ECON 899b - Problem Set 2

Due Monday, Nov. 22nd

Consider the following dynamic model of loan repayment. Let $T_i \in \{1,2,3,4\}$ denotes the observed loan duration, and Y_{it} denotes an indicator variable equal to one if the loan is pre-paid at the end of period t. The relationship between the two variables is given by:

$$T_i = \begin{cases} 1 & \text{If } Y_{i0} = 1 \\ 2 & \text{If } Y_{i0} = 0 \text{ and } Y_{i1} = 1 \\ 3 & \text{If } Y_{i0} = Y_{i1} = 0 \text{ and } Y_{i2} = 1 \\ 4 & \text{If } Y_{i0} = Y_{i1} = Y_{i2} = 0 \end{cases} \tag{1}$$

At each period t, a loan is repaid if:

$$Y_{it} = 1 \quad \text{if } \alpha_t + X_i\beta + Z_{it}\gamma + \epsilon_{it} > 0$$

The vector X_i is a time-invariant vector borrower characteristics, Z_{it} is a vector of time varying characteristics (i.e. FICO score), and $\epsilon_{it} = \rho\epsilon_{it-1} + \eta_{it}$ if $t > 1$ and $\eta_{it} \sim N(0,1)$. The initial residual is drawn from the stationary distribution: $\epsilon_{i0} \sim N(0, \sigma_0^2)$ where $\sigma_0^2 = \frac{1}{(1-\rho)^2}$.

The likelihood associated with duration T_i is given by:

$$\begin{aligned}
\Pr(T_i|X_i, Z_i, \theta) &= \begin{cases} \Pr\left(\epsilon_{i0} < -\alpha_0 - X_i\beta - Z_{it}\gamma\right) & \text{If } T_i = 1 \\ \Pr\left(\epsilon_{i0} < \alpha_0 + X_i\beta + Z_{it}\gamma, \eta_{i1} < -\alpha_1 - X_i\beta - Z_{it}\gamma - \rho\epsilon_{i0}\right) & \text{If } T_i = 2 \\ \Pr\left(\epsilon_{i0} < \alpha_0 + X_i\beta + Z_{it}\gamma, \epsilon_{i1} < \alpha_1 + X_i\beta + Z_{it}\gamma, \epsilon_{i2} < -\alpha_2 - X_i\beta - Z_{it}\gamma\right) & \text{If } T_i = 3 \\ \Pr\left(\epsilon_{i0} < \alpha_0 + X_i\beta + Z_{it}\gamma, \epsilon_{i1} < \alpha_1 + X_i\beta + Z_{it}\gamma, \epsilon_{i2} < \alpha_2 + X_i\beta + Z_{it}\gamma\right) & \text{If } T_i = 4 \end{cases} \\
&= \begin{cases} \Phi\left((-\alpha_0 - X_i\beta - Z_{it}\gamma)/\sigma_0\right) & \text{If } T_i = 1 \\ \int_{-\infty}^{\alpha_0 + X_i\beta + Z_{it}\gamma} \Phi\left((-\alpha_1 - X_i\beta - Z_{it}\gamma - \rho\epsilon_{i0})\right) \frac{\phi(\epsilon_{i0}/\sigma_0)}{\sigma_0} d\epsilon_{i0} & \text{If } T_i = 2 \\ \int_{-\infty}^{\alpha_0 + X_i\beta + Z_{it}\gamma} \int_{-\infty}^{\alpha_1 + X_i\beta + Z_{it}\gamma} \Phi\left(-\alpha_2 - X_i\beta - Z_{it}\gamma - \rho\epsilon_{i1}\right) \phi(\epsilon_{i1} - \rho\epsilon_{i0}) \frac{\phi(\epsilon_{i0}/\sigma_0)}{\sigma_0} d\epsilon_{i1} d\epsilon_{i0} & \text{If } T_i = 3 \\ \int_{-\infty}^{\alpha_0 + X_i\beta + Z_{it}\gamma} \int_{-\infty}^{\alpha_1 + X_i\beta + Z_{it}\gamma} \Phi\left(\alpha_2 + X_i\beta + Z_{it}\gamma - \rho\epsilon_{i1}\right) \phi(\epsilon_{i1} - \rho\epsilon_{i0}) \frac{\phi(\epsilon_{i0}/\sigma_0)}{\sigma_0} d\epsilon_{i1} d\epsilon_{i0} & \text{If } T_i = 4 \end{cases}
\end{aligned}$$

The data for this problem is a sample from the National Survey of Mortgage Originations Public Use File. A version of this data-set is available in STATA format: Mortgage_performance_data.dta. The data-set includes 16,401 individual residential 30-year mortgages originated between 2013 and 2017. The characteristics (X_i) include:

```
score_0 rate_spread i_large_loan i_medium_loan i_refinance  age_r cltv dti
cu first_mort_r i_FHA  i_open_year2-i_open_year5
```

The time-varying characteristics Z_{it} include the risk score of the borrower:

```
score_0 score_1 score_2
```

Note that the initial risk score (score_0) enters both Z and X. This is to capture the fact that the initial score captures a permanent source of heterogeneity across borrowers, while the other two capture changes in consumers' balance sheets.

The main outcome variables are indicator variables equal to one if the loan is pre-paid in each year: Y_{it} =i_close_t (for $t = 0, 1, 2$). The variable "duration" corresponds to the T_i variable above. The Stata file contains the variable labels for most variables. The do-file "PS2.do" describes the main variables, and estimate a series probit model of the probability of repayment.

1. Using your favorite software (not Stata or R), write a routine that evaluates the log-likelihood function using the quadrature method. You can use the nodes/weights with precision 20. See attached spreadsheets.

2. Using your favorite software (not Stata or R), write a routine that evaluates the simulated log-likelihood function using the GHK method. Set the number of simulation draws to 100.

3. Using your favorite software (not Stata or R), write a routine that evaluates the simulated log-likelihood function Accept/Reject method. Set the number of simulation draws to 100.

4. Using the following parameter values:

$$\alpha_0 = 0, \alpha_1 = -1, \alpha_2 = -1, \beta = 0, \gamma = 0.3, \rho = 0.5$$

   compare the predicted choice probabilities calculated using the three methods above.

5. Maximize the log-likelihood function using the Gaussian Quadrature integration method. You can use the BFGS optimization method.

Additional notes on quadrature integration

For this problem, we will use the KPU sparse-grid quadrature nodes/weights (available here: http://www.sparse-grids.de). Each grid point u defined over the $(0,1)$ interval, we therefore need to transform it to the proper range. Let $\rho(u)$ denotes this transformation function, and $\rho'(x)$ denotes the derivative. The approximation of an integral of dimension K is done as follows:

$$\begin{aligned}
\bar{m} &= \int_{a_1}^{b_1} \cdots \int_{a_K}^{b_K} m(x_1,\ldots,x_K) f(x_1,\ldots,x_K) dx \\
&\approx \sum_r w_r \cdot \underbrace{m(\rho(u_1^r),\ldots,\rho(u_K^r))}_{\text{Function}} \cdot \underbrace{f(\rho(u_1^r),\ldots,\rho(u_K^r))}_{\text{Density}} \cdot \underbrace{[\rho'(u_1^r)\cdot\rho'(u_2^r)\cdot\cdots\cdot\rho'(u_K^r)]}_{\text{Jacobian}}
\end{aligned}$$

The transformation depends on the range of integration. I like to use a log/logistic transformations:

- Bounded support: (a,b)
$$\rho(u) = (b-a)\cdot u + a$$
$$\rho'(u) = (b-a)$$
- Unbounded support: $(-\infty,\infty)$
$$\rho(u) = -\ln\frac{1-u}{u}$$
$$\rho'(u) = -\frac{u}{1-u}\left(\frac{-1}{u} - \frac{(1-u)}{u^2}\right)$$
- Lower bound: (a,∞)
$$\rho(u) = -\ln(1-u) + a$$
$$\rho'(u) = \frac{1}{1-u}$$
- Upper bound: $(-\infty,b)$
$$\rho(u) = \ln u + b$$
$$\rho'(u) = \frac{1}{u}$$

The Ox function *quad_transform.ox* is attached to the problem-set. You just need to write a version using your programming language. Here is a pseudo-code using this function to evaluate the choice-probability for duration $T=3$:

```
transform(&mEps0,&mJac0,mGrid2,-1,M_INF_NEG,a0);
transform(&mEps1,&mJac1,mGrid2,-1,M_INF_NEG,a1);
density=densn(mEps1-r*mEps0).*densn(mEps0/s)/s;
vProb=(probn(a2-r*mEps1).*density.*mJac0.*mJac1)*vWeight2;
```

The quadrature grid and weight are denoted by: $vWeight2$ and $mGrid2$.