Supplementary material to: On Lasso and Adaptive Lasso for non-random sample in credit scoring

Emmanuel O. Ogundimu
Department of Mathematical Sciences, Durham University, UK

Appendix

A.1. **American Express credit card data**

Table A1: Variables used in analysis of American Express credit card data

Indicators	
CARDHLDR	=1 for cardholders, 0 for denied applicants.
DEFAULT	=1 for defaulted on payment, 0 if not.
Demographic and Socioeconomic, from Application	
AGE	= age in years and twelfths of a year.
OWNRENT	= indicators = 1 if own home, 0 if rent.
INCOME	= primary income.
SELFEMPL	= 1 if self employed, 0 if otherwise.
ADEPCNT	= number of dependents.
Constructed Variables	
INCPER	= income per family member=(income+additional income)/(1+dependents).
EXP_INC	= average expenditure for 12 months/average monthly income
Miscellaneous Application Data	
ACADMOS	= months living at current address.
AEMPMOS	= months employed.
CREDMAJR	= 1 if first credit card indicated on application is a major credit card.
Types of Bank Accounts	
BANKSAV	= 1 if only savings account, 0 otherwise.
BANKCH	= 1 if only checking account, 0 else.
BANKBOTH	= 1 if both savings and checking, 0 else.
Derogatories and Other Credit Data	
MAJORDRG	= count of major derogatory reports (long delinquencies) from credit bureau.
MINORDRG	= count of minor derogatories from credit bureau.
TRADACCT	= number of open, active trade lines.
Credit Bureau Data	
CBURDEN	= ratio of the monthly credit expenditures to the number of household members.

A.2. **Derivation of Gradients and Hessian**

Define $r = (1-\rho^2)$; $\omega_1 = \Big(\boldsymbol{\gamma}^T\mathbf{w}_i - \rho\boldsymbol{\beta}^T\mathbf{x}_i\Big)\Big/\sqrt{1-\rho^2}$; $\omega_2 = \Big(-\boldsymbol{\beta}^T\mathbf{x}_i + \rho\boldsymbol{\gamma}^T\mathbf{w}_i\Big)\Big/\sqrt{1-\rho^2}$; $K_1 = \boldsymbol{\beta}^T\mathbf{x}_i$; $K_2 = \boldsymbol{\gamma}^T\mathbf{w}_i$

The gradient of the bivariate probit selection model is:

$$
\begin{aligned}
\frac{\partial l}{\partial \beta} &= \sum_{i=1}^{n} S_i(1-Y_i)\left\{\frac{-\mathbf{x}\phi(-K_1)\Phi(\omega_1)}{\Phi_2(K_2,-K_1;-\rho)}\right\} + S_iY_i\left\{\frac{\mathbf{x}\phi(K_1)\Phi(\omega_1)}{\Phi_2(K_2,K_1;\rho)}\right\}\\
\frac{\partial l}{\partial \gamma} &= \sum_{i=1}^{n}(1-S_i)\left\{\frac{-\mathbf{w}\phi(-K_2)}{\Phi(-K_2)}\right\} + S_i(1-Y_i)\left\{\frac{\mathbf{w}_i\phi(K_2)\Phi(\omega_2)}{\Phi_2(K_2,-K_1;-\rho)}\right\} + S_iY_i\left\{\frac{\mathbf{w}_i\phi(K_2)\Phi(-\omega_2)}{\Phi_2(K_2,K_1;\rho)}\right\}\\
\frac{\partial l}{\partial \rho} &= \sum_{i=1}^{n} S_i(1-Y_i)\left\{\frac{-\phi_2(K_2,-K_1;-\rho)}{\Phi_2(K_2,-K_1;-\rho)}\right\} + S_iY_i\left\{\frac{\phi_2(K_2,K_1;\rho)}{\Phi_2(K_2,K_1;\rho)}\right\}
\end{aligned}
$$

The derivative of the bivariate normal CDF is calculated using equation (4) given in **?**. In general, if $x = (x_1^T, x_2^T)^T \in \mathbb{R}^d$, $x_1 \in \mathbb{R}^{d_1}$, $x_2 \in \mathbb{R}^{d_2}$, $d_1 + d_2 = d$, with corresponding decomposition of the covariance matrix Σ:, then

$$
\frac{\partial}{\partial x_1}\Phi_d(x;\Sigma) = \phi_{d_1}(x_1;\Sigma_{11})\Phi_{d_2}(x_2 - \Sigma_{21}\Sigma_{11}^{-1}x_1; \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}).
$$

To speed up the computation of the gradient and the Hessian matrix, we decompose the PDF of bivariate normal distribution into univariate PDFs. For example,

$$
\begin{aligned}
\phi_2(\boldsymbol{\gamma}^T\mathbf{w}_i, -\boldsymbol{\beta}^T\mathbf{x}_i; -\rho) =& \phi(\boldsymbol{\gamma}^T\mathbf{w}_i)(1-\rho^2)^{-1/2}\phi\Big(\frac{-\boldsymbol{\beta}^T\mathbf{x}_i + \rho\boldsymbol{\gamma}^T\mathbf{w}_i}{\sqrt{1-\rho^2}}\Big)\\
=& \phi(\boldsymbol{\beta}^T\mathbf{x}_i)(1-\rho^2)^{-1/2}\phi\Big(\frac{-\boldsymbol{\gamma}^T\mathbf{w}_i - \rho\boldsymbol{\beta}^T\mathbf{x}_i}{\sqrt{1-\rho^2}}\Big).
\end{aligned}
$$

The elements of the Hessian matrix are:

$$
\begin{aligned}
\frac{\partial^2 l}{\partial \beta^2} &= \sum_{i=1}^{n} S_i(1-Y_i)\Bigg\{\Bigg[\frac{\mathbf{x}_i^2\phi(K_1)\Big(K_1\Phi(\omega_1)+\rho r^{-1/2}\phi(\omega_1)\Big)}{\Phi_2(K_2,-K_1;-\rho)}\Bigg]-\Bigg[\Bigg(\frac{\mathbf{x}_i\phi(K_1)\Phi(\omega_1)}{\Phi_2(K_2,-K_1;-\rho)}\Bigg)^2\Bigg]\Bigg\} \\
&+S_iY_i\Bigg\{\Bigg[\frac{-\mathbf{x}_i^2\phi(K_1)\Big(K_1\Phi(\omega_1)+\rho r^{-1/2}\phi(\omega_1)\Big)}{\Phi_2(K_2,K_1;\rho)}\Bigg]-\Bigg[\Bigg(\frac{\mathbf{x}_i\phi(K_1)\Phi(\omega_1)}{\Phi_2(K_2,K_1;\rho)}\Bigg)^2\Bigg]\Bigg\} \\
\frac{\partial^2 l}{\partial \gamma^2} &= \sum_{i=1}^{n}(1-S_i)\Bigg\{\frac{\mathbf{w}_i^2K_2\phi(K_2)}{\Phi(-K_2)}-\Bigg(\mathbf{w}_i\frac{\phi(K_2)}{\Phi(-K_2)}\Bigg)^2\Bigg\} \\
&+S_i(1-Y_i)\Bigg\{\Bigg[\frac{\mathbf{w}_i^2\phi(K_2)\Big(-K_2\Phi(\omega_2)+\rho r^{-1/2}\phi(\omega_2)\Big)}{\Phi_2(K_2,-K_1;-\rho)}\Bigg]-\Bigg[\Bigg(\frac{\mathbf{w}_i\phi(K_2)\Phi(\omega_2)}{\Phi_2(K_2,-K_1;-\rho)}\Bigg)^2\Bigg]\Bigg\} \\
&+S_iY_i\Bigg\{\Bigg[\frac{-\mathbf{w}_i^2\phi(K_2)\Big(K_2\Phi(-\omega_2)+\rho r^{-1/2}\phi(-\omega_2)\Big)}{\Phi_2(K_2,K_1;\rho)}\Bigg]-\Bigg[\Bigg(\frac{\mathbf{w}_i\phi(K_2)\Phi(-\omega_2)}{\Phi_2(K_2,K_1;\rho)}\Bigg)^2\Bigg]\Bigg\} \\
\frac{\partial^2 l}{\partial \rho^2} &= \sum_{i=1}^{n} S_i(1-Y_i)\Bigg\{\Bigg[\frac{-\phi(K_1)\phi(\omega_1)r^{-3/2}\Big(\rho-r^{-1/2}\omega_1(\rho K_2-K_1)\Big)}{\Phi_2(K_2,-K_1;-\rho)}\Bigg]-\Bigg[\Bigg(\frac{\phi(K_1)r^{-1/2}\phi(\omega_1)}{\Phi_2(K_2,-K_1;-\rho)}\Bigg)^2\Bigg]\Bigg\} \\
&+S_iY_i\Bigg\{\Bigg[\frac{\phi(K_1)\phi(\omega_1)r^{-3/2}\Big(\rho-r^{-1/2}\omega_1(\rho K_2-K_1)\Big)}{\Phi_2(K_2,K_1;\rho)}\Bigg]-\Bigg[\Bigg(\frac{\phi(K_1)r^{-1/2}\phi(\omega_1)}{\Phi_2(K_2,K_1;\rho)}\Bigg)^2\Bigg]\Bigg\} \\
\frac{\partial^2 l}{\partial \beta\partial\gamma} &= \sum_{i=1}^{n} S_i(1-Y_i)\Bigg\{\Bigg[\frac{-\mathbf{x}_i\mathbf{w}_i\phi(-K_1)r^{-1/2}\phi(\omega_1)}{\Phi_2(K_2,-K_1;-\rho)}\Bigg]-\Bigg[\frac{-\mathbf{x}_i\mathbf{w}_i\phi(-K_1)\Phi(\omega_1)\phi(K_2)\Phi(\omega_2)}{(\Phi_2(K_2,-K_1;-\rho))^2}\Bigg]\Bigg\} \\
&+S_iY_i\Bigg\{\Bigg[\frac{\mathbf{x}_i\mathbf{w}_i\phi(K_1)r^{-1/2}\phi(\omega_1)}{\Phi_2(K_2,K_1;\rho)}\Bigg]-\Bigg[\frac{\mathbf{x}_i\mathbf{w}_i\phi(K_1)\Phi(\omega_1)\phi(K_2)\Phi(-\omega_2)}{(\Phi_2(K_2,K_1;\rho))^2}\Bigg]\Bigg\} \\
\frac{\partial^2 l}{\partial \beta\partial\rho} &= \sum_{i=1}^{n} S_i(1-Y_i)\Bigg\{\Bigg[\frac{-\mathbf{x}_i\phi(K_2)r^{-1}\omega_2\phi(\omega_2)}{\Phi_2(K_2,-K_1;-\rho)}\Bigg]-\Bigg[\frac{\mathbf{x}_i\phi(K_2)r^{-1/2}\phi(\omega_2)\phi(-K_1)\Phi(\omega_1)}{(\Phi_2(K_2,-K_1;-\rho))^2}\Bigg]\Bigg\} \\
&+S_iY_i\Bigg\{\Bigg[\frac{\mathbf{x}_i\phi(K_2)r^{-1}\omega_2\phi(-\omega_2)}{\Phi_2(K_2,K_1;\rho)}\Bigg]-\Bigg[\frac{\mathbf{x}_i\phi(K_2)r^{-1/2}\phi(\omega_2)\phi(K_1)\Phi(\omega_1)}{(\Phi_2(K_2,K_1;\rho))^2}\Bigg]\Bigg\} \\
\frac{\partial^2 l}{\partial \gamma\partial\rho} &= \sum_{i=1}^{n} S_i(1-Y_i)\Bigg\{\Bigg[\frac{\mathbf{w}_i\phi(K_1)r^{-1}\omega_1\phi(\omega_1)}{\Phi_2(K_2,-K_1;-\rho)}\Bigg]-\Bigg[\frac{-\mathbf{w}_i\phi(K_1)r^{-1/2}\phi(\omega_1)\phi(K_2)\Phi(-\omega_2)}{(\Phi_2(K_2,-K_1;-\rho))^2}\Bigg]\Bigg\} \\
&+S_iY_i\Bigg\{\Bigg[\frac{-\mathbf{w}_i\phi(K_1)r^{-1}\omega_1\phi(\omega_1)}{\Phi_2(K_2,K_1;\rho)}\Bigg]-\Bigg[\frac{\mathbf{w}_i\phi(K_1)r^{-1/2}\phi(\omega_1)\phi(K_2)\Phi(-\omega_2)}{(\Phi_2(K_2,K_1;\rho))^2}\Bigg]\Bigg\}
\end{aligned}
$$

A.3. **Gaussian and AMH (Ali–Mikhail–Haq) copulas**

Gaussian copula:

$$C(u_1, u_2; \theta) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_1)} \frac{1}{2\pi\sqrt{1-\theta^2}} \left\{ \exp\left[-\frac{s^2 + 2\theta st + t^2}{2(1-\theta^2)} \right] \right\} \mathrm{dsdt}$$
$$= \Phi_2\Big(\Phi^{-1}(u_1)), \Phi^{-1}(u_2); \theta\Big),$$

where Φ is the CDF of the standard normal distribution, Φ_2 is the CDF of the standard bivariate normal distribution and θ is the correlation parameter. The Gaussian copula is flexible in that it allows for equal degree of positive and negative dependence that includes both Fretchet bounds in its permissible range.

AMH copula:

$$C(u_1, u_2; \theta) = \frac{u_1 u_2}{1 - \theta(1-u_1)(1-u_2)},$$

where the dependence parameter θ is defined on the interval [-1,1]. The copula tends to the product copula as θ approaches 0. Dependence in the tails is weak and so it is often used where dependence is relatively modest. It does not include both Fretchet bounds in its permissible range.

A.4. **Tables of results based on data generated via AMH copula and additional American Express credit card data analysis**

Table A2: Results on the covariate selection based on selection model and corresponding complete case analyses - AMH data generation

Method	Sensitivity	Specificity	Sensitivity	Specificity
	Outcome Equation		Selection Equation	
		$\rho = 0$		
BPSSM_P-value	0.742	0.940	0.849	0.963
BPSSM_Lasso	0.959	0.702	0.991	0.568
BPSSM_ALasso	0.833	0.927	0.914	0.931
CBSSM_Lasso	0.967	0.725	0.991	0.592
CBSSM_ALasso	0.831	0.929	0.911	0.938
PROBIT_P-value	0.748	0.939	-	-
PROBIT_Lasso	0.963	0.730	-	-
PROBIT_ALasso	0.821	0.950	-	-
		$\rho = 0.2$		
BPSSM_P-value	0.716	0.955	0.851	0.940
BPSSM_Lasso	0.971	0.653	0.987	0.585
BPSSM_ALasso	0.836	0.919	0.914	0.904
CBSSM_Lasso	0.973	0.671	0.987	0.591
CBSSM_ALasso	0.835	0.932	0.909	0.910
PROBIT_P-value	0.720	0.953	-	-
PROBIT_Lasso	0.971	0.679	-	-
PROBIT_ALasso	0.825	0.936	-	-
		$\rho = 0.5$		
BPSSM_P-value	0.714	0.954	0.843	0.941
BPSSM_Lasso	0.968	0.663	0.987	0.605
BPSSM_ALasso	0.838	0.924	0.901	0.929
CBSSM_Lasso	0.973	0.670	0.988	0.608
CBSSM_ALasso	0.836	0.936	0.898	0.940
PROBIT_P-value	0.723	0.954	-	-
PROBIT_Lasso	0.972	0.675	-	-
PROBIT_ALasso	0.826	0.937	-	-

Table A3: Simulation results for the number of times each covariate is selected (out of 200) with both weak and moderate covariate effects (Outcome equation)- AMH data generation.

Method	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
					$\rho = 0$						
BPSSM_P-value	120	78	92	12	14	11	10	13	200	200	200
BPSSM_Lasso	186	180	185	62	62	49	56	69	200	200	200
BPSSM_ALasso	146	120	133	20	15	13	10	15	200	200	200
CBSSM_Lasso	190	183	187	59	56	43	53	64	200	200	200
CBSSM_ALasso	144	122	131	19	15	12	11	14	200	200	200
Probit_P-value	124	79	94	13	15	11	10	12	200	200	200
Probit_Lasso	188	181	187	57	56	43	48	66	200	200	200
Probit_ALasso	137	120	128	14	11	6	10	9	200	200	200
					$\rho = 0.2$						
BPSSM_P-value	99	73	87	16	8	5	9	7	200	200	200
BPSSM_Lasso	188	187	190	75	71	60	77	64	200	200	200
BPSSM_ALasso	131	133	139	20	12	16	19	14	200	200	200
CBSSM_Lasso	190	187	191	71	68	62	69	59	200	200	200
CBSSM_ALasso	133	134	135	19	10	11	15	13	200	200	200
Probit_P-value	98	74	92	16	9	5	9	8	200	200	200
Probit_Lasso	188	188	189	68	66	59	71	57	200	200	200
Probit_ALasso	127	132	131	19	8	9	13	15	200	200	200
					$\rho = 0.5$						
BPSSM_P-value	137	118	125	14	13	8	11	13	200	200	200
BPSSM_Lasso	190	185	187	69	73	64	69	62	200	200	200
BPSSM_ALasso	132	137	136	22	10	11	18	15	200	200	200
CBSSM_Lasso	191	186	191	71	70	59	71	59	200	200	200
CBSSM_ALasso	130	136	137	20	8	8	14	14	200	200	200
Probit_P-value	100	75	92	16	8	5	9	8	200	200	200
Probit_Lasso	188	188	190	68	66	60	73	58	200	200	200
Probit_ALasso	128	131	132	19	8	8	13	15	200	200	200

Table A4: Results of the optimism corrected model performance- AMH data generation

Method	AUROC	AUPRC	Brier	ECE	MCE
			$\rho = 0$		
BPSSM_P-value	0.929	0.652	0.056	0.014	0.057
BPSSM_Lasso	0.931	0.661	0.056	0.019	0.078
BPSSM_ALasso	0.931	0.658	0.056	0.015	0.057
CBSSM_Lasso	0.931	0.661	0.056	0.019	0.076
CBSSM_ALasso	0.931	0.658	0.055	0.014	0.056
Probit_P-value	0.929	0.653	0.056	0.014	0.057
Probit_Lasso	0.932	0.662	0.056	0.018	0.071
Probit_ALasso	0.931	0.658	0.056	0.014	0.053
			$\rho = 0.2$		
BPSSM_P-value	0.928	0.646	0.057	0.014	0.055
BPSSM_Lasso	0.932	0.660	0.056	0.019	0.080
BPSSM_ALasso	0.931	0.656	0.056	0.016	0.061
CBSSM_Lasso	0.932	0.661	0.056	0.019	0.077
CBSSM_ALasso	0.931	0.656	0.056	0.015	0.058
Probit_P-value	0.928	0.649	0.057	0.014	0.057
Probit_Lasso	0.932	0.662	0.056	0.018	0.072
Probit_ALasso	0.931	0.657	0.056	0.014	0.056
			$\rho = 0.5$		
BPSSM_P-value	0.928	0.647	0.057	0.013	0.052
BPSSM_Lasso	0.932	0.660	0.057	0.020	0.081
BPSSM_ALasso	0.930	0.655	0.057	0.015	0.060
CBSSM_Lasso	0.932	0.660	0.056	0.019	0.078
CBSSM_ALasso	0.930	0.656	0.056	0.015	0.059
Probit_P-value	0.928	0.648	0.057	0.014	0.057
Probit_Lasso	0.932	0.661	0.056	0.018	0.071
Probit_ALasso	0.931	0.656	0.056	0.015	0.057