



THE MEASURE OF FIT AND CUT-OFF POINT FOR A BINOMIAL LOGIT MODEL. AN OVERLOOKED CONTRIBUTION OF JAN SALOMON CRAMER

Marek Gruszczyński

SGH Warsaw School of Economics

e-mail: marek.gruszczyński@sgh.waw.pl

ORCID: 0000-0001-8946-0713

Received 14.05.2025, Revised 16.01.2026, Accepted 16.01.2026

Abstract

Research background: This paper revisits the contribution of Jan Salomon (Mars) Cramer (1928–2014) to binomial logit modelling. Cramer proposed both a novel measure of fit for binomial logit models and a cut-off point for binary classification, particularly suitable for unbalanced samples.

Purpose: To present and demonstrate J.S. Cramer's often-overlooked proposals for a new measure of fit and a cut-off point in binomial logit models. Additionally, the paper discusses the relevance of Cramer's contributions in the current discourse on binomial logit modelling.

Research methodology: The study is based on a literature analysis aimed at establishing the current state of knowledge regarding J.S. Cramer's contributions to binomial logit modelling, including the citation analysis using Scopus.

Results: We highlight that the measure of fit proposed by Cramer in 1999, known as Cramer's *lambda*, has gained recognition in the literature under a different name – Tjur's *D* – introduced a decade later (2009). We also note that Cramer's proposed cut-off point, *alpha*, is rarely cited as an alternative to more commonly referenced thresholds. These arguments are illustrated with a numerical example.

Novelty: Both of J.S. Cramer's proposals remain relatively unknown today. We hope this paper contributes to re-establishing Mars Cramer's legacy in econometrics. Cramer's *lambda* and Cramer's *alpha* deserve recognition and adoption as legitimate standards for binomial logit modelling.

Keywords: goodness of fit, binomial logit, predicted probabilities, cut-off point

JEL classification: C25, C52

Introduction

This paper presents the overlooked contribution of Jan Salomon (Mars) Cramer, a microeconomerician from the University of Amsterdam to binomial logit modelling. His contribution consists of a novel measure of fit and a cut-off point in binary classification. We examine the fate of both proposals in practical applications, noting that Cramer's measure of fit *lambda* is currently in use under the different name – Tjur's *D* – while Cramer's cut-off point *alpha* has been largely forgotten, despite being a natural standard, particularly for unbalanced samples. There is no clear justification for why the current software standards (e.g. Stata) continue to rely on the conventional cut-off value of 0.5 rather than incorporating Cramer's *alpha*.

The paper is organized as follows. Section 1 introduces Cramer's measure of fit, *lambda*, its treatment in econometrics textbooks, and its re-emergence in the literature under the name Tjur's *D*. Section 2 explores the connection of Cramer's measure of fit to the confusion matrix, and critiques a misguided proposal in the literature to directly link *lambda* (or *D*) to Youden's Index *J*. Section 3 discusses Cramer's cut-off point, *alpha*, its lack of recognition in the literature, and its current relevance to practical applications of binomial logit models. Section 4 provides examples illustrating both of Cramer's proposals for binomial logit models, alongside commonly used competing indices and cut-off points. Section 5 addresses binomial logit modelling for rare events (i.e., highly unbalanced samples), with the focus on the application of Firth's penalized maximum likelihood estimation method. Section 6 repeats examples of section 4, with the binomial models estimated with the Firth method. In the Conclusion, we emphasize the continuing relevance and importance of Cramer's proposals in today's applied econometric practice.

1. Cramer's *lambda*

In 1999 Mars (J.S.) Cramer proposed a new measure of fit for the binomial (binary) logit model, particularly useful in cases involving unbalanced samples (Cramer 1999). He referred to it as a *measure of discrimination*. Cramer's proposal did gain widespread recognition but remained largely overlooked for years. The measure was later "rediscovered" by Tjur (2009) who introduced it without referencing Cramer's original contribution.

In this section we briefly present Cramer's concept, along with some relevant historical context.

Assume the binary outcome $Y_i = 1$ or $Y_i = 0$ with p_i representing the probability that $Y_i = 1$ for observations $i = 1, 2, \dots, n$. The estimated form of the binomial logit model is given by:

$$\hat{p}_i = F(\mathbf{x}_i' \hat{\boldsymbol{\beta}}) = \frac{\exp(\mathbf{x}_i' \hat{\boldsymbol{\beta}})}{1 + \exp(\mathbf{x}_i' \hat{\boldsymbol{\beta}})} = \frac{1}{1 + \exp(-\mathbf{x}_i' \hat{\boldsymbol{\beta}})} \quad (1)$$

and

$$\frac{\hat{p}_i}{1 - \hat{p}_i} = \exp(\mathbf{x}_i' \hat{\boldsymbol{\beta}}) \quad \text{and} \quad \ln \frac{\hat{p}_i}{1 - \hat{p}_i} = \mathbf{x}_i' \hat{\boldsymbol{\beta}}$$

where, typically, $\mathbf{x}_i' \hat{\boldsymbol{\beta}}$ is a linear function of explanatory variables X_j ($j = 1, 2, \dots, k$):

$$\mathbf{x}_i' \hat{\boldsymbol{\beta}} = (1 X_{1i}, X_{2i}, \dots, X_{ki}) \cdot (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k)'$$

and $\hat{\beta}_j$ ($j = 0, 1, 2, \dots, k$) are maximum likelihood estimates of logit model's parameters β_j .

The $\hat{p}_i$'s are estimates of probabilities p_i . Let us now denote:

$\hat{p}_{i1}$ is the estimated probability of the observed outcome $Y_i = 1$ ($i = 1, 2, \dots, n_1$),

$\hat{p}_{i0}$ is the estimated probability of the observed outcome $Y_i = 0$ ($i = 1, 2, \dots, n_0$);

$$n_1 + n_0 = n.$$

Now, let us assume that α is the relative share of outcomes $Y_i = 1$ in the sample, and therefore $1 - \alpha$ is the relative share of outcomes $Y_i = 0$ in the sample. Cramer writes: "whenever the two sample shares are unequal, α is by convention the larger share and the corresponding outcome is labelled $Y_i = 1$ ". We shall see the importance of this in a moment.

Note that for the logit model the overall average of estimated probabilities $\hat{p}_i$ is equal to relative share α :

$$\bar{P} = \frac{1}{n} \sum \hat{p}_i = \alpha \quad (2)$$

Cramer refers to this unique property of binomial logit as the "equality of the means". The average predicted probabilities $\hat{p}_i$ for the two outcome groups $Y_i = 1$ and $Y_i = 0$ are:

$$\bar{P}^1 = \frac{1}{n_1} \sum \hat{p}_{i1} \quad \text{and} \quad \bar{P}^0 = \frac{1}{n_0} \sum \hat{p}_{i0}$$

Then the average predicted probability $\bar{P}$ is a weighted average of the group-specific averages:

$$\alpha \bar{P}^1 + (1 - \alpha) \bar{P}^0 = \bar{P} = \alpha \quad (3)$$

Note that $\alpha = \frac{n_1}{n}$, and thus

$$\frac{1 - \bar{P}^1}{\bar{P}^0} = \frac{n_1}{n_0} \quad (4)$$

Cramer argues that “if the fitted model has any explanatory power, $\bar{P}^1$ exceeds $\bar{P}^0$, and the two will lie on either side of their (weighted) average α .” By convention, α corresponds to the larger share, attributed to the outcome $Y_i = 1$. The average probabilities $\bar{P}^1$ and $\bar{P}^0$ are thus constrained: $\bar{P}^1$ within interval $[\alpha, 1)$ and $\bar{P}^0$ within interval $(0, \alpha]$. Consequently, $\bar{P}^1$ may be written as a linear combination of α and 1 with nonnegative weights $1 - \lambda$ and λ :

$$\bar{P}^1 = (1 - \lambda)\alpha + \lambda \quad (5)$$

Similarly, $\bar{P}^0$ may be written as a linear combination of 0 and α (due to formula (3)):

$$\bar{P}^0 = (1 - \lambda)\alpha \quad (6)$$

So, we have

$$\bar{P}^1 - \bar{P}^0 = \lambda \quad (7)$$

Cramer proposed λ as a “crude measure of fit since it indicates the discrimination of $\hat{p}_i$ between the two observed outcomes.” Further analysis in Cramer’s paper demonstrates that λ “varies between 0 for the null model and 1 for perfect prediction, it reflects the differences between $\bar{P}^1$ and $\bar{P}^0$, it measures the proportion of the total variation of Y that is explained”. It should be noted that Cramer’s λ is positive only if α “represents the larger share [of ones]”. In practice, the λ is the absolute value of the difference between $\bar{P}^1$ and $\bar{P}^0$ (Allison, 2013).

Thus, since 1999 we have had a potentially useful measure of fit for logit models. Despite this, and although the original paper has received 129 citations in Scopus (as of March 17, 2025), Cramer’s *lambda* or *Cramer’s measure of discrimination* is rarely mentioned in the relevant literature,¹ and is not implemented in widely used software for estimating logit models, such as Stata.

Among classical econometrics textbooks, Cramer’s *lambda* is mentioned only in William H. Greene’s *Econometric Analysis* (7th ed. 2012, 8th ed. 2020).² In the 8th edition

¹ Two exemplary citations in the field of economics and econometrics are Clithero (2018) and Leppel (2007). Cramer’s *lambda* is also presented in Gruszczyński (2001, in Polish).

² Also checked: Wooldridge “Introductory Econometrics. A Modern Approach”, and “Econometric Analysis of Cross Section and Panel Data”, Cameron and Trivedi “Microeconometrics. Methods and Applications”, Winkelmann and Boes “Analysis of Microdata”, Stock and Watson “Introduction to Econometrics”, Maddala and Lahiri “Introduction

Greene categorizes goodness-of-fit measures for binomial (binary choice) models into two groups: (a) fit measures based on the fitting criterion, and (b) fit measures based on predicted values. Focusing on the second category, Greene identifies three such measures: R-squared measures of Efron (1978) and of Ben Akiva and Lehrman (1985), as well as Cramer's *lambda*.³ He notes that *lambda* "heavily penalizes the incorrect predictions, and because each proportion is taken within the subsample, it is not unduly influenced by large proportionate size of the group of more frequent outcomes."

Greene also notes that the same measure – originally proposed by Cramer – appears in the paper by Tjur (2009). In that work, Tjur "introduces" the same measure of fit as Cramer's and calls it the *coefficient of discrimination*. Unfortunately, Tjur does not acknowledge the precedence of Cramer's contribution; there is no mention of J.S. Cramer in the references.

Tjur's version of the coefficient of discrimination has gained much wider recognition than Cramer's original *lambda*. It is now widely known as Tjur's *D*. As of March 30, 2025, Tjur's paper has received 617 citations in Scopus, primarily in fields such as agricultural and biological sciences, environmental science, and medicine.⁴ Tjur's *D* has also been implemented in some logit procedures in Stata.

As noted by Allison (2013), Tjur's *D* "has a lot of intuitive appeal, its upper bound is 1.0, and it's closely related to R² definitions for linear models. It's also easy to calculate. The definition is very simple. For each of the two categories of the dependent variable, calculate the mean of the predicted probabilities of an event. Then, take the [absolute value of the] difference between those two means. That's it!"

Tjur (2009) also shows that:

$$\lambda = D = \frac{R_{ss}^2 + R_{mod}^2}{2} \quad (8)$$

where

$$R_{ss}^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{p}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (9)$$

to Econometrics", Stock and Watson "Introduction to Econometrics", Dougherty "Introduction to Econometrics", Angrist and Pischke "Mostly Harmless Econometrics", Hosmer, Lemeshow, Sturdivant "Applied Logistic Regression", Cunningham "Causal Inference: The Mixtape", Croissant "Microeconometrics with R" (references to be found e.g. on the webpage <https://www.econometricsbooks.com>).

³ There are also other fit measures based on predicted values, e.g. Morrison (1972) (and Goldberger, 1973) who proposed squared correlation between Y_i and $\hat{p}_i$ and Maddala (1992) – the squared correlation between Y_i and $\hat{Y}_i$ (i.e. value of Y_i predicted from the model).

⁴ Two exemplary citations in the field of economics and econometrics are Croissant (2025) and Szarucki et al. (2016). In the extensive survey paper on measures of fit for logistic regression Allison (2014) broadly presents the Tjur's *D*, without reference to Cramer's *lambda*.

and

$$R_{mod}^2 = \frac{\sum_{i=1}^n (\hat{p}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (10)$$

In his book on logit models Cramer (2003) also presents the proof that – approximately – the Cramer's *lambda* is equal to Efron's measure of fit (Efron, 1978), i.e.

$$R_E^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{p}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (11)$$

In the formula (8) Efron's R^2 is equal to R_{ss}^2 . More discussion on Tjur's D (Cramer's *lambda*) is presented in Williams (2022).

Cramer's *lambda* (Tjur's D) is generalized to nominal and original logit regression models by Smith et al. (2020).

2. Connection of Cramer's *lambda* (Tjur's D) to the confusion matrix

Measures based on predicted probabilities also include those that derive from a 2×2 contingency table (confusion matrix, diagnostic table) which displays two numbers of correct predictions and two numbers of incorrect predictions. Let us denote the elements of this table as follows:

	True $Y_i = 1$	True $Y_i = 0$	Total
$\hat{Y}_i = 1$	a	b	$a + b$
$\hat{Y}_i = 0$	c	d	$c + d$
Total	n_1	n_0	n

Symbols a , b , c , and d are the numbers of predicted outcomes: a and d for correct predictions, c , and b for incorrect predictions. The contingency table allows us to compute various fit measures, such as:

$$\text{sensitivity} = \frac{a}{n_1} \quad \text{specificity} = \frac{d}{n_0} \quad (12)$$

These are the proportions of correctly predicted ones (*sensitivity*) and zeroes (*specificity*). Another commonly used measure derived from the confusion matrix is the *count* R-squared: the ratio of $a + c$ and n .

A natural question arises: Can we calculate Cramer's *lambda* using only the values a , b , c , and d ? According to Hughes (2017), in a note published in the *Annals of Epidemiology*, the answer is yes. The author repeats this in Hughes et al. (2019) in another note, this time in *Phytopathology*. The finding is that Cramer's *lambda* (Tjur's D) is equal to the Youden's Index. The Youden's Index J originally proposed by Youden (1950), is defined as:

$$J = \text{sensitivity} + \text{specificity} - 1 \quad (13)$$

The J index is another measure of classification accuracy, commonly associated with the concept of ROC (Receiver Operating Characteristic) curve, which plots the *true positive rate* (*sensitivity*) against *false positive rate* ($1 - \text{specificity}$).

Hughes states that *Youden's J = Cramer's λ = Tjur's D* . This would suggest that Cramer's *lambda* can be easily calculated from a confusion matrix.

Unfortunately, this is not the case, as can be quickly demonstrated by numerical examples. Formal proof is provided by Hoessly (2025), however it also has some flaws. The question remains: Can *lambda* be expressed as a function of a , b , c , and d – the counts of correct and incorrect predictions?

In general, Youden's J is not equal to Cramer's *lambda*. This is because the probabilities $\hat{p}_i$ ($i = 1, \dots, n$) obtained from the estimated logit mode are fixed – determined solely by the model and estimation method – so Cramer's *lambda* is fixed. In contrast, sensitivity and specificity depend on how the predicted outcomes $\hat{Y}_i$ are derived from $\hat{p}_i$ values. The standard (default) procedure is:

- $\hat{Y}_i = 1$ if $\hat{p}_i \geq 0.5$,
- $\hat{Y}_i = 0$ if $\hat{p}_i < 0.5$.

Here the cut-off point (or cut-point) is 0.5. In the next section, we will discuss alternative choices for the cut-off.

Once predictions are made for all $i = 1, \dots, n$, the number of correct predictions a (cases where $\hat{Y}_i = 1$ and $Y_i = 1$) and d (cases where $\hat{Y}_i = 0$ and $Y_i = 0$) – can be determined, allowing for the computation of Youden's J . However, the value will only equal Cramer's *lambda* by coincidence.

Moreover, changing the cut-off point alters the values of a and d , and therefore changes Youden's J , while the Cramer's *lambda* remains unchanged.

3. Cramer's cut-off point

The choice of cut-off point for predicting Y_i is not obvious. Greene (2020) notes that “changing [cut-off point] will always reduce the probability of one type of error while increasing the probability of the other; there is no correct answer as to the best value to choose; it depends on the setting and on the criterion function upon which the prediction rule depends”. The commonly used cut-off point of 0.5 is clearly inappropriate for the samples with a larger share of one category (unbalanced).

In his 1999 paper in the section *Unfortunate effects for unbalanced samples* Cramer proposes that the “optimal” cut-off point for predicting Y_i is α – the relative share of ones in the sample. His reasoning is as follows.

Recall that in the logit model the average of estimated probabilities $\hat{p}_i$ equals the relative share α (see equation (2)). The estimated probability of observing the actual outcome for observation i is given by:

$$\Pr(i) = Y_i \hat{p}_i + (1 - Y_i)(1 - \hat{p}_i) \quad (14)$$

where $\hat{p}_i$ is defined by equation (1). In the *null model*, i.e. the logit model with no regressors, the predicted probabilities $\hat{p}_i$ are all equal α . Therefore, for the null model:

- when $Y_i = 1$, $\Pr(i) = \alpha$,
- when $Y_i = 0$, $\Pr(i) = 1 - \alpha$.

By introducing explanatory variables (regressors) into the model we aim to have higher values of $\Pr(i)$ as compared to the null model. The ratio of $\Pr(i)$ under the fitted model to that under the null model Cramer defines as *performance index* for i -th observation:

$$performanceindex(i) = Y_i \frac{\hat{p}_i}{\alpha} + (1 - Y_i) \frac{1 - \hat{p}_i}{1 - \alpha} \quad (15)$$

He argues that “it is natural to use a prediction that is optimal in the sense that, for given $\hat{p}_i$, it maximizes the *performance index* (i) of the equation (15), and hence the fit of $\hat{Y}_i$ to the given $\hat{p}_i$; this achieved by a cut-off point of α ”. According to Cramer, using the cut-off point α “more evenly spreads” the success rate over two outcome categories when compared to the conventional 0.5 cut-off.

The *optimal cut-off point* α , like the *measure of discrimination* λ , is not widely recognized in the literature as Mars Cramer's legacy.⁵ In his econometrics textbook Wooldridge (2013)

⁵ Cramer mentions his proposal of cut-off point α in the book on logit models (Cramer, 2003, p. 67). A few papers that make use of Cramer's *alpha* include e.g. Śmigielski et al. (2010) and Gruszczyński (2019).

clearly defines the cut-off point α , yet he does not attribute this idea to Cramer: “Some have criticized the prediction rule just described for using a threshold value of 0.5, especially when one of the outcomes is unlikely. (...) One alternative is to use the fraction of successes in the sample as the threshold. (...) Using this rule will certainly increase the number of predicted successes, but not without cost (...)”

There is substantial body of work devoted to identifying the “optimal” cut-off point in binomial logit modelling, particularly in biostatistics. For example, Hassanzad and Hajian-Tilaki (2024) present a survey of optimal cut-off point approaches within ROC analysis, especially in the context of biomarker evaluation in diagnostic medicine. The goal is to determine “the best” point on the ROC curve. Techniques include:

- Youden’s J : Maximizes *sensitivity* + *specificity* at various cut-off points, i.e. it maximizes the vertical distance from the ROC curve to the diagonal:

$$J = \text{sensitivity} + \text{specificity} - 1$$

- Euclidean distance: Identifies the point closest to the ideal corner (0,1) in ROC space:

$$\sqrt{(1 - \text{sensitivity})^2 + (1 + \text{specificity})^2} \quad (16)$$

In Stata, this is known as “nearest to (0, 1) method”.

- Index of Union: Minimizes the absolute value of differences between the diagnostic metrics (*sensitivity* and *specificity*) and the AUC (area under the ROC curve)

$$IU = |\text{sensitivity} - \text{AUC}| + |\text{specificity} - \text{AUC}| \quad (17)$$

- Cost based approaches: Incorporate the benefits of correct classifications and the costs of misclassifications. In economics and finance similar approaches were proposed by Palepu (1986) and Theodossiou et al. (1996).
- Maximum product of sensitivity and specificity. A method known as Liu (2012) method (available in Stata):

$$Liu = \text{sensitivity} \times \text{specificity} \quad (18)$$

Additional techniques considered by Hassanzad and Hajian-Tilaki (2024) include the positive likelihood ratio and negative likelihood ratio, number needed to misdiagnose, analytical method, diagnostic odds ratio and minimum p -value approach.

The methods that are available in Stata include: the Youden index, nearest to (0,1) criterion, and Liu’s method. Gerke and Zapf (2022) evaluate these techniques through a simulation exercise and on real-life data (Agatston scores for coronary calcification based on a CT scans).

However, these procedures could be usefully augmented by Cramer's *alpha*, a cut-off point that is immediately and easily available. The simplicity of *alpha* justifies its inclusion as the standard option in widely used estimation packages such as Stata. In our view, *alpha* (rather than the conventional value of 0.5) can serve as a natural starting point for a further exploration of cut-off points. Unlike other approaches that require optimizing an objective function over the ROC curve, Cramer's cut-off is derived analytically and offers a practical alternative – particularly in unbalanced classification settings.

4. Example 1

To illustrate Cramer's *measure of discrimination* λ and *optimal cut-off point* a we use the dataset `womenwk.dta` from the textbook of Baum (2006). The dataset contains information on 2,000 women, of whom 657 are not wage earners. Our Y variable **work** is coded as 0 for nonworking women and 1 for working women (i.e. those reporting positive wages). The explanatory variables X_j are **age** (in years), **education** (in years), **married** (zero–one), **children** (number). Number of working women is 1,343, i.e. the sample mean of the Y variable is 0.6715.

In addition to the full dataset of 2,000 observations, we use also four random subsamples, each consisting of 500 observations. The first subsample contains 500 observations randomly selected from the full dataset. The remaining three subsamples represent unbalanced datasets: the second subsample includes 500 observations randomly drawn from the full dataset with 10% of observations having **work** = 0; the third subsample includes 500 observations with 5% of **work** = 0; and the fourth subsample includes 500 observations with 2% of **work** = 0.

Table 1 presents the maximum likelihood (ML) estimates of the logit model with **work** as the dependent (explained) variable – first for the full dataset and then for each of the four 500-observation subsamples. These subsets were selected from the original `womenwk.dta` dataset. Note that the line “*alpha* = mean of Y ” shows the share of working women in a dataset. All of the results are obtained in Stata 18.6 (StataNow).

Table 2 presents measures of fit based on predicted probabilities for all five datasets. It shows the Cramer's *lambda* (Tjur's D), Efron R^2 as well as measures obtained by predicting Y with

- cut-off point 0.5,
- cut-off point equal to the share of ones in the dataset (Cramer's *alpha*),
- cut-off point that maximises the product of the sensitivity and specificity (Liu method),
- cut-off point that maximises the sum of the sensitivity and specificity (Youden method).

Table 1. Estimated parameters of the logit models for **work** (method of maximum likelihood)

Variable/Model	(1)	(2)	(3)	(4)	(5)
age	0.0579 (0.01)	0.0483 (0.02)	0.0711 (0.02)	0.0626 (0.03)	0.1041 (0.05)
education	0.0983 (0.02)	0.0301 (0.04)	0.1603 (0.07)	0.1641 (0.08)	0.1174 (0.12)
married	0.7418 (0.13)	0.7859 (0.24)	0.4945 (0.39)	0.2613 (0.50)	1.8751 (0.80)
children	0.7645 (0.05)	0.6960 (0.10)	0.8491 (0.16)	0.8036 (0.22)	1.3809 (0.44)
constant	-4.1592 (0.33)	-2.8715 (0.63)	-3.8357 (1.00)	-2.6700 (1.26)	-3.9154 (2.13)
<i>n</i>	2000	500	500	500	500
<i>alpha</i> = mean of <i>Y</i>	0.6715	0.6700	0.9000	0.9500	0.9800
McFadden's R ²	0.1882	0.1436	0.2069	0.1480	0.2617
AUC	0.7806	0.7480	0.8297	0.7824	0.8891

(1) full sample of 2,000 observations,

(2) 500 observations randomly selected from the full sample,

(3) 500 observations randomly selected from the full sample with 10% of **work** = 0,

(4) 500 observations randomly selected from the full sample with 5% of **work** = 0,

(5) 500 observations randomly selected from the full sample with 2% of **work** = 0.

Standard errors in brackets.

Source: own elaboration.

Cramer's *lambda* for the ML estimates of the logit model is obtained using Stata command suggested by Allison 2013 (see e.g. Williams, 2022). By default, Stata only reports Cramer's *lambda* (Tjur's *D*) for penalized ML estimates, i.e. for the Firth (1993) method. Later in this example we also apply the Firth approach.

As shown in Table 2, Cramer's *lambda* is quite close to Efron's R² across all samples – an observation that reduces the distinctiveness of *lambda*, given the broader familiarity and widespread use of Efron R². However, for highly unbalanced samples (models 3, 4, and 5) the *lambda* diverges more substantially from Efron's R².

Table 2. Measures of fit based on predicted probabilities (method of maximum likelihood)

Measure/Model	(1)	(2)	(3)	(4)	(5)
1	2	3	4	5	6
Efron's R²	0.216	0.166	0.147	0.099	0.118
Cramer's λ (Tjur's <i>D</i>)	0.218	0.167	0.154	0.088	0.112
Cut-off point 0.5	0.500	0.500	0.500	0.500	0.500
Sensitivity	0.876	0.875	0.987	1.000	1.000
Specificity	0.451	0.333	0.080	0.000	0.000
Youden index	0.327	0.208	0.067	0.000	0.000

1	2	3	4	5	6
Cut-off point <i>alpha</i>	0.6715	0.6700	0.9000	0.9500	0.9800
Sensitivity	0.6830	0.6660	0.7440	0.6910	0.7940
Specificity	0.7120	0.7090	0.8200	0.6800	0.8000
Youden index	0.3950	0.3750	0.5640	0.3710	0.5940
Cut-off point Liu	0.6942	0.6759	0.8598	0.9326	0.9830
Sensitivity	0.6400	0.6600	0.8200	0.8000	0.7600
Specificity	0.7700	0.7300	0.7600	0.6800	0.9000
Youden index	0.4100	0.3900	0.5800	0.4800	0.6600
Cut-off point Youden	0.7015	0.6759	0.8598	0.9326	0.9830
Sensitivity	0.6300	0.6600	0.8200	0.8000	0.7600
Specificity	0.7900	0.7300	0.7600	0.6800	0.9000
Youden index	0.4200	0.3900	0.5800	0.4800	0.6600

(1) full sample of 2,000 observations,

(2) 500 observations randomly selected from the full sample,

(3) 500 observations randomly selected from the full sample with 10% of **work** = 0,

(4) 500 observations randomly selected from the full sample with 5% of **work** = 0,

(5) 500 observations randomly selected from the full sample with 2% of **work** = 0.

Source: own elaboration.

More precise conclusions would require a rigorous analysis using alternative resampling strategies applied to a wider range of larger and smaller datasets (beyond the illustrative example considered here). Our objective in this section is to present a numerical example that illustrates the central message of the paper: Cramer's cut-off point *alpha* provides a valid and convenient choice for unbalanced samples – without a deeper search along the ROC curve. The results in Table 2 show that *alpha* performs comparably to more complex techniques, unless additional objectives or constraints are imposed.

5. Rare events and binomial logit

For the case of rare events, i.e. highly imbalanced data it is advised to apply the penalized ML of Firth. Heinz Leitgöb, at the European Survey Research Association conference in 2013 presented the widely cited paper “The Problem of Modeling Rare Events in ML-Based Logistic Regression” (Leitgöb 2013; see also notes by Richard Williams, 2022). Our paper compares three methods:

- exact logistic regression,
- prior correction method of King and Zeng (2001),⁶
- penalized maximum likelihood estimation (PMLE) of Firth (1993).

⁶ For details see e.g. Gruszczyński (2019), section 5.

Users of Stata may apply these methods with commands, respectively, *exlogistic*, *relogit* and *firthlogit*. Exact logistic regression “was first introduced by Cox and Snell (1989) as an alternative to using maximum likelihood estimation, which can perform poorly for small sample sizes.” (quote from Stata manual). It is only applicable for small samples (less than 200 observations) and small numbers of covariates. For the King/Zeng correction and Firth’s PMLE Leitgöb (2013) shows with the Monte Carlo simulation that:

- the King and Zeng prior correction method “is somewhat overcorrecting bias in MLEs as n is getting small (<200)”,
- PMLEs “seem unbiased”, even when n is small and with very few number of events; also, PLME is always converging.

Thus, when estimating the logit model for rare events it is recommended to apply PLME.

6. Example 2

Applying the penalized ML of Firth gives estimates presented in Table 3. For the Firth method (Stata command: *firthlogit*) the list of goodness of fit measures includes Tjur’s D (Stata command: *firthfit*). The Firth estimates differ from the ones from Table 1, especially for more unbalanced cases.

Table 3. Estimated parameters for the logit models for **work**
(Firth method of penalized maximum likelihood)

Variable/Model	(1)	(2)	(3)	(4)	(5)
age	0.0577 (0.01)	0.0475 (0.01)	0.0692 (0.02)	0.0598 (0.03)	0.0963 (0.05)
education	0.0979 (0.02)	0.0296 (0.04)	0.1546 (0.07)	0.1557 (0.08)	0.0969 (0.11)
married	0.7394 (0.13)	0.7764 (0.24)	0.4872 (0.38)	0.2751 (0.49)	1.7470 (0.76)
children	0.7614 (0.05)	0.6852 (0.10)	0.8227 (0.16)	0.7698 (0.22)	1.2317 (0.41)
constant	−4.1427 (0.33)	−2.8253 (0.62)	−3.7093 (0.99)	−2.5353 (1.23)	−3.4711 (2.01)
n	2000	500	500	500	500
α = mean of Y	0.6715	0.6700	0.9000	0.9500	0.9800
McFadden’s R^2	0.1900	0.1480	0.2180	0.1600	0.2990
AUC	0.7806	0.7480	0.8297	0.7821	0.8887

(1) full sample of 2,000 observations,

(2) 500 observations randomly selected from the full sample,

(3) 500 observations randomly selected from the full sample with 10% of **work** = 0,

(4) 500 observations randomly selected from the full sample with 5% of **work** = 0,

(5) 500 observations randomly selected from the full sample with 2% of **work** = 0.

Standard errors in brackets.

Source: own elaboration.

Table 4 presents the same measures of fit as Table 2 – for the Firth method of penalized maximum likelihood. For the Firth method the values of sensitivity and specificity are obtained using the advice of Joseph Coveney on Stata Forum (on 4th September 2019).

Table 4. Measures of fit based on predicted probabilities
(Firth method of penalized maximum likelihood)

Measure/Model	(1)	(2)	(3)	(4)	(5)
Efron's R2	0.216	0.166	0.148	0.097	0.115
Cramer's λ (Tjur's D)	0.218	0.166	0.154	0.089	0.118
Cut-off point 0.5	0.5	0.5	0.5	0.5	0.5
Sensitivity	0.876	0.875	0.989	1.000	1.000
Specificity	0.451	0.333	0.080	0.000	0.000
Youden index	0.327	0.208	0.067	0.000	0.000
Cut-off point α	0.6715	0.6700	0.9000	0.9500	0.9800
Sensitivity	0.681	0.666	0.718	0.655	0.731
Specificity	0.714	0.721	0.840	0.720	0.900
Youden index	0.395	0.387	0.558	0.375	0.631
Cut-off point Liu	0.6942	0.6759	0.8598	0.9326	0.9830
Sensitivity	0.642	0.657	0.809	0.766	0.712
Specificity	0.776	0.733	0.760	0.680	0.900
Youden index	0.419	0.390	0.569	0.446	0.612
Cut-off point Youden	0.7015	0.6759	0.8598	0.9326	0.9830
Sensitivity	0.631	0.657	0.809	0.766	0.712
Specificity	0.790	0.733	0.760	0.680	0.900
Youden index	0.421	0.390	0.569	0.446	0.612

(1) full sample of 2,000 observations,

(2) 500 observations randomly selected from the full sample,

(3) 500 observations randomly selected from the full sample with 10% of **work** = 0,

(4) 500 observations randomly selected from the full sample with 5% of **work** = 0,

(5) 500 observations randomly selected from the full sample with 2% of **work** = 0.

Source: own elaboration.

Similarly to the results obtained using the ML method, Cramer's cut-off point – along with the Liu and Youden cut-off points, substantially improves the sensitivity-specificity pair compared to the default ("standard") cut-off point of 0.5. This provides a strong rationale for considering Cramer's *alpha* as the default choice in practical applications.

Cramer's *lambda*, or Tjur's *D* (obtained directly by the Stata command *firthfit*) does not differ significantly from the values reported in Table 2, except for highly imbalanced sample (models 3, 4 and 5).

Conclusions

This paper revisits Mars Cramer's contributions to logit modelling, focusing on his often-overlooked proposals for of a new measure of fit and a cut-off point in predicting values of the explained variable. Both were introduced in his 1999 paper (Cramer, 1999). We have highlighted that Cramer's *lambda* – a goodness of fit measure – has gained recognition in the literature under a different name, Tjur's *D*, which was introduced a decade later (2009). We also point out that Cramer's cut-off point, *alpha* has not been widely considered among standard choices used by researchers.

We hope this paper helps to re-establish Mars Cramer legacy in econometrics. Both Cramer's *lambda* and Cramer's *alpha* deserve recognition and adoption as legitimate standards for binomial logit modelling.

References

- Allison, P. (2014). *Measures of Fit for Logistic Regression*. Proceedings of the SAS Global Forum 2014 Conference, Paper 1485–2014.
- Allison, P. (2013). *What's the Best R-Squared for Logistic Regression?* Retrieved from <http://www.statisticalhorizons.com/r2logistic>.
- Baum, C.F. (2006). *An Introduction to Modern Econometrics using Stata*. Stata Press.
- Ben-Akiva, M., Lerman S. (1985). *Discrete Choice Analysis*. London: MIT Press.
- Clithero, J.A. (2018). Improving out-of-sample predictions using response times and a model of the decision process. *Journal of Economic Behavior & Organization*, 148, 344–375. DOI: 10.1016/j.jebo.2018.02.007.
- Cox, D.R., Snell, E.J. (1989). *Analysis of Binary Data*. 2nd ed. London: Chapman & Hall.
- Cramer, J.S. (1999). Predictive Performance of the Binary Logit Model in Unbalanced Samples. *Journal of the Royal Statistical Society, Series D (The Statistician)*, 48, 85–94. DOI: 10.1111/1467-9884.00173.
- Cramer, J.S. (2003). *Logit Models from Economics and Other Fields*. Cambridge University Press. DOI: 10.1017/CBO9780511615412.

- Croissant, Y. (2025). *Microeconometrics with R*. Chapman & Hall. DOI: 10.1201/9781003100263.
- Efron, B. (1978). Regression and ANOVA with Zero-One Data: Measures of Residual Variation. *Journal of the American Statistical Association*, 73(361), 113–212. DOI: 10.2307/2286531.
- Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1), 27–38. DOI: 10.2307/2336755.
- Gerke, O., Zapf, A. (2022). Convergence Behavior of Optimal Cut-Off Points Derived from Receiver Operating Characteristics Curve Analysis: A Simulation Study. *Mathematics*, 10(22), 4206. DOI: 10.3390/math10224206.
- Goldberger, A.S. (1973). Correlations between Binary Outcomes and Probabilistic Predictions. *Journal of the American Statistical Association*, 68(341), 84. DOI: 10.1080/01621459.1973.10481338.
- Greene, W.H. (2012). *Econometric Analysis* (7th edition), Pearson Education.
- Greene, W.H. (2020). *Econometric Analysis* (8th edition), Pearson Education.
- Gruszczyński, M. (2001). *Modele i prognozy zmiennych jakościowych w finansach i bankowości*. Warszawa: Oficyna Wydawnicza SGH.
- Gruszczyński, M. (2019). On Unbalanced Sampling in Bankruptcy Prediction. *International Journal of Financial Studies*, 7(2), 28. DOI: 10.3390/ijfs7020028.
- Hassanzad, M., Hajian-Tilaki, K. (2024). Methods of determining optimal cut-point of diagnostic biomarkers with application of clinical data in ROC analysis: an update review. *BMC Medical Research Methodology*, 24, 84. DOI: 10.1186/s12874-024-02198-2.
- Hoessly, L. (2025). Youden index and Tjur's R² in 2 × 2 tables. *Annals of Epidemiology*, 103, 28–29. DOI: 10.1016/j.annepidem.2025.01.012.
- Hughes, G. (2017). Tjur's R² for logistic regression models is the same as Youden's index for a 2 × 2 diagnostic table. *Annals of Epidemiology*, 27(12), 801–802. DOI: 10.1016/j.annepidem.2017.10.009.
- Hughes, G., Choudhury, R.A., McRoberts, N. (2019). Summary Measures of Predictive Power Associated with Logistic Regression Models of Disease Risk. *Phytopathology*, 109(5), 712–715. DOI: 10.1094/PHYTO-09-18-0356-LE.
- King, G., Zeng, L. (2001). Logistic Regression in Rare Events Data. *Political Analysis*, 9(2), 137–163. DOI: 10.1093/oxfordjournals.pan.a004868.
- Leitgöb, H. (2013). The Problem of Rare Events in Maximum Likelihood Logistic Regression – Assessing Potential Remedies, European Survey Research Association. Ljubljana. Retrieved from http://www.europeansurveyresearch.org/conf/uploads/494/678/167/PresentationLeitg_b.pdf.
- Leppel, K. (2007). Home-ownership among opposite- and same-sex couples in the US. *Feminist Economics*, 13(1), 1–30. DOI: 10.1080/13545700601075070.

- Liu, X. (2012). Classification accuracy and cut point selection. *Statistics in Medicine*, 31(23), 2676–2686. DOI: 10.1002/sim.4509.
- Maddala, G. (1992). *Introduction to Econometrics*. 2nd ed. New York: Macmillan.
- Morrison, D.G. (1972). Upper Bounds for Correlations between Binary Outcomes and Probabilistic Predictions. *Journal of the American Statistical Association*, 67(337), 68–70. DOI: 10.1080/01621459.1972.10481207.
- Palepu, K.G. (1986). Predicting takeover targets: a methodological and empirical analysis. *Journal of Accounting and Economics*, 8, 3–35. DOI: 10.1016/0165-4101(86)90008-X.
- Smith, T.J., Walker, D.A., McKenna, C.M. (2020). A coefficient of discrimination for use with nominal and ordinal regression models. *Journal of Applied Statistics*, 48(16), 3208–3219. DOI: 10.1080/02664763.2020.1796940.
- Szarucki, M., Brzozowski, J., Stankevičienė, J. (2016). Determinants of self-employment among Polish and Romanian immigrants in Germany. *Journal of Business Economics and Management*, 17(4), 598–612. DOI: 10.3846/16111699.2016.1202313.
- Śmigielski, J., Majdzińska, A., Śmigielski (2010). Using ROC curves to find the cut-off point in logistic regression with unbalanced data samples. *Statistics in Transition*, 11, 381–402. DOI: 10.59170/stattrans-2010-023.
- Theodossiou, P., Kahya, E., Saidi, R., Philippatos, G. (1996). Financial distress and corporate acquisitions: further empirical evidence. *Journal of Business Finance and Accounting*, 23, 699–719. DOI: 10.1111/j.1468-5957.1996.tb01149.x.
- Tjur, T. (2009). Coefficients of Determination in Logistic Regression Models – A New Proposal: The Coefficient of Discrimination. *The American Statistician*, 63(4), 366–372. DOI: 10.1198/tast.2009.08210.
- Williams, R. (2022). Scalar Measures of Fit: Pseudo R² and Information Measures (AIC & BIC), University of Notre Dame. Retrieved from <https://www3.nd.edu/~rwilliam> (3.01.2022).
- Wooldridge, J.M. (2013). *Introductory Econometrics. A Modern Approach* (5th edition), South-Western, Cengage Learning.
- Youden, W.J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32–5. DOI: 10.1002/1097-0142(1950)3:1<32::aid-cnrcr2820030106>3.0.co;2-3. PMID: 15405679.

Citation

- Gruszczyński, M. (2026). The Measure of Fit and Cut-off Point for a Binomial Logit Model. An Overlooked Contribution of Jan Salomon Cramer. *Folia Oeconomica Stetinensia*, 26(1), 70–86. DOI: 10.2478/fole-2026-0004.