ECON 5P04 Panel Analysis Assignment
Option 1: Charitable Contributions

Madeline Arnott
Greg DeVilliers
Zack Lansfield

March 17, 2020

a)

Model 1: Pooled OLS, using 470 observations
Included 47 cross-sectional units
Time-series length = 10
Dependent variable: charity

	Coefficient	Std. Error	t-ratio	p-value
const	−4.67422	1.29813	−3.601	0.0004
income	1.03578	0.128944	8.033	0.0000
price	0.483092	0.207703	2.326	0.0205
age	1.54727	0.216955	7.132	0.0000
ms	−0.00803641	0.184849	−0.04348	0.9653
deps	0.175368	0.0426421	4.113	0.0000

Mean dependent var	6.577150	S.D. dependent var	1.313659
Sum squared resid	627.6639	S.E. of regression	1.163067
R^2	0.224488	Adjusted R^2	0.216131
$F(5, 464)$	26.86280	P-value(F)	7.06e–24
Log-likelihood	−734.8801	Akaike criterion	1481.760
Schwarz criterion	1506.677	Hannan–Quinn	1491.563
$\hat{\rho}$	0.791334	Durbin–Watson	0.405122

Model 2: Fixed-effects, using 470 observations
Included 47 cross-sectional units
Time-series length = 10
Dependent variable: charity

	Coefficient	Std. Error	t-ratio	p-value
const	−2.08997	1.13112	−1.848	0.0654
income	0.838810	0.111267	7.539	0.0000
price	0.366080	0.124294	2.945	0.0034
age	0.102249	0.208039	0.4915	0.6233
ms	0.199833	0.263890	0.7573	0.4493
deps	−0.0863524	0.0534826	−1.615	0.1072

Mean dependent var	6.577150	S.D. dependent var	1.313659
Sum squared resid	191.6735	S.E. of regression	0.677163
LSDV R^2	0.763177	Within R^2	0.134974
$F(51, 418)$	26.41239	P-value(F)	2.4e–101
Log-likelihood	−456.1204	Akaike criterion	1016.241
Schwarz criterion	1232.183	Hannan–Quinn	1101.198
$\hat{\rho}$	0.341431	Durbin–Watson	1.113096

Joint test on named regressors –
Test statistic: $F(5, 418) = 13.0445$
with p-value = $P(F(5, 418) > 13.0445) = 8.21589\text{e-}12$

Test for differing group intercepts –
Null hypothesis: The groups have a common intercept
Test statistic: $F(46, 418) = 20.6697$
with p-value = $P(F(46, 418) > 20.6697) = $ 2.3426e-81

Model 3: Random-effects (GLS), using 470 observations
Included 47 cross-sectional units
Time-series length = 10
Dependent variable: charity

	Coefficient	Std. Error	z	p-value
const	−2.37057	1.11486	−2.126	0.0335
income	0.852996	0.108734	7.845	0.0000
price	0.370199	0.125398	2.952	0.0032
age	0.277063	0.201695	1.374	0.1695
ms	0.199669	0.233954	0.8535	0.3934
deps	−0.0362540	0.0492893	−0.7355	0.4620

Mean dependent var	6.577150	S.D. dependent var	1.313659
Sum squared resid	698.6427	S.E. of regression	1.225748
Log-likelihood	−760.0567	Akaike criterion	1532.113
Schwarz criterion	1557.030	Hannan–Quinn	1541.916
$\hat{\rho}$	0.341431	Durbin–Watson	1.113096

$$\hat{\sigma}_v^2 = 0.866645$$
$$\hat{\sigma}_\varepsilon^2 = 0.458549$$
$$\theta = 0.775831$$

Joint test on named regressors –
Asymptotic test statistic: $\chi^2(5) = 70.9941$
with p-value = 6.36362e-14

Breusch-Pagan test –
Null hypothesis: Variance of the unit-specific error = 0
Asymptotic test statistic: $\chi^2(1) = 772.645$
with p-value = 4.78359e-170

Hausman test –
Null hypothesis: GLS estimates are consistent
Asymptotic test statistic: $\chi^2(5) = 16.3469$
with p-value = 0.00592049

b)

In a pooled OLS regression all data is treated as if there was a single index i.e no differences among the individual cross section units. The coefficients produce one regression and indicate that all independent variables expect for marital status are statistically significant. The coefficient for income is positive, as expected and indicates that controlling for all other variables a one-unit increase (dollars) in total income will increase charitable donations by \$1.04. Under the same assumption, donations per year increase by \$.49 per unit change in tax rate. Being over 65 (Age) increases charitable donations by \$1.55 compared to those under the age if 65. A surprising result is that the dependent variable (deps) indicates that having an additional dependent increase donations per year by \$.17 hold all else constant.

Now using a fixed effects regression, the intercepts are able to vary among the cross section units. The slopes however are the same across all cross section units. Fixed effects also allow the model to account for factors that vary among the cross sectional units but do not vary over time. The coefficients of age (65+), Marital status and dependents are no longer statistically different from zero. The coefficients of income and tax rate remain highly statistically significant. All else equal a one unit increase in total income (\$) will increase charitable contributions by \$.84 on average. A one unit increase in the tax rate will increase charitable donations by \$.37 per year on average. As mentioned above the remaining independent variables are insignificant.

Random effects regression treats the intercepts as random variables which have probability distribution. A second difference is that this model assumes the covariance between the error term and the independent variables is 0 but allows the inclusion of the variables eliminated by the fixed effects regression. Coefficients in the random effects model are very similar to those in the fixed effects. All else equal income and price now increase charitable donations by \$.85 and \$.37 respectively. This is on average and with respect to a one unit increase in the total income or tax rate variable. All other variables (age, marriage, dependents) are insignificant or have no relationship to charitable donations per year.

c)

As displayed in part a) the quasi demeaned theta is equal to 0.775831. Since random effects lies somewhere between the classical pooling (all variables included) and fixed effects (variables that don't vary over time excluded). The model is much closer to a fixed effects model (closer to the value of 1). If it were a lower value, then theta would indicate the classical pooling model is a better fit.

d)

The purpose of conducting an F-test is to determine is to determine if a pooled or fixed effects regression is a better decision. This is accomplished by using an unrestricted regression incorporating the dummy variables for the relevant fixed effects (varying intercepts). This is compared a restricted model or the classical pooling model (model 1). The null hypothesis is for the pooled regression meaning the dummy variables producing the varying intercepts are not statistically significant as a group. The alternative hypothesis indicates a fixed effects regression is better since as a group the intercepts are significant. The F-statistic is calculated the standard way and generates:

$$F(5, 464) = 26.86280 > 1.14$$

F(5, 464) critical value is equal to 1.14 using a 5% level of significance.

Since the critical value is greater than the critical value, we reject the null hypothesis. This result would indicate the fixed effects regression is preferred for reasons mentioned above.

e)

The purpose of performing of Breusch-Pagan test is to determine is a pooled or random effects model is better suited. The null hypothesis would indicate a classical pooled model. This is because the BP test is determining if there is variation among the cross-section units. If this is the case, they can be treated as if there was a single index and hence a classical pooled model. Calculated using Gretl the Breusch-Pagan statistic is 772.645. This large value corresponds with a p-value of approximately 0 indicating that the hypothesis is to be rejected. This means that the preferred model is the random effects model.

f)

Lastly a Hausman test is conducted to determine if a Fixed or Random effects model is preferred. This is done by determining the nature of the time-invariant variables. If they are unique to the individual and should not be correlated with other individual characteristics meaning. This test determines whether the error terms are correlated and if so, the above assumption is violated, and random effects should be used. The Hausman Test statistic found is 19.24 and has a corresponding p-value = 0.00173. This is less than the %5 level of significance (alpha) thus we reject the null hypothesis (fixed effects) in favor of the alternative (random effects). This means that the error terms are correlated and a better model would be the random effects.

g)

According to the above tests the most appropriate model to use is the random effects model. The F-test indicates that fixed effects regression is preferred over the pool but the Breusch Pagan and Hausman tests both indicate Random Effects. Without further testing I would conclude that Random effects is the preferred model for estimation.

h)

In a fixed-effects model, subjects serve as their own controls. The justification for this is that the effects that the omitted variables estimate for the subject at one point in time will also have the same effect at a later time period; thus, their effects will be constant, or fixed. Age and MS are both tim-variant values and presumably have time-varying effects. Therefore, Age is included because an individual's age varies over time, at the speed of time, and presumably influences the charity contributions they will make. Intuitively, one matures over their life and accumulates experience that will influence their donation likelihood, as well as magnitude. Marital Status is also included since one's marital status might adjust from one time period to the next, thus the effects will not be "fixed". Also, one's spouse might have influence over one's charitable donations for reasons similar to the effects of age in the model.

i)

The existence of unit roots is, at first glance, unlikely. This is due to the relative dimensions of the panel; there exists a relatively small T dimension (years, in this case) compared to N (the individual). This can be tested for the heterogeneous and homogeneous cases to determine if there exists a unit root for some individuals, or for all individuals. In either case, we find this initial speculation to be true by obtaining p-values $< 2.2e-16$ for the LLN and IPS tests used individually on each of the four continuous variables. These tested the homogeneous and heterogeneous cases, respectively. It's unsurprising that this is the case when the alternative hypothesis is trend-stationarity; income is known to increase over time due to gains in experience and natural inflation. It is worth noting that when the exogenous intercept variable is used in place of trend, the results remain the same. In conclusion, there is evidence that a unit root does not exist in the data.

Panel cointegration is of little worry here as well, the time dimension of the panel is only 10 compared to the N dimension of 47. This can be confirmed via the Pedroni test (1999) via incorporation of trend as an independent variable.

Again, being a micro panel, there is little worry for autocorrelation. This is confirmed via a Breusch-Godfrey/Wooldridge test for serial correlation. The result is a p-value of $5.43e-11$ and $1.141e-09$ against the null hypothesis of serial correlation in error terms for the fixed and random effects models, respectively.

j)

Including a lagged-Charity variable would is a process known as first differencing, used to remove unobserved heterogeneity also underlies the family of estimators that have been developed for dynamic panel data models. These models contain one or more lagged dependent variables, allowing for the modeling of a partial adjustment mechanism. However, some issues from bias-implications may result from running fixed and random effect estimators once including lagged variables, as discussed in the following question.

k)

Fixed effects and random effects estimators would be inappropriate for estimating the parameters of the modified regression in Part (j) due to a resulting Nickel bias, as Nickell shows (Nickell, 1981). This bias arises due to the demeaning process, whereby the individuals' mean Charitable donation as well as the mean of the independent variables is subtracted from each respective observable characteristic/variable, creating a correlation between the regressors and error; the demeaning operation yields regressors that cannot be distributed independently of the error term.

l)

The following GMM coefficients were estimated via the Blundell-Bond system.

Variable	Coefficient
income	0.298503
price	0.204637
age	-0.275822
ms	0.300632
deps	-0.060764
3	0.129998
4	0.352498
5	0.381998
6	0.428491
7	0.585987
8	0.474582
9	0.428373
10	0.305329

m)

It is known that Generalized Method of Moments (GMM) implement more instruments when the time dimension of the panel is increased. With a large time dimension, it may eventually become impossible to apply GMM. Due to this particular panel only having a relatively low time dimension of 10, it is not a problem. If the time dimension was large, applying any of the three GMMs would likely fail.

n)

The distinction between Generalized Method of Moments (GMM) and Method of Moments (MOM) lies under the specification in which each method should be used. The MOM estimation technique is used when the number of moment conditions is equal to the number of parameters in the model. Suppose there are K estimator parameters in a given model, resulting in K sample moments being computed for the dependent variable Y, followed by setting them equal to the population moments derived from the assumed probability distribution. However, when the number of parameters exceeds the number of parameters exceeds the number of observations, which is often the case in econometric modelling, the Generalized Method of Moments is used. Under GMM, the parameter vector is estimated by minimizing the sum of the squares differences between the population and sample moments, whereby the variances of the moments are used as a metric. This would be the minimum variance estimator.

References

[1] Nickell, S. (1981). *"Biases in Dynamic Models with Fixed Effects"*. Econometrica. 49(6) 1417-26.