SAMPSEL (非線形オプション) probit 従属変数 probit 独立変数
| 回帰従属変数 回帰独立変数 ;

機能:

SAMPSEL は一般化 Tobit モデルあるいはサンプルセレクションモデルを推定します. 回帰変数と, 選択を予測する潜伏変数の両方とも外生変数の線形回帰関数になります. (選択されない観測値に対しては観測しない回帰変数の) センサー化か, (観測されない全ての変数の) 切捨てモデルの一方を推定できます.

使用法:

推定されるモデルは Amemiya(雨宮) で書かれている Tobit タイプ II モデル, あるいは Maddala (参考文献を参照) で書かれている確率的閾値を持ったセンサー化回帰モデルです. モデルは次のように書くことができます,

$$
\begin{aligned}
y_{2i} &= x_{2i}b + e_i \qquad y_{1i} > 0 \quad (\text{回帰方程式}) \\
&= 0 \quad \text{その他} \\
y_{1i} &= x_{1i}d + u_{1i} \qquad (\mathit{probit}\,\text{方程式})
\end{aligned}
$$

e_i と u_i は分散が σ^2 と 1(正規化の仮定) で相関係数が ρ の同時正規分布をしていると仮定されています.

このモデルの推定にプロシジャーを使うには, 観測値が観測されたかされないか ($y_{1i} > 0$) を表す 0/1 変数の名前を probit 従属変数とし, 回帰因子 x_{1i} を probit 独立変数, y_{2i} を回帰従属変数, そして x_{1i} を回帰独立変数として与えます. 回帰変数に対する欠測値は $D(y_{1i} > 0)$ がゼロの観測値に対しては許されます.

もし $D(y_{1i} > 0)$ が常に 1 なら, 切り捨てられた (条件付き) モデルが推定されます (Bloom and Killingsworth). これは, “LATENT SELECTION VARIABLE” と言うメッセージが表れます. HVOV=B は切捨てされた場合に得られる唯一の標準誤差です. Probit 方程式での定数項以外の変数についての識別条件はチェックしません.

オプション:

標準的な非線形オプション– NONLINEAR を参照して下さい.

HITER=B, HCOV=N が既定値 (切捨てモデルが推定される時を除く – 上記参照).

例:

```
SAMPSEL (PRINT, MAXIT=50, HCOV=NBW) IY C Z | Y C X ;
```

アウトプット:

SAMPSEL のアウトプットは方程式タイトルに続いて従属変数名が出力されます. 初期値とイタレーション中の診断結果がプリントされ, 最終の収束状態がプリントされます.

続いて従属変数の平均, 正の観測値数, 残差平方和, R^2, 右辺の変数名の表, 係数推定値, 標準誤差とその t-値が出力されます.

SAMPSEL はこれらの結果のいくつかを後で利用できるように, データ領域にも保存します. 以下の表は SAMPSEL コマンドの後で利用できる結果のリストです.

変数名	タイプ	長さ	変数の記述
@LHV	リスト	1	従属変数名
@RNMS	リスト	変数の数	右辺の変数名
@YMEAN	スカラー	1	従属変数の平均
@NOB	スカラー	1	観測値数
@NPOS	スカラー	1	正の観測値数
@SSR	スカラー	1	残差平方和
@RSQ	スカラー	1	R^2
@LOGL	スカラー	1	対数尤度関数
@IFCONV	スカラー	1	収束すれば 1, その他は 0
@NCOEF	スカラー	1	右辺の変数の数
@NCID	スカラー	1	識別された係数の数
@COEF	ベクトル	変数の数	係数推定値
@SES	ベクトル	変数の数	標準偏差
@T	ベクトル	変数の数	t 統計値
@GRAD	ベクトル	変数の数	収束時における log L の勾配
@VCOV	行列	変数の数×変数の数	係数推定値の分散共分散
@DPDX	行列	変数の数×2	平均確率微係数
@FIT	系列	観測値数	推定された確率
@RES	系列	観測値数	残差
@MILLS	系列	観測値数	逆ミルス比

モデルが PDL あるいは SDL 変数を含んでいる場合には, 次のものも保存されます.

@SLAG	スカラー	1	ラグ係数の和
@MLAG	スカラー	1	ラグ係数の平均
@LAGF	ベクトル	ラグの数	『分解』後の推定されたラグ係数

回帰方程式に対する@SESは, 観測された標本だけに対して保存されます.

方法:

使用した方法は, HITER オプションで与えられるヘッシアン近似をつかった勾配法によって尤度関数の最大化をおこないます. この尤度関数はしばしば複数個の局所最大化があることが知られているので, Nawataの方法 (1994, 1995, 1996) を用いて大域最大化を行います. RHO についての格子探索を行い, 局所最大値を見つけます. 使用する (27) 個の格子点は, `0 .1 .2 ... .8 .85 .9 .95 .99 .9999, -.1 -.2 ... -.9999` です. 各格子点で, RHO は固定したままで, $\log L$ は他のパラメータに関して最大化されます. 各局所最大値に対して, 最後のイタレーションでは RHO が動いても構いません. これらのイタレーションは精度が良く, 最大値になるように修正され, 大域的な最大値を選択することを可能にします.

しばしば大域的最大値が RHO=1.0000 あるいは -1.0000 であることを示します. これらの場合には, RHO の実際の推定値は絶対値で僅かに 1 より小さく, 残差共分散行列は特異に近くなります. RHO の標準誤差と他の変数との共分散は, この場合はゼロと置かれます. この場合にモデルをどう解釈するかは明らかではありません. プロビット方程式は尤度関数が大きいときこのようになり, 定数項が爆発するように見えます. しばしば, もっと多くの変数を回帰式に加えてその当てはまりをよくすると, RHO の推定値は極端にはなりません.

参考文献:

Amemiya, takeshi, **Advanced Econometrics,** Harvard University Press, Cambridge, Massachusetts, 1985, Chapter 13.

Bloom, David E., and Mark R. Killingsworth,"Correcting for truncation Bias caused by a Latent Truncation Variable", **Journal of Econometrics,** 1985, pp.131-135.

Griliches, Z., B. H. Hall, and J. A. Hausman, "Missing Data and Self- selection in Large Panels", **Annales de l'Insee,** Avril-Sept 1978, pp.137-176.

Maddala, G. S., **Limited-Dependent and Qualitative Variables in Econometrics,** Cambridge University Press, Cambridge, 1983, Chapter 6.

Nawata, Kazumitsu, "Estimation of Sample Selection Bias Models by the Maxmum Likelihood Estimator and Heckman's two-step Estimator," **Economics Letters,** 45, 1994, pp. 33-40.

Nawata, Kazumitsu, "Estimation of Sample Selection Models by the Maxmum Likelihood Method," **Mathematics and Computers in Simulation,** 39, 1995, pp. 299-303.

Nawata, Kazumitsu, and Nagase Nobuko,"Estimation of Sample Selection Bias Models," **Econometric Reviews,** 15, 1996, pp. 387-400.

Olsen, R.J., "Distributional Tests for Selectivity Bias and a More Robust Likelihood Estimator," **International Economic Review** 23, 1982, pp. 223-240.