

Econ 753 HW2

Evan Wasner

10/13/2020

Part 1 - Berndt Exercises

In part 1, we analyze a hedonic price model constructed by Gregory Chow in the rental market for computers. Chow constructs a quality-controlled model of price indices using memory time, multiplication time, and access time as the set of qualitative characteristics. We are asked to replicate Chow's results and perform further analysis.

Question 3a

In question 3a, we construct a correlation matrix for the regression variables, both for the 1954-1959 and 1960-1965 time periods.

Table 1: Correlation Matrix for 1954-1959

| | lnrent | lnmult | lnaccess | lnadd | lnmem |
|----------|------------|------------|------------|------------|------------|
| lnrent | 1.0000000 | -0.8673364 | -0.8450792 | -0.8397468 | 0.8756174 |
| lnmult | -0.8673364 | 1.0000000 | 0.8626912 | 0.9771873 | -0.7371050 |
| lnaccess | -0.8450792 | 0.8626912 | 1.0000000 | 0.8457563 | -0.6129908 |
| lnadd | -0.8397468 | 0.9771873 | 0.8457563 | 1.0000000 | -0.7295567 |
| lnmem | 0.8756174 | -0.7371050 | -0.6129908 | -0.7295567 | 1.0000000 |

Table 2: Correlation Matrix for 1960-1965

| | lnrent | lnmult | lnaccess | lnadd | lnmem |
|----------|------------|------------|------------|------------|------------|
| lnrent | 1.0000000 | -0.6331708 | -0.5849575 | -0.6316314 | 0.8825522 |
| lnmult | -0.6331708 | 1.0000000 | 0.7739316 | 0.9405762 | -0.5707293 |
| lnaccess | -0.5849575 | 0.7739316 | 1.0000000 | 0.8064029 | -0.4648790 |
| lnadd | -0.6316314 | 0.9405762 | 0.8064029 | 1.0000000 | -0.5633702 |
| lnmem | 0.8825522 | -0.5707293 | -0.4648790 | -0.5633702 | 1.0000000 |

We can see that the variables are all highly correlated with each other, with many of the correlation coefficients coming close to 1. This indicates that Chow's concern over collinearity was a valid concern, as several of the independent variables maintained high correlation coefficients (e.g. the correlation between lnmult and lnaccess is 0.863 in the 1954-1959 range and 0.774 in the 1960-1965 range).

The correlation coefficients maintained the same signs throughout the two periods, but changed somewhat significantly in values, with the change in magnitudes of some of the correlation coefficients changing by as much as 30%. Therefore, it was correct for Chow to perform separate regressions for the two time periods,

as the relation between qualitative characteristics changed somewhat significantly between the two price periods.

Question 3b

In part b, we run a regression for the years 1960-1965 of $\ln\text{rent}$ on $\ln\text{mult}$, $\ln\text{mem}$, $\ln\text{access}$, and a list of dummy variables indicating if the observation takes place in a particular year (excluding the year 1960).

Table 3: Quality-Controlled Regression of $\ln\text{rent}$ on Quality-Variables

| | <i>Dependent variable:</i> |
|--|----------------------------|
| | $\ln\text{rent}$ |
| Constant | −0.104 (0.315) |
| d61 | −0.140 (0.166) |
| d62 | −0.489*** (0.174) |
| d63 | −0.594*** (0.166) |
| d64 | −0.925*** (0.166) |
| d65 | −1.163*** (0.166) |
| $\ln\text{mult}$ | −0.065** (0.028) |
| $\ln\text{mem}$ | 0.579*** (0.035) |
| $\ln\text{access}$ | −0.141*** (0.029) |
| Observations | 82 |
| R^2 | 0.908 |
| Adjusted R^2 | 0.898 |
| Residual Std. Error | 0.384 (df = 73) |
| F Statistic | 90.522*** (df = 8; 73) |
| <i>Note:</i> *p<0.1; **p<0.05; ***p<0.01 | |

The regression successfully replicated Chow's results. Each of the qualitative variables result in the correctly predicted signs; multiplication time and access time both correlate negatively with computer rental price (the longer the multiplication times and access times, the slower the computer and thus the lower the price), while memory correlates positively with rental price (more memory means higher quality computer). Furthermore, each of the qualitative variables are statistically significant.

| | | | | | |
|--------------------|---------------------------|---------------------------|---------------------------|---------------------------|-------|
| | | (0.0779) | (0.0821) | (0.0999) | |
| Pooled Run: | | | | | |
| 1960-1965 | -0.1045 | -0.0654 | 0.5793 | -0.1406 | 0.908 |
| | | (0.0284) | (0.0354) | (0.0293) | |
| | -0.1398 · D ₆₁ | -0.4891 · D ₆₂ | -0.5938 · D ₆₃ | -0.9248 · D ₆₄ | |
| | (0.1665) | (.1738) | (0.1661) | (0.1663) | |

Figure 1: Chow's Original Regression

The coefficients on the year dummy variables decrease as the year increases, which corresponds to the rental price of computers lowering over time. This intuitively makes sense, as computers have consistently improved in quality while lowering in price for the past 50 years. After running the regression, a list of price indices for the years 1960-1965 is generated by taking the antilogarithm of the dummy variable coefficients. Since the regression was run on the natural logarithm of the rental price, and since the coefficients are relatively small, the antilogarithm of the dummy variable coefficients approximately equals the percent change in the rental price from the year 1960 to the year of the dummy variable. The resulting price indices are identical to Chow's price indices, located in the middle column of the screenshot of the Berndt textbook.

Table 4: Replicated Price Indices

| | Year | Estimated_Coefficient | Price_Index |
|-------------|------|-----------------------|-------------|
| (Intercept) | 1960 | -0.1044553 | 1.0000000 |
| d61 | 1961 | -0.1397971 | 0.8695346 |
| d62 | 1962 | -0.4891062 | 0.6131742 |
| d63 | 1963 | -0.5938483 | 0.5521982 |
| d64 | 1964 | -0.9248183 | 0.3966035 |
| d65 | 1965 | -1.1631690 | 0.3124943 |

Table 4.2 Quality-Adjusted Price Indexes for Computers, 1960-1965. Based on Chow's Dummy Variable Coefficient Estimates and on Chow's Arithmetic-Weighted Procedure

| Year | Estimated Coefficient | Price Index (Antilogarithm) | Price Index* (Arithmetic-Weighted) |
|------|-----------------------|-----------------------------|------------------------------------|
| 1960 | | 1.0000 | 1.0000 |
| 1961 | -0.1398 | 0.8695 | 0.8438 |
| 1962 | -0.4891 | 0.6132 | 0.6414 |
| 1963 | -0.5938 | 0.5522 | 0.5330 |
| 1964 | -0.9248 | 0.3966 | 0.3906 |
| 1965 | -1.163 | 0.3125 | 0.3188 |

*Entries in the final column are taken from Chow [1967, Table 2, p. 1124], normalized to unity in 1960.

Figure 2: Chow's Original Price Indices

Question 3e

In part e, we are asked to perform a generalized least squares (GLS) regression, using the square root of the number of observations (volume) as the weight. Using the number of observations as a weight for our regression attempts to overcome the issue of heteroskedasticity, as OLS ceases to be the best linear unbiased estimator (BLUE) in the case of heteroskedasticity.

Table 5 displays a comparison between the OLS and GLS regressions. None of the coefficients change significantly: all retain their signs and remain more or less within range of the OLS regression. Furthermore,

Table 5: Comparison of OLS and GLS for Hedonic Price Model

| | <i>Dependent variable:</i> | |
|-------------------------------|-----------------------------|----------------------|
| | lnrent | |
| | OLS (1) | WLS (2) |
| Constant | −0.104 (0.315) | −0.544 (0.335) |
| d61 | −0.140 (0.166) | −0.197 (0.160) |
| d62 | −0.489*** (0.174) | −0.483** (0.198) |
| d63 | −0.594*** (0.166) | −0.401*** (0.147) |
| d64 | −0.925*** (0.166) | −0.802*** (0.166) |
| d65 | −1.163*** (0.166) | −0.969*** (0.154) |
| lnmult | −0.065** (0.028) | −0.024 (0.028) |
| lnmem | 0.579*** (0.035) | 0.605*** (0.038) |
| lnaccess | −0.141*** (0.029) | −0.160*** (0.030) |
| Observations | 82 | 82 |
| R ² | 0.908 | 0.892 |
| Adjusted R ² | 0.898 | 0.880 |
| Residual Std. Error (df = 73) | 0.384 | 0.768 |
| F Statistic (df = 8; 73) | 90.522*** | 75.574*** |
| <i>Note:</i> | *p<0.1; **p<0.05; ***p<0.01 | |

Generally, performing a log-log regression tends to reduce the impact of heteroskedacity, as the logarithm function “pulls in” large numbers, thus reducing the spread between error terms. However, when variables are small numbers, log-log regressions can actually worsen the problem of heteroskedacity, as taking the logarithm of small numbers tends to increase the spreads. In our case, most of the variables we are dealing with are greater than 1, and thus the problem of heteroskedacity is probably mitigated a bit by performing the log-log regression.

Part 3 - Replication Paper Topic

Clark LP, Millet DB, Marshall JD (2014). National Patterns in Environmental Injustice and Inequality: Outdoor NO_2 Air Pollution in the United States. PLoS ONE 9(4): e94431. doi:10.1371/journal.pone.0094431

For my replication paper, I will replicate the paper titled “National Patterns in Environmental Injustice and Inequality: Outdoor NO_2 Air Pollution in the United States” by Clark et al. The paper describes spatial patterns in environmental injustice and inequality for nitrogen dioxide (NO_2) concentrations in the United States. It uses Census demographic data and a recently published dataset of outdoor (NO_2) concentrations. It compares (NO_2) concentrations in populations of different demographics and in different locations (e.g. urban vs rural).

There are three tables and two figures. Table 1 provides data on (NO_2) concentration for different populations. Table 2 offers comparisons of specific populations. Table 3 provides “environmental justice” and “environmental inequality” indices for different areas. Figure 1 plots a regression of (NO_2) concentration on household income for different regions. Figure 2 projects the environmental justice and inequality indices onto a map of the United states. I am interested in replicating all tables and figures, although I will likely be unable to replicate the figure of the US map. Finally, I am interested in taking a deeper look at the data and offer any other revelations I come across.