 residual less than 25%).

First of all, we can notice that the 1982 residual is rather large. Thus, by again changing the sample range, we may be able to improve the fit of the regression. Remember, though, that we should have a minimum of 8 observations in the regression, and abandoning the 1982 observation will leave us with exactly 8. Thus, we can not delete any more observations after this. Using the command SMPL 1983 1990, and then rerunning the regressions LS SEPP C APRILO MAYILO and LS SEPP C MAYILO APRROX, we come up with values of R^2 of 0.74 and 0.72 respectively. Thus, the overall fit of both regressions is still good. Let us now look at the residual plots, however, which is why we are running these regressions in the first place (Tables 8 and 9). From these plots, we can notice that the 1990 residual has declined from 17% to 14% and 16% respectively for the two regressions. These are not major improvements, but they are better than nothing, especially since the R^2 values have increased slightly in the process and the t-statistics are all greater than 2. Thus, as our final regression, we will retain LS SEPP C APRILO MAYILO over the range SMPL 1983 1990. The regression output for this regression is shown in Table 8.

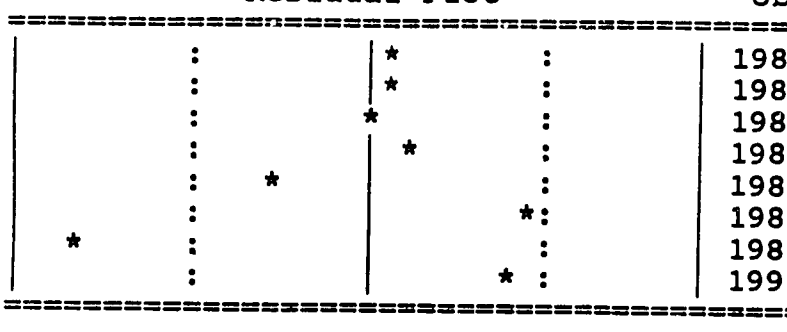
Table 8

LS // Dependent Variable is SEPP
Date: 11-14-1991 / Time: 15:19
SMPL range: 1983 - 1990
Number of observations: 8

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	77665.544	30522.860	2.5445041	0.052
MAYILO	554.88374	171.53980	3.2347230	0.023
APRILO	452.31329	192.85579	2.3453446	0.066

R-squared	0.744245	Mean of dependent var	177958.3
Adjusted R-squared	0.641944	S.D. of dependent var	73192.69
S.E. of regression	43796.91	Sum of squared resid	9.59E+09
Durbin-Watson stat	3.348162	F-statistic	7.274995
Log likelihood	-94.97004		

Residual Plot



obs	RESIDUAL	ACTUAL	FITTED
1983	3668.00	87635.0	83967.0
1984	7338.69	176145.	168806.
1985	1851.30	233405.	231554.
1986	9601.83	148585.	138983.
1987	-25092.4	78401.0	103493.
1988	41536.3	267384.	225848.
1989	-75649.2	168149.	243798.
1990	36745.4	263962.	227217.

Table 9

LS // Dependent Variable is SEPP
Date: 11-14-1991 / Time: 15:21
SMPL range: 1983 - 1990
Number of observations: 8

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	75296.264	32656.954	2.3056732	0.069
MAYILO	552.37390	178.34606	3.0972027	0.027
APRROX	487.81426	224.78153	2.1701706	0.082

R-squared	0.723410	Mean of dependent var	177958.3
Adjusted R-squared	0.612774	S.D. of dependent var	73192.69
S.E. of regression	45545.95	Sum of squared resid	1.04E+10
Durbin-Watson stat	3.154020	F-statistic	6.538656
Log likelihood	-95.28331		

```
==========================================================================
            Residual Plot                 obs RESIDUAL   ACTUAL    FITTED
==========================================================================
|     :              |  *          :     | 1983  6414.70  87635.0  81220.3
|    *:              |             :     | 1984 -48925.6  176145.  225071.
|     :              |           * :     | 1985  36869.9  233405.  196535.
|     :              |  *          :     | 1986  10742.6  148585.  137842.
|     :      *       |             :     | 1987 -21503.2  78401.0  99904.2
|     :              |         *   :     | 1988  32449.9  267384.  234934.
|  *  :              |             :     | 1989 -57107.6  168149.  225257.
|     :              |            *:     | 1990  41059.2  263962.  222903.
==========================================================================
```

J. Adjusted Rainfall Series and the "Right Sign" of the Independent Variable

Before describing how the annual updating occurs and how the regressions are used to make forecasts, we still need to go through some more examples to demonstrate other strategies for arriving at suitable regressions for predicting palay production. The first of these involves the sign on the independent variables in the regression, and the use of what we will call "adjusted" rainfall variables.

In the regression above, the independent variables were APRILO and MAYILO. When the regressions were run with the LS command, TSP generated a group of statistics, of which we have discussed R^2, the t-statistic and the residual plot. Now we will discuss the signs on the coefficients of the independent variables. Referring back to Table 8, we see that the coefficient for both of the independent variables is positive. Essentially this means that the regression line drawn through the scatter plot (refer back to Figure 2) will slope upward. If this regression line slopes upward, this is equivalent to saying that more rainfall in the growing season corresponds to a larger crop harvested. This is in line with what one would expect intuitively. It is true, however, that more rain can sometimes mean less crop harvested (e.g. a typhoon is accompanied by lots of rain, but this is not good for production). How can one take account of this in the regression analysis?

One way to do this is by forming what we will call an adjusted rainfall variable. This variable measures not absolute levels of rainfall, but deviations from normal rainfall. With this variable, one would expect that small deviations from normal are good for production, while large deviations, possibly representing either drought or typhoon, are bad for production. Thus, the expected coefficient of this variable would be negative. In other

words, small deviations from normal imply small values of the adjusted variable, which means relatively large production. Conversely, large deviations from normal imply large values of the adjusted variable, which corresponds to relatively small production. Thus, an inverse relationship between the adjusted variable and production exists, which implies the sign of the coefficient should be negative.

In order to name adjusted variables, we will simply place an A in front of the name of the absolute rainfall series. For example, AAUGDAV will represent Adjusted AUGust DAVao, where the corresponding absolute series is simply AUGDAV. While absolute rainfall data is simply gathered from PAGASA and entered into the TSP files (exactly how this is done will be described later), adjusted series need to be calculated before they can be used as independent variables in regressions. The formula for calculating an adjusted series in TSP is as follows:

SMPL 1975 1991

followed by:

GENR AAUGDAV = ABS(AUGDAV - @MEAN(AUGDAV))

This formula simply defines normal rainfall as the average rainfall at that station from 1975 to the present (calculated with @MEAN), and then measures deviations by taking the absolute value of the difference between normal rainfall at that station and the actual rainfall in any given year. The GENR stands for "generate," and it is used anytime one needs to generate a data series by means of a calculation. It is very important to note that adjusted rainfall series should be generated when the sample range is set at its maximum limits (in this case, 1975-1991). If this is not done, then the @MEAN calculation will only consider rainfall values in the current sample range when calculating average rainfall. Thus, if one is moving back and forth between different sample ranges, one will have to calculate different adjusted rainfall series for every different sample range. This will lead to a huge proliferation of the number of independent variables in the worksheet that will be difficult to name and keep track of. Thus, calculation of adjusted rainfall series should always be preceded by the appropriate SMPL command.

The following is an example of a regression using an adjusted series: SMPL 1980 1989, LS DECP C AAUGDAV. Although the sample range for the regression is 1980-1989,

the adjusted series AAUGDAV was generated over the sample range 1975-1990. The statistics for this regression are shown in Table 10. As can be seen there, the coefficient on AAUGDAV is negative as expected. This leads us to another rule of thumb, namely, that if an independent variable is to be included in a regression, it must be of the right sign. This is in addition to the requirement that it have a reasonably high t-statistic in absolute value.

The "right sign" for an independent variable in this model is positive for absolute rainfall series (an exception to this is discussed below) and negative for "adjusted" rainfall series. The reason why adjusted rainfall series should have negative coefficients was discussed above. The reason why absolute rainfall series should have positive coefficients is that it is generally true that more rainfall means more production. While this may not be true for very high levels of rainfall (e.g. typhoons), it is unlikely to be false over the whole range of rainfall values of the past 15 years. If a negative sign emerged for an absolute rainfall series, this would imply one of two things. One possibility is that more rain always means less production, even when the additional rain is not anywhere close to typhoon levels. The other is that the negative sign came about because of random chance. The second seems the more likely possibility, so we will stipulate that the "right sign" for an absolute rainfall series is positive.

Table 10

```
LS // Dependent Variable is DECP
Date: 11-14-1991 / Time: 15:23
SMPL range: 1980  -  1989
Number of observations: 10
==================================================================
     VARIABLE   COEFFICIENT    STD. ERROR      T-STAT.    2-TAIL SIG.
==================================================================
        C        199370.57     9658.2464     20.642523      0.000
     AAUGDAV    -527.82140     109.93807    -4.8010792      0.001
==================================================================
R-squared                0.742354    Mean of dependent var    159968.7
Adjusted R-squared       0.710148    S.D. of dependent var    29909.14
S.E. of regression       16102.45    Sum of squared resid     2.07E+09
Durbin-Watson stat       1.339204    F-statistic              23.05036
Log likelihood          -109.9409
==================================================================
```

```
==================================================================
          Residual Plot                obs RESIDUAL  ACTUAL   FITTED
==================================================================
|         :*        |        :        | 1980 -13572.3  183080.  196652.
|         :         |*       :        | 1981  1550.05  157824.  156274.
|         :         |    *   :        | 1982  10114.5  178370.  168255.
|         :        *        :        | 1983  167.236  162300.  162133.
|         :         |   *    :        | 1984  8104.01  135295.  127191.
|         :         |  *     :        | 1985  6606.27  133375.  126769.
|        *:         |        :        | 1986 -16482.5  108650.  125132.
|  *      :         |        :        | 1987 -28567.8  145441.  174009.
|         :         |    *   :        | 1988  9970.43  188624.  178654.
|         :         |        :   *    | 1989  22110.0  206728.  184618.
==================================================================
```

Variables that deviate from these expected signs are not behaving in accord with intuition, and if their t-statistic is relatively high in absolute value, it is likely due simply to random chance (as noted above). If the correlation between an independent variable and the dependent variable is due to random chance, the independent variable is unlikely to serve as a good predictor of the dependent variable in the future. Since the purpose of this model is to predict the future, <u>variables with the wrong sign should be discarded from all final regressions</u>.

Given these restrictions on the signs of independent variables, a hint may prove useful at this point. If an absolute rainfall variable has a high t-statistic, but has the wrong sign, it is worthwhile to try the corresponding adjusted rainfall series in place of the absolute series. Often, the adjusted series will also be significant and will retain the negative sign. Since a negative sign is the right sign for an adjusted series, however, this is OK. This may seem like we are just playing games, but there is an underlying rationale for this "hint." Sometimes, one or two large values of rainfall may correspond with low values of production, and these one or two points may be enough to make the slope of the whole regression line negative even if the regression line would have a positive slope if these two outlying points were excluded from the scatter diagram (in order to see this, it may help to visualize a scatter diagram like the one in Figure 2, but with one or two points at the lower right hand corner of the plot, these points corresponding to high rainfall and low production). Thus, using an adjusted series in place of an absolute series will properly take these one or two outlying points into account. Similarly, if an adjusted rainfall series has

the wrong sign (i.e. a positive one), but is highly significant, then it is worthwhile to try using the corresponding absolute series in its place.

As noted parenthetically above, there can be exceptions to the rule that absolute rainfall series must always have a positive sign. For example, let us look at the regression SMPL 1975 1989, LS ONP C AMJILO, which regresses October and November production in zone 4 on the total rainfall in Iloilo in April, May and June. The results are presented in Table 11, and we can see there that the coefficient on AMJILO is -398, which means that each additional millimeter of rain that falls in those three months means a loss of 398 metric tons of palay in October and November.

Table 11

```
LS // Dependent Variable is ONP
Date: 11-14-1991 / Time: 15:34
SMPL range: 1975  -  1989
Number of observations: 15
==================================================================
     VARIABLE   COEFFICIENT   STD. ERROR      T-STAT.   2-TAIL SIG.
==================================================================
        C        626226.99     51622.168     12.130970     0.000
     AMJILO     -398.00427     94.375828    -4.2172268     0.001
==================================================================
R-squared                0.577716    Mean of dependent var   423154.6
Adjusted R-squared       0.545233    S.D. of dependent var   106849.2
S.E. of regression       72055.25    Sum of squared resid    6.75E+10
Durbin-Watson stat       2.156154    F-statistic             17.78500
Log likelihood          -187.9886
==================================================================
```

```
==================================================================
            Residual Plot                obs RESIDUAL  ACTUAL   FITTED
==================================================================
|            :      |*    :            | 1975  6547.87  365913.  359365.
|            :      |  *  :            | 1976  21459.0  427510.  406051.
|            :   *  |     :            | 1977 -31017.1  495868.  526885.
|            :      |*    :            | 1978  16880.5  503806.  486926.
|            :     *|     :            | 1979  3163.91  489174.  486010.
|            :      |     :         *  | 1980  188276.  544895.  356619.
|            :*     |     :            | 1981 -58843.1  354770.  413613.
|            :      |    *:            | 1982  53174.8  434470.  381295.
|            :      |  *  :            | 1983  20426.8  569640.  549213.
|            : *    |     :            | 1984 -47659.5  305855.  353514.
|            :   *  |     :            | 1985 -25007.4  288030.  313037.
|            : *    |     :            | 1986 -51803.0  425650.  477453.
|            :      |   *:            | 1987  36386.9  579033.  542646.
|            :     *|     :            | 1988 -2884.33  318551.  321435.
|        *   :      |     :            | 1989 -129101.  244154.  373255.
==================================================================
```

Although this result may be surprising, it is much less so if one recalls the numbers presented in Table 8, where it can be seen that an additional millimeter of rain in April means an additional 452 metric tons of palay in September and an additional millimeter of rain in May means an additional 555 metric tons of palay in September. What is happening here is that if rains come early in April and May for the wet season crop in the Western Visayas, farmers plant early instead of late. This will increase production in September for two reasons. One reason is that more rain is a good thing, and production will be greater than if rainfall had been less. Another is that some farmers will shift their planting schedule forward so that they harvest in September instead of October or November, which implies that production in the latter two months will suffer. Thus, part of the increase in September production due to the increased rain in April and May is not really a net gain to the country, but simply a reallocation from one month to another.

Using the coefficients on the independent variables to quantify things, we see that an additional millimeter of rain increases production in September by 452 metric tons, but that 398 of this is simply a shift out of October and November, which means a net gain of 54 metric tons for each additional millimeter of rain in April. Similarly, an additional millimeter of rain in May implies a net gain of 157 metric tons.

In this case, a negative sign on an absolute rainfall series makes sense because of the shifting of production from one month to another. Thus, the "right sign" here is negative because one can tell a "good story" about why additional rainfall implies less production in October and November. This story could not be told if the regression for September production did not also have April and May rainfall as independent variables. In sum, one need not stick to the requirement that absolute rainfall series have a positive sign <u>if and only if</u> one can tell a convincing story about why this is so.

K. The Use of Time Trends

Up to this point, we have only used rainfall (either absolute or adjusted) series as independent variables. Occasionally, though, it also makes sense to use some others. A prime example of this is the use of time trends, which will be signified in TSP by TIME.

A time trend variable is a variable that increases by one fixed unit every year. For example, it equals 1 in 1975, 2 in 1976 and so on until 16 in 1990.

One example of where time trends can be helpful is if production is increasing over time due to continually increasing yields (which in turn is due to constantly improving technology). Technology is a difficult variable to measure, but since it usually improves more or less constantly over time, it is often proxied by a time trend. If a time trend is not used in a regression to crudely account for technology, and the trend of production is distinctly upward, it will be impossible to capture that upward trend using only rainfall variables (since rainfall does not tend to increase constantly over 15 year time periods). Thus, use of time trends can be an essential component of some regressions.

An example of a regression with a time trend is SMPL 1980 1990, LS JASP5 C APRZAM AFEBDIP TIME, the results of which are presented in Table 12 (JASP5 stands for July August September Production zone 5, i.e. third quarter production in zone 5). One can see that the TIME variable has a positive sign and a significant t-statistic. In addition, one should notice that the sample range has been changed since data from the 1970's did not fit very well, that all variables are highly significant (i.e. have t-statistics greater than 2), that all rainfall variables have the right sign, and that the R^2 is very high. Thus, this is likely to be a good regression for predicting future palay production.

Table 12

```
LS // Dependent Variable is JASP5
Date: 11-14-1991 / Time: 15:41
SMPL range: 1980  -  1990
Number of observations: 11
```

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	228611.77	37485.961	6.0985969	0.000
APRZAM	1491.0878	367.25555	4.0600823	0.005
AFEBDIP	-2365.1116	471.95005	-5.0113601	0.002
TIME	10689.715	3449.8734	3.0985820	0.017

R-squared	0.904627	Mean of dependent var	323122.9
Adjusted R-squared	0.863752	S.D. of dependent var	84148.64
S.E. of regression	31060.74	Sum of squared resid	6.75E+09
Durbin-Watson stat	3.203536	F-statistic	22.13191
Log likelihood	-126.9031		

```
==================================================================
            Residual Plot                 obs RESIDUAL  ACTUAL   FITTED
==================================================================
|       :           *         :       |  1980  351.011  268109.  267758.
|       :           |     *   :       |  1981  18508.0  371777.  353269.
|       *           |         :       |  1982 -29849.5  241400.  271250.
|       :           |    *    :       |  1983  16970.4  220455.  203485.
|       :  *        |         :       |  1984 -24908.3  210980.  235888.
|       :           |  *      :       |  1985  7997.49  333380.  325383.
|       :           |         :    *  |  1986  46356.7  332910.  286553.
|  *    :           |         :       |  1987 -45935.9  292516.  338452.
|       :           |  *      :       |  1988  10594.3  406923.  396329.
|       :       *   |         :       |  1989 -9424.21  461292.  470716.
|       :           |  *      :       |  1990  9339.97  414610.  405270.
==================================================================
```

Time trends need not always be positive, however. In other words, the "right sign" for a TIME variable can be either positive or negative. Negative time trends are not due to technology that is getting worse and worse over time, but can be due to changes in cropping patterns. For example, let us look at the regression SMPL 1978 1989, LS OCTP C TIME AJULZAM D1986, the results of which are shown in Table 13. This regression is trying to predict palay production in October in zone 5 (for the moment, ignore the D1986 variable - this will be discussed shortly). Notice that the time trend here has a negative sign. The time trend for the November production regression in zone 5 also has a negative sign, and the fact that both October and November production exhibit this trend likely means that an important structural change in rice production is occurring in zone 5.

The negative trend is most likely due to the effect of corn on rice cropping patterns in Mindanao (which is where zone 5 is located). Corn hectarage has been increasing in Mindanao for a long time now, and the main crop is harvested in the third quarter of the year. If a farmer harvests a corn crop in the third quarter of the year, then it will not be possible to also harvest a rice crop in the fourth quarter. Instead, rice planting will have to take place in the fourth quarter, and harvest will be delayed until the first quarter of the following year. If this theory is correct, one should also observe time trends in first quarter palay production in zone 5, but the sign should be positive. This is in fact the case, so the theory postulated here has some support. In sum, the right sign for a time trend can be either positive or negative.

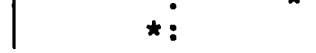
Table 13

```
LS // Dependent Variable is OCTP
Date: 11-14-1991 / Time: 15:44
SMPL range: 1978  -  1989
Number of observations: 12
```

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	428608.56	46863.972	9.1458010	0.000
TIME	-9337.3679	3380.3216	-2.7622721	0.025
AJULZAM	-1025.9027	455.20357	-2.2537228	0.054
D1986	153092.88	41539.153	3.6855080	0.006

R-squared	0.745401	Mean of dependent var	294040.6
Adjusted R-squared	0.649926	S.D. of dependent var	65018.10
S.E. of regression	38469.28	Sum of squared resid	1.18E+10
Durbin-Watson stat	1.629364	F-statistic	7.807317
Log likelihood	-141.2859		

```
Residual Plot
```



obs	RESIDUAL	ACTUAL	FITTED
1978	6511.17	306933.	300422.
1979	24888.2	377732.	352844.
1980	-39200.9	276183.	315384.
1981	-56915.0	231704.	288619.
1982	15653.4	266620.	250967.
1983	31665.6	319140.	287474.
1984	16560.1	289170.	272610.
1985	3217.04	312745.	309528.
1986	0.00000	426610.	426610.
1987	54513.5	287711.	233197.
1988	-16640.2	260990.	277630.
1989	-40252.9	172949.	213202.

L. Dummy Variables

Now let us return to the D1986 variable. The D stands for what is known as a dummy variable, and this variable takes on the value of 1 in 1986 and 0 in all other years. Such a variable can be useful if there is one year in the sample (in this case it is 1986) that has a very abnormally large or small production that can be traced to no readily apparent reason. To see this, let us run the regression SMPL 1978 1989, LS OCTP C TIME AJULZAM (this is the regression from Table 13 without D1986). The results are shown

in Table 14. Let us look at the residual plot first. Notice that the residual for 1986 is very large. As it turns out, almost every regression that tries to explain October production in zone 5 has the same very large positive residual. If one compares the regression statistics in Table 13 with those in Table 14, one can see that not using the dummy variable for 1986 forces the R^2 and the t-statistics for the other independent variables to decrease substantially. Thus, this one year alone is seriously impairing the statistical fit (and likely the predictive power as well) of the regression. Inclusion of a dummy variable solves the problem.

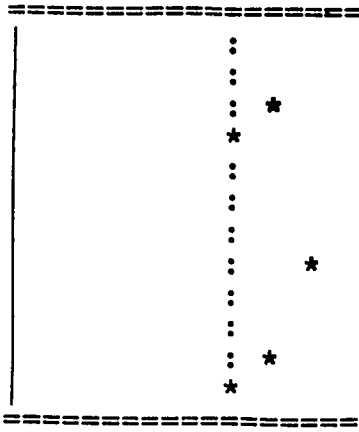
Table 14

LS // Dependent Variable is OCTP
Date: 11-14-1991 / Time: 15:45
SMPL range: 1978 - 1989
Number of observations: 12

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	432546.90	72553.857	5.9617355	0.000
TIME	-7072.8938	5147.5063	-1.3740428	0.203
AJULZAM	-1248.0406	698.71350	-1.7861979	0.108

R-squared	0.313124	Mean of dependent var	294040.6
Adjusted R-squared	0.160485	S.D. of dependent var	65018.10
S.E. of regression	59572.85	Sum of squared resid	3.19E+10
Durbin-Watson stat	1.580943	F-statistic	2.051405
Log likelihood	-147.2406		

Residual Plot



obs	RESIDUAL	ACTUAL	FITTED
1978	13183.9	306933.	293749.
1979	15923.7	377732.	361808.
1980	-44340.5	276183.	320523.
1981	-60545.5	231704.	292249.
1982	15889.4	266620.	250731.
1983	19710.4	319140.	299430.
1984	3537.12	289170.	285633.
1985	-22086.0	312745.	334831.
1986	131301.	426610.	295309.
1987	37165.6	287711.	250545.
1988	-47895.4	260990.	308885.
1989	-61843.7	172949.	234793.

Dummy variables should be used very sparingly and only when necessary because there is typically no good explanation for why they are needed in the regression. If one knew why the harvest in October of 1986 in Mindanao was so good, one could be a little more comfortable in employing the dummy variable. Short of such an explanation, however, one should be skeptical of using dummy variables.

If the abnormal year is early in the sample range, the sample range can simply be shortened to exclude that particular year, eliminating the need for a dummy variable. This in fact was done with the regression for November production in zone 5, where both 1976 and 1978 were very abnormal years. Instead of having a regression such as SMPL 1975 1989, LS NOVP C TIME JULGNS D1976 D1978, it was simply estimated as SMPL 1979 1989, LS NOVP C TIME JULGNS. Clearly, however, this technique will not work with October production since the abnormal year is so late in the sample. In other words, one should not estimate the regression SMPL 1987 1989, LS OCTP C TIME AJULZAM, because there are not enough observations to estimate anything (recall the restriction that there needs to be a minimum of 8 observations). In cases such as these, use of a dummy variable may be unavoidable.

The other important use of a dummy variable in the model is DUMMY, which takes on the value 1 for all years from 1987 on and zero for all years prior to 1987. This is used because there appears to have been a large shift in palay production out of April and into March beginning in 1987 and continuing to the present (in all zones except zone 5). Thus, DUMMY takes on a positive value when explaining first quarter production and a negative value when explaining second quarter production. For example, let us look at the regression LS APRP C TIME DUMMY for zone 2, whose results are shown in Table 15. Notice the negative sign on DUMMY, which indicates that there has been a shift of production out of April and into March beginning in 1987. At this point, I am unaware of any good reason for this shift, but there is no doubt that it is present in the data. Since the shift coincides with the first full year of the Aquino administration and occurs throughout the country, it is possible it is due to some change in statistical procedures, although I am far from sure of this. If this dummy variable proves not to fit future data, then it should be discarded.

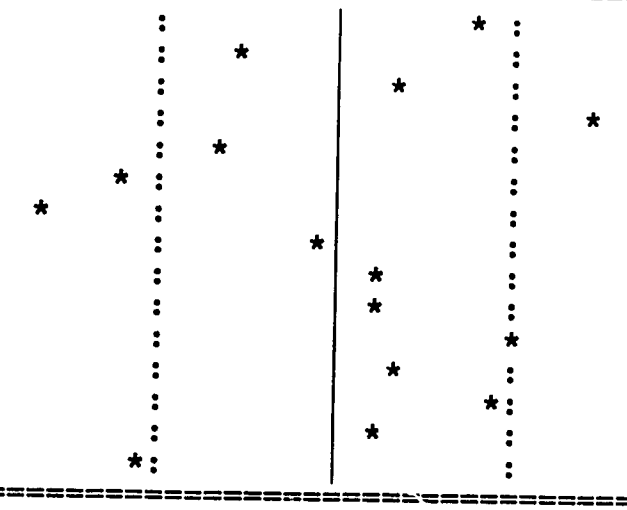
Table 15

LS // Dependent Variable is APRP
Date: 11-14-1991 / Time: 15:49
SMPL range: 1975 - 1989
Number of observations: 15

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	78507.291	12132.206	6.4709826	0.000
TIME	8920.3655	1645.9119	5.4197102	0.000
DUMMY	-67199.075	17777.880	-3.7799261	0.003

R-squared	0.709961	Mean of dependent var	136430.4
Adjusted R-squared	0.661622	S.D. of dependent var	34071.34
S.E. of regression	19819.40	Sum of squared resid	4.71E+09
Durbin-Watson stat	1.286572	F-statistic	14.68690
Log likelihood	-168.0268		

Residual Plot



obs	RESIDUAL	ACTUAL	FITTED
1975	15214.3	102642.	87427.7
1976	-10471.0	85877.0	96348.0
1977	7163.61	112432.	105268.
1978	28781.2	142970.	114189.
1979	-14690.1	108419.	123109.
1980	-25459.5	106570.	132029.
1981	-34518.8	106431.	140950.
1982	-3210.21	146660.	149870.
1983	3799.42	162590.	158791.
1984	5039.05	172750.	167711.
1985	20718.7	197350.	176631.
1986	7633.32	193185.	185552.
1987	18136.0	145409.	127273.
1988	5185.67	141379.	136193
1989	-23321.7	121792.	145114.

M. Creation of other Independent Variables

Sometimes the statistical fit of regressions can be improved by creating new production or rainfall variables. As far as production is concerned, the only way to do this is by adding together production in two or three months from the same quarter. For example, one could form the variable JASP2, which adds together production in zone 2 for

all months of the third quarter (the command would be GENR JASP2=JULP+AUGP+SEPP with the sample range set at its widest limits). One could also add just two months together, e.g. ONP (GENR ONP=OCTP+NOVP), which is the sum of October and November production. It would not be possible to add together March and April production to form a variable MAP, however, because those two months are in different quarters.

With respect to rainfall variables, there is a little more room for creativity. It is possible to add together rainfall from two or three months in the same quarter at one station, e.g. JFLEG, which is the sum of January and February rainfall in Legaspi City. It is also possible to add together rainfall for the same month at different stations in the same zone, e.g. MARAT, which is the sum of March rainfall in Aparri and Tuguegarao.

Another possibility is to take the square root of a rainfall series, e.g. SRJFLEG, which is the square root of the sum of rainfall in January and February in Legaspi City. This would be done in an attempt to capture the idea of diminishing "returns" to rainfall. In other words, additional rainfall always implies more production, but as the amount of rainfall increases, there is progressively less of an effect on palay production. This can be seen in Figures 1 and 2, where the data point for 1250 mm of rain lies below the regression line. This implies that production does not increase as rapidly with additional rainfall once high levels of rainfall are reached. Thus, the regression equation that estimates third quarter palay production in zone 3 (which is the variable on the vertical axis of Figures 1 and 2) uses SRJFLEG as the independent variable, not JFLEG (see Appendix C).

N. Use of (-1) for Rainfall Variables

If one examines the complete list of regressions used for forecasting in Appendix C, one will notice a fair number of them contain rainfall variables (both absolute and adjusted series) that have a (-1) attached to them (e.g. SMPL 1976 1989, LS JANP C DUMMY SEPDAV(-1)). This is necessary when estimating production for months early in the calendar year because January production will be influenced by September rainfall from the year before, not September rainfall of the same calendar year. For example, production in

January of 1985 will clearly depend on rainfall in September of 1984, not September of 1985. Thus, when estimating production for months early in the calendar year (usually the first two quarters), one needs to attach a (-1) to some of the rainfall variables so that in the regression equation production is postulated to be influenced by past rainfall, not future rainfall.

O. Regressions with only a Constant Term

Occasionally, it will be difficult to find any variables at all that can explain production in a given month. If this is the case, it is possible to run a regression with only a constant C. When this is done, the predicted value of future production is simply the average production over the sample range. Such a strategy should be avoided if at all possible, but in some cases it may be the best one can do. This is acceptable for forecasting production in months with very little production, for example January in zone 1 (see Appendix C). It is completely unacceptable however for months with large production, e.g. the November harvest in zone 1.

If one is forced to use a simple forecasting equation such as SMPL 1980 1989, LS JANP C, then one will notice that the R^2 is exactly equal to zero. In this one case, the low R^2 is nothing to worry about. An equation such as this is to be used only as a last resort, however, and in any event only for months with low production. If at all possible, try to find some variables that will help to explain production in that month.

P. Another Example

In the final section of this chapter, we will discuss another example that shows how to approach the problem of finding a suitable regression to explain production without simply running regressions at random. As noted above in section C of this chapter, this is an important skill to develop if the model is to be used properly. The above example of trying to explain September production in zone 4 hopefully helped to develop this skill a little bit, but there we were also busy with trying to explain quite a few other things, and the

process of how to critically examine TSP's output may have been a little obscured. Thus, this example will try to focus more on how to think about the data as opposed to explaining what the meaning of various regression statistics is or the various criteria by which one gauges whether or not a final regression is satisfactory. We will, however, introduce and make use of two new (but very simple) TSP commands in the process.

As our example, we will try to explain September production in zone 1. The first step is to LOAD A:\ZONE1. After this, perhaps the best thing to do when starting off is to make a graph of the data one is trying to explain in order to get a feel for it. This is done with the PLOT command, e.g. PLOT SEPP. If that command is typed in, TSP will prompt the user with a small table asking what kind of scaling is desired on the graph. For now, the best choice is A for automatic (if a description of the other scaling options is desired, it is best to consult the TSP users manual, a copy of which is at the DA). Once the A option is selected, the graph shown in Figure 3 appears on the screen. From this, one can see that, in general, production was much higher in the early part of the sample, specifically 1975-1981. It is possible that this was entirely caused by higher rainfall in that period, but it is not likely, and it would appear that production has gone through two "phases," one from 1975-1981, and the other from 1982-1990. Fortunately, the second phase is of sufficient length that we can do some analysis with it (remember a minimum of 8 observations is required).

Thus, let us reset the sample range from 1982-1990 with the SMPL command. From the graph, it should be obvious that over this interval, there is a distinct upward trend over time. This argues for the inclusion of a TIME variable, so let us run such a regression. This is done by LS SEPP C TIME (we have already set SMPL 1982 1990), and the results of that regression are shown in Table 16. The R^2 is equal to 0.41, which is good for a first step, but it most likely can be improved. To look for clues as to how it might be improved, a good thing to do is to look at the residual plot for this regression, which is also shown in Table 16. From this, one can notice that the large positive residuals are in 1982 and 1988, while the large negative residuals are in 1983 and 1987. Instead of just trying regressions at random using the LS command, let us examine some relevant rainfall data first.

Figure 3

The rainfall variables most likely to influence September production in zone 1 are those from Cabanatuan and Dagupan for the months of March, April and May. These weather stations are in the heart of Central Luzon and Pangasinan, which is where most of the palay in zone 1 is harvested. In addition, these months are 4,5 and 6 months before the September harvest, which is when rainfall is most likely to influence the crop. To examine these variables, we will make use of the SHOW command. In this case, it is SHOW

MARCAB MARDAG APRCAB APRDAG MAYCAB MAYDAG, and this produces Table 17 on the screen.

Table 16

```
LS // Dependent Variable is SEPP
Date: 11-14-1991 / Time: 15:54
SMPL range: 1982  -  1990
Number of observations: 9
==================================================================
     VARIABLE   COEFFICIENT   STD. ERROR      T-STAT.   2-TAIL SIG.
==================================================================
        C        3458.0000     18503.659     0.1868820     0.857
      TIME       3340.3333     1507.4712     2.2158522     0.062
==================================================================
R-squared                0.412259    Mean of dependent var   43542.00
Adjusted R-squared       0.328296    S.D. of dependent var   14247.40
S.E. of regression       11676.82    Sum of squared resid    9.54E+08
Durbin-Watson stat       2.571534    F-statistic             4.910001
Log likelihood          -95.92778
==================================================================

==================================================================
            Residual Plot              obs RESIDUAL  ACTUAL   FITTED
==================================================================
|         :          |         :     *  | 1982  18878.3  49059.0  30180.7
|      *  :          |         :        | 1983 -14986.0  18535.0  33521.0
|         :       *  |         :        | 1984 -1306.33  35555.0  36861.3
|         :     *    |         :        | 1985 -3261.67  36940.0  40201.7
|         :       *  |         :        | 1986 -1812.00  41730.0  43542.0
|         *          |         :        | 1987 -11518.3  35364.0  46882.3
|         :          |         :*       | 1988  12457.3  62680.0  50222.7
|         :     *    |         :        | 1989 -5090.00  48473.0  53563.0
|         :          |     *   :        | 1990  6638.67  63542.0  56903.3
==================================================================
```

Table 17

obs	MARCAB	MARDAG	APRCAB	APRDAG	MAYCAB	MAYDAG
1982	12.60000	1.100000	44.10000	163.4000	36.80000	178.4000
1983	1.000000	2.000000	2.500000	0.400000	67.50000	61.30000
1984	29.00000	20.80000	120.7000	91.40000	201.9000	137.7000
1985	0.000000	20.20000	66.80000	99.50000	15.00000	136.7000
1986	1.500000	13.00000	1.800000	0.000000	338.3000	203.0000
1987	0.000000	11.00000	9.800000	11.20000	79.60000	61.60000
1988	0.300000	0.000000	41.00000	70.20000	145.9000	284.6000
1989	80.90000	97.40000	7.400000	18.20000	164.7000	155.6000
1990	0.000000	0.000000	2.000000	0.000000	163.5000	183.1000

In this table, we are looking for a rainfall series that is at its peak (or close to it) for the years 1982 and 1988 (the years of large positive residuals), and at its trough for 1983 and 1987 (the years of large negative residuals). A series that is high when the residuals from Table 16 are high, and low when the residuals from Table 16 are low, will help to explain the variation in September production that is not explained by a time trend. From Table 17, we can see that MARCAB and MARDAG do not fit this requirement very well. For MARCAB, 1982 and 1988 are both low rainfall years, while the same is true of MARDAG. APRCAB shows some promise, however, as both 1982 and 1988 are relatively high. In addition, 1983 and 1987 are relatively low. At this point it should be noted that, in the context of regression analysis, a high or low observation is always relative to other observations for the same series, not relative to other series. For example, while the value of 44.1 for APRCAB is high for APRCAB, it is low for MAYCAB. The former comparison is what is relevant for our purposes, and the latter has literally no meaning whatsoever. This will always be true for whatever analysis we are doing in this model. Thus, APRCAB may be a good variable to try in our regression for September production.

Looking at the rest of the variables in Table 17, we also find that APRDAG and MAYDAG have high values in 1982 and 1988 and low values in 1983 and 1987. Thus, the next step is to run some trial regressions. If we run the regressions LS SEPP C TIME APRCAB, LS SEPP C TIME APRDAG and LS SEPP C TIME MAYDAG, we find that the values of R^2 are 0.45, 0.77 and 0.81 (Tables 18, 19 and 20). As far as Table 18 is concerned, an R^2 of 0.45 is not much of an improvement on 0.41, which was the value when TIME was the only independent variable (which is reflected in the low value of the t-statistic for APRCAB). The values of R^2 from Tables 19 and 20 are substantial improvements, however (again reflected in the high values of the t-statistic for APRDAG and MAYDAG). Since both of these latter two variables appear to work well, it is worthwhile to use both of them in the regression and see if they still have significant t-statistics. Unfortunately, however, it was stated earlier that for regressions with 8 or 9 observations one should use at a maximum two independent variables other than the constant (so that the number of independent variables does not approach too closely the number of observations - see section G for a discussion). Thus, the only way in which we

could utilize the data from both APRDAG and MAYDAG would be to construct a new variable that adds the rainfall from the two months together. This makes sense because the two months are adjacent on the calendar and because the data is for the same weather station, Dagupan. To construct this variable, we would use the command GENR AMDAG = APRDAG + MAYDAG, after which we can run the regression LS SEPP C TIME AMDAG, whose results are shown in Table 21. The R^2 is now 0.90, which is very high indeed. Also note that the t-statistics for both TIME and AMDAG are significant and that the coefficient on AMDAG is positive, which is the "right sign." Looking at the residual plot may at first cause some alarm, since the residual for the most recent observation (1990) appears to be large. As explained in section I of this chapter, however, what is important is not how large the residual is relative to other residuals, but how large it is in some absolute sense. A quick calculation shows that the residual for 1990 is less than 12% of the actual value for that year, which is very acceptable. Since all of these considerations point to a regression that is likely to be of use in forecasting, we will stop here and consider this to be our final regression.

In thinking about this section, one should remember not the variables used in the final regression, but rather the thought process that was used to arrive at the final regression. An understanding of this thought process and the ability to apply it for other months and other zones is of crucial importance if the model is to produce reliable forecasts.

Table 18

LS // Dependent Variable is SEPP
Date: 11-14-1991 / Time: 15:59
SMPL range: 1982 - 1990
Number of observations: 9

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	-4363.1137	23097.964	-0.1888960	0.856
TIME	3791.5421	1737.0729	2.1827190	0.072
APRCAB	73.149213	117.65329	0.6217354	0.557

R-squared	0.447832	Mean of dependent var	43542.00
Adjusted R-squared	0.263777	S.D. of dependent var	14247.40
S.E. of regression	12224.76	Sum of squared resid	8.97E+08
Durbin-Watson stat	2.206864	F-statistic	2.433134
Log likelihood	-95.64682		

```
==================================================================
            Residual Plot                 obs RESIDUAL  ACTUAL   FITTED
==================================================================
|         :           |         :      *  | 1982  19863.9  49059.0  29195.1
|         *           |         :         | 1983 -11408.6  18535.0  29943.6
|         :    *      |         :         | 1984 -6826.42  35555.0  42381.4
|         :     *     |         :         | 1985 -5290.22  36940.0  42230.2
|         :           *         :         | 1986  462.941  41730.0  41267.1
|         :*          |         :         | 1987 -10279.8  35364.0  45643.8
|         :           |        *:         | 1988  10962.4  62680.0  51717.6
|         :      *    |         :         | 1989 -4578.32  48473.0  53051.3
|         :           |     *   :         | 1990  7094.14  63542.0  56447.9
==================================================================
```

Table 19

```
LS // Dependent Variable is SEPP
Date: 11-14-1991 / Time: 16:05
SMPL range: 1982  -  1990
Number of observations: 9
==================================================================
    VARIABLE   COEFFICIENT   STD. ERROR      T-STAT.    2-TAIL SIG.
==================================================================
       C       -31484.539    16851.124    -1.8683940      0.111
    TIME        5504.7957    1234.2996     4.4598538      0.004
   APRDAG       177.68196    57.752219     3.0766257      0.022
==================================================================
R-squared                  0.771982   Mean of dependent var   43542.00
Adjusted R-squared         0.695975   S.D. of dependent var   14247.40
S.E. of regression         7855.791   Sum of squared resid    3.70E+08
Durbin-Watson stat         2.657981   F-statistic             10.15683
Log likelihood            -91.66691
==================================================================
```

```
==================================================================
            Residual Plot                 obs RESIDUAL  ACTUAL   FITTED
==================================================================
|       :             |            *:     | 1982  7471.94  49059.0  41587.1
|       :             |*            :     | 1983  405.305  18535.0  18129.7
|       :      *      |             :     | 1984 -4248.55  35555.0  39803.5
|  *    :             |             :     | 1985 -9807.57  36940.0  46747.6
|       :             |            *:     | 1986  7156.99  41730.0  34573.0
|       : *           |             :     | 1987 -6703.84  35364.0  42067.8
|       :             |     *       :     | 1988  4624.13  62680.0  58055.9
|       :  *          |             :     | 1989 -5848.21  48473.0  54321.2
|       :             |            *:     | 1990  6949.81  63542.0  56592.2
==================================================================
```

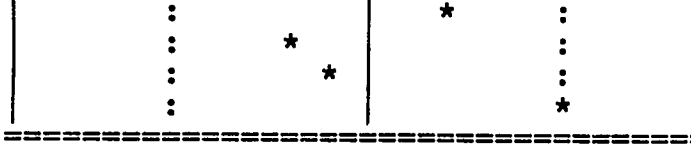
Table 20

LS // Dependent Variable is SEPP
Date: 11-14-1991 / Time: 16:06
SMPL range: 1982 - 1990
Number of observations: 9

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	-3699.5020	11410.083	-0.3242310	0.757
TIME	2139.5190	974.14463	2.1963053	0.070
MAYDAG	138.44898	38.421654	3.6034101	0.011

R-squared	0.814247	Mean of dependent var	43542.00
Adjusted R-squared	0.752329	S.D. of dependent var	14247.40
S.E. of regression	7090.443	Sum of squared resid	3.02E+08
Durbin-Watson stat	1.898570	F-statistic	13.15044
Log likelihood	-90.74438		

Residual Plot



obs	RESIDUAL	ACTUAL	FITTED
1982	10943.1	49059.0	38115.9
1983	-5508.09	18535.0	24043.1
1984	-1205.11	35555.0	36760.1
1985	-1821.18	36940.0	38761.2
1986	-8349.87	41730.0	50079.9
1987	2721.30	35364.0	32642.7
1988	-2976.34	62680.0	65656.3
1989	-1462.94	48473.0	49935.9
1990	7659.19	63542.0	55882.8

Table 21

LS // Dependent Variable is SEPP
Date: 11-14-1991 / Time: 16:09
SMPL range: 1982 - 1990
Number of observations: 9

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	-20951.205	9564.0493	-2.1906208	0.071
TIME	3685.1439	686.03152	5.3716830	0.002
AMDAG	98.283310	18.538784	5.3014971	0.002

R-squared	0.896603	Mean of dependent var	43542.00
Adjusted R-squared	0.862137	S.D. of dependent var	14247.40
S.E. of regression	5290.042	Sum of squared resid	1.68E+08
Durbin-Watson stat	1.287978	F-statistic	26.01435
Log likelihood	-88.10809		

```
Residual Plot
```

obs	RESIDUAL	ACTUAL	FITTED
1982	6935.82	49059.0	42123.2
1983	255.830	18535.0	18279.2
1984	-2861.94	35555.0	38416.9
1985	-5859.90	36940.0	42799.9
1986	-1492.03	41730.0	43222.0
1987	1253.31	35364.0	34110.7
1988	-2831.73	62680.0	65511.7
1989	-2934.59	48473.0	51407.6
1990	7535.23	63542.0	56006.8

II. IMPLEMENTATION OF THE MODEL

Chapter I discussed in some detail the techniques used for generating the regression equations to be used in making the quarterly forecasts of palay production. The entire set of these regression equations as of December 1991, for all months and all zones, is listed in Appendix C. The next task is to use those equations to make the actual forecast, and this will be the subject of the rest of the manual. The techniques of Chapter I will be an integral part of model implementation, however, a point which will become clear shortly. Thus, for anyone involved in making the forecasts, the material in Chapter I is not simply useful background, but rather essential material.

The reason why this material is essential is that the regression equations in Appendix C absolutely must be updated every time a forecast is made. For example, before making the next forecast for the fourth quarter, all equations in all zones for the fourth quarter must be checked to insure they are still satisfactory for making predictions, and if they are not, they must be updated so that they are satisfactory. Before embarking on a discussion of how to update the regression equations, however, we first need to discuss how to update the data.

A. Updating the Data

In order to update the regression equations so that they consider more years of data, and then to make predictions from current rainfall data, one needs to input the most recent production and rainfall data available. The production data necessary is palay production by calendar quarter by region (not necessarily by province), and this data comes from the Bureau of Agricultural Statistics (BAS). The production data by region must then be added according to the scheme laid out in Table 1 to obtain production data by agro-climatic zone, which is the form in which the data is used in the model. The rainfall data necessary is monthly rainfall at every one of the stations listed in Appendix B, and this data comes from PAGASA.

The procedure for entering data in TSP makes use of the DATA command. As an example, let us explain how to enter new production data for zone 1. The first step is to load the appropriate file, which is done with the LOAD command followed by the name of the file, e.g. LOAD A:\ZONE1 (assuming the file ZONE1 is on a diskette in the A drive). Once this file is loaded, the command to use is the DATA command, followed by the name(s) of the variables that need to be updated, e.g. DATA JANP FEBP MARP APRP MAYP JUNP (the maximum number of variables that TSP can read with a single DATA command is six. If more data than that needs to be entered, the DATA command must be used repeatedly). After entering this command, type N followed by the year of the data that one wishes to enter, e.g. N1991 (while typing this in, the command will start to write over data on the screen. This is nothing to be worried about. The command is only showing up on the screen, and is not actually writing over any data). This will move the cursor to the appropriate "cell," after which one can simply enter the number on the keyboard. After entering the number for January production (in metric tons), press ENTER and the cursor will move to the cell for February 1991, at which point one enters the February data, and so on. When the data for all six months has been entered, press X, followed by ENTER, and one will be returned to the regular TSP screen. If more data needs to be entered, one can use the DATA command again now.

A couple of points should be made at this juncture. One is that, while entering data, TSP displays a list of commands at the top of the screen (including the N and X commands discussed above) to help the user if a mistake is made in entering the data. A more important point is that the data needs to be entered in the proper units, which is metric tons for palay production and millimeters for rainfall. THESE UNITS ARE NOT NOTED ANYWHERE IN THE TSP FILE, SO THE USER MUST BE AWARE OF THE PROPER UNITS.

A similar procedure is followed for entering rainfall data, e.g. DATA JANLEG FEBLEG MARLEG APRLEG MAYLEG JUNLEG would be used for entering rainfall data for Legaspi City for the first six months of the year.

After entering all new production and rainfall data, other variables used in the regressions will also need to be updated, e.g. TIME, all dummy variables, adjusted rainfall

series and any other variables that were created using the GENR command (as in section M of Chapter I). For all TIME variables, one uses the command DATA TIME, after which the TIME variable is updated by entering a number for the most recent observation that is equal to one greater than the value for the preceding observation. For example, if TIME for 1990 is equal to 16, then TIME for 1991 should be 17. Dummy variables are updated in a similar manner. For the variable DUMMY, regardless of which file it is in (e.g. ZONE1, ZONE5), new values always take on the number 1, since DUMMY is equal to 1 for all years after and including 1987 (see Chapter I, section L). For other dummy variables, e.g. D1984 or D1981, new values will take on the number zero, since these dummies are only equal to 1 in the exact year specified (again see Chapter I, section L).

In the case of adjusted rainfall series (see Chapter I, section J), one will always have to generate the entire series over again from scratch instead of simply calculating one new value and entering it. This is because the presence of a new rainfall observation will change the average rainfall over the entire sample range. In other words, average rainfall from 1975-1990 will not be the same as average rainfall from 1975-1991. Since the average rainfall has now changed, the deviation from average will have also changed for all years in the sample. Thus, the GENR command will have to be used as in Chapter I, section J to create a new adjusted series (remember to set the sample range at its widest limits before doing this).

Finally, for any other variables that were created using the GENR command, such as JAP (July August Production) or JFLEG (January February Legaspi rainfall), the GENR command will have to be used once again to generate the most recent value for the series (e.g. GENR JAP=JULP+AUGP, GENR JFLEG=JANLEG+FEBLEG with the appropriate sample range).

Once all of this updating has been completed, be sure to save the new file so that the data is not lost. This is done with the SAVE command, followed by the name of the file, e.g. SAVE A:\ZONE1. Unfortunately, TSP does not prompt the user with the name of the file that you loaded (as most software does), so one must be very careful of which file name that you enter so as to not mistakenly overwrite another file. If you forget which file you loaded, inspection of the names of the rainfall variables will tell you what the current file

is. For example, presence of the variable MARAPA in the file means the name of the file is ZONE2, since Aparri is in zone 2. By using these clues, you can be assured of not making a mistake such as overwriting ZONE2 with ZONE1.

B. Updating the Regression Equations

Once the new data has been entered, the next step is to update the regression equations used for making predictions, a list of which is given in Appendix C. The regression equations must be revised every year, but the extent of revision required will depend on how well the original equations continue to track the data as new data becomes available. DO NOT EVER ASSUME THAT THE OLD REGRESSION EQUATION IS PROBABLY GOOD ENOUGH. IT MUST ALWAYS BE CHECKED AS NEW DATA BECOMES AVAILABLE.

At the very least, the sample range for the regression equations must be extended to incorporate new data. For example, when the model was updated in order to make a prediction for the fourth quarter of 1991, the regressions pertaining to the fourth quarter in Appendix C were at that time used data only up until 1989. Thus, the regressions were run once again using a new sample range that extended up until 1990, e.g. SMPL 1975 1990 or SMPL 1981 1990.

Using a new sample range will change all of the regression statistics, including the R^2, the signs of the coefficients on the independent variables, the t-statistics, and so on. If the new regression with the new sample range displays satisfactory statistics (as discussed in Chapter I), then this new regression (not the old one) can be used to make predictions for the new quarter (this process will be discussed in the next section), and there is no more updating to do. On the other hand, if the new regression suffers from a low R^2, wrong signs on the independent variables, low t-statistics, or an unacceptably large most recent residual, then a new, more satisfactory, regression equation involving new independent variables and/or a new sample range must be found. In order to do this, the user must apply the procedures discussed in Chapter I above.

Once all of the regression equations for the quarter for which we are trying to make predictions have been updated in this fashion, the user is ready to make the actual forecast. This is the topic of the next section.

C. Making the Forecast

Making a forecast of palay production using past rainfall data is done essentially by using the procedure shown in Figure 2 of Chapter I. A regression line is estimated, and using the coefficients on the independent variables (rainfall) generated by the regression, along with the most recent values for those independent variables (which were entered as data), an estimate of the dependent variable (palay production) is made.

In order to make this a bit clearer, let us do a specific example using a regression equation with several independent variables. The equation we will use is SMPL 1980 1990, LS JASP5 C APRZAM AFEBDIP TIME, whose regression statistics were presented in Table 12. In order to calculate a forecast for third quarter palay production in zone 5 for 1991, we need to know the 1991 values for all of the independent variables, after which we multiply these values by the coefficient on the respective independent variable:

1991 estimate of JASP5 =

$$228{,}612 + 1{,}491(69.9) - 2365(41.2) + 10{,}690(17)$$
$$= 417{,}080$$

Thus, 417,080 metric tons is the model's estimate of third quarter palay production in zone 5 in 1991.

TSP allows the user to automate this process so that it does not become too cumbersome when there are many different equations that need to be used to come up with the national forecast of palay production. The command that makes the forecast in TSP is the FIT command. In order to use the FIT command, one first runs the regression that has emerged from the update procedure described above in section B (with the appropriate sample range). The regression is being run once again, not so that the user can see the regression statistics once again (these should already be known since they were used to select the regression in the first place), but so that the regression can be stored in TSP's

memory. Immediately after running the regression, the sample should be changed by adding one year, the year for which the forecast is being made. For example, for the case discussed in the preceding paragraph, the user would type:

SMPL 1980 1990
LS JASP5 C APRZAM AFEBDIP TIME
SMPL 1980 1991

The first two statements run the regression and place it in memory. The third statement sets the sample range in preparation for making the forecast, which is done with the FIT command, e.g. FIT FJASP5. FJASP5 is the name of a new variable that has been created by TSP (the F in FJASP5 stands for Forecast). It is the value of "predicted" past palay production or predicted future production, depending on which year we are referring to. The predictions are made according to the regression equation using actual values of past rainfall. This procedure was described in the preceding paragraph for the most recent rainfall observation, but the mathematics can be carried out using rainfall data from any year whatsoever, and this is precisely what TSP does for all observations in the sample range.

Let us now use the SHOW command to examine what the FIT procedure has done. If one types SHOW JASP5 FJASP5, the numbers in Table 22 will be displayed on the screen. The JASP5 column shows actual values of third quarter production in zone 5, while the FJASP5 column shows predicted values of production, the predicted values being made with the regression. The value of FJASP5 for 1991 is the prediction for third quarter production in zone5, and this is the number that will go into the model forecast. Earlier values of FJASP5, in comparison with the corresponding values for JASP5, show how well the model has tracked production in the past. This is perhaps seen more easily by means of a graph, and if one types PLOT JASP5 FJASP5, the resulting graph is as shown in Figure 4. From Figure 4, one can see that the model tracks the data rather well.

Before turning to a discussion of batch files, which can automate the forecast generation process even further, it is necessary to discuss some problems that may arise in

the generation of the forecasts. The first of these potential problems is a lack of the rainfall data necessary to make a forecast. For example, with the above regression LS JASP5 C APRZAM AFEBDIP, if the rainfall data for April in Zamboanga is not available when the third quarter forecast is made in early July, it will simply not be possible to use this equation to forecast third quarter palay production in the current year, regardless of how well the equation may track the data in the past. In such a case, there are only two things to do. One is to stay in touch with PAGASA continuously so that the data can be accessed as soon as it is processed. The other is to abandon the regression, and try to find another one that will predict third quarter production without using variables for which no current data is available.

Table 22

obs	JASP5	FJASP5
1980	268109.0	267758.0
1981	371777.0	353269.0
1982	241400.0	271249.5
1983	220455.0	203484.6
1984	210986.0	235888.3
1985	333380.0	325382.5
1986	332910.0	286553.3
1987	292516.0	338451.8
1988	406923.0	396328.7
1989	461292.0	470716.2
1990	414610.0	405270.0
1991	NA	417079.7

The other potential problem arises because regressions do not tend to be as good at making predictions when the new values of the independent variables are very far removed from their past values. Referring back to Figure 2 in Chapter I, if the value for the independent variable (JFLEG) is less than about 50 mm or greater than about 1250 mm, then using it to predict palay production (JASP5) may result in some problems because extrapolation of regression equations outside the bounds of the data that was used to calculate them in the first place may result in poor forecasts. As a further example, let us

look at DECDAG and MARZAM, which are the values of rainfall in Dagupan in December and in March in Zamboanga. Using the SHOW command (SHOW DECDAG while in the file ZONE1, and SHOW MARZAM while in the file ZONE5) produces the results shown in Table 23. Notice that the value of rainfall in 1990 (52.4 mm), the most recent observation for DECDAG, is higher than any previous value recorded. Similarly, notice that 1991 was the first time in Zamboanga that it has not rained in March over the entire sample range. Using either of these series to forecast production in 1991 could pose problems, since the most recent values are outside the bounds of past experience. Using these variables to make production forecasts in 1991 is not a good idea.

Figure 4

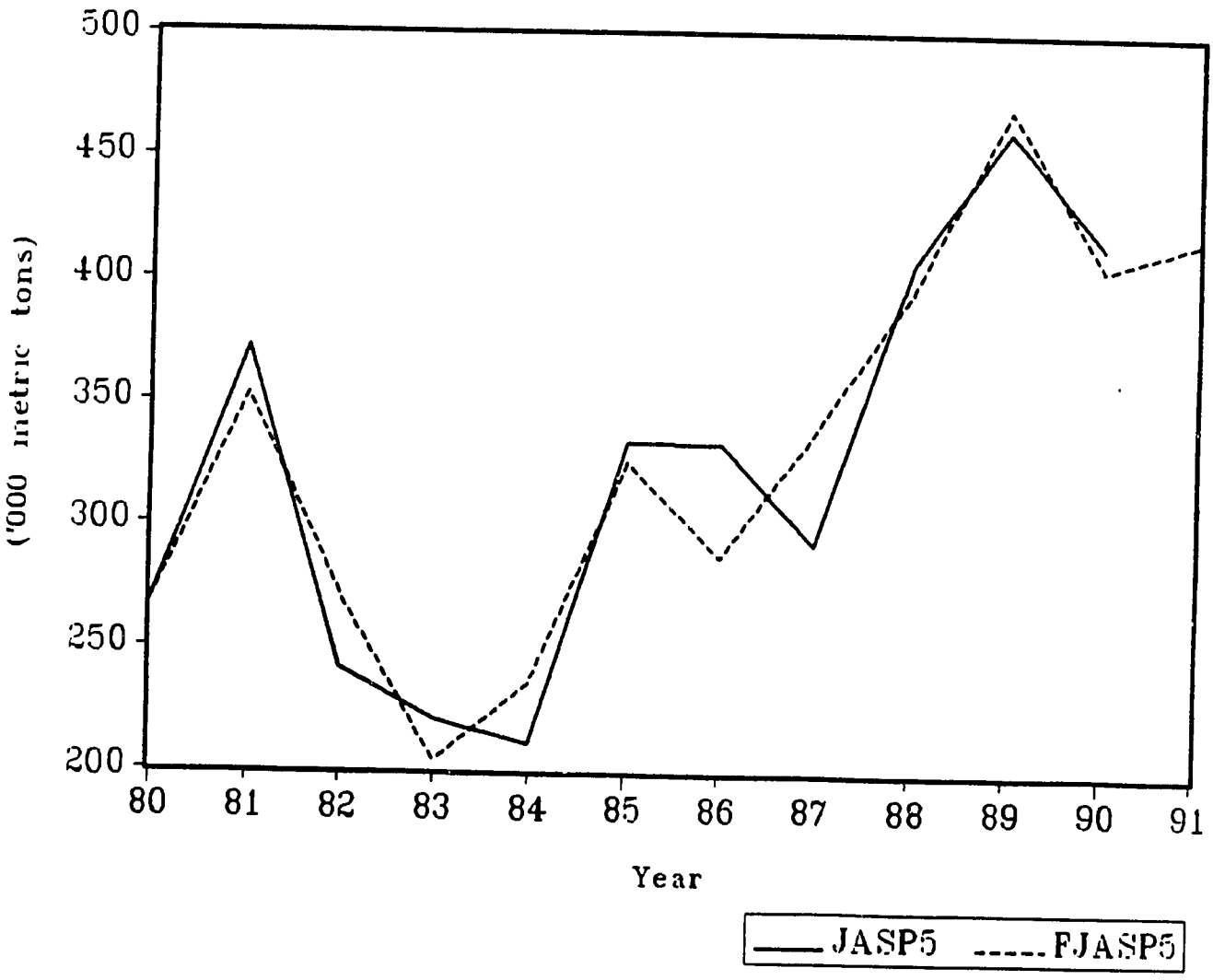


Table 23

obs	DECDAG
1975	10.70000
1976	0.000000
1977	0.000000
1978	1.000000
1979	3.100000
1980	4.200000
1981	1.400000
1982	45.90000
1983	7.200000
1984	1.000000
1985	0.800000
1986	16.30000
1987	30.30000
1988	0.000000
1989	0.600000
1990	52.40000
1991	NA

obs	MARZAM
1975	103.7000
1976	62.50000
1977	22.10000
1978	12.40000
1979	2.500000
1980	7.900000
1981	85.70000
1982	10.10000
1983	2.000000
1984	13.40000
1985	22.70000
1986	98.80000
1987	39.00000
1988	50.60000
1989	80.00000
1990	13.30000
1991	0.000000

The fact that they should not be used for making forecasts in 1991 does not disqualify them in the future, however. If, for example, June production in zone 1 was found to depend on rainfall in Dagupan the December before, and a regression that includes both December 1990 Dagupan rainfall (the high and unprecedented observation) and June 1991 zone 1 production comes out with good regression statistics, then it can be used to forecast June 1992 production in zone 1 (assuming that the December 1991 Dagupan rainfall observation is not another unprecedented high or low). A problem only arises if the <u>most recent</u> rainfall observation (i.e. the one being used to make a prediction about the future) is either an all-time sample high or low, and care should be taken to avoid using such observations to make predictions about the future.

D. Batch Files to Automate Forecast Generation

A batch file is a file that contains a list of commands that are executed automatically by the computer without a human operator having to sit by and press all the keys. The computer executes the commands much more rapidly than is possible for a human being, and this can avoid tedious, time consuming work in some instances. Appendix D contains

the current batch files QTR11991.BAT, QTR21991.BAT, QTR31991.BAT, and QTR41991.BAT, which are used for making forecasts for the respective quarters.

As one can see from looking at the content of these files, they consist of a list of commands that generate all the individual monthly or quarterly forecasts by zone (using SMPL, LS and FIT commands), and at the very end, sum up these individual zonal forecasts to come up with a national forecast for any given quarter. These batch files need to be edited every year since the regressions used for making predictions will always change. Even if the choice of independent variables does not change, the sample ranges will have to change, so the batch file will have to be updated every year. The editing can not be done in TSP, but is most easily accomplished in a word processing program such as Wordstar® or Word Perfect. If Wordstar is used, the file must be opened and saved as a nondocument, and if using Word Perfect, the file must be saved as a generic Word Perfect file (Ctrl F5, 3, 1). In order to make this a little bit clearer, Appendix D also contains an old version of QTR31991.BAT called QTR31990.BAT, which is the one that was used to make the third quarter 1990 forecast. The newer version is the result of the annual update procedure described in this chapter, and it was used to make the third quarter forecast in 1991. By examining these two files closely, one can see that many of the commands are identical, but many have also been changed.

The procedure for running a batch file is very simple. Execution is done while in TSP, and the command to use is simply RUN followed by the batch file name, e.g. RUN QTR31991.BAT. While the file is being executed, the sequence of commands will flash on the screen, indicating the work that is going on. The batch file will print out the national forecast, as well as the regression output for all the regressions used to make the forecast. In addition, it will also create a Lotus® 1-2-3® worksheet file that contains past values of quarterly and crop year production. This 1-2-3 file provides the data necessary to create the graph for the "Quarterly Production Outlook". The batch file will also create a new file in TSP which contains each of the zonal forecasts, along with past values of actual zonal production, and is a convenient way of storing all of these series in one place. In the case of RUN QTR31991, the new TSP file created will have the name FCST91_3, indicating that this file contains forecasts for the third quarter of 1991.

Once the batch file has been executed and all of the regression output has been printed, it is a good idea to check all of the regressions to make sure that they all have reasonable values of R^2. It is possible that when the editing of the batch file was done, a sample range was not changed that should have been. If this happens, the model will be using the wrong regression to make the forecast. Such a mistake is easily detected if the incorrect regression has a very low R^2, which most likely will be the case. Checking the batch file output will guard against this mistake. After checking all of the regression output, the new regressions should replace the old ones in Appendix C. Appendix D should also be updated by adding the newly created batch file.

E. Printing of the "Quarterly Palay Production Outlook"

The final "Outlook" report is the document that contains the model forecast for the given quarter and is given to the Secretary and other individuals both inside and outside the DA. In addition to the national forecast of palay production, which is just a single number, the outlook report should also contain several other things. First, there should be a brief description of the model to familiarize others with it. Second, there should also be a graph that shows actual past values of quarterly production for the given quarter, as well as a graph that shows how production in the current crop year compares to that in past crop years. Such a graph will provide some context to the solitary forecast number, which all by itself means little to most people. Finally, a calculation that compares the model forecast for the present quarter with actual production for the same quarter a year ago should also be included, along with a similar calculation for the current crop year. A sample outlook report that includes all of this information is presented in Appendix F, along with a description of the commands in Word Perfect that are necessary in order to create the report. The graph in the report is created in Lotus 1-2-3 or Borland Quattro-Pro®, and it is assumed that the user of the model is already familiar with that software. The data necessary to create the graph is stored in the Lotus 1-2-3 file created by the execution of the batch file.

We thus have all of the data required to produce the "Quarterly Production Outlook" except for an estimate of production for the current crop year. The technique for creating this estimate will depend on which quarter of the year is current, and the reason for this will become obvious in the process of explaining how to make the estimates.

Let us start with the fourth quarter, since this is the beginning of the October-September crop year. In order to provide an estimate for production for the current crop year in this case, we need to have estimates for the first, second and third quarters of the following calendar year, while the model will supply the estimate for the fourth quarter (the current quarter). The estimate for the first quarter of the following calendar year will be obtained directly from the BAS planter's intentions estimate. Estimates for the second and third quarters, however, are not available from either BAS or the regression model. Thus, it is suggested that estimates for these quarters be constructed by taking an average for quarterly production in these quarters over the past two or three years. The exact method used for constructing these estimates should not matter too much since they are low production quarters, which means that the crop year estimate will be dominated by the fourth and first quarter forecasts (the former from the model, the latter from BAS).

As a concrete example, when making a crop year 1991/92 estimate in the fourth quarter of 1991, one uses the model estimate for the fourth quarter of 1991, the BAS planter's intentions estimate for the first quarter of 1992, and a constructed average of past production for the second and third quarters of 1992.

The crop year estimate needed when the first quarter outlook report is issued is constructed in a similar fashion. Here, the value for fourth quarter production should come from the BAS preliminary estimate for that quarter, since that quarter has already passed. The first quarter estimate comes from the regression model itself, and the second quarter estimate used is the BAS planter's intentions estimate. Finally, the third quarter estimate can be constructed by taking an average of the past two or three actual values for the third quarter as described above.

Let us continue with the second quarter report. Here, the crop year estimate will be the sum of BAS preliminary estimates for the preceding fourth and first quarters, the model estimate for the second quarter, and the BAS planter's intentions estimate for the third

quarter. Finally, the crop year estimate in the third quarter report will be the sum of BAS preliminary estimates for the fourth. first and second quarters and the model estimate for the third quarter.

Thus, in constructing these crop year estimates, the estimate for the current quarter always comes from the model. After that, any BAS preliminary estimates that are available for past quarters are used. If necessary, BAS planter's intentions estimates are then used for quarters for which no other estimates are available. As a last resort, averages of past quarterly production are used for quarters for which there are no model or BAS estimates whatsoever.

F. A Timetable for Generating and Issuing the Outlook Reports

It is suggested that the outlook reports be issued in the third week of the first month of each quarter, or roughly at the same time as BAS issues its reports. As outlined in the preceding section, some of the information contained in the BAS reports will be necessary in order to construct the crop year estimates for the model report (e.g. preliminary production estimates for past quarters and planter's intentions estimates for future quarters). Remember that this information from BAS is not being inputted into the model, but rather is being used to provide context to the forecast generated by the model, which is derived solely from rainfall data. This information should be available internally to DA personnel before BAS releases their official report, which will improve the timeliness of the issuance of the model report. In other words, it should be possible to release the model report at more or less the same time that BAS officially releases their report despite the fact that some information from the BAS report needs to be incorporated into the model report.

In order to release the model report at this time, work must begin earlier in order to update all of the data and the regressions as described in earlier sections of this Chapter. This will depend on the workload of the personnel involved in making the forecast, as well as the number of individuals helping in the generation of the forecast. It is suggested that a different person be responsible for each of the five zones, after which one of them or the chief of the Policy Analysis Division can be the one to edit and run the batch file and

prepare the report. If this procedure is followed, a reasonable time to begin the process of collecting the data from PAGASA and updating the regressions is the first day of the first month of the given quarter. This should allow sufficient time to produce the forecast so that it is ready by the third week of the first month of each quarter.

APPENDIX A

List of Restrictions on the Choice of Independent Variables

1. A constant C must always be used as one of the independent variables (p.11, footnote 4).

2. In predicting production from any given zone, only rainfall data from that zone can be used (p.9).

3. Rainfall data 4,5 or 6 months in advance of production should be used to predict production. Data 2 or 3 months in advance may also be used if it is available in time to make the quarterly forecast (p.10).

4. The absolute maximum number of independent variables (in addition to the constant) that can be used in any one regression is 4, although 3 is a better limit to observe if at all possible. If the number of observations is small, however (e.g. 8 or 9), the maximum number that can be used is two (again, in addition to the constant) (p.18-19).

5. Any independent variable included in the final regression must have an acceptably high t-statistic. In most cases, this means at a minimum greater than 1.5 in absolute value (p.21).

6. The coefficient on any independent variable must be of the "right sign," which in most cases means positive for absolute rainfall series and negative for adjusted rainfall series (p.26-29). This rule can be violated if one can tell a convincing story about why the sign is different (see p.30-31 for an example of this).

Restrictions on the Observations Used in a Regression

1. The number of observations must be greater than or equal to 8 (p.19).

2. The observations must be consecutive and include the most recent observation available (p.19).

Other Considerations in Selecting a Final Regression

1. The R^2 statistic should be acceptably high, which in most cases means at a very minimum 0.40, although one should not be happy unless a value of at least 0.55 is achieved (p.12-13).

2. The size of the most recent residual should be less than 25% of the corresponding actual value, although exceptions to this can be made if everything else about the regression is very satisfactory (p.23-24).

APPENDIX B

List of Weather Stations and their
Abbreviations in the Model Files;
Conventions for Naming Other Variables

Zone 1

BAG - Baguio City
CAB - Cabanatuan (Muñoz)
DAG - Dagupan

Zone 2

APA - Aparri
TUG - Tuguegarao

Zone 3

DAE - Daet
LEG - Legaspi
MLB - Malaybalay
TAC - Tacloban

Zone 4

ILO - Iloilo
ROX - Roxas City

Zone 5

DAV - Davao
DIP - Dipolog
GNS - General Santos
ZAM - Zamboanga

Rainfall variables are named using the first three letters of the appropriate month followed by the three letters of the rainfall station, e.g. FEBDAE for rainfall in February in Daet. "Compound" rainfall variables which combine rainfall data across stations or months are named in a similar fashion, e.g. AMDAV for the sum of April and May rainfall in Davao or JANDD for the sum of rainfall in Davao and Dipolog in January. All adjusted

rainfall series are named simply by placing an A in front of the corresponding absolute series, e.g. AJANDD for adjusted JANDD. If one takes the square root of a rainfall series, the new variable is named by placing SR in front of the original variable, e.g. the square root of AUGROX would be named SRAUGROX.

Production variables are named in a similar fashion, the variable name being the first three letters of the month followed by a P, e.g. APRP for April production. If production is combined across months, the first letter of each month is used, e.g. ONP for the sum of October and November production. When naming a variable that represents production for an entire quarter, the number of the zone is placed after it, e.g. JASP5 for production in July, August and September in zone 5.

Dummy variables that take on the value 1 in only one year are named with the letter D followed by the year, e.g. D1984. The dummy variable that takes on the value 1 from 1987 on is arbitrarily called DUMMY.

APPENDIX C

List of Current Regression Equations Used for Forecasting

Zone 1

SMPL 1980 1989	LS JANP C
SMPL 1980 1989	LS FEBP C DUMMY
SMPL 1976 1989	LS MARP C DUMMY SEPDAG(-1)
SMPL 1976 1989	LS AMJP1 C TIME ADECCAB(-1) AOCTBAG(-1)
SMPL 1982 1990	LS JAP C MARCAB
SMPL 1982 1990	LS SEPP C TIME AMDAG
SMPL 1975 1990	LS ONDP1 C TIME JULCAB

Zone 2

SMPL 1976 1989	LS JANP C SEPTUG(-1)
SMPL 1976 1989	LS FEBP C TIME AOAPA(-1)
SMPL 1975 1989	LS MARP C DUMMY
SMPL 1975 1989	LS APRP C DUMMY TIME
SMPL 1976 1989	LS MAYP C DJFAT D1984
SMPL 1975 1989	LS JUNP C TIME
SMPL 1975 1990	LS JASP2 C AJANAPA MARAT MAYAPA
SMPL 1975 1990	LS ONDP2 C TIME JUNTUG

Zone 3

SMPL 1975 1989	LS JFMP3 C DUMMY TIME
SMPL 1976 1989	LS AMJP3 C DUMMY TIME DECTAC(-1)
SMPL 1983 1990	LS JASP3 C SRJFLEG
SMPL 1975 1990	LS ONDP3 C TIME JULLEG

Zone 4

SMPL 1976 1989	LS JANP C TIME SEPROX(-1)
SMPL 1976 1989	LS FEBP C TIME OCTROX(-1)
SMPL 1976 1989	LS MARP C DUMMY AOCTROX(-1) NOVROX(-1)
SMPL 1976 1989	LS APRP C DUMMY OCTROX(-1) ANOVILO(-1) ADECILO(-1)
SMPL 1976 1989	LS MJP C NOVILO(-1) AJANILO
SMPL 1978 1990	LS JAP C MAYROX
SMPL 1983 1990	LS SEPP C APRILO MAYILO
SMPL 1981 1990	LS ONP C AMJILO
SMPL 1977 [illegible]0	LS DECP C MAYILO

Zone 5

```
SMPL 1976 1989        LS JANP  C  DUMMY SEPDAV(-1)
SMPL 1976 1989        LS FEBP  C  TIME ASEPZAM(-1)
SMPL 1976 1989        LS MARP  C  TIME AAUGGNS(-1)
SMPL 1977 1989        LS APRP  C  TIME ADECGNS(-1)
SMPL 1975 1989        LS MAYP  C  TIME JANDD
SMPL 1976 1989        LS JUNP  C  TIME DECDAV(-1) JANDAV
SMPL 1980 1990        LS JASP5  C  TIME APRZAM AFEBDIP
SMPL 1980 1990        LS OCTP  C  D1986
SMPL 1979 1990        LS NOVP  C  TIME JULGNS
SMPL 1980 1990        LS DECP  C  AAUGDAV
```

ZONE1, First Quarter

```
LS // Dependent Variable is JANP
Date: 11-15-1991 / Time: 12:50
SMPL range: 1980  -  1989
Number of observations: 10
==================================================================
     VARIABLE   COEFFICIENT   STD. ERROR      T-STAT.   2-TAIL SIG.
==================================================================
        C        38553.600     4466.2607   . 8.6321876     0.000
==================================================================
R-squared                0.000000    Mean of dependent var   38553.60
Adjusted R-squared       0.000000    S.D. of dependent var   14123.56
S.E. of regression       14123.56    Sum of squared resid    1.80E+09
Durbin-Watson stat       1.568315    Log likelihood         -109.2186
==================================================================
```

```
LS // Dependent Variable is FEBP
Date: 11-15-1991 / Time: 12:51
SMPL range: 1980  -  1989
Number of observations: 10
==================================================================
     VARIABLE   COEFFICIENT   STD. ERROR      T-STAT.   2-TAIL SIG.
==================================================================
        C        48643.571     3335.3602     14.584204     0.000
      DUMMY      35544.095     6089.5066     5.8369417     0.000
==================================================================
R-squared                0.809840    Mean of dependent var   59306.80
Adjusted R-squared       0.786070    S.D. of dependent var   19079.03
S.E. of regression       8824.534    Sum of squared resid    6.23E+08
Durbin-Watson stat       2.572477    F-statistic             34.06989
Log likelihood          -103.9266
==================================================================
```

```
LS // Dependent Variable is MARP
Date: 11-15-1991 / Time: 12:52
SMPL range: 1976  -  1989
Number of observations: 14
==================================================================
     VARIABLE   COEFFICIENT   STD. ERROR      T-STAT.   2-TAIL SIG.
==================================================================
        C        128616.43     20667.645     6.2230811     0.000
      DUMMY      100839.33     19121.413     5.2736338     0.000
   SEPDAG(-1)    100.36512     61.151678     1.6412489     0.129
==================================================================
R-squared                0.771671    Mean of dependent var   182777.6
Adjusted R-squared       0.730156    S.D. of dependent var   54711.42
S.E. of regression       28420.67    Sum of squared resid    8.89E+09
Durbin-Watson stat       2.142301    F-statistic             18.58800
Log likelihood          -161.7452
==================================================================
```

ZONE1, Second Quarter

```
LS // Dependent Variable is AMJP1
Date: 11-15-1991 / Time: 12:56
SMPL range: 1976  -  1989
Number of observations: 14
```

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	491588.91	79170.570	6.2092380	0.000
TIME	13091.826	8037.2509	1.6288936	0.134
ADECCAB(-1)	-2050.1132	1057.2611	-1.9390793	0.081
AOCTBAG(-1)	-1097.4872	518.27028	-2.1175963	0.060

R-squared	0.544308	Mean of dependent var	454255.4
Adjusted R-squared	0.407601	S.D. of dependent var	126017.9
S.E. of regression	96992.79	Sum of squared resid	9.41E+10
Durbin-Watson stat	1.623621	F-statistic	3.981551
Log likelihood	-178.2633		

ZONE1, Third Quarter

LS // Dependent Variable is JAP
Date: 11-15-1991 / Time: 12:58
SMPL range: 1982 - 1990
Number of observations: 9

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	22593.162	3504.2127	6.4474288	0.000
MARCAB	340.39541	121.00361	2.8131012	0.026

R-squared	0.530628	Mean of dependent var	27332.22
Adjusted R-squared	0.463575	S.D. of dependent var	12585.98
S.E. of regression	9218.106	Sum of squared resid	5.95E+08
Durbin-Watson stat	0.916867	F-statistic	7.913538
Log likelihood	-93.79986		

LS // Dependent Variable is SEPP
Date: 11-15-1991 / Time: 13:00
SMPL range: 1982 - 1990
Number of observations: 9

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	-20951.205	9564.0493	-2.1906208	0.071
TIME	3685.1439	686.03152	5.3716830	0.002
AMDAG	98.283310	18.538784	5.3014971	0.002

R-squared	0.896603	Mean of dependent var	43542.00
Adjusted R-squared	0.862137	S.D. of dependent var	14247.40
S.E. of regression	5290.042	Sum of squared resid	1.68E+08
Durbin-Watson stat	1.287978	F-statistic	26.01435
Log likelihood	-88.10809		

ZONE1, Fourth Quarter

LS // Dependent Variable is ONDP1
Date: 11-18-1991 / Time: 17:05
SMPL range: 1975 - 1990
Number of observations: 16

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	881021.42	91798.246	9.5973666	0.000
JULCAB	489.94305	226.88348	2.1594479	0.050
TIME	44232.291	8025.6922	5.5113367	0.000

R-squared	0.782570	Mean of dependent var	1420089.
Adjusted R-squared	0.749119	S.D. of dependent var	281123.6
S.E. of regression	140809.3	Sum of squared resid	2.58E+11
Durbin-Watson stat	1.062466	F-statistic	23.39462
Log likelihood	-210.7245		

ZONE2, First Quarter

```
LS // Dependent Variable is JANP
Date: 11-18-1991 / Time: 17:16
SMPL range: 1976  -  1989
Number of observations: 14
==================================================================
     VARIABLE   COEFFICIENT   STD. ERROR      T-STAT.   2-TAIL SIG.
==================================================================
        C         22209.826     4107.5494     5.4070746     0.000
   SEPTUG(-1)     49.322871     18.672220     2.6415109     0.022
==================================================================
R-squared                0.367675    Mean of dependent var   31042.14
Adjusted R-squared       0.314981    S.D. of dependent var   10785.52
S.E. of regression       8926.730    Sum of squared resid    9.56E+08
Durbin-Watson stat       1.759384    F-statistic             6.977580
Log likelihood          -146.1414
==================================================================

LS // Dependent Variable is FEBP
Date: 11-18-1991 / Time: 17:26
SMPL range: 1976  -  1989
Number of observations: 14
==================================================================
     VARIABLE   COEFFICIENT   STD. ERROR      T-STAT.   2-TAIL SIG.
==================================================================
        C         62512.301     11972.781     5.2212012     0.000
    AOAPA(-1)     33.658826     16.980303     1.9822277     0.073
      TIME       -2834.1370     1074.0725    -2.6386832     0.023
==================================================================
R-squared                0.446698    Mean of dependent var   56423.36
Adjusted R-squared       0.346098    S D. of dependent var   19438.72
S.E. of regression       15718.97    Sum of squared resid    2.72E+09
Durbin-Watson stat       1.347926    F-statistic             4.440325
Log likelihood          -153.4537
==================================================================

LS // Dependent Variable is MARP
Date: 11-18-1991 / Time: 17:27
SMPL range: 1975  -  1989
Number of observations: 15
==================================================================
     VARIABLE   COEFFICIENT   STD. ERROR      T-STAT.   2-TAIL SIG.
==================================================================
        C        112742.11     21902.667     5.1474147     0.000
      DUMMY      121193.05     32094.989     3.7760740     0.003
      TIME       4383.3931     2971.4186     1.4751853     0.166
==================================================================
R-squared                0.795482    Mean of dependent var   172047.9
Adjusted R-squared       0.761395    S.D. of dependent var   73250.12
S.E. of regression       35780.62    Sum of squared resid    1.54E+10
Durbin-Watson stat       2.344563    F-statistic             23.33723
Log likelihood          -176.8879
==================================================================
```

ZONE2, Second quarter

LS // Dependent Variable is APRP
Date: 11-18-1991 / Time: 17:28
SMPL range: 1975 - 1989
Number of observations: 15

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	78507.291	12132.206	6.4709826	0.000
DUMMY	-67199.075	17777.880	-3.7799261	0.003
TIME	8920.3655	1645.9119	5.4197102	0.000

R-squared	0.709961	Mean of dependent var	136430.4
Adjusted R-squared	0.661622	S.D. of dependent var	34071.34
S.E. of regression	19819.40	Sum of squared resid	4.71E+09
Durbin-Watson stat	1.286572	F-statistic	14.68690
Log likelihood	-168.0268		

LS // Dependent Variable is MAYP
Date: 11-19-1991 / Time: 10:45
SMPL range: 1976 - 1989
Number of observations: 14

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	23673.492	6110.1152	3.8744755	0.003
DJFAT	22.978945	12.673073	1.8132102	0.097
D1984	26246.802	9598.6275	2.7344328	0.019

R-squared	0.484470	Mean of dependent var	35549.36
Adjusted R-squared	0.390737	S.D. of dependent var	11837.75
S.E. of regression	9239.992	Sum of squared resid	9.39E+08
Durbin-Watson stat	2.118221	F-statistic	5.168630
Log likelihood	-146.0152		

LS // Dependent Variable is JUNP
Date: 11-19-1991 / Time: 10:46
SMPL range: 1975 - 1989
Number of observations: 15

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	6719.9238	2243.1753	2.9957195	0.010
TIME	974.64286	246.71639	3.9504585	0.002

R-squared	0.545552	Mean of dependent var	14517.07
Adjusted R-squared	0.510594	S.D. of dependent var	5901.229
S.E. of regression	4128.355	Sum of squared resid	2.22E+08
Durbin-Watson stat	1.442104	F-statistic	15.60612
Log likelihood	-145.0953		

ZONE2, Third Quarter

LS // Dependent Variable is JASP2
Date: 11-18-1991 / Time: 17:12
SMPL range: 1975 - 1990
Number of observations: 16

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	102758.07	13496.146	7.6138822	0.000
AJANAPA	-206.41839	101.27083	-2.0382808	0.064
MAYAPA	88.710688	51.265023	1.7304330	0.109
MARAT	189.39637	94.294079	2.0085712	0.068

R-squared	0.519597	Mean of dependent var	113594.
Adjusted R-squared	0.399496	S.D. of dependent var	29516.7
S.E. of regression	22873.20	Sum of squared resid	6.28E+0
Durbin-Watson stat	2.520115	F-statistic	4.32634
Log likelihood	-181.0051		

ZONE2, Fourth Quarter

LS // Dependent Variable is ONDP2
Date: 11-19-1991 / Time: 10:49
SMPL range: 1975 - 1990
Number of observations: 16

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	109394.29	24130.990	4.5333528	0.001
TIME	15879.314	2461.3226	6.4515372	0.000
JUNTUG	260.90267	108.28333	2.4094444	0.032

R-squared	0.843954	Mean of dependent var	288439.8
Adjusted R-squared	0.819947	S.D. of dependent var	98454.43
S.E. of regression	41776.82	Sum of squared resid	2.27E+10
Durbin-Watson stat	1.280609	F-statistic	35.15442
Log likelihood	-191.2835		

ZONE3, First Quarter

LS // Dependent Variable is JFMP3
Date: 11-19-1991 / Time: 11:08
SMPL range: 1975 - 1989
Number of observations: 15

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	177552.92	23207.746	7.6505886	0.000
TIME	17789.948	3148.4717	5.6503440	0.000
DUMMY	237133.80	34007.381	6.9730098	0.000

R-squared	0.956076	Mean of dependent var	367299.3
Adjusted R-squared	0.948756	S.D. of dependent var	167479.4
S.E. of regression	37912.62	Sum of squared resid	1.72E+10
Durbin-Watson stat	1.752241	F-statistic	130.6008
Log likelihood	-177.7561		

ZONE3, Second Quarter

LS // Dependent Variable is AMJP3
Date: 11-19-1991 / Time: 11:40
SMPL range: 1976 - 1989
Number of observations: 14

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	490249.44	71865.670	6.8217473	0.000
TIME	16576.365	7661.2072	2.1636753	0.056
DUMMY	-277457.21	75951.065	-3.6531049	0.004
DECTAC(-1)	177.53381	110.32542	1.6091832	0.139

R-squared	0.650013	Mean of dependent var	635504.1
Adjusted R-squared	0.545017	S.D. of dependent var	120193.9
S.E. of regression	81073.70	Sum of squared resid	6.57E+10
Durbin-Watson stat	1.908054	F-statistic	6.190839
Log likelihood	-175.7534		

ZONE3, Third Quarter

LS // Dependent Variable is JASP3
Date: 11-19-1991 / Time: 11:43
SMPL range: 1983 - 1990
Number of observations: 8

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	34426.077	51799.922	0.6645971	0.531
SRJFLEG	9498.5316	2400.9055	3.9562289	0.007

R-squared	0.722886	Mean of dependent var	229774.4
Adjusted R-squared	0.676700	S.D. of dependent var	77878.56
S.E. of regression	44281.31	Sum of squared resid	1.18E+10
Durbin-Watson stat	1.649461	F-statistic	15.65175
Log likelihood	-95.78732		

ZONE3, Fourth Quarter

LS // Dependent Variable is ONDP3
Date: 11-19-1991 / Time: 11:44
SMPL range: 1975 - 1990
Number of observations: 16

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	833194.58	49031.672	16.992987	0.000
JULLEG	251.22314	141.30863	1.7778329	0.099
TIME	19940.214	3048.6199	6.5407347	0.000

R-squared	0.772156	Mean of dependent var	1067932.
Adjusted R-squared	0.737103	S.D. of dependent var	109085.6
S.E. of regression	55931.94	Sum of squared resid	4.07E+10
Durbin-Watson stat	1.879998	F-statistic	22.02832
Log likelihood	-195.9522		

ZONE4, First Quarter

LS // Dependent Variable is JANP
Date: 12-09-1991 / Time: 15:13
SMPL range: 1976 - 1989
Number of observations: 14

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT	2-TAIL SIG.
C	475.65060	12542.875	0.0379220	0.970
TIME	4781.6348	791.53031	6.0410003	0.000
SEPROX(-1)	76.664245	42.246303	1.8146971	0.097

R-squared	0.772245	Mean of dependent var	57676.29
Adjusted R-squared	0.730835	S.D. of dependent var	22738.10
S.E. of regression	11796.78	Sum of squared resid	1.53E+09
Durbin-Watson stat	2.316051	F-statistic	18.64875
Log likelihood	-149.4352		

LS // Dependent Variable is FEBP
Date: 12-09-1991 / Time: 15:14
SMPL range: 1976 - 1989
Number of observations: 14

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	22166.916	11996.372	1.8478016	0.092
TIME	8639.8536	1213.2341	7.1213408	0.000
OCTROX(-1)	76.592826	34.443457	2.2237264	0.048

R-squared	0.878311	Mean of dependent var	117097.1
Adjusted R-squared	0.856186	S.D. of dependent var	44583.47
S.E. of regression	16907.29	Sum of squared resid	3.14E+09
Durbin-Watson stat	2.363638	F-statistic	39.69736
Log likelihood	-154.4740		

LS // Dependent Variable is MARP
Date: 12-09-1991 / Time: 15:16
SMPL range: 1976 - 1989
Number of observations: 14

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	155610.23	15055.206	10.335974	0.000
NOVROX(-1)	40.880813	15.913234	2.5689820	0.028
AOCTROX(-1)	-245.40569	117.87216	-2.0819649	0.064
DUMMY	30804.195	22664.091	1.3591630	0.204

R-squared	0.489372	Mean of dependent var	149392.1
Adjusted R-squared	0.336183	S.D. of dependent var	37964.94
S.E. of regression	30931.92	Sum of squared resid	9.57E+09
Durbin-Watson stat	1.548211	F-statistic	3.194573
Log likelihood	-162.2634		

ZONE4, Second Quarter

LS // Dependent Variable is APRP
Date: 11-19-1991 / Time: 16:55
SMPL range: 1976 - 1989
Number of observations: 14

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	48646.558	6691.5701	7.2698272	0.000
DUMMY	-28128.807	4617.2510	-6.0921113	0.000
OCTROX(-1)	42.579747	17.971922	2.3692372	0.042
ANOVILO(-1)	-88.430383	37.836617	-2.3371641	0.044
ADECILO(-1)	-200.46650	94.360126	-2.1244831	0.063

R-squared	0.870071	Mean of dependent var	40591.07
Adjusted R-squared	0.812325	S.D. of dependent var	14332.08
S.E. of regression	6208.863	Sum of squared resid	3.47E+08
Durbin-Watson stat	1.915843	F-statistic	15.06721
Log likelihood	-139.0446		

LS // Dependent Variable is MJP
Date: 11-19-1991 / Time: 11:53
SMPL range: 1976 - 1989
Number of observations: 14

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	11119.954	5638.9193	1.9720008	0.074
NOVILO(-1)	81.340991	25.493273	3.1906844	0.009
AJANILO	-281.36550	160.63952	-1.7515335	0.108

R-squared	0.544226	Mean of dependent var	20613.50
Adjusted R-squared	0.461358	S.D. of dependent var	12229.23
S.E. of regression	8975.306	Sum of squared resid	8.86E+08
Durbin-Watson stat	1.981389	F-statistic	6.567377
Log likelihood	-145.6083		

ZONE4, Third Quarter

```
LS // Dependent Variable is JAP
Date: 11-19-1991 / Time: 11:55
SMPL range: 1978  -  1990
Number of observations: 13
==========================================================================
     VARIABLE   COEFFICIENT   STD. ERROR      T-STAT.   2-TAIL SIG.
==========================================================================
        C        13738.461     6373.3213     2.1556203     0.054
     MAYROX      266.10948     38.652128     6.8847304     0.000
==========================================================================
R-squared                0.811642    Mean of dependent var    47059.46
Adjusted R-squared       0.794519    S.D. of dependent var    32982.90
S.E. of regression       14951.16    Sum of squared resid     2.46E+09
Durbin-Watson stat       1.907671    F-statistic              47.39951
Log likelihood          -142.3234
==========================================================================
```

```
LS // Dependent Variable is SEPP
Date: 11-19-1991 / Time: 16:56
SMPL range: 1983  -  1990
Number of observations: 8
==========================================================================
     VARIABLE   COEFFICIENT   STD. ERROR      T-STAT.   2-TAIL SIG.
==========================================================================
        C        77665.544     30522.860     2.5445041     0.052
     APRILO      452.31329     192.85579     2.3453446     0.066
     MAYILO      554.88374     171.53980     3.2347230     0.023
==========================================================================
R-squared                0.744245    Mean of dependent var    177958.3
Adjusted R-squared       0.641944    S.D. of dependent var    73192.69
S.E. of regression       43796.91    Sum of squared resid     9.59E+09
Durbin-Watson stat       3.348162    F-statistic              7.274995
Log likelihood          -94.97004
==========================================================================
```

ZONE4, Fourth Quarter

LS // Dependent Variable is ONP
Date: 11-19-1991 / Time: 11:57
SMPL range: 1981 - 1990
Number of observations: 10

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	668930.02	50958.100	13.127060	0.000
AMJILO	-533.91187	83.466895	-6.3966902	0.000

R-squared	0.836460	Mean of dependent var	365999.1
Adjusted R-squared	0.816017	S.D. of dependent var	138711.7
S.E. of regression	59497.91	Sum of squared resid	2.83E+10
Durbin-Watson stat	2.060530	F-statistic	40.91765
Log likelihood	-123.0106		

LS // Dependent Variable is DECP
Date: 11-19-1991 / Time: 16:58
SMPL range: 1977 - 1990
Number of observations: 14

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	24494.782	4965.8462	4.9326501	0.000
MAYILO	282.50495	38.801983	7.2806833	0.000

R-squared	0.815408	Mean of dependent var	52206.50
Adjusted R-squared	0.800026	S.D. of dependent var	26686.43
S.E. of regression	11933.77	Sum of squared resid	1.71E+09
Durbin-Watson stat	2.692526	F-statistic	53.00835
Log likelihood	-150.2059		

ZONE5, First Quarter

LS // Dependent Variable is JANP
Date: 11-19-1991 / Time: 12:08
SMPL range: 1976 - 1989
Number of observations: 14

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	20690.913	1987.0731	10.412759	0.000
SEPDAV(-1)	47.950799	10.875152	4.4092074	0.001
DUMMY	59017.643	2017.0272	29.259717	0.000

R-squared	0.989922	Mean of dependent var	41824.50
Adjusted R-squared	0.988090	S.D. of dependent var	26764.42
S.E. of regression	2920.899	Sum of squared resid	93848141
Durbin-Watson stat	1.711477	F-statistic	540.2528
Log likelihood	-129.8921		

LS // Dependent Variable is FEBP
Date: 11-19-1991 / Time: 12:08
SMPL range: 1976 - 1989
Number of observations: 14

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	19910.563	15402.631	1.2926728	0.223
ASEPZAM(-1)	-398.82730	137.88700	-2.8924213	0.015
TIME	9224.7830	1332.9461	6.9205971	0.000

R-squared	0.836896	Mean of dependent var	72597.21
Adjusted R-squared	0.807240	S.D. of dependent var	45792.11
S.E. of regression	20104.76	Sum of squared resid	4.45E+09
Durbin-Watson stat	0.712347	F-statistic	28.22072
Log likelihood	-156.8900		

LS // Dependent Variable is MARP
Date: 11-19-1991 / Time: 12:09
SMPL range: 1976 - 1989
Number of observations: 14

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	57290 182	14348.774	3.9926883	0.002
AAUGGNS(-1)	-468.70990	183.82920	-2.5497033	0.027
TIME	12014.450	1347.1058	8.9187124	0.000

R-squared	0.886369	Mean of dependent var	142098.4
Adjusted R-squared	0.865709	S.D. of dependent var	55445.06
S.E. of regression	20318.25	Sum of squared resid	4.54E+09
Durbin-Watson stat	1.576415	F-statistic	42.90237
Log likelihood	-157.0469		

ZONE5, Second Quarter

LS // Dependent Variable is APRP
Date: 11-19-1991 / Time: 12:11
SMPL range: 1977 - 1989
Number of observations: 13

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	92846.085	28614.264	3.2447483	0.009
ADECGNS(-1)	-1403.9608	655.97617	-2.1402619	0.058
TIME	5981.5832	2504.1076	2.3887086	0.038

R-squared	0.511977	Mean of dependent var	115642.0
Adjusted R-squared	0.414372	S.D. of dependent var	44136.16
S.E. of regression	33775.79	Sum of squared resid	1.14E+10
Durbin-Watson stat	2.383977	F-statistic	5.245411
Log likelihood	-152.2983		

LS // Dependent Variable is MAYP
Date: 11-19-1991 / Time: 12:12
SMPL range: 1975 - 1989
Number of observations: 15

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	-11462.918	16466.302	-0.6961440	0.500
TIME	6074.3592	1179.9644	5.1479174	0.000
JANDD	158.14504	37.017439	4.2721768	0.001

R-squared	0.746610	Mean of dependent var	80909.67
Adjusted R-squared	0.704378	S.D. of dependent var	34925.89
S.E. of regression	18989.59	Sum of squared resid	4.33E+09
Durbin-Watson stat	1.678724	F-statistic	17.67891
Log likelihood	-167.3852		

LS // Dependent Variable is JUNP
Date: 11-19-1991 / Time: 16:59
SMPL range: 1976 - 1989
Number of observations: 14

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	-5235.0077	11982.477	-0.4393923	0.670
DECDAV(-1)	133.61868	52.565850	2.5419295	0.029
TIME	3369.8700	890.16602	3.7856646	0.004
JANDAV	113.87012	48.291286	2.3579849	0.040

R-squared	0.711759	Mean of dependent var	51269.00
Adjusted R-squared	0.625287	S.D. of dependent var	21356.57
S.E. of regression	13073.17	Sum of squared resid	1.71E+09
Durbin-Watson stat	1.394813	F-statistic	8.231083
Log likelihood	-150.2063		

ZONE5, Third Quarter

```
LS // Dependent Variable is JASP5
Date: 11-19-1991 / Time: 12:13
SMPL range: 1980 - 1990
Number of observations: 11
==================================================================
     VARIABLE   COEFFICIENT   STD. ERROR      T-STAT.    2-TAIL SIG.
==================================================================
        C        228611.77     37485.961     6.0985969     0.000
     APRZAM      1491.0878     367.25555     4.0600823     0.005
     AFEBDIP    -2365.1116     471.95005    -5.0113601     0.002
      TIME       10689.715     3449.8734     3.0985820     0.017
==================================================================
R-squared                0.904627    Mean of dependent var    323122.9
Adjusted R-squared       0.863752    S.D. of dependent var    84148.64
S.E. of regression       31060.74    Sum of squared resid     6.75E+09
Durbin-Watson stat       3.203536    F-statistic              22.13191
Log likelihood          -126.9031
==================================================================
```

ZONE5, Fourth Quarter

LS // Dependent Variable is OCTP
Date: 11-19-1991 / Time: 12:14
SMPL range: 1980 - 1990
Number of observations: 11

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	271843.30	13695.145	19.849610	0.000
D1986	154766.70	45421.659	3.4073326	0.008

R-squared	0.563317	Mean of dependent var	285913.0
Adjusted R-squared	0.514797	S.D. of dependent var	62173.42
S.E. of regression	43307.85	Sum of squared resid	1.69E+10
Durbin-Watson stat	2.029221	F-statistic	11.60992
Log likelihood	-131.9416		

LS // Dependent Variable is NOVP
Date: 11-19-1991 / Time: 12:15
SMPL range: 1979 - 1990
Number of observations: 12

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	268377.72	22708.552	11.818355	0.000
TIME	-5087.8259	1991.3989	-2.5549004	0.031
JULGNS	208.84393	61.747458	3.3822272	0.008

R-squared	0.655686	Mean of dependent var	237785.7
Adjusted R-squared	0.579172	S.D. of dependent var	36662.39
S.E. of regression	23783.36	Sum of squared resid	5.09E+09
Durbin-Watson stat	2.364759	F-statistic	8.569462
Log likelihood	-136.2221		

LS // Dependent Variable is DECP
Date: 11-19-1991 / Time: 12:16
SMPL range: 1980 - 1990
Number of observations: 11

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	185111.60	12244.689	15.117705	0.000
AAUGDAV	-398.32332	145.99277	-2.7283770	0.023

R-squared	0.452689	Mean of dependent var	157567.5
Adjusted R-squared	0.391877	S.D. of dependent var	29470.70
S.E. of regression	22981.91	Sum of squared resid	4.75E+09
Durbin-Watson stat	1.849906	F-statistic	7.444041
Log likelihood	-124.9717		

APPENDIX D

Batch Files Used to Generate the Forecasts

QTR11991.BAT

```
pon
load zone1
smpl 1975 1991
genr jfmp1=janp+febp+marp
genr amjp1=aprp+mayp+junp
genr jasp1=julp+augp+sepp
genr ondp1=octp+novp+decp
smpl 1976 1989
ls marp c dummy sepdag(-1)
smpl 1976 1991
fit fmar
smpl 1980 1989
ls febp c dummy
smpl 1980 1991
fit ffeb
smpl 1980 1989
ls janp c
smpl 1980 1991
fit fjan
genr fjfmp1=fjan + ffeb + fmar
store jfmp1 fjfmp1 amjp1 jasp1 ondp1

load zone2
smpl 1975 1991
genr jfmp2=janp+febp+marp
genr amjp2=aprp+mayp+junp
genr jasp2=julp+augp+sepp
genr ondp2=octp+novp+decp
genr aoapa=augapa+octapa
smpl 1976 1989
ls janp c septug(-1)
smpl 1976 1991
fit fjan
smpl 1975 1989
ls marp c dummy time
smpl 1975 1991
fit fmar
smpl 1976 1989
```

```
ls febp c aoapa(-1) time
smpl 1976 1991
fit ffeb
genr fjfmp2=fjan + ffeb + fmar
store jfmp2 fjfmp2 amjp2 jasp2 ondp2

load zone3
smpl 1975 1991
genr jfmp3=janp+febp+marp
genr amjp3=aprp+mayp+junp
genr jasp3=julp+augp+sepp
genr ondp3=octp+novp+decp
smpl 1975 1989
ls jfmp3 c time dummy
smpl 1975 1991
fit fjfmp3
store jfmp3 fjfmp3 amjp3 jasp3 ondp3

load zone4
smpl 1975 1991
genr jfmp4=janp+febp+marp
genr amjp4=aprp+mayp+junp
genr jasp4=julp+augp+sepp
genr ondp4=octp+novp+decp
genr aoctrox=abs(octrox-@mean(octrox))
smpl 1976 1989
ls janp c time seprox(-1)
smpl 1976 1991
fit fjan
smpl 1976 1989
ls febp c time octrox(-1)
smpl 1976 1991
fit ffeb
smpl 1976 1989
ls marp c novrox(-1) aoctrox(-1) dummy
smpl 1976 1991
fit fmar
genr fjfmp4=fjan + ffeb + fmar
store jfmp4 fjfmp4 amjp4 jasp4 ondp4

load zone5
smpl 1975 1991
genr jfmp5=janp+febp+marp
genr amjp5=aprp+mayp+junp
```

```
genr jasp5=julp+augp+sepp
genr ondp5=octp+novp+decp
genr asepzam=abs(sepzam-@mean(sepzam))
genr aauggns=abs(auggns-@mean(auggns))
smpl 1976 1989
ls janp c sepdav(-1) dummy
smpl 1976 1991
fit fjan
smpl 1976 1989
ls febp c asepzam(-1) time
smpl 1976 1991
fit ffeb
smpl 1976 1989
ls marp c aauggns(-1) time
smpl 1976 1991
fit fmar
genr fjfmp5=fjan + ffeb + fmar
store jfmp5 fjfmp5 amjp5 jasp5 ondp5

create a 1975 1991
fetch  jfmp1 jfmp2 jfmp3 jfmp4 jfmp5 fjfmp1 fjfmp2 fjfmp3 fjfmp4 fjfmp5 amjp1 amjp2
amjp3 amjp4 amjp5 jasp1 jasp2 jasp3 jasp4 jasp5 ondp1 ondp2 ondp3 ondp4 ondp5
genr jfmp= jfmp1 + jfmp2 + jfmp3 + jfmp4 + jfmp5
genr fjfmp =  fjfmp1 + fjfmp2 + fjfmp3 + fjfmp4 + fjfmp5
genr amjp=amjp1+amjp2+amjp3+amjp4+amjp5
genr jasp=jasp1+jasp2+jasp3+jasp4+jasp5
genr ondp=ondp1+ondp2+ondp3+ondp4+ondp5
smpl 1975 1990
genr cropyrp=ondp+jfmp(+1)+amjp(+1)+jasp(+1)
smpl 1975 1991
write(w) c:\zone\fcst91_1.wk1 jfmp fjfmp cropyrp
print jfmp fjfmp
save fcst91_1
```

```
pon
load zone1
smpl 1975 1991
genr amjp1=aprp+mayp+junp
genr jfmp1=janp+febp+marp
genr jasp1=julp+augp+sepp
genr ondp1=octp+novp+decp
genr adeccab=abs(deccab-@mean(deccab))
genr aoctbag=abs(octbag-@mean(octbag))
smpl 1976 1989
ls amjp1 c time adeccab(-1) aoctbag(-1)
smpl 1976 1991
fit famjp1
store amjp1 famjp1 jfmp1 jasp1 ondp1

load zone2
smpl 1975 1991
genr amjp2=aprp+mayp+junp
genr jfmp2=janp+febp+marp
genr jasp2=julp+augp+sepp
genr ondp2=octp+novp+decp
genr decat=decapa+dectug
genr janat=janapa+jantug
genr febat=febapa+febtug
genr djfat=decat(-1)+janat+febat
smpl 1975 1989
ls aprp c time dummy
smpl 1975 1991
fit fapr
smpl 1976 1989
ls mayp c djfat d1984
smpl 1976 1991
fit fmay
smpl 1975 1989
ls junp c time
smpl 1975 1991
fit fjun
genr famjp2=fapr + fmay + fjun
store amjp2 famjp2 jfmp2 jasp2 ondp2

load zone3
smpl 1975 1991
```

```
genr amjp3=aprp+mayp+junp
genr jfmp3=janp+febp+marp
genr jasp3=julp+augp+sepp
genr ondp3=octp+novp+decp
smpl 1976 1989
ls amjp3 c time dummy dectac(-1)
smpl 1976 1991
fit famjp3
store amjp3 famjp3 jfmp3 jasp3 ondp3

load zone4
smpl 1975 1991
genr amjp4=aprp+mayp+junp
genr jfmp4=janp+febp+marp
genr jasp4=julp+augp+sepp
genr ondp4=octp+novp+decp
genr mjp=mayp+junp
genr adecilo=abs(decilo-@mean(decilo))
genr anovilo=abs(novilo-@mean(novilo))
genr ajanilo=abs(janilo-@mean(janilo))
smpl 1976 1989
ls aprp c dummy octrox(-1) anovilo(-1) adecilo(-1)
smpl 1976 1991
fit fapr
smpl 1976 1989
ls mjp c novilo(-1) ajanilo
smpl 1976 1991
fit fmjp
genr famjp4=fapr+fmjp
store amjp4 famjp4 jfmp4 jasp4 ondp4

load zone5
smpl 1975 1991
genr amjp5=aprp+mayp+junp
genr jfmp5=janp+febp+marp
genr jasp5=julp+augp+sepp
genr ondp5=octp+novp+decp
genr jandd= jandip+jandav
genr adecgns=abs(decgns-@mean(decgns))
smpl 1975 1989
ls mayp c time jandd
smpl 1975 1991
fit fmay
smpl 1977 1989
```

```
ls aprp c adecgns(-1) time
smpl 1977 1991
fit fapr
smpl 1976 1989
ls junp c decdav(-1) time jandav
smpl 1976 1991
fit fjun
genr famjp5=fjun + fapr + fmay
store amjp5 famjp5 jfmp5 jasp5 ondp5

create a 1975 1991
fetch  amjp1 amjp2 amjp3 amjp4 amjp5 famjp1 famjp2 famjp3 famjp4 famjp5 jfmp1 jfmp2
jfmp3 jfmp4 jfmp5 jasp1 jasp2 jasp3 jasp4 jasp5 ondp1 ondp2 ondp3 ondp4 ondp5
genr amjp= amjp1 + amjp2 + amjp3 + amjp4 + amjp5
genr famjp =  famjp1 + famjp2 + famjp3 + famjp4 + famjp5
genr jfmp=jfmp1+jfmp2+jfmp3+jfmp4+jfmp5
genr jasp=jasp1+jasp2+jasp3+jasp4+jasp5
genr ondp=ondp1+ondp2+ondp3+ondp4+ondp5
smpl 1975 1990
genr cropyrp=ondp+jfmp(+1)+amjp(+1)+jasp(+1)
smpl 1975 1991
write(w) c:\zone\fcst91_2.wk1 amjp famjp cropyrp
print amjp famjp
save fcst91_2
```

```
pon
load zone1
smpl 1975 1991
genr jasp1=julp+augp+sepp
genr jfmp1=janp+febp+marp
genr amjp1=aprp+mayp+junp
genr ondp1=octp+novp+decp
genr jap=julp+augp
genr amdag=aprdag+maydag
smpl 1982 1990
ls jap c marcab
smpl 1982 1991
fit fjap
smpl 1982 1990
ls sepp c time amdag
smpl 1982 1991
fit fsep
genr fjasp1=fjap+fsep
store jasp1 fjasp1 jfmp1 amjp1 ondp1

load zone2
smpl 1975 1991
genr jasp2=julp+augp+sepp
genr jfmp2=janp+febp+marp
genr amjp2=aprp+mayp+junp
genr ondp2=octp+novp+decp
genr ajanapa=abs(janapa-@mean(janapa))
genr marat=marapa+martug
smpl 1975 1990
ls jasp2 c ajanapa mayapa marat
smpl 1975 1991
fit fjasp2
store jasp2 fjasp2 jfmp2 amjp2 ondp2

load zone3
smpl 1975 1991
genr jasp3=julp+augp+sepp
genr jfmp3=janp+febp+marp
genr amjp3=aprp+mayp+junp
genr ondp3=octp+novp+decp
genr jfleg=janleg+febleg
genr srjfleg=sqr(jfleg)
```

```
smpl 1983 1990
ls jasp3 c srjfleg
smpl 1983 1991
fit fjasp3
store jasp3 fjasp3 jfmp3 amjp3 ondp3

load zone4
smpl 1975 1991
genr jasp4=julp+augp+sepp
genr jfmp4=janp+febp+marp
genr amjp4=aprp+mayp+junp
genr ondp4=octp+novp+decp
genr jap=julp+augp
smpl 1978 1990
ls jap c mayrox
smpl 1978 1991
fit fja
smpl 1983 1990
ls sepp c aprilo mayilo
smpl 1983 1991
fit fsep
genr fjasp4=fja+fsep
store jasp4 fjasp4 jfmp4 amjp4 ondp4

load zone5
smpl 1975 1991
genr jasp5=julp+augp+sepp
genr jfmp5=janp+febp+marp
genr amjp5=aprp+mayp+junp
genr ondp5=octp+novp+decp
genr afebdip=abs(febdip-@mean(febdip))
smpl 1980 1990
ls jasp5 c aprzam afebdip time
smpl 1980 1991
fit fjasp5
store jasp5 fjasp5 jfmp5 amjp5 ondp5

create a 1975 1991
fetch jasp1 jasp2 jasp3 jasp4 jasp5 fjasp1 fjasp2 fjasp3 fjasp4 fjasp5 jfmp1 jfmp2 jfmp3 jfmp4
jfmp5 amjp1 amjp2 amjp3 amjp4 amjp5 ondp1 ondp2 ondp3 ondp4 ondp5
genr jasp=jasp1+jasp2+jasp3+jasp4+jasp5
genr fjasp=fjasp1+fjasp2+fjasp3+fjasp4+fjasp5
genr jfmp=jfmp1+jfmp2+jfmp3+jfmp4+jfmp5
genr amjp=amjp1+amjp2+amjp3+amjp4+amjp5
```

```
genr ondp=ondp1+ondp2+ondp3+ondp4+ondp5
smpl 1975 1990
genr cropyrp=ondp+jfmp(+1)+amjp(+1)+jasp(+1)
smpl 1975 1991
write(w) c:\zone\fcst91_3.wk1 jasp fjasp cropyrp
print jasp fjasp
save fcst91_3
```

```
pon
load zone1
smpl 1975 1990
genr jasp1=julp + augp + sepp
genr jfmp1=janp + febp + marp
genr amjp1=aprp + mayp + junp
genr ondp1=octp + novp + decp
genr djbag=decbag(-1) + janbag
smpl 1975 1989
ls jasp1 c djbag decdag(-1)
smpl 1975 1990
fit fjasp1
store jasp1 fjasp1 jfmp1 amjp1 ondp1

load zone2
smpl 1975 1990
genr jasp2 = julp + augp + sepp
genr jfmp2 = janp + febp + marp
genr amjp2 = aprp + mayp + junp
genr ondp2 = octp + novp + decp
genr marat = marapa + martug
smpl 1975 1989
ls sepp c time mayapa juntug
smpl 1975 1990
fit fsep
smpl 1975 1989
ls julp c
smpl 1975 1990
fit fjul
smpl 1981 1989
ls augp c marat
smpl 1981 1990
fit faug
genr fjasp2=fjul + faug + fsep
store jasp2 fjasp2 jfmp2 amjp2 ondp2

load zone3
smpl 1975 1990
genr jasp3=julp+augp+sepp
genr jfmp3=janp+febp+marp
genr amjp3=aprp+mayp+junp
```

```
genr ondp3=octp+novp+decp
smpl 1976 1989
ls jasp3 c febleg aprmlb
smpl 1976 1990
fit fjasp3
store jasp3 fjasp3 jfmp3 amjp3 ondp3

load zone4
smpl 1975 1990
genr jasp4=julp+augp+sepp
genr jfmp4=janp+febp+marp
genr amjp4=aprp+mayp+junp
genr ondp4=octp+novp+decp
genr afebilo =abs(febilo-@mean(febilo))
genr aaprrox =abs(aprrox-@mean(aprrox))
genr jap=julp + augp
smpl 1975 1989
ls jap c janrox afebilo mayrox
smpl 1975 1990
fit fja
smpl 1975 1989
ls sepp c aprilo junilo aaprrox
smpl 1975 1990
fit fsep
genr fjasp4=fja + fsep
store jasp4 fjasp4 jfmp4 amjp4 ondp4

load zone5
smpl 1975 1990
genr jasp5=julp + augp + sepp
genr jfmp5=janp + febp + marp
genr amjp5=aprp + mayp + junp
genr ondp5=octp + novp + decp
smpl 1975 1989
ls julp c mardip
smpl 1975 1990
fit fjul
smpl 1980 1989
ls augp c marzam aprdav
smpl 1980 1990
fit faug
smpl 1980 1989
ls sepp c marzam aprdav
smpl 1980 1990
```

```
fit fsep
genr fjasp5=fjul + faug + fsep
store jasp5 fjasp5 jfmp5 amjp5 ondp5

create a 1975 1990
fetch jasp1 jasp2 jasp3 jasp4 jasp5 fjasp1 fjasp2 fjasp3 fjasp4 fjasp5 jfmp1 jfmp2 jfmp3 jfmp4
jfmp5 amjp1 amjp2 amjp3 amjp4 amjp5 ondp1 ondp2 ondp3 ondp4 ondp5
genr jasp=jasp1+jasp2+jasp3+jasp4+jasp5
genr fjasp=fjasp1+fjasp2+fjasp3+fjasp4+fjasp5
genr jfmp=jfmp1+jfmp2+jfmp3+jfmp4+jfmp5
genr amjp=amjp1+amjp2+amjp3+amjp4+amjp5
genr ondp=ondp1+ondp2+ondp3+ondp4+ondp5
smpl 1975 1989
genr cropyrp=ondp+jfmp(+1)+amjp(+1)+jasp(+1)
smpl 1975 1990
write(w) c:\zone\fcst90_3.wk1 jasp fjasp cropyrp
print jasp fjasp
save fcst90_3
```

```
pon
load zone1
smpl 1975 1991
genr ondp1=octp+novp+decp
genr jfmp1=janp+febp+marp
genr amjp1=aprp+mayp+junp
genr jasp1=julp+augp+sepp
smpl 1975 1990
ls ondp1 c julcab time
smpl 1975 1991
fit fondp1
store ondp1 fondp1 jfmp1 amjp1 jasp1

load zone2
smpl 1975 1991
genr ondp2=octp+novp+decp
genr jfmp2=janp+febp+marp
genr amjp2=aprp+mayp+junp
genr jasp2=julp+augp+sepp
smpl 1975 1990
ls ondp2 c time juntug
smpl 1975 1991
fit fondp2
store ondp2 fondp2 jfmp2 amjp2 jasp2

load zone3
smpl 1975 1991
genr ondp3=octp+novp+decp
genr jfmp3=janp+febp+marp
genr amjp3=aprp+mayp+junp
genr jasp3=julp+augp+sepp
smpl 1975 1990
ls ondp3 c julleg time
smpl 1975 1991
fit fondp3
store ondp3 fondp3 jfmp3 amjp3 jasp3

load zone4
smpl 1975 1991
genr ondp4=octp+novp+decp
genr jfmp4=janp+febp+marp
```

```
genr amjp4=aprp+mayp+junp
genr jasp4=julp+augp+sepp
genr onp=octp+novp
genr amjilo=aprilo+mayilo+junilo
smpl 1981 1990
ls onp c amjilo
smpl 1981 1991
fit fon
smpl 1977 1990
ls decp c mayilo
smpl 1977 1991
fit fdec
genr fondp4=fon+fdec
store ondp4 fondp4 jfmp4 amjp4 jasp4


load zone5
smpl 1975 1991
genr ondp5=octp+novp+decp
genr jfmp5=janp+febp+marp
genr amjp5=aprp+mayp+junp
genr jasp5=julp+augp+sepp
genr aaugdav=abs(augdav-@mean(augdav))
smpl 1980 1990
ls octp c d1986
smpl 1980 1991
fit foct
smpl 1979 1990
ls novp c time julgns
smpl 1979 1991
fit fnov
smpl 1980 1990
ls decp c aaugdav
smpl 1980 1991
fit fdec
genr fondp5=foct+fnov+fdec
store ondp5 fondp5 jfmp5 amjp5 jasp5


create a 1975 1991
fetch ondp1 ondp2 ondp3 ondp4 ondp5 fondp1 fondp2 fondp3 fondp4 fondp5 jfmp1 jfmp2
jfmp3 jfmp4 jfmp5 amjp1 amjp2 amjp3 amjp4 amjp5 jasp1 jasp2 jasp3 jasp4 jasp5
genr ondp=ondp1+ondp2+ondp3+ondp4+ondp5
genr fondp=fondp1+fondp2+fondp3+fondp4+fondp5
```

```
genr jfmp=jfmp1+jfmp2+jfmp3+jfmp4+jfmp5
genr amjp=amjp1+amjp2+amjp3+amjp4+amjp5
genr jasp=jasp1+jasp2+jasp3+jasp4+jasp5
smpl 1975 1990
genr cropyrp=ondp+jfmp(+1)+amjp(+1)+jasp(+1)
smpl 1975 1991
write(w) c:\zone\fcst91_4.wk1 ondp fondp cropyrp
print ondp fondp
save fcst91_4
```

APPENDIX E

A Summary of the Steps Required to Update the Model Every Quarter

1.Collect new rainfall data from PAGASA for all of the stations listed in Appendix B and the most recent production data from BAS for all regions. Aggregate the regional production data into zones according to Table 1 in the text and then enter all of this data into the TSP files Zone1, Zone2, Zone3, Zone4, and Zone5 as described in Chapter II, section A.

2.Update all of the regression equations for all zones for the quarter for which the forecast is being made. This procedure is described in detail in Chapter I. Most likely, a different individual will be responsible for updating the regression equations in each zone. Before doing the updating, however, all variables created using the GENR command must also be updated as described in section A of Chapter II.

3.After updating all regression equations, appropriate corrections must be made to the batch file. The batch file can then be executed as described in section D of Chapter II.

4.Finally, once the forecast has been made with the batch file, the "Quarterly Palay Production Outlook" can be printed and sent to the Undersecretary.

APPENDIX F

By Rayna Lyn Roxas

1. You will have to download **FONTS** first before changing the sub-directory to **WP51**. At the **C:>** type **CD\FONTS, ENTER** then at the **CD\FONTS>** type **HPDF**. You will be asked what filename to download so you type in **PORTDOC** and hit **ENTER** twice. Go to the **WP51** program by typing **CD\WP51** at the **C:FONTS>**. When the prompt is already **C:WP51>** type in **WP** and hit **ENTER**.

2. You will be "Welcomed" into the **WP** version 5.1 program after which you will get a blank blue colored screen with this message at the bottom right corner of the screen:

Doc 1 Pg 1 Ln 1" Pos 1"

3. To make the **Quarterly Palay Production Outlook** Memo, it is necessary that you familiarize yourself with the following commands and function keys that will often be used:

Function key	Feature
F1	**Cancel Command**
F3	**Help**
Shift-F10	**Retrieve a file**
F7	**Exit WP51(with option to save document before exiting)**
F10	**Save document**
Shift-F6	**Center a line**
F6	**Bold**
F8	**Underline**
Alt-F3, F11	**Reveal Codes**
Shift-F7	**Print Menu/Options**
Shift-F7,6	**View Document**
Shift-F7,1	**Print full document**
Shift-F7,2	**Print page**
Shift-F7,4	**Control Printer (to cancel print jobs)**
Ctrl-F8,4	**Base Font Feature (To change the font sizes of the text, just choose the desired size and hit ENTER. The font size currently in use will be preceded by an * .)**
Alt-F4	**Highlight a block**
Ctrl-F4	**Options to Move, Copy, Delete, Append the block**
Home, Home, ↑	**Move back to top of document**
Home, Home, ↓	**Move down to end of document**
Ctrl-Home	**Go to (page no.)**
Insert	**Typeover**

4. Font sizes that you will likely use are:

TmsRmn 18pt Bold (AD) for the 1st **Title Line**(e.g.QUARTERLYPALAY...)
TmsRmn 14pt Bold (AD) for the 2nd **Title Line** (e.g. Historical Palay ...)
TmsRmn 12pt Bold (AD) for the **Text** and **data explaining the graph** composed of 4 lines:

a) Production Estimate, Crop Year 90/91 **TAB 5X** to **Pos 6"** and type in the **number of metric tons**;

b) % Change from Crop Year 89/90 **TAB 5X** to **Pos 6"** and type in the **%change**. Please use the **Space Bar** to **center** the number between the **number of metric tons**. In order to make sure if you have already centered this item, **view** the document and **move** the **space bar** again, if necessary. Take note of the position **(Pos)** of the **% change** because you will be referring back to this for the 4th line.

To illustrate:

Production Estimate, Crop Year 90/91 9,338,460 metric tons
% Change from Crop Year 89/90 -18.9%
The number **9,338,460 metric tons** started at **Pos 6"** and the **-18.9%** started at **Pos 6.5"**;

c) Model Forecast, _____ Quarter 1991 (same as in (a));

d) % Change from _____ Quarter 1991 (same as in (b)).

TmsRmn 10pt Bold (AD) for the **Legend** descriptions

5. Lotus/Quattro graphs are not imported into the **WP** document because they become condensed (too small) for our purposes. This is the reason why the **text** is done in **WP** and the graph in **Lotus/Quattro**, therefore **running** the document **twice** through the printer. Thus, when making the **Text** in **WP** you have to allow a **wide blank space** after the 2nd **Title Line** by hitting **ENTER** a couple of times (e.g. **10X**) in order to fit the **Lotus/Quattro graph** later on. The **Legend** is done in **Quattro-Pro** and you have to edit the size of the font by getting into **Graph, Text, Font, Legend, Point Size** then choose **12 Pt** so that you can fit most of the words like: **Prices** and **Forecast**, which are found in the legend.

6. Before you start typing the 1st word in the data explaining the graph (4 lines), you have to change the Font size to **TmsRmn 12pt Bold (AD)** by **Ctrl-F8,4**. Please refer to the instructions in **#4** for aligning the numbers in this portion.

*7. If you are limited to using two pages only to make this Memo, try to fit all the graph data in the front page and put the text in the second page to make it easier for you to estimate how much space to leave blank for the graph.

8. The **Text** of the Memo is done in **TmsRmn 12pt Bold (AD)**. After typing in the text, hit **ENTER** about **3X** before you make a **Table** containing information on **Who** made the Production Outlook and **Who** to contact for further information.

9. a) The **Table** can be generated by **Alt-F7,2,1** after which you will be asked **Number of Columns** (a number will already be there e.g. 3, but we will only be needing **1 column**) to which you will type **1,ENTER** and followed by another **Number of Rows** to which you will type **1,ENTER**. You will get this message:

_ The inside of this box will be colored red with the cursor blinking at the left corner of the box.

Table Edit: Press **Exit** when done **Cell A1 Doc 1 Pg 2 Ln 1.72" Pos 1.12"**

Ctrl-Arrows Column Widths; **Ins** Insert; **Del** Delete; **Move** Move/Copy;
1 Size; **2** Format; **3** Lines; **4** Header; **5** Math; **6** Options; **7** Join; **8** Split: **0**

b) To get rid of the double border, Choose **#3** Lines then you will get another set of choices:

Top = Double; Left = Double; Bottom = Double; Right = Double
Lines: **1** left; **2** Right; **3** Top; **4** Bottom; **5** Inside; **6** Outside; **7** All; **8** Shade: **0**

c) Choose **#7** All and you will get another set of options:

Top = Double; Left = Double; Bottom = Double; Right = Double
1 None; **2** Single; **3** Double; **4** Dashed; **5** Dotted; **6** Thick; **7** Extra Thick: **0**

d) Choose **#2** Single and you will see that the table will have a **single border.**

e) You can then press **F7** to exit from the **Table Edit** mode. Go into the box using the **arrow key**, change the Font size to **TmsRmn 10pt Bold (AD)** and on **Pos 1.19"** type in the necessary information. You can use the **Space Bar** to align the ends of the text inside the box, when you automatically move down to the 2nd Line move the **Space Bar** again to **Pos 1.19"** before starting the second line of the text, and the same for the 3rd Line. When you see that you are going past the right border, don't mind it and don't hit **ENTER**. It will automatically reposition itself inside the box.

10. After all this, you can **Save** the file under **OUTLOOK**. It would also be advisable to **save** and **re-save** the file as you go along so that you don't have to re-do the whole thing over and over again because you made a slight mistake. If ever there is some information or data or command you don't need or shouldn't have made, you can use **Reveal Codes (Alt-F3 or F11)** and delete unnecessary information from there before **Exiting** without saving. To get out of the **Reveal Codes** mode, you just press **Alt-F3 or F11** again to restore screen.

11. Now, you are ready to print out your Memo (without the graph yet) in **WP** (this is after you have **edited** and **viewed** your document over and over). To print **Shift-F7,1** and the printer will process your command.

12. After **WP51** you will now go into the **Lotus Printgraph** or **Quattro Pro GraphPrint** (wherever your graph is located). In **Lotus**, after you have chosen what .pic file to print, you go into **Settings**, **Image**, **Size** and choose **Half**. You will see that the numbers in the lower portion under **Settings** will change. After choosing **Half** you again choose **Manual** and change the **Top Margin** and the **Height** according to the space you have allotted in the **WP** document. It is advisable to print out samples and try to fit it in the blank space of your document first, before you actually run it on your finished **WP** print out of the Text until you get the exact **Top Margin** and **Height** measurement. Make sure to choose **Align** before you hit **Go**.

13. If you are using **Quattro Pro GraphPrint**, you will need to go into the **Layout** portion and enter the necessary measurements **Top Edge**, **Left Edge**, **Height**, **Width** before you hit **Name** (name of graph to print) then **Go**. Also run a couple of samples before you actually run on the finished **WP** print out of your Text.

QUARTERLY PALAY PRODUCTION OUTLOOK

Historical Palay Production and Model Estimates for the Third Quarter of 1991

Production Estimate, Crop Year 90/91	9,653,500 metric tons
% Change from Crop Year 89/90	+8.9%
Model Forecast, Third Quarter 1991	1,030,000 metric tons
% Change from Third Quarter 1990	-18%

This model attempts to forecast palay production on a quarterly basis. The model is based on regression analysis of rainfall data from stations throughout the Philippines. The national third quarter estimate presented here is the sum of the regional estimates for each of five agro-climatic zones. The crop year estimate is based on actual Bureau of Agricultural Statistics data for the first three quarters of the crop year and the model estimate for the last quarter of the current crop year. The forecast should be used as a complement to the standing crop estimates of the Bureau of Agricultural Statistics.

This monthly report is produced by the Policy Analysis Division, Planning and Monitoring Services, Dept. of Agriculture. For inquiries, call Preceles H. Manzo at tel. nos: 95-28-70 or 99-87-41 to 51 loc 210.

APPENDIX G

List of Common TSP Commands

DATA: Allows one to change old data or enter new data for the series named after the command (p.49).

EXIT: Exits from TSP.

FIT : Makes a forecast of the dependent variable in the immediately preceding regression. The name given to this forecast is specified by the user after the command (p.52-53).

GENR: Generates a new series by performing algebraic manipulations (e.g. addition, multiplication) on other variables (p.27).

LOAD: Loads a file into memory. The file name (including the path) must be specified after the command (p.11).

LS : Runs an ordinary least squares regression. The name of the dependent variable follows the command, followed next by a list of the independent variables (including the constant) (p.11).

PLOT: Plots a graph of all series named after the command (p.40).

SAVE: Saves the file named after the command (p.50).

SHOW: Shows the data for all series named after the command (p.42).

SMPL: Sets the sample range (p.16).

If a more detailed discussion of any of these commands is desired, please consult the TSP user's manual, a copy of which is at the Policy Analysis Division.

APPENDIX H

Performance of the Model vs. BAS Forecasts

Quarter	Actual Prod ('000 mt)	Model Forecast ('000 mt)	BAS Forecast ('000 mt)
1990-1	2146	2388 (+11%)	2468 (+15%)
1990-2	1337	1522 (+14%)	1288 (- 4%)
1990-3	1260	923 (- 27%)	1326 (+ 5%)
1990-4	4576	4219 (- 8%)	4424 (- 3%)
1991-1	2533	2416 (- 6%)	2758 (+9%)
1991-2	1515	1600 (+6%)	1410 (- 7%)
1991-3	1181	1030 (-13%)	1125 (- 5%)