Problem Set 1 Solutions
EE426 Semester 2/2014

1. (Difference-in-differences)

1.1 If building the incinerator reduces the value of homes closer to the site, the sign of δ_1 should be positive, that is the longer distance of homes from the incinerator site (that was built in 1981), the higher value of the home. **>> the difference in elasticity of price of house with respect to the distance to the incinerator site between 1981 and 1978**

β_1 measures the location effect that is not due to the presence of the incinerator. (δ_1 explains the location effect that is due to the presence of the incinerator since the incinerator was built in 1981.) If $\beta_1 > 0$, on average with or without the incinerator, the prices of the house that is located far away from the supposed incinerator site increase with the distance from the supposed incinerator site.

1.2 The results are reported in column (1). We see that, with the presence of the incinerator in 1981, there is a positively relationship between the distance and the housing prices. If the house is located one percent far away from the incinerator, the housing price increases 0.048 percent in year 1981 higher than the increase in year 1978. However, the coefficient is not statistically significant even at a 10% significance level.

	(1)	(2)
	b/se	b/se
y81	-0.011	-0.225
	(0.8051)	(0.4947)
ldist	0.317***	0.001
	(0.0515)	(0.0446)
y81ldist	0.048	0.062
	(0.0818)	(0.0503)
age		-0.008***
		(0.0014)
agesq		0.000***
		(0.0000)
rooms		0.046***
		(0.0173)
baths		0.101***
		(0.0278)
lintst		-0.060*
		(0.0317)
lland		0.095***
		(0.0247)
larea		0.351***
		(0.0519)
_cons	8.058***	7.674***
	(0.5084)	(0.5016)
r2	0.396	0.787
N	321	321

```
use KIELMC.dta

reg lprice y81 ldist y81ldist
est sto reg1

reg lprice y81 ldist y81ldist age agesq rooms baths lintst lland larea
est sto reg2

estout reg1 reg2, cells(b(star fmt(3)) se(par fmt(4))) stats(r2 N, fmt(3
0)) starlevels(* 0.10 ** 0.05 *** 0.01)
```

1.3 Column (2) reports the results when adding other control variables. When we control for other factors variables, the coefficient of y81*log(dist) becomes 0.062. With other things being equal, one percent of the house location away from the incinerator increases the housing price in 1981 for 0.062 percent higher than that in 1978. Still, the coefficient is not statistically significant. However, the main explanations on the housing price are from other control variables as they are all statistically significant. Hence, model (1) seems to omit some important variables.

2. Pooled OLS vs. First Differences

2.1 Pooled OLS:

$\log(wage)_{it} = \beta_0 + \beta_1 educ_{it} + \beta_2 black_{it} + \beta_3 hisp_{it} + \beta_4 exper_{it} + \beta_5 married_{it} + \beta_6 union_{it} + \gamma_1 d81 + \gamma_2 d82 + \gamma_3 d83 + \gamma_4 d84 + \gamma_5 d85 + \gamma_6 d86 + \gamma_7 d87 + u_{it}$

```
STATA command:
reg lwage educ black hisp exper married union d81- d87
est sto pool1
```

If they are married, wage will increase by 11.1 percent on average, compared to those who are not married with other controls being equal. Additionally, if they have a membership in a union, their wage will be 18.6 percent higher than those who are not in the union, ceteris paribus. These coefficients are all statistically significant at 1% significance level.

2.2-2.3 Do the first difference:

Equation: $\Delta\log(wage)_{it} = \beta_0 + \beta_5 \Delta married_{it} + \beta_6 \Delta union_{it} + \delta_2 d82 + \delta_3 d83 + \delta_4 d84 + \delta_5 d85 + \delta_6 d86 + \delta_7 d87 + \Delta u_{it}$

```
sort nr year
by nr: gen l_lwage = lwage[_n-1]
by nr: gen l_educ = educ[_n-1]
by nr: gen l_black = black[_n-1]
by nr: gen l_hisp = hisp[_n-1]
by nr: gen l_exper = exper[_n-1]
by nr: gen l_married = married[_n-1]
by nr: gen l_union = union[_n-1]

gen d_lwage = lwage-l_lwage
gen d_educ = educ-l_educ
gen d_black = black-l_black
gen d_hisp = hisp-l_hisp
gen d_exper = exper-l_exper
gen d_married = married-l_married
gen d_union = union-l_union
```

```
reg d_lwage d_educ d_black d_hisp d_exper d_married d_union d82- d87
est sto fdiff1
note: d_educ omitted because of collinearity
note: d_black omitted because of collinearity
note: d_hisp omitted because of collinearity
note: d_exper omitted because of collinearity
```

You can see that these time-invariant variables are all dropped out of the estimation because of collinearity after doing the first difference. You can browse to see how these variables look like:

```
browse year nr d_educ d_black d_hisp d_exper
```
>> They are all constant across time.

	(1)	(2)
	b/se	b/se
educ	0.093***	
	(0.0052)	
black	-0.137***	
	(0.0236)	
hisp	0.014	
	(0.0208)	
exper	0.030***	
	(0.0055)	
married	0.111***	
	(0.0157)	
union	0.186***	
	(0.0171)	
d_married		0.040*
		(0.0229)
d_union		0.042**
		(0.0197)
d81	0.078***	
	(0.0296)	
d82	0.097***	-0.060**
	(0.0311)	(0.0269)
d83	0.107***	-0.071***
	(0.0334)	(0.0269)
d84	0.140***	-0.047*
	(0.0363)	(0.0269)
d85	0.160***	-0.067**
	(0.0398)	(0.0269)
d86	0.188***	-0.056**
	(0.0437)	(0.0269)
d87	0.211***	-0.052*
	(0.0478)	(0.0269)
_cons	0.156**	0.115***
	(0.0753)	(0.0191)
r2	0.188	0.005
N	4360	3815

For you reference, what if we do the changes in year dummy variables.

```
*using loop
forvalues y = 1(1)7 {
by nr: gen l_d8`y' = d8`y'[_n-1]
gen d_d8`y' = d8`y' - l_d8`y'
}
browse d_d8*
```

*Take a look on how it looks like when we do changes in year dummy: Δd81, Δd82,…, Δd87

When using the changes in year dummy, don't forget that we don't have a constant. (Compare to our lecture note in class on "Differencing with more than 2 periods")

```
reg d_lwage d_married d_union d_d81 d_d82 d_d83 d_d84 d_d85 d_d86 d_d87,
nocons
est sto fdiff2

estout fdiff1 fdiff2, cells(b(star fmt(3)) se(par fmt(4))) stats(r2 N,
fmt(3 0)) starlevels(* 0.10 ** 0.05 *** 0.01)
```

	fdiff1	fdiff2
	b/se	b/se
d_married	0.040*	0.040*
	(0.0229)	(0.0229)
d_union	0.042**	0.042**
	(0.0197)	(0.0197)
d82	-0.060**	
	(0.0269)	
d83	-0.071***	
	(0.0269)	
d84	-0.047*	
	(0.0269)	
d85	-0.067**	
	(0.0269)	
d86	-0.056**	
	(0.0269)	
d87	-0.052*	
	(0.0269)	
d_d81		0.115***
		(0.0191)
d_d82		0.171***
		(0.0271)
d_d83		0.215***
		(0.0334)
d_d84		0.284***
		(0.0387)
d_d85		0.333***
		(0.0432)
d_d86		0.392***
		(0.0474)
d_d87		0.455***
		(0.0512)
_cons	0.115***	
	(0.0191)	

```
--------------------------------------------
r2                    0.005           0.027
N                      3815            3815
--------------------------------------------
```

If you predict the fitted values from these two models (fdiff1 and fdiff2), you will see that the fitted values are exactly the same.

```
reg d_lwage d_married d_union d82- d87
predict y1

reg d_lwage d_married d_union d_d81 d_d82 d_d83 d_d84 d_d85 d_d86 d_d87,
nocons
predict y2

sum y1 y2
```

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
          y1 |      3815    .0675718    .0300554  -.0335739   .1975348
          y2 |      3815    .0675718    .0300554  -.0335739   .1975348
```

2.4 For the first difference, if they are married, wage will increase by 4.03 percent, compared to those who are not married with other controls being equal. Additionally, if they have a membership in a union, their wage will be 4.2 percent higher than those who are not in the union, ceteris paribus. (The interpretation on coefficients for FD is still the same as we use level. See Example 13.5 and Section 13.4)

The first difference estimation provides smaller estimates compared to the levels estimates. We may say that without taking into account individual unobserved effects (that are constant across time), Pooled OLS overestimates the premium on marriage and the effect of union membership on log(wage).

3. FE, RE, and pooled OLS

3.1-3.2 Pooled OLS vs. RE

```
*3.1 Pooled OLS
reg lwage educ black hisp
est sto ols1

*3.2 RE
xtset nr year

xtreg lwage educ black hisp, re
est sto re1

estout ols1 re1, cells(b(star fmt(3)) se(par fmt(4))) stats(r2 N, fmt(3
0)) starlevels(* 0.10 ** 0.05 *** 0.01)
```

```
--------------------------------------------
                    ols1             re1
                    b/se             b/se
--------------------------------------------
educ                0.077***         0.077***
                  (0.0046)         (0.0092)
black              -0.123***        -0.123**
                  (0.0247)         (0.0497)
hisp                0.025            0.025
                  (0.0222)         (0.0447)
_cons               0.752***         0.752***
                  (0.0554)         (0.1114)
--------------------------------------------
r2                  0.070
N                    4360             4360
--------------------------------------------
```

The RE and pooled OLS estimates of β_j are all the same.

3.3 The RE standard errors are larger than the pooled OLS standard errors. If you do BP test (xttest0), it rejects the null hypothesis that var(a_i) =0. So, it means that we have significant heterogeneous (unobserved) effects a_i, and that transformation on the data (GLS for RE) is preferred.

```
xttest0

Breusch and Pagan Lagrangian multiplier test for random effects

        lwage[nr,t] = Xb + u[nr] + e[nr,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                   lwage |   .2836728       .5326094
                       e |   .1499484       .3872317
                       u |   .1148867       .3389494

        Test:   Var(u) = 0
                                 chibar2(01) =  2843.93
                              Prob > chibar2 =   0.0000
```

3.4-3.5 Add a full set of year dummies, do pooled OLS, RE, and FE.

```
*3.4
reg lwage educ black hisp d81- d87
est sto ols2

xtreg lwage educ black hisp d81- d87, re
est sto re2

*3.5
xtreg lwage educ black hisp d81- d87, fe
est sto fe

estout ols2 re2 fe, cells(b(star fmt(3)) se(par fmt(4))) stats(r2 N, fmt(3
0)) starlevels(* 0.10 ** 0.05 *** 0.01)
```

```
------------------------------------------------------------
                    ols2            re2             fe
                    b/se            b/se            b/se
------------------------------------------------------------
educ                0.077***        0.077***
                  (0.0044)        (0.0092)
black              -0.123***       -0.123**
                  (0.0237)        (0.0497)
hisp                0.025           0.025
                  (0.0213)        (0.0447)
d81                 0.119***        0.119***        0.119***
                  (0.0299)        (0.0215)        (0.0215)
d82                 0.178***        0.178***        0.178***
                  (0.0299)        (0.0215)        (0.0215)
d83                 0.226***        0.226***        0.226***
                  (0.0299)        (0.0215)        (0.0215)
d84                 0.297***        0.297***        0.297***
                  (0.0299)        (0.0215)        (0.0215)
d85                 0.346***        0.346***        0.346***
                  (0.0299)        (0.0215)        (0.0215)
d86                 0.406***        0.406***        0.406***
                  (0.0299)        (0.0215)        (0.0215)
d87                 0.473***        0.473***        0.473***
                  (0.0299)        (0.0215)        (0.0215)
_cons               0.497***        0.497***        1.393***
                  (0.0567)        (0.1123)        (0.0152)
------------------------------------------------------------
r2                  0.145                           0.163
N                    4360            4360            4360
------------------------------------------------------------
```

- The coefficients of pooled OLS and RE on educ, black, and hisp, after adding year dummy, are still the same, but the constant is smaller as all year dummies are statistically significant.
- With FE, it drops all explanatory variables, except year dummies. The FE coefficients on the year dummies are the same as the RE coefficients on the year dummies, but the constant coefficients are largely different.

3.6

* year dummies as intercepts and interactions with educ

```
xtreg lwage d81- d87 c.educ#d81 c.educ#d82 c.educ#d83 c.educ#d84 c.educ#d85
c.educ#d86 c.educ#d87 , fe

Fixed-effects (within) regression               Number of obs      =      4360
Group variable: nr                              Number of groups   =       545

R-sq:  within  = 0.1647                         Obs per group: min =         8
       between = 0.1183                                        avg =       8.0
       overall = 0.1056                                        max =         8

                                                F(14,3801)         =     53.54
corr(u_i, Xb)  = 0.0588                         Prob > F           =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         d81 |  -.0283826   .1463373    -0.19   0.846    -.3152897    .2585246
         d82 |  -.0129925   .1463373    -0.09   0.929    -.2998996    .2739146
```

```
         d83 |   .0100348   .1463373     0.07   0.945    -.2768723    .2969419
         d84 |   .0917698   .1463373     0.63   0.531    -.1951374    .3786769
         d85 |   .0397244   .1463373     0.27   0.786    -.2471827    .3266316
         d86 |   .0525314   .1463373     0.36   0.720    -.2343757    .3394386
         d87 |   .0930808   .1463373     0.64   0.525    -.1938263     .379988
             |
  d81#c.educ |
         1   |   .0125583   .0123016     1.02   0.307    -.0115601    .0366766
             |
  d82#c.educ |
         1   |   .0162474   .0123016     1.32   0.187     -.007871    .0403657
             |
  d83#c.educ |
         1   |   .0183354   .0123016     1.49   0.136     -.005783    .0424537
             |
  d84#c.educ |
         1   |   .0174258   .0123016     1.42   0.157    -.0066926    .0415441
             |
  d85#c.educ |
         1   |   .0260227   .0123016     2.12   0.034     .0019044    .0501411
             |
  d86#c.educ |
         1   |   .0300596   .0123016     2.44   0.015     .0059412    .0541779
             |
  d87#c.educ |
         1   |   .0322871   .0123016     2.62   0.009     .0081688    .0564055
             |
       _cons |   1.393477   .0151875    91.75   0.000     1.363701    1.423253
-------------+----------------------------------------------------------------
     sigma_u |  .38055118
     sigma_e |  .35455536
         rho |  .53531917   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(544, 3801) =     8.65           Prob > F = 0.0000
```

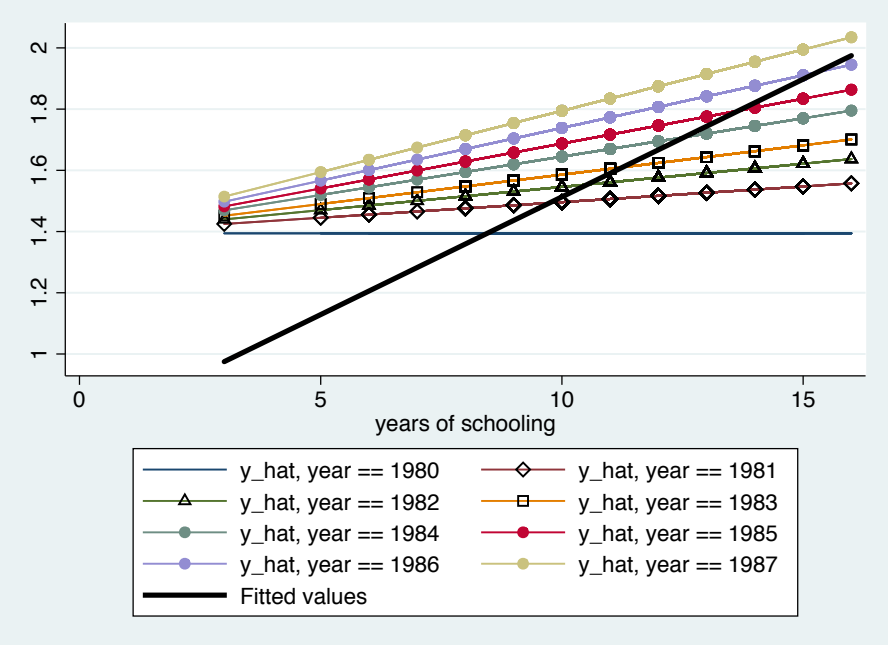
* Without year dummies as intercepts, but only interactions with educ

```
xtreg lwage c.educ#d81 c.educ#d82 c.educ#d83 c.educ#d84 c.educ#d85 c.educ#d86
c.educ#d87 , fe

Fixed-effects (within) regression               Number of obs      =      4360
Group variable: nr                              Number of groups   =       545

R-sq:  within  = 0.1644                         Obs per group: min =         8
       between = 0.1183                                        avg =       8.0
       overall = 0.1089                                        max =         8

                                                F(7,3808)          =    107.04
corr(u_i, Xb)  = 0.0635                         Prob > F           =    0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------|
  d81#c.educ |
         1   |   .0101982   .0018042     5.65   0.000     .0066609    .0137355
             |
  d82#c.educ |
         1   |    .015167   .0018042     8.41   0.000     .0116297    .0187043
             |
  d83#c.educ |
         1   |   .0191698   .0018042    10.63   0.000     .0156325    .0227071
             |
  d84#c.educ |
         1   |   .0250567   .0018042    13.89   0.000     .0215194     .028594
             |
```

```
  d85#c.educ |
          1  |   .0293259   .0018042    16.25   0.000     .0257886    .0328633
             |
  d86#c.educ |
          1  |   .0344277   .0018042    19.08   0.000     .0308904     .037965
             |
  d87#c.educ |
          1  |    .040027   .0018042    22.19   0.000     .0364897    .0435644
             |
       _cons |   1.394139   .0150326    92.74   0.000     1.364666    1.423611
-------------+----------------------------------------------------------------
     sigma_u |  .37939091
     sigma_e |  .35429438
         rho |  .53416615   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(544, 3808) =     9.10                 Prob > F = 0.0000
```

Graph from the latter model (without year dummies, only interactions)

```
predict y_hat
separate y_hat, by(year)
twoway connected y_hat1980- y_hat1987 educ, msymbol(none diamond_hollow
triangle_hollow square_hollow) msize(medium) mcolor(black black black black) ||
lfit lwage educ, clwidth(thick) clcolor(black)
```



With both year dummy variables and their interactions with education, there is no effect of year as intercepts on log(wage). Only from year 1985 onwards that education shows positive relationship with log(wage). With only interactions model, we see increasing association between each year education and log(wage), reflecting the effects of schooling on log(wage) has changed over time.

3.7 General conclusions?

From BP test, we prefer RE model. Since we have all time-invariant variables, using FE will drop all these variables we want to estimate. Hence, in this case, RE will be our preferred model.

4. Show that $corr(\triangle u_{it}, \triangle u_{i,t+1}) = -0.5$ when $u_{it} \sim iidN(0, \sigma_u^2)$, hence we have $E(u_{it}, u_{is}) = 0$ for any $t \neq s$

$corr(\triangle u_{it}, \triangle u_{i,t+1}) = \frac{E(u_{it}-u_{i,t-1}, u_{i,t+1}-u_{it})}{[E(u_{it}-u_{i,t-1})^2 E(u_{it+1}-u_{i,t})^2]^{1/2}}$

$E(u_{it}-u_{i,t-1}, u_{i,t+1}-u_{it}) = E[u_{it}u_{i,t+1}-u_{it}^2-u_{i,t-1}u_{i,t+1}+u_{i,t+1}u_{it}] = -E[u_{it}^2] = -\sigma_u^2$

$E(u_{it} - u_{i,t-1})^2 = E[u_{it}^2 - 2u_{it}u_{i,t-1} + u_{i,t-1}^2] = 2\sigma_u^2$

$E(u_{i,t+1} - u_{it})^2 = E[u_{i,t+1}^2 - 2u_{i,t+1}u_{it} + u_{it}^2] = 2\sigma_u^2$

$\therefore corr(\triangle u_{it}, \triangle u_{i,t+1}) = \frac{-\sigma_u^2}{(2\sigma_u^2 \cdot 2\sigma_u^2)^{1/2}} = -\frac{1}{2}$

5. The composite error $v_{it} = a_i + u_{it}$, $E[a_i, u_{it}] = 0$, $E[u_{it}^2] = \sigma_u^2$, and $E(u_{it}, u_{is}) = 0$ for any $t \neq s$. Define $e_{it} = v_{it} - \lambda \bar{v}_i$, where $\lambda = 1 - [\sigma_u^2/(\sigma_u^2 + T\sigma_a^2)]^{1/2}$.

(5.1) Show that $E(e_{it}) = 0$.

$E[v_{it} - \lambda \bar{v}_i] = E[v_{it}] - \lambda E[\bar{v}_i] = E[a_i + u_{it}] - \lambda E[\bar{a}_i + \bar{u}_i]$

$= E[a_i] + E[u_{it}] - \lambda E[\bar{a}] - \lambda E[\bar{u}] = 0$

(5.2) Show that $Var(e_{it}) = \sigma_u^2$, t = 1, ..., T.

$Var(e_{it}) = E[(v_{it} - \lambda \bar{v}_i)^2] = E(v_{it}^2) - 2\lambda E(v_{it}, \bar{v}_i) + \lambda^2 E(\bar{v}_i^2)$

$E(v_{it}^2) = E[a_i^2 + 2a_i u_{it} + u_{it}^2] = \sigma_a^2 + \sigma_u^2$

$E(\bar{v}_i^2) = E[\bar{a}_i^2 + 2\bar{a}_i\bar{u}_i + \bar{u}_i^2] = \sigma_a^2 + \frac{1}{T^2}\sum E(u_{it}^2) = \sigma_a^2 + \frac{\sigma_u^2}{T}$

$E(v_{it}, \bar{v}_i) = E[a_i + u_{it}, \bar{a}_i + \bar{u}_i] = E[a_i\bar{a}] + E[a_i\bar{u}] + E[u_{it}\bar{a}] + E[u_{it}\bar{u}]$

$= \sigma_a^2 + \frac{\sigma_u^2}{T}$

Note: (i) a_i is the same for every t within each cross-sectional unit i, so $a_i = \bar{a}$.

$E[a_i\bar{a}] = Var(a_i) = \sigma_a^2$

(ii) $E[\bar{u}_i^2] = E\left(\frac{\sum u_i}{T}\right)^2 = \frac{1}{T^2}\sum(Eu_i^2) = \frac{T}{T^2}\sigma_u^2 = \frac{\sigma_u^2}{T}$

.

$\therefore Var(e_{it}) = \sigma_a^2 + \sigma_u^2 - 2\lambda\left[\sigma_a^2 + \frac{\sigma_u^2}{T}\right] + \lambda^2\left[\sigma_a^2 + \frac{\sigma_u^2}{T}\right]$

$= (1 - 2\lambda + \lambda^2)\sigma_a^2 + (1 - 2\lambda + \lambda^2)\frac{\sigma_u^2}{T} + (T-1)\frac{\sigma_u^2}{T}$

$= (1-\lambda)^2(\sigma_a^2 + \frac{\sigma_u^2}{T}) + (T-1)\frac{\sigma_u^2}{T}$
$= \left(1 - 1 + \left(\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_a^2}\right)^{1/2}\right)^2 \left(\frac{T\sigma_a^2 + \sigma_u^2}{T}\right) + \sigma_u^2 - \frac{\sigma_u^2}{T}$
$= \frac{\sigma_u^2}{T} + \sigma_u^2 - \frac{\sigma_u^2}{T} = \sigma_u^2$

(5.3) Show that for $t \neq s$, $Cov(e_{it}, e_{is}) = 0$.

$Cov(e_{it}, e_{is}) = E[(v_{it} - \lambda\bar{v}_i), (v_{is} - \lambda\bar{v}_i)] = E[v_{it}v_{is} - \lambda v_{it}\bar{v}_i - \lambda v_{is}\bar{v}_i + \lambda^2\bar{v}_i^2]$

$E[v_{it}v_{is}] = E[a_i^2 + a_i u_{is} + a_i u_{it} + u_{it}u_{is}] = \sigma_a^2$

$\therefore Cov(e_{it}, e_{is}) = \sigma_a^2 - 2\lambda\left(\sigma_a^2 + \frac{\sigma_u^2}{T}\right) + \lambda^2\left(\sigma_a^2 + \frac{\sigma_u^2}{T}\right) = (1 - 2\lambda + \lambda^2)\left(\sigma_a^2 + \frac{\sigma_u^2}{T}\right) - \frac{\sigma_u^2}{T}$

$= \frac{\sigma_u^2}{\sigma_u^2 + T\sigma_a^2} \cdot \frac{T\sigma_a^2 + \sigma_u^2}{T} - \frac{\sigma_u^2}{T} = 0$