

TO: William T. Lake, Media Bureau Chief
FROM: Octavian Carare
DATE: May 5, 2016
RE: Review of "Hispanic Television Study"

In response to your request, I have reviewed a copy of the Hispanic Television Study dated April 2016. I think the study is a major step toward achieving a better understanding of the role played by Hispanic ownership on the programming choices of Hispanic television stations in the United States, and on the viewing choices of these stations' potential audiences. I think the study reflects an impressive and admirable effort on the part of several economists in the Media Bureau and in the Office of Strategic Planning. In the attached document I discuss some of the merits of the study, and I formulate a few recommendations that, I hope, would help improve the quality and accessibility of the study.

Please note that I have submitted my 2015 Confidential Financial Disclosure Forms, and that I am not aware of any actual or perceived conflict of interest with my role of reviewer of this study.

May 5, 2016

Review of the “Hispanic Television Study”
by Octavian Carare

The Hispanic Television Study, referred henceforth as the study, analyzes the effects of Hispanic ownership on the programming choices of Hispanic television stations and on the viewing choices of these stations’ audiences. The study relies on a new and valuable data set painstakingly assembled using a variety of data sources. On the consumer side, the main findings of the study indicate that Hispanic viewers favor the major Spanish-language networks, watch more local, Spanish-language news than English-language news, and watch more telenovelas than other program types. The study also finds that Hispanic-owned stations are less likely than other Hispanic-oriented stations to show telenovelas, and more likely to show paid and local programming. The data also indicate that Hispanic ownership is associated with higher ratings among Hispanics, but that, relative to the viewership of large Spanish-language networks, the viewership of Hispanic-owned stations is small.

I think that the graphs and the summary statistics presented in the study are very informative. In particular, I like how the market-level data are summarized, and the clever way of presenting program characteristics by rating deciles in Table 22.

My main observation about the study concerns its presentation. It is very clear to me, and I hope it will be clear to most of the readers of the study that the effort required to put together the data for this study was quite significant. Indeed, as the authors themselves state, organizing the data for the study presented an unanticipated challenge. This effort notwithstanding, I think that aggregating into a more comprehensive technical data appendix all the details of putting together the data set would make the study more accessible to a wider audience. I think that the practitioners interested in the minute details of putting together such a valuable data set (including, e.g., most of the steps described in paragraphs 30, 31, 35, 39, 40, 43, 44, etc.) would be pleased to find all these details in one place, and the wider audience interested in the analysis and its results would not be encumbered by a lengthy description of some of the finer data points. Deciding which piece of information is best included in the technical appendix and which piece of information is essential for a clear understanding of the results is not a trivial process, so I urge the authors of the study to give this some serious consideration.

The remainder of this review focuses primarily on some of the technical details of the data analysis.

The market-level regressions described by the equations on p. 38 seek to measure the relationship between the number of programming minutes and the average Hispanic audience ratings on one hand, and, on the other hand, various market-level characteristics that include a measure of the presence in the market of one or more Hispanic-owned stations. The linear-in-logarithms structure of these equations implies that the number of programming minutes (and the average ratings) can be expressed

May 5, 2016

multiplicatively in terms of a collection of variables including – *inter alia* – the number of stations (Spanish and otherwise), the number of Hispanic TV households, and a measure of income. The structure of these estimating equations seems ad-hoc; indeed, I see no *a priori* reason why, for example, doubling the income in a market would bring about a doubling, on average, of the number of programming minutes (and of the average ratings). Furthermore, as it is apparent from the discussion preceding Table 16, the estimating equations do not accommodate markets in which any of the numbers of stations is equal to zero. I think that taking the logarithm of the left-hand-side variables might be a good idea in principle, primarily because as a result one can deal with the potential complications arising from the presence of truncated error terms when the dependent variables are in levels.¹ However, I am not sure that taking logarithms of the right-hand-side variables is appropriate, and I urge the authors of the study to experiment with various other specifications that might include polynomials, etc. More importantly, I think that providing an explanation for the functional form of the estimating equations and a formal interpretation of the coefficients that precedes the presentation of the results would increase significantly the quality of the study.

One last point about the market-level regressions is worth discussing here. In general, the error term in a regression reflects the effect on the dependent variable of one or more unobserved variables. Suppose that one of the potential variables unobserved by the econometrician is weather (hot or rainy weather is likely to be associated with more TV viewing, all other things equal). In regions with hot or rainy weather, Hispanic (and non-Hispanic) households likely watch more TV. This, in turn, might allow more stations to operate a given market. If this chain of reasoning is correct, then the market-level estimates presented in Tables 16-17 are biased.² One possibility to correct for this bias is to use instrumental variables. It is conceivable that a binary variable equal to 1 if a market is adjacent to Mexico, and zero otherwise, might be a valid instrument for the Hispanic ownership variable. If not, I urge the authors to investigate the patterns, geographic and otherwise, of Hispanic ownership in search of a good instrument.

My next few comments concern the household-level regressions presented in the study. I think it is worth mentioning here that the empirical framework outlined on top of p. 50 does not permit the identification of the Hispanic viewers' preferences. Because the viewing choices observed in the data arise as the interplay between supply (the programs offered at any point in time by the stations competing in a particular market) and demand (i.e., viewer preferences), regressions similar to that on p. 50 only allow an examination of the Hispanic viewers' *observed choices*. I think that is just fine for the purposes of this study, but I also think that one does not want to oversell the study's results.

¹ Since minutes and ratings are positive, the error terms when the left-hand-side variables are in levels are left-truncated. The logarithms of these values may take on negative values.

² Unobserved weather-related variables are only one of the potential examples of missing variables that may be correlated with one or more of the independent variables. It is worth noting that, in general, any unobserved variables that tend to increase or decrease viewership in a market are also likely to be correlated with the number of stations in that market.

May 5, 2016

I had a hard time understanding the meaning of the dependent variables in the regressions outlined at the top of p. 50. I take it from context that the dependent variable represents some household-level measure of viewing. If so, the assertion made at the bottom of p. 50 that the dependent variable is censored does not make sense. I think that, since by definition a measure of viewership such as minutes or ratings must be greater than or equal to zero, the error term in the regression on top of p. 50 is left-truncated (because values of the error term smaller than some value would be associated with negative ratings, or negative minutes watched, an impossibility). The fact that observationally the dependent variable appears to be censored (i.e., there is a mass point at zero in the distribution of viewership, and all other observations entail positive ratings) does not mean that it is actually censored. I think that the analytical framework that applies best to these data is a two-part model (the first part looks at whether the dependent variable is zero or not, the second part models the value of the dependent variable if it is positive).

I am puzzled about the regressions outlined at the bottom of p. 53 (the programming choice regressions). It looks like that the dependent variables in these logit regressions are binary variables that characterize the ownership and Spanish network affiliation of a particular station, and that the independent variables are program characteristics and a set of interaction terms. I think the dependent and independent variables in these regressions are in fact reversed; indeed, I think one best views a station's programming choices as affected by the station's ownership structure, rather than the other way around. As a result, I am having trouble interpreting the estimates provided in Table 25. At a minimum, I would like to see included in the study a detailed explanation of why this seemingly non-standard approach yields estimates that can be readily interpreted.

The regressions outlined at the bottom of p. 56 aim to establish an empirical relationship between program ratings and Hispanic ownership. As above, I do not think that the dependent variables in these regressions are censored.³ It is also worth mentioning that, at any point in time, the ratings of a station in a given market are determined jointly with the contemporaneous ratings of the other stations in that market. This means that the estimates obtained by analyzing one station's ratings as disjoint from the other stations' contemporaneous ratings might be biased. One way to avoid all these complications is to estimate a model of discrete television choice (see e.g., Goettler & Shachar, *RJE* 32(4), 2001).

Overall, I think the study brings an important contribution to our understanding of the role played by Hispanic ownership on programs and ratings. I think that, while some of the regression analyses in the study may need to be revamped, overall the study reflects an impressive research effort that is commendable and the results of the study are informative.

³ Since ratings are bounded from above, the error term in the equations on p. 56 might indeed be viewed as doubly truncated.