Instituto Complutense de Análisis Económico
UNIVERSIDAD COMPLUTENSE
Campus de Somosaguas
28223 MADRID

Documento de Trabajo

Outliers in Binary Choice Models

Mercedes Gracia-Díez
Gregorio R. Serrano

No.9505

Mayo 1995

Instituto Complutense de Análisis Económico
UNIVERSIDAD COMPLUTENSE

OUTLIERS IN BINARY CHOICE MODELS

Mercedes Gracia-Díez*
Gregorio R. Serrano*
Departamento de Economía Cuantitativa
Facultad de Ciencias Económicas
Universidad Complutense
28223 Madrid, Spain
Ph. 34-1-3942370, FAX 34-1-3942611

ABSTRACT

This paper focuses on the problem of outliers in binary choice models. It is show that identifying outliers as observation with a residual close to one in absolute value might be misleading and outlier detection procedures should rely on influence measures on the fit. Two scalar measures are derived to evaluate the influence of each observation as well as the influence of a group of observations on i) the vector of estimated parameters and ii) the vector of estimated probabilities. Also, the method proposed by Peña and Yohai (1995) to treat the problem of masking in linear models has been generalized to the case of binary choice models. A small Monte Carlo study analyzes the performance of all measures and an empirical application presents a diagnostic strategy for detection of outliers.

RESUMEN

Este trabajo trata el problema de la presencia de observaciones extremas en modelos de elección binaria. Se muestra que identificar como atípicas observaciones con residuos próximos a uno, en valor absoluto, puede resultar erróneo y que los procedimientos de detección de estas observaciones deberían basarse en medidas de influencia sobre el ajuste. Se proponen dos medidas para evaluar la influencia de cada observación así como para un grupo de observaciones sobre: i) El vector de parámetros estimados y ii) el vector de las probabilidades estimadas. También, se ha generalizado al caso de modelos de elección binaria, el método propuesto por Peña y Yohai (1995) para tratar el problema de enmascaramiento en modelos lineales. Con un pequeño estudio de Monte Carlo se analizan las propiedades de las medidas propuestas y una aplicación empírica presenta una estrategia para la detección de medidas extremas.

Key Words: outliers, detection, residuals, influence measures, masking problem, diagnostic strategy.

* We are grateful to Alfonso Novales and Daniel Peña for helpful suggestions.
The responsibility for all errors is only ours.
Finance for this research was provided by the DGICYT, Spain, under grant PB92-1072.

1. Introduction

The problem of outlying observations in linear regression models has been a subject of considerable research in the statistics literature. Less attention has been paid to this topic in respect of binary response models. Although these models have been used for decades, strategies for outlier detection have been developed only recently. In previous work (Pregibon, 1981; Cook and Weisberg, 1982; Copas, 1988; among others) an outlier is defined as the success with the lowest estimated probability of success or the failure with the highest estimated probability of success. That is, an outlier is identified as an observation with a residual close to one in absolute value. Therefore, most of diagnostic techniques rely on residual analyses. Since normal approximations are clearly inappropriated for binary data, the research has mainly focussed on developing the distribution of different sets of residuals (Jennings, 1986; Pierce and Schafer, 1986; Williams, 1987; Bedrick and Hill, 1990).

In this paper the standard concept of outliers in econometrics is applied to binary response models: an outlier is defined as an observation that has not been generated by the stochastic model assumed. Using this definition and due to the censoring of the dependent variable in this type of models, we show that outliers are not necessarily observations with a residual close to one in absolute value. Further, an observation from the tails of the distribution might present a residual larger than that associated to a true outlier. This suggests that diagnostics based on residual analyses might be misleading and outlier detection procedures should rely on techniques for judging the influence of an observation on a particular aspect of the fit, such as those developed by Pregibon (1981) for logistic regression and McCullagh and Nelder (1983) for generalized linear models.

This paper proposes diagnostic measures to evaluate the influence of an individual observation as well as the joint influence of a group of observations on the estimation of the model. In particular, we develop two scalar measures; the first one evaluates the influence on the maximum likelihood (ML) estimate of the parameter vector, while the second one evaluates the effect on the estimated probability vector. Also, a computational efficient procedure to treat the problem of masking is presented. This procedure generalizes the method of Peña and Yohai (1995) for linear regression to the case of binary response models. All the measures proposed in this paper take into account the specific features of binary choice models and they can be easily computed for the probit and logit cases as well as for any other case in which a continuous distribution is assumed. A small Monte Carlo study illustrates the performance of all measures and an empirical application to the sample used by Dhillon et. al. (1987) presents a diagnostic strategy for detecting outliers in binary choice models. In this particular application, where the choice between fixed and adjustable rate mortgages is examined, the exclusion of a small number of outliers from the sample produces different results from those obtained in the original article in which the existence of outlying observations is ignored.

The paper is organized as follows. Section 2 contains a formal definition of outliers in binary response models. In section 3 the problem of using residual analyses for detecting outliers is discussed. Section 4 presents two scalar measures to evaluate the influence of each individual observation on the fit. In section 5 both measures are generalized to analyze the joint influence of a group of observations. Also measures to evaluate the influence of a group of observations on a subset of coefficients in the model are provided. Section 6 contains the procedure to deal with the problem of masking. Section 7 is devoted to the Monte Carlo study. Section 8 presents the empirical application to the analysis of Dhillon et al. (1987) and, finally, section 9 concludes the study.

2. Definition of outliers in binary choice models

To deal with the problem of outliers in binary choice models, we depart from the standard concept of outliers in econometrics: an outlier is an observation which is suspected of being partially or wholly irrelevant because it is not generated by the stochastic model assumed (see for instance Box and Tiao, 1968). We also use the fact that a binary choice model can be derived

from a linear regression model with latent dependent variable (McFadden, 1983):

$$y_i^* = x_i^T\beta + u_i \qquad i=1,\dots,n \qquad [1]$$

where x_i is the vector of observations of the k explanatory variables for individual i, β is the common vector of coefficients and u_i are iid continuous random variables with zero mean and known σ_0^2 variance. Then, introducing an indicator function $y_i = z(y_i^*)$ such that: $y_i=1$ if $y_i^*\geq 0$ and $y_i=0$ if $y_i^*<0$, and being $F(u_i|x_i)$ the cumulative distribution function of the disturbances, usually the standard normal (probit model) or the standard logistic (logit model), we have:

$$P_i = P(y_i=1\,|\,x_i) = P(y_i^* \geq 0\,|\,x_i) = P(u_i \geq -x_i^T\beta) = F(x_i^T\beta) \qquad [2]$$

which is the response probability model.

Note that equation [1] is a linear regression model that fulfills the usual hypotheses. In particular, this equation states that the variables y_i^* have been generated by *the same* stochastic model; that is, y_i^* comes from a $F(y_i^*|x_i)$ cumulative distribution with $Ey_i^*=x_i^T\beta$ and $var(y_i^*)=\sigma_0^2$ for all i. Then, according to the above definition of outliers, a variable y_i will be anomalous when its associated latent variable y_i^* had not been generated by model [1].

3. The problem of using residual analyses for detecting outliers

In empirical work a battery of diagnostic checks is used to detect observations which are very unlikely of being generated by the stochastic model assumed. The objective of this section is to show that residual analysis is a misleading diagnostic tool for detection of outliers in binary choice models. This is an important difference between this type of models and linear models, where residual analysis is a useful check. Take the usual definition of residual for observation i:

$$e_i = y_i - \hat{P}_i \qquad [3]$$

such that e_i is bounded between (-1,1), $E(e_i) = 0$ and $var(e_i) = \hat{P}_i(1-\hat{P}_i)$.

In previous literature (Pregibon, 1981; Cook and Weisberg, 1982; Williams, 1987; Pierce and Schafer, 1986; Copas, 1988; Bedrick and Hill, 1990) no definition of outlier is considered and, typically, outlier detection procedures identify as outliers those observations with a residual e_i close to one in absolute value. Jennings (1986, pp.988) defines an outlier as an observation such that $P[y_i = s \mid x_i] < \alpha$ where $s = 0, 1$ and α is small. Based on this definition, he claims that "outliers are necessary for logistic regression estimation" since the fitting of these models strongly depends on small probability events. Our argument is that outliers and small probability events are two different concepts. If an observation shows a residual e_i close to one in absolute value, it does not necessarily imply that y_i^* had not been generated by the distribution assumed; it could mean that the value of y_i in the sample has small probability because y_i^* comes from the tails of the distribution.

To see this, we use Figure 1. The lower part of the figure contains the (y_i^*, x_i^T) points, where $x_i^T = (1, x_i)$, and the true regression line $x_i^T\beta$ of model [1]. The upper part of the figure shows the probabilities associated to model [2] for the probit case, as well as the sample values of y_i, such that $y_i = 1$ if $y_i^* \geq 0$ and $y_i = 0$ if $y_i^* < 0$. Note that in practice the available sample consists on the (y_i, x_i) points and the lower part of the figure is not observed. However, this example might be useful to illustrate the following situation.

Consider point A corresponding to a high positive value of y_i^*. In linear regression, where y_A^* is observed, this point will show a large positive residual and it would be identified as an outlier. In a binary response model the realization of $y_A^* > 0$ is $y_A = 1$ and, hence, the associated residual as defined in [3] will be positive and close to one. Consider now point B corresponding to a value of y_i^* which comes from the distribution assumed in model [1]. Since the realization of $y_B^* > 0$ is $y_B = 1$, once the probit model is estimated this point will also present a large positive residual; in fact *identical* to the residual of point A. Therefore, an observation with a residual close to one in absolute

value is not necessarily an outlier; point A is an outlier, while point B is an observation from the tails of the distribution. Further, the residual associated to point C, which also belongs to the tails of the distribution, is *larger* than the residual of point A. As a consequence, when observations with a large residual are eliminated from the sample, we might be censoring observations of small probability, which are crucial for the fitting of the model. Only outliers should be eliminated.

4. Influence measures for individual observations

The first measure evaluates the influence of each observation in a given sample on the ML estimate of β in [2]. This measure is based on the distance introduced by Cook (1977) for linear regression models and it is given by:

$$\hat{c}_i = (\hat{\beta} - \hat{\beta}_{(i)})^T I(\hat{\beta}) (\hat{\beta} - \hat{\beta}_{(i)}) \qquad i = 1, \ldots, n \qquad [4]$$

where $\hat{\beta}$ is the ML estimate of β, $\hat{\beta}_{(i)}$ is the ML estimate after eliminating the ith observation and $I(\hat{\beta})$ is the information matrix evaluated at $\hat{\beta}$.

In order to compute efficiently the measure in [4], we use the fact that given an initial estimate $\hat{\beta}^r$, an iteration by the method of scoring to calculate the ML estimate of β can be obtained from the following OLS linear regression (Amemiya, 1981, 1985):

$$\tilde{y}_i = \tilde{x}_i^T \beta + u_i \qquad [5]$$

where:

$$\tilde{y}_i = \frac{y_i + \hat{f}_i x_i^T \hat{\beta}^r - \hat{F}_i}{[\hat{F}_i(1-\hat{F}_i)]^{1/2}} \qquad [6]$$

$$\tilde{x}_i = \frac{\hat{f}_i}{[\hat{F}_i(1-\hat{F}_i)]^{1/2}} x_i \qquad [7]$$

$\hat{F}_i = F(x_i^T \hat{\beta}^\tau)$, $\hat{f}_i = f(x_i^T \hat{\beta}^\tau)$ and $var(u_i) = 1$ for all i.

Then, considering that convergence is achieved at iteration τ, i.e. $\hat{\beta}^\tau = \hat{\beta}$ is the ML estimate of β, and using the expression of the OLS estimator in [5] when observation i is deleted, after some algebraic manipulation, we have:

$$\hat{\beta}_{(i)} = \hat{\beta} + (\tilde{X}^T\tilde{X})^{-1}\tilde{x}_i[1 - \tilde{x}_i(\tilde{X}^T\tilde{X})^{-1}\tilde{x}_i]^{-1}(\tilde{x}_i^T\hat{\beta} - \tilde{y}_i) \qquad [8]$$

where $\tilde{X}$ is a $n \times k$ matrix which rows are given by expression [7].

Substituting [8] into [4] and given that $I(\hat{\beta}) = (\tilde{X}^T\tilde{X})$ and:

$$e_i^* = \frac{e_i}{[\hat{P}_i(1 - \hat{P}_i)]^{1/2}} = \tilde{y}_i - \tilde{x}_i^T\hat{\beta} \qquad [9]$$

where e_i is the residual of observation i as defined in [3] and e_i^* is the standardized residual, expression [4] can be written as:

$$\hat{c}_i = \frac{e_i^{*2}\,\tilde{x}_i^T(\tilde{X}^T\tilde{X})^{-1}\tilde{x}_i}{[1 - \tilde{x}_i^T(\tilde{X}^T\tilde{X})^{-1}\tilde{x}_i]^2} = \frac{e_i^{*2}\,\tilde{h}_i}{(1 - \tilde{h}_i)^2} \qquad [10]$$

where in general $\tilde{h}_{ij} = \tilde{x}_i^T(\tilde{X}^T\tilde{X})^{-1}\tilde{x}_j$ and particularly $\tilde{h}_{ii} \equiv \tilde{h}_i$.

As model [2] is a member of the class of the generalized linear models (Nelder and Wedderburn, 1972), expression [10] is a special case of the general influence measure for those models proposed by McCullagh and Nelder (1989). However, this general measure does not specify the particular features of each model, while measure [10] takes into account the specific characteristics of binary choice models and it is computationally efficient. Pregibon (1981) introduced a similar influence measure, but it only applies to the case of logistic regression.

Next we propose a second scalar measure, which considers the effect of each observation on the estimated probability vector ($\hat{\boldsymbol{P}}$). In linear regression models, the influence of an observation on the estimated parameter vector is identical to the influence of that observation on the vector $\hat{y}$ of fitted

values (Cook, 1977). However, since binary choice models are nonlinear, the previous result does not hold. Then it might be interesting to derive a scalar measure to evaluate the influence of the ith observation on vector $\hat{\boldsymbol{P}}$. Using a first order Taylor expansion, the estimated probability for the jth observation when observation i is deleted can be approximated by:

$$F(x_j^T\hat{\beta}_{(i)}) \approx F(x_j^T\hat{\beta}) + f(x_j^T\hat{\beta})x_j^T(\hat{\beta}_{(i)} - \hat{\beta}) \qquad [11]$$

Then, the difference between the vector of estimated probabilities when observation i is included in the sample and the vector of estimated probabilities when observation i is deleted, can be written as:

$$\hat{\boldsymbol{F}} - \hat{\boldsymbol{F}}_{(i)} = -\hat{\psi}X(\hat{\beta}_{(i)} - \hat{\beta}) = -\hat{\Psi}^{1/2}\tilde{X}(\hat{\beta}_{(i)} - \hat{\beta}) \qquad [12]$$

where $\hat{\psi}$ and $\hat{\Psi}$ are $n \times n$ diagonal matrices with generic elements $\hat{f}_i$ and $\hat{F}_i(1 - \hat{F}_i)$ respectively.

From [12] an influence measure on the vector of estimated probabilities is given by:

$$\hat{c}_i(P) = (\hat{\boldsymbol{F}} - \hat{\boldsymbol{F}}_{(i)})^T(\hat{\boldsymbol{F}} - \hat{\boldsymbol{F}}_{(i)}) = \frac{e_i^{2*}\,\bar{h}_i}{(1 - \tilde{h}_i)^2} \qquad [13]$$

where:

$$\bar{h}_{ij} = \tilde{x}_i^T(\tilde{X}^T\tilde{X})^{-1}(\tilde{X}\hat{\Psi}\tilde{X})(\tilde{X}^T\tilde{X})^{-1}\tilde{x}_j \quad \text{and} \quad \bar{h}_i \equiv \bar{h}_{ii} \qquad [14]$$

The difference between measures [10] and [13] is matrix $\hat{\Psi}$ in expression [14]. The presence of this matrix is because in a nonlinear model, the change in the function arguments does not have to be equal to the variation in the function. More specifically, equation [12] indicates that changes in the estimated probabilities do not only depend on the change in the vector of parameters but also on the values of the estimated probabilities with the whole sample, in such a way that the change is bigger as long as those probabilities are closer to 0.5. Therefore, the use of measure [13] is specially interesting when the main objective is aggregate forecasting in large populations.

Taking into account that the following inequality holds (Rao, 1973):

$$\lambda_{min}(\hat{\Psi}) \leq \frac{(\hat{\beta} - \hat{\beta}_{(i)})^T \tilde{X}^T \hat{\Psi} \tilde{X} (\hat{\beta} - \hat{\beta}_{(i)})}{(\hat{\beta} - \hat{\beta}_{(i)})^T \tilde{X}^T \tilde{X} (\hat{\beta} - \hat{\beta}_{(i)})} \leq \lambda_{max}(\hat{\Psi}) \qquad [15]$$

where $\lambda_{min}(\hat{\Psi})$ and $\lambda_{max}(\hat{\Psi})$ are respectively the minimum and maximum eigenvectors of the diagonal matrix $\hat{\Psi}$ such that $\lambda_{min}(\hat{\Psi}) \geq 0$ and $\lambda_{max}(\hat{\Psi}) \leq 1/4$, we have that the ratio between $\hat{c}_i(P)$ and $\hat{c}_i$ is bounded between the limits:

$$0 \leq \frac{\hat{c}_i(P)}{\hat{c}_i} \leq \frac{1}{4} \qquad [16]$$

Expression [16] indicates that $\hat{c}_i(P)$ always takes a smaller value than $\hat{c}_i$. Note that this ratio reaches a maximum for observations with a little influence on the estimate of β but with a big effect on the estimated probability vector.

5. Influence measures for groups of observations and other particular cases

In this section we extent the results of the previous section to a situation in which the joint influence of a group of observations is analyzed. Also we provide measures to evaluate the influence of a group of observations on a subset of estimated parameters in the model.

A similar measure to [10] is derived to evaluate the influence of a group I of p observations on the ML estimate of vector β. This measure is as follows:

$$\begin{aligned} \hat{c}_I &= (\hat{\beta}_{(I)} - \hat{\beta})^T (\tilde{X}^T \tilde{X}) (\hat{\beta}_{(I)} - \hat{\beta}) \\ &= e_I^{*T} (I_p - \tilde{H}_I)^{-1} \tilde{H}_I (I_p - \tilde{H}_I)^{-1} e_I^* \end{aligned} \qquad [17]$$

where $\hat{\beta}_{(I)}$ denotes the ML of β when group I is eliminated, $\tilde{H}_I = \tilde{X}_I (\tilde{X}^T \tilde{X})^{-1} \tilde{X}_I^T$, $\tilde{X}_I$ is the $p \times k$ submatrix of $\tilde{X}$ that contains the observations of group I and e_I^* is the $p \times 1$ vector of standardized residuals.

Also, we obtain a measure to evaluate the influence of a group of observations I on the estimated probability vector. This measure is a generalization of the one in [13] and can be written as:

$$\begin{aligned} \hat{c}_I(P) &= (\hat{\beta}_{(I)} - \hat{\beta})^T (\tilde{X}^T \hat{\Psi} \tilde{X}) (\hat{\beta}_{(I)} - \hat{\beta}) \\ &= e_I^{*T} (I_p - \tilde{H}_I)^{-1} \bar{H}_I (I_p - \tilde{H}_I)^{-1} e_I^* \end{aligned} \qquad [18]$$

where $\bar{H}_I = \tilde{X}_I (\tilde{X}^T \tilde{X})^{-1} (\tilde{X}^T \hat{\Psi} X) (\tilde{X}^T \tilde{X})^{-1} \tilde{X}_I^T$.

Finally we derive an expression to evaluate the influence of a group of observations I on a linear combination of parameters of vector $\hat{\beta}$. Given the $m \times 1$ vector $\theta = R^T\beta$, where R^T is a $m \times k$ matrix of known constants with rank$(R) = m \leq k$, and being $R^T[I(\hat{\beta})]^{-1}R$ the covariance matrix of the ML estimates of θ, this measure is:

$$\begin{aligned} \hat{c}_I(\theta) &= (\hat{\theta}_{(I)} - \hat{\theta})^T \left[R^T (\tilde{X}^T \tilde{X})^{-1} R\right]^{-1} (\hat{\theta}_{(I)} - \hat{\theta}) \\ &= e_I^{*T} (I_p - \tilde{H}_I)^{-1} \tilde{X}_I \tilde{N} \tilde{X}_I^T (I_p - \tilde{H}_I)^{-1} e_I^* \end{aligned} \qquad [19]$$

where $\tilde{N} = (\tilde{X}^T \tilde{X})^{-1} R \left[R^T (\tilde{X}^T \tilde{X})^{-1} R\right]^{-1} R^T (\tilde{X}^T \tilde{X})^{-1}$, which is easy to compute since $\tilde{N}$ does not depend on the excluded observations.

The measure in [19] is specially interesting since binary choice models frequently contains several sets of qualitative explanatory variables. Therefore, it might be useful to evaluate the effect of a group of observations on a subset of components of $\hat{\beta}$. In this case, without lost of generality we can assume that the m last components of vector β are the parameters of interest. Then:

$$R^T\beta = (0_{m \times (k-m)} \mid I_m)\beta = (\beta_{k-m+1}, \ldots, \beta_k)^T \qquad [20]$$

and measure [19] reduces to:

$$\hat{c}_I(\theta) = \hat{c}_I - e_I^{*T} (I_p - \tilde{H}_I)^{-1} \tilde{G}_I (I_p - \tilde{H}_I)^{-1} e_I^* \qquad [21]$$

where $\tilde{G}_I = \tilde{X}_2 (\tilde{X}_2^T \tilde{X}_2)^{-1} \tilde{X}_2$ and $\tilde{X}_2$ is a submatrix that contains the last m columns of $\tilde{X}$.

6. The masking problem

Masking occurs when a single outlier is not detected by using individual influence measures because of the presence of other outliers. The problem arises because these influential subsets can only be identified by looking at them jointly. Measures derived in section 5 can be used to analyze the influence of any subset of observations but, obviously, the identification of influential groups would involve serious combinatorial difficulties. Therefore, the objective of this section is to propose a procedure to identify influential subsets in an efficient way. Specifically, we extent the method suggested by Peña and Yohai (1995) for linear regression to the case of binary choice models. This method is based on looking at the eigenvalues of an influence matrix M, which is defined as the uncentred covariance of a set of vectors which represent the effect on the fit of the deletion of each data point. In linear regression, the ijth element of the $n \times n$ influence matrix M is given by $m_{ij} = (\hat{y} - \hat{y}_{(i)})^T (\hat{y} - \hat{y}_{(j)})$, where $\hat{y}_{(i)}$ and $\hat{y}_{(j)}$ are the vectors of fitted values when the ith and jth data points are deleted. Based on a heuristic justification, Peña and Yohai show that the eigenvectors linked to non-null eigenvalues of the influence matrix can be used to identify influential subsets. This method has been tested in linear models and it seems to work very well. It is also computationally efficient in comparison to other procedures proposed in the literature for linear models.

To construct an analogous matrix $\tilde{M}$ to identify influential subsets in binary choice models, we just have to apply the procedure suggested by Peña and Yohai (1995) to equation [5]. Therefore, $\tilde{M}$ must be understood in terms of the fit of $\tilde{y}_i$ in [5]. This is appropriated because equation [5] is a good linear approximation to model [2]. Note that [5] is obtained by expanding $F(x_i^T\beta)$ in a first order Taylor series around $\hat{\beta}$ and higher order terms evaluated at $\hat{\beta}$ converge in probability to zero. From [8] and [9] we obtain that:

$$\hat{\beta} - \hat{\beta}_{(i)} = \frac{e_i^* (\tilde{X}^T\tilde{X})^{-1}\tilde{x}_i}{1 - \tilde{h}_i} \qquad [22]$$

Consequently, if we denote by $\hat{\tilde{y}}_j$ the fitted value in [5] for observation j and by $\hat{\tilde{y}}_{(i)j}$ the new fitted value when observation i is deleted, we have:

$$\hat{\tilde{y}}_j - \hat{\tilde{y}}_{(i)j} = \frac{e_i^* \tilde{h}_{ij}}{1 - \tilde{h}_i} \qquad [23]$$

Therefore if $\hat{\tilde{y}}_{(i)} = (\hat{\tilde{y}}_{(i)1} \dots \hat{\tilde{y}}_{(i)n})^T$, the vector $\tilde{t}_i = \hat{\tilde{y}} - \hat{\tilde{y}}_{(i)}$ summarizes the effect on the fit of deleting observation i.

Then, to detect possible sets of influential observations having similar or opposite effects on the fit, we look at the uncentred covariance matrix of the $\tilde{t}_i$. Being $\tilde{T}$ the $n \times n$ matrix $\tilde{T} = (\tilde{t}_1 \dots \tilde{t}_n)$ whose columns are the vector $\tilde{t}_i$, the $n \times n$ influence matrix $\tilde{M}$ is defined as $\tilde{M} = \tilde{T}^T\tilde{T}$. Therefore, the ijth element of $\tilde{M}$ is given by:

$$\tilde{m}_{ij} = \tilde{t}_i^T\tilde{t}_j = \frac{e_i^* e_j^* \tilde{h}_{ij}}{(1 - \tilde{h}_i)(1 - \tilde{h}_j)} \qquad [24]$$

Observe that the diagonal elements of $\tilde{M}$ are measure [10] for each observation in the sample. Note also that measure $\hat{c}_I$ in [17] can be approximated as $\hat{c}_I \approx \sum_{i \in I} \sum_{j \in I} m_{ij}$.

Once matrix $\tilde{M}$ is specified, the procedure to identify influential subsets is analogous to the case of linear models. Since the rank of $\tilde{M}$ is k, this procedure is as follows. Step 1: find the eigenvectors corresponding to the k non-null eigenvalues of the influence matrix $\tilde{M}$. Step 2: consider the eigenvectors corresponding to the m larger eigenvalues for $m \leq k$. Then, a group of candidate outliers is obtained by searching on those eigenvectors either for a set of co-ordinates with relatively large weight and the same sign (observations with similar effects on the fit) or for a set of co-ordinates with relatively large weight and different signs (observations with opposite effects on the fit). Step 3: given the sets of candidate outliers, use measures in section 5 to evaluate the influence of those sets.

7. A Monte Carlo study

In this section the behavior of previous measures is analyzed by using simulated data. Consider a probit model with the following data generating mechanism:

$$y_i^* = \beta_1 + \beta_2 x_i + u_i \quad [25]$$

$$P_i = P(y_i = 1 \mid x_i) = P(y_i^* \geq 0 \mid x_i) = F(\beta_1 + \beta_2 x_i) \quad [26]$$

where $u_i \sim iid\, N(0,1)$, $x_i \sim iid\, N(0,1)$, $\beta^T = (-0.65, 1)$ and F is the standard normal cumulative distribution. Each sample contains data generated by this mechanism as well as a fixed proportion ω of outliers, being the total sample size of 200. Two cases are considered. *Case 1*: outliers come from a normal distribution with the same variance as the good observations but a different mean; in particular, $y_i^* \sim iid\, N(x_i^T\gamma, 1)$ with $\gamma = (1, -0.5)$. *Case 2*: outliers come from a normal distribution with the same mean as the good observations but with variance greater than 1; in particular, $y_i^* \sim iid\, N(x_i^T\beta, 7)$.

We first analyze the behaviour of measure $\hat{c}_i$ in [10]. Similarly to the case of Cook's statistic for linear models, this measure does not have a known distribution. Therefore, to determine critical points we use a method based on a type I error probability: the frequency of rejecting the null hypothesis when it is true. Specifically, in a sample without outliers, we find the values $\hat{c}_\alpha^*$ such that $\hat{c}_i > \hat{c}_\alpha^*$ for a *given proportion* α of observations. To find these values samples have been generated as described above while components of vector β come from standard normal variables. The results based on 500 replications are reported in Table 1. Note that the values of $\hat{c}_\alpha^*$ corresponding to $\alpha=1\%$ and $\alpha=10\%$ imply respectively changes of 5% and 1% in the confidence region of a χ_2^2 distribution.

Table 2 contains the Monte Carlo results for $\hat{c}_i$ based on 500 replications. For each ω and the critical values of $\alpha=1\%$ and $\alpha=5\%$, the mean of the following measures are included: the mean squared error of the model once observations for which $\hat{c}_i > \hat{c}_\alpha^*$ have been eliminated ($MSE_{(I)}$) and, in parentheses, the MSE of the model with the whole sample; the number of observations identified as outliers by $\hat{c}_i$ (OBS); and the number of true outliers among those identified by $\hat{c}_i$ (ANOBS). From this table we remark the following points:

i) The design of the experiment obligates $\hat{c}_i$ to detect a fixed number of observations as outliers for all ω. Then, when there is no outlier in the sample ($\omega=0$), $MSE_{(I)}$ is higher than MSE. This is because $\hat{c}_i$ identifies as outliers observations in the tails of the distribution, being those observations essential for the estimation of the model. Obviously, the difference between both MSE is greater for $\alpha=5\%$ than for $\alpha=1\%$ since in the first case $\hat{c}_i$ must eliminate a higher number of observations.

ii) In *case I*, where outliers are more dangerous, the behaviour of $\hat{c}_i$ seems satisfactory. $MSE_{(I)}$ is lower than MSE for $\omega \geq 5$, while as ω increases the ratio between ANOBS and OBS is getting close to one, meaning that the outliers detected by $\hat{c}_i$ are true outliers. For $\omega=20$ and $\omega=30$, $\hat{c}_i$ detects less outliers than for smaller values of ω, which might be due to a masking effect.

iii) In *case II*, where outliers are less dangerous, the results of the experiment are different. For small values of ω, $MSE_{(I)}$ is higher than MSE because $\hat{c}_i$ still classifies observations in the tails of the distribution as outliers. The results improve as the proportion of outliers in the sample increases, being considerably better for $\alpha=1\%$ with respect to $\alpha=5\%$.

These results come from a very simple simulated experiment and they do not intend to be conclusive. However they do show that, for a fixed critical value, the effectiveness of $\hat{c}_i$ depends on both the number and magnitude of outliers in the sample. Then, in empirical work, the best way to use this measure seems to consider no critical value, but instead to analyze the observations with relatively highest values of $\hat{c}_i$. It is also important to note that in this experiment some of the true outliers detected by $\hat{c}_i$ show a large residual, but it is also the case of a proportion of observations generated by the true model. This fact supports the discussion of section 3. An analogous experiment has been carried out for measure $\hat{c}_i(P)$ in [13]. The results are not presented here, but they suggest that $\hat{c}_i(P)$ behaves analogously to $\hat{c}_i$.

With respect to masking, the behaviour of the method based on the influence matrix $\bar{M}$ in [24] is analyzed. The data generating mechanism is also given by [25] and [26] and the same types of outliers as in the previous experiment are considered. However, this method requires the intervention of the analyst to identify influential subsets, which complicates any simulation. Then a way to identify a group I of potential outliers is as follows: $i \in I$ if in any of the two eigenvectors corresponding to the non-null eigenvalues of $\bar{M}$, observation i shows a component greater than a fixed constant c in absolute value. This is not the correct way to apply this method, but it is enough to show that it works well in the case of binary choice models.

Table 3 presents for c=0.15 the mean of the following measures for 500 replications: the value of $\hat{c}_I$ in [17] for the identified group I; the $MSE_{(I)}$ of the model once set I is eliminated and, in parentheses, the MSE including group I; the number of observations in group I (OBS); and the number of true outliers among those in OBS (ANOBS). From this table, it is clear that, even if the method is not used properly, it works well in situations where masking is likely to exists; i.e. when ω is high. Particularly, in both *case I* and *II*, for ω=25% and ω=30% the MSE reduces to half when group I is eliminated from the sample. However, when the proportion of outliers is small or zero, the bad use of the method eliminates good observations in the tails of the distribution, making $MSE_{(I)}$ to be higher than MSE. Note that, since the eigenvectors are normalized, a systematic number of observations (approximately 12) is eliminated even if there is no outlier in the sample.

Finally, it is important to note that in all the experiments similar results are obtained for the logit case. Also for a sample side of 400 observations and/or for values of x_i being generated by a continuous uniform distribution.

8. Application: Choosing between fixed and adjustable rate mortgages

Using a probit model, Dhillon et al. (1987) examine the choice between fixed and adjustable rate mortgages as a function of borrower characteristics. The main objective of the study is to analyze empirically the two views that have dominated the theoretical literature: one that emphasizes the irrelevance of borrower characteristics in the mortgage choice decision given the prices and terms of the contract, and one that focuses on the potential impact of borrower characteristics assuming asymmetric information. The explanatory variables are defined in Table 4 and the sample contains 78 observations with 46 fixed rate and 32 adjustable rate loans.

Table 5 contains the ML estimates of the three model specifications considered in the original article: the unconstrained model which includes all the variables in the analysis (model 3) and two constrained models, one excluding the LA and STL variables (model 1) and one excluding all the borrower characteristics (model 2). The results can be summarized as follows:

i) The price variables are clearly significant and have the expected sign. The only exceptions are YLD and MAT that are insignificant in the unconstrained model.

ii) Borrower characteristics are not significant individually, although taken together have a significant effect. The likelihood ratio tests between models 2 and 3 rejects the hypothesis of a zero joint influence.

iii) The coefficients of the borrower economic variables (LA and STL) are insignificant. In addition, the hypothesis that both coefficients are jointly zero cannot be rejected (the likelihood ratio between models 1 and 3 is 1.60).

The estimation results are then inconclusive about the impact of borrower characteristics. The authors point out that "...the pricing variables play a dominant role in the choice decision. The evidence also suggests that households with co-borrowers, married couples, and short expected housing tenures have a greatest probability of taking out adjustable rate mortgages.

However, in general, borrower characteristics do not significantly influence the choice" (Dhillon et al., 1987, pp. 265).

In order to analyze if this sample contains outliers, we apply the following diagnostic strategy. In a first step, we consider similar instruments to those used in linear regression (Belsley et al., 1981; Peña and Ruiz-Castillo, 1984) to measure the distance of an observation x_i to the center of gravity of the scatter of points X. In particular $h_i = x_i^T(X^TX)^{-1}x_i$ together with $\tilde{h}_i = \tilde{x}_i^T(\tilde{X}^T\tilde{X})^{-1}\tilde{x}_i$ that considers the transformed variables in [7] evaluated at the ML estimates. Although $\tilde{h}_i$ is analogous to h_i, it may incorporate additional information due to the nonlinearity of the model. These measures are not conclusive, but sample points with relatively high h_i or $\tilde{h}_i$ are potentially influential, so we may pay special attention to them in further steps. Next we obtain $\hat{c}_i$ in [10] and $\hat{c}_i(P)$ in [13]. High values of these measures indicate the presence of individual outliers. Finally, it should be convenient to consider the possibility of masking. Groups of candidate outliers can be obtained by the method of section 6 using the matrix in [24], while the influence of those groups can be measured by expressions [17] and [18].

Table 6 presents the values of the diagnostic instruments for a selected set of observations worthy of attention. These instruments have been computed from the estimation of the unconstrained model (model 3). By columns, this table contains: the distance h_i, the distance for the transformed variables $\tilde{h}_i$, the residual e_i, the individual influence measures $\hat{c}_i$ and $\hat{c}_i(P)$, and the components of the eigenvectors associated to the two largest eigenvalues of the influence matrix $\tilde{M}$. Since both h_i and $\tilde{h}_i$ are only valid for continuous variables, they have been computed from the continuous explanatory variables in the sample. The following should be emphasized:

i) Observations 14 and 37 are the potentially more influential with the largest h_i and $\tilde{h}_i$ values.

ii) The atypical nature of observation 37 is confirmed by the value of $\hat{c}_i$ which is extremely large. The $\hat{c}_i$ value of observation 14 is considerably smaller, but it is also large in relation to other sample points. In addition, given that the mean of $\hat{c}_i$ for the whole sample is around 4 and the same mean when observations 14 and 37 are eliminated is approximately 1.5, it seems clear that both observations 14 and 37 are highly influential. Taken together, the joint influence value $\hat{c}_I$ is 176.7. The measure $\hat{c}_i(P)$ also presents the largest values for both observations.

iii) Although points 14 and 37 can be classified as outliers, the associated residuals e_i are not large, specially for observation 37 which is -0.23. On the contrary, observations with a residual closed to one in absolute value (numbers 22, 26, 55, 68, and 76) do not show any evidence of being outliers when looking at the influence measures.

iv) From the last two columns of the table, it is evident that both eigenvectors are respectively dominated by observations 14 and 37. Therefore the two outliers are again clearly identified: they appear in the eigenvector corresponding to the largest eigenvalues with large values and the same sign, whereas the rest of the points are given zero weight. The only exception is point 61 with a component value of 0.1577 in the first eigenvector. This observation does not show any evidence of being anomalous by looking at the individual influence measures, but when observations 14 and 37 are eliminated from the sample, the $\hat{c}_i$ value for observation 61 rises to 3.71, which is large in relation to the remainder sample points. This suggests that the presence of points 14 and 37 were masking the influence of point 61.

In view of this information, we reestimate the probit model without observations 14, 34 and 61, obtaining the results presented in Table 7. The main differences with respect to estimates in Table 5 are the following:

i) Borrower characteristics have a stronger influence on the mortgage choice decision. The variables BS, CB, SE, MOB and NW become significant at the 95% level in the unconstrained model. Also the value of the likelihood test between models 2 and 3 is almost double than the one obtained in Table 5.

ii) The borrower economic variables (LA and STL) are still insignificant individually, but the hypothesis that both coefficients are jointly zero is rejected with a likelihood ratio of 25.32.

iii) Estimates of model 2 do not show significant changes with respect to those in Table 5, which suggests that outliers come from borrower characteristics. In fact observation 37 is a "rich" borrower with extremely high values of the NW, LA y STL variables that chooses a fixed rate mortgage, while observation 14 is a relatively old individual with a low net worth, first-time homebuyer, that prefers an adjustable rate mortgage.

These results suggest that borrower characteristics do have a relevant influence on the mortgage choice decision, which is consistent with the notion of asymmetric information or incomplete capital markets. This analysis does not try to present evidence about the choice between fixed and adjustable rate mortgages; it is only an example to illustrate a diagnostic strategy and to show how outliers can mislead the estimation results in a probit model.

9. Final remarks

This paper has focused on the problem of outlying observations in binary response models. Both a Monte Carlo study and an empirical application have shown the following points: 1) Outliers are not necessarily observations with a residual close to one in absolute value. In fact highly influential observations may show a very small residual, while observations from the tails of the true distribution may present a large residual. 2) The two individual measures proposed to evaluate the effect of each observation on the fit seem to behave well. The procedure based on the influence matrix to detect masked observations also works well, being useful to identify influential subsets when masking is present.

The empirical application illustrates that few outliers in the sample might affect seriously the estimation results. This example is also used to present a diagnostic strategy for detection of outliers in binary response models. First, potentially influential points are detected by measuring the distance of each data point to the center of gravity of the scatter of points. Then individual influence measures are computed to identify those points that are truly influential. Finally, the possibility of masking should be checked.

References

Amemiya, T. (1981), "Qualitative Response Models: A Survey", *Journal of Economic Literature*, XIX, 1483-1536.

Amemiya, T. (1985), *Advanced Econometrics*, Oxford, Basil Blackwell Ltd.

Bedrick, E. J., and Hill, J. R. (1990), "Outlier Tests for Logistic Regression, a Conditional Approach", *Biometrika*, 77, 4, 815-827.

Belsley, D. A., Kuh, E. and Welsch, R. E. (1981), *Regression Diagnostics. Identifying Influential Data and Sources of Collinearity*, New York, John Wiley & Sons.

Box, G. E. P., and Tiao, G. C. (1968), "A Bayesian Approach to some Outlier Problems", *Biometrika*, 55, 1, 119-129.

Cook, R. D. (1977), "Detection of Influential Observation in Linear Regression", *Technometrics*, 19, 1, 15-18.

Cook, R. D. and Weisberg, S. (1982), *Residuals and Influence in Regression*, New York: Chapman and Hall.

Copas, J. B. (1988), "Binary Regression Models for Contaminated Data", *Journal of the Royal Statistical Society*, B, 50, 2, 225-265.

Dhillon, U. S., Shilling, J. D. and Sirmans, C.F. (1987), "Choosing between Fixed and Adjustable Rate Mortgages", *Journal of Money, Credit and Banking*, 19, 1, 260-267.

Guttman, I. (1973), "Premium and Protection of Several Procedures for Dealing with Outliers when Sample Sizes are Moderate to Large", *Technometrics*, 15, 385-404.

Jennings, D. E. (1986), "Outliers and Residual Distributions in Logistic Regression", *Journal of the American Statistical Association*, 81, 396, 987-990.

McCullagh, P. and Nelder, J. A. (1983), *Generalized Linear Models*, London: Chapman and Hall, Inc.

McFadden, D.L. (1983), "Econometric Analysis of Qualitative Response Models", *Handbook of Econometrics*, Griliches, I. and Intrilligator, M. eds., 2, 24, 1396-1457.

Nelder, J. A. and Wedderburn, R. W. M. (1972), "Generalized Linear Models", *Journal of the Royal Statistical Society*, A, 135, 370-384.

Peña, D. and Ruiz-Castillo, J. (1984), "Robust Methods of Building Regression Models - An Application to the Housing Sector", *Journal of Business and Economic Statistics*, 2, 1, 10-20.

Peña, D. and Yohai, V. J. (1995), "The Detection of Influential Subsets in Linear Regression using an Influence Matrix", *Journal of the Royal Statistical Society*, B, 57, 2.

Pierce, D. A. and Schafer, D.V. (1986), "Residuals in Generalized Linear Models", *Journal of the American Statistical Association*, 81, 977-986.

Pregibon, D. (1981), "Logistic Regression Diagnostics", *The Annals of Statistics*, 9, 4, 705-724.

Rao, C. R. (1973), *Linear Statistical Inference and its Applications*, 2nd. ed., John Wiley.

Williams, D. A. (1987), "Generalized Linear Model Diagnostics: The Deviance and Single Case Deletion", *Applied Statistics*, 36, 2, 181-191.

Figure 1. Examples of points with a large residual in a probit model

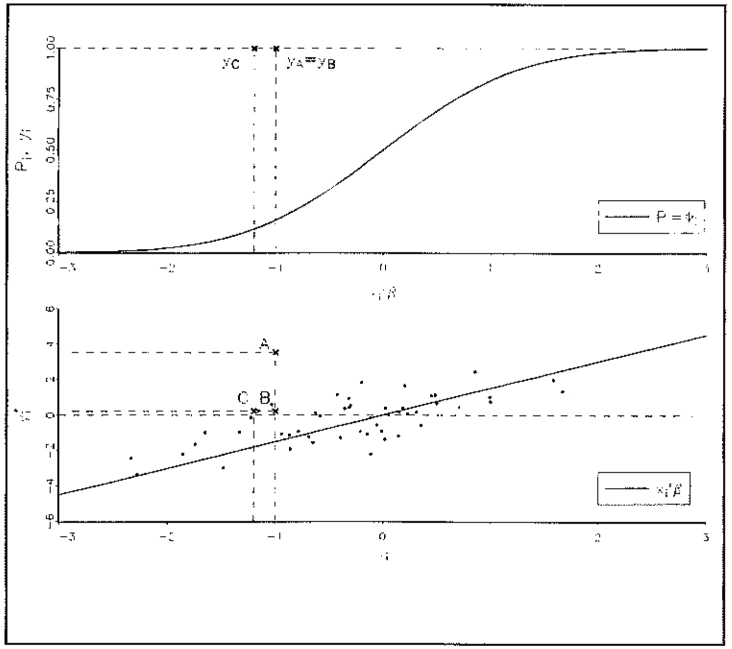

Table 1. Values of $\hat{c}_\alpha^*$ for different α

α	$\hat{c}_\alpha^*$
1.0	0.1133 (0.0459)
2.5	0.0636 (0.0226)
5.0	0.0382 (0.0127)
10.0	0.0202 (0.0055)

Table 2. Monte Carlo results for $\hat{c}_t$

	ω%	α = 1%, $\hat{c}_\alpha^*$ = 0.1133			α =5%, $\hat{c}_\alpha^*$ = 0.0382		
		ECM_m (ECM)	OBS	ANOBS	ECM_m (ECM)	OBS	ANOBS
	0	0.10 (0.03)	2.55	--	0.26 (0.03)	9.95	--
Case 1	5	0.04 (0.08)	2.68	1.59	0.07 (0.07)	9.52	3.84
	10	0.10 (0.18)	2.56	2.04	0.06 (0.18)	9.63	5.50
	15	0.20 (0.31)	2.52	2.31	0.11 (0.31)	9.34	6.68
	20	0.35 (0.44)	2.09	1.96	0.20 (0.44)	9.16	7.25
	30	0.67 (0.72)	1.19	1.13	0.48 (0.72)	8.17	6.95
Case 2	5	0.07 (0.04)	2.59	0.57	0.20 (0.04)	9.87	1.44
	10	0.05 (0.04)	2.51	1.11	0.14 (0.04)	9.66	2.57
	15	0.05 (0.06)	2.47	1.30	0.11 (0.06)	9.72	3.54
	20	0.05 (0.07)	2.44	1.54	0.08 (0.07)	9.65	4.56
	30	0.07 (0.12)	2.35	1.85	0.06 (0.12)	9.58	6.00

Table 3. Monte Carlo results for masking

	ω%	$\hat{c}_t$	ECM_m (ECM)	OBS	ANOBS
Case 1	0	10.70	0.356 (0.035)	14.34	--
	5	11.05	0.109 (0.077)	12.89	3.77
	10	11.03	0.067 (0.180)	11.80	6.12
	15	11.10	0.105 (0.307)	12.64	7.55
	20	11.93	0.182 (0.443)	13.20	9.15
	25	10.13	0.289 (0.574)	11.35	8.86
	30	9.96	0.395 (0.723)	12.46	10.00
Case 2	0	10.55	0.346 (0.034)	14.43	--
	5	10.47	0.255 (0.037)	13.73	1.68
	10	10.87	0.207 (0.043)	13.58	3.29
	15	10.57	0.127 (0.056)	13.28	4.55
	20	16.06	0.092 (0.070)	13.44	5.98
	25	11.16	0.086 (0.094)	13.68	6.36
	30	10.55	0.067 (0.120)	13.02	7.36

Table 4. Variables in the mortgage choice decision model

Dependent variable	
ADJ	takes value one if borrower chooses an adjustable rate and zero otherwise
Exogenous variables: market conditions and terms of the contract	
FI	Fixed interest rate
MAR	Margin on the adjustable rate mortgage
YLD	The difference between the 10-years Treasury rate less 1-year Treasure rate
PTS	The ratio of points paid on adjustable to fixed rate mortgages
MAT	The ratio of maturities on adjustable to fixed rate mortgages
Exogenous variables: individual borrower characteristics	
BA	Age of the borrower
BS	number of years of school
FTB	A dummy variable of one if the borrower is a first-time homebuyer, zero otherwise
CB	A dummy variable of one if there is a co-borrower, zero otherwise
MC	A dummy variable of one is the borrower is married, zero otherwise
SE	A dummy variable of one is the borrower is self-employed, zero otherwise
MOB	Number of years at present address
NW	Net worth of the borrower
Exogenous variables: economic borrower characteristics	
LA	Liquid assets
STL	Short-term liabilities

Table 5. Estimation of the original models in Dhillon et al. (1987)

Variable	Model 1	Model 2	Model 3
Constant	-3.4855	-1.8774	-3.1077
	(5.2870)	(4.2249)	(5.8775)
FI (Fixed interest rate)	0.9786	0.4987	1.0081
	(0.3911)	(0.2772)	(0.4107)
MAR (Margin)	-0.6268	-0.4310	-0.7052
	(0.2588)	(0.1736)	(0.2723)
YLD (r_{10}-r_1)	-2.2381	-2.3840	-2.5251
	(1.4310)	(1.0880)	(1.5881)
PTS (Points)	-0.7226	-0.2999	-0.8303
	(0.3753)	(0.2415)	(0.3977)
MAT (Maturity)	-1.1366	-0.0592	-1.1644
	(0.8927)	(0.6147)	(0.8946)
BA (Age)	-0.0031	--	-0.0040
	(0.0390)		(0.0429)
BS (School)	-0.1094	--	-0.1083
	(0.0967)		(0.0998)
FTB (First-time homebuyer)	0.2398	--	0.1434
	(0.5208)		(0.5583)
CB (Co-borrower)	-0.8061	--	-1.0666
	(0.6044)		(0.6922)
MC (Married couples)	-1.0358	--	-1.0586
	(0.6557)		(0.6728)
SE (Self-employed)	-0.5906	--	-1.1275
	(1.2238)		(1.5598)
MOB (Mobility)	-0.0882	--	-0.0930
	(0.0521)		(0.0550)
NW (Net worth)	0.1349	0.0838	0.1288
	(0.0901)	(0.0422)	(0.1053)
LA (Liquid assets)	--	--	0.0146
			(0.0350)
STL (Short-term liab.)	--	--	0.0161
			(0.0283)
ln ℓ (log likelihood)	-31.53	-39.21	-30.73
LRT (likelihood ratio)	1.60	16.96	

Numbers in parentheses are estimated standard errors

Table 6. Diagnostic instruments for a selected set of observations

i	h_i	$\tilde{h}_i$	e_i	$\hat{c}_i$	$\hat{c}_i(P)$	λ_i^1	λ_i^2
5	0.2006	0.3072	0.5854	0.9040	0.1877	0.0460	-0.0010
14	0.3737	0.6298	0.6990	10.6700	2.0910	-0.9762	0.0118
15	0.1491	0.5046	0.6070	3.1750	0.6207	-0.0025	-0.0075
22	0.1905	0.2090	-0.8115	1.4380	0.2280	0.0339	0.0001
23	0.1531	0.2305	-0.5074	0.4009	0.0887	0.0094	0.0007
24	0.1531	0.2305	-0.5074	0.4009	0.0887	0.0094	0.0007
25	0.1531	0.2305	-0.5074	0.4009	0.0887	0.0094	0.0007
26	0.1905	0.2090	-0.8115	1.4380	0.2280	0.0339	0.0001
35	0.0987	0.3097	-0.5198	0.7037	0.1162	-0.0264	-0.0004
37	0.8889	0.9675	-0.2322	277.2000	49.1000	-0.0103	-0.9998
45	0.3373	0.0315	-0.0008	0.0000	0.0000	0.0001	0.0000
46	0.3579	0.0570	-0.0020	0.0001	0.0000	0.0015	0.0000
53	0.1101	0.3102	-0.3445	0.3426	0.0643	-0.0076	0.0001
55	0.0604	0.1820	-0.8855	2.1030	0.2963	0.0194	0.0053
58	0.1353	0.4689	-0.6062	2.5580	0.4804	0.0160	0.0070
59	0.4819	0.5944	-0.1261	0.5211	0.0636	0.0338	0.0013
61	0.1650	0.3417	-0.3911	0.5066	0.0938	0.1577	0.0021
62	0.1456	0.4145	-0.4997	1.2070	0.2240	-0.0379	0.0000
63	0.3541	0.0832	-0.0042	0.0004	0.0001	0.0028	0.0000
64	0.0637	0.2432	-0.4694	0.3757	0.0715	0.0250	0.0021
67	0.1159	0.4599	-0.5995	2.3600	0.4387	0.0022	0.0027
68	0.0976	0.1798	0.8874	2.1060	0.3470	0.0815	0.0047
69	0.0996	0.3206	0.6719	1.4220	0.2577	0.0102	-0.0023
71	0.1208	0.1890	0.7418	0.8257	0.1454	0.0478	0.0037
76	0.1211	0.2318	0.8253	1.8560	0.3052	-0.0180	0.0036
77	0.1890	0.4327	0.4205	0.9757	0.2028	-0.0072	0.0010
78	0.1482	0.3224	0.3790	0.4286	0.0790	0.0343	0.0010

Table 7. Estimation results when observations 14, 37 and 61 are eliminated

Variable	Model 1	Model 2	Model 3
Constant	-6.5240	-0.9841	-5.5329
	(5.6853)	(4.3099)	(6.2820)
FI (Fixed interest rate)	1.0681	0.5474	1.1339
	(0.4212)	(0.2869)	(0.4488)
MAR (Margin)	-0.1585	-0.5126	-0.2064
	(0.2764)	(0.1992)	(0.3082)
YLD (r_{10}-r_1)	-0.8826	-3.0185	-0.5817
	(1.7310)	(1.2795)	(1.9498)
PTS (Points)	-1.1446	-0.3821	-1.1011
	(0.4885)	(0.2769)	(0.4857)
MAT (Maturity)	-0.0807	-0.2079	-0.2943
	(1.1683)	(0.6590)	(1.1719)
BA (Age)	-0.0127	--	-0.0602
	(0.0565)		(0.0664)
BS (School)	-0.1728	--	-0.2311
	(0.1086)		(0.1156)
FTB (First-time homebuyer)	-0.9465	--	-1.4168
	(0.7223)		(0.9045)
CB (Co-borrower)	-1.1802	--	-1.8438
	(0.6920)		(0.9051)
MC (Married couples)	-0.2378	--	-0.4495
	(0.7130)		(0.7130)
SE (Self-employed)	-0.6103	--	-4.2153
	(1.4365)		(2.4431)
MOB (Mobility)	-0.7847	--	-0.7308
	(0.2765)		(0.2861)
NW (Net worth)	0.5290	0.0803	0.6143
	(0.1982)	(0.0439)	(0.2305)
LA (Liquid assets)	--	--	0.0699
			(0.0762)
STL (Short-term liab.)	--	--	0.0614
			(0.0372)
ln ℓ (log likelihood)	-42.80	-39.41	-22.39
LRT (likelihood ratio)	25.32	32.10	

Numbers in parentheses are estimated standard errors