

UC Berkeley

Earlier Faculty Research

Title

Does Generalizing Density Functions Better Explain Urban Commuting? Some Evidence from the Los Angeles Region

Permalink

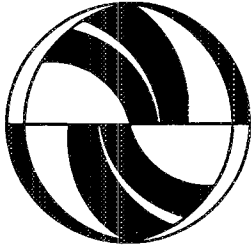
<https://escholarship.org/uc/item/353631r6>

Author

Song, Shunfeng

Publication Date

1994



**Does Generalizing Density Functions
Better Explain Urban Commuting?
Some Evidence from the Los Angeles Region**

Shunfeng Song

Working Paper
UCTC No. 197

**The University of California
Transportation Center**

University of California
Berkeley, CA 94720

**The University of California
Transportation Center**

The University of California Transportation Center (UCTC) is one of ten regional units mandated by Congress and established in Fall 1988 to support research, education, and training in surface transportation. The UC Center serves federal Region IX and is supported by matching grants from the U.S. Department of Transportation, the California Department of Transportation (Caltrans), and the University.

Based on the Berkeley Campus, UCTC draws upon existing capabilities and resources of the Institutes of Transportation Studies at Berkeley, Davis, Irvine, and Los Angeles; the Institute of Urban and Regional Development at Berkeley; and several academic departments at the Berkeley, Davis, Irvine, and Los Angeles campuses. Faculty and students on other University of California campuses may participate in

Center activities. Researchers at other universities within the region also have opportunities to collaborate with UC faculty on selected studies.

UCTC's educational and research programs are focused on strategic planning for improving metropolitan accessibility, with emphasis on the special conditions in Region IX. Particular attention is directed to strategies for using transportation as an instrument of economic development, while also accommodating to the region's persistent expansion and while maintaining and enhancing the quality of life there.

The Center distributes reports on its research in working papers, monographs, and in reprints of published articles. It also publishes *Access*, a magazine presenting summaries of selected studies. For a list of publications in print, write to the address below.



**University of California
Transportation Center**

108 Naval Architecture Building
Berkeley, California 94720
Tel: 510/643-7378
FAX: 510/643-5456

The contents of this report reflect the views of the author who is responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California or the U.S. Department of Transportation. This report does not constitute a standard, specification, or regulation.

**Does Generalizing Density Functions
Better Explain Urban Commuting?
Some Evidence from the Los Angeles Region**

Shunfeng Song

Department of Economics
University of Nevada at Reno
Reno, Nevada 89557-0016

*Working Paper
January 1994*

prepared for the 33rd annual meeting of the Western Regional Association, Tucson, Arizona

UCTC No. 197

The University of California Transportation Center
University of California at Berkeley

ABSTRACT

The assumption that urban workers economize on commuting is implicit in urban economic theory. Yet it has been challenged by some recent studies. This paper estimates commute flows implied by three urban density functions: monocentric, polycentric, and dispersive. It finds that an urban density function better predicting the actual spatial patterns also better explains the actual commuting behavior. This finding helps us to preserve the assumption that urban workers make attempts to economize on commuting in their location choices.

This research was supported by the U.S. Department of Transportation and California Department of Transportation to the University of California Transportation Center. I alone am responsible for the analysis that is presented here.

1. Introduction

The assumption that urban workers economize on their commuting is implicit in urban economic theory, and is reflected in the trade-off behavior between land rents and transportation costs in location decisions. Such fundamental assumption, however, has been challenged by Hamilton (1982). Based on the standard urban economic model, Hamilton shows that 87 percent of the actual commute is "wasteful" in typical U.S. metropolitan areas. He concludes "it is not clear that the trade-off between commuting and land rent plays any significant role at all in location decisions." (page 1050) Small and Song (1992) verify his finding, using 1980 small-zone journey-to-work census data for Los Angeles County.

Hamilton's finding, however, could be misleading if the monocentric model inadequately represents the actual spatial structure. In fact, the standard urban economic model has been proved a poor description of reality for large metropolitan areas. Several recent studies have demonstrated the presence of employment subcenters in large American cities (McDonald, 1987; Cervero, 1989; Giuliano and Small, 1991). Studies by Gordon et al (1986), McDonald and Prather (1991), and Small and Song (1993) further show that polycentric models fit statistically better the actual spatial patterns than the monocentric model. Polycentric models as well as more general urban density functions, however, have not been incorporated into the "wasteful commuting" analyses. Hence, we are unable to distinguish whether Hamilton's finding is an indictment of the monocentricity or of the more fundamental assumption on commuting behavior that urban workers economize on commuting in their location choice. An indictment of the monocentricity might not be surprising. But an indictment of the assumption on commuting behavior is drastic. It implies a need to re-construct the theory of urban economics.

Using 1980 small-zone journey-to-work data for the five-county greater Los Angeles region, this paper first investigates the spatial patterns of employment and worker residences with three density functions: monocentric, polycentric, and dispersive. A polycentric density function generalizes the monocentric model by assuming that employment and worker residences are distributed in a pattern consistent with several employment centers, not just one. A density function, which we call "dispersive", further generalizes the polycentric model by assuming that urban residents not only value access to the employment centers but also value access to the overall job opportunities in their location choices. We test that generalized density functions fit better the actual spatial patterns.

This paper then estimates commute flows implied by these three urban models. It therefore, for the first time, determines the effects of polycentric and more general spatial structures on the urban commuting. Since the assumption on the commuting behavior is implicit in all three models, a model better explaining the actual spatial structure is expected to better explain the actual commuting behavior. In this way, we can test the validity of the assumption that urban workers economize on commuting in their location choices.

2. Three Density Functions

The economic model underlying urban density functions is a static equilibrium model, in which firms and households value access to urban center(s) and maximize their profits and utilities by trading off transportation costs and land rents. In the case of residential location, urban households choose locations to maximize their welfare according to commuting costs, space consumption, and their income. In the case of firm location, firms trade off agglomeration

economies against transportation costs and land rents. Accessibility to center(s) is reflected in land rents. Hence, the closer to the center(s), the higher the rent is. In turn, higher rents reduce space consumption, leading to higher densities.

This section presents three density functions for three different urban forms: monocentric, polycentric, and dispersive. For a monocentric urban form, we use the negative exponential density and write here as

$$D_i = D_0 e^{-gr_i} e^{u_i}, \quad i = 1, 2, \dots, I, \quad (1)$$

where D_i is the worker residence or employment density at distance r_i to the single urban center; e^{u_i} is a multiplicative error term associated with zone i ;¹ D_0 and g are parameters to be estimated from the data by ordinary least squares after taking the natural logarithm of equation (1). Theoretically, D_0 is the density extrapolated to the urban center, and g is the density gradient measuring the percentage fall off in density for a unit increase in distance from the central business district (CBD).

For a polycentric urban form, we use an additive extension of the monocentric density function, as in Gordon et al (1986) and Small and Song (1993). It is written here as

$$D_i = \sum_{n=1}^N a_n e^{-b_n r_{ni}} + v_i, \quad i = 1, 2, \dots, I, \quad (2)$$

¹The literature on the negative exponential density function have used two specifications of the error term in estimation. One assumes the error term is multiplicative and estimates the two coefficients of the model by ordinary least squares after taking logarithm of the model; whereas the other assumes an additive error term and estimates the density function by nonlinear least squares. Greene and Barnbrock (1978) show that a multiplicative error term is more appropriate with respect to the criterion of homoscedasticity of the error term in regression models.

where N is the number of employment centers in an urban area; r_{ni} is the distance from center n to zone i ; v_i is the error term associated with zone i ; a_n and b_n are parameters to be estimated for each employment center n by nonlinear least squares. This specification of polycentric model assumes that the density at any location i is the vertical summation of the N negative exponential density functions, each reflecting the influence of a center on that location.²

For a dispersive urban form, we further generalize the polycentric model by assuming that workers value access to the overall job opportunities in their location choices, not just to employment centers. Several recent studies have shown that there is a small share of employment located in the CBD. Gordon et al. (1989) show that the Census-defined CBD has only 3 percent of jobs in Los Angeles County and an average of 7.4 percent in the ten Largest U.S. cities; Giuliano and Small (1991) show that only 32 percent of employment are located in 32 centers in the Los Angeles region in 1980. These facts may indicate that the importance of employment centers on the residential location choice is limited.

In this paper, we formulate a more general worker residence density function as following,

$$D_i = e^{a_1} A_i^{a_2} e^{\epsilon_i}, \quad i=1, \dots, I, \quad (3)$$

where A_i is the accessibility of zone i to employment opportunities in all zones in an urban

²Small and Song (1993) argue that the sum of center-specific functions is superior to the upper envelope or the product of such functions. All these three specifications are discussed in Heikkila et al. (1989).

area; ϵ_i is the error term; α_1 and α_2 are parameters to be estimated by ordinary least squares, after taking the natural logarithm of equation (3).

The variable A_i measures the accessibility to the overall employment opportunities at location (zone) i . It is defined as a negative exponential function of commuting distance to other locations weighted by the size of employment in destinations,

$$A_i = \frac{\sum_j E_j e^{-\alpha r_{ij}}}{E}, \quad (4)$$

where α is a parameter to be estimated, measuring the resistance of space separation; r_{ij} is the commuting distance from i to j ; and $E = \sum_j E_j$ is the total employment of the urban area. This specification is recommended by Ingram (1971) as the most suitable form for determining the accessibility at a given location, and is used by Dalvi and Martin (1976) and Williams and Senior (1978).

Equations 1-3 define three density functions for three different urban forms: monocentric, polycentric, and dispersive.³ If the spatial structure in an urban follows a polycentric form, the polycentric model will fit the actual distributions statistically better than the monocentric model. If the spatial structure in an urban follows a dispersive form, the dispersive model will be the best. Hence, we can use these three density functions to test which urban form that a metropolitan area follows.

³We call Equation 3 a density function because A_i depends heavily on distance r_{ij} . Song (1993) shows the conditions that Equation 3 collapses to the monocentric and polycentric models.

3. Estimates of Required and Excess Commute

This paper uses a linear assignment model to calculate the average commute implied by a density function, as in White (1988) and Small and Song (1992). Knowing the distributions of jobs and worker residences and the commuting distances among locations, a linear assignment model is to swap jobs or workers so that total commuting distance is minimized.

Let n_{ij} be the number of commuters from zone i to zone j , r_{ij} be the corresponding network commuting distance. A linear program is specified to find n_{ij} to

$$\min \bar{c} = \frac{1}{W} \sum_i \sum_j r_{ij} n_{ij}, \quad (5)$$

subject to the constraints

$$\sum_j n_{ij} = W_i, \quad \sum_i n_{ij} = E_j, \quad n_{ij} \geq 0, \quad (i, j = 1, 2, \dots, I) \quad (6)$$

where $W = \sum_i W_i = \sum_j E_j$ is the number of commuters in the urban area, while W_i is the number of worker residences in zone i and E_j is the number of jobs in zone j .

In principle, a linear assignment model can be applied to any set of densities, predicted or actual, to estimate the minimum required commute. In this way, we can calculate the commute implied by different density functions. The difference between the actual commute and the minimum required commute is the excess commute ('wasteful commuting' in Hamilton's terminology, we prefer 'excess commute', a more normatively neutral term), which is expected to be smaller if a density function better explain actual distributions and the assumption that urban workers economize on their commuting is valid.

4. Study Area and Data

The study area consists of five counties in the greater Los Angeles region, covering the urban parts of Los Angeles, Orange, Riverside, San Bernardino, and Ventura Counties. The geographical unit is the traffic analysis zones (AZ), defined by the Southern California Association of Governments (SCAG). Like census tracts, traffic analysis zones are aggregates of census blocks and have their boundaries determined by functional traffic characteristics but need not have a fixed population, and hence reduce the "census-tract delineation bias" observed by Frankena (1978) in density estimation. The study area consists of 1124 AZs, after deleting 161 very low-density zones for simplicity.⁴ The 1124 zones cover 3,401 square miles.

We use the journey-to-work data from the 1980 Census of the Los Angeles region, provided by the SCAG. The data include aggregate zone-to-zone commute flows. Information on zone-to-zone travel distances is extracted from the data created for the Urban Transportation Planning Package (UTPP), which is calibrated based on a peak-period representation of the road network. The journey-to-work census data not only allow us to examine the spatial patterns of worker residences and employment, but also allow us to examine the urban commuting patterns. This research analyzes the 4.53 million workers who both live and work in the 1124-AZ study area. Since the location theory really only considers resident workers and only employed individuals commute to work, this research analyzes resident workers rather than population.

⁴All the deleted zones are remote from the highly developed parts of the region, with the exception of 11 zones which have both zero worker residence and employment and 11 largely undeveloped zones in the Santa Monica mountain which separates the densely developed West Los Angeles corridor (roughly, Hollywood to Santa Monica) from the more suburban San Fernando Valley.

5. Empirical Results

This section first estimates the three density functions presented in section 2. It then calculates the minimum commutes implied by these density functions. Specifically, we examine the following questions: Does generalizing density function better explain the actual urban spatial patterns? Does generalizing density function better explain the actual urban commuting patterns?

Does generalizing density function better explain the actual urban spatial patterns?

The monocentric density function is estimated by ordinary least squares, after taking the natural logarithm of equation (1). Downtown Los Angeles is the monocentric center.⁵ In estimation, seven zones with zero density are deleted because $\log(0)$ is undefined. The polycentric density function is estimated by non-linear least squares, with respect to five employment centers: Downtown Los Angeles, UCLA/Santa Monica, LA Airport, West Hollywood, and Santa Ana.⁶ The dispersive density function is estimated by ordinary least squares, after taking the natural logarithm of equation (3), with parameter α obtained by grid search.⁷ As in the monocentric density function estimation, seven zero-density zones are deleted

⁵Small and Song (1993) show that Downtown Los Angeles is the statistical monocentric center to the region.

⁶Giuliano and Small (1991) present the procedures to identify employment centers. In this paper, we use criteria $\bar{D}=17$ and $\bar{E}=37,000$. Five employment centers are identified. When we lower the criteria to $\bar{D}=15$, and $\bar{E}=35,000$, one additional center, Pasadena, is identified. This center, however, is insignificant in explaining the distributions of worker residences and employment, and thus we exclude it.

⁷Equation (2) is a non-linear regression model with respect to parameters $(\alpha_0, \alpha_1, \alpha)$, because α is an unknown parameter in the definition of A . Direct estimation on this model is extremely difficult, because the independent variable A is the sum over 1124 zonal terms and

here.

Table 1 presents the results on the tests between these three density functions in their explaining of actual spatial structure. Performing F -tests between the monocentric and the polycentric models, we have F -statistic values of 34.30 for the worker residence distribution and 30.19 for the employment distribution. These results, with (8, 1114) degrees of freedom, indicate that the null hypothesis (H_0 : Model is the monocentric model) is soundly rejected at a significance level of 0.0001. Therefore, we conclude that the polycentric model statistically explains the actual distributions better than the monocentric model.

Between the monocentric and dispersive models, we use the likelihood-ratio test for non-nested hypotheses developed by Vuong (1989). Vuong suggests a simple test for model selection between a pair of competing non-nested models F_θ (the dispersive model) and G_γ (the monocentric model). Test procedures are shown in the Appendix. Our results show that the Vuong's value is 6.05, suggesting that the null hypothesis is soundly rejected at a significance level of 0.0001 in favor of the dispersive model. Therefore, we conclude that the dispersive model is statistically superior to the monocentric model in explaining the actual worker residence distribution.

Between the polycentric and dispersive models, we cannot test them statistically due to the different dependent variables. However, we observed that the dispersive model not only has

each of those terms contains parameter α in its exponent. A grid search is used here, considering α is the single parameter that causes the non-linearity in the regression model (Greene, chpt. 11, 1990). It minimizes sum of squared residuals (SSR) for all of the parameters by scanning over values of α for the one that gives the lowest SSR . Hence, we determine the optimal value of α , with the associated least squares estimates of parameters (α_0 , α_1) and their standard errors.

a better accuracy in predicting the total worker residences but also has many fewer unknown parameters than the polycentric model. The dispersive model underpredicts the actual worker residences by 8.69 percent, whereas the polycentric model overpredicts the actual total by 35.09 percent. Hence, the former is superior to the latter based on the criterion of prediction accuracy which is one of the two criteria used by McDonald and Bowman (1976). In addition, we observe that the dispersive model has many fewer unknown parameters than the polycentric model. The former has only three (α , α_1 , α_2); the latter has ten for a five-center model. Therefore, it is much easier to have reliable estimates for the dispersive density function than for the polycentric density function.

In summary, we found that the polycentric density function is superior to the monocentric density function in explaining the actual distributions of worker residences and employment. We also found that the dispersive density function, for worker residences, explains better the actual distribution than both the monocentric and polycentric density functions. Therefore, we conclude that generalizing density function better explains the actual urban spatial patterns.

Does generalizing density function better explain the actual urban commuting patterns?

The linear assignment model is applied to all three sets of densities, predicted by the monocentric, polycentric, and dispersive models. Table 2 reports the estimates on the average commutes implied by these models. For the monocentric density function, its average required commute is 1.99 miles. Comparing the actual average commute of 10.81 miles, the monocentric model explains 18.41 percent of the actual commute and has an excess commute of 81.59 percent. These results indicate that the monocentric model is poor at explaining the actual

commuting behavior, confirming the results in Hamilton (1982) and Small and Song (1992). They further indicate that the monocentric model is unsuitable in analyzing urban spatial patterns, since this model is based on the assumption that urban workers economize on their commuting.

The minimum average commute implied by the polycentric density function is 4.42 miles, and the corresponding excess commute is 59.11 percent. Comparing the monocentric estimates, the polycentric model has considerably higher required commute and lower excess commute. These results suggest that the polycentric model explains the actual commuting patterns much better than the monocentric model.

The required and excess commutes for the dispersive density function are also presented in Table 2. The results show that the dispersive model has an average required commute of 5.07 miles and an excess commute of 53.10 percent. Clearly, the dispersive density function has a highest estimate of required commute and a lowest amount of excess commute. Therefore, we conclude that the dispersive model best explains the actual commuting patterns.

The results in Table 1 show that generalizing density function better explains the urban spatial patterns. The results in Table 2 show that generalizing density function better explains the urban commuting patterns. These findings together suggest that an urban model better predicting the spatial patterns also better explains the commuting patterns. Put differently, the ability of an urban model to explain the actual commuting patterns is positively related to its ability to explain the actual spatial patterns. This conclusion is consistent with the assumption that urban workers economize on their commuting, because all these three urban models are based on the same behavior that assumes urban workers value accessibility to employment and

trade off land rents and commuting costs in their location choices.

6. Conclusions

Using the 1980 small zone journey-to-work census data for the Los Angeles region, this paper has examined the spatial patterns with three urban density functions: monocentric, polycentric, and dispersive. We found that the polycentric model explains the actual distributions of worker residences and employment better than the monocentric model. For the worker residence distribution only, we found that the dispersive model best explains the actual spatial pattern. Hence, we conclude that generalizing urban density function better explains the actual spatial patterns.

This paper also calculated the minimum urban commutes implied by these three density functions. We found that the polycentric model has a higher required commute and a lower excess commute than the monocentric model, suggesting that it better explains the actual commuting behavior. The dispersive best explains the actual urban commuting patterns, because it has a highest required commute and a lowest excess commute. Therefore, we conclude that generalizing density function improves the ability to explain the actual urban commuting patterns.

These results together suggest that an urban model better explaining the actual spatial pattern also better explains the actual commuting patterns. This finding helps us to preserve the assumption that urban workers economize on their commuting, because this assumption is implicit in all these three urban models. In turn, it indicates that the trade-off between commuting costs and land rents plays a role in location decisions.

Table 1. Tests Between Density Functions

Monocentric vs Polycentric Models

H_0 : Model is monocentric.

H_A : Model is polycentric.

Test statistic: F -statistic, with degrees of freedom (8,1114)

Results: F -value is 30.19 for employment and 34.30 for worker residence.

Conclusion: H_0 is rejected at a significance level of 0.0001.

Monocentric vs Dispersive Models^a

H_0 : Dispersive model is equivalent to monocentric model.

H_A : Dispersive model is superior to the monocentric model.

Test statistic: Standard normal random variable (z).

Results: $z=6.05$ for worker residence distribution.

Conclusion: H_0 is rejected at a significance level of 0.0001.

Polycentric vs Dispersive Models^a

No formal test performed because the two models have different dependent variables.

The polycentric model overpredicts the actual total worker residences by 35.09 percent.

The dispersive model underpredicts the actual total worker residences by 8.69 percent.

The polycentric model has ten unknown parameters for a five-center model.

The dispersive model has three unknown parameters, easier to have reliable estimates.

Note:

a. Employment distribution is given in the dispersive density function.

Table 2. Estimates on Average Required and Excess Commute

Density Function	Ave. Required commute (miles)	Ave. Excess Commute (%)
Monocentric	1.99	81.59
Polycentric	4.42	59.11
Dispersive	5.07	53.10

Note:

The average actual commuting distance is 10.81 miles.

References

Cervero, R. (1989), *America's Suburban Centers: The Land Use - Transportation Link*, Boston: Unwin Hyman.

Dalvi, M.Q., and K.M. Martin (1976), "The Measurement of Accessibility: Some Preliminary Results," *Transportation*, 5, pp. 17-42.

Frankena, Mark W. (1978), "A Bias in Estimating Urban Population Density Functions," *Journal of Urban Economics*, 5, pp. 35-45.

Giuliano, Genevieve, and Kenneth A. Small (1991), "Subcenters in the Los Angeles Region," *Journal of Regional Science and Urban Economics*, 21, pp. 163-182.

Gordon, Pete, Ajay Kumar, and Harry H. Richardson (1989), "Congestion, Changing Metropolitan Structure, and City Size in the United States," *International Regional Science Review*, Vol. 12, No. 1, pp. 45-56.

Gordon, Pete, Harry W. Richardson, and H.L. Wong (1986), "The Distribution of Population and Employment in a Polycentric City: The Case of Los Angeles," *Environment and Planning A*, 18, pp. 161-173.

Greene, David L., and John Barnbrock (1978), "A Note on Problem in Estimating Exponential Urban Density Models," *Journal of Urban Economics*, 5, pp. 285-290.

Greene, William H. (1990), *Econometric Analysis*, New York: MacMillan Publishing Company.

Hamilton, Bruce W. (1982), "Wasteful Commuting," *Journal of Political Economy*, 90, pp. 1035-1053.

Heikkila, E., P. Gordon, J.I. Kim, R.B. Peiser, H.W. Richardson, and D. Dale-Johnson (1989), "What happened to the CBD-Distance Gradient?: Land Values in a Policentric City," *Environment and Planning A*, 21, pp. 221-232.

Ingram, D.R. (1971), "The Concept of Accessibility: A Search for an Operational Form," *Regional Studies*, Vol. 5, pp. 101-107.

McDonald, John F. (1987), "The Identification of Urban Employment Subcenters," *Journal of Urban Economics*, 21, pp. 242-258.

McDonald, John F. and H.W. Bowman (1976), "Some Tests of Alternative Urban Population Density Functions," *Journal of Urban Economics*, 3, pp. 242-252.

McDonald, John F. and Paul Prather (1991), "A Polycentric Employment Density Model for the Chicago Urbanized Area," presented at the 38th North American Meetings of the RSAI, New Orleans, LA (Nov. 7-10, 1991).

Small, Kenneth A., and Shunfeng Song (1992), "'Wasteful' Commuting: A Resolution," *Journal of Political Economy*, 100 (4), pp. 888-898.

Small, Kenneth A., and Shunfeng Song (1993), "Population and Employment Densities: Structure and Change," *Journal of Urban Economics*, forthcoming.

Song, Shunfeng (1993), "Modelling Worker Residence Distribution in the Los Angeles Region," Paper presented at the 40th North American RSAI Meetings, Houston, Texas, Nov. 11-14, 1993.

Suh, Seoung Hwan (1990), "Wasteful Commuting: An Alternative Approach," *Journal of Urban Economics*, 28, pp. 277-286.

Vuong, Quang H. (1989), "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57(2), pp. 307-333.

White, Michelle J. (1988), "Urban Commuting Journeys Are Not 'Wasteful'," *Journal of Political Economy*, 96(5), pp. 1097-1110.

William, Huw C.W.L., and Martyn L. Senior (1978), "Accessibility, Spatial Interaction and the Spatial Benefit Analysis of Land Use - Transportation Plans," in *Spatial Interaction Theory and Planning Models*, ed. by Anders Karlqvist et al., New York: North-Holland and Publishing Company.

Appendix

Vuong (1989) develops a likelihood-ratio test for non-nested hypotheses. He suggests a simple test for model selection between a pair of competing non-nested models F_θ and G_γ . Let $LR_n(\hat{\theta}_n, \hat{\gamma}_n)$ be the likelihood ratio statistic for the model F_θ against the model G_γ . That is:

$$LR_n(\hat{\theta}_n, \hat{\gamma}_n) = \sum_{i=1}^n \log \frac{f(Y_i | Z_i; \hat{\theta}_n)}{g(Y_i | Z_i; \hat{\gamma}_n)},$$

where n is the sample size, and $\hat{\theta}_n$ and $\hat{\gamma}_n$ are the ML estimates; f_i and g_i are the values taken by the corresponding probability densities for observation i , evaluated in each case at the corresponding maximum-likelihood parameter estimate. Under the null hypothesis (H_0 : F_θ and G_γ are equivalent), Vuong's value is asymptotically distributed according to a central normal distribution.

To test the null hypothesis, one chooses a critical value c from the standard normal distribution for some significance level. If the value of the statistic $n^{1/2}LR_n(\hat{\theta}_n, \hat{\gamma}_n)/\omega_n$ is higher than c , then one rejects the null hypothesis in favor of F_θ being better than G_γ , where ω_n is the square root of the variance of $\log[f(Y_i | Z_i; \hat{\theta}_n)/g(Y_i | Z_i; \hat{\gamma}_n)]$. If $n^{1/2}LR_n(\hat{\theta}_n, \hat{\gamma}_n)/\omega_n$ is smaller than $-c$, then one rejects the null hypothesis in favor of G_γ being better than F_θ . If $|n^{1/2}LR_n(\hat{\theta}_n, \hat{\gamma}_n)/\omega_n| \leq c$, then one cannot reject the null hypothesis. ω is defined as

$$\omega^2 = \frac{1}{T} \sum_{i=1}^T [\log(f_i/g_i)]^2 - \left[\frac{1}{T} \sum_{i=1}^T \log(f_i/g_i) \right]^2.$$