SIMULATION OF THE HIRFINDAHL-HIRSCHMAN INDEX: THE CASE OF THE ST. LOUIS BANKING GEOGRAPHIC MARKET

ABSTRACT

This paper proposes three types of functional specifications of firm size distribution in an attempt to simulate the Hirfindahl-Hirschman Index or HHI. The three types of specifications are arithmetic progression, linear function, and a group of non-linear specifications in which the firm size is a function of the firm rank. Empirical simulations are conducted for the St. Louis banking market using non-linear specifications and bank data from 1994 and 2002. Simulation results show that relatively higher accuracy is achieved for inverse, power, log and quadratic specifications.

INTRODUCTION

The Hirfindahl-Hirschman Index (HHI) is a standard index used in analyzing the degree of market concentration of a particular industry in a particular geographic market. The HHI belongs to a family of indices that also includes the Rosenbluth Index and the Entropy Index (Jacquemin, 1987). All indices measuring concentration utilize percentage shares of individual firms in a geographic market. The difference resides in how to weigh such percentages (Shepard, 1979). From this point of view, all indices of distribution and concentration belong to a broad family. For example, the standard concentration ratio, which calculates cumulative shares of the largest *m* firms in an industry with a total *n* firms, can be seen as a weighted sum of firm market shares, with weight being 1 for the largest *m* firms and 0 for the rest of *n-m* firms. The standard Lorenz curve involves a comparison between the cumulative market share and cumulative shares of the number of firms (Devine, *et al*, 1974). Different indices vary in terms of their emphasis of different aspects of market structure. For example, while the HHI gives weight to the influence of large firms, the entropy index tends to emphasize small firms in the shaping of the overall index. In addition, the Rosenbluth Index includes not only the firm market share, but also the firm rank. These three indices are similar in that, unlike the Lorenz curve in which the number of firms does not affect the evaluation of evenness of distribution, the number of firms in an industry matters. Studies also found strong correlations among different measures of concentration indices (Nelson, 1963). This has made the selection of different indices dependent on the ease in collecting data and of calculation. Advances in computation technologies further reduces the difficulties in handing large data sets, and leaves the matter to the researchers' personal preference.

The HHI was first used by Hirschman in a trade study in the 1940s where the square root of the sum of squared market shares were calculated (Hirschman, 1980). In 1950, Herfindahl used the version of the HHI as it is used today in his Ph.D. dissertation, and in 1959 in a study of the international copper industry (Herfindahl, 1959). The index came to be known as the Herfindahl Index after studies by Rosenbluth (1955, 1957). In 1964, Hircshman published a short article in the *American Economic Review* claiming original ownership of the index (Hirschman, 1964). The index has been known by its current name since that time. Since 1982, the Antitrust Division of the U.S. Department of Justice has been using the HHI as a measure of market concentration in antitrust issues, along with the four firm concentration ratio. The HHI has also become an ingredient in constructing some more complex indices for particular types of markets such as an oligopolistic market (Cabral, 2000).

Although widely used as a measure of industrial structure, so far the HHI for a given year is only calculated by using that same year's data. It has not been used to predict and simulate the market concentration likely to occur in the future. This is largely due to the fact that the market shares used in calculating an HHI are unpredictable. Market shares are determined by the size distribution of an industry in a geographic market. For industries that undergo significant restructuring, it may be particularly difficult to predict the size distribution of the industry. However, when ample data are available, it may allow a simulation of the size distribution using data for geographic markets that have numerous firms. The unpredictability due to a few reorganizations may be overwhelmed or evened out by a large number of units that maintain existing patterns of behavior. Alternatively, well-specified functional forms may capture the essential structural relationship that goes beyond a single year. This paper takes a step in this direction. It specifies the mathematical forms of firm size distribution in a geographic market, and also conducts empirical simulations using selective specifications proposed. Three types of specifications are discussed. The first involves specifying firm size distribution as an arithmetic progression. The second specification assumes that there exists a linear relationship between firms of different sizes. Finally, a group of specifications are discussed. The commonality within the group is that all specifications are functions of firm size ranks. The empirical simulations are conducted by using banking data for the St. Louis metropolitan area.

THE HHI MODEL SPECIFICATIONS

HHIs with even and semi-even distributions

The following gives the definition of the HHI,

$$HHI_{1...n} = \sum_{n} S_n^2 \qquad (1)$$

where S_n is the market share of firm n where n=1, 2, … and n. The maximum value of the index is 10,000 when all industry is concentrated in one firm (10,000=100 ×100). The theoretical values of minimum HHI depend on the distribution of the market shares. Assume a completely even distribution where all firms in a market are the same size and have the same share of the market. With a market having two firms, $HHI_{1..2} = 100^2(\frac{1}{2^2} + \frac{1}{2^2})$. For a market with three firms, $HHI_{1..3} = 100^2(\frac{1}{3^2} + \frac{1}{3^2} + \frac{1}{3^2})$. In general with a market having n firms,

$$HHI_{1...n} = 100^2 \times n \times \frac{1}{n^2} = \frac{100^2}{n} \qquad (2)$$

Equation (2) gives the general formula for a market with an even distribution. Note that if n=1, then HHI=10,000, the maximum value possible. Now suppose that out of n firms, there are m dominant firms. The m firm concentration ratio is therefore $S_{tm}=S_1+S_2+..+S_m$. For the remaining n-m firms, market share is 1- S_{tm} Assume first that within both the m dominant firms and the rest of the n-m firms, there is even distribution. Applying Equation (2) in both groups results in

$$HHI_{1..n} = HHI_{1...m} + HHI_{m+1...n-m} = \frac{S_{tm}^2}{m} + \frac{(100 - S_{tm})^2}{n-m} \qquad (3)$$

In order to avoid a situation where the size of firms in the remaining group is larger than that in the dominant group, the condition, $S_{tm}>m/n$, is set. Notice that in Equation (3), if $S_{tm}=0$, Equation (3) changes into Equation (2). While Equation (2) describes even distributions of different numbers of firms, Equation (3) describes semi-even distributions in that between the dominant and the remaining groups, there are different firm sizes. However, within each group, there is a uniform firm size. (One exception occurs where $S_{tm}=m/n$. When this happens, the firm sizes of both groups are equal.) The reason for dividing the n firms into two separate groups is that within an industry or a market, there may be a dominant group of firms and the remaining marginal group of small firms. The two groups may demonstrate different size distributions and thus require different specifications regarding their size distribution.

Firm market shares as an arithmetic progression

Even and semi-even distributions can be used for approximations either for the dominant or the remaining group, depending on industries and/or markets. In general, we would expect firms to have different sizes. Three different size distributions are specified here. The first specification assumes an arithmetic progression for firms within an industry; the second specification adopts a linear function that describes size relationships among firms; and the last is a group of specifications where size is a function of the firm rank.

For an industry where firm market shares constitute an arithmetic progression, firm sizes are S_1, S_2, S_3, …S_n. Here $S_2=S_1+\Delta$, $S_3=S_2+\Delta$, …$S_n=S_{n-1}+\Delta$, where Δ is a common difference. Substituting $S_2=S_1+\Delta$ in $S_3=S_2+\Delta$ leads to $S_3=S_1+2\Delta$. Similar substitution in other terms leads to $S_4=S_1+3\Delta$, … $S_n=S_1+(n-1)\Delta$. Since HHI is the sum of the square of these market shares, then

$$\begin{aligned} HHI_{1...n} &= S_1^2 + S_2^2 + S_3^2 + ... + S_n^2 \\ &= S_1^2 + (S_1+\Delta)^2 + (S_1+2\Delta)^2 + (S_1+3\Delta)^2 + ... + (S_1+(n-1)\Delta)^2 \end{aligned} \qquad (4)$$

Expanding the squared terms and collecting relevant terms in Equation (4), gives

$$HHI_{1...n} = nS_1^2 + \Delta S_1(2+4+6+...+2(n-1)) + \Delta^2(1+4+9+...+(n-1)^2)$$

In the above equation, the second term on the right-hand side contains a sum of an arithmetic sequence and the last term contains the sum of squares of n-1's. Appling the relevant summation formula results in

$$HHI_{1\ldots n} = nS_1^2 + \Delta S_1 n(n-1) + \frac{n(n-1)(2n-1)}{6}\Delta^2 \qquad (5)$$

Equation (5) gives the HHI for an industry with n firms forming an arithmetic progression in their market shares. It is a form of uneven distribution not only in that firms have different sizes, but also that the Lorenz curve for such an industry will show an uneven distribution. The market share for the first firm is $S_1/(nS_1+(n-1)\Delta)$. An even distribution in a Lorenz sense means the market share is equal to that of the number of firms. That is $S_1/(nS_1+(n-1)\Delta)=1/n$ or $nS_1=(nS_1+(n-1)\Delta)$. However, it is easy to see that as long as $\Delta\neq 0$, there should be $nS_1\neq(nS_1+(n-1)\Delta)$. Therefore the assumption of an even distribution is false.

Whether or not there is any industry with such a firm size distribution is an empirical matter. The advantage of this distribution is that it is easy to calculate and can be used in approximation of a certain group in the entire distribution of a market.

Firm market shares forming a linear function

This specifies that the sizes of firms in an industry or market form a sequence in which the size of firm m can be written as a linear function of the size of firm m-1. Specifically, for S_1, S_2, S_3, …S_n, $S_2=C+\alpha S_1$, $S_3=C+\alpha S_2$, … and $S_n=C+\alpha S_{n-1}$, where C is a constant, and α a slope coefficient. Again, substitute $S_2=C+\alpha S_1$ into $S_3=C+\alpha S_2$, there is $S_3=(1+\alpha)C+\alpha^2S_1$. Similar substitution in other terms results in $S_4=(1+\alpha+\alpha^2)C+\alpha^3S1$… $S_n=(1+\alpha+\alpha^2+\ldots\alpha^{n-2})C+\alpha^{n-1}S_1$. As the sum of these market shares, the HHI is

$$\begin{aligned}
HHI &= S_1^2 + S_2^2 + \ldots S_n^2 \\
&= S_1^2 + (C+\alpha S_1)^2 + [(1+\alpha)C + \alpha^2 S_1]^2 + [(1+\alpha+\alpha^2)C + \alpha^3 S_1]^2 + \ldots \\
&+ [(1+\alpha+\alpha^2+\alpha^3+\ldots+\alpha^{n-2})C + \alpha^{n-1}S_1]^2
\end{aligned}$$

Expanding and collecting the terms results in

$$HHI_{1\ldots n} = C\Sigma_I + 2CS_1\Sigma_{II} + S_1^2\Sigma_{III} \qquad (6)$$

where

$$\Sigma_I = 1 + (1+\alpha)^2 + (1+\alpha+\alpha^2)^2 + \ldots + (1+\alpha+\alpha^2+\ldots+\alpha^{n-2})^2$$

$$\Sigma_{II} = 1 + (1+\alpha)\alpha^2 + (1+\alpha+\alpha^2)\alpha^3 + \ldots + (1+\alpha+\alpha^2+\ldots+\alpha^{n-2})\alpha^{n-1}$$

$$\Sigma_{III} = 1 + \alpha^2 + \alpha^4 + \alpha^6 + \ldots + \alpha^{2(n-1)}$$

Σ_I, Σ_{II}, and Σ_{III} are functions of α and n and can be reduced to various extents as sums of geometric progressions. Specifically,

$$\Sigma_I = 1 + (\frac{1-\alpha^2}{1-\alpha})^2 + (\frac{1-\alpha^3}{1-\alpha})^2 + ... + (\frac{1-\alpha^{n-1}}{1-\alpha})^2 = \sum_n (\frac{1-\alpha^{n-1}}{1-\alpha})^2$$

$$\Sigma_{II} = 1 + (\frac{1-\alpha^2}{1-\alpha})\alpha^2 + (\frac{1-\alpha^3}{1-\alpha})\alpha^3 + ... + (\frac{1-\alpha^{n-1}}{1-\alpha})\alpha^{n-1} = \sum_n (\frac{1-\alpha^{n-1}}{1-\alpha})\alpha^{n-1}$$

$$\Sigma_{III} = \frac{1-\alpha^{2n-1}}{1-\alpha^2}$$

Equation (6) therefore becomes

$$HHI_{1...n} = C\sum_n (\frac{1-\alpha^{n-1}}{1-\alpha})^2 + 2C\sum_n (\frac{1-\alpha^{n-1}}{1-\alpha})\alpha^{n-1}S_1 + \frac{1-\alpha^{2n-1}}{1-\alpha^2}S_1^2 \qquad (7)$$

The advantage of the arithmetic progression and linear specifications is that both can be written as a function of the size of one firm. This helps analysis, which looks into how an individual firm may influence the HHI of a market. A related advantage is to relate HHI to a group of firms. For example, by definition, the m firm concentration ratio is $S_{tm}=S_1+S_2+..+S_m$. From the condition $S_n=(1+\alpha+\alpha^2+\ldots\alpha^{n-2})C+\alpha^{n-1}S_1$

$$S_{tm}=S_1+(C+\alpha S_1)+[(1+\alpha)C+\alpha^2 S_1]+\ldots+[(1+\alpha+\alpha^2+\ldots+\alpha^{m-2})C+\alpha^{m-1} S_1]$$

That is, $S_{tm}=S_1(1+\alpha+ \alpha^2+\alpha 3+\ldots\alpha^{m-1}) + C[1+(1+\alpha)+(1+\alpha+\alpha^2)+\ldots(1+\alpha+\alpha^2+\ldots+\alpha^{m-2})$, or

$$S_1 = \frac{S_{tm} - C[1+(1+\alpha)+(1+\alpha+\alpha^2)+...+(1+\alpha+\alpha^2+...+\alpha^{m-2})]}{1+\alpha+\alpha^2+...+\alpha^{m-1}}$$

If the above equation is substituted into (7) and the relevant terms written as the sums of geometric progressions, (7) becomes

$$HHI_{1\ldots n} = C\sum_{n}\left(\frac{1-\alpha^{n-1}}{1-\alpha}\right)^2$$
$$+2C\sum_{n}\left(\frac{1-\alpha^{n-1}}{1-\alpha}\right)\alpha^{n-1}\frac{S_{tm}-C\sum_{m}\frac{1-\alpha^{m-1}}{1-\alpha}}{\sum_{m}\frac{1-\alpha^{m}}{1-\alpha}} \qquad (8)$$
$$+\frac{1-\alpha^{2n-1}}{1-\alpha^2}\left[\frac{S_{tm}-C\sum_{m}\frac{1-\alpha^{m-1}}{1-\alpha}}{\sum_{m}\frac{1-\alpha^{m}}{1-\alpha}}\right]^2$$

Equation (8) gives a hyperbolic relationship between the HHI and m firm concentration ratio, indicating a positive correlation with complex non-linear characteristics.

Firm market shares as functions of firm size ranks

The distribution of firm sizes can be specified as a function of the size ranks of firms in an industry or market. The parameters in the functional specifications can be statistically estimated using existing data. In essence, this method states that there is an inherent relationship between the size of market share and the rank of the firms, and the later pattern follows the early pattern, reflected in the constants of parameters estimated.

One such empirical pattern is a variant of the rank-size rule, which is used in urban geography. The rank-size rule is a simple empirical observation regarding the size distribution of places in an urban system. It states that given the size of the largest city of an urban system, the size of a particular place in the hierarchy can be predicted by dividing the rank of the place into the size of the largest city. That is, the sizes of places are given by a sequence $Size_1$, $Size_1/2$, $Size_1/3$, … $Size_1/n$, where $Size_1$ is the size of the largest place. This is the original rank-size rule. In this sequence, dividing each term by the total size of the entire system, there is S_1, $S_1/2$, $S_1/3 \ldots S_1/n$ where S_1 is the share of the largest place. That is, the market share of a unit is the relative size of the largest unit divided by the rank of the unit. Therefore, the HHI of the entire system is

$$HHI_{1\ldots n} = S_1^2\left(1+\frac{1}{2^2}+\frac{1}{3^2}+\ldots.+\frac{1}{n^2}\right) \qquad (9)$$

In general, there is the inverse model $S_n=a+bn^{-1}$ where a and b are constants, and can be statistically estimated. The related HHI is

$$HHI_{1\ldots n} = na^2 + 2ab\sum_{n}\frac{1}{n} + b^2\sum_{n}\frac{1}{n^2} \qquad (10).$$

As seen from the above, if a=0 and b=S_1, then $S_n=a+bn^{-1}=a+S_1n^{-1}=S_1n^{-1}$. The last relationship is exactly what the rank-size specifies. As a result, (10) becomes (9). In other words, the rank-size rule is a special case of the inverse function.

In a power form, firm market shares can be specified as a power function of firm ranks. That is $S_n=an^b$ where b<0. The resultant HHI is

$$HHI_{1...n} = a^2(1+2^{2b}+3^{2b}+...+n^{2b}) = a^2\sum_n n^{2b} \quad (11)$$

When a=S_1 and b=−1, (11) turns into (9). This shows a connection between these two non-linear specifications. The estimation of values of a and b can be obtained statistically by regression equation LN*S*=LN*a*+*b*LN*n*, where LN indicates the natural log.

In an exponential function, market share is expressed in a function $S_n=ae^{bn}$ and a and b are constant with b<0. The HHI is

$$HHI_{1...n} = a^2e^2(e^b+e^{2b}+e^{3b}+...+e^{nb}) = a^2e^2\sum_n e^{nb} \quad (12)$$

The constants a and b can be statistically estimated using LNS=LNa+bn.

Using the logarithmic function S_n=a+bLNn, the squared market share is $S_n^2=(a+bLNn)^2$ and the HHI can be expressed as

$$\begin{aligned} HHI_{1...n} &= \sum_n (a+nLNn)^2 \\ &= n(a^2+2abLNn!+b^2\sum_n (LNn)^2) \end{aligned} \quad (13)$$

Finally, the size distribution can be specified as quadratic function $S_n=a+bn+cn^2$ where a, b, and c are constants to be statistically estimated. The HHI is

$$HHI_{1...n} = na^2+2(a+b+ab)(1+n)\frac{n}{2}+(b^2+2c)\sum_n n^2+c^2\sum_n n^4 \quad (14).$$

HHI SIMULATIONS USING ST. LOUIS BANKING DATA

This section discusses simulations that adopt a mix of alternative specifications discussed in the previous section. Equation (3) indicates an HHI as a combination of component indices calculated for different segments of a market. As stated earlier, the reason for dividing a market into segments is to capture the behavioral patterns demonstrated by firms of different sizes. For example, dominant firms in a market may be described by an arithmetic sequence represented by Equation (5) while the remaining firms may be better described by an even distribution as in Equation (2), or a linear function series reflected in Equation (7). Alternatively, various non-linear specifications and their combinations can be considered. The choice of actual specifications should be based on a study of existing data related to a particular industry.

This paper simulates the HHI for the banking industry in the St. Louis metropolitan area. The St. Louis metropolitan area contains eight Missouri counties (Crawford (part), Franklin, Jefferson, Lincoln, St. Charles, St. Louis, and Warren, two Missouri cities (Sullivan city and St. Louis city), and five Illinois counties (Clinton, Jersey, Madison, Monroe and St. Clair). This area differs from the St. Louis bank market defined by the Federal Reserve Bank of St. Louis. The St. Louis bank market includes the city of St. Louis, Missouri counties such as Jefferson, Lincoln, St. Louis, St. Charles, Warren, and Franklin (excluding the area around the town of Berger), and Illinois counties such as Madison, St. Clair and Monroe, plus Sugar Creek and Looking Glass townships in Clinton County. However, as Zhou (1997) discusses, bank deposits in the St. Louis bank market account for over 90% of the deposits in the St. Louis metropolitan area. Since the purpose of the simulation is mainly to evaluate the effectiveness of the simulation techniques proposed, instead of market structure and characteristics of a geographic bank market per se, the study uses the St. Louis metropolitan area as an approximation of the St. Louis bank market.

Bank deposit data for the St. Louis metropolitan area from 1994 to 2001 demonstrates several characteristics. First, the largest banking firm accounts for nearly 20% or even higher of bank deposits, and 40% or above of the HHI in the entire bank market; Second, just under to just over 70% of bank deposits, and 98 to 99% of HHI are concentrated in the top 10% of firms; scatter plots of market share distribution from 1994 to 2002 consistently depict non-linear curves. These characteristics show that the crucial element in simulations is to accurately predict the distribution of the top 10% of the bank market in the St. Louis area, as well as to use non-linear specification.

Given the characteristics of the existing data, this study adopts the following strategy in simulation. First, it is assumed that the size of the largest banking firm is known for the future. This is a reasonable assumption in that the largest banking firm in a large bank market is always closely watched. Information is readily available to make reasonably accurate estimations regarding its influence and size in a particular geographic market. Secondly, the simulation will separate the top 10% of dominant firms from the remaining firms. Essentially, this is to separate the St. Louis banking industry into three groups in the simulation. The largest banking firm is in the first group. The rest of the top 10% of the banking firms are in the second group. The remaining of the banking firms are in the last group. That is

$$HHI_{1..n} = HHI_1 + HHI_{2...m} + HHI_{m+1...n} \qquad (15)$$

where subscripts indicate the ranks of banking firms. The subscript 2…m implies the rest of the firms in the top 10% of the banking firms, and m+1…n implies the remaining 90% of the firms. This means that simulations in this study focus on the firms 2…n.

Thirdly, in choosing specifications, the study adopts various combinations of non-linear specifications for top 2…m firms and the bottom m+1…n firms. The following table lists these combinations.

Table 1 Combinations of Model Specifications Used in Simulations

	Top 2…m firm model specification					
Bottom m+1…n firm model specification		Inverse	Power	Log	Exponential	Quadratic
	Inverse					
	Power					
	Log					
	Exponential					
	Quadratic					

The simulations of the HHI for firms 2…n are contained in the following tables. The headers indicate combinations of specifications used for the top 10% of firms and the bottom 90% of firms. For example, inverse-inverse means that the size distribution of firms 2…m of the top 10% of firms is specified as in an inverse function, and so is the bottom 90% of firms. Alternatively, inverse-exp means that the size distribution of firms 2…m of the top 10% firms is specified as in an inverse function while that of the bottom firms in an exponential function. The number shown in the tables are the sum of the HHIs for both the top and bottom groups. Simulations are conducted for every year from 1994 to 2001.

Table 2. HHI Simulations for n-1 firms

a. Top firms and bottom firms using the same specifications

	inverse-inverse	exp-exp	log-log	power-power	quadratic-quadratic
1994	460.13	414.33	460.02	556.77	462.86
1995	440.41	393.76	440.85	521.39	444.07
1996	491.89	422.39	484.92	573.95	488.22
1997	594.06	429.92	568.17	660.02	578.56
1998	410.89	288.56	393.69	429.67	404.34
1999	382.32	267.35	368.41	399.95	378.78
2000	383.67	274.24	367.23	395.09	374.84
2001	387.93	307.1	378.96	431.76	384.57

b. Top firms using an inverse function

	inverse-exp	inverse-log	inverse-power	inverse-quadratic
1994	458.07	459.45	464.16	459.46
1995	437.63	439.59	443.94	439.81
1996	489.76	491.12	497.35	491.31
1997	592.59	593.66	599.5	593.87
1998	410.25	410.33	421.37	410.46
1999	381.33	381.78	392.83	381.78
2000	382.93	382.93	398.67	382.87
2001	386.81	387.01	405.85	386.99

c. Top firms using an exponential function

	exp-inverse	exp-log	exp-power	exp-quadratic
1994	416.39	415.71	420.42	415.72
1995	396.54	395.72	400.07	395.94
1996	424.52	423.75	429.98	423.94
1997	431.39	430.99	436.83	431.2
1998	289.2	288.64	299.68	288.77
1999	268.34	267.8	278.85	267.8
2000	274.98	274.24	289.98	274.18
2001	308.22	307.3	326.14	307.28

d. Top firms using a logarithmic function

	log-inverse	log-exp	log-power	log-quadratic
1994	460.7	458.64	464.73	460.03
1995	441.67	438.89	445.2	441.07
1996	485.69	483.56	491.15	485.11
1997	568.57	567.1	574.01	568.38
1998	394.25	393.61	404.73	393.82
1999	368.95	367.96	379.46	368.41
2000	367.97	367.23	382.97	367.17
2001	379.88	378.76	397.8	378.94

e. Top firms using a power function

	power-inverse	power-exp	power-log	power-quadratic
1994	552.74	550.68	552.06	552.07
1995	517.86	515.08	517.04	517.26
1996	568.49	566.36	567.72	567.91
1997	654.58	653.11	654.18	654.39
1998	419.19	418.55	418.63	418.76
1999	389.44	388.45	388.9	388.9
2000	380.09	379.35	379.35	379.29
2001	413.84	412.72	412.92	412.9

f. Top firms using a quadratic function

	quadratic-inverse	quadratic-exp	quadratic-log	quadratic-power
1994	463.53	461.47	462.85	467.56
1995	444.67	441.89	443.85	448.2
1996	488.8	486.67	488.03	494.26
1997	578.75	577.28	578.35	584.19
1998	404.77	404.13	404.21	415.25
1999	379.32	378.33	378.78	389.83
2000	375.64	374.9	374.9	390.64
2001	385.51	384.39	384.59	403.43

To obtain a sense of the accuracy in the simulations, HHI simulations given in the above tables and the actual HHI should be compared. Since the purpose of the study is to predict future HHI using existing data, it is assumed that a simulation based on a previous year's data is used to predict the next year's HHI. That is, a simulation based on the 1994 data is used to predict 1995's HHI, simulation based on the 1995 data is used to predict 1996's HHI, … and simulation based on the 2001 data is used to predict 2002's HHI. To find the difference between the simulations and actual HHIs, we subtract actual 1995 HHI from the simulation based on 1994 data, subtract actual 1996 HHI from the simulation based on 1995 data… and subtract actual 2002 HHI from the simulation based on 2001 data. These differences are displayed in the following tables.

Table 3. Differences between simulations and the actual HHIs

a. Top firms and bottom firms using the same specifications

year of data used in simulation	year to be simulated	inverse-inverse	exp-exp	log-log	power-power	quadratic-quadratic
1994	1995	11.69	-34.11	11.58	108.33	14.42
1995	1996	-57.22	-103.87	-56.78	23.76	-53.56
1996	1997	-113.3	-182.8	-120.27	-31.24	-116.97
1997	1998	180.79	16.65	154.9	246.75	165.29
1998	1999	25.1	-97.23	7.9	43.88	18.55
1999	2000	-6.43	-121.4	-20.34	11.2	-9.97
2000	2001	-8.47	-117.9	-24.91	2.95	-17.3
2001	2002	-14.26	-95.09	-23.23	29.57	-17.62

b. Top firms using an inverse function

year of data used in simulation	year to be simulated	inverse-exp	inverse-log	inverse-power	inverse-quadratic
1994	1995	9.63	11.01	15.72	11.02
1995	1996	-60	-58.04	-53.69	-57.82
1996	1997	-115.43	-114.07	-107.84	-113.88
1997	1998	179.32	180.39	186.23	180.6
1998	1999	24.46	24.54	35.58	24.67
1999	2000	-7.42	-6.97	4.08	-6.97
2000	2001	-9.21	-9.21	6.53	-9.27
2001	2002	-15.38	-15.18	3.66	-15.2

c. Top firms using an exponential function

year of data used in simulation	year to be simulated	exp-inverse	exp-log	exp-power	exp-quadratic
1994	1995	-32.05	-32.73	-28.02	-32.72
1995	1996	-101.09	-101.91	-97.56	-101.69
1996	1997	-180.67	-181.44	-175.21	-181.25
1997	1998	18.12	17.72	23.56	17.93
1998	1999	-96.59	-97.15	-86.11	-97.02
1999	2000	-120.41	-120.95	-109.9	-120.95
2000	2001	-117.16	-117.9	-102.16	-117.96
2001	2002	-93.97	-94.89	-76.05	-94.91

d. Top firms using a logarithmic function

year of data used in simulation	year to be simulated	log-inverse	log-exp	log-power	log-quadratic
1994	1995	12.26	10.2	16.29	11.59
1995	1996	-55.96	-58.74	-52.43	-56.56
1996	1997	-119.5	-121.63	-114.04	-120.08
1997	1998	155.3	153.83	160.74	155.11
1998	1999	8.46	7.82	18.94	8.03
1999	2000	-19.8	-20.79	-9.29	-20.34
2000	2001	-24.17	-24.91	-9.17	-24.97
2001	2002	-22.31	-23.43	-4.39	-23.25

e. Top firms using a power function

year of data used in simulation	year to be simulated	power-inverse	power-exp	power-log	power-quadratic
1994	1995	104.3	102.24	103.62	103.63
1995	1996	20.23	17.45	19.41	19.63
1996	1997	-36.7	-38.83	-37.47	-37.28
1997	1998	241.31	239.84	240.91	241.12
1998	1999	33.4	32.76	32.84	32.97
1999	2000	0.69	-0.3	0.15	0.15
2000	2001	-12.05	-12.79	-12.79	-12.85
2001	2002	11.65	10.53	10.73	10.71

f. Top firms using a quadratic function

year of data used in simulation	year to be simulated	quadratic-inverse	quadratic-exp	quadratic-log	quadratic-power
1994	1995	15.09	13.03	14.41	19.12
1995	1996	-52.96	-55.74	-53.78	-49.43
1996	1997	-116.39	-118.52	-117.16	-110.93
1997	1998	165.48	164.01	165.08	170.92
1998	1999	18.98	18.34	18.42	29.46
1999	2000	-9.43	-10.42	-9.97	1.08
2000	2001	-16.5	-17.24	-17.24	-1.5
2001	2002	-16.68	-17.8	-17.6	1.24

Finding the absolute values for numbers in Table 3, and summing up the absolute values along the columns, results in Table 4, which gives aggregate values of inaccuracy in the simulations.

Table 4 Aggregate Sums of Differences Between Simulations and the Actual HHIs

	Top 2…m firm model specification					
Bottom m+1…n firm model specification		Inverse	Power	Log	Exponential	Quadratic
	Inverse	417.26	460.33	417.76	760.06	411.51
	Power	413.33	497.68	385.29	698.57	383.68
	Log	419.41	457.92	419.91	764.69	413.66
	Exponential	420.85	454.74	421.35	769.05	515.10
	Quadratic	419.43	458.34	419.93	764.43	413.68

The simulation results summarized in Tables 3 and 4 show that simulation errors obtained from estimations that used an exponential specification are significantly larger than those of other specifications. These other forms, when used to specify the top firm distribution, along with various combinations with any other specification, tend to give estimations with smaller divergence. Table 5 lists simulation errors as percentages of the actual HHIs for different specifications, averaged over simulations from 1995 to 2002. The average error over a total of 200 simulations is 13.6%. Again, specifications with an exponential form as the top firm size distribution generally have significantly larger error percentages than those with other forms as the top firm size distribution.

Table 5. Average Simulation Errors as Percentages of the Actual HHIs (%)

		Top 2…m firm model specification				
Bottom m+1…n firm model specification		Inverse	Power	Log	Exponential	Quadratic
	Inverse	11.30	13.32	11.29	21.37	11.12
	Power	11.25	14.52	10.33	19.51	10.31
	Log	11.36	13.25	11.34	21.50	11.18
	Exponential	11.39	13.15	11.37	21.61	11.20
	Quadratic	11.37	13.26	11.35	21.49	11.18

Of the total 200 simulations, 80, or 40%, have a simulation error within 5% of the actual HHIs; over half, or 57.5%, have a simulation error within 10% of the actual HHIs. When excluding simulations with exponential forms as the top firm distribution and considering only those 160 simulations with inverse, power, log and quadratic forms as the top firm distribution, 47.5% of simulations have a simulation error within 5% of the actual HHIs, and 65.6% have a simulation error within 10% of the actual HHIs. Table 6 displays simulation errors for alternative specifications. The higher accuracy from using inverse, power, log and quadratic forms is clear, especially compared with the poor performance of the exponential form. Of inverse, power, log and quadratic forms themselves, there is no clear advantage of one over the other. For example, while the power form produces the highest percentage of less than 10% of simulation errors, it has the highest percent of more than 20% simulation errors. On balance, inverse, power, log and quadratic forms generate similar accuracy in simulation, with two thirds to 3 quarters of simulations having less than 10% simulation errors, and 12.5% to a quarter of simulations having more than 20% of simulation errors.

Table 6. Simulations Errors for Alternative Specifications (%)

	Top 2…m firm model specification				
Simulation Error*	Inverse	Power	Log	Exponential	Quadratic
<5%	50.0	47.5	32.5	10.0	60.0
5-10%	12.5	25.0	30.0	15.0	5.0
10-15%	12.5	2.5	12.5	0.0	10.0
15-20%	12.5	0.0	10.0	5.0	12.5
>20%	12.5	25.0	15.0	70.0	12.5

* Measured in simulation discrepancy as a percent of the actual HHI

Some additional observations may be made regarding the simulations. An exponential specification tends to under-estimate the HHIs while a power functional model tends to overestimate the HHIs. Many simulations with large distortions seem to occur in 1996, 1997 and 1998. In quite a few specifications, simulations for these several years diverge from the actual HHIs by over 100 points. A possible reason for such a large difference is that the actual HHI jumps by 70 points from 1996 to 1997, and over 100 points from 1997 to 1998. Such a large change in HHI causes significant changes in size distribution. As a result, the parameter estimates based on the previous year's data cannot adequately capture the changes, leading to simulations widely off the mark. However, such reasoning cannot be applied to simulations in 2000 and 2001. The actual HHI decreases by over 100 points from 1999 and 2000, and again from 2000 to 2001. However, there is no consistent significant error in simulation. The cause for significant distortions remains an issue for further investigation.

CONCLUDING REMARKS

This paper simulates HHIs by specifying functions of firm size distributions. This is an attempt to go a step beyond calculating the HHI using data from the same year. The paper proposes three types of alternative specifications concerning the size distribution of firms in an industry or a geographic market. The advantage of arithmetic progression and linear function is that HHIs resultant can be expressed as a function of the size of one (largest) firm. A group of non-linear specifications where the firm size is a function of firm rank are used in the empirical simulations. Simulation for the St. Louis banking market is conducted using non-linear specifications, with various combinations of specifications for dominant firms and remaining firms. Simulation results show that relatively higher accuracy is achieved for inverse, power, log and quadratic specifications for the St. Louis bank market between 1994 and 2002.

REFERENCES

Cabral, L.M.B. 2000. *Introduction to Industrial Organization*. Cambridge, Massachusetts: The MIT Press

Devine, P.J., Jones, R.M., Lee, N, and Tyson, W.J. 1974. *An Introduction to Industrial Economics*. London: George Allen & Unwin Ltd

Herfindahl, O.C. 1959. *Copper Costs and Prices: 1870 – 1957*. Baltimore: The John Hopkins Press

Hirschman, A.O. 1964. The Paternity of an Index. 1964. American Economic Review Vol. 54, pp 761-762.

------ 1980. *National Power and the Structure of Foreign Trade*. Los Angles: University of California Press

Jacquemin, A. 1987. *The New Industrial Organization: Market forces and strategic behavior*. Cambridge, Massachusetts: the MIT Press

Rosenbluth, G. 1955. Measures of Concentration, In *Business Concentration and Price Policy* by National Bureau of Economic Research. Princeton: Princeton University Press

---------. 1957. *Concentration in Canadian Manufacturing Industries*. Princeton: Princeton University Press

Shepard, W. G. 1979. *The Economics of Industrial Organization*. Englewood Cliff, New Jersey: Prencice-Hall, Inc.

Zhou, B. 1997. Chaning banking market structure amid banking restrcutring: a case of St. Louis. *Papers and Proceedings of the Applied Geography Conferences 20*: 110-118