The distribution of the uncitedness factor and its functional relation with the impact factor

L .Egghe

Universiteit Hasselt (UHasselt), Campus Diepenbeek, Agoralaan, B-3590 Diepenbeek, Belgium

The uncitedness factor (U) of a journal is its fraction of uncited articles. Given a set of journals (e.g. in a field) we can determine the rank-order distribution of these uncitedness factors. Hereby we use the Central Limit Theorem which is valid for uncitedness factors since it are fractions, hence averages.

A similar result was proved earlier for the impact factors (IF) of a set of journals. Here we combine the two rank-order distributions, hereby eliminating the rank, yielding the functional relation between the impact factor (IF) and the uncitedness factor (U).

The rank-order distribution of the impact factor IF

In an earlier paper we studied the rank-order distribution of impact factors (IF) say of a set of journals (e.g. in a field). Remarking that IFs are averages (average number of citations per article in a journal) we can use the Central Limit Theorem (CLT) for the distrbution of IFs over these journals:

The normal (or Gaussian) distribution

$$\varphi(x) = Ae^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where $x = IF \geq 0$ and where the constant A is such that

$$\int_0^\infty \varphi(x)dx = T$$

the total number of journals.

Ranking the journals in decreasing order of their IFs we have that

$$r = \int_x^{\infty} \varphi(y)dy$$

with x=IF(r), the IF of the journal at rank r (continuous argument). Denoting

$$F(x) = \int_0^x \varphi(y)dy$$

, the cumulative normal distribution, we have

$$r = T - \int_0^x \varphi(y)dy$$

$$r = T - F(x)$$

Hence, denoting by F^{-1} the inverse of the injective function F,

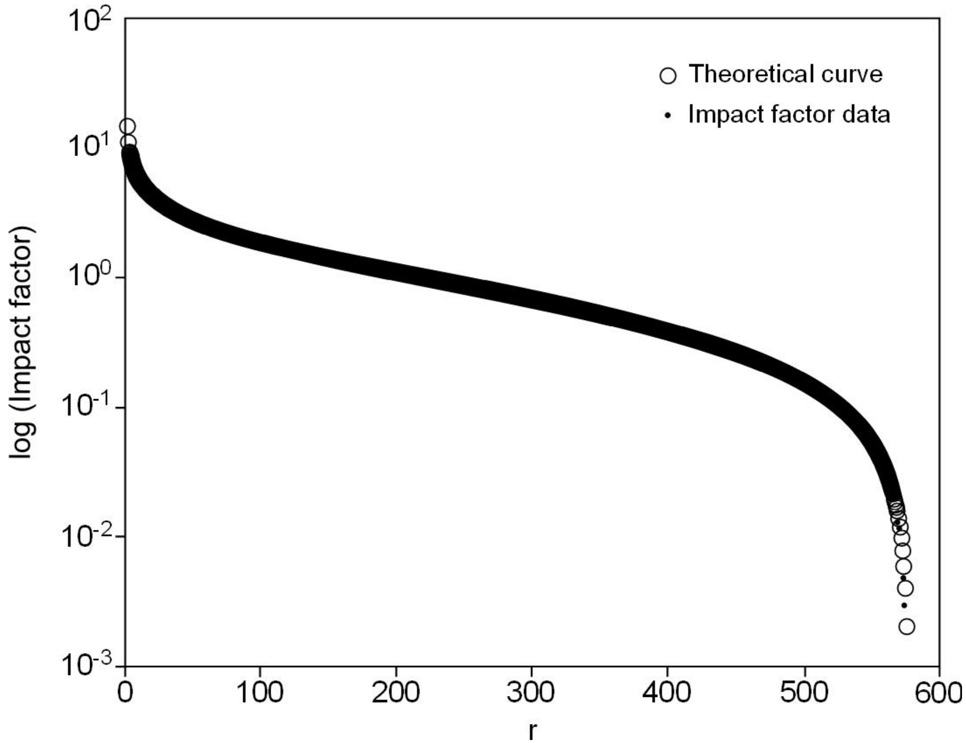
$$IF(r) = x = F^{-1}(T - r)$$

yielding the desired distribution.

Theorem: The function

$$f(r) = \ln(IF(r)) = \ln(F^{-1}(T - r))$$

is strictly decreasing, first, on an interval $[0, r_0]$ convexly and on the interval $[r_0, T]$ concavely. Furthermore we have that $IF(r_0) > \mu$ in the inflection point $(r_0, \ln(IF(r_0)))$. This is proved based on properties of the Gaussian curve.



Rank-order distribution of ln(IF). Reprinted from Mansilla, Köppen, Cocho and Miramontes (2007) with kind permission from Elsevier.

Note that the inflection point is indeed in a rank such that $\mathrm{IF}(\mathrm{r}_0) > \mu$ (= average of the IFs, i.e. of the Gaussian distribution of IF), i.e. the convex part is "smaller" than the concave part.

We can also show:

<u>Theorem</u>: The function

$$g(r) = IF(r) = F^{-1}(T - r)$$

is strictly decreasing, first on an interval $[0, r_1]$ convexly and on the interval $[r_1, T]$ concavely. Furthermore we have that $IF(r_1) = \mu$ in the inflection point $(r_1, IF(r_1))$.

The rank-order distribution of the uncitedness factor U

Similar as what we did for the IF, we suppose that U is distributed according to a normal (Gaussian) distribution:

$$\psi(y) = Be^{-\frac{(y-\mu')^2}{2\sigma'^2}}$$

where B is such that

$$\int_0^1 \psi(y)dy = T$$

(note that $y \in [0,1]$ being an uncitedness factor). Indeed, also for U we can apply the CLT since U is a fraction: the fraction of the papers in the journal which are uncited.

Now we rank the journals in increasing order r of their uncitedness factors U. The defining relation for U(r) is now, evidently,

$$r = \int_0^y \psi(y')dy'$$

where $y = U(r)$. Denote $G(y)$ this integral (which is the cumulative normal distribution) we simply obtain

$$r = G(U)$$

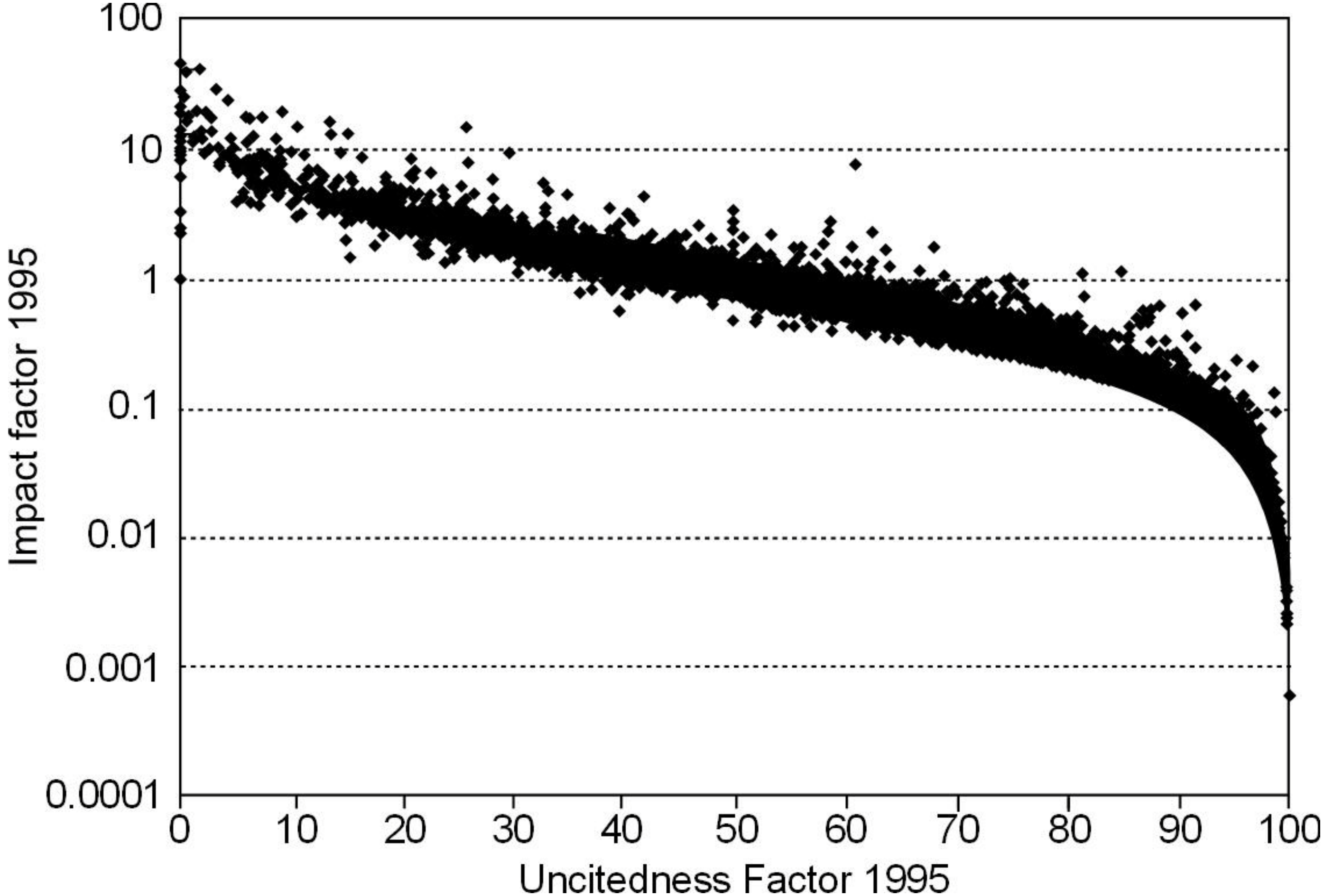
hence

$$U(r) = G^{-1}(r)$$

where G^{-1} is the inverse of the injective function G.

Using properties of the Gaussian distribution we can prove that the function $\mathrm{U(r)}$ is first concavely increasing followed by a convex increase. The inflection point is in $(\mathrm{r}, \mathrm{U(r)})$ where $\mathrm{U}\left(\mathrm{r}\right) = \mu'$ the average of the Us, i.e. of the Gaussian distribution of U.

The functional relation between the impact factor IF and the uncitedness factor U.



ln(IF(U)), experimentally: SCI (all fields), 1995, van Leeuwen and Moed (2005)

In this section we make one assumption: the rank r occuring in the IF-distribution is the same as the rank r in the U-distribution. This is model-theoretically acceptable since this assumption is equivalent by supposing that IF decreases with U (note that IF(r) decreases in r and U(r) increases in r).

We proved already

$$r = T - F(x)$$

with x=IF(r) and

$$r = G(y)$$

with $y = U(r)$.

$$T - F(x) = G(y)$$

$$F(x) = T - G(y)$$

$$x = F^{-1}(T - G(y))$$

$$\mathbb{IF}(U) = \mathbb{IF} = F^{-1}(T - G(U))$$

involving two different Gaussian distributions.

$$\mathrm{IF}(U) = F^{-1}\left(T - G(U)\right)$$

$$\mathrm{IF}'(U) = \frac{-G'(U)}{F'\left(F^{-1}\left(T - G(U)\right)\right)}$$

$$= \frac{-\psi(U)}{\varphi\left(\mathrm{IF}(U)\right)} < 0$$

$$U = 0 \Rightarrow \mathrm{IF}(0) = F^{-1}(T) = +\infty$$

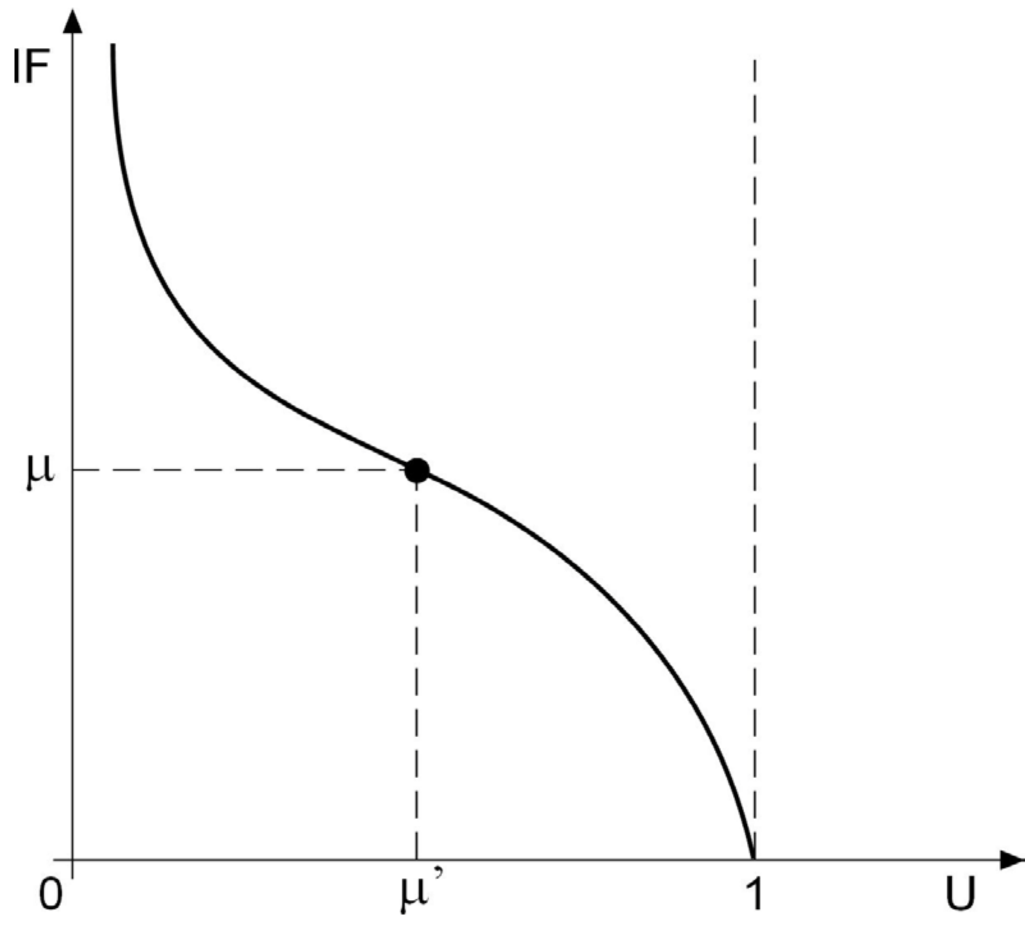
$$U = 1 \Rightarrow \mathrm{IF}(1) = F^{-1}(0) = 0$$

$$\mathrm{IF}''(\mathrm{U}) = -\frac{1}{\left(\mathrm{F}'\left(\mathrm{F}^{-1}\left(\mathrm{T}-\mathrm{G}(\mathrm{U})\right)\right)\right)^2}\frac{\mathrm{F}''\left(\mathrm{F}^{-1}\left(\mathrm{T}-\mathrm{G}(\mathrm{U})\right)\right)}{\mathrm{F}'\left(\mathrm{F}^{-1}\left(\mathrm{T}-\mathrm{G}(\mathrm{U})\right)\right)}\left(-\mathrm{G}'(\mathrm{U})\right)\left(-\mathrm{G}'(\mathrm{U})\right)$$

$$-\frac{1}{\mathrm{F}'\left(\mathrm{F}^{-1}\left(\mathrm{T}-\mathrm{G}(\mathrm{U})\right)\right)}\mathrm{G}''(\mathrm{U})$$

$$\mathrm{IF}''(\mathrm{U}) = -\frac{1}{\varphi(\mathrm{IF})^3}\varphi(\mathrm{IF})\left(-\frac{\mathrm{IF}-\mu}{\sigma^2}\right)\psi^2(\mathrm{U}) - \frac{1}{\varphi(\mathrm{IF})}\psi(\mathrm{U})\left(-\frac{\mathrm{U}-\mu'}{\sigma'^2}\right)$$

$\Rightarrow \mathrm{IF}''(\mu') = 0$ in the point (μ, μ') on the curve $(\mathrm{IF} = \mu, \mathrm{U} = \mu')$



IF
μ
0
μ'
1
U

$$f(U) = \ln(IF(U)) = \ln\left(F^{-1}(T - G(U))\right)$$

Now $f''(U)$ has the sign of

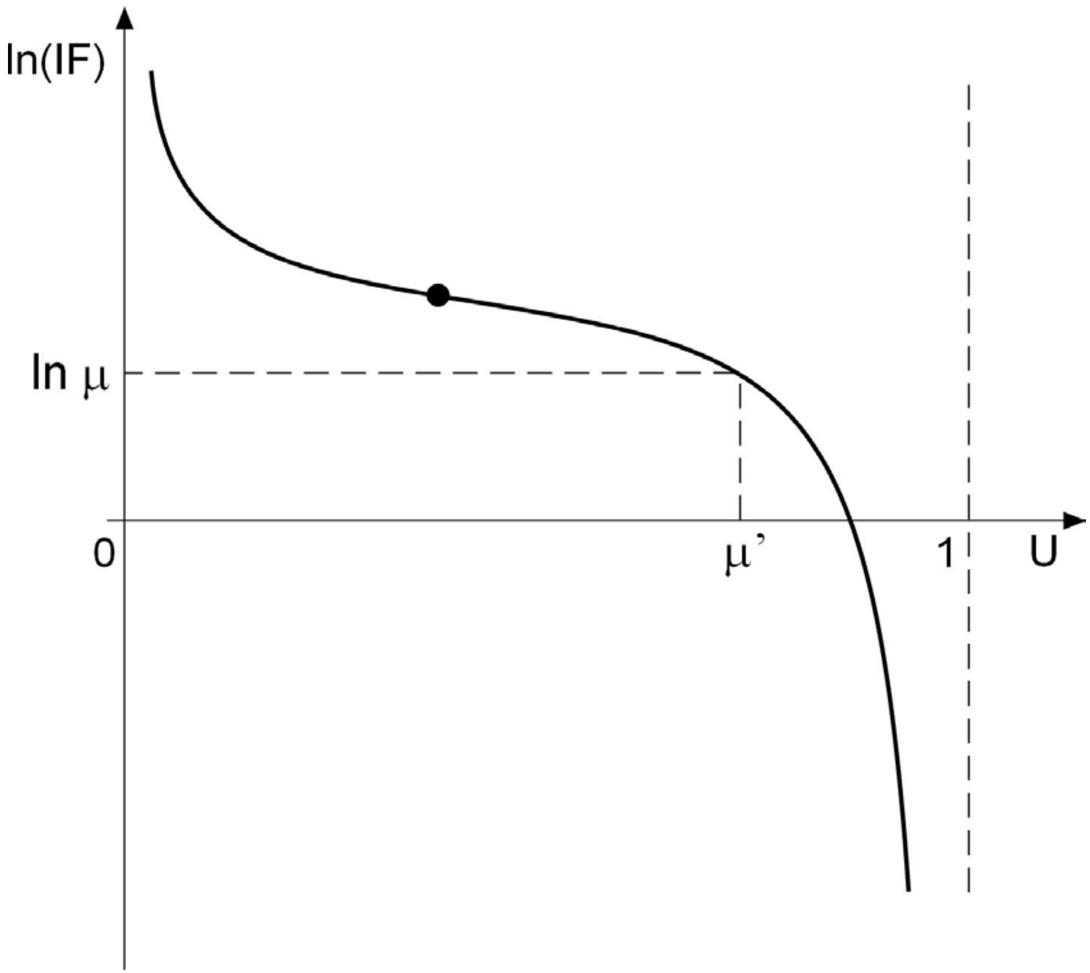
$$-\frac{\psi(U)}{IF(U)\varphi(IF(U))} + \frac{\psi(U)\left(\frac{IF(U)-\mu}{\sigma^2}\right)}{\varphi(IF(U))} + \frac{U-\mu'}{\sigma'^2}$$

We now see that in the point $(\mu', \ln\mu)$(i.e. for $U = \mu'$ and $IF = \mu$) that $f''(U)$ has the sign of

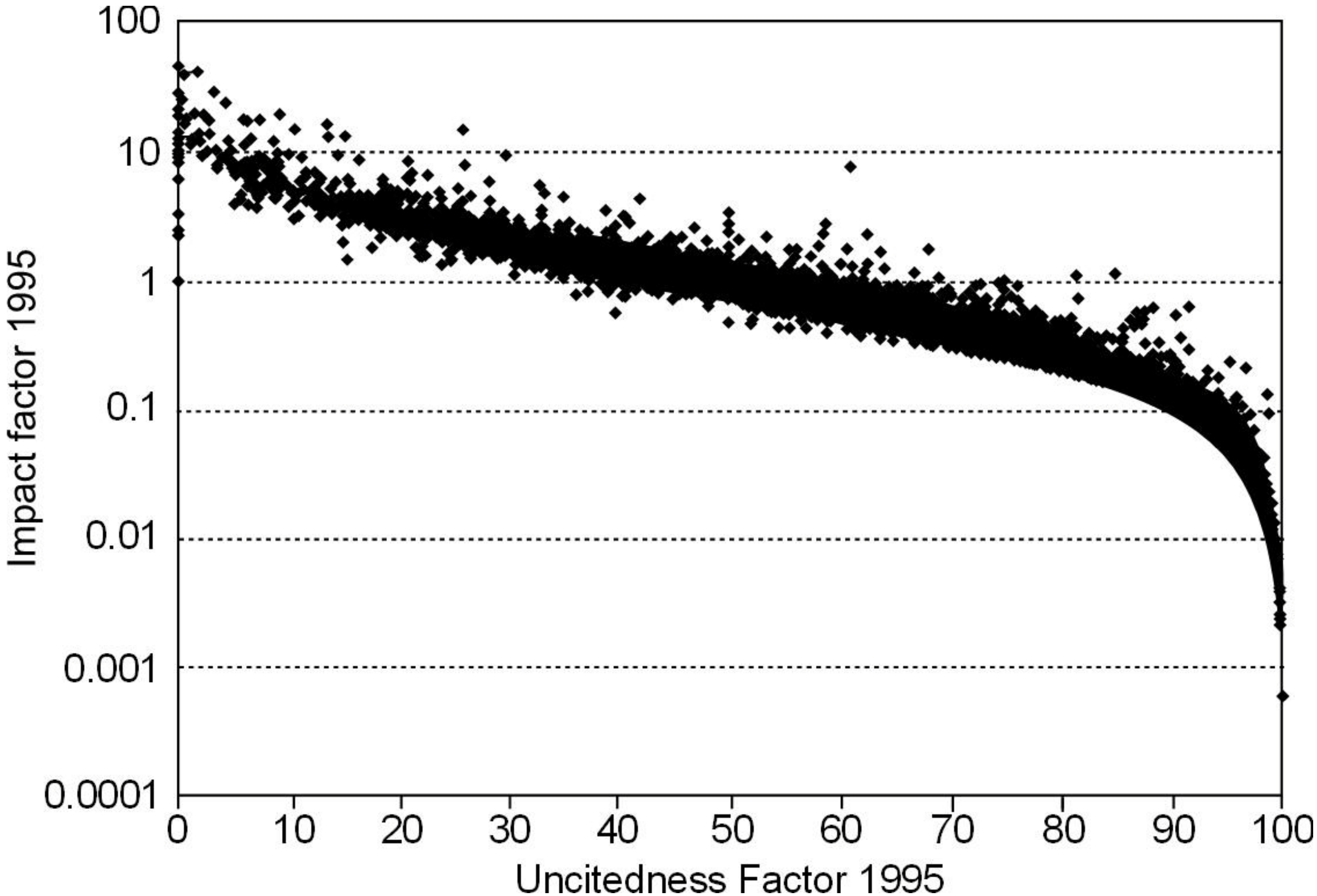
$$\frac{-\psi(U)}{IF(U)\varphi(IF(U))} < 0$$

hence strictly negative.

This implies, when compared to $IF(U)$ (where in this point one had an inflection point), that $\ln(IF(U))$ already concavely decreasing in this point.



ln(IF)
ln μ
0
μ'
1
U



ln(IF(U)), experimentally: SCI (all fields), 1995, van Leeuwen and Moed (2005)