Douglas Locke

April 20th, 2018

Advanced Quant Methods – Spring 2018

The paper I chose for this assignment is called “The Ins and Outs of Homeschooling: The Determinants of Parental Motivations and Student Achievement” by Ed Collom, University of Southern Maine; Published in Education & Urban Society, May 2005.

I am interested in alternative schooling approaches as a way to better understand student achievement in non-typical schooling environments. The previous California Schools assignment revealed a very well researched and known phenomenon, namely that income explains much of the variance in student achievement test scores. I thought it would be interesting to find papers that explored similar data but instead “change the environment”, namely the public school system itself, to see if the household income or other factors would still play a significant role. There is still a low population of home schooled students relative to both public and private schooled students (in 2013, the US Department of Education estimated 3.4% of all students are homeschooled). Given this fact, it would also be interesting to understand how data collection methods could produce a reliable study given the assumption that home school environments are decentralized and vary widely in structure and operation.

Collom’s research solves some of these data collection issues by collecting & exploring original survey data from an organized group of homeschoolers. His major conclusions are divided into learnings about parental motivators, and student achievement. Regarding parental motivators, Collom concludes “Academic and pedagogical concerns are most important, and it appears that the religious base of the movement is subsiding. Several major demographic variables have no impact upon parental motivations, indicating that this is a diverse group.”

Collom concludes that for student achievement: “Parents educational attainment and political identification are consistent predictors of their student’s achievement. Race and class – the two major divides in public education – are not significant determinants of standardized test achievement, suggesting that homeschooling is efficacious.” And in his conclusion: “Minority students and those from low-income families have consistently been found to be at disadvantage in the public education system. Home schooling apparently levels the playing field, ameliorating the negative affects [sic] that race and class subordination have shown in the public schools.” In reading these claims I thought they were provocative in a way that challenges common societal norms or conventional thinking.

In a final summarizing statement Collom adds, “It is concluded that homeschoolers are a heterogeneous population with varying and overlapping motivations.”

To summarize the data set employed, there is a data set with a large sample size (n=235 parents representing 391 students) and “the response rate is the highest of any known homeschooling sample (71%).” The data was assembled from a southern California K-12 charter school and “is essentially an organized homeschooling operation. Its educational charter identifies parents as the primary instructors.” Further, as a charter school, students are required to participate in California’s Standardized Testing & Reporting Program.

As Collom notes, historically homeschooling has two major roots in the United States: 1) 1960s and 70s left-liberals generally seeking less schooling structure and 2) 1980s and 1990s right-religious seeking religious values education and generally higher levels of structure. By using a charter school that is essentially a homeschool, Collom could be reasonably certain that solve some of the challenges in data collection methods amongst a diverse and difficult to measure group. By having a large enough sample with enough demographic diversity, he also could generalize his findings to “homeschoolers” not just homeschoolers of one of these left/right paradigms, or in a narrow income bands, or with a particular religious outlook, etc. Further, having data from standardized test results consistent with the state of California, presumably comparisons could also be made against other types of students in the state of California. I felt the paper missed this opportunity however. That, may be deliberate, given the paper also suggests that homeschoolers who take standardized tests may not represent homeschoolers in general. In other words, homeschoolers who take standardized tests are likely to be only from the more “conservative” side, and not be representative of homeschoolers in geneal.

I liked this paper particularly because it included several techniques discussed in our class. The research neatly combines factor analysis techniques and models together. For example, principal component factor analysis was used to determine and explain parent motivational patterns. These patterns in turn were then used in regression models as dependent variables to identify demographics of parents who homeschool for various reasons. This I felt was more interesting then some research papers which only rely on a single statistical technique.

The principal component analysis was used to analyze the parental survey data. This was appropriate given the survey was constructed using Likert scales on 16 response items. The principals component analysis helped the author to identify four broad themes regarding parental motivations: “dissatisfaction with public schools, academic & pedagogical concerns, religious values, and family life.” It was important for Collom to establish that these motivations matched some prior homeschool research studies; thus he was able to determine that his sample of homeschoolers was reflective of homeschoolers in general, and didn’t have bias in his sample towards any particular group. The distinguishing item in Collom’s study from priors is the richer data set that allows for multi-variate modeling analysis, as we’ll see.

The paper includes the results of the factor analysis and the Cronbach alpha values. For the 4 factors identified, the Cronbach’s alphas ranged from .58 to .79. Field (p.799) discusses cut-offs of .7, however Field makes warnings that Cronbach alpha values should be interpreted carefully given that the alpha increases with the number of items on the scale. Thus given my own interpretation of the factors & the survey items in the paper, the factors seem very reasonable and indeed, reliable. It is informative to study and contrast the highest alpha value with the weakest. The highest .79 (Attracted to Home Charter Schools) one can clearly see the very consistent focus in the survey items, as each begins with “Home Charter….” and then relates a comment about the program strength, such as “Home Charter’s educational program is of superior quality” and “Home Charter’s strength and focus on science education.” One item that appeared to be a stretch however was including a survey item in this first factor with only a .06 alpha “The opportunity to give my child religious instruction.” It doesn’t seem to belong with the others and I suspect it may be an error in the table (Table 1 page 318).

However, in the lowest alpha value .59 (Family Life) the survey items are certainly less focused. One begins “The scheduling of other schools is too inflexible” and another “My child(ren) have special learning needs that cannot be met.” From these we can see how these questions together while related to “Family Life” are not as tightly related as the factor previously discussed.

The author takes an interesting approach then in combining the outcomes of the factor analysis and using them as dependent variables. In Table 2 , he establishes that for this n=235 sample of parents, descriptive statistics give us a type of “score” or “index” on each parent across the 4 motivational factors. So each motivational factor ranges from 4-20, and he provides mean and standard deviation. I think it would have been also helpful to include measures of skew to understand how normal these distributions are. Is it possible there could be something bi-modal here? We don’t know for sure from the paper. These 4 motivational factor dependent variables serve will drive the creation of 4 models to understand the demographic factors driving them.

Because 6 of the 13 parent demographic variables are actually dummy variables, it would be interesting also to run the analysis as regression trees. The author took time to quantify variables like “religiosity” and “political affiliation”. While this is worthwhile, it would also be interesting to have categorical data on items such as these. For example, if religious data included splits on Agnostics, Atheists, Christians, Jewish, Muslims, etc. or political party affiliation (Democrat, Independent, Republican, etc.) it would be interesting to see if the models picked up on any differences amongst these, and perhaps provide a deeper insight into the data. Also its not clear why the author picked the bins he used for Parent Age, but trees would be a convenient way to help determine these.

In Table 3, he shows the results of 4 Ordinary Least Squares (OLS) models, 1 each for the motivational factors of parents. While only a few demographic factors were deemed significant across the models, the results are interesting. Some results are expected, for example Religiosity explains “Ideological Reasons [for Homeschooling]” and Teaching Experience explains “Family and Children Needs.” Perhaps something more interesting is that “Minority” (when value =1 or True) and “Spouse Involvement” explains “Critical of Public Schools”. The R2 values of each these models is low (.10 to .28). My understanding is this shows the difficulty in identifying the all the possible demographics and factors underlying something like a motivation, which is essentially a very conceptual & personal thought concept, like an incentive. So there are likely many third variables lurking here. One item might be to include # of children in the household, as its possible larger families may find it actually to lower the cost per student to have a parent teach at home. There was little discussion on OLS assumptions, although there is a mention of VIF being < 2 (highest is 1.49 in the enrollment motivations analysis and 2.35 in the achievement analysis), so basically no multi-collinearity. I thought it would be interesting to show all the necessary assumptions being met, at least in the Appendix.

In reviewing the parts of the paper on achievement, again Collom employs OLS models, using standardized achievement tests as dependent variables. Again, one of the key finding in this paper is that income (in binned variables) & race (represented as a dummy variable where minority = 1 when true) seem to not effect student achievement. One realization I had in reviewing this paper where a major conclusion is based on a lack *not* finding something places a great emphasis on the data collection and ensuring it is not a biased sample. The author states that in the data, 16.6% represent minority parent/teachers, which he states is in line with previous studies that show 80-90% of homeschool families are white.

There are a few concepts I would want to build on this author’s approach. First, not all parents within the same family may have the same motivating factors for homeschooling. It would be interesting to see if one parent perhaps had greater influence in this. Second, there seems to be no accommodation for analyzing mixed race families. Approximately 15% of new marriages in 2010 were of mixed-races in the U.S. according to the U.S. Census. To better understand minority effects, this additional data may shed additional light. Third, without longitudinal data, its not clear how the motivational factors may shift over time. For example, perhaps a homeschooling parent may initially be dissatisfied with public schools, but over time as they gain experience as a teacher, they shift towards appreciation of various pedagogical methods. Fourth, it would be interesting to see if the findings are the same across a more national geography.

From a statistical approach, I would like to see various model approaches compared against each other. However, the size of the data is small, and I do believe tree modeling such as CART was available at the time. The author makes no mention of attempting any other methods.

In summary, through this exercise I was able to appreciate a deeper, applied examination of factor analysis & OLS regression techniques working together. I did not find any egregious errors that I felt invalidated any of the findings. The research was clearly laid out, presented in a logical order, and included helpful tables with many statistical details, although I would say it was not completely exhaustive. There are some deeper levels of questions left begging for future study.