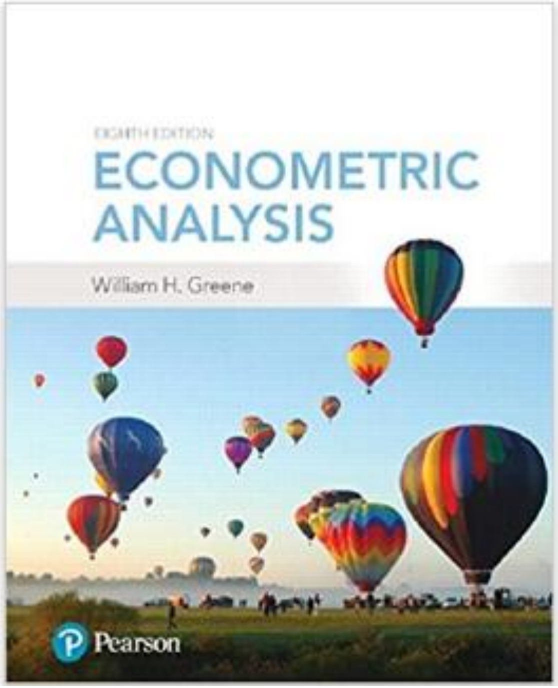
Econometrics I

Professor William Greene

Stern School of Business

Department of Economics



Econometrics I

Part 4 – Partial Regression and Partial Correlation

Frisch-Waugh (1933) 'Theorem'

Context: Model contains two sets of variables:

$\mathbf{X}$ = [(1,time) : (other variables)]

= $[\mathbf{X}_1 \;\; \mathbf{X}_2]$

Regression model:

$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon$ (population)

$= \mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_2\mathbf{b}_2 + \mathbf{e}$ (sample)

Problem: Algebraic expression for the second set of least squares coefficients, $\mathbf{b}_2$

Partitioned Solution

Method of solution (Why did F&W care? In 1933, matrix computation was not trivial!)

Direct manipulation of normal equations produces

$$(\mathbf{X}'\mathbf{X})\mathbf{b} = \mathbf{X}'\mathbf{y}$$

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2] \text{ so } \mathbf{X}'\mathbf{X} = \begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{bmatrix} \text{ and } \mathbf{X}'\mathbf{y} = \begin{bmatrix} \mathbf{X}_1'\mathbf{y} \\ \mathbf{X}_2'\mathbf{y} \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X})\mathbf{b} = \begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{bmatrix} \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} = \begin{bmatrix} \mathbf{X}_1'\mathbf{y} \\ \mathbf{X}_2'\mathbf{y} \end{bmatrix}$$

$$\mathbf{X}_1'\mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_1'\mathbf{X}_2\mathbf{b}_2 = \mathbf{X}_1'\mathbf{y}$$

$$\mathbf{X}_2'\mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_2'\mathbf{X}_2\mathbf{b}_2 = \mathbf{X}_2'\mathbf{y} \Longrightarrow \mathbf{X}_2'\mathbf{X}_2\mathbf{b}_2 = \mathbf{X}_2'\mathbf{y} - \mathbf{X}_2'\mathbf{X}_1\mathbf{b}_1$$

$$= \mathbf{X}_2'(\mathbf{y} - \mathbf{X}_1\mathbf{b}_1)$$

Partitioned Solution

Direct manipulation of normal equations produces

$$\mathbf{b}_2 = (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'(\mathbf{y} - \mathbf{X}_1\mathbf{b}_1)$$

What is this? Regression of $(\mathbf{y} - \mathbf{X}_1\mathbf{b}_1)$ on $\mathbf{X}_2$

If we knew $\mathbf{b}_1$, this is the solution for $\mathbf{b}_2$.

Important result (perhaps not fundamental). Note the result if $\mathbf{X}_2'\mathbf{X}_1 = \mathbf{0}$.

Useful in theory: Probably

Likely in practice? Not at all.

Partitioned Inverse

Use of the partitioned inverse result produces a fundamental result: What is the southeast element in the inverse of the moment matrix?

$$\begin{bmatrix} \mathbf{X_1'X_1} & \mathbf{X_1'X_2} \\ \mathbf{X_2'X_1} & \mathbf{X_2'X_2} \end{bmatrix}^{-1}$$

Partitioned Inverse

The algebraic result is:

$$[\;]^{-1}_{(2,2)} = \{[\mathbf{X}_2'\mathbf{X}_2] - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\,\mathbf{X}_1'\mathbf{X}_2\}^{-1}$$

$$= [\mathbf{X}_2'(\mathbf{I} - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1')\mathbf{X}_2]^{-1}$$

$$= [\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2]^{-1}$$

- Note the appearance of an "**M**" matrix. How do we interpret this result?
- Note the implication for the case in which $\mathbf{X}_1$ is a single variable. (Theorem, p. 37)
- Note the implication for the case in which $\mathbf{X}_1$ is the constant term. (p. 38)

Frisch-Waugh (1933) Basic Result
Lovell (JASA, 1963) did the matrix algebra.

Continuing the algebraic manipulation:

$$\mathbf{b}_2 = [\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2]^{-1}[\mathbf{X}_2'\mathbf{M}_1\mathbf{y}].$$

This is Frisch and Waugh's famous result - the "double residual regression."

How do we interpret this? A regression of residuals on residuals.

"We get the same result whether we (1) detrend the other variables by using the residuals from a regression of them on a constant and a time trend and use the detrended data in the regression or (2) just include a constant and a time trend in the regression and not detrend the data"

"Detrend the data" means compute the residuals from the regressions of the variables on a constant and a time trend.

Important Implications

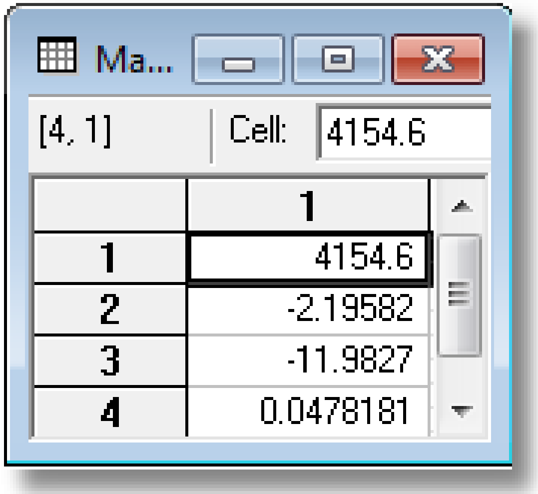
- Isolating a single coefficient in a regression. (Corollary 3.2.1, p. 37). The double residual regression.
- Regression of residuals on residuals – 'partialling' out the effect of the other variables.
- It is not necessary to 'partial' the other **X**s out of **y** because $\mathbf{M}_1$ is idempotent. (This is a very useful result.) (i.e., $\mathbf{X}_2'\mathbf{M}_1'\mathbf{M}_1\mathbf{y} = \mathbf{X}_2'\mathbf{M}_1\mathbf{y}$)
- (Orthogonal regression) Suppose $\mathbf{X}_1$ and $\mathbf{X}_2$ are orthogonal; $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$. What is $\mathbf{M}_1\mathbf{X}_2$?

Applying Frisch-Waugh

Using gasoline data from Notes 3.

X = [**1, year**, **PG, Y**], **y** = *G* as before.

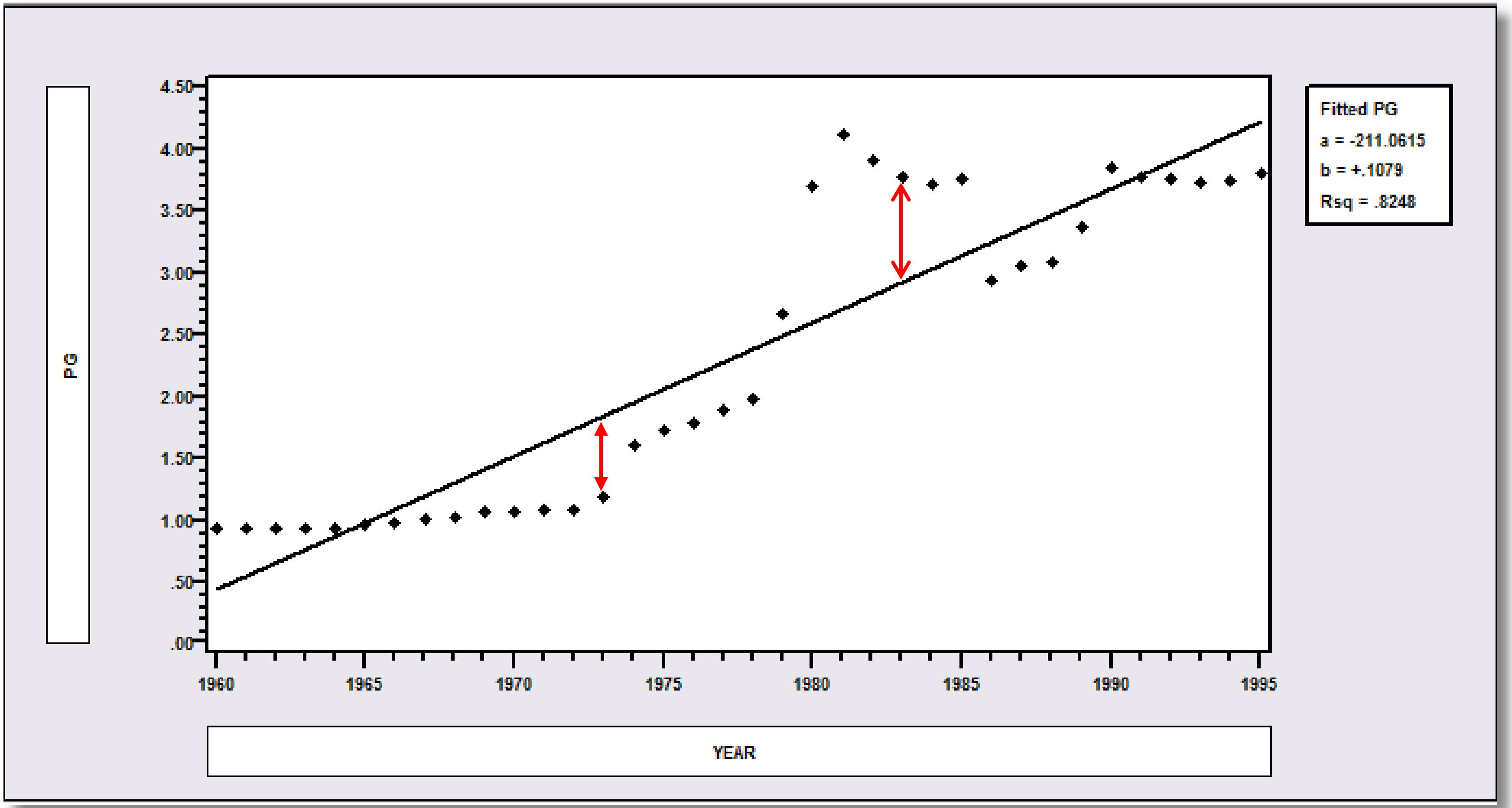
Full least squares regression of **y** on **X.**



Partitioned regression strategy:

1. Regress PG and Y on (1,Year) (detrend them) and compute residuals PG* and Y*
2. Regress G on (1,Year) and compute residuals G*. (This step is not actually necessary.)
3. Regress G* on PG* and Y*. (Should produce -11.9827 and 0.0478181.

Detrending the Variables – Pg
Pg* are the residuals from this regression



Regression of detrended G on detrended Pg and detrended Y

```
namelist;x1=one,year$
namelist;x2=bg,y$
regress;lhs=pg;rhs=x1;res=pgstar;quiet$
regress;lhs=y ;rhs=x1;res=ystar ;quiet$
regress;lhs=g ;rhs=x1;res=gstar ;quiet$
regress;lhs=gstar;rhs=pgstar,ystar$
-----------------------------------------------------------------------
Ordinary    least squares regression ............
LHS=GSTAR   Mean                   =        .00000
            Standard deviation     =      16.40826
            Number of observs.     =            36
Model size  Parameters             =             2
            Degrees of freedom     =            34
Residuals   Sum of squares         =       1244.71
            Standard error of e    =       6.05056
Fit         R-squared              =        .86791
            Adjusted R-squared     =        .86402
Model test  F[  1,    34] (prob) =   223.4(.0000)
Not using OLS or no constant. Rsqrd & F may be < 0
Model was estimated on Jul 19, 2012 at 10:01:40 PM
--------+--------------------------------------------------------------
        |                 Standard            Prob.      95% Confidence
   GSTAR|  Coefficient      Error       t    |t|>T*         Interval
--------+--------------------------------------------------------------
  PGSTAR|   -11.9827***   2.11719    -5.66  .0000    -16.1323   -7.8330
   YSTAR|     .04782***    .00453    10.56  .0000      .03895    .05669
--------+--------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
-----------------------------------------------------------------------
```

Ma...

[4, 1] Cell: 4154.6

	1
1	4154.6
2	-2.19582
3	-11.9827
4	0.0478181

Partial Regression

Important terms in this context:

Partialing out the effect of $\mathbf{X}_1$.

Netting out the effect …

"**Partial regression coefficients**."

To continue belaboring the point: Note the interpretation of partial regression as "net of the effect of …"

This is the (*very powerful*) Frisch – Waugh Theorem. This is what is meant by "controlling for the effect of $\mathbf{X}_1$."

Now, follow this through for the case in which $\mathbf{X}_1$ is just a constant term, column of ones.

What are the residuals in a regression on a constant. What is $\mathbf{M}_1$?

Note that this produces the result that we can do linear regression on data in mean deviation form.

'Partial regression coefficients' are the same as 'multiple regression coefficients.' It follows from the Frisch-Waugh theorem.

Partial Correlation

Working definition. Correlation between sets of residuals.

Some results on computation: Based on the **M** matrices.

Some important considerations:

Partial correlations and coefficients can have signs and magnitudes that differ greatly from gross correlations and simple regression coefficients.

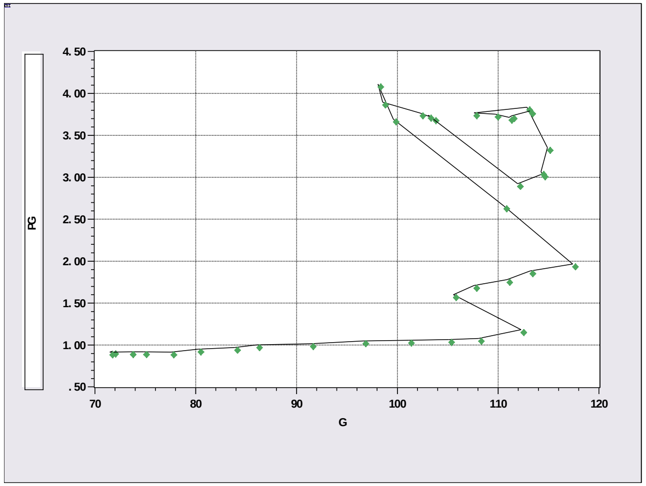
Compare the simple (gross) correlation of G and PG with the partial correlation, net of the time effect. (**Could you have predicted the negative partial correlation?**)

CALC;list;Cor(g,pg)$

Result = .7696572

CALC;list;cor(gstar,pgstar)$

Result = -.6589938



Partial Correlation

```
namelist;x1=one,year$
regress;lhs=pg;rhs=x1;res=pgstar;quiet$
regress;lhs=y ;rhs=x1;res=ystar ;quiet$
regress;lhs=g ;rhs=x1;res=gstar ;quiet$

calc ; list ; cor(gstar,pgstar)$

|-> calc ; list ; cor(gstar,pgstar)$
[CALC]          =          -.6589938
```

<u>THE</u> Most Famous Application of Frisch-Waugh: The Fixed Effects Model

A regression model with a dummy variable for each individual in the sample, each observed T_i times. There is a constant term for each individual.

$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{d}_i\alpha_i + \boldsymbol{\varepsilon}_i$, for each individual

N columns (spanning the $\mathbf{d}$ columns)

$$\begin{pmatrix}\mathbf{y}_1\\ \mathbf{y}_2\\ \vdots\\ \mathbf{y}_N\end{pmatrix} = \begin{bmatrix}\mathbf{X}_1 & \mathbf{d}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{X}_2 & \mathbf{0} & \mathbf{d}_2 & \mathbf{0} & \mathbf{0}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \mathbf{X}_N & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{d}_N\end{bmatrix}\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\alpha}\end{pmatrix} + \boldsymbol{\varepsilon}$$

N may be thousands. I.e., the regression has thousands of variables (coefficients).

$$= [\mathbf{X},\mathbf{D}]\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\alpha}\end{pmatrix} + \boldsymbol{\varepsilon}$$

$$= \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon}$$

Application – Health and Income

German Health Care Usage Data, 7,293 Individuals,
Varying Numbers of Periods
Variables in the file are

Data downloaded from *Journal of Applied Econometrics* Archive. This is an unbalanced panel with N = 7,293 individuals. There are altogether n = 27,326 observations. The number of observations ranges from 1 to 7 per family.
(Frequencies are: 1=1525, 2=2158, 3=825, 4=926, 5=1051, 6=1000, 7=987).
The dependent variable of interest is

DOCVIS = number of visits to the doctor in the observation period

HHNINC = household nominal monthly net income in German marks / 10000.
(4 observations with income=0 were dropped)
HHKIDS = children under age 16 in the household = 1; otherwise = 0
EDUC = years of schooling
AGE = age in years

We desire also to include a separate family effect (7293 of them) for each family. This requires 7293 dummy variables in addition to the four regressors.

Estimating the Fixed Effects Model

The FE model is a linear regression model with <u>*many*</u> independent variables

$$\begin{pmatrix} \mathbf{b} \\ \mathbf{a} \end{pmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{D} \\ \mathbf{D}'\mathbf{X} & \mathbf{D}'\mathbf{D} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{D}'\mathbf{y} \end{bmatrix}$$

Using the Frisch-Waugh theorem

$$\mathbf{b} \quad = [\mathbf{X}'\mathbf{M}_{\mathbf{D}}\mathbf{X}]^{-1} \left[\mathbf{X}'\mathbf{M}_{\mathbf{D}}\mathbf{y}\right]$$

Fixed Effects Estimator (cont.)

$$\mathbf{M_D} = \mathbf{I} - \mathbf{D}(\mathbf{D'D})^{-1}\mathbf{D'}$$

$$\mathbf{D} = \begin{bmatrix} \mathbf{d_1} & \mathbf{0} & & \mathbf{0} \\ \mathbf{0} & \mathbf{d_2} & & \mathbf{0} \\ & & \cdots & \\ \mathbf{0} & \mathbf{0} & & \mathbf{d_N} \end{bmatrix}$$

$$\mathbf{DD'} = \begin{bmatrix} \mathbf{d_1 d_1'} & \mathbf{0} & & \mathbf{0} \\ \mathbf{0} & \mathbf{d_2 d_2'} & & \mathbf{0} \\ & & \cdots & \\ \mathbf{0} & \mathbf{0} & & \mathbf{d_N d_N'} \end{bmatrix}$$

Fixed Effects Estimator (cont.)

$$\mathbf{M_D} = \mathbf{I} - \mathbf{D}(\mathbf{D'D})^{-1}\mathbf{D'}$$

$$\mathbf{M_D} = \begin{bmatrix} \mathbf{M_D^1} & \mathbf{0} & & \mathbf{0} \\ \mathbf{0} & \mathbf{M_D^2} & & \mathbf{0} \\ & & & \\ \mathbf{0} & \mathbf{0} & & \mathbf{M_D^N} \end{bmatrix}$$ (The dummy variables are orthogonal)

$$\mathbf{M_D^i} = \mathbf{I_{T_i}} - \mathbf{d_i}(\mathbf{d_i'd_i})^{-1}\mathbf{d'} = \mathbf{I_{T_i}} - (1/T_i)\mathbf{d_i d'}$$

$$= \begin{bmatrix} 1 - \frac{1}{T_i} & - \frac{1}{T_i} & \ldots & - \frac{1}{T_i} \\ - \frac{1}{T_i} & 1 - \frac{1}{T_i} & \ldots & - \frac{1}{T_i} \\ \ldots & \ldots & \ldots & \ldots \\ - \frac{1}{T_i} & - \frac{1}{T_i} & \ldots & 1 - \frac{1}{T_i} \end{bmatrix}$$

'Within' Transformations

$$\mathbf{M_D^i X}_i = \left(\mathbf{I_{T_i}} - \mathbf{d_i}(\mathbf{d_i'd_i})^{-1}\mathbf{d}'\right)\mathbf{X}_i = \left(\mathbf{I_{T_i}} - (1/T_i)\mathbf{d_i d}'\right)\mathbf{X}_i$$

$$= \mathbf{X}_i - (1/T_i)\mathbf{d_i}\left(\mathbf{d}'\mathbf{X}_i\right) = \mathbf{X}_i - (1/T_i)\mathbf{d_i}(T_i\bar{\mathbf{x}}_i') = \mathbf{X}_i - \mathbf{d}_i\bar{\mathbf{x}}_i'$$

$$(T_i \times K) - (T_i \times 1) * (1 \times K)$$

$$\mathbf{X}'\mathbf{M_D X} = \Sigma_{i=1}^{N}\mathbf{X}_i'\mathbf{M_D^i X}_i,$$

$$\left\{\mathbf{X}_i'\mathbf{M_D^i X}_i\right\}_{k,l} = \Sigma_{t=1}^{T_i}(x_{it,k} - \bar{x}_{i.,k})(x_{it,l} - \bar{x}_{i.,l}) \quad \text{(k,l element)}$$

$$\mathbf{X}'\mathbf{M_D y} = \Sigma_{i=1}^{N}\mathbf{X}_i'\mathbf{M_D^i y}_i,$$

$$\left\{\mathbf{X}_i'\mathbf{M_D^i y}_i\right\}_{k} = \Sigma_{t=1}^{T_i}(x_{it,k} - \bar{x}_{i.,k})(y_{it} - \bar{y}_{i.})$$

Least Squares Dummy Variable Estimator

- **b** is obtained by 'within' groups least squares (group mean deviations)
- Normal equations for **a** are **D'Xb**+**D'Da**=**D'y**

a = (D'D)$^{-1}$D'(y – Xb)   (see slide 5)

$$a_i = (1/T_i)\Sigma_{t=1}^{T_i}(y_{it} - \mathbf{x}_{it}'\mathbf{b}) = \bar{e}_i$$

A Fixed Effects Regression

```
----------------------------------------------------------------------
LSDV          least squares with fixed effects ....
LHS=DOCVIS    Mean                  =     3.18352
              Standard deviation    =     5.68969
----------    No. of observations   =       27326  DegFreedom    Mean square
Regression    Sum of Squares        =     450364.       7296       61.72755
Residual      Sum of Squares        =     434216.      20029       21.67939
Total         Sum of Squares        =     884581.      27325       32.37258
----------    Standard error of e   =     4.65611  Root MSE          3.98626
Fit           R-squared             =      .50913  R-bar squared      .33032
Model test    F[***, 20029]         =     2.84729  Prob F > F*        .00000
Estd. Autocorrelation of e(i,t)     =    -.220319
--------------------------------------------------
Panel:Groups Empty        0,      Valid data     7293
             Smallest     1,      Largest           7
             Average group size in panel         3.75
Variances    Effects a(i)        Residuals e(i,t)
                 19.776379              21.679385
Rho squared: Residual variation due to ai .477048
Within groups variation in DOCVIS       435764.4262
R squared based on within group variation  .003552
--------+-------------------------------------------------------------
        |                Standard            Prob.      95% Confidence
  DOCVIS|  Coefficient     Error        z    |z|>Z*        Interval
--------+-------------------------------------------------------------
  HHNINC|   -.94188***     .30603    -3.08  .0021    -1.54168   -.34207
  HHKIDS|   -.02274        .12225     -.19  .8525     -.26235    .21687
    EDUC|   -.10269        .11552     -.89  .3740     -.32909    .12372
     AGE|    .09984***     .01206     8.28  .0000      .07620    .12347
--------+-------------------------------------------------------------
Note: ***, **, * ==>  Significance at 1%, 5%, 10% level.
----------------------------------------------------------------------
```

Constant terms

[7293, 1] Cell: -3.27015

	1
1	-3.27015
2	-2.33104
3	-1.48175
4	3.07553
5	-0.560508
6	3.22613
7	-0.832392
8	1.82994
9	1.96091
10	-0.832094
11	-1.06964
12	2.32093
13	1.39914
14	-1.76044
15	0.22136
16	-2.05827
17	4.98343
18	24.3749
19	-0.355367
20	0.161649
21	-1.9608
22	-1.9459
23	-3.09324
24	-3.04032
25	-0.396102
26	-1.0917
27	-0.794878
28	-2.72324
29	2.00065
30	-0.644711
31	0.40696

What happens to a time invariant variable such as gender?

$\mathbf{f}$ = a variable that is the same in every period (FEMALE)

$\mathbf{f}'\mathbf{M}_\mathbf{D}\mathbf{f} = \Sigma_{i=1}^{n}\mathbf{f}_i\mathbf{M}_\mathbf{D}^i\mathbf{f}_i = \Sigma_{i=1}^{n}\Sigma_{t=1}^{T_i}(f_{it}-\overline{f}_{i.})^2$

but $f_{it} = f_i$, so $f_{it} = \overline{f}_{i.}$ and $(f_{it}-\overline{f}_{i.}) = 0$ for all i and t

This is a simple case of multicollinearity.

$\mathbf{f} = \text{diag}(f_i)*\mathbf{D}$

```
----------------------------------------------------
Panel:Groups Empty       0,      Valid data     7293
             Smallest    1,      Largest           7
             Average group size in panel        3.75
Variances    Effects a(i)           Residuals e(i,t)
                  23.803375                21.680476
Std.Devs.          4.878870                 4.656230
Rho squared: Residual variation due to ai    .523337
Within groups variation in DOCVIS        .43576D+06
R squared based on within group variation    .003552
Between group variation in DOCVIS        .44882D+06
****************************************************
These  1 variables have no within group variation.
FEMALE
They are  not included in the fixed effects model.
```

DOCVIS	Coefficient	Standard Error	z	Prob. |z|>Z*	95% Confidence Interval	
INCOME	-.94193***	.30618	-3.08	.0021	-1.54202	-.34183
HHKIDS	-.02281	.12226	-.19	.8520	-.26243	.21681
EDUC	-.10216	.11553	-.88	.3765	-.32859	.12427
AGE	.09982***	.01206	8.28	.0000	.07619	.12345
FEMALE	0.0	.....(Fixed Parameter).....				

```
***, **, * ==>  Significance at 1%, 5%, 10% level.
Fixed parameter ... is constrained to equal the value or
had a nonpositive st.error because of an earlier problem.
```