Fall 2003 Econ120B

Answers to HW2

Part I

9.

Procedure goes as follows:

1. Omit *localtax*;
2. Omit *unemprt*;
3. Omit *statetax*;
4. Omit *density*.

The rationale to omit each variable is that the coefficient of each variable has least significance, that is, it is most probable that it is not different from zero indicated by the largest value of p-value among others. In addition to that, those variables seem to cause multicollinearity as inspected from the correlation matrix obtained in Step 3.

10.

The model is given by

$$\begin{aligned} housing_t = \beta_1 &+ \beta_2 density_t + \beta_3 value_t + \beta_4 income_t \\ &+ \beta_5 popchang_t + \beta_6 unemprt_t + \beta_7 localtax_t + \beta_8 statetax_t + u_t. \end{aligned} \tag{1}$$

The description of variables in the equation is available at p.646 in the textbook.

The dependent variable $house_t$ is the number of new private housing units authorized by building permits. Since the number of newly authorized housing cannot be negative (even in the middle of nowhere), we expect β_1 to be nonnegative.

$density_t$ represents population density per square mile. When population density is increased, people would demand more comfortable housing. Thus β_2 is expected to be positive.

$value_t$ is median value of owner-occupied homes. When the median value of owner-occupied homes is increased, that is, the price of a house is increased, the demand for housing would be less. β_3 will be negative.

$income_t$ is median household income. When people become richer, it is quite natural to buy a house. So β_4 will be positive.

Percent increase in population is denoted by $popchang_t$. If population is increased and if population density is held constant, the number of housing should increase. Thus β_5 should be positive.

$unemprt_t$ represents unemployment rate. When the unemployment rate increases, there are more workers whose economic condition gets worse. This implies that the increased unemployment rate has negative effect on the demand for housing, that is, $\beta_6 < 0$.

$localtax_t$ and $statetax_t$ are average local tax and state tax per capita, respectively. When taxes are increased, disposal income is decreased so that demand for housing will be dropped. Also increased tax would make owing a house relatively less attractive. So both β_7 and β_8 are expected to be negative.

The estimated result for Model (1) is given in Table 1. Estimated results are consistent with our intuition. The correlation matrix obtained in Step 3. is presented in Table 2.

variable	coefficient	expected sign
constant	813.368	≥ 0
density	0.0752928	> 0
value	-0.855036	< 0
income	110.411	> 0
popchang	26.7659	> 0
unemprt	-76.5464	< 0
localtax	-0.0611163	< 0
statetax	-1.00593	< 0

Table 1: Estimated Coefficients

There are some evidences suggesting multicollinearity. For example, *density* and *localtax* are highly correlated (correlation coefficient = 0.7320).

density	*value*	*income*	*popchang*	*unemprt*	*localtax*	*statetax*
1.0000	0.4574	0.0447	-0.3572	0.2733	0.7320	0.5475
	1.0000	0.7136	0.2663	-0.3136	0.3327	0.6222
		1.0000	0.5674	-0.6500	0.1585	0.4114
			1.0000	-0.5790	-0.2652	0.0304
				1.0000	-0.0237	-0.0853
					1.0000	0.2846
						1.0000

Table 2: Correlation Matrix

11.

The last model is

$$housing_t = \beta_1 + \beta_3 value_t + \beta_4 income_t + \beta_5 popchang_t + v_t. \tag{2}$$

The null hypothesis and the alternative are given by

$$\begin{aligned} H_0 : &\quad \beta_2 = \beta_6 = \beta_7 = \beta_8 = 0 \\ H_1 : &\quad \text{at least one of them is different from zero} \end{aligned}$$

We need to compute the following test statistic:

$$F_c = \frac{(R_U^2 - R_R^2)/(k-m)}{(1-R_U^2)/(n-k)} \quad \sim \quad F_{k-m,n-k} \tag{3}$$

where R_U^2 corresponds to R^2 obtained from Model (1) and R_R^2 is the one from Model (2). We have 8 coefficients in the unrestricted model, i.e., $k = 8$, and 4 coefficients in the restricted model, i.e., $m = 4$. Then we have

$$F_c = \frac{(0.349371 - 0.311826)/(8-4)}{(1-0.349371)/(40-8)} \approx 0.46$$

and the critical value at 5% level is $F^*_{4,32}(0.05) \in (2.61, 2.69)$. So we fail to reject H_0, implying that those coefficient are statistically insignificantly different from zero at 5% level.

Part II

1.

$$\begin{aligned} \textit{True Model:} \quad Y_t &= \beta X_t + u_t && (4) \\ \textit{Estimated Model:} \quad Y_t &= \alpha + \beta X_t + u_t && (5) \end{aligned}$$

OLS estimator of β, using the wrong model, is given by

$$\hat{\beta} = \frac{\sum X_t Y_t - \frac{1}{n}(\sum X_t)(\sum Y_t)}{\sum X_t^2 - \frac{1}{n}(\sum X_t)^2}. \tag{6}$$

Using (5), we have

$$\begin{aligned} \hat{\beta} &= \frac{\beta \sum X_t^2 + \sum X_t u_t - \frac{1}{n}\beta(\sum X_t)^2 - \frac{1}{n}(\sum X_t)(\sum u_t)}{\sum X_t^2 - \frac{1}{n}(\sum X_t)^2} && (7) \\ &= \beta + \frac{1}{\sum X_t^2 - \frac{1}{n}(\sum X_t)^2}\left(\sum X_t u_t - \frac{1}{n}\sum X_t \sum u_t\right) && (8) \end{aligned}$$

Taking expectation of (8) and using Assumption3.3 ($E[u_t] = 0$) and Assumption3.4 (non-randomness of X_t), we have

$$E[\hat{\beta}] = \beta \tag{9}$$

Thus $\hat{\beta}$ is unbiased, implying adding an irrelevant constant term does not bias the remaining coefficients.

2.

This statement is correct. Suppose there are m linear combinations to test. To obtain the restricted model, we would solve each restriction for one of the parameters (different each time) and substitute in the original model. This would mean m fewer parameters in the restricted model. The difference in the d.f. between the restricted and unrestricted model would therefore be also m.

3.

The statement is wrong. Although t-tests might indicate individual insignificance, several variables may be jointly significant. If all the insignificant variables are dropped, we are likely to introduce serious omitted variable bias.

4.

This statement is also wrong. Although multicollinearity does rise the standard errors, the estimates are unbiased and consistent and the t- and F-distributions are valid. Therefore the tests are valid.

5.

False. The opposite is true. Multicollinearity increases the standard errors and *lowers* t-statistics. A lower t-statistic is likely to make a variable insignificant rather than significant.

6.

This statement is not valid because high multicollinearity does not affect the assumptions made on the model and hence the properties of unbiasedness, consistency, and efficiency are unaffected by multicollinearity.

7.

It was pointed out in Property 4.4 that adding an irrelevant variable still yields estimates that are unbiased and consistent. But because their variance will be higher than that using the true model, the estimates are inefficient. Multicollinearity, in contrast, does not affect any of the properties and hence estimates are also BLUE, that is, efficient. Hence the statement is only partly true.