Statistical Models for Market Approach to Domain Name Valuation

Alex Tajirian
June 20, 2010

1 Introduction

To describe various types of domain-name valuation methodologies, I compare domain names with commercial and residential real estate and art. I then focus on statistical methods to identify comparables used to appraise a domain name. I also outline the intuitive appeal, advantages, and limitations of statistical models, and the erroneous arguments against the models.

2 Analogy with Real Estate and Art

In the commercial real-estate market, values are primarily driven by occupancy rates and lease prices. With an estimate of these two quantities in hand, a given property's market value can be easily quantified. If the ongoing lease rate is $30 per square foot, no rational tenant will pay $50 per square foot for a comparable property. Developed or undeveloped, a property within a commercial zoning designation is valued according to the cash flows it is expected to generate. Different buyers can easily quantify that value even if they don't always agree on a figure—for instance, because of their different projections of the revenue, construction costs, or risk. Hence, an income approach—based on discounted cash flows (DCF)—is not just feasible but a superior approach to valuation.

Residential real estate, a market where values are driven by sales of comparable property, is trickier because there are no cash flows to estimate. Moreover, sellers and buyers can attach different values to various decision variables. Nonetheless, you can still develop quantitative statistical models to capture the subjectivity.

Let's look at an example of subjectivity in residential real estate. You are selling your house. In your mind, one of the selling points of the house is its proximity to a shopping district. On the other hand, the buyer tells you that she is looking for something more "secluded." Thus, this aspect of the house's location has no value to the buyer, irrespective of how important it is to you. However, you can point out that the house is on a two-acre lot in a residential zoned area, meaning there's little chance that a developer will build an industrial park nearby. In this case, "secluded," as a price predictor, can be quantified in terms of acreage and zoning.

The frequently made statement that the "price is what the buyer is willing to pay" works for art and for made-up or personal-use domain names with unquantifiable cash flows

and unquantifiable price predictors. Of course, such a statement is not a model predicting for sale prices; it is descriptive of how the sale price is determined.

If a domain name is more suitable for generating profit, its value is primarily driven by quantifiable cash flows or quantifiable comparables. If the name is more suitable for personal use, or if it's made up, its market price may be driven by the "what the buyer is willing to pay" principle. One should keep in mind that the intrinsic value of an asset is based on the asset's best use. Thus, if a buyer wants to use a commerce-viable domain name for a personal Web site, either the buyer is paying too much or has a personal valuation that is different than the market's.

The above discussion is summarized by the table below:

Table 1

Domain name use	Analogy	Appraisal Method	Sale Process
Investment/Business	Commercial real estate	DCF	Auction or Negotiate
	Residential real estate	Quantifiable subjective predictors	
Nonprofit	Art	Hard-to-quantify predictors	Two Dealers

3 Statistical Techniques: Motivation

The market valuation approach is based on the average price of comparable assets. But, how do you identify value predictors and comparable domain names?

3.1 Issue

To put the problem into perspective, consider the following example in the table below:

Table 2

Domain Name	Key Word	Predictors	
		Number of Google Search Results	PPC ($)
Domain1.com	Domain 1	800,000	0.5
Domain2.com	Domain 2	600,000	1.5
Domain3.com	Domain 3	1,000,000	0.6
Domain1.net	Domain 1	800,000	0.5

Which of the above four domains is more valuable and by how much? It is obvious that Domain3.com is more valuable than Domain1.com because the latter dominates on all prediction variables. How about Domain1.com and Domain1.net? You correctly guessed that ".com" is more valuable, but by how much? As you can imagine, identifying the contribution of tens of factors for thousands of domain names is not a trivial task.

3.2 Identifying Comparables

You should keep in mind the distinction between comparables and substitutes. Substitutes are comparables, but not all comparables are substitutes.

At the heart of our regression-tree statistical model is a set of predictors of demand for domain names and historical prices of sold names. The estimated model yields the best relationship between the specified predictors and value. This relationship is used to identify comparables whose price is averaged out—as in the approach involving real estate comparables—to predict the price of a similar domain name at a specific time. One of the obvious predictors is the domain extension. Another of the quantifiable predictors is the number of search results in Google for the key word embedded in a domain name. A third can be the number of tweets for the key word.

4 Advantages of Statistical Models

Not having the right tool to pound a nail can be frustrating. But it sure beats pounding the nail with your head. Similarly, we don't know the ideal model for generating domain name prices, but we can do a lot with what we have and know. Statistical models provide a systematic analysis of a huge amount of data. The advantages are:

4.1 Eliminate pitfalls in selecting predictors.

- Some of the variables that some appraisers use seem like dubious candidates for quantification. This is not necessarily a problem, as appraisers are only trying to predict value overall, not measure the impact of each variable. But because the models are black boxes, it is not possible to tell whether or not such variables are adding noise to value estimates and thereby making them less precise. Moreover, if appraisers can actually estimate, say, brandability, and thus include it as one of their predictors, why should companies pay branding agencies millions of dollars to come up with a name? Why not toss the domain name appraisers a few bucks and get the same result? Furthermore, if these appraisers can estimate resale value, won't the price of an inactive domain, for example, be easily derived by simply discounting the resale value to the present?

- Domainers are suspicious of some of the information provided in appraisal reports. For example, reports by some of the prominent professional services include a list of comparable sales. If they are able to accurately identify comparable sales, then mission accomplished! All they need to do is calculate the average and/or the median of the prices of the comparables, and voila, they get a very good estimate of value. I conjecture that this provides a better estimate of value and range than their unsubstantiated appraisal methodology.

4.2 Possibility of greater predictive power than experts.[1]

[1] For a condensed version of the study, see Andrew D. Martin, et al., "Competing Approaches to Predicting Supreme Court Decision Making," *Perspectives on Politics*, 2, vol. 4, 2004, pp. 761–767. The study compares legal experts' predictive accuracy with that of statistical regression. The selected experts for the

4.3 Confirm which of the intuitive predictors are relevant to valuation.

4.4 Determine which domain names are comparable.

4.5 Obtain more precise estimates of value and range.

4.6 Bring in transparency of methodology and data sources, including a list of significant predictive variables. The appraisal can be duplicated by anyone, which reduces suspicion.

4.7 Detect data outliers and choose the best techniques to deal with them.

4.8 Estimate the size of the premiums of ".com" over other extensions of comparable domain names.[2]

4.9 Estimate value appreciation.[3]

4.10 Enable post-valuation feedback.[4]

4.11 Advantages over manual appraisals[5]

- Humans like to "control" outcomes. Control can be achieved through human involvement in appraisals. Thus, an appraised value controlled by a human can be erroneously preferred over a potentially superior one generated by an automated system.

- Humans overvalue events that they think they control. Thus, they unjustifiably place more value on a "human appraisal."

- Some customers project from a narrow experience. I constantly hear things like: "I have seen some automated real estate appraisals. They are worthless" and "I

---

study include scholars (five of whom had been law school deans), practitioners, and pundits. The regression model used only six factors: (1) the circuit court of origin, (2) the issue area of the case, (3) the type of petitioner, (4) the ideological direction of the lower court ruling, (5) the type of respondent, and (6) whether the petitioner argued that a law or practice is unconstitutional. The model's data set was composed of 628 cases previously decided by the nine justices. The regression model predicted 75% of the affirm/reverse results correctly, while the legal experts collectively got 59.1% right. The model also predicted Justice O'Connor's vote (a swing vote) correctly 70% of the time, while the experts' success rate was only 61%.

[2] See Alex Tajirian (2006), "Branding Hierarchy and Premiums Among gTLDs," DomainMart.

[3] See Alex Tajirian (2006), "94% Annual Domain Name Price Appreciation," DomainMart.

[4] For the importance of performance feedback, see Alex Tajirian (2008), "LandingBoeing777.com Appraisal," DomainMart.

[5] See, for example, Alex Tajirian (2006), "'Human' vs. Machine Appraisals," DomainMart.

have seen few of these domain name appraisals. They are based on a machine that spits random numbers."

- A human appraisal can potentially be based on an appraiser "randomly picking numbers from a hat." However, it is very likely that's not the mental image that customers form of a human appraiser. They imagine due diligence. Thus, the mental image of the process and what an appraiser actually does can be out of whack. Moreover, no matter how many times an appraiser "picks numbers from a hat," it does not make him a significant appraisal expert.

- Human appraisals can have major disadvantages. For example, in a scientific study, Americans were asked which countries were most similar to each otherr—Ceylon and Nepal or West Germany and East Germany. Most picked the latter. But when asked which countries were most dissimilar, most Americans also picked the same pair. How can a pair of countries be similar and dissimilar? Does this imply that logical analytics is always superior to gut feeling? Not necessarily! However, when it comes to domain names, due to the massive amount of data, analytical models are superior in detecting comparables.

It should be noted that regression-tree models require human tweaking to determine the optimal tree size. Thus, at least in this sense, they are human based.

5 Advantages of Regression Trees

We use regression-tree statistical models[6] to identify comparables and predict value. Their advantages are:

5.1 They group domain names that are judged to be similar because of the names' predictors. Thus, a domain name's predicted value is estimated to be the average price of that group of similar names. This is no different than the "heart of the market" approach.

5.2 Traditional linear-regression techniques don't yield satisfactory results, as the relationship between the extensions and market price is most likely to be nonlinear. Moreover, because the domain name market is relatively new and not very active, more robust statistical techniques are needed.

5.3 Robust to data outliers.

6 Limitations of Statistical Models

6.1 Human input is necessary in the following areas:

[6] For a description of these models as applied to domain names, see Alex Tajirian (2005), "Valuing Domain Names: Methodology," DomainMart.

- Identifying complements to a domain name's key words. When analyzing, for example, BrooklynCollege.com, AdWords' suggestion tool does not include complementary products/services, such as books and student loans.[7]

- Incorporating extraneous public information. Yale University, for example, is pursuing legal means to secure "similar" domain names. Machines alone cannot yet identify potentially relevant extraneous information. Thus, when appraising, for example, BrooklynCollege.com, Yale's action can be an extremely valuable signal regarding possible change in the environment of education-related domain names.

- Unlike determining comparability of real estate based on, say, location, number of bedrooms, square footage, etc., the drivers of domain name similarity are not well known. Moreover, even if a domain has the name, say, of an Indian tribe, that does not mean that the domain name and tribal name are comparables. For example, one of the criteria for Indian casino comparables is that the tribe must be on the federal list of recognized Indian tribes and must have a casino.

- Improving the accuracy of the predictive model by narrowing down the list of possible explanatory variables. A machine has to go over an extremely large number of variables to determine whether they contribute significantly to value prediction. Moreover, the benefits of meaningful and parsimonious predictive variables would be lost with pure machine appraisal.

- The value estimate depends on the length of the period of the sample used. To overcome this estimation issue, you should use the data length that provides the "best" predictive model.

- When appraising real estate, extrapolation can be used to adjust sales data. Instead of finding out sales prices for the exact size of the home being considered, you can compare price per square foot. Extrapolation is a lot trickier with domain names and their explanatory variables; therefore, adjustments related to time synchronicity should be avoided and data integrity should be ascertained. Of course, these restraints limit the viability of using macroeconomic data, which is typically undated quarterly or annually.

- Ground rules that can be confusing and counterintuitive, especially those for nonlinear models such as regression trees and neural networks. For example, the familiar fixed number of relevant explanatory variables for different groups of similar domain names breaks down, leaving no clear answer to the common question, "How many significant variables do you use?" R^2, which is widely recognized as a goodness-of-fit measure, is no longer appropriate.

[7] You may suspect that I am adding variables only to improve the model's fit. True. But from a forecasting standpoint, I only care about greater predictive power, not the goodness of fit per se. Nevertheless, on a theoretical level, including additional explanatory variables may increase goodness of fit and also parameter estimation accuracy.

6.2 The need to regularly collect and update a lot of data.

6.3 Low utility in judging made-up words and high-value names because of the limited availability of price data for such names. When judging made-up names, the best move is simply to ask a large number of random citizens (crowdsourcing, as properly applied). For high-value names, it is more appropriate to use a discounted cash flow model.[8]

6.4 Valuing brand-based domain names. BestBuy.com and BestSell.com are not comparables because the former is a brand name. The income approach is better suited to value brand and trademark-based domain names.

7 Erroneous Arguments Against Statistical Models

Unjustifiable skepticism over the use of statistical models for prediction is not confined to domain name valuation. Other fields include law, wine rating, sports, and medicine, to name a few.[9]

Some of the skeptics' most common complaints: too many variables, only "experts" can do it, not enough data, and confusing statistical results. Let's address each separately.

- Too many variables. This criticism does not refer to situations where there are too many variables to estimate compared to observations. Rather, the argument maintains that there are too many hard-to-quantify variables and that not all variables have an impact on all domain names—i.e., there is a large number of domain name–specific factors. To counter this argument, the Supreme Court case above demonstrates that six predictors, a relatively small number, perform better than the experts.

- Only "experts" who have buying and selling experience can appraise. A Supreme Court study[10] also provides a counterexample using regressions.

- Not enough data. A significant amount of data on sale prices is publicly available, although the publicly available data does becomes thinner for domains sold for greater than $250,000. Nevertheless, there are sources of data, other than sales prices, that can shed even more information on domain name values.[11]

---

[8] "Valuing Domain Names: Methodology," supra.

[9] Ian Ayres, *Super Crunchers: Why Thinking-by-Numbers Is the New Way to Be Smart,* Bantam, 2007.

[10] Martin et al., supra

[11] See Alex Tajirian, "Food for Thought: Appraisal Dataset," DomainMart.

- Statistical results are confusing. The group voicing this particular complaint does believe in the power of statistics, but they hear contradictory messages and are left to wonder who is right and who is wrong. "Coffee is good for you," "Coffee is bad for you," "Domain guru says appraisals are useless," "Domain guru says statistical appraisals are, in general, the most robust." Who should you believe?

When it comes time to pick a side, to go pro-statistical or anti-statistical, you'll find that there is no easy way to make the choice. One solution is what Andy Grove, the former chairman of Intel, did when he had to decide on the method of treatment for his cancer. He dug into the relevant literature to better understand it. Obviously, not everyone has the learning ability, time, or desire to follow such an approach. Although a domain name's appraisal is not a matter of life and death, you want to make sure that you are not taken for a ride. Thus, at a minimum, you need to be aware of some common sale pitches and their weaknesses:

- "We have more stats and data." Two questions to ask them: How do they know they have more stats than their competitors, and how much difference in value precision does their additional data lead to?

- If they believe that the length of the domain name is a strong predictor of value, ask them why their belief contradicts other statistical studies.[12]

- When someone tells you "I believe company X is the best appraiser," ask them why. Would that be based on statistical prediction accuracy or the closeness of the appraisals to the recommender's estimates? ■

Related Essays:

- Injecting Valuation Standards in Domain Name Transactions.html

- For additional references on domain name appraisal, click here.

[12] See, for example, Alex Tajirian, "Length of Domain Name Is Irrelevant!," DomainMart.