FACULTY OF LAW, ECONOMICS AND FINANCE

CREA Discussion Paper 2017-08

Economics

Center for Research in Economics and Management
University of Luxembourg

Monotonicity of Probit Weights

available online : http://wwwfr.uni.lu/recherche/fdef/crea/publications2/discussion_papers

Martin Schumann, CREA, Université du Luxembourg
Gautam Tripathi, CREA, Université du Luxembourg

May, 2017

For editorial correspondence, please contact: crea@uni.lu
University of Luxembourg
Faculty of Law, Economics and Finance
162A, avenue de la Faïencerie
L-1511 Luxembourg

UNIVERSITÉ DU LUXEMBOURG

MONOTONICITY OF PROBIT WEIGHTS

MARTIN SCHUMANN AND GAUTAM TRIPATHI

Abstract. We demonstrate that the probit weight function is U-shaped on $\mathbb{R}$, i.e., it is strictly decreasing on $(-\infty, 0)$, strictly increasing on $[0, \infty)$, and strictly convex on $\mathbb{R}$. Knowledge of the shape of the probit weight function can resolve any confusion that may arise from a result in the classic paper of Sampford (1953).

The probit model for a binary outcome $Y \in \{0, 1\}$, conditional on a (column) vector of non-degenerate regressors X, is given by

$$\Pr(Y = 1|X) = \Phi(\alpha_0 + X^\top \beta_0) = \Phi(\tilde{X}^\top \theta_0) \qquad \text{w.p.1}, \tag{1}$$

where Φ is the standard normal cumulative distribution function, $\tilde{X} := [\begin{smallmatrix} 1 \\ X \end{smallmatrix}]$, and $\theta_0 := [\begin{smallmatrix} \alpha_0 \\ \beta_0 \end{smallmatrix}]$. Let $(Y_1, X_1), \dots, (Y_n, X_n)$ be i.i.d. copies of (Y, X). The maximum likelihood estimator (MLE) of θ_0 is the maximizer (over the parameter space containing θ_0) of the loglikelihood

$$\ell_n(\theta) := n^{-1} \sum_{j=1}^{n} Y_j \log \Phi(\tilde{X}_j^\top \theta) + (1 - Y_j) \log \Phi(-\tilde{X}_j^\top \theta).$$

Hence, the MLE can be obtained as a zero of the score function $\theta \mapsto \nabla \ell_n(\theta)$, i.e., by finding θ in the parameter space that solves the first order condition

$$n^{-1} \sum_{j=1}^{n} [Y_j - \Phi(\tilde{X}_j^\top \theta)] \tilde{X}_j G(\tilde{X}_j^\top \theta) = 0,$$

where

$$G(v) := \frac{\phi(v)}{\Phi(v)\Phi(-v)}, \qquad v \in \mathbb{R},$$

and $\phi := \Phi'$ denotes the standard normal density.

We call G the probit weight function because it is present multiplicatively in the expression for the probit first order conditions. Additional justification behind this terminology can be obtained from the theory of optimal instruments (Newey, 1993, Section 2). Begin by observing that (1) is equivalent to the conditional moment restriction model

$$\mathbb{E}[Y - \Phi(\tilde{X}^\top \theta_0)|X] = 0 \qquad \text{w.p.1}. \tag{2}$$

`Source: pw-1n.tex. Compiled: 2nd May, 2017 at 12:41 Noon.`

Key words and phrases. Probit weight function, Shape.

JEL classification: C18, C25.

If $v(X)$ denotes a vector-valued function of X, say, e.g., $v(X) = \tilde{X}$, then (1) implies that $\mathbb{E}[Y - \Phi(\tilde{X}^\top\theta_0)]v(X) = 0$. Hence, provided this moment condition identifies θ_0, the generalized method of moments (GMM) estimator can be used to estimate θ_0. An interesting problem is to determine the "optimal instrument" $v^*(X)$ such that the GMM estimator of θ_0 based on the moment condition $\mathbb{E}[Y - \Phi(\tilde{X}^\top\theta_0)]v^*(X) = 0$ is semiparametrically efficient. Following equation 2.6 in Newey's paper, it is easy to show that the optimal instrument in the probit model is $v^*(X) := -\tilde{X}G(\tilde{X}^\top\theta_0)$. Therefore, $G(\tilde{X}^\top\theta_0)$ can be interpreted as the weight that transforms the "natural instrument" $\tilde{X}$, suggested by (2), into the optimal instrument.

Although we refer to G as the probit weight function (and justify why it should be so called), this usage is not uniform in the literature. For instance, the classic paper of Sampford (1953), which showed that the Gaussian hazard function is strictly increasing and strictly convex on $\mathbb{R}$, defines the function

$$\psi(v) := \frac{e^{-v^2}}{\int_{-\infty}^{v} e^{-t^2/2}\,dt \int_{v}^{\infty} e^{-t^2/2}\,dt}, \qquad v \in \mathbb{R},$$

and states that "ψ is well known as the weight function in probit analysis" (Sampford, Section 4, p. 132).[1] However, it is easy to see that $\psi(v) = \phi(v)G(v) \neq G(v)$, $v \in \mathbb{R}$. Similarly, Hammersley (1950, p. 206), who is cited by Sampford, refers to (in his notation) $Y(v) := 4e^{v^2}\Phi(v)(1 - \Phi(v)) = 2/[\pi\psi(v)]$ as "a function which occurs in probit analysis." But, clearly, Hammersley's function $v \mapsto Y(v)$ is different from both ψ and G! This lack of consensus on terminology may cause confusion because these different weight functions have different properties. For instance, Sampford shows that ψ is strictly decreasing on $[0, \infty)$, which is exactly opposite of how G behaves on this interval (Figure 1). Moreover, the monotonicity and shape properties of G cannot be inferred from those of ψ, because $G = \psi/\phi$ and both ψ and ϕ are strictly decreasing on $[0, \infty)$; instead, a separate proof is required.

The following result can help eliminate any potential confusion by determining the shape of the probit weight function G.

Lemma 1 (Shape of G)**.** (i) *G is strictly decreasing on $(-\infty, 0)$, and strictly increasing on $[0, \infty)$.* (ii) *G is strictly convex on $\mathbb{R}$.*

Strict convexity of G can be used to show that $|G'| < 1$ on $\mathbb{R}$ (Lemma 3), which makes precise the boundedness of G' alluded to in Park and Phillips (2000, p. 1254).

For the remainder of the paper, let $h(v) := \phi(v)/\Phi(-v)$, $v \in \mathbb{R}$, denote the Gaussian hazard function. The reciprocal of the Gaussian hazard function, i.e., $1/h$, is known in the literature as Mills' ratio. The reflection of h about the origin, namely, $h(-v) = \phi(v)/\Phi(v)$, is called the "reverse" hazard function (Marshall and Olkin, 2007, Equation 5, p. 13), although

[1] In describing the history of probit regression, McCulloch (1999, Section 1) notes that the weight function ψ was used by Bliss (1935) to estimate the probit model as an iteratively reweighted linear regression.

in some older works, e.g., Barlow, Marshall, and Proschan (1963, p. 379), it is also referred to as the "dual" hazard function.

Proof of Lemma 1. (i) Since G is symmetric about the origin, i.e., $G(-v) = G(v)$, $v \in \mathbb{R}$, it suffices to show that G is strictly increasing on $[0, \infty)$. We give two proofs of this result.

The first proof begins by observing that

$$G(v) = \phi(v)[\frac{1}{\Phi(v)} + \frac{1}{\Phi(-v)}] = h(-v) + h(v), \qquad v \in \mathbb{R}.$$

Hence,

$$G'(v) = h'(v) - h'(-v), \qquad v \in \mathbb{R}. \tag{3}$$

But, $u > 0 \implies -u < u \implies h'(-u) < h'(u)$ because h' is strictly increasing on $\mathbb{R}$.[2] Hence, $G' > 0$ on $(0, \infty)$, which implies that G is strictly increasing on $(0, \infty)$. Differentiability of G on $[0, \infty)$ then reveals that it is also strictly increasing on $[0, \infty)$.[3]

The second proof uses straightforward differentiation of $v \mapsto \phi(v)/[\Phi(v)\Phi(-v)]$ and the facts that $\phi(-v) = \phi(v)$ and $\phi'(v) = -v\phi(v)$, $v \in \mathbb{R}$, to show that

$$G'(v) = G(v)[h(v) - h(-v) - v], \qquad v \in \mathbb{R}.$$

Then, $G' > 0$ on $(0, \infty)$ follows immediately from Lemma 2.

(ii) By (3), $G''(v) = h''(v) + h''(-v)$, $v \in \mathbb{R}$. Consequently, $G'' > 0$ follows from the strict convexity of the Gaussian hazard function. □

Lemma 2 (Tate)**.** *If* $u > 0$, *then* $h(u) - h(-u) > u$.

Proof of Lemma 2. Tate (1953, Lemma 1) has shown that

$$\Phi(u)\Phi(-u)u < [2\Phi(u) - 1]\phi(u), \qquad u > 0. \tag{4}$$

But, since $\Phi(u) + \Phi(-u) = 1$,

$$\begin{aligned} (4) &\iff \Phi(u)\Phi(-u)u < [\Phi(u) - \Phi(-u)]\phi(u) \\ &\iff u < [\frac{1}{\Phi(-u)} - \frac{1}{\Phi(u)}]\phi(u) \\ &\iff u < h(u) - h(-u). \end{aligned}$$

□

Remark 1. Tate's inequality, as stated in Lemma 2, is equivalent to $h(u) > u + h(-u)$, $u > 0$, which tightens the well known lower bound for the Gaussian hazard function (due to Gordon, 1941), namely, that $h(u) > u$, $u > 0$.

---

[2]This follows from the fact that $h'' > 0$ on $\mathbb{R}$ (cf. equation (4) in Sampford's paper), which implies that the Gaussian hazard function is strictly convex on $\mathbb{R}$.

[3]Since we know that $G' > 0$ on $(0, \infty)$, to demonstrate that G is strictly increasing on $[0, \infty)$ it suffices to show that $G(u) > G(0)$ for $u \in (0, \infty)$. But this is immediate because $G(u) - G(0) = G'(\lambda u)u$ for some $\lambda \in (0, 1)$ by the mean-value theorem, and we have already shown that $G' > 0$ on $(0, \infty)$.

Lemma 3. $|G'(v)| < 1$ *for* $v \in \mathbb{R}$.

Proof of Lemma 3. Strict convexity of G on $\mathbb{R}$ (Lemma 1(ii)) implies that G' is strictly increasing on $\mathbb{R}$. Hence, to prove that $-1 < G' < 1$, it suffices to show that $\lim_{v\to\infty} G'(v) = 1$ and $\lim_{v\to-\infty} G'(v) = -1$. However, since $G'(-v) = -G'(v)$ by (3), it is enough to show that $\lim_{v\to\infty} G'(v) = 1$. Begin by observing that

$$h'(v) = \frac{-v\phi(v)\Phi(-v) + \phi^2(v)}{\Phi^2(-v)} \implies h'(-v) = \frac{v\phi(v)\Phi(v) + \phi^2(v)}{\Phi^2(v)}. \tag{5}$$

Hence, $\lim_{v\to\infty} h'(-v) = 0$. Therefore, by (3), $\lim_{v\to\infty} G'(v) = \lim_{v\to\infty} h'(v)$. But,[4]

$$\begin{aligned}
\lim_{v\to\infty} h'(v) &= \lim_{v\to\infty} \frac{-v\phi(v)\Phi(-v) + \phi^2(v)}{\Phi^2(-v)} && \text{((5))} \\
&= \lim_{v\to\infty} \frac{\Phi(-v)(1-v^2) + v\phi(v)}{2\Phi(-v)} && \text{(L'Hospital)} \\
&= \lim_{v\to\infty} \frac{v\Phi(-v)}{\phi(v)} && \text{(L'Hospital)} \\
&= \lim_{v\to\infty} 1 - \frac{\Phi(-v)}{v\phi(v)} && \text{(L'Hospital)} \\
&= \lim_{v\to\infty} 1 - \frac{1}{v^2-1} && \text{(L'Hospital)} \\
&= 1. && \square
\end{aligned}$$

ACKNOWLEDGEMENTS

We thank Andreas Irmen for helpful conversations.

References

Barlow, R. E., A. W. Marshall, and F. Proschan (1963): "Properties of probability distributions with monotone hazard rate," *Annals of Mathematical Statistics*, 34, 375–389.

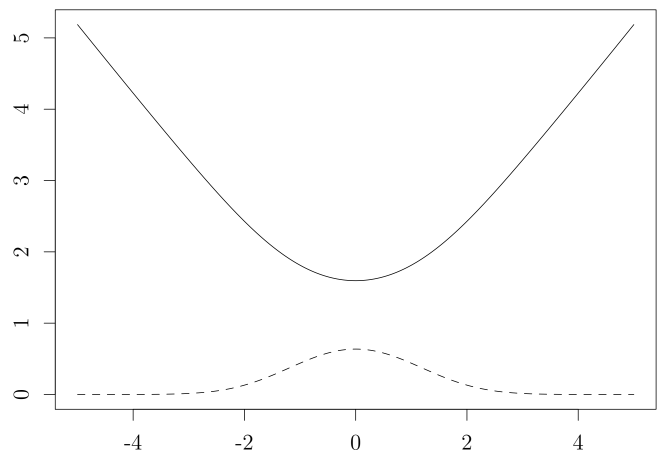
Bliss, C. I. (1935): "The calculation of the dosage-mortality curve," *Annals of Applied Biology*, 22, 134–167.

Gordon, R. D. (1941): "Values of Mills' ratio of area to bounding ordinate and of the normal probability integral for large values of the argument," *Annals of Mathematical Statistics*, 12, 364–366.

Hammersley, J. M. (1950): "On estimating restricted parameters," *Journal of the Royal Statistical Society, Series B*, 12, 192–240.

[4]In applying L'Hospital's rule to obtain the third and the fourth equalities, we used the fact that $\lim_{v\to\infty} v^2\Phi(-v) = 0$ and $\lim_{v\to\infty} v\Phi(-v) = 0$. These results follow from the inequality $\Phi(-u) < \phi(u)/u$, $u > 0$, which is a restatement of Gordon's lower bound for the Gaussian hazard function (Remark 1).

Figure 1. The probit weight function G (solid) and Sampford's function ψ (dashed).



Marshall, A. W., and I. Olkin (2007): *Life distributions: Structure of nonparametric, semiparametric, and parametric families*. Springer, New York.

McCulloch, C. E. (1999): "Generalized linear models," Biometrics Unit Technical Report BU-1449-M, https://ecommons.cornell.edu/bitstream/handle/1813/32046/BU-1449-M.pdf.

Newey, W. K. (1993): "Efficient estimation of models with conditional moment restrictions," in *Handbook of Statistics, vol. 11*, ed. by G. S. Maddala, C. R. Rao, and H. D. Vinod, pp. 2111–2245. Elsevier Science B.V., The Netherlands.

Park, J. Y., and P. C. B. Phillips (2000): "Nonstationary binary choice," *Econometrica*, 68, 1249–1280.

Sampford, M. R. (1953): "Some inequalities on Mill's ratio and related functions," *Annals of Mathematical Statistics*, 24, 130–132.

Tate, R. F. (1953): "On a double inequality of the normal distribution," *Annals of Mathematical Statistics*, 24, 132–134.

Center for Research in Economics and Management (CREA), Faculty of Law, Economics and Finance, University of Luxembourg, L-1511, Luxembourg.
E-mail address: martin.schumann@uni.lu

Center for Research in Economics and Management (CREA), Faculty of Law, Economics and Finance, University of Luxembourg, L-1511, Luxembourg.
E-mail address: gautam.tripathi@uni.lu