Panel Analysis

Lucky

Data is collected in 2001 from the transactions of 754 Mexican sex workers. There is information on four transactions per worker. Labels ID and TRANS are used to describe a particular woman and a particular transaction.

There are three categories of variables.

1.Sex worker characteristics: AGE, ATTRACTIVE = 1 if the worker is attractive, and SCHOOL if she has completed secondary school or higher.

2.Client characteristics: REGULAR = 1 if the client is a regular, RICH = 1 if the client is rich, and ALCOHOL if the client has consumed alcohol before the transaction.

3.Transaction characteristics: log of the price of the transaction LNPRICE, NOCONDOM = 1 if a condom was not used, and BAR = 1 if the transaction originated in a bar and STREET if the transaction originated in the street.

```
# Load libraries
library(haven)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(plm)
```

```
##
## Attaching package: 'plm'

## The following objects are masked from 'package:dplyr':
##
##     between, lag, lead
```

```
library(broom)
# Load data set
mex <- read_dta("mexican.dta")
# Convert data into panel data
mex_p <- pdata.frame(mex,index=c("id","trans"))
```

a) Estimate a fixed effects model with LNPRICE as the dependent variable and client characteristics as explanatory variables, and remaining are transaction characteristics.

```
# Fixed effects model
model1 <- plm(lnprice~ regular + rich + alcohol + nocondom + bar + street,
              model="within",data=mex_p)
summary(model1)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = lnprice ~ regular + rich + alcohol + nocondom +
##     bar + street, data = mex_p, model = "within")
##
## Balanced Panel: n = 754, T = 4, N = 3016
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -2.844848 -0.038538  0.000000  0.020045  1.664766
##
## Coefficients:
##           Estimate Std. Error t-value  Pr(>|t|)
## regular   0.037219   0.016849  2.2090 0.0272770 *
## rich      0.082636   0.020528  4.0254 5.875e-05 ***
## alcohol  -0.056856   0.026139 -2.1751 0.0297261 *
## nocondom  0.170282   0.025817  6.5957 5.256e-11 ***
## bar       0.298455   0.134450  2.2198 0.0265299 *
## street    0.455159   0.130465  3.4887 0.0004946 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    111.72
## Residual Sum of Squares: 107.6
## R-Squared:      0.03688
## Adj. R-Squared: -0.28715
## F-statistic: 14.3978 on 6 and 2256 DF, p-value: 3.5477e-16
```

```
tidy(model1)
```

```
## # A tibble: 6 x 5
##   term     estimate std.error statistic  p.value
##   <chr>       <dbl>     <dbl>     <dbl>    <dbl>
## 1 regular    0.0372    0.0168      2.21 2.73e- 2
## 2 rich       0.0826    0.0205      4.03 5.87e- 5
## 3 alcohol   -0.0569    0.0261     -2.18 2.97e- 2
## 4 nocondom   0.170     0.0258      6.60 5.26e-11
## 5 bar        0.298     0.134       2.22 2.65e- 2
## 6 street     0.455     0.130       3.49 4.95e- 4
```

Why did we omit the sex worker characteristics?

They are omitted as they are time-invariant over time in which the 4 transactions took place. Their effect can't be separated from individual effects given by coefficients of the fixed-effects dummy variables.

What coefficient estimates are significantly different from zero at a 5% level of significance?

All coefficient estimates are significantly different from zero at a 5% level.

Some people argue that coefficient of NOCONDOM is a risk premium, some workers are willing to take risk because of extra price some clients are willing to pay. What is your estimate of the risk premium?

Estimated risk premium for not using a condom is approximately 17%.

How is the price affected when clients are rich, are regular, and have consumed alcohol?

Price is 3.7% higher for regular and 8.3% higher for rich customers. Price is 5.7% lower for customers who have consumed alcohol

How does the location of the transaction influence the price?

Origin of transaction has a relatively large effect on the price. Transactions originated in a bar is a 29.8% premium while transactions in the street the premium is 45.5%.

b) Estimate the model assuming random effects with the characteristics of workers added to the model.

```
# Random effects model
model2 <- plm(lnprice~ regular + rich + alcohol + nocondom + bar + street + age + attractive + school,
              model="random",data=mex_p)
summary(model2)
```

```
## Oneway (individual) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = lnprice ~ regular + rich + alcohol + nocondom +
##     bar + street + age + attractive + school, data = mex_p, model = "random")
##
## Balanced Panel: n = 754, T = 4, N = 3016
##
## Effects:
##                   var std.dev share
## idiosyncratic 0.04769 0.21839  0.14
## individual    0.29337 0.54163  0.86
## theta: 0.8024
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -3.042579 -0.102652 -0.010356  0.094877  1.782238
##
## Coefficients:
##               Estimate Std. Error z-value  Pr(>|z|)
```

```
## (Intercept)  5.9103651  0.1303194 45.3529 < 2.2e-16 ***
## regular      0.0236290  0.0161849  1.4599    0.1443
## rich         0.1160067  0.0200346  5.7903 7.026e-09 ***
## alcohol      0.0148896  0.0249556  0.5966    0.5507
## nocondom     0.1389842  0.0250266  5.5535 2.801e-08 ***
## bar          0.4642454  0.0998912  4.6475 3.360e-06 ***
## street       0.1032864  0.1010769  1.0219    0.3068
## age         -0.0257651  0.0027534 -9.3574 < 2.2e-16 ***
## attractive   0.2768274  0.0602379  4.5956 4.316e-06 ***
## school       0.2161494  0.0453396  4.7673 1.867e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    167.65
## Residual Sum of Squares: 149.03
## R-Squared:      0.11104
## Adj. R-Squared: 0.10838
## Chisq: 375.488 on 9 DF, p-value: < 2.22e-16
```

```
tidy(model2)
```

```
## # A tibble: 10 x 5
##    term        estimate std.error statistic  p.value
##    <chr>          <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)   5.91     0.130      45.4   0
##  2 regular       0.0236   0.0162      1.46  1.44e- 1
##  3 rich          0.116    0.0200      5.79  7.03e- 9
##  4 alcohol       0.0149   0.0250      0.597 5.51e- 1
##  5 nocondom      0.139    0.0250      5.55  2.80e- 8
##  6 bar           0.464    0.0999      4.65  3.36e- 6
##  7 street        0.103    0.101       1.02  3.07e- 1
##  8 age          -0.0258   0.00275    -9.36  8.17e-21
##  9 attractive    0.277    0.0602      4.60  4.32e- 6
## 10 school        0.216    0.0453      4.77  1.87e- 6
```

How have the coefficients of the common variables changed?

Rich clients are now to pay 11.6% instead of 8.3%. Clients have consumed alcohol are now to pay a higher price instead of a lower price, although this coefficient is not significantly different from zero.

How do the sex worker characteristics affect the price of commercial sex?

Premium for not using a condom has declined to 13.9%. There have been large changes in the coefficients of BAR and STREET. Random effects suggests that transactions in a bar are much more expensive than on the street. Price of commercial sex is lower for older sex workers, higher for attractive workers, and higher for secondary educated sex workers.

How much extra does a client have to pay to have unprotected sex with an attractive secondary educated sex worker?

Compared with protected sex with an unattractive uneducated sex worker, extra percentage premium is 63.2%

c) Using t-test and a 5% significance level to test whether there are any significant differences between the fixed effects and random effects on estimates of coefficients on NOCONDOM, RICH, REGULAR, ALCOHOL, BAR, and STREET.

```
model_fe <- plm(lnprice~ regular + rich + alcohol + nocondom + bar + street,model="within",data=mex_p)
summary(model_fe)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = lnprice ~ regular + rich + alcohol + nocondom +
##     bar + street, data = mex_p, model = "within")
##
## Balanced Panel: n = 754, T = 4, N = 3016
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -2.844848 -0.038538  0.000000  0.020045  1.664766
##
## Coefficients:
##           Estimate Std. Error t-value  Pr(>|t|)
## regular   0.037219   0.016849  2.2090 0.0272770 *
## rich      0.082636   0.020528  4.0254 5.875e-05 ***
## alcohol  -0.056856   0.026139 -2.1751 0.0297261 *
## nocondom  0.170282   0.025817  6.5957 5.256e-11 ***
## bar       0.298455   0.134450  2.2198 0.0265299 *
## street    0.455159   0.130465  3.4887 0.0004946 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    111.72
## Residual Sum of Squares: 107.6
## R-Squared:      0.03688
## Adj. R-Squared: -0.28715
## F-statistic: 14.3978 on 6 and 2256 DF, p-value: 3.5477e-16
```

```
model_re <- plm(lnprice~ regular + rich + alcohol + nocondom + bar + street,model="random",data=mex_p)
summary(model_re)
```

```
## Oneway (individual) effect Random Effect Model
##    (Swamy-Arora's transformation)
##
## Call:
## plm(formula = lnprice ~ regular + rich + alcohol + nocondom +
##     bar + street, data = mex_p, model = "random")
##
## Balanced Panel: n = 754, T = 4, N = 3016
##
## Effects:
##                    var std.dev share
## idiosyncratic 0.04769 0.21839 0.128
## individual    0.32396 0.56918 0.872
## theta: 0.8116
```

```
##
## Residuals:
##      Min.   1st Qu.    Median   3rd Qu.      Max.
## -3.016131 -0.102352 -0.011579  0.087671  1.723286
##
## Coefficients:
##             Estimate Std. Error z-value  Pr(>|z|)
## (Intercept) 5.223965   0.102926 50.7545 < 2.2e-16 ***
## regular     0.011973   0.016399  0.7301    0.4653
## rich        0.126309   0.020299  6.2224 4.896e-10 ***
## alcohol     0.025377   0.025312  1.0025    0.3161
## nocondom    0.128353   0.025381  5.0570 4.259e-07 ***
## bar         0.565921   0.102605  5.5155 3.478e-08 ***
## street      0.073388   0.104054  0.7053    0.4806
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    162.55
## Residual Sum of Squares: 152.43
## R-Squared:      0.062259
## Adj. R-Squared: 0.060389
## Chisq: 199.775 on 6 DF, p-value: < 2.22e-16
```

```
# Hausman Test
phtest(model_fe,model_re)
```

```
##
##  Hausman Test
##
## data:  lnprice ~ regular + rich + alcohol + nocondom + bar + street
## chisq = 337.92, df = 6, p-value < 2.2e-16
## alternative hypothesis: one model is inconsistent
```

Should we rely on the fixed effects estimates or on the random effects estimates? Explain your choice.

At a 5% level of significance, there's a significant difference between all coefficients except for BAR. Thus, we reject a null hypothesis that the individual random effects are uncorrelated with the variables in the model. Fixed effects estimates are more reliable in this instance because they are consistent.
