# EC475 Problem set 6

## Semiparametric Analysis of LDV Models

Antoine Goujard

10/12/09

- Var of interest: the propensity of a country $i$ to experience repayments problems in period $t$.
- Filtered information about this propensity:
  1. a censored variable if we observe the external debt level of the country;
  2. a binary variable, if an indication of financial difficulties is observed or not.
- In **(1)** we assume that there is a latent (panel) linear model:
  $y_{it}^* = x_{it}'\beta + \varepsilon_{it}$
  $y_{it} = max(0, y_{it}^*) = y_{it}^*.1_{y_{it}^*>0}$ is the observed variable.
- We disregard the panel structure of the data and assume $(\varepsilon_{it})$ iid over $i$ and $t$.

- Assume $\varepsilon|X \sim \mathcal{N}_S(0, \sigma^2.I_S)$, normal, iid and homoscedastic disturbance term.
- Then we can work out the MLE estimator.
- $P(y_{it} = 0|x_{it}) = P(y^*_{it} \leq 0|x_{it}) = P(\varepsilon^*_{it} \leq -x'_{it}\beta|x_{it}) = \Phi(\frac{-x'_{it}\beta}{\sigma})$
- For $y > 0$, $f_{y_{it}|x_{it}}(y) = f_{y^*_{it}|x_{it}}(y) = \frac{1}{\sigma}\phi(\frac{y-x'_{it}\beta}{\sigma})$
- Thus $f_{y_{it}|x_{it}}(y) = \Phi(\frac{-x'_{it}\beta}{\sigma}).1_{y=0} + \frac{1}{\sigma}\phi(\frac{y-x'_{it}\beta}{\sigma}).1_{y>0}$
- $LL(\beta, \sigma) = \sum_{i,t} ln(f_{y_{it}|x_{it}}(y))$
- This *LL* can be made strictly concave by considering a change of variables $b = \beta/\sigma$ and $s = 1/\sigma$ and we can use the usual maximization methods **NR** and **BHHH** as *LL* is $C^2(\mathbb{R}^{k+1}, \mathbb{R})$ in the parameters.
- The MLE estimator is based on normality and homoscedasticity. And if one of them fails, the estimator becomes inconsistent.

- Powell (1984) proposes an alternative **LAD=Least Absolute Deviations** estimator.
- Powell (1984) assumes $Med(\varepsilon_{it}|x_{it}) = 0$
- $Med(X) = inf_x\{P(X \leq x) \geq 0.5\}$.
- Then, $m_\beta(x_{it}) = Med(y^*_{it}|x_{it}) = x'_{it}\beta$.
- And $\beta = argmin_b E(|y^*_{it} - x'_{it}b||x_{it})$.
- Thus Powell shows that if the regressors have some minimal properties:
  $\beta = argmin_b E(|y^*_{it} - x'_{it}b|)$.
- And if $y^*_{it}$ is observed it is natural to estimate $\beta$ using:
  $\hat{\beta} = argmin_b \frac{1}{S} \sum_{i,t} |y^*_{it} - x'_{it}b|$
- But if $g()$ is a non-decreasing function then it can be shown that:
  $\hat{\beta} = argmin_b \frac{1}{S} \sum_{i,t} |g(y^*_{it}) - g(x'_{it}b)|$
  because $Med(g(z)) = g(med(z))$.
- Here, $g(z) = max(0, z)$, non decreasing in $z$.

- Problem, due to absolute value, the objective function is not everywhere differentiable wrt $b$ ($g$ has a Kink at 0).
- We can not rely on the usual optimization algorithms based on derivatives, gradient and Hessian to find $\hat{\beta}_{LAD}$.
- `clad` in `STATA` implements Powell estimator using Buchinsky's iterative linear programming algorithm (1994).
- This does not require the computation of derivatives. As the simplex algorithm (Nelder-Mead$=$downhill simplex method$=$amoeba) that we have seen last week.
- This estimator is robust to heteroscedasticity and departure from the normal distribution.
- The usual solution to compute the ses of $\hat{\beta}_{LAD}$ is the bootstrap. (Powell -1984- shows that under some conditions, the LAD estimator is consistent and AN-CAN-, but the asymptotic variance depends on the true law of $\varepsilon|X$).

Descriptive statistics for two violations of MLE assumptions:

1. **cryhet** normal but heteroscedastic error term ;
2. **cryevd** extreme value error term ($F(u) = e^{-e^{-u/\sigma}}$).

| Var | Mean | SD | Min | Max |
|---|---|---|---|---|
| cry | 1.801 | 1.927 | 0 | 8.497 |
| cryhet | 1.865 | 2.15 | 0 | 13.376 |
| cryevd | 1.604 | 2.147 | 0 | 11.808 |
| one | 1 | 0 | 1 | 1 |
| debtlxl | 1.004 | 0.871 | 0.052 | 8.481 |
| reslimpl | 0.302 | 0.254 | 0.009 | 1.429 |
| N | 600 | 600 | 600 | 600 |

Censored variable, heteroscedastic case.

| | cry | | | cryhet | | |
|---|---|---|---|---|---|---|
| | OLS | MLE | LAD | OLS | MLE | LAD |
| Var | Est/(se) | Est/(se) | Est/(se) | Est/(se) | Est/(se) | Est/(se) |
| debtlxl | 0.137 | 0.235 | 0.241 | 0.174 | 0.268 | 0.235 |
| | (0.092) | (0.128) | (.249) | (0.103) | (0.134) | (0.140) |
| reslimpl | -0.648 | -0.855 | -0.750 | -0.402 | -0.665 | -0.528 |
| | (0.315) | (0.446) | (0.667) | (0.353) | (0.462) | (0.332) |
| Intercept | 1.860 | 1.308 | 1.199 | 1.812 | 1.359 | 1.096 |
| | (0.166) | (0.234) | (0.453) | (0.186) | (0.242) | (0.239) |

LAD's ses computed with 50 replications.

- In the ideal case (`cry`), the MLE point estimates are quite close to the CLAD estimates and have smaller ses. OLS slopes' estimates appear biased towards 0.
- Comparing `cry` and `cryhet`, the inconsistency of the MLE estimator does not appear obvious. Given the large ses, it is difficult to distinguish the 2 cases.

- For OLS, MLE and CLAD, the main change concerns `reslimpl` whose estimates shift towards 0, the other point estimates are more or less unaffected.

## Censored variable, extreme value case.

| | cry | | | cryevd | | |
|---|---|---|---|---|---|---|
| | OLS | MLE | LAD | OLS | MLE | LAD |
| Var | Est/(se) | Est/(se) | Est/(se) | Est/(se) | Est/(se) | Est/(se) |
| debtlxl | 0.137 | 0.235 | 0.241 | 0.117 | 0.300 | 0.250 |
| | (0.092) | (0.128) | (.249) | (0.103) | (0.154) | (0.177) |
| reslimpl | -0.648 | -0.855 | -0.750 | -0.680 | -1.076 | -0.995 |
| | (0.315) | (0.446) | (0.667) | (0.352) | (0.548) | (0.895) |
| Intercept | 1.860 | 1.308 | 1.199 | 1.693 | 0.788 | 0.648 |
| | (0.166) | (0.234) | (0.453) | (0.185) | (0.286) | (0.426) |

LAD's ses computed with 50 replications.

- Here there are much more change between the MLE on `cry` and `cryevd`.
- Except for the `intercept` the LAD point estimates using `cryevd` are close to the point estimates of `cry` (MLE).
- As expected, CLAD appears less sensible than MLE to the distribution assumption.

- The same types of argument apply to the binary case:
  $y_{it}^* = x_{it}'\beta + \varepsilon_{it}$
  $y_{it} = 1_{y_{it}^*>0}$ is the observed variable.
- The MLE is based on strong distributional assumptions, normality and homoscedasticity in the case of the probit.
- Manski (1975,1984,1988) proposes a LAD estimator in the same spirit as Powell (1984) estimator. This estimator is called the Maximum Score estimator.
- Manski (1988) assumes $Med(\varepsilon_{it}|x_{it}) = 0$ and other regularity conditions.
- If $y_{it}^*$ is observed it is natural to estimate $\beta$ using:
  $\hat{\beta} = argmin_b \frac{1}{S} \sum_{i,t} |y_{it}^* - x_{it}'b|$
- Here we can define $g(z) = 1_{z>0} = sgn(z)$ non decreasing in $z$.
- We get $\hat{\beta}_{MS} = argmin_b \frac{1}{S} \sum_{i,t} |y_{it} - 1_{x_{it}'b>0}|$

- We can rewrite:
  $\hat{\beta}_{MS} = argmax_{b:||b||=1} \sum_{i,t}(y_{it}.1_{x'_{it}b>0} + (1 - y_{it}).1_{x'_{it}b\leq 0}) \equiv S_1(b)$
  $\hat{\beta}_{MS} = argmax_{b:||b||=1} \sum_{i,t}[(2.y_{it} - 1).1_{x'_{it}b>0}] \equiv S_2(b)$
- $S_1(b)$ is the initial definition of the maximum score estimator (Manski, 1975, 1984).
- The objective function, $S_1$, is the **score**. The score is the number of correct predictions we would make if we predicted $y_{it}$ to be 1 whenever $x'_{it}b > 0$ and 0 otherwise.
- Maximization of the non-continuous score is (very) difficult. See Vassilis' handout 8 (p.95, 101, 108). The gradient based methods fail and the Nelder-Mead algorithm does not provide good results.
- Inference in this model is complicated. Manski (1984) proves consistency. Kim & Pollard (1990) show that the estimator has a non normal asymptotic distribution (reached at a slow rate). Abrevaya & Huang (2005) show that the usual bootstrap is not consistent in this case. Delgado et al. (2001) use sub-sampling to evaluate the distribution of the estimator.

Descriptive statistics for two violations of MLE assumptions:

1. **bcyhet** normal but heteroscedastic error term ;
2. **bcyevd** extreme value error term ($F(u) = e^{-e^{-u/\sigma}}$).

| Var | Mean | SD | Min | Max |
|---|---|---|---|---|
| bcy | 0.668 | 0.471 | 0 | 1 |
| bcyhet | 0.692 | 0.462 | 0 | 1 |
| bcyevd | 0.635 | 0.482 | 0 | 1 |
| one | 1 | 0 | 1 | 1 |
| debtlxl | 1.004 | 0.871 | 0.052 | 8.481 |
| reslimpl | 0.302 | 0.254 | 0.009 | 1.429 |
| N | 600 | 600 | 600 | 600 |

## Heteroscedastic case

| | bcy | | | bcyhet | | |
|---|---|---|---|---|---|---|
| | Probit | Logit | MS | Probit | Logit | MS |
| debtlxl | 0.247 | 0.448 | 0.143 | 0.231 | 0.440 | 0.125 |
| | (0.077) | (0.139) | | (0.078) | (0.145) | |
| reslimpl | -2.367 | -3.989 | -0.939 | -2.520 | -4.399 | -0.939 |
| | (0.276) | (0.490) | | (0.275) | (0.513) | |
| Intercept | 0.935 | 1.531 | 0.312 | 1.079 | 1.802 | 0.317 |
| | (0.130) | (0.225) | | (0.133) | (0.236) | |

Score maximization using `STATA` is difficult. `optimize` in `mata` has a Nelder-Mead algorithm. The algorithm converges but does not seem to reach the true maximum. So I used random search on the unit hyper-sphere.

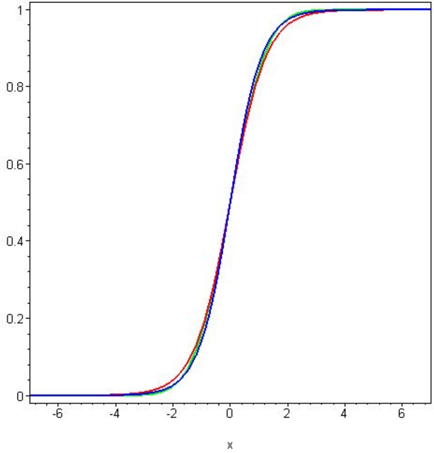
## Extreme value case

| | bcy | | | bcyevd | | |
|---|---|---|---|---|---|---|
| | Probit | Logit | MS | Probit | Logit | MS |
| debtlxl | 0.247 | 0.448 | 0.143 | 0.340 | 0.593 | 0.118 |
| | (0.077) | (0.139) | | (0.081) | (0.143) | |
| reslimpl | -2.367 | -3.989 | -0.939 | -2.489 | -4.239 | -0.946 |
| | (0.276) | (0.490) | | (0.286) | (0.513) | |
| Intercept | 0.935 | 1.531 | 0.312 | 0.782 | 1.293 | 0.293 |
| | (0.130) | (0.225) | | (0.130) | (0.223) | |

Score maximization using `STATA` is difficult. `optimize` in `mata` has a Nelder-Mead algorithm. The algorithm converges but does not seem to reach the true maximum. So I used random search on the unit hyper-sphere.

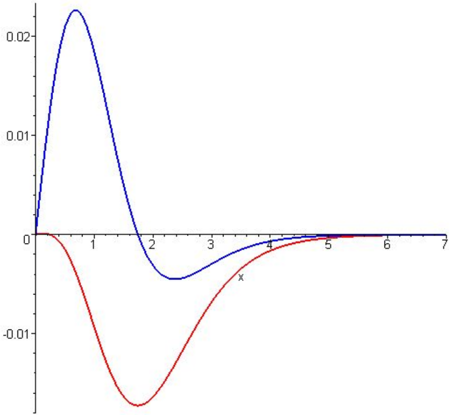
- The different set of rough point estimates are not easy to compare.
- A first idea is to apply Amemiya's rule $\hat{\beta}_L \simeq 1.6\hat{\beta}_P$ to compare logit and probit estimates, or even better to compare only the marginal effects, but this will not work for the maximum score estimator.
- A second idea is to rescale the estimates and compare $\hat{\beta}/||\hat{\beta}||$.
- The key is to remember that all our binary outcomes models are identified up to a treshold and a scale parameters:
  $y_i^* = x_i'\beta + \varepsilon_i$ with $\varepsilon_i | x_i \sim (0, \sigma^2)$.
  $y_i = 1_{x_i'\beta + \varepsilon_i > \tau}$ is the observed variable.
- Then we can not estimate $\tau$ and $\beta_1$ the parameter for the constant:
  $1_{\beta_1 + x_{i-1}'\beta_{-1} + \varepsilon_i > \tau} = 1_{\beta_1 - \tau + x_{i-1}'\beta_{-1} + \varepsilon_i > 0}$
- So we set $\tau = 0$, but then we can not estimate $\beta$ and $\sigma > 0$:
  $1_{x_i'\beta + \varepsilon_i > 0} = 1_{\frac{x_i'\beta + \varepsilon_i}{\sigma} > 0}$
  So in the probit case, we set $\sigma^2 = 1$ and in the logit case, we set $\sigma^2 = \pi^2/3$. In other words, from the latent model, we are only able to identify, $\frac{\beta_1 - \tau}{\sigma}$ and $\frac{\beta_{-1}}{\sigma}$

## Amemiya's rule $\hat{\beta}_L \simeq 1.6\hat{\beta}_P$ to compare logit and probit estimates




Normal cdf, Modified logistic cdf ($F_\lambda(x) = \frac{1}{1+e^{-x.\lambda}}$) with $\lambda = \pi/\sqrt{3}$, with $\lambda = 1.6$.

# Amemiya's rule $\hat{\beta}_L \simeq 1.6\hat{\beta}_P$, difference between modified logistics and standard normal cdf




$\frac{1}{1+e^{-x.\lambda}} - \Phi(x)$ with $\lambda = \pi/\sqrt{3}$, with $\lambda = 1.6$.

Rescaled parameters' estimates ($\hat{\beta}/||\hat{\beta}||$)

| | bcy | | | bcyhet | | |
|---|---|---|---|---|---|---|
| | Probit | Logit | MS | Probit | Logit | MS |
| debtlxl | 0.097 | 0.104 | 0.138 | 0.090 | 0.094 | 0.136 |
| reslimpl | -0.926 | -0.928 | -0.938 | -0.861 | -0.836 | -0.939 |
| intercept | 0.366 | 0.356 | 0.315 | 0.340 | 0.321 | 0.315 |
| | bcy | | | bcyevd | | |
| | Probit | Logit | MS | Probit | Logit | MS |
| debtlxl | 0.097 | 0.104 | 0.138 | 0.094 | 0.100 | 0.110 |
| reslimpl | -0.926 | -0.928 | -0.938 | -0.900 | -0.892 | -0.947 |
| intercept | 0.366 | 0.356 | 0.315 | 0.355 | 0.342 | 0.297 |