BEFORE THE
POSTAL RATE COMMISSION

POSTAL RATE AND FEE CHANGES, 2006	:	**DOCKET NO. R2006-1**

WORKPAPER
UPS-T1-NEELS-WP-1
Revised

Workpaper in Support of Dr. Neels' Testimony (UPS-T-1),
one (1) CD-Rom of the Source Data and Programs
for His Data Analysis on IOCS and Mail Processing Variability

Contents

1 Introduction 1

2 Source Data 1

2.1 MODS Data 1
2.1.1 *vv9905.csv* 1
2.1.2 *vvscreens.csv* 2
2.1.3 *add9905.csv* 2
2.1.4 *opqtrFHP9905* 2
2.1.5 *MERGED_vv9905 and add9905 and FHP06.dta* 2
2.2 IOCS Data 2
2.2.1 *prcsas05.sas7bdat* 2
2.2.2 *prc04.sas7bdat* 2
2.3 RPW Data 3
2.4 Other 3
2.4.1 MODS Data/Data Errors 3
2.4.2 MODS Data/plant reporting/inputs 3
2.4.3 IOCS 3
2.4.4 Analysis of USPS Data 3

3 Analysis 4

3.1 MODS Data 4
3.1.1 Data Errors 4
3.1.2 Plant Activity Mix 8
3.1.3 Plant Reporting 8
3.2 IOCS 9
3.2.1 *iocs_2005_analysis.sas* 9
3.2.2 *iocs_2004_analysis.sas* 11
3.3 Other 11
3.3.1 Modeling of MP Costs 11
3.3.2 Plant Growth 11
3.3.3 Plant Size 12
3.4 Analysis of USPS Models 12
3.4.1 Chow Tests 12
3.4.2 Tests of Instruments 13
3.4.3 Analysis of Fixed Effects 13
3.5 Plant Level Model 14
3.6 Logit Models 14

1 INTRODUCTION

These work papers contain background material for the data analysis reported in my testimony on IOCS and mail processing variability analysis. The first section describes data sources. The second section describes the various analyses that I have performed using these data sources.

In the interest of minimizing the size of these work papers, I have not included copies of datasets if the data are already available in the same or similar form in this docket. The source data section contains details on the origins, sources and content of these datasets.

Most of the analysis contained here was performed using STATA version 9 for PCs. The file extensions affiliated with STATA are **.do* for executable program code, **.dta* for datasets, **.log* for log files containing output from programs. The Plant Level Model is estimated using "ado" files that are not included in the standard STATA package. The ado file called "xtivreg2" requires another ado file called "ivreg2." These are available for download by typing "ssc install ivreg2" and "ssc install xtivreg2" at the command prompt in STATA while the computer is connected to the internet. They are also available at http://ideas.repec.org/. I have also included them in the electronic back up to this document under the directory "ado."

A portion of the analysis is performed on the IOCS data using SAS version 9.1.3 running on the SunOS Operating System release 5.8. The executable program files have **.sas* extensions, the log files **.log* contain any return messages from running the program, **.lst* files contain output from commands. Datasets created from the SAS programs have **.csv* extensions.

The order that programs are listed below is the order in which they should be run.

2 SOURCE DATA

2.1 MODS Data

2.1.1 vv9905.csv

- *vv9905.csv* is a comma-separated file containing the data found in *vv9905.xls*. *vv9905.xls* is the analysis dataset used by Dr. Bozzo. It contains measures of piece handlings, volume, capital, labor hours, delivery points, and RPW volumes for the MODS plants used in Dr. Bozzo's analysis, quarterly, for the period 1999 to 2005. It is described in USPS-LR-L-56 Section 1.A.1, page 2. It was produced in Section 1 of USPS-LR-L-56.

2.1.2 vvscreens.csv

- *vvscreens.csv* is a comma-separated file containing the data found in *vvscreens.xls*. *vvscreens.xls* contains the number of counts of good and bad weeks and accounting periods (APS) of MODS data. These data are used to screen the MODS data for bad records. Dr. Bozzo describes that data in Section VII.E., page 96, of USPS-T-12. It was produced in Section 1 of USPS-LR-L-56.

2.1.3 add9905.csv

- *add9905.csv* is a comma-separated file containing the data found in *add9905.xls*. *add9905.xls* contains measures of piece handlings and labor hours for additional MODS groups used by Dr. Roberts. It was produced in Section 4 of USPS-LR-L-56.

2.1.4 opqtrFHP9905

- *opqtrFHP9905* is a text file containing FHP for the three-digit MODS codes for each plant and quarter from 1999 to 2005. Program *WP_Merge Data.do* reads in this data and creates the variable FHP for cost pool number 6 (FSM 1000) which is saved as *FHP06.dta*. *opqtrFHP9905* was produced in Section 4 of USPS-LR-L-56.

2.1.5 MERGED_vv9905 and add9905 and FHP06.dta

The program *WP_Merge Data.do* merges together *vv9905.csv*, *add9905.csv* and *opqtrFHP9905* to create *MERGED_vv9905 and add9905 and FHP06.dta.*

2.2 IOCS Data

2.2.1 prcsas05.sas7bdat

This is the dataset of responses collected by Postal Service data collectors when conducting the 2005 IOCS survey. It reflects the new methodology employed by the Postal Service in this survey. The new methodology is discussed by Dr. Bozzo in his USPS-T-46 testimony. The dataset is described in USPS-LR-L-9, USPS-LR-L-9_R2006-1_IOCS.pdf. The IOCS manual is available in USPS-LR-L-21, Handbook F-45, Data Collection User's Guide for the In-Office Cost System. The dataset was produced in USPS-LR-L-9.

2.2.2 prc04.sas7bdat

This is the dataset of responses collected by Postal Service data collectors when conducting the 2004 IOCS survey. It was conducted under the old IOCS methodology. The dataset is described in USPS-LR-K-9, IOCSLR9.DOC. The old IOCS manual is available in USPS-LR-I-14, Handbook F-45, Data Collection User's Guide for the In-Office Cost System. The dataset was produced in USPS-LR-K-9.

2.3 RPW Data

- *ups-14c-odisrpw.dta* is a STATA dataset derived directly from the dataset *ups-14c-odisrpw.xls* provided by Dr. Bozzo in USPS-LR-L-164. The dataset contains RPW volumes of destinating and originating volumes by shape and class of mail for each combination of IDNUM, fiscal year, and quarter found in *vv9905.xls*.

2.4 Other

2.4.1 MODS Data/Data Errors

- *bozzo_productivity_screens.csv* contains the high and low productivity cutoffs used by Dr. Bozzo. These screens can be found in the TSP programs in Section 1 of USPS-LR-L-56 and are described in Docket No. R2000-1, USPS-T-15, page 112, Table 5.
- *bozzo_rogg.csv* contains records for each IDNUM where Dr. Bozzo indicated that gaps in the reporting of MODS activity were related to seasonality of operations at those plants. This information is found in Dr. Bozzo's response to UPS/USPS-T12-38.
- *cost_pool_map.csv* contains a mapping of operations to cost pools.
- *shape_map.csv* contains a mapping of operations to shapes.

2.4.2 MODS Data/plant reporting/inputs

- *input.csv* is a comma-separated dataset which summarizes Dr. Bozzo's responses to UPS/USPS-T12-39 and 40 at the plant level. The information is first hand-entered and reformatted within the file *WP Summary of Plant Reporting - Yearly.xls, and* then is saved as this *.csv* file.

2.4.3 IOCS

- *mod_grps.txt* is a mapping of 3-digit MODS codes to the MODS groups that Dr. Bozzo uses in his USPS-T-12 testimony. Its source is *oper-grp-maps.xls* from USPS-LR-L-56, Section 4.
- *mods_fins.txt* is a list of all MODS-related finance numbers. The list is sourced from the material supporting Table 6 of Dr. Bozzo's USPS-T-46 testimony in USPS-LR-L-128.
- *mods_shape_xwalk.txt* is a mapping of 3-digit MODS codes to mailpiece shapes. This mapping was obtained from the Appendix of the MODS manual (USPS-LR-L-150).

2.4.4 Analysis of USPS Data

These datasets containing Dr. Bozzo's final analysis samples (one for each cost pool analyzed by Dr. Bozzo) were generated using TSP. They contain the final variables

used in Dr. Bozzo's models and contain only the records which were used in the final regressions. They have been converted to STATA format. The numbers included in the name of each dataset refer to the number of the cost pool represented.

- *MANUAL11.dta*
- *MANUAL12.dta*
- *MANUAL13.dta*
- *MANUAL14.dta*
- *MANUAL18.dta*
- *NONMANUAL04.dta*
- *NONMANUAL06.dta*
- *NONMANUAL34.dta*
- *NONMANUAL39.dta*
- *NONMANUALac.dta*
- *NONMANUALbd.dta*

3 ANALYSIS

3.1 MODS Data

3.1.1 Data Errors

Create Error Flags Data.do merges various data sources described in Section 2 for use in identifying and categorizing various errors in the analysis dataset used by Dr. Bozzo. These datasets are:

- *MERGED_vv9905 and add9905 and FHP06.dta*
- *vvscreens.xls*
- *cost_pool_map.csv*
- *bozzo_productivity_screens.csv*
- *shape_map.csv*

Flag Errors.do identifies errors in the analysis dataset used by Dr. Bozzo and prepares tables containing the results regarding the prevalence of these errors. The program also creates two datasets containing flags for bad data by shape. The errors are identified in most cases at the "operation" level. In some cases an operation is a cost pool, but in other cases two operations are combined to create a cost pool. An error is indicated for a particular cost pool if either or both of the operations that contribute to that cost pool exhibit that error.[1] The mapping of operations to cost pools is listed below.

[1] The productivity cutoffs for flag 4 and flag 5 (explained below) are assessed at the cost pool level for AC, BD, and 34, mirroring the method used by Dr. Bozzo.

Operation Number	Operation Name	Cost Pool Number	Cost Pool Name
4	OCR	4	OCR
6	FSM 1000	6	FSM 1000
9	SPBS Regular	34	SPBS Total
10	SPBS Priority	34	SPBS Total
11	Manual Flats	11	Manual Flats
12	Manual Letters	12	Manual Letters
13	Manual Parcels	13	Manual Parcels
14	Manual Priority	14	Manual Priority
18	Cancellations	18	Cancellations
72	MPBCS Outgoing	AC	Outgoing D/BCS
74	DBCS Outgoing	AC	Outgoing D/BCS
71	MPBCS Incoming	BD	Incoming D/BCS
73	DBCS Incoming	BD	Incoming D/BCS
93	AFSM Incoming	39	AFSM 100
94	AFSM Outgoing	39	AFSM 100

The program constructs a series of flags which represent whether a record exhibits a certain type of error. These flags are listed below.

- FLAG0 = 1 if HRS, FHP, TPF **and** TPH = zero (for automated operations). FLAG0 = 1 if HRS, FHP, **and** TPH = zero (for manual operations). FLAG0 = 1 if HRS, **and** TPH = zero (for cancellations). These are “true” zeroes.
- FLAG1 = 1 if HRS, TPF, **or** TPH = zero (for automated operations). FLAG1 = 1 if HRS **or** TPH = 0 (for manual operations). [And FLAG0 = 0].
- FLAG2 = 1 if TPH > TPF (for automated operations only [see page 107 of Docket No. R2000-1 USPS-T-15]).
- FLAG3 = 1 if FHP > TPH.
- FLAG13 = 1 if FHP > TPF (for automated operations only).
- FLAG4 = 1 if productivity is above or below cutoff used by Dr. Bozzo.
- FLAG5 = 1 if TPH data has less than 12 good weeks for quarters 1-3 or less than 16 good weeks for quarter 4, or if TPH data has less than 3 good pay periods for quarters 1-3 or less than 4 good pay periods for quarter 4.
- FLAG6 = 1 if TPF data has less than 12 good weeks for quarters 1-3 or less than 16 good weeks for quarter 4, or if TPF data has less than 3 good pay periods for quarters 1-3 or less than 4 good pay periods for quarter 4.
- FLAG7 *[NOT USED]*
- FLAG8 *[NOT USED*]

- FLAG9 = 1 if HRS, FHP, TPF and TPH = 0 (for automated operations) or HRS, FHP, and TPH = 0 (for manual operations) in this period but one of these variables was greater than zero at an earlier and later time.
- FLAG10 = 1 if Dr. Bozzo reviewed a “zero” and indicated it was a true “zero.”
- FLAG11 *[NOT USED*]
- FLAG12 = 1 if HRS, FHP, TPH, or TPF < 0.
- FLAG14 = 1 if (FLAG0 = 1 & FLAG9 = 0) | (FLAG0 = 1 & FLAG9 = 1 & FLAG10 = 1).

These flags are summarized to generate tables showing the following for each cost pool:

- *table_2.csv* contains this information as raw counts (*Data Error Counts.xls* formats this data for presentation).
- *table_6.csv* contains this information as the cumulative result.

That is, each line is the count of records that exhibit the condition described for that line, but do not exhibit the condition of a previous line in the table (*Cum. Data Error Counts*.xls formats this data for presentation).

Shape level counts of the cumulative effect of each line described below are found in *table_shape_cum.csv*. *Analysis Sample.xls* formats this data and shows how the final analysis sample was achieved.

Line	Label	Contents For cost pools that are single operation	Contents For cost pools that are two operations
1	Total records in Dr. Bozzo's Analysis Dataset	total obs: 10, 304	total obs: 10, 304
2	Valid zeros: Count of records where operation was not present at plant, plant was not reporting data to MODS system, or plant was not in operation	count of "true zeros" (FLAG0 = 1 & FLAG9 = 0) | (FLAG0 = 1 & FLAG9 = 1 & FLAG10 = 1) [FLAG14]	Max FLAG14 = 1 & min FLAG14 = 1
3	Valid Sample	Line 1- Line 2	Line 1- Line 2
4	Zeros caused by apparent gaps in the data	FLAG0 = 1 & FLAG9 = 1 & FLAG14 = 0	Max FLAG0 = 1 & max FLAG9 = 1 min FLAG14 = 0
5	HRS, TPF, FHP, or TPH < 0	count of FLAG12	Max of FLAG12 = 1 & in valid sample
6	HRS, TPF, or TPH = 0	count of FLAG1	Max of FLAG1 = 1 & in valid sample
7	Record fails Dr. Bozzo's Threshold or Productivity Screens at the quarterly level	count of FLAG4	Max of FLAG4 = 1 & in valid sample
8	TPH fails Dr. Bozzo's Threshold or Productivity Check at the AP or weekly level	count of FLAG5	Max of FLAG5 = 1 & in valid sample
9	TPF fails Dr. Bozzo's Threshold or Productivity Check at the AP or weekly level (for automated operations)	count of FLAG6	Max of FLAG6 = 1 & in valid sample
10	FHP > TPH	count of FLAG13	Max of FLAG13 = 1 & in valid sample
11	FHP > TPF (for automated operations)	count of FLAG3	Max of FLAG3 = 1 & in valid sample
12	TPH > TPF (for automated operations)	count of FLAG2	Max of FLAG2 = 1 & in valid sample

The program also creates two datasets containing flags for bad data. The first dataset *plant_and_shape_level_data_w_error_flags_V1.dta* has an indicator variable for each shape that equals "1" if the record exhibits any of the issues indicated in lines 4 through 9 in the above table. The second dataset *plant_and_shape_level_data_w_error_flags_V2.dta*, contains a looser definition of bad data, where a problem is indicated if the record exhibits any of the conditions described in lines 4 through 8.

3.1.2 Plant Activity Mix

The program under Plant Activity Mix is designed to (1) determine whether or not a plant is reporting for each of Dr. Bozzo's MODS Cost Pools, and (2) summarize by shape, sorting categories letters, flats, parcels, and priority mail plant reporting by MODS Cost Pools. Results are then presented in an Excel spreadsheet.

WP change in activity mix.do uses the previously described dataset *MERGED_vv9905 and add9905 and FHP06.dta.* First, records for quarter four only of each year and plant are kept to represent end-of-year reporting status. Then, total TPH, FHP, and HRS are calculated for each of Dr. Bozzo's MODS Cost Pools as well as totals for all pools. A plant is defined as reporting if any of the three variables TPH, FHP, or HRS show a non-zero value, and as not reporting otherwise. Output from this program is contained in *activity mix.csv.*

WP Plant Activity Mix.xls presents the results from *WP change in activity mix.do* in table format.

3.1.3 Plant Reporting

The program and other files under Plant Reporting are designed to (1) determine whether or not a plant reports for each year, and (2) determine whether or not reportage or non-reportage is due to a start or stop in operations based on Dr. Bozzo's responses to UPS/USPS-T12-39 and 40. Results are then presented in an Excel spreadsheet.

WP reporting flags - MODS grp - yearly.do uses the previously described datasets *MERGED_vv9905 and add9905 and FHP06.dta* and *input.csv*. First, the dataset is collapsed to yearly totals for TPH, FHP, and HRS. Then, TPH, FHP, and HRS are totaled across all of Dr. Bozzo's MODS Cost Pools. A plant is defined as reporting if any of the three variables TPH, FHP, or HRS shows a non-zero value, and shows as not reporting otherwise. Instances in which a plant starts reporting for the first time and stops reporting are identified. Then, these start and stop indicators are further distinguished in on/off/start/stop flags to indicate respectively plants going online, going offline, discontinuing MODS reporting with continued operations, and initiating MODS reporting with continued operations, according to Dr. Bozzo's explanations. Plants not explained by Dr. Bozzo are assumed to have continuing operation. Output from this program is contained in *reporting flags - MODS grp - yearly.csv.*

WP Summary of Plant Reporting - Yearly.xls compiles Dr. Bozzo's interrogatory responses as previously described, and presents the results from *WP change in activity mix.do* in table format.

3.2 IOCS

3.2.1 iocs_2005_analysis.sas

This SAS program performs three functions:

3.2.1.1 Calculate dollar tallies by type (direct, mixed, tall pallet boxes, or nothandling) and by MODS group

The SAS program selects all records from the IOCS Tally File which meet the following criteria:

- Employee is a clerk or mailhandler or city carrier (based on the value of variable f257)
- Employee is not at a CAG K office (based on the value of variables f261 and f264)

The IOCS tallies are then divided into three facility groups, based on the tally finance numbers:

- MODS (matched to values in *mods_fins.txt*)
- BMCS (matched to BMC finance numbers identified in *mbcLOG.rtf* of USPS-LR-L-55)
- NONMODS (residual)

The above step is necessary as the programming processing tasks are organized and performed separately for each of the above three groups.

The next task is to identify tallies that involve Mail Processing, Window Service, and City Carrier operations. To do this, the program begins by associating an LDC to each 3-digit MODS code. MODS code numbers are stored in variable Q18A03. Mail processing tallies include those in Function 1 (LDC = 1X), most of those in Function 4 (LDC = 41-44, LDC = 48-49) and those with LDC = 79. Window Service tallies include those in LDC = 45. Tallies for Claims and Inquiries include those for LDC = 75. The remaining tallies are assigned to Administrative Services. This task uses the same code and logic as employed by the Postal Service in program module MODS05 in USPS-LR-L-9.

Tallies with no MODS codes or invalid MODS codes are then remapped using the same criteria employed by the Postal Service. See REMAP05 module in USPS-LR-L-9. For these tallies (BMC and NONMODS tallies), those tallies that were remapped to the "2WINDOW" pool are designated Window Service tallies. Special service activity codes are also reassigned based on MODS cost pools as done by the Postal Service in module MOD1POOL of USPS-LR-L-9.

The program then classifies tallies as direct, mixed, tall pallet boxes, and nothandling based on whether the tally is a MODS, BMC or NONMODS tally.

For mail processing tallies, the MODS tallies are classified using the logic in the MOD1DIR module, the NONMODS tallies using the NONMOD1 module, and the BMC tallies using the BMC1 module provided by the Postal Service in USPS-LR-L-9.

Window Service tallies are classified based on the tally's activity code (variable f9806). Window service tallies with an activity code between 10 and 4950 are direct tallies, those with an activity code between 5610 and 5750 are mixed tallies, while the rest are nothandling tallies.

Carrier costs tallies are determined based on the value in variable f257. (Please see the IOCS handbook for this roster designation.) These tallies are also classified by type based on the tally's activity code. Window service tallies with an activity code between 10 and 4950 are direct tallies, those with an activity code between 5300 and 5750 are mixed tallies, while the rest are nothandling tallies.

Finally, Dr. Bozzo's MODS groups, as used in his IOCS testimony, USPS-T-12, are assigned to each tally based on the mapping from *mod_grps.txt*.

The processed dataset is collapsed by tally type and cost pool resulting in *tally_type_by_cost_pool_06.csv* and is combined with *tally_type_by_cost_pool_05.csv* in the workbook *IOCS Tally distribution by type.xls* which presents final results.

3.2.1.2 Estimate misclocking error in FY2005 in Cost Segment 3

The program estimates the level of misclocking in the IOCS dataset. That is, this program compares the MODS activity an employee is clocked into at the time the tally is taken with the activities that employee is actually observed performing.

To determine this, the data are first divided into clerk and mailhandler tallies only. City carrier tallies are outside the scope of our interest. The program then assigns actual observed activity variables to each tally based on the recorded response to various questions. For example, if the value in variable Q18C02 is "A", then the employee was observed actually taking part in a D/BCS operation. These observed activity variables are then juxtaposed with the clocked MODS codes to determine if misclocking occurred.

The output from this analysis is *migration_by_cost_pool_06.cvs.* This data is further manipulated and presented in the excel workbook *Actual activity vs clocked MODS.xls*

3.2.1.3 Estimate shape distribution for manual priority

The purpose of this section of the program is to estimate the proportion of letters, flats and parcels in the manual priority operation. To determine this, the program first subsets the data into only tallies that have a manual priority MODS code as determined

by the mapping in *mod_grps.txt*. These are 3-digit MODS codes that map to group 14 in *mod_grps.txt*.

Each of the 3-digit MODS codes in manual priority is associated with a shape as determined in *mods_shape_xwalk.txt*. The program then collapses dollar tally information by shape and outputs *prioshape.csv.* These values are read into *Shape composition of Manual Priority.xls* which computes the final distribution key for manual Priority by shape. These values are used in the Plant Level Modeling effort.

3.2.2 iocs_2004_analysis.sas

This SAS program mirrors *iocs_2005_analysis.sas* but only calculates dollar tallies by type (direct, mixed, tall pallet boxes, or nothandling) and by MODS group. The code in this program reflects the different variable names used in *prc04.sas7bdat* as well as any differences in Postal Service modules in USPS-LR-K-9 (Docket No. R2005-1) from those in USPS-LR-L-9 (Docket No. R2006-1). The output is contained in *tally_type_by_cost_pool_05.csv* and is combined with *tally_type_by_cost_pool_06.csv* in the workbook *IOCS Tally distribution by type.xls* which presents final results.

3.3 Other

3.3.1 Modeling of MP Costs

The excel workbook *Modeling of Mail Processing Costs.xls* calculates the percent of mail processing costs modeled by Dr. Bozzo in Docket Nos. R2001-1, R2005-1, and R2006-1.

3.3.2 Plant Growth

The program under "plant growth" is designed to categorize growth for each plant. Ultimately, a plant is labeled "POS" if it shows positive growth, "NEG" if it shows negative growth, and "-" if growth is not identified.

WP growth by plant.do uses the dataset *ups-14c-odisrpw.dta,* previously described, calculates total RPW volume for each plant and quarter, and then fits the simple regression model:

$$TtlRPW_i = \alpha_0 + \beta_1 Time_i + \varepsilon,$$

in which i = Plant, and Time = Time Trend 1-28. RPW volume is summed across destinating values only, to represent the RPW volumes used in Dr. Bozzo's analysis. Regression results are then stored in the output files *rpw-growth_by_plant_1-200.csv* and *rpw-growth_by_plant_201-368.csv.*

WP growth_by_plant.xls uses the regression results generated by *WP growth by plant.do* to categorize each plant. A plant is first categorized as a growing plant if the

coefficient on the independent variable is at least significant at a 5% or 1% level of confidence. Plants demonstrating growth are then categorized as "POS" or "NEG" depending on the sign of the coefficient of the independent variable. The final result by plant is then saved as a *.csv* file of the same name.

3.3.3 Plant Size

The program under "plant growth" is designed to categorize the relative size of each plant. Ultimately, plants are labeled as "BIG," "SMALL," or "N/A" (if not enough RPW data are reported).

WP plant size.do uses the dataset *ups-14c-odisrpw.dta* previously described, calculates total RPW volume for each plant and quarter, and replaces zero values ("0") as missing (".."). RPW volume is summed across destinating values only, to represent the RPW volumes used in Dr. Bozzo's analysis. The RPW totals are then averaged by plant and compared with the median of these averages. A plant is categorized as "BIG" if its average is greater than or equal to the median of all plant averages, and "SMALL" if its average is less than the median of all plant averages. Results are then stored in the output file *plant size.dta.*

3.4 Analysis of USPS Models

The analysis contained in this folder analyzes the models estimated by Dr. Bozzo. The data and models used by Dr. Bozzo are explained in USPS-T-12 and USPS-LR-L-56 Section 1. In each case, the analysis uses the same model specification as Dr. Bozzo and is run on the same observations used by Dr. Bozzo.

3.4.1 Chow Tests

The three programs in this directory perform F-tests and Likelihood Ratio Tests to determine whether a particular subset of the data can be pooled with the rest of the data.[2] These tests are conducted by creating an indicator variable for a subset of records. All the explanatory variables (except the fixed effects indicator variables) are interacted with the indicator variable and added to the original set of explanatory variables and the model is reestimated. The joint significance of the interacted variables is then tested. If the variables are jointly significant, the conclusion is that the data cannot be pooled for model estimation. Each program takes as its starting point the analysis datasets originally used by Dr. Bozzo. These are listed in Section 2.4.4. The estimation procedure used exactly matches that of Dr. Bozzo, and it is performed using *xtivreg2.ado* with fixed effects and robust standard errors (limited information, maximum likelihood).

[2] Note that the use of the term "Chow Test" to refer to these tests is merely shorthand for the concept of testing whether subsets of data can be pooled together.

- *WP_MergeSampleFlags.do* creates a set of indicator variables (one for each cost pool) that equals "1" if a record was included in Dr. Bozzo's regressions for that particular cost pool. These indicator variables are found in *Witness Bozzo sample Flags.dta.*
- *WP_Create Chow Test Datasets.do* merges *WP growth_by_plant.dta*, *plant size.dta* and *Witness Bozzo sample Flags.dta* to the regression data sets used by Dr. Bozzo.
- *WP_Chow_Big vs Rest.do* tests whether big plants can be pooled with the rest of the plants. The "BIG" designation is taken from the file *plant size.dta* found under "plant growth/output." It is created by the program *WP plant size.do.*
- *WP_Chow_Growing vs Static.do* tests whether growing plants can be pooled with the rest of the plants. The "growing" designation is taken from the file *WP growth_by_plant.csv* found under "Other/plant growth/tables." It is created by the program *WP growth by plant.do*.
- *WP_Chow_Technology Pooling.do* tests whether plants that are active in multiple equipment types for the same shapes can be pooled together with the plants that are active in only one of those equipment types. The information on which plants have which equipment is taken from *Witness Bozzo sample Flags.dta.* The following subsets of technologies are tested:
    - Can plants that are active in both AFSM100 and FSM1000 technologies be pooled with those that are only active in FSM1000?
    - Can plants that are active in both AFSM100 and manual flats technologies be pooled with those that are only active in manual flats?
    - Can plants that are active in both SPBS and manual parcels technologies be pooled with those that are only active in manual parcels?

3.4.2 Tests of Instruments

WP_Examination of Bozzo IV Models.do uses Dr. Bozzo's Manual Regression Samples listed in Section 2.4.4. The program replicates Dr. Bozzo's Instrumental Variable analysis and produces statistical diagnostics that are used to examine the relevancy and validity of the instruments that are used by Dr. Bozzo. The estimation procedure used exactly matches that of Dr. Bozzo; it is performed using *xtivreg2* with fixed effects and robust standard errors (limited information, maximum likelihood).

3.4.3 Analysis of Fixed Effects

WP_fixed effects.do runs the fixed effects models used by Dr. Bozzo and outputs the coefficient results to *manual_results.csv* and *nonmanual_results.csv*. The workbook *Table of Fixed Effects.xls* reads in these coefficient estimates, calculates the antilog of each of the fixed effects parameters, identifies the largest and smallest values for those fixed effects, and presents the ratio of the smallest to the largest, for each cost pool analyzed by Dr. Bozzo.

3.5 Plant Level Model

WP_Plant-Level_Regression_final.do merges together *vv9905.dta*, *FHP06.dta*, *Witness Bozzo sample Flags.dta*, *plant_and_shape_level_data_w_error_flags_V1.dta,* and *plant_and_shape_level_data_w_error_flags_V2.dta*, and creates *Bozzo Regression Samples Flagged in vv9905_fhp06_Brattle.dta*

The program creates the shape level variables from the FHP, TPH, and HRS variables which are at the cost pool level. The shape level FHP, TPH, and HRS variables are rescaled for the extra AP at the 4th quarter for all years. ODIS destinating volume variables are adjusted for the 4th quarter for years prior to 2004. The program allocates manual priority to the other shapes by using the distribution keys provided in *Shape composition of Manual Priority.xls*. The program runs the plant level model where total hours are regressed on three shape-based volume variables, delivery point and time trend (all in logs but time trend) on the samples defined by the strict and loose definitions of errors, using the xtivreg2 command described in the Introduction to this document. For these models, I continue to use xtivreg2 with fixed effects and robust standard errors (limited information, maximum likelihood).

3.6 Logit Models

Logit analysis.do uses *Bozzo Regression Samples Flagged in vv9905_fhp06_Brattle.dta* to analyze the automation response to volume growth by estimating simple logit models. The program creates 3 binary variables: TECH06 which is equal to 1 if the plant has the FSM1000 technology, TECH39 which is equal to 1 if the plant has the AFSM100 technology, and TECH34 which is equal to 1 if the plant has the SPBS technology. The program runs logit models using lntph_FLATS as the independent variable for each of TECH06 and TECH39, and lntph_PARCELS as the independent variable for SPBS. The logit is run using the loose definition of errors.