Notes on the PROCISUR Study

Robert E. Evenson
Preliminary Background Paper
PROCISUR Evaluation Planning Meeting
Montevideo - December 19-21, 1988

1. General Conceptual Issues

The PROCISUR program is a specialized program designed to provide a number of support services to facilitate national program conduct. It is not generally building and supporting commodity research program per se. This type of support is also important within countries where national programs support and facilitate state and regional programs. In an international context this support may enable better exchange of technology and germplasm (see below for a more general definition and treatment of germplasm then usually offered), between national programs and between national and International Agricultural Research Centers (IARC's).

Evidence from a single country can be utilized to assess PROCISUR impact but one must be realistic as to the possibilities for measuring an impact. This is because the PROCISUR program will typically constitute a small part of the resources in a commodity program in a given country. Unless PROCISUR has clearly enabled a specific readily identified line of research it will be very difficult to separate the PROCISUR contribution from national program investment.

Nonetheless there is much merit in country studies for the following reasons:

1) There may be cases where one can identify PROCISUR impacts. Even if these are part of a more general impact these studies will be valuable.

2) They will provide institutional richness to an international study.

3) They will provide a data base for an international study.

The likelihood that one can identify PROCISUR effects in international data, i.e., pooled data from the relevant countries is greater than in most country studies, because the PROCISUR program has many multiple country impacts

(just as the IARC's have multiple country impacts.) The potential for an international study is indicated in two recent studies. The first is a World Bank study of several commodities that sought to estimate IARC impacts in several Latin American countries. The second is a recent extension of a study of technology transfer in Brazil. This study develops new techniques for estimating types of technology transfer between regions and countries and opens up new possibilities for estimating the impacts of transfer enhancing activities such as the PROCISUR programs.

2. **Proposed Work Program (tentative)**

1. December Workshop. Montevideo

This workshop will discuss data requirements from each country and national and international methods. Selection of commodities (corn, wheat, soybeans, others?).

2. January-February. National Data Base Completion.

Each participating group will compile a data set for each commodity for the years 1966 to the most recent period 1988, if possible, and for regions within the country.

3. March-April. National studies - International Studies.

Each participating group will complete statistical estimates to be presented and discussed at May workshop in Brazil.

Elmar R. da Cruz and R.E. Evenson will estimate an international system (at Yale) based on national data, IARC data and PROCISUR impact data.

4. May - Workshop in Brazil.

National studies will be finalized and reported. International results will be reported and discussed.

3. **General Data Requirements by Commodity**

We have two alternatives regarding data. 1) We could attempt a "Total

Factor Productivity" (TFP) measure for different regions, 2) we could attempt partial productivity analysis (with certain price adjustments). For Brazil, we have some TFP measures for detailed regions (muncipios and, of course, states) for the years 1970-75-80 and 85. We probably do not have them for other countries. It seems, however, that we are wisest to pursue the partial productivity analysis of selected commodities for two reasons. First, the transfer questions differ considerably by commodity. Second, we can obtain more recent data and recent data is important because the PROCISUR program is probably having an impact on production (if it has had an impact) only after 1985 or so.

The commodities in order of importance would seem to be wheat, corn, soybean, cattle (possible other commodities are rice and barley). The data to analyze cattle productivity gains are probably not really available. Thus it appears to make best sense to concentrate on wheat, corn and soybeans.

Thus for each country and commodity the following should be collected by year (1966 to date), and by province or state (i.e., the smallest geographic unit).

1. Area planted (area harvested also if possible)
2. Production
3. Yield
4. Varieties used - if possible
5. Fertilizer used on the crop or other crops if possible
6. Price rations - fertilizer/crop, wages/crop, where crop prices are prices received.
7. Extension services supplied
8. Public and private research expenditure (over recent years since 1960)
9. Farm schooling -other indicators
10. Use of ICAR assistance, varieties, etc.
11. PROCISUR specific program inputs with other countries
12. Geo-climate classification by regions
13. Weather indexes.

It will not be possible to obtain all of the data listed above. Obviously production and area are central. We can probably work out the PROCUSUR inputs at the meeting in December....

We will illustrate data for Brazil and Paraguay at the meeting. We will also discuss a method for measuring weather impacts - see below.

In each country a germplasm-plant breeding analysis might be undertaken. Experiment station trial data may prove to be useful.

4. Methods - National Studies

A simple PFP (Partial Factor Productivity) analysis.

The first step is to construct a PFP measure. The best is probably a simple yield index.

$$PFP_t = Y_t/Y_{66-68}$$

where Y_{66-68} is the average yield for 1966-68. This index has three limitations.

(1) It doesn't take non-land inputs (especially fertilizer and irrigation) into account.

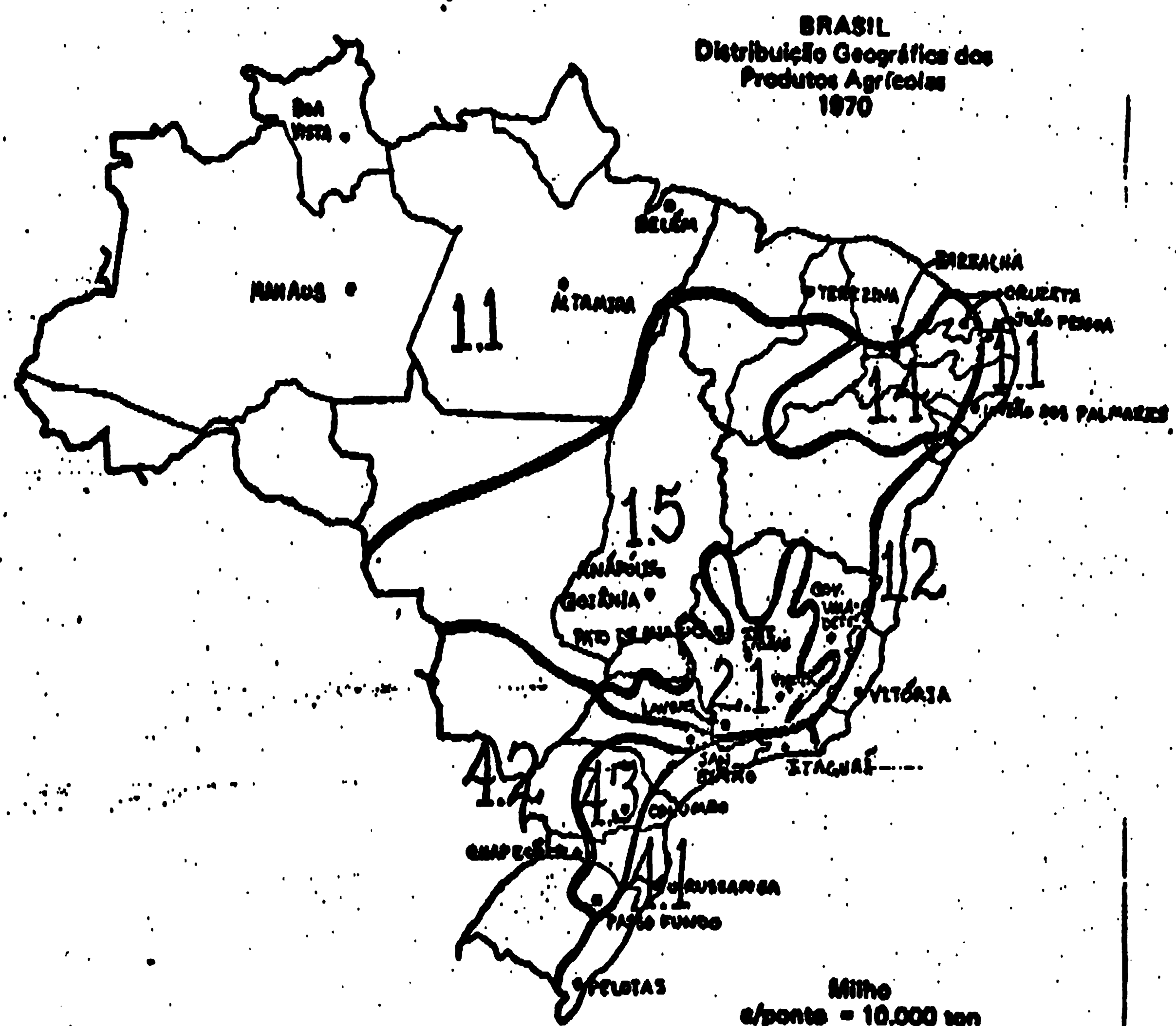
(2) It is affected by acreage shifts because new land planted to the crop may be better (or worse) than existing land planted to the crop.

(3) It is affected by weather problem.

One can do something about each of these problems. Problem (1) can probably best be dealt with by some price ratio variables. Unfortunately with the inflation rates in the region in recent years one may have very poor measures. It would be useful to actually construct some inflation indexes to control for these things. Fertilizer data could be used even if it isn't available by commodity.

The second problem could be dealt with by constructing acreage variables and using them as regressions. This is problematic because acreage is an endogenous variable much the way yield is, i.e., it is affected by research programs. (Prices could be used to "instrument" acreage - possibly we could use international prices.) This problem is helped by obtaining data on relatively small regions.

Figure 18


BRASIL
Distribuição Geográfica dos
Produtos Agrícolas
1970
MANAUS
ALTAMIRA
BELÉM
TERESINA
GOIÂNIA
ANÁPOLIS
VITÓRIA
PELOTAS
PASSO FUNDO
CHAPECÓ
1.1
1.4
1.5
1.2
2.1
4.2
4.3
4.1
Milho
c/ponto = 10.000 ton

learning by the farmer before the full impact of the research investment will be realized. Some research projects are unsuccessful. Some produce new technology, but also produce new intellectual capital that enhances further research projects. Some are not designed to produce technology per se, but have pre-technology science objectives.

Furthermore, technology once adopted by farmers may experience a real depreciation in value or TFP impact. This stems from two sources. The first, and probably most important in agriculture, is through real "deterioration" from exposure to pests and pathogens. This is a common problem with new crop varieties and to some extent with animal improvements as well. The second is through "replacements with incomplete additivity." New inventions are continuously replacing older inventions because the new inventions are superior. In some cases they build upon or add to the older invention. In these cases the TFP impact of the older replaced invention does not deteriorate but is an actual part of the new invention. But this additivity may not be complete. The new invention may have emerged from a different technology core or sequence of inventions. In this case it will not contain the full effect of the replaced invention.

From the perspective of specifying a research variable to be associated with TFP change in a given time period one must look backward in time and include the research investment or activities that are effectively "contained" in the TFP index. If the TFP index is in the form of an annual change this can include negative weights.

Set	a	b	c
1	5	6	15
2	7	6	20
3	9	6	30

Extension programs also have a time lag, but it differs considerably from the research lag. Extension programs have direct and relatively quick impacts because of direct contact with farmers. Because of an education and learning process, these impacts will have a rising component over time. They will also have a falling component because there are good substitutes for public extension programs. Markets supply information to farmers; private firms also supply information. Much of the public extension effect is to enable processing and conversion of technical and price information into managerial decision-making earlier and more effectively. Alternative sources of information are of the "replacement with incomplete additivity" type. Hence, much, perhaps most, of the extension impacts deteriorate within a relatively short period.

Given the burden of estimation of other parameters in this study the extension time weights were not estimated. Instead they were imposed to last only 3 periods with time weights of .5, .25 and .25. Schooling impacts were specified to be permanent. Again, schooling-associated skills may deteriorate over time but this study does not attempt to estimate this effect--except in interaction effects (see below). (Government price variables are treated as having a 3-year impact.)

Spatial-Geo-climate Dimensions (Spill-in of Technology)

Since the unit of observation is productivity in a specific time period, for this study, research, extension and other variables must be "matched" with the unit of observations. For some variables, it may be argued that there is no appreciable spatial issue, because the variable is closely associated with farm producers. This is the case for schooling and also for extension, but not for research.

If one could actually measure technology in use by farmers directly one

could possible trace it to its origins. For example, technology in use in a given state may have originated (i.e., been invented) in another state or even in another country. If so it can be said to have "spilled-in" to the state in question and "spilled-out" of the origin state. Using this spill-in and spill-out information, one could attribute the value of technology to its originating institution.

Some technology spills far and wide. For example, a chemical herbicide may be more valuable than the next best alternative in every region. In economic terms it is the best technology in a broad range of locations. If all agricultural technology had this characteristic one would specify a single national (or international) research stock utilizing the time shape weights noted above. But most agricultural technology does not spill far and wide. Spilling is inhibited by soil, climate, and even economic factors. The biological performance of a variety of corn, for example, is inhibited by changes in day length and length of growing season.

As crop and animal husbandry priorities were developed, "husbandry selection" modified many crop and animal species through selection for economically valuable characteristics. Considerable improvement in economic species occurred over the centuries prior to the modern agricultural research period. Some of the natural inhibitors were reduced in scope and importance so that economic species exhibited much less fine tailoring to small niches than non-economic species. Nonetheless, the basic pattern of tailoring through location specific husbandry selection was maintained.

With the advent of modern plant breeding and research practices, further selection to reduce inhibiting effects has taken place (e.g., modern high-yielding rice varieties in Asia have been selected for lower photoperiod sensitivity). At the same time, the existence of inhibitions (sometimes

referred to as geno-type-environment intervention (see Herdt, Kauffman, et.al.)) has become a central feature of the organization and design of agricultural research systems. (Englander, 1987.)

Given the fact that most inhibitors are soil and climate factors, information on these factors can assist in the specification of spill-in and the matching of research programs with state TFP observations. Geo-climate regions and sub-regions can be used to estimate. Figure 1 is a geo-climate region map for Brazil. Each state contains one or more regions and subregions. A spill-in specification can be estimated for research by estimating the parameter α in the following expressions.

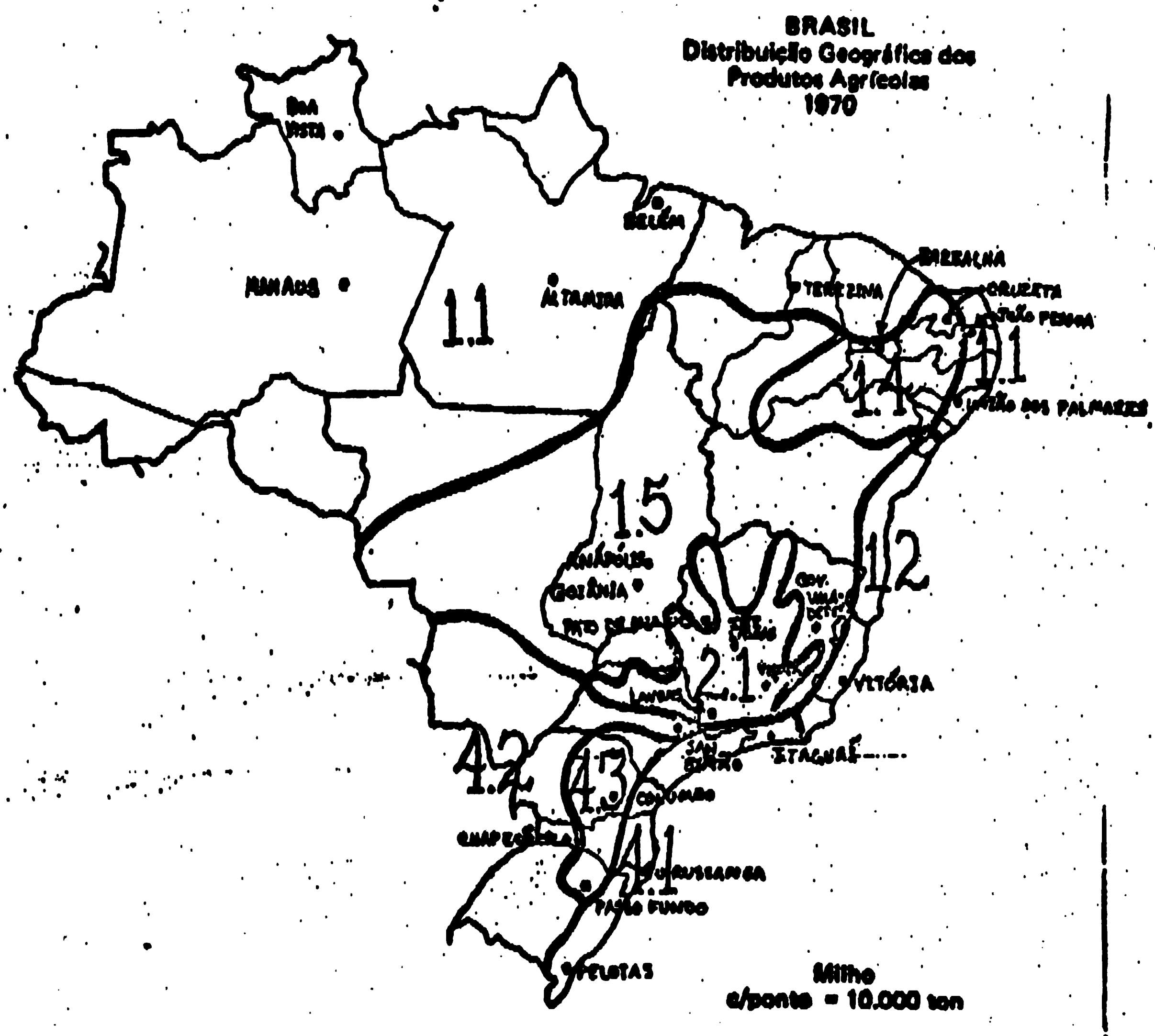
$$R_i^* = R_i + (1-\alpha)R_{gs} + (1-2\alpha)R_{gR}$$

where R_i^* is the constructed research stock for state i, R_i is the state's own research stock, R_{gs} is the research in other states conducted in similar sub-regions to those in state i; and R_{gR} is the research stocks conducted in other sub-regions in similar regions to those in state i. If state i has more than one region the above expression is computed for every region and weighted by the region's importance.

This procedure implicitly subsumes both direct and indirect spill-in of research. The specification imposes a spatial structure on the weights. Some research products produced in similar sub-regions (or regions) may be directly transferable and these products are "substitutes" for a state's own research program. Other findings outside the state indirectly spill in by complementing the states research program. The specification indicated above does not seek to identify substitutability or complementary. It seeks to measure net spill-in. (See below for more.)

Consider the case where a single commodity is being produced in a single homogenous region with no spill-in. In this case a research stock should not

Figura 18


BRASIL
Distribuição Geográfica dos Produtos Agrícolas 1970
MANAUS
ALTAMIRA
BELÉM
TERESINA
GOIÂNIA
VITÓRIA
PASSO FUNDO
PELOTAS
1.1
1.4
1.1
1.5
1.2
2.1
4.2
4.3
4.1
Milho
c/ponto = 10.000 ton

DEC

be

whe

abs

Lar

be deflated at all. The form of the productivity specification would depend on whether the dependent variable was measured as a rate of change or as an absolute quantity (e.g., output-inputs). It would not matter, however, how large the homogenous region in question was.

Now suppose that the region is not homogeneous, but that there are sub-regions in it and that we have two states each with a different number of sub-regions. Each state has a station that seeks to tailor technology to each sub-region. How do we define a meaningful research stock variable for the two states?

Consider two extremes. One is that the sub-region characteristics do not inhibit technology spilling from one region to the other. In this case the sub-region would not matter. At the other extreme, no significant spill-over, even indirectly, takes place between sub-regions. In this case each sub-region would require a separate research program and the aggregate program research stock could be defined as:

$$\sum_i S_i R_i$$

where S_i is the share of production in the ith sub-region. This "deflated" aggregate research stock presupposes not only that no spill-over between sub-regions occurs but that the system is optimally allocating research between sub-regions in proportion to the size of the sub-region.

Some spill-over is likely to occur between sub-regions and not all sub-region allocation will be optimal. This would require a mixed or composite deflator:

$$\alpha \sum_i R_i + \beta \sum S_i^C R_i + \lambda (S_i^C - S_i^R)^2 R_i$$

where S_i^C is the share of commodity production in subregion i and S_i^R the share of research directed to sub-region i. The third term of this expression is a correction for non-optimal sub-region allocation of

research effort. The first and second constitute a composite deflator with weights α and β.

Typically data to a priori specify the third term or the weight α and β do not exist. It is possible, however, to actually include two research stock variables in a decomposition specification and implicitly estimate α and β.

The analyst has an option to estimate the basic specification with OLS methods. There is merit, however, in estimating 3 equations in a seemingly unrelated regression system where error correlation in the 3 equations (say for wheat, corn, and soybeans) are considered in the estimation.

5. **Methods International Studies**

The Appendix to this note reports the results of a recent international study that effectively "pooled" data for a number of countries to enable a more meaningful analysis of the IARC programs of research. It demonstrates the merits of the approach and it was generally successful in identifying IARC effects in Latin America. Since the PROCISUR program is another type of IARC program with some differences, of course, this approach appears more promising. Its promise, however, is further improved by recent developments in technology transfer specifications.

Evenson (1988) developed the basic approach in a study of U.S. agricultural productivity. Evenson and da Cruz have recently applied it to Brazilian soybean data (a summary is presented below).

The basic idea is that for each geographic unit, say a state, one can define geo-climate neighbors. In the U.S. and Brazil studies, geo-climate zone data were used to define "close" neighbors, i.e., "similar sub-regions", and regions which are close to the close regions, i.e., similar regions. These regions can thus extend across international boundaries. Variables that can be defined for a state (Vg) then can be defined in analogous way for similar c

ch

"t

pr

sui

sci

type

desi

conc

endo

The m

(

I

P

P

P

T

the pr

the sim

sub-regions (V_{ss}) and for similar regions (V_{sR}. This approach defines 3 types of technology transfer:

<u>Direct</u>: As when technology produced in region i and which has a certain cost advantage in region i also has the same cost advantage in region j.

<u>Semi-direct</u>: As when the technology produced in region i does not have the same cost advantage in region j, but where it serves as a kind of "technological germplasm" for researchers in region j. That is, research programs in region j can adapt and modify the technology from i to make it suitable to conditions in j.

<u>Indirect</u>: As when research findings in i constitute a type of general or scientific germplasm that facilitates research programs in region j. This is a type of "pre-technology" germplasm transfer.

It can readily be seen from these definitions that the PROCISUR program is designed to facilitate and support all 3 types of transfer but is particularly concerned with semi-direct and indirect transfer.

The statistical and analytic model then is one in which there are six endogenous or jointly determined variables and a number of exogenous variables. The model is a six equation model:

P_s: $P_{ss}, P_{ss}R_s, R_s, R_{ss}, R_{sR}R_s, R_sR_{sR}, E_s$, etc.

P_{ss}: $P_s R_s, P_sR_{ss}, R_s, R_{ss}R_{sR}, R_s, E_s$

P_{sR}: $R_{ss}, R_{ss}R_{sR}, R_s, E_s$

$P_{ss}R_s$: R_sR_{sR}, E_s

P_sR_{ss}: R_sR_s, E_s

P_sR_{sR}: R_sR_s, E_s

$P_S = \frac{P_i}{P_{OP}}$

The 6 left-hand side variables are the endogenous variables that measure the productivity indexes in each state and its similar sub-region neighbors and the similar neighbors to those sub-region (i.e., sub-region).

The term P_{gg} in the first equation is then indexing direct transfer or spill-in from the similar sub-regions. The term $R_{gg}R_g$, i.e., the productivity indexed in similar sub-regions interacted with the states own research variable in measuring Semi-Direct Transfer or spill-in.

The terms R_gR_{gg} and R_gR_{gR} are then measuring Indirect Transfer.

Table 1 illustrates this model and shows the restrictions that one can place on the estimates. Note that in this case an extension variable is used. It shows that the direct spill-in is facilitated by extension investment. Note that the negative R_gR_{gg} term reflects the fact that nearby research can be a substitute for the state research. The R_gR_{gR} term indicates Indirect Transfer. The relevance of this model to the PROCISUR evaluation is that if we have measures of PROCISUR (and IARC) investments they can be included in the model probably in interactions with the Semi-Direct and Indirect measures, e.g., a variable PR x R_g x S_{gg} or PR x P_{gg} x R_g can be added to the model. It is possible that this model could then measure a PROCISUR impact. Clearly the more detailed the information on PROCISUR program by commodity and by country pair the better these prospects are.

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	APPROX PROB>|T|	VARIABLE LABEL
INTERCEPT	1	-28311.6	10434.82	-2.7132	0.0074	
ES	1	1.263202	0.033395	37.8258	0.0001	
ESS	1	-0.00222925	0.083906	-0.0266	0.9788	
ESR	1	0.135413	0.095469	1.4184	0.1581	
RS	1	-0.00114624	0.0005996964	-1.9114	0.0579	
RSS	1	-0.00100186	0.0003400123	-2.9465	0.0037	
RSR	1	-0.00137307	0.000352849	-3.8914	0.0001	
RSRSS	1	1.19055E-09	2.88968E-09	0.4120	0.6809	
RSRSR	1	1.73741E-08	1.25068E-08	1.3892	0.1668	
RSSRSR	1	1.27217E-08	5.69404E-09	2.2342	0.0269	
YEAR	1	14.359488	5.311725	2.7034	0.0077	

MODEL: D5
DEP VAR: PSESS

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	APPROX PROB>|T|	VARIABLE LABEL
INTERCEPT	1	-17435	11241.6	-1.5509	0.1230	
ES	1	0.123919	0.039627	3.1272	0.0021	
ESS	1	1.221819	0.087809	13.9145	0.0001	
ESR	1	0.110775	0.104531	1.0597	0.2910	
RS	1	-0.00107568	0.000661044	-1.6272	0.1058	
RSS	1	-0.00142376	0.0003666718	-3.8829	0.0002	
RSR	1	-0.000960725	0.0003886845	-2.4717	0.0146	
RSRSS	1	3.08035E-09	3.20776E-09	0.9603	0.3385	
RSRSR	1	1.35685E-08	1.35786E-08	0.9993	0.3193	
RSSRSR	1	1.75833E-08	5.90113E-09	2.9797	0.0034	
YEAR	1	8.821748	5.720926	1.5420	0.1252	

MODEL: D6
DEP VAR: PSRESS

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	APPROX PROB>|T|	VARIABLE LABEL
INTERCEPT	1	-20196.8	6890.995	-2.9309	0.0039	
ES	1	0.008034384	0.017803	0.4513	0.6524	
ESS	1	1.465118	0.058419	25.0797	0.0001	
ESR	1	-0.173139	0.069176	-2.5029	0.0134	
RS	1	-0.0026177	0.0003722965	-7.0312	0.0001	
RSS	1	-0.00106401	0.0002018683	-5.2708	0.0001	
RSR	1	-0.000915259	0.0002302712	-3.9747	0.0001	
RSRSS	1	1.82077E-08	1.83697E-09	9.9118	0.0001	
RSRSR	1	2.53882E-08	7.61423E-09	3.3343	0.0011	
RSSRSR	1	6.25970E-09	3.81692E-09	1.6400	0.1031	
YEAR	1	10.251459	3.507068	2.9231	0.0040	

SYSTEM: SYST001

T H I R D   S T A G E

MODEL: D1
DEP VAR: PS

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	APPROX PROB>|T|	VARIABLE LABEL
INTERCEPT	1	-60.472298	9.397875	-6.4347	0.0001	
B1.PSS	1	-0.158343	0.054748	-2.8922	0.0044	
B1.PSSES	1	0.0002638155	.00004534051	5.8185	0.0001	
ES	1	-.0000189687	.00006663497	-0.2847	0.7763	
RS	1	-5.03031E-07	1.99479E-07	-2.5217	0.0127	
RSRSS	1	-7.04328E-12	1.10971E-12	-6.3469	0.0001	
RSRSR	1	8.29306E-12	3.29974E-12	2.5132	0.0130	
YEAR	1	0.031295	0.004789917	6.5334	0.0001	

MODEL: D2
DEP VAR: PSS

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	APPROX PROB>|T|	VARIABLE LABEL
INTERCEPT	1	-59.940723	9.165481	-6.5398	0.0001	
B1.PS	1	-0.158343	0.054748	-2.8922	0.0044	
B1.PSESS	1	0.0002638155	.00004534051	5.8185	0.0001	
ESS	1	-.0000189687	.00006663497	-0.2847	0.7763	
RSS	1	-5.03031E-07	1.99479E-07	-2.5217	0.0127	
RSRSS	1	-7.04328E-12	1.10971E-12	-6.3469	0.0001	
RSSRSR	1	3.49660E-12	3.82852E-12	0.9133	0.3625	
YEAR	1	0.031037	0.004679619	6.6324	0.0001	

MODEL: D3
DEP VAR: PSR

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	APPROX PROB>|T|	VARIABLE LABEL
INTERCEPT	1	-38.777886	9.157986	-4.2343	0.0001	
B1.PSS	1	-0.158343	0.054748	-2.8922	0.0044	
B1.PSRESS	1	0.0002638155	.00004534051	5.8185	0.0001	
ESR	1	-.0000189687	.00006663497	-0.2847	0.7763	
RSR	1	-5.03031E-07	1.99479E-07	-2.5217	0.0127	
RSRSR	1	8.29306E-12	3.29974E-12	2.5132	0.0130	
RSSRSR	1	3.49660E-12	3.82852E-12	0.9133	0.3625	
YEAR	1	0.020278	0.004676828	4.3359	0.0001	

MODEL: D4
DEP VAR: PSSES

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	APPROX PROB>|T|	VARIABLE LABEL