Appendix C
DIFFERENCES BETWEEN TSP 4.4 AND EARLIER VERSIONS

This appendix lists the features added to TSP since the release of 4.3. These features include general changes, and enhancements to specific commands. The last section lists the incompatibilities with earlier versions.

C.1. General changes released in 4.4 only

Panel data -- You can specify global panel data IDs and information in options to the FREQ command, and share this information with subsequent PANEL, AR1(TSCS), and PRINT commands. TSP 4.4 handles missing values, lags, and leads correctly for such panel data, so you no longer have to check for missing values "by hand" for PANEL, or set up strange gapped SMPL commands for AR1(TSCS)! (See FREQ below.) PRINT prints panel data with a dual label.

Long variable names -- Names can be as long as 64 characters (the previous limit was 8 characters). This should help make TSP more compatible with databases like OECD which use such long names. It also helps with commands like DOT, FORM and DIFFER which can easily create names longer than 8 characters (in the past, such names were truncated arbitrarily to 8 characters). The long names can be imported/exported in spreadsheets and TSP databank files.

LMHET heteroskedasticity test -- This test uses a regression of squared residuals on squared fitted values, and it is computed for most models that lack RHS endogenous variables (OLSQ, PANEL, VAR, LSQ without instruments) (not for AR1, 2SLS, 3SLS, GMM, FIML).

P-values -- In estimation output, most statistics which have standard distributions will now have their P-values printed by default. This is most noticable for t-statistics, the Durbin-Watson statistic. Similarly, any regression diagnostics requested with the REGOPT command will include P-values by default (previously, the PVPRINT option was required to obtain them).

Numeric warning handling -- By default, only 10 numeric warnings (such as log(zero)) will be printed per command. Numeric warnings are only converted into "serious numeric errors" when analytic derivatives are being computed, or when the number of numeric warnings is over 100000. (See OPTIONS LIMWNUMC=n; below.)

Options error handling -- If you request an option which does not exist, a list of valid options will be printed.

Speed improvements -- Some significant speed gains have been obtained by optimizing the computation of quadratic forms and non-orthonormalized regressions:

- GMM is much faster for large models (due to quadratic forms). Similar speed gains apply to the MATRIX and ANALYZ commands.
- Regression diagnostics that involve auxillary regressions (such as LMAR, Durbin's h alt., White het. test, etc.) are done with non-orthonormalized regressions.
- Global OPTIONS FAST; command which makes all regression calculations skip orthonormalization. This is a simple (X'X)"(X'y) type calculation,where (X'X)" is computed with the Cholesky factorization. Without orthonormalization, OLSQ still yields about 7 correct significant digits in the Longley(1967) regression benchmark, which is about equal to the precision of the Longley data. The default TSP regression calculations (still) yield about 10 correct significant digits (when the Longley data is stored in double precision, and treated as if it has many significant zeros to the right of the input digits).
- Redundant orthonormalization removed from estimators with instruments (2SLS, LIML, 3SLS, GMM, AR1).
- 2SLS is computed in a partitioned manner (like LIML), which yields significant speed and memory gains, especially when there are relatively few RHS endogenous variables.

- Calculations for ROBUST standard errors are faster.

Equation printing -- PRINT, DBCOPY, DIFFER(PRINT), FORM(PRINT). TSP 4.4 prints the full function names, like CNORM(), instead of the 4-character abbreviations like CNRM() . (See also non-upward compatibilities below.)

C.2. General changes released in some copies of 4.3

Faster DO loops and PROCs (especially for long programs). These speed gains were about 25% for some long programs.

Memory usage by PROCs and changing SMPLs improved.

New nonlinear options:

- HITER=F and HCOV=F for BFGS method (for use with numeric or analyticfirst derivatives)
- GRAD = ANALYTIC or C2 or C4 etc. method of obtaining first derivatives
- TOLS = value convergence tolerance on squeezed parameter changes. The default is zero, i.e. only perform convergence test on full change vector, not on squeezed version.

Nonlinear (numeric gradients). If there is a numerical warning in computing the numeric gradient, try a smaller EPS (up to 10 times).

Subscripted list variables. Examples: (See also the LIST(DROP) command below.)

```
LIST X A B C;
PRINT X(2);                    ? same as print b;
LIST X(3) = D;          ? same as list x a b d; (replace 3rd list item)
```

PDL variables -- adjusts estimation sample to handle missing values, when they are due to initial observations for lags.

CNORM() and CNORMI() functions improved to 14 digit accuracy. Before, in worst cases, they were accurate to about 7 digits and 4 digits respectively. P-values for t-statistics are more accurate for degrees of freedom larger than 500.

Nonlinear/Monte Carlo -- Don't print any non-convergence warning messages when the SILENT option is on and all other output is being SUPRESed by the user (SUPRES LOGL NOB COEF;). The user should be checking the @IFCONV variable in such a situation.

Support subscripted matrices in FRMLs for LSQ, FIML, ML, etc. I.e. treat like a scalar CONST.

C.3. Command-specific improvements released in 4.4 only

CDF -- new options for finite sample critical values (Dickey-Fuller)

- NOB = number of observations in regression (DICKEYF) same as the old DF option.
- NLAGS = number of augmenting lags in regression -- The Cheung and Lai BJES 7/95 response surface is used to calculate finite sample critical values for the augmented Dickey-Fuller test. If NVAR>1, the MacKinnon 1991 response surface is used to calculate critical values for the Engle-Granger test. P-values are interpolated from these critical values (at 5%, and at 1% or 10%).

COINT(FINITE) calculates finite sample P-values (for Dickey-Fuller and Engle-Granger, as described above under CDF). These are usually slightly larger than the default asymptotic P-values. The P-values are labelled either P-valFin or P-valAsy, so you can tell which tests are using the finite sample versions.

FORM(VARPREF=prefix) allows coefficient names to be based on the variable names. For example:

```
FORM(VARPREF=B) EQ Y C X Z ;        creates          FRML EQ Y = B0 + BX*X + BZ*Z;
```

FREQ -- New global options for panel data:

- ID = identification series (useful for unbalance or balanced data)
- T = number of time periods (useful for balanced data)
- N = number of individuals (OK for balanced data, or to check total number of individuals calculated from ID series)
- TIME = series which specifies individual time periods for each observation (coded as 1 for first period, 2 for second, etc.) This is only used for printing at present.
- START = date for period 1. Used for printing. Can be used in conjunction with a time series frequency, so that panel data can have an internal FREQ A, Q, or M, etc. These FREQ options simply specify a way to interpret standard FREQ N series as panel data. They can be used with existing databanks. In TSP 5.0 we hope to provide capabilities to automatically merge regular time series variables with panel series.

KEEP procname; give an error message for this, since PROCs can't be stored in databanks.

LIST(DROP) old_list args; drops arguments from an existing list.

LOGIT (multinomial) has a Kullback-Leibler R-squared. This is the same as McFadden's pseudo R-squared, or the likelihood ratio index (for multinomial logit and binary probit).

MATRIX EIGVAL() and EIGVEC() now output sorted eigenvalues. Real eigenvalues are simply sorted (from high to low), while complex eigenvalues are sorted by their norm.

MMAKE(VERT) stacks series or matrices vertically instead of horizontally.

OLSQ has some additional default regression diagnostics (besides LMHET):

- RESET2 - Ramsey's RESET test of order two; checks for missing quadratic terms on the RHS
- JB - Jarque-Bera normality test on residuals

OPTIONS FAST; performs regression calculations without orthonormaliztion. Mentioned above under speed improvements. Should provide the largest speed gains when there are over 1000 observations, or in Monte Carlo loops.

OPTIONS LIMWNUMC=n; limits the number of warning messages for numeric errors in any particular command. The default value is 10. This means that each command will produce at most 10 numeric warning messages, and then the remainder for that command will be suppressed.

OPTIONS SECONDS=2.1; when a fractional number of seconds is given, within-command timings for regression calculations will be reported. Such within-command profile timings will eventually be extended to commands like ML, GMM, and MATRIX to investigate which parts of certain commands are slow, for use in improving speed. See also the DATE command below for timing across several commands.

PANEL

- Besides LMHET, the Durbin-Watson is now computed (with BOUNDS P-value), correctly allowing for gaps between individuals.
- The R-squared and residuals for VARCOMP are now based on the original dependent variable (rather than on the transformed dependent variable). As a side effect, the mean and std.dev. of the dependent variable are now printed (also true for WITHIN).

PLOT on DOS/Win, Mac, allow arbitrary missing values in hi-res PLOT (including SMPL gaps, some series longer than others, etc.)

PLOT/GRAPH on DOS/Win:

- A4 option supports A4 paper size (originally for LJ3 drivers only, but also for PostScript drivers in 4.4). Can patch this option permanently with the PATCH.EXE program.
- A long time bug in hardcopy graphics was fixed, which had sometimes made small (1" x 2") plots, or refused to make more than one hardcopy during a run.

PROBIT has a Kullback-Leibler R-squared. See LOGIT above.

REGOPT

- New test for OLSQ (besides LMHET and RESET mentioned above): SWILK - Shapiro-Wilk normality test (has good small sample performance) With P-value, of course. This involves sorting the residuals, so it can be rather slow in very large samples.
- New options:
  - DWPVALUE = APPROX or BOUNDS or EXACT -- the method of computing the Durbin-Watson P-value. The default depends on the current FREQ.
  - RESETORD = order of RESET test (default 2)

C.4. Command-specific improvements released in some copies of 4.3

AR1 stores @IFCONV

BJEST

- EXACTML option for exact ML estimation (vs. conditional ML which is the default). Only imposes invertibility restriction for NMA=1; could be improved to impose stationarity and invertibility for all model orders. This would be helpful, since it is when the parameters are close to these constraints that the EXACTML estimates are most advantageous. Doesn't support NSMA= and NSAR= . Still useful for many ARMA models, though.

- Computes stationarity (@ARSTAT) and invertibitilty (@MAINV) checks for final (or input) parameter estimates.

- Store @QSTAT, %QSTAT; use missing instead of zero for first NAR+NMA P-values.

- Use @START for starting values

BJFRCST

- EXP option forecasts Y, after you have used BJEST to fit log(Y). Includes the EXP(sigmasq(t)/2) term in the forecasts to make them unbiased.

- Store @FITSE (SE for fitted value). @FITSE is for log version of model when EXP option is used, since bounds are not symmetric around the forecast in this case.

BJIDENT -- In autocorrelation plots, print number of lags and values of autocorrelations.

DATE seconds; If you supply an argument to the DATE command, it will store the number of seconds since midnight. This is useful for timing the execution of a single command or series of commands. I.e. use DATE before and after, and take the difference to see the elapsed time.

DIFFER

DEPVARPR=prefix option allows for a different dependent variable name pattern than just Dyn. Example:

FRML R E = Y - X*B; DIFFER(DEP=Q) R B; creates FRML DR1 Q1 = -X;

Default unnormalized output frml if input frml is unnormalized. This is helpful to prevent having to create separate names for each FRML and dependent variable name. Example:

FRML E Y - X*B; DIFFER E B; creates FRML DE1 -X;

EQSUB works with lagged endogenous variables. The best way to see this is by example. For a conditional AR(1) model:

```
FRML U Y - (B0 + B1*X);
FRML E U - RHO*U(-1);
EQSUB E U;
```

This will correctly lag y and x when they enter the e equation via u(-1). It requires that all variables in the u equation be defined (as series or params, etc., when the eqsub command is executed, so that eqsub will know which variables to lag).

FIND allow search for abbreviated command, like FIND OLS

FORCST after AR1: for (default) DYNAM option, look backwards for presample data, to calculate a presample residual. Use it with a power of RHO to help forecast the first and subsequent observations. Previously, the presample residual was treated as zero, and this results in a forecast that ignores RHO; i.e. yhat = X*B. This can still occur if a complete presample observation is not available.

FORM

- RESIDUAL option for unnormalized Y - (A+BX) type eqn. Example:
  OLSQ Y C X; FORM(RESID,PREFIX=B) EQ1; creates FRML EQ1 Y - (B0 + B1*X);
- CONST option with list of equations . Converts all scalar variables (CONST, PARAM) to their current values. Useful for printing out an equation, or for storing one permanently in a databank, without having to worry about storing the associated PARAMs.

FREQ variablename; , store @FREQ . These features are useful for saving and restoring the FREQ in a PROC.

GMM(NOOPTCOV,COVOC=mat,INST=...) computes correct VCOV for "one-step" estimator, using Hansen Theorem 3.1, i.e. doesn't assume that the user-supplied COVOC is "optimal". Default: NOOPTCOV.

LAD(QUANTILE=q) now produces correct standard errors for the case when q takes values other than .5 . These assume the Laplace distribution for the errors.

LIST(PREFIX=chars) can be used to make non-numeric lists. Example:

```
LIST X A B D;
LIST(PREFIX=E) EQS X;                    ? same as   list eqs ea eb ed;
```

ML PROC -- ML estimation where the LOGL is calculated by a Proc instead of in a single FRML. See the TSP web page www.tspintl.com for several examples. It has potential uses in recursive time series (GARCH), hyperparameter estimation for KALMAN, simulation models like multinomial probit, and even simple things like concentrated log likelihood functions. The syntax is ML(options) Procname list_of_params ; Procname computes @LOGL as a function of the params. Numeric first derivatives are used, so HITER=F,D, or B are possible. For HITER=B or HCOV=B, Procname must also store the series LOGL, so that numeric derivatives at each observation can be computed. Procname can include other iterative estimators, such as SOLVE (useful for recursive evaluation) or even ML. The Proc makes it fairly easy to impose inequality constraints on the input parameters, or on functions of them, such as h(t) in GARCH models. This is because a separate command can be used to check each constraint, and set @LOGL to a penalty value (@MISS) if the constraint is violated; this can't be done easily in the single FRML approach of standard ML.

OPTIONS APPEND; close and reopen output file during each nonlinear iteration. Only works on some systems, like Sun, and this feature allows you to scan the tail of the output file for iteration progress.

PANEL(ROBUST) heteroskedastic-robust SEs for TOTAL and WITHIN

C.5. Non-upward compatibilities

Long names that were automatically truncated to 8 characters in the past will no longer be truncated. This would only be a problem if you have programs which make explicit reference to the truncated versions of these names.

Most Durbin-Watson calculations now ignore SMPL gaps. This allows them to use the standard tables for determining the P-value. The exception is that gaps between individuals in Panel data (PANEL and AR1(TSCS) commands) -- they are considered significant.

PANEL(TIME=period_variable) is no longer considered sufficient to identify the separation between individuals. This

is because one individual might have time periods 1,2,3 and the next individual might have time periods 4,5,6 . PANEL(ID=series) and PANEL(T=value) are upward compatible.

Certain regression diagnostics have been eliminated in cases where they are inappropriate (inconsistent). This includes the Durbin-Watson when there are lagged dependent variables (Durbin's h and Durbin's h alternative are already printed in this case). It also includes some bogus F-statistics generated by nonlinear commands like BJEST and 2SLS.

Several undocumented alternative function names are no longer supported. This means users can define variables like CNRM, DLCN, and LG10 without being given warning messages about reserved function names.

Function	Old Undocumented Synonym(s)
LOG	ALOG
CNORMI	CNMI
CNORM	CNRM
NORM	NORMAL
LCNORM	LCNM
DLCNORM	DLCN
LNORM	LNRM
GAMFN	GAMF
LGAMFN	LGAM
DIGAMMA	DIGAMF, DIGAMFN, DIGM (note: DLGAMFN *is* supported)
TRIGAMMA	TRIGAMFN, TRIGAM, TRGM
LFACT	LFAC
LOG10	LG10
POS	POSITIVE
INT	FLOOR, TRUNC
CEIL	CEILING
ROUND	ROUN
MISS	MISSING

(One) documented function synonym that is still supported:

Function	Synonym
DIGAMMA	DLGAMFN