

Relationship Science

Andrew Oh

Econ 482

Fall 2007

oha@u.washington.edu

Executive Summary

Members of the opposite sex have baffled each other for as long as anyone can remember. From old schoolyard crushes to shows like VH1's The Pickup Artist – the secrets of attraction have long been pondered. While completely solving the mysteries of attraction and breaking down the art of the pickup is beyond this econometric student's insight and intellect, I can at least try to figure out part of what it is that people are attracted to.

By creating fake personal ads with photographs pulled from www.hotornot.com and posting them to a craigslist region, I was able to collect data with parameters such as physical attractiveness (as determined by at least 100 of hotornot.com's voters – the Central Limit Theorem suggests that with more voters, the more accurate the number), height, age, weight, hobbies, personality, drinking and smoking habits, whether or not there was a child in the single's life, level of education, hair color, and what kind of relationship they were searching for (casual? long-term? friends first?). All of these dummy variables were independently assigned through use of a random number generator, if not already independently supplied.

After attaching each picture to and mentioning the previously listed parameters in a personal ad, I released the batch to Vancouver's craigslist's w4m, m4w, and casual encounters personals sections. After three days, I pulled the ads and checked the number of hits each ad received. I then regressed those hits against the parameters, and conducted a series of f-tests to find out which parameters actually mattered. Upon weeding out the insignificant right hand side variables, I then ran a series of forecasts to check how accurate the model was.

The forecast was fairly accurate, much to my pleasure. This meant I was at least on the right track with the variables I was tracking. The social implications, however, were a bit disillusioning: nobody on the internet cared about your hobbies, personality, whether or not you have kids, whether or not you drink or smoke, or your level of education. The only things that mattered were your physical attractiveness (for both men and women), your weight (for women exclusively) and your gender (women are in MUCH higher demand than men on the internet). To be approached often in online dating, one must be a female, thin, and hot (as opposed to "not"). While this sounds *suspiciously* like real-world dating, that is a study for another project and another student.

Data Collection, Modeling, Regressions, Restrictions, and Forecasting

What is it that attracts two people to each other? Should you pose this question, you are sure to get a laundry list of responses ranging from physical attractiveness, hobbies, personalities, and more.

In this study I hope to explore a set of characteristics in order to find that which makes a person most attractive to a member of the opposite sex. While this study will not say anything as to the potential longevity of a relationship or how well two people will work as a couple, it does say something about how many men or women will be attracted to a person with a given combination of certain characteristics.

Data collection for this project occurred on www.hotornot.com and craigslist. First, I took a sample of photographs with at least 100 votes from www.hotornot.com to get a reasonable estimate of the absolute physical attractiveness of a person. Then I attached that picture to a personal ad template with variations in height, weight, personalities, hobbies, levels of education, whether or not they had kids, and their drinking and smoking habits. Each picture/personal ad combination would be released to the Vancouver, BC <http://www.craigslist.org> w4m, m4w, and casual encounters personals sections. Each response to a given personal ad would be classified as a "hit." The more hits a personal ad received, the more attractive that personal ad was. Regressions were first gender-neutral in an effort to find certain "universal" characteristics that make a person more appealing. After that, they were broken down by gender. Finally, I did a series of forecasts, in an attempt to check the accuracy of my model.

The data gathering process was a long and arduous one. Each personal ad was comprised of a picture, a height, and several other characteristics. They all had to be convincingly written, while maintaining some semblance of uniformity so as to minimize error due to differences in presentation. This meant keeping the ads as simple as possible, while making them realistic enough to avoid those who would have them flagged and deleted, thereby ruining the data collection process. The picture and rating were independently supplied by the hotornot.com website. Using a random number generator, I assigned a height, weight, age, parental status, drinking/smoking habits, hobbies, personalities, level of education, and what kind of relationship a person was seeking to each personal advertisement.

Initially, I ran ordinary least squares with many, many parameters. Naturally, I felt that some of these were useless, so I began cropping them out through a series of f-tests. The fact that these were all independently assigned greatly helped, since that ensured no covariance between any parameters and

allowed for flexibility in choosing which parameters to test – a great convenience when dealing with a sample size as small as this one. I arrived at the following list of significant parameters for men, women, and the general population (all probabilities provided by the Wald f-test):

- General Population
 - Rating ($p=.0265$)
 - Gender ($p=.0409$)
 - Weight ($p=.0322$)
- Men only
 - Rating ($p=.0107$)
- Women only
 - Rating ($p=.0004$)
 - Weight ($p=.0040$)

Using these parameters in my regression, I came up with the following equations:

$$\mathbf{HITS_{GEN} = 1.463770 * RATING + 56.93626 * FEMALE - 0.566644 * WEIGHT}$$

$$\mathbf{HITS_{FEM} = 4.850498 * RATING - 2.351904 * WEIGHT}$$

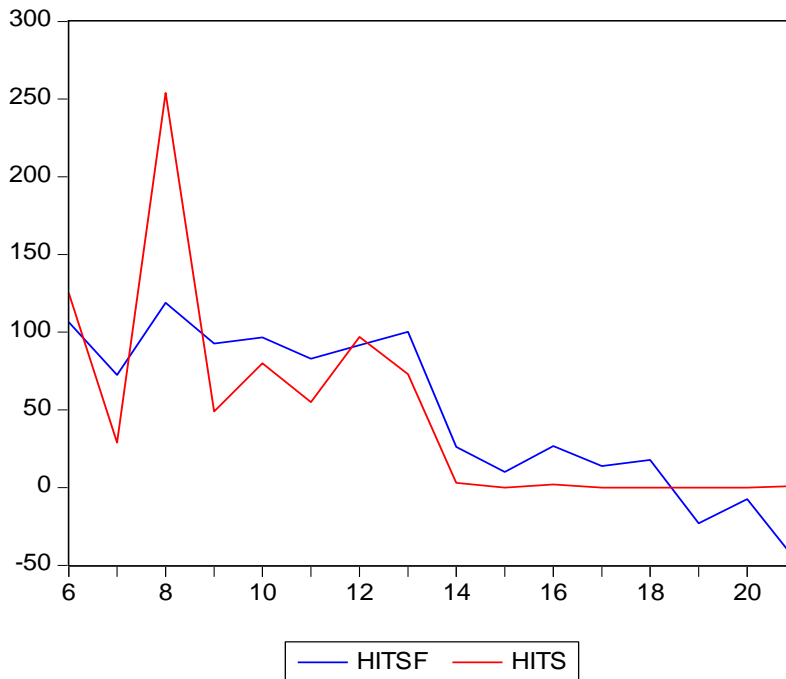
$$\mathbf{HITS_{MEN} = 0.011516 * RATING}$$

As it turns out, personality is overrated on the internet, as is talking about your hobbies. The only things that matter are your gender, your rating, and your weight. In order to be approached often in the world of online dating, one should be hot, female, and thin.

For men, the prognosis is grim. Chances are you won't be finding anyone at all on the internet. Even if you are a "perfect ten," according to this model you will get next to no replies. Should you be unable to date and meet women through conventional means, you are screwed... figuratively.

Women, on the other hand, enjoy a virtual buffet of potential suitors. Simply by being a woman, you will on average receive about 57 hits. Should you move your rating up by one point out of a possible hundred, you will gain around 5 more. Should you gain a pound you will lose two or three suitors. As crushing and painful as this may initially seem, keep in mind you will on average have 160 charming Craigslist Casanovas chasing after you.

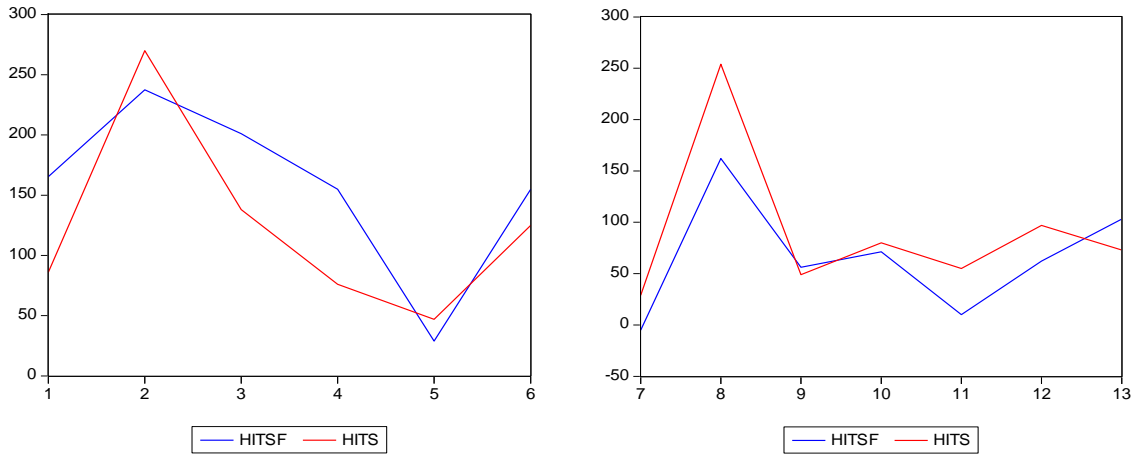
In order to test the accuracy of this model, I made the decision to forecast across the general sample, the female sample, and the male sample. To forecast for the general population, I took the middle portion of the data (so as to remove an equal number of males and females) and forecast the ends, producing the following graph:



One should note that the forecasted hits is usually higher than the actual hits. I believed this discrepancy to be a result of both removing particularly notable data points, as well as trying to forecast the two genders at once. After this, I ran four separate out-of-sample forecasts – one using the first half of the data within the sample, and another with the second half for each gender.. The results are shown below, with the female forecast being shown first:

Female Forecasting

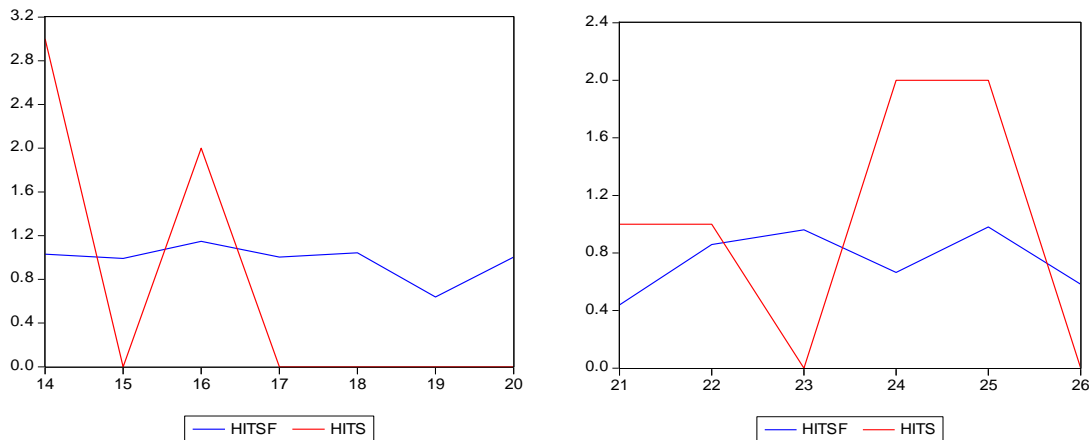
(results of first out-of-sample forecast on the left, second on the right)



The female forecast looks to be accurate enough – the parameters of weight and rating explain enough of the dependent variable hits to provide at least a reasonable estimate, as the out-of-sample forecasting shows. For men, it is a similar story, with a slight twist:

Male Forecasting

(results of first out-of-sample forecast on the left, second on the right)



While at first appearing to be fairly accurate (after accounting for scale in the visual representations) certain meaningful discrepancies do exist. The forecast line is particularly flat – much more so than the female forecast. This may be the result of a real-world limitation on the data: there could exist some personals within my sample that are so hideous (rating was the only significant parameter for the male regression) that they would receive negative hits if it were possible. As this is a practical

limitation rather than a theoretical one, there is little to be done in order to address how it may affect the results of this analysis.

In short, the universal characteristics that determine how attractive a person is are apparently gender, weight, and rating. For males, one's physical attractiveness is the sole determining factor in how often you are approached in online dating. Keep in mind, however, that men are very fortunate to be approached at all. For women rating and weight will both play a role. Your hobbies, your personality, whether you drink or smoke, how tall you are, how well-educated you are, your hair color, and whether or not you have kids is irrelevant. The internet may be a shallow place, but at the very least it should be easy for any woman to find a date.

Technical Appendices: EViews Outputs

General Sample

Dependent Variable: HITS
Method: Least Squares
Date: 12/05/07 Time: 19:41
Sample: 1 26
Included observations: 26

	Coefficient	Std. Error	t-Statistic	Prob.
RATING	1.463770	0.617255	2.371417	0.0265
WEIGHT	-0.566644	0.248486	-2.280390	0.0322
FEMALE	56.93626	26.27825	2.166669	0.0409
R-squared	0.607833	Mean dependent var		53.46154
Adjusted R-squared	0.573731	S.D. dependent var		74.98866
S.E. of regression	48.95955	Akaike info criterion		10.72803
Sum squared resid	55131.86	Schwarz criterion		10.87320
Log likelihood	-136.4644	Hannan-Quinn criter.		10.76984
Durbin-Watson stat	2.384915			

Female Sample

Dependent Variable: HITS
Method: Least Squares
Date: 12/05/07 Time: 18:44
Sample: 1 26 IF FEMALE=1
Included observations: 13

	Coefficient	Std. Error	t-Statistic	Prob.
RATING	4.850498	0.965499	5.023826	0.0004
WEIGHT	-2.351904	0.648271	-3.627967	0.0040
R-squared	0.666440	Mean dependent var		106.0769
Adjusted R-squared	0.636116	S.D. dependent var		75.60364
S.E. of regression	45.60624	Akaike info criterion		10.61860
Sum squared resid	22879.22	Schwarz criterion		10.70552
Log likelihood	-67.02093	Hannan-Quinn criter.		10.60074
Durbin-Watson stat	2.328212			

Male Sample

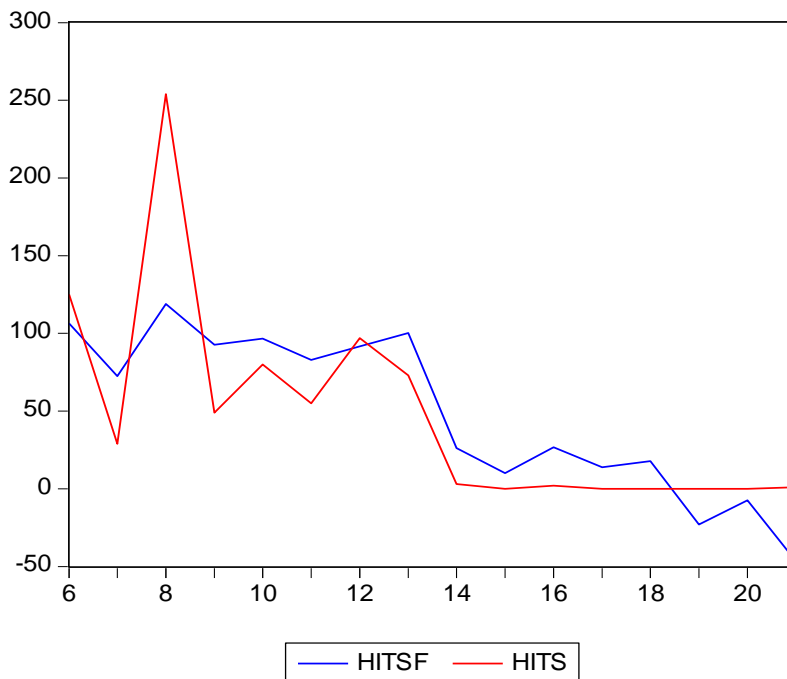
Dependent Variable: HITS
Method: Least Squares
Date: 12/05/07 Time: 18:46
Sample: 1 26 IF FEMALE=0
Included observations: 13

	Coefficient	Std. Error	t-Statistic	Prob.
RATING	0.011516	0.003815	3.018638	0.0107

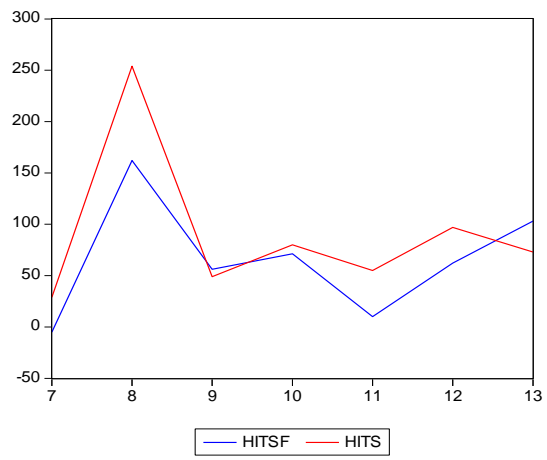
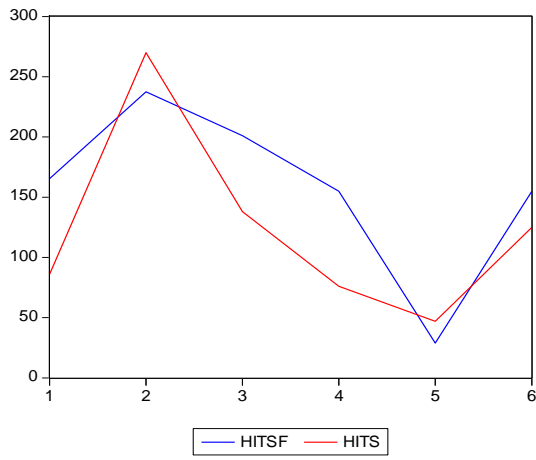
R-squared	0.045229	Mean dependent var	0.846154
Adjusted R-squared	0.045229	S.D. dependent var	1.068188
S.E. of regression	1.043752	Akaike info criterion	2.997325
Sum squared resid	13.07303	Schwarz criterion	3.040782
Log likelihood	-18.48261	Hannan-Quinn criter.	2.988392
Durbin-Watson stat	2.094785		

Forecast Scatterplots

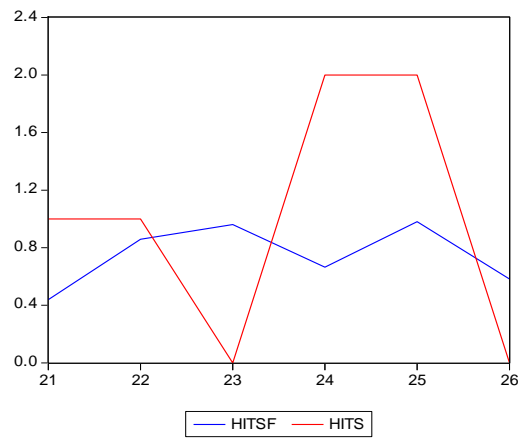
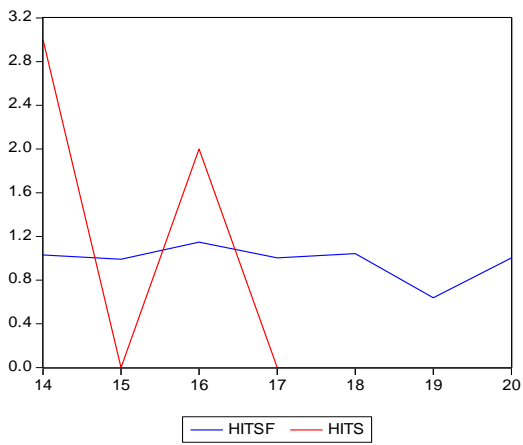
General



Women

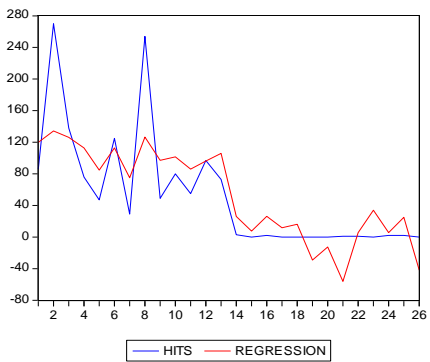


Men

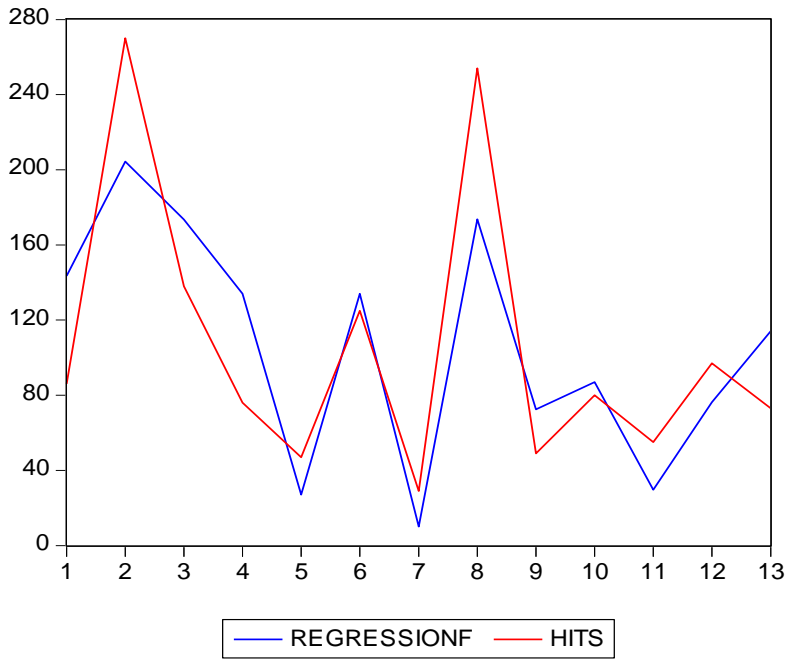


Ordinary Least-Squares Graphs:

General:



Women:



Men:

