Gini's coefficient of mean difference as a measure of adoption speed: theoretical issues and empirical evidence from India

Hans G.P. Jansen

Asian Vegetable Research and Development Center, Shanhua, Tainan, Taiwan, China

(Accepted 13 February 1992)

ABSTRACT

Jansen, H.G.P., 1992. Gini's coefficient of mean difference as a measure of adoption speed: theoretical issues and empirical evidence from India. *Agric. Econ.*, 7: 351–369.

This paper represents a variant on the empirical analysis of technological adoption by Griliches. It follows up on previous research regarding the potential of the Gini coefficient of mean difference as a distribution-free measure of adoption speed. A formal proof is supplied for the derivation of an empirically workable definition of the Gini. Data regarding the uptake of high-yielding wheat varieties in India are used to compare the performance of the Gini relative to that of the logistic and Weibull adoption speed coefficients. The results suggest that, for the purpose of ex-post analyses of monotonically increasing adoption processes, use of the Gini in measuring adoption speed can result in a better explanation of aggregate adoption behavior.

INTRODUCTION

Since Griliches' (1957) pathbreaking study on the adoption of hybrid corn in the U.S.A., many studies have followed that analyze long-run adoption behavior of innovations with aggregate time-series data (Globerman, 1975; Jarvis, 1981; Martinez, 1973; Rapoport, 1978; Wattleworth, 1980). A common feature of these studies is that they all use Griliches' two-stage method. This method estimates a growth function in the first stage and then explains the rate of adoption parameter using explanatory economic variables in the second stage. The first stage summarizes, for

---

Correspondence to: Hans G.P. Jansen, AVRDC, P.O. Box 42, Shanhua, Tainan, 74199 Taiwan, China. Tel. 886-6-5837801. Fax 886-6-5830009.

each cross-sectional entity, the adoption process in two parameters: the speed of adoption and the adoption ceiling. The adoption ceiling is defined as the long-run upper limit on aggregate adoption (Jansen et al., 1990); adoption speed is defined as the rate of adjustment towards the ceiling level.

It has been argued that alternative functional specifications such as the Weibull (Hernes, 1976; Sharif and Islam, 1980) or the Gompertz (Dixon, 1980) may trace the historical diffusion path more accurately. Different functional forms often fit the data equally well but can result in widely diverging adoption speed estimates (Jansen, 1988; Trajtenberg and Yitzhaki, 1982). However, there does not exist a satisfactory criterion that can be used to choose between different functional forms in aggregate adoption analysis. In any case, the search for an 'appropriate' functional form to describe the adoption process is a rather futile exercise since the adoption speed coefficients of different functional forms are not directly comparable (Appendix 1).

Accurate measurement of adoption speed and its determinants is important because the slowness of the diffusion process itself can be as important a cause of stagnant productivity as the slowness of the research process itself (Wattleworth, 1980). This paper follows up on Trajtenberg and Yitzhaki (1982) in arguing that, in addition to fitting one or more functional forms in the first stage, it can be worthwhile to also calculate the Gini coefficient of mean difference (hereafter referred to as the Gini) as a distribution-free measure of adoption speed. Different measurement of adoption speed is useful when it allows more accurate measurement of factors influencing adoption behavior which, in turn, may lead to more effective adoption policies. Because the Gini can be calculated directly from the time-series data, an unbiased measurement of adoption speed in the first stage can be obtained. This, in turn, can lead to a better explanation of cross-sectional variation in adoption speed in the second stage. The more accurately past adoption experience can be explained, the more effective future agricultural technology development and transfer are likely to be in accelerating adoption, avoiding potential losses of potential returns to research, and reducing undesirable distribution effects of differential adoption (Mueller and Jansen, 1988). The argument is illustrated empirically with data regarding the uptake of high-yielding varieties (HYVs) of wheat (*Triticum aestivum*) in India.

LOGISTIC ADOPTION SPEED PARAMETER

The diffusion path of aggregate adoption of a new technology often resembles a sigmoid curve, largely reflecting the dynamics of the spread of

information (Feder et al., 1982). Most previous studies have followed Griliches and specified a logistic curve in the first stage. The first-stage logistic cumulative probability distribution function (CPDF) is:

$$F_i(t) = K_i / (1 + \exp(-a_i - b_i t)) \tag{1}$$

where, in this study, $F_i(t)$ is the cumulative percentage of wheat area sown with HYVs for production region i and year t, K is the ceiling coefficient or long-run equilibrium adoption level, b is the adoption speed coefficient, and a is a constant of integration that positions the curve on the time scale.

ALTERNATIVES TO THE LOGISTIC SPECIFICATION

Although one of the main attractions of the logistic specification is that it can be easily estimated with most commonly available econometric computer software packages, it imposes certain a priori restrictions on the adoption process the validity of which is not always formally tested. Such implicit restrictions include the greatest adoption rate to occur when cumulative adoption equals half the ceiling level (i.e. $0.5\ K$) as well as symmetry around the point of inflection, thus excluding asymmetry in adoption behavior over time (Hernes, 1976).

Dixon (1980) argues in favor of the Gompertz, defined as follows (the subscript i is omitted for ease of notation):

$$F(t) = Ka^{b^t} \qquad (b < 1) \tag{2}$$

as an alternative for the logistic on the basis of a test for skewness applied to the distribution of first differences in cumulative adoption figures, which "should approximate a normal distribution" (Dixon, 1980, p. 1456). However, arguing in favor of or against a specific functional form on the basis of a skewness measure is a doubtful practice, particularly in the context of diffusion data (Jansen, 1988). Also, while it is unlikely that Dixon had more than 25 observations for any of the states he analyzed, time series of at least 30 to 40 observations are needed for a robust skewness measure.

A second point of critique of Dixon is the use of the Kolmogorov–Smirnov (K–S) test for goodness of fit. Since the maximum difference between the empirically observed and the hypothetical cumulative level of diffusion of both the logistic and the Gompertz in Dixon's case is unlikely to have exceeded 0.25 (25%), the lower bound to the power of the K–S test was probably less than 10% (Brunk, 1965; Massey, 1951).

A third point of critique stems from empirical experience gained during the present study. Although it was reported by Dixon that the Gompertz outperformed the logistic for most states in the analysis (both in terms of R^2 and significance of the estimated coefficients), this need not be the case

in general. In case of below-100% adoption, the empirical performance of the Gompertz is poor, both in terms of goodness of fit as well as in terms of convergence problems in non-linear estimation procedures. Dixon did not face this problem since he analyzed fully completed adoption processes and therefore could a priori set all K values equal to one.

In view of the above considerations, the Weibull class of distributions (Hernes, 1976; Sharif and Islam, 1980) was considered as a potentially more flexible way of measuring adoption speed. The Weibull is defined as follows (Walpole and Meyers, 1972):

$$F(t) = 1 - \exp(-at^b)$$

or, allowing for a variable ceiling level K:

$$F(t) = K(1 - \exp(-at^b)) \tag{3}$$

The Weibull CPDF curve changes in shape considerably for different values of the adoption speed parameter b. It can be symmetrical as well as non-symmetrical, with the point of inflection in either the earlier phase or the later phase of development. Depending on the value of b, the Weibull CPDF becomes left-skewed, symmetrical or right-skewed. It can be either S-shaped or concave over its entire domain.

GINI COEFFICIENT OF MEAN DIFFERENCE

The Gini was included in the analysis as a third measure of adoption speed. The Gini measures the expected time difference between any two adoption levels over the entire adoption process. The smaller (larger) the Gini, the faster (slower) the adoption process (on average). Although the concept of the Gini relates to waiting time rather than speed, its inverse can be readily interpreted as a measure of adoption speed. The Gini is defined as follows (Kendall and Stuart, 1977):

$$G = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x - y| \, \mathrm{d}F(x) \, \mathrm{d}F(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x - y| \, f(x) \, f(y) \, \mathrm{d}x \, \mathrm{d}y \tag{4}$$

where x and y are two independent, identically distributed random variables. In the context of adoption, the variable of interest is time, t_1 and t_2 being the dates (years in this study) of any two cumulative adoption levels. After reordering, (4) can be written as follows:

$$G = \int_{-\infty}^{\infty} \int_{t_1}^{\infty} (t_1 - t_2) \, f(t_2) \, f(t_1) \, \mathrm{d}t_1 \, \mathrm{d}t_2 \tag{5}$$

Alternatively, equation (4) can be rewritten as:

$$G = \int_{-\infty}^{\infty} F(t) \, (1 - F(t)) \, \mathrm{d}t \tag{6}$$

where $F(t)$ represents the CPDF (a proof is supplied in Appendix 2). Equation (6) can be rewritten by using partial integration:

$$G = t \, F(t) \, (1 - F(t)) \big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} t \, f(t) \, (1 - 2F(t)) \, \mathrm{d}t \tag{7}$$

Whenever the Gini exists (i.e. whenever the expected value of $t(\mathrm{E}(t)) < \infty$), the first term vanishes (see Appendix 2). Rearranging the second term in (7) results in the following:

$$G = 2 \int_{-\infty}^{\infty} t \, (F(t) - 0.5) \, f(t) \, \mathrm{d}t \tag{8}$$

or (using $\mathrm{d}F = \mathrm{f}(t) \, \mathrm{d}t$ and changing the integration limits accordingly):

$$G = 2 \int_{0}^{1} t(F - 0.5) \, \mathrm{d}F = 2 \, \mathrm{cov}(t, F) \tag{9}$$

The derivation of (9) from (8) is facilitated by noting that:
(a) $\mathrm{cov}(t, F) = \mathrm{E}((t - \mathrm{E}(t)) \, (F - \mathrm{E}(F))) = (t - \mathrm{E}(t)) \, (F - \mathrm{E}(F)) \, \mathrm{d}F$
    $= \mathrm{E}(t \, (F - \mathrm{E}(F))) - \mathrm{E}(t) \, \mathrm{E}(F - \mathrm{E}(F)) = \mathrm{E}(t \, (F - \mathrm{E}(F)))$, and
(b) by letting x be a random variable and $F(x)$ its cumulative distribution, F is a uniformly distributed variable. That is, $P(F < g) = g$ and hence $\mathrm{E}(F) = 0.5$.
Combining (a) and (b) results in:

$$\int_{0}^{1} t(F) \, (F - 0.5) \, \mathrm{d}F = \int_{0}^{1} (t(F) - \mathrm{E}(t \, (F))) \, (F - 0.5) \, \mathrm{d}F = \mathrm{cov}(t, F)$$

AN APPLICATION: ADOPTION OF HYVs OF WHEAT IN INDIA

Wheat HYVs were released throughout India in the 1960s. They represented a substantial change from local varietal types. HYVs were photo-period-insensitive, fertilizer-responsive, and short-statured. Grown with good management on fertile soil with access to reliable rainfall and/or irrigation facilities, they gave markedly higher grain yields than traditional wheat varieties. By the mid 1980s, the wheat HYVs had largely completed their diffusion process (Table 1). Adoption is measured as the proportion of total wheat area planted to HYVs.

Concepts and methods for the first stage. Equations (1) and (3) were both estimated with Marquardt's numerical optimization method (Judge et al.,

TABLE 1

Area, production, yield and HYV adoption of wheat in India (selected years)

Crop year	Area (million ha)	Production (million t)	Yield (t/ha)	HYV adoption (% of area)
1949–50	9.9	6.3	0.64	–
1956–57	12.4	9.3	0.75	–
1960–61	13.0	10.8	0.83	–
1964–65	13.5	12.3	0.91	–
1967–68	14.9	16.6	1.10	15.2
1970–71	18.2	23.8	1.30	41.1
1974–75	18.0	24.1	1.30	74.7
1980–81	22.3	36.3	1.60	72.3
1984–85	23.6	44.2	1.90	83.1

Sources: Gov. India, various issues of *Indian Agriculture in Brief; Agricultural Situation in India;* and *Economic Survey;* and Dalrymple (1986).

1980). Values for the Gini were obtained via equation (9). In order to be able to compare the logistic, Weibull and Gini adoption speed measures, they need to be brought under a common denominator. Appendix 3 explains how the Gini-equivalent of the logistic and Weibull adoption speed parameters were obtained. These are then directly comparable to the values of equation (9).

Concepts and methods for the second stage. In the second stage, the results of two regression models were compared. Both models aim at an ex-post explanation of the interregional variation in the speed of adoption of wheat HYVs in India and correspond to the second stage in Griliches' two-stage methodology. In one model the dependent variable is the logistic adoption speed parameter (model 1). In the other model the dependent variable is the Gini coefficient of mean difference (model 2). The model with the Weibull adoption speed parameter as the dependent variable was statistically insignificant.

The two most commonly used functional forms in the literature for the purpose of explaining cross-sectional variation in adoption speed are the simple linear relation and the loglinear specification. The loglinear form allows the effect on the dependent variable of each independent variable to depend on the level of the other independent variable(s) (Mansfield, 1961). Since the effect of yield variability is presumably smaller the larger the average yield difference between traditional varieties and HYVs, the loglinear specification was considered to be more appropriate than a linear form.

In order to attain interdistrict comparability, the estimated adoption speed coefficients (b) in model 1 were multiplied by their corresponding ceiling levels (K), resulting in bK (rather than b) as the dependent variable (Griliches, 1957). Similarly, in model 2 with the Gini (G) as the dependent variable, using p for the adoption level reached in 1984/85 (the last year for which data were available), the dependent variable was G/p rather than G. Since $p \leqslant 1$, the Gini gets 'inflated' for those districts where the adoption process has not reached completion.

In addition, in model 1, the inverses of the standard errors of the estimated adoption speed coefficients were used in a weighted least squares procedure to derive more accurate parameter estimates in the second stage (Wattleworth, 1980). Each error term e_i was assumed normally distributed with variance σ_i^2, where $\text{Var}(e_i) = \text{E}(e_i^2) = \sigma_i^2$ is not constant across production regions [1].

DATA

The analysis is based on secondary annual crop area and production data from the 65 most important wheat-growing districts in five states in India (Punjab, Haryana, Uttar Pradesh, Madhya Pradesh and Rajasthan). The data were assembled by the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) and the World Bank largely from the State Season and Crop Reports and Statistical Abstracts. A district was included in the sample if it accounted for at least 0.5% of the average total production of wheat during the agricultural years 1981/82, 1982/83 and 1983/84. The study districts accounted for over 99% of the all-India wheat production. The time series spanned 1966/67 to 1984/85.

DETERMINANTS OF INTERDISTRICT VARIATION IN ADOPTION SPEED

Interregional differences in the adoption speed parameter are usually interpreted as differences in the rate of adjustment of demand to the new equilibrium position (Feder et al., 1982). In India, however, supply restrictions are likely to play a role as well, even though reliable data are lacking. Although well-adapted wheat HYVs have been available for irrigated areas since the mid 1960s, such varieties have not yet been developed for most rainfed conditions (Jansen, 1988). The regional intensity of advertising and

---

[1] The dependent variable in the loglinear second-stage regression models is $\log(bK)$, the variance and standard deviation of which are $1/b^2 \text{var}(b)$ and $1/b \, \text{std}(b)$, respectively (Kendall and Stuart, 1977, p. 247).

promotional activities of seed companies and regional expenditures of extension agencies might also significantly influence adoption speed.

Due to data limitations, however, supply-related factors had to be largely ignored and only conditions of demand could be taken into account. The explanatory variables in the models included yield increase as a proxy measure for profitability and annual yield variance prior to the introduction of HYVs as a proxy measure for risk [2]. Since data regarding interregional variation in cost of cultivation and prices were also not available, the profitability of HYVs was approximated by the absolute yield increase over traditional varieties which, in turn, was approximated by the difference in average yield during the period 1979/80 to 1983/84 and the average yield during the period 1961/62 to 1964/65 (measured in metric tonnes per hectare) [3]. The more profitable an innovation, the faster its expected adoption (other things being equal). Therefore, the sign of the profitability proxy was expected to be positive in model 1 and negative in model 2 (recall that the dependent variable in model 2 can be interpreted as the inverse of adoption speed).

Production risk is mainly determined by physical and climatic conditions which do not change as a result of the introduction of HYVs. Since intertemporal yield variability is a prime indicator of production risk, the latter was approximated by the coefficient of variation of pre-HYV wheat yields (based on yield data for the period spanning the period 1956/57 to 1965/66 and measured as a percentage) [4]. The sign of the production risk proxy was expected to be negative in model 1 (higher risk slows down the adoption process) and positive in model 2 (the higher the risk, the longer the average waiting time between two adoption levels).

First-stage results. The average value for the Gini, as calculated directly from the time-series data via equation (9), came to 2.6 years (Table 2). The

---

[2] Only in the absence of uncertainty can it be expected that any positive difference in expected return between the use of a HYV versus a traditional variety would induce a complete switch to the new variety. However, assuming that at least some producers are risk-averse, the presence of production risk becomes a rationale for a less than complete switch.

[3] The implicit assumption here is that the entire yield increase is due to the use of HYVs. Although it is recognized that other factors such as increased use of fertilizers played an important role as well, the assumption is considered a workable one in cross-sectional comparisons.

[4] Martinez (1973) suggests the variance of the ratio of the area actually harvested and the total area seeded as a useful indicator of production risk, this variance being essentially determined by agroclimatic conditions. However, separate data regarding the areas seeded and harvested were not available. Similarly, data regarding differences in intertemporal yield variability between the traditional varieties and HYVs were also not available.

TABLE 2

Summary of first-stage estimation results [a]

Variable [b]	Mean	Coefficient of variation	Minimum value	Maximum value
Gini (G)	2.64	0.25	0.76	3.82
Ginilog	2.88	0.24	1.17	5.39
Giniweib	2.70	0.25	0.95	4.15
|Ginilog − Gini|	0.60	1.00	0.01	3.08
|Giniweib − Gini|	0.43	0.95	0.00	2.21
Adoption ceiling (K)	0.84	0.23	0.13	1.00
G/p	3.47	0.50	0.76	9.63

[a] Number of observations is 65.
[b] Variable definitions as follows:
Gini, Gini coefficient of mean difference (G) in years, calculated via equation (9);
Ginilog, Gini-equivalent of the estimated logistic adoption speed coefficient, in years, calculated via equation (A10) in Appendix 3;
Giniweib, Gini-equivalent of the estimated Weibull adoption speed coefficient, in years, calculated via equation (A11) in Appendix 3;
\|Ginilog − Gini\|, absolute value of the difference between Ginilog and Gini, in years.
\|Giniweib − Gini\|, absolute value of the difference between Giniweib and Gini, in years.
G/p, Gini coefficient of mean difference (G) divided by the cumulative adoption level attained in 1984/85.

variables Ginilog and Giniweib in Table 2 represent the Gini-equivalent of the estimated logistic and Weibull adoption speed parameters, respectively (calculation methods are explained in Appendix 3). The performance of the estimated logistic and Weibull adoption speed parameters is assessed by their absolute deviation from the raw data Gini, and measured by the absolute values of (Ginilog − Gini) and (Giniweib − Gini), respectively. Of the 65 districts in the sample, (Giniweib − Gini) exceeded (Ginilog − Gini) for 34 districts. Thus, although both the logistic and the Weibull adoption speed parameter estimates deviate from the Gini adoption speed measure by about 20% on average, there is little difference between the performance of the two functional specifications (Table 2).

This result did not come as a surprise. Most wheat is irrigated, and suitable HYVs are available for most irrigated environments including most districts in the Punjab, Haryana and western Uttar Pradesh. Adoption patterns in these districts often conform well to the logistic S-shaped pattern and rapidly approached ceiling levels of 90% or more. In irrigated districts the yield difference between HYVs and traditional wheat varieties is well above 1 t/ha and can be as high as 2 t/ha. Seed supply is not generally a major factor inhibiting adoption. Although adoption of wheat HYVs in the rainfed districts of the eastern and northern parts of Uttar

TABLE 3

Loglinear regressions of adoption speed parameters on profitability and production risk variables: Wheat

Variables	Variable statistics				Logistic-based (Model 1)		Gini-based (Model 2)	
	Mean	CV	Min	Max	Coeff.	SE	Coeff.	SE
Independent								
Intercept					−0.81	0.68	1.15	0.37
Profitability (t/ha)	1.04	0.37	0.19	1.99	0.77	0.13	−0.97	−0.09
Production risk	19.02	0.29	0.06	0.30	−0.04	−1.09	−0.01	−0.12
Dependent								
Gini-based adoption speed measure (G/p)	3.79	0.82	0.77	24.31				
Logistic-based adoption speed measure (bk)	0.44	0.51	0.03	1.58				
N					65		65	
$\bar{R}^2$					0.38		0.65	
F					20.30		61.11	
CORR [a]					0.39		0.90	

[a] CORR refers to the Pearson correlation coefficient between actual and predicted values.

Pradesh, Rajasthan and Madhya Pradesh was substantially slower and much more erratic, resulting in a poorer fit of both the logistic and the Weibull, these districts formed only a minor part of the total sample.

Second-stage results. In the second stage both the logistic-based and the Gini-based adoption speed parameters were related to profitability and production risk variables. The parameters of the independent variables are, as a group, significantly different from zero in both second-stage models (Table 3). The models are able to explain up to 65% of the total variation in adoption speed between districts. The results of model 1 are well in line with those obtained by Griliches (1957) and Martinez (1973), who analyzed adoption of maize hybrids under similar conditions (i.e. adoption had been more or less monotonically increasing in all regions with no major supply restrictions). Risk-related considerations do not seem to have had much impact on the adoption of wheat HYVs which is mainly conditioned by profitability. This is not surprising since wheat is mostly grown under

relatively stable, irrigated conditions which has resulted in average yield increases of over 1 t/ha (Table 3) [5].

The $\bar{R}^2$ statistic may not be legitimately used to compare the performance of model 1 with model 2 because the dependent variables are different (Rao and Miller, 1971). Model 2 may, however, be regarded as empirically preferable to model 1 because it fits the data better and because its profitability coefficient which is statistically significant in model 1 is also significant in model 2 but with a lower standard error. These regression results provide empirical support for the Gini coefficient of mean difference as a potentially useful alternative measure of adoption speed, since it might allow for a more accurate explanation of cross-sectional variation in adoption speed.

CONCLUSION

The empirical results presented for the diffusion of wheat HYVs in India suggest that, for the purpose of analyzing aggregate adoption behavior in the case of monotonically increasing diffusion processes, the Gini coefficient of mean difference is worthy of being considered as an alternative measure of adoption speed. Thus, analysts of such diffusion processes should include the Gini in their search for appropriate measures and explanations of adoption speed. In cases where the diffusion process does not exhibit a strictly increasing trend over time and/or is still relatively far from completion, the Gini coefficient of mean difference performs less well, and functional-form dependent adoption speed measures are usually preferable (Jansen, 1988). When the researcher's interest focuses on adoption ceilings rather than on adoption speed (e.g. Jansen et al., 1990), a certain functional form will have to be employed since it is impossible to estimate a ceiling without a functional form. In such cases, comparisons of the Gini calculated directly from the time-series data with the Gini-equivalent of functional form-dependent adoption speed parameters (and taking the ceiling level from the functional form whose Gini-equivalent of the adoption speed parameter exhibits minimum absolute deviation from the Gini calculated directly from the data) can assist in choosing among alternative functional forms. Thus, in addition to its use as a non-parametric measure of adoption speed, the Gini might also be employed as an alternative to residual variance measures like the coefficient of determination (R^2) in choosing between alternative functional forms in diffusion analysis.

---

[5] In the Punjab, most of Haryana, and western Uttar Pradesh the intertemporal coefficient of variation of wheat yields is well below 10%.

ACKNOWLEDGMENTS

APPENDIX 1

Meaning of diffusion speed coefficients

The implicit meaning of the diffusion speed coefficient b for each of the three following functional forms is considered in this appendix: the logistic, the Weibull and the Gompertz. This is done by assessing the coefficient b in calculating the time for diffusion to go from 20% to 80%. In order to simplify the algebra, K was set to 1 (or 100%).

Logistic.   The logistic function is defined as:

$$F(t) = K/(1 + \exp(-a - bt))$$

For $K = 1$,

$$F(t) = \ln(F(t)/1 - F(t)) = a + bt$$

Thus:

$$t = 1/b \, \ln(F(t)/1 - F(t)) - a$$

$$t(F2) - t(F1) = 1/b \, (\ln(F2/1 - F2) - \ln(F1/1 - F1))$$

$$t(F2) - t(F1) = 1/b \, \ln((F2/1 - F2)(1 - F1)/F1))$$

Let $F2 = 0.8$ and $F1 = 0.2$; it then takes $2.77/b$ time periods for diffusion to go from 20% to 80%.

Weibull.   The Weibull function is defined as:

$$F(t) = K(1 - \exp(at^b))$$

For $K = 1$,

$$F(t) = 1 - \exp(-at^b)$$

Thus:

$$1 - F(t) = \exp(-at^b)$$

$$F(t)/(1 - F(t)) = \exp(at^b) - 1$$

$$\exp(at^b) = F(t)/(1 - F(t)) + 1 = 1/(1 - F(t))$$

$$t = (1/a \; \ln(1/(1 - F(t))))^{1/b}$$

Therefore:

$$t(F2) - t(F1) = a^{1/b}(\ln((1 - F1)/(1 - F2)))^{1/b}$$

Let $F2 = 0.8$ and $F1 = 0.2$; it then takes $(1/a)^{1/b}((\ln(5))^{1/b} - \ln(1.25))^{1/b})$ time periods for diffusion to go from 20% and 80%.

Gompertz. The Gompertz function is defined as:

$$F(t) = Ka^{b^t} \quad (b < 1)$$

For $K = 1$,

$$F(t) = a^{b^t} \quad (b < 1)$$

$$\ln F(t) = b^t \; \ln(a) = \exp(\ln(b^t)) \; \ln(a) = \exp(t \; \ln(b)) \; \ln(a)$$

$$\ln(\ln(F(t)) = t \; \ln(b) + \ln(\ln(a))$$

Thus:

$$t = (\ln \ln(F(t)) - \ln \ln(a))/\ln(b)$$

and

$$t(F2) - t(F1) = \ln(\ln(F2)/\ln(F1))/\ln(b)$$

Let $F2 = 0.8$ and $F1 = 0.2$; it then takes $-1.98/\ln(b)$ time periods for diffusion to go from 20% to 80% (note that, as $b < 1$, $\ln(b) < 0$).

To conclude, diffusion speed coefficients from different functional forms are not directly comparable.

APPENDIX 2

Re-expressing the Gini in a convenient form

The Gini coefficient of mean difference is proved in this appendix and is defined as:

$$G = 0.5 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x - y| \; f(x) \; f(y) \; \mathrm{d}x \; \mathrm{d}y \tag{A1}$$

which can be written as:

$$G = \int_{-\infty}^{\infty} F(t) \; (1 - F(t)) \; \mathrm{d}t \tag{A2}$$

Proof:

$$G = 0.5 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x - y| \; f(x)\, f(y)\, \mathrm{d}x\, \mathrm{d}y$$

$$[\text{noting that } 0.5\,|x-y| = (x+y)/2 - \min(x,\,y)]$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ((x+y)/2 - \min(x,\,y))\; \mathrm{f}(x)\, \mathrm{f}(y)\, \mathrm{d}x\, \mathrm{d}y$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 0.5x\, f(x)\, f(y)\, \mathrm{d}x\, \mathrm{d}y + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 0.5y\, f(x)\, f(y)\, \mathrm{d}x\, \mathrm{d}y$$

$$- \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \min(x,\,y)\, f(x)\, f(y)\, \mathrm{d}x\, \mathrm{d}y$$

$$= \int_{-\infty}^{\infty} 0.5x\, f(x)\, \mathrm{d}x \int_{-\infty}^{\infty} f(y)\, \mathrm{d}y + \int_{-\infty}^{\infty} 0.5y\, f(y)\, \mathrm{d}y \int_{-\infty}^{\infty} f(x)\, \mathrm{d}x$$

$$- \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \min(x,\,y)\, f_{x,y}(x,\,y)\, \mathrm{d}x\, \mathrm{d}y$$

Letting $m \equiv \min(x,\,y)$, i.e. $\mathrm{E}(\min(x,\,y)) \equiv \mathrm{E}(m)$, gives:

$$G = \int_{-\infty}^{\infty} 0.5x\, f(x)\, \mathrm{d}x + \int_{-\infty}^{\infty} 0.5y\, f(y)\, \mathrm{d}y - \mathrm{E}(\min(x,\,y))$$

The dummy y can be called x:

$$G = \int_{-\infty}^{\infty} 0.5x\, f(x)\, \mathrm{d}x + \int_{-\infty}^{\infty} 0.5x\, f(x)\, \mathrm{d}x - \mathrm{E}(m)$$

$$= \int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x - \mathrm{E}(m) \tag{A3}$$

Consider $\mathrm{E}(m)$:

$$\mathrm{E}(m) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \min(x,\,y)\, f(x)\, f(y)\, \mathrm{d}x\, \mathrm{d}y$$

Alternatively $\mathrm{E}(m)$ can be computed as:

$$\mathrm{E}(m) = \int_{-\infty}^{\infty} m\, f(m)\, \mathrm{d}m \qquad \text{where } f(m) \text{ is the PDF of } m \tag{A4}$$

Equation (A4) can be derived by:
- noting that $f(x)\, f(y) = f_{x,y}(x,\,y)$ (i.e. the joint density of $(x,\,y)$)
- changing variables from $(x,\,y)$ to $(m,\,M)$ (let $M \equiv \max(x,\,y)$)
- integrating out M.

That is,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \min(x, y) \, f(x) \, (f(y) \, \mathrm{d}x \, \mathrm{d}y$$

$$= \int_{-\infty}^{\infty} \int_{\infty}^{\infty} \min(x, y) \, f(x) \, f(y) \, \mathrm{d}x \, \mathrm{d}y$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \min(m, M) \, f_{m,M}(m, M) \, \mathrm{d}m \, \mathrm{d}M = \int_{-\infty}^{\infty} m \, f(m) \, \mathrm{d}m$$

Note also that $f_m(m) = 2(1 - F(m)) \, f(m)$, where $F(m)$ and $f(m)$ are the CPDF and the PDF of x and y, respectively. Here an application of theorem 11 in Mood et al. (1974) for the $n = 2$ case is used:

$$F_y(y) = \binom{2}{1} (F(y)^1 \, (1 - F(y))^{2-1} + \binom{2}{1} (F(y))^2 = 2 \, F(y) - F(y)^2$$

and

$$f_y(y) = \mathrm{d}F_y(y)/\mathrm{d}y = 2 \, f(y) \, (1 - F(y))$$

Thus:

$$\mathrm{E}(m) = \int_{-\infty}^{\infty} 2m \, (1 - F(m)) \, f_m(m) \, \mathrm{d}m \qquad (A5)$$

Substituting (A5) into (A3) gives:

$$G = \int_{-\infty}^{\infty} x \, f(x) \, \mathrm{d}x - \int_{-\infty}^{\infty} m(2(1 - F(m)) \, f(m)) \, \mathrm{d}m$$

Again, the dummy y can be called x:

$$G = \int_{-\infty}^{\infty} x \, f(x) \, \mathrm{d}x - \int_{-\infty}^{\infty} 2x \, (1 - F(x)) \, f(x) \, \mathrm{d}x$$

$$= \int_{-\infty}^{\infty} (x \, f(x) - 2x \, f(x) + 2x \, F(x) \, f(x)) \, \mathrm{d}x$$

$$= \int_{-\infty}^{\infty} x \, (2 \, F(x) - 1) \, f(x) \, \mathrm{d}x \qquad (A6)$$

Using the partial integration rule,

$$\int_a^b v \, \mathrm{d}u = uv \, |_a^b - \int_a^b \mathrm{d}v \, u$$

and considering $x = u \, [\mathrm{d}x = \mathrm{d}u]$ and $\mathrm{d}v = 2 \, F(x) - 1 \, [v = F(x)^2 - F(X)]$:

$$G = x \left(F(x)^2 - F(x) \right) |_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \left(F(x)^2 - F(x) \right) \, \mathrm{d}x$$

$$= x \, F(x) \, (F(x) - 1) |_{-\infty}^{\infty} + \int_{-\infty}^{\infty} F(x) \, (1 - F(x)) \, \mathrm{d}x \qquad (A7)$$

Thus, a sufficient condition for (A1) to equal (A2) is that if the first term in (A7) vanishes. This is indeed the case for a large class of distributions:

$$
\begin{aligned}
x\,F(x)\,(F(x)-1)\,|_{-\infty}^{\infty} &= \lim_{x\to\infty} x\,F(x)\,(F(x)-1) - \lim_{x\to\infty} x\,F(x)\,(F(x)-1) \\
&= \left(\lim_{x\to\infty} x\,(F(x)-1)\right)\left(\lim_{x\to\infty} F(x)\right) \\
&\quad - \left(\lim_{x\to-\infty}(F(x)-1)\right)\left(\lim_{x\to-\infty} x\,F(x)\right) \\
&= \lim_{x\to\infty} x\,(F(x)-1) + \lim_{x\to-\infty} x\,F(x) \\
&= \lim_{x\to\infty}(x\,(F(x)-F(-x)-1)) \\
&= \lim_{x\to\infty}(-x(1-F(x)+F(-x)))
\end{aligned}
\tag{A8}
$$

A sufficient condition for (A8) to be zero is that $E(x) < \infty$, or:

$$
\int_{-\infty}^{\infty} |x|\,\mathrm{d}F(x) + \int_{-\infty}^{\infty} |x|\,f(x)\,\mathrm{d}x = E(x) < \infty
$$

This can be seen as follows:

$$
\text{if}\quad \int_{-\infty}^{\infty} |x|\,\mathrm{d}F(x) < \infty \quad\text{then}\quad \lim \int_{|x|>t}^{\infty} |x|\,\mathrm{d}F(x) = 0
$$

i.e. for the integral to be finite, the 'tail' area of the probability distribution function must go to zero. Although there exist distributions for which this condition is not satisfied (e.g. the Cauchy distribution), it is satisfied for most other distributions.

For $t > 0$,

$$
\int_{|x|>t} |x|\,\mathrm{d}F(x) \geq \int_{|x|>t} |t|\,\mathrm{d}F(x) = |t|\int_{|x|>t} \mathrm{d}F(x)
$$

$$
= |t|\left(\int_{-\infty}^{-t}\mathrm{d}F(x) + \int_{t}^{\infty}\mathrm{d}F(x)\right)
$$

Therefore:

$$
\int_{|x|>t} |x|\,\mathrm{d}F(x) \geq |t|(F(-t)-F(-\infty)+F(\infty)-F(t))
$$

and

$$
\int_{|x|>t} |x|\,\mathrm{d}F(x) \geq t(F(-t)-0+1-F(t)) = t(1-F(t)+F(-t)) \geq 0
$$

$$
0 = \lim_{t\to\infty} \int_{|x|>t} |x|\,\mathrm{d}F(x) \geq \lim_{t\to\infty} t(1-F(t)+F(-t)) \geq 0
$$

$$\lim_{t \to \infty} t\big((1 - F(t) + F(-t))\big) = 0$$

and

$$\lim_{t \to \infty} -t(1 - F(t) + F(-t)) = 0$$

Thus:

$$x\,F(x)\,(F(x) - 1)\,|_{-\infty}^{\infty} = 0$$

if $F(x)$ is such that $\mathrm{E}(x) < \infty$. *End proof.*

APPENDIX 3

Gini-equivalent of the logistic and Weibull adoption speed parameters

This appendix shows how to calculate the Gini from both the logistic and the Weibull adoption speed coefficients.

Gini-equivalent of the logistic adoption speed parameter. The logistic is defined as follows:

$$F(t) = (1 - \exp(-a - bt))^{-1}$$
$$1 - F(t) = (\exp(-a - bt))/(1 + \exp(-a - bt))$$

Thus,

$$G = \int_{-\infty}^{T'} F(1 - F)\,\mathrm{d}t = \int_{-\infty}^{T'} (\exp(-a - bt))/(1 + \exp(-a - bt))\,\mathrm{d}t$$

$$= 1/(b(1 + \exp(-a - bt)))\,|_{-\infty}^{T'} = p/b \qquad\qquad (A9)$$

Here $p \equiv F(T')$ and T' represents the last time period for which there are data available.

In case of a variable ceiling level K, the logistic is defined as follows:

$$F(t) = K(1 + \exp(-a - bt))^{-1}$$
$$1 - F(t) = 1 - K/(1 + \exp(-a - bt))$$

$$G = \int_{-\infty}^{T'} F(1 - F)\,\mathrm{d}t = K(1 - K)\int_{-\infty}^{T'} 1/(1 - \exp(-a - bt))^2\,\mathrm{d}t$$

$$+ K\int_{-\infty}^{T'} \Big(\exp(-a - bt)(1 + \exp(-a - bt))^{-2}\Big)\,\mathrm{d}t$$

[using (A9)]

$$= K(1 - K)\int_{-\infty}^{T'} \Big(1 + \exp(-a - bt)^{-2}\Big)\,\mathrm{d}t + K\,p/b \qquad\qquad (A10)$$

The solution to the first part was found with the help of MACSYMA, a mathematical library, and can be written as follows:

$$K(1 - K)((\exp(bt + a) + 1) \ln(\exp(-bt - a)(\exp(bt + a) + 1)))$$

$$+ (\exp(a)bt + (a - 1) \exp(a))(\exp(bt) + bt + a)/(b \exp(bt + a) + b)$$

Gini-equivalent of the Weibull adoption speed parameter. The Weibull with variable ceiling is defined as follows:

$$F(t) = K - K \exp(-at^b) = K(1 - \exp(-at^b))$$

$$1 - F(t) = 1 - K + K \exp(-at^b)$$

$$F(1 - F) = K((1 - K) + (2K - 1) \exp(-at^b) - K \exp(-2at^b))$$

Thus:

$$G = \int_0^{T'} F(1 - F) \, dt$$

$$= K \int_0^{T'} ((1 - K) + (2K - 1) \exp(-at^b) - K \exp(-2at^b)) \, dt$$

$$= K \left(\int_0^{T'} (1 - K) \, dt + (2K - 1) \int_0^{T'} \exp(-at^b) \, dt - K \exp(-2at^b) \, dt \right)$$

$$= K \left((1 - K)t \big|_0^{T'} + (2K - 1) \int_0^{T'} \exp(-at^b) \, dt - K \exp(-2at^b) \, dt \right)$$

$$= K \left((1 - K)T' + 2(K - 0.5) \int_0^{T'} \exp(-at^b) \, dt - K \exp(-2at^b) \, dt \right)$$

$$\tag{A11}$$

The last two parts of this expression can be evaluated by numerical integration for given values of the coefficients a, b and K, with the help of MACSYMA.

REFERENCES

Brunk, H.D., 1965. An Introduction to Mathematical Statistics (2nd Edition). Blaisdell, Waltham, MA.
Dalrymple, D., 1986. Development and spread of high-yielding wheat varieties in developing countries. Bureau for Science and Technology, Agency for International Development, Washington, DC.
Dixon, R., 1980. Hybrid corn revisited. Econometrica, 48: 1451–1461.
Feder, G., Just, R. and Zilberman, D., 1982. Adoption of agricultural innovations in developing countries: a survey. Staff Work. Pap. 542, World Bank, Washington, DC.

Globerman, S., 1975. Technological diffusion in the Canadian tool and die industry. Rev. Econ. Stat., 56: 428–434.

Gov. India, various years. Agricultural Situation in India, various volumes. Ministry of Agriculture and Irrigation, Directorate of Economics and Statistics, New Delhi.

Gov. India, various years. Indian Agriculture in brief, various volumes. Department of Agriculture and Cooperation, New Delhi.

Gov. India, various years. Economic Survey, various volumes. Economic Division, Ministry of Finance, New Delhi.

Griliches, Z., 1957. Hybrid corn: an exploration in the economics of technological change. Econometrica, 25: 501–522.

Hernes, G., 1976. Diffusion and growth – the non-homogeneous case. Scand. J. Econ., 78: 427–436.

Jansen, H.G.P., 1988. Adoption of modern cereal cultivars in India: Determinants and implications of interregional variation in the speed and ceiling of diffusion. Ph.D. thesis, Cornell University, Ithaca, NY.

Jansen, H.G.P., Walker, T.S. and Barker, R., 1990. Adoption ceilings and modern coarse cereal cultivars in India. Am. J. Agric. Econ., 72: 653–663.

Jarvis, L.S., 1981. Predicting the diffusion of improved pastures in Uruguay. Am. J. Agric. Econ., 63: 495–502.

Judge, G.G., Griffiths, W.E., Hill, H.C. and Lee, T.C., 1980. The Theory and Practice of Econometrics. Wiley, New York.

Kendall, M.G. and Stuart, A., 1977. The advanced Theory of Statistics. Three-volume edition, volume 1, 4th edition, Griffin, London.

Mansfield, E., 1961. Technical change and the rate of imitation. Econometrica, 29: 741–766.

Martinez, J.C., 1973. On the economics of technological change: induced innovation in Argentine agriculture. Ph.D. thesis, Iowa State University, Ames, IA.

Massey, F.J., 1951. The Kolmogorov–Smirnov test for goodness of fit. J. Am. Stat. Assoc., 49: 68–78.

Mood, A.M., Graybill, F.A., and Boes, D.C., 1974. Introduction to the Theory of Statistics (3rd Edition). McGraw-Hill, New York.

Mueller, R.A.E. and Jansen, H.G.P., 1988. Farmer and farm concepts in measuring adoption lags. J. Agric. Econ., 39: 121–124.

Rao, P. and Miller, R.L., 1971. Applied Econometrics. Wadsworth, Belmont, CA.

Rapoport, J., 1978. Diffusion of technological innovation among non-profit firms: a case study of radioisotopes in US hospitals. J. Econ. Bus., 30: 108–118.

Sharif, M. and Islam, M., 1980. The Weibull distribution as a general model for forecasting technological change. Technol. Forecast. Social Change, 18: 247–256.

Trajtenberg, M. and Yitzhaki, S., 1982. The diffusion of innovations: a methodological reappraisal. Work. Pap. 1008, National Bureau for Economic Research, Cambridge, MA.

Walpole, R.E. and Meyers, R.H., 1972. Probability and Statistics for Engineers and Scientists. MacMillan, New York.

Wattleworth, M.A., 1980. Diffusion of technology in Malaysia's rubber estate sector. Ph.D. thesis, University of California, Berkeley, CA.