

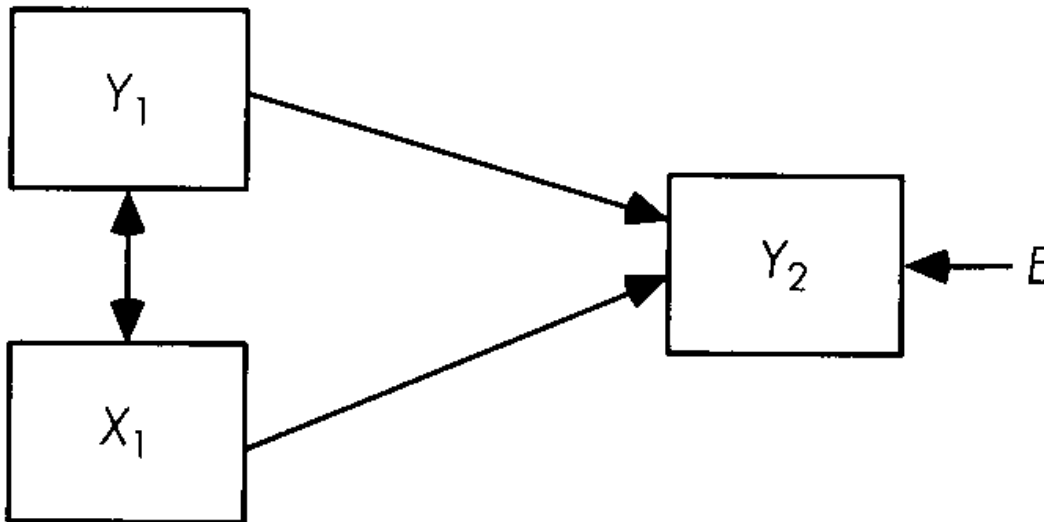
Longitudinal Data Analysis

Instructor: Natasha Sarkisian

Change Models: Analyzing Longitudinal Data with Two Times Points

Lagged Dependent Variable (LDV) approach

This approach is also known as regressor variable approach. The idea is to predict time 2 outcome using time 1 independent variables while controlling for stability in the outcome variable by including the dependent variable from time 1 into the model.



```
. use http://www.sarkisian.net/sc706/crossnat.dta
```

```
. xtset v1 v2, yearly
      panel variable:  v1 (strongly balanced)
      time variable:  v2, 1972 to 1973
      delta: 1 year
```

```
. drop if v2<1972
```

```
. des v66 v76 v7 v41 v36 v82 v85
```

variable name	storage type	display format	value label	variable label
v66	float	%9.0g		television sets
v76	int	%8.0g		pri + sec enr per capita
v7	int	%8.0g		urb 100,000+ per capita
v41	float	%8.0g		size of military
v36	float	%8.0g		% gdp indust activity
v82	int	%8.0g		% literate
v85	float	%8.0g		gdp per capita

```
. reg v66 l.v66
```

Source	SS	df	MS	Number of obs =	113
				F(1, 111) =	70.47

Model	1.5317e+13	1	1.5317e+13	Prob > F	=	0.0000
Residual	2.4128e+13	111	2.1737e+11	R-squared	=	0.3883
-----				Adj R-squared	=	0.3828
Total	3.9445e+13	112	3.5218e+11	Root MSE	=	4.7e+05

v66	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
v66						
L1.	.3459818	.0412157	8.39	0.000	.26431	.4276535
_cons	106675.7	45222.49	2.36	0.020	17064.28	196287

```
. reg v66 l.v66 l.v76 l.v7 l.v41 l.v36 l.v82 l.v85
```

Source	SS	df	MS	Number of obs =	113
Model	2.5535e+13	7	3.6478e+12	F(7, 105) =	27.54
Residual	1.3910e+13	105	1.3247e+11	Prob > F	= 0.0000
-----				R-squared	= 0.6474
-----				Adj R-squared	= 0.6238
Total	3.9445e+13	112	3.5218e+11	Root MSE	= 3.6e+05

v66	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
v66						
L1.	.0414205	.061761	0.67	0.504	-.0810402	.1638811
v76						
L1.	-182.7192	68.92729	-2.65	0.009	-319.3893	-46.04915
v7						
L1.	764.4433	306.5997	2.49	0.014	156.5127	1372.374
v41						
L1.	918.3877	148.9893	6.16	0.000	622.9695	1213.806
v36						
L1.	5967.985	2917.97	2.05	0.043	182.1905	11753.78
v82						
L1.	378.3224	159.7066	2.37	0.020	61.65378	694.991
v85						
L1.	-60.96188	52.6678	-1.16	0.250	-165.3924	43.46863
_cons	-121335.8	103159.4	-1.18	0.242	-325881.7	83210.19

We can do the same thing in wide format:

```
. reshape wide v3 - v134 , i(v1) j(v2)
(note: j = 1972 1973)
```

Data	long	->	wide
Number of obs.	226	->	113
Number of variables	114	->	224
j variable (2 values)	v2	->	(dropped)
xij variables:			
	v3	->	v31972 v31973
	v4	->	v41972 v41973
[output omitted]			
	v134	->	v1341972 v1341973

```
. reg v661973 v661972 v761972 v71972 v411972 v361972 v821972 v851972
```

Source	SS	df	MS	Number of obs = 113		
Model	2.5535e+13	7	3.6478e+12	F(7, 105) =	27.54	
Residual	1.3910e+13	105	1.3247e+11	Prob > F =	0.0000	
Total	3.9445e+13	112	3.5218e+11	R-squared =	0.6474	
				Adj R-squared =	0.6238	
				Root MSE =	3.6e+05	

v661973	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
v661972	.0414205	.061761	0.67	0.504	-.0810402	.1638811
v761972	-182.7192	68.92729	-2.65	0.009	-319.3893	-46.04915
v71972	764.4433	306.5997	2.49	0.014	156.5127	1372.374
v411972	918.3877	148.9893	6.16	0.000	622.9695	1213.806
v361972	5967.985	2917.97	2.05	0.043	182.1905	11753.78
v821972	378.3224	159.7066	2.37	0.020	61.65378	694.991
v851972	-60.96188	52.6678	-1.16	0.250	-165.3924	43.46863
_cons	-121335.8	103159.4	-1.18	0.242	-325881.7	83210.19

This format also allows us to examine interactions of the effects of each of the variables of interest with the lagged DV:

```
. for var v661972 v761972 v71972 v411972 v361972 v821972 v851972: sum X \ gen Xm=X-r(mean)
```

```
-> sum v661972
```

Variable	Obs	Mean	Std. Dev.	Min	Max
v661972	113	267416.8	1068867	0	9900000

```
-> gen v661972m=v661972-r(mean)
```

```
-> sum v761972
```

Variable	Obs	Mean	Std. Dev.	Min	Max
v761972	113	1572.124	645.4242	214	3214

```
-> gen v761972m=v761972-r(mean)
```

```
-> sum v71972
```

Variable	Obs	Mean	Std. Dev.	Min	Max
v71972	113	179.4336	138.0003	0	664

```
-> gen v71972m=v71972-r(mean)
```

```
-> sum v411972
```

Variable	Obs	Mean	Std. Dev.	Min	Max
v411972	113	148.0619	406.0215	0	3375

```
-> gen v411972m=v411972-r(mean)
```

```
-> sum v361972
```

Variable	Obs	Mean	Std. Dev.	Min	Max
----------	-----	------	-----------	-----	-----

```

-----+-----
      v361972 |          113      26.38053      13.935          7          73
-> gen v361972m=v361972-r(mean)
-> sum v821972
      Variable |          Obs          Mean      Std. Dev.          Min          Max
-----+-----
      v821972 |          113      570.9912      345.7928          45          998
-> gen v821972m=v821972-r(mean)
-> sum v851972
      Variable |          Obs          Mean      Std. Dev.          Min          Max
-----+-----
      v851972 |          113      818.3628      1022.956          45          5047
-> gen v851972m=v851972-r(mean)
. for var v761972 v71972 v411972 v361972 v821972 v851972: gen Xstart=X*v661972
-> gen v761972start=v761972*v661972
-> gen v71972start=v71972*v661972
-> gen v411972start=v411972*v661972
-> gen v361972start=v361972*v661972
-> gen v821972start=v821972*v661972
-> gen v851972start=v851972*v661972
. drop v761972start - v851972start
. for var v761972 v71972 v411972 v361972 v821972 v851972: gen Xmstart=Xm*v661972
-> gen v761972mstart=v761972m*v661972
-> gen v71972mstart=v71972m*v661972
-> gen v411972mstart=v411972m*v661972
-> gen v361972mstart=v361972m*v661972
-> gen v821972mstart=v821972m*v661972
-> gen v851972mstart=v851972m*v661972
. reg v661973 v661972 v761972 v71972 v411972 v361972 v821972 v851972 v761972mstart
v71972mstart v411972mstart v361972mstart v821972mstart v851972mstart
> 2mstart v821972mstart v851972mstart

```

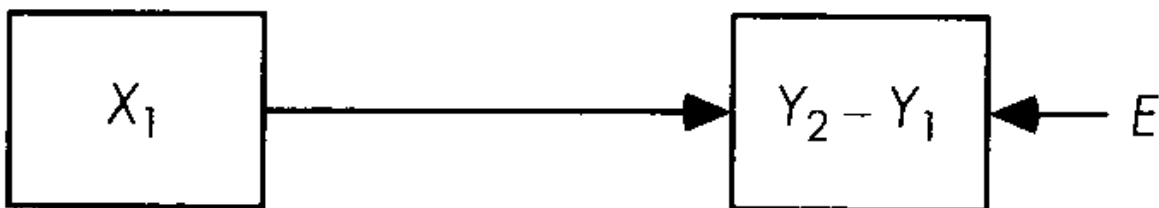
Source	SS	df	MS	Number of obs =	113
Model	3.9286e+13	13	3.0220e+12	F(13, 99) =	1881.60
Residual	1.5900e+11	99	1.6061e+09	Prob > F =	0.0000
				R-squared =	0.9960
				Adj R-squared =	0.9954
Total	3.9445e+13	112	3.5218e+11	Root MSE =	40076

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
v661973						
v661972	.9436762	.0664993	14.19	0.000	.811727	1.075625
v761972	31.63962	8.182788	3.87	0.000	15.4032	47.87605
v71972	7.285169	38.45911	0.19	0.850	-69.02604	83.59638
v411972	37.69741	28.56211	1.32	0.190	-18.97601	94.37083
v361972	-1852.277	364.935	-5.08	0.000	-2576.387	-1128.167
v821972	-91.00154	18.91816	-4.81	0.000	-128.5393	-53.4638
v851972	37.16843	6.292473	5.91	0.000	24.6828	49.65406
v761972mst~t	-.0001489	.0000493	-3.02	0.003	-.0002467	-.0000511
v71972mstart	-.0002934	.0001004	-2.92	0.004	-.0004925	-.0000943
v411972mst~t	-.0001035	.0000136	-7.63	0.000	-.0001305	-.0000766
v361972mst~t	.0054472	.0023584	2.31	0.023	.0007677	.0101267
v821972mst~t	.0010411	.0002177	4.78	0.000	.0006092	.001473
v851972mst~t	-.0002177	9.24e-06	-23.56	0.000	-.0002361	-.0001994
_cons	14449.97	11901.55	1.21	0.228	-9165.282	38065.23

Here we can see that the higher the starting value of DV in 1972, the smaller the effect of each DV. This illustrates the problem of regression to the mean for IV coefficients.

Difference score approach

This approach is also known as the change score approach. There has been a lot of controversy surrounding this approach.



```
. gen v66diff= v661973-v661972
```

```
. reg v66diff v761972 v71972 v411972 v361972 v821972 v851972
```

Source	SS	df	MS	Number of obs = 113		
Model	3.3038e+13	6	5.5063e+12	F(6, 106)	=	12.74
Residual	4.5822e+13	106	4.3229e+11	Prob > F	=	0.0000
				R-squared	=	0.4189
				Adj R-squared	=	0.3861
Total	7.8860e+13	112	7.0411e+11	Root MSE	=	6.6e+05

v66diff	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
v761972	-304.8255	123.698	-2.46	0.015	-550.0689	-59.58205
v71972	499.7393	552.9908	0.90	0.368	-596.6188	1596.097
v411972	-930.7858	161.6021	-5.76	0.000	-1251.178	-610.3939
v361972	15149.34	5161.63	2.93	0.004	4915.906	25382.78
v821972	840.3095	283.4424	2.96	0.004	278.3574	1402.262
v851972	-447.5426	83.82893	-5.34	0.000	-613.7417	-281.3436
_cons	-54057.21	186184.7	-0.29	0.772	-423186.5	315072.1

For many years, difference scores were criticized. One reason is their presumed unreliability – if the DV for time 1 and time 2 are positively correlated (which is pretty much always the case), then the difference score will have lower reliability than each of the time points individually, and if the correlation across time is high, that decrease in reliability will be substantial.

But more recently, Paul Allison (1990) has argued that it is not a problem – “low reliability results from the fact that in calculating the change score we differ out all the stable between-subject variation.” He showed that what matters is error variance, not unreliability.

The second critique is that difference score models do not account for the regression to the mean effect—the phenomenon when extremely low initial scores will be followed by an increase, and extremely high scores – by a decrease. So the initial level might shape change, but if we add the lagged DV to this change score model, we are back to the LDV model, so this strategy is not useful:

```
. reg v66diff v661972 v761972 v71972 v411972 v361972 v821972 v851972
```

Source	SS	df	MS	Number of obs =	113
Model	6.4950e+13	7	9.2786e+12	F(7, 105) =	70.04
Residual	1.3910e+13	105	1.3247e+11	Prob > F =	0.0000
				R-squared =	0.8236
				Adj R-squared =	0.8119
Total	7.8860e+13	112	7.0411e+11	Root MSE =	3.6e+05

v66diff	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
v661972	-.9585796	.061761	-15.52	0.000	-1.08104 - .8361189
v761972	-182.7192	68.9273	-2.65	0.009	-319.3893 -46.04915
v71972	764.4434	306.5997	2.49	0.014	156.5127 1372.374
v411972	918.3878	148.9893	6.16	0.000	622.9695 1213.806
v361972	5967.985	2917.97	2.05	0.043	182.1905 11753.78
v821972	378.3224	159.7066	2.37	0.020	61.65378 694.991
v851972	-60.96188	52.6678	-1.16	0.250	-165.3924 43.46863
_cons	-121335.8	103159.4	-1.18	0.242	-325881.7 83210.19

But Allison argued that regression to the mean does not always happen (although it is common) – mostly if there are ceiling and/or floor effects; the correlation between the initial score and the increase does not have to be negative – it can be positive and then the variance of scores increases with time. Allison argues that regression to the mean is not a problem when we compare stable groups, and in such cases difference score approach may produce better results (less bias) than LDV approach.

Evaluating regression to the mean empirically by examining a group with high scores at time 1 and examining their distance from the mean at time 1 and time 2:

```
. for var v661972: sum X, det \ scalar Xmean1=r(mean) \ gen sample=1 if X>r(p75)\ sum X if X>r(p75)\di r(mean)-Xmean1
```

```
-> sum v661972, det
```

1972 v66

Percentiles		Smallest		
1%	0	0		
5%	0	0		
10%	0	0	Obs	113
25%	720	0	Sum of Wgt.	113
50%		8500	Mean	267416.8
75%		95600	Std. Dev.	1068867
90%		482000	Variance	1.14e+12
95%		1227901	Skewness	7.259941
99%		4540001	Kurtosis	61.94597

```
-> scalar v661972mean1=r(mean)
```

```
-> gen sample=1 if v661972>r(p75)
(85 missing values generated)
```

```
-> sum v661972 if v661972>r(p75)
```

Variable	Obs	Mean	Std. Dev.	Min	Max
v661972	28	1038998	1979140	98000	9900000

```
-> di r(mean)-v661972mean1
771581.47
```

```
. for var v661973: sum X \ scalar Xmean1=r(mean) \ sum X if sample==1\di r(mean)-Xmean1
```

```
-> sum v661973
```

Variable	Obs	Mean	Std. Dev.	Min	Max
v661973	113	199197	593450.7	0	4920001

```
-> scalar v661973mean1=r(mean)
```

```
-> sum v661973 if sample==1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
v661973	28	759401.6	1013181	99500	4920001

```
-> di r(mean)-v661973mean1
560204.62
```

The difference between overall mean and the mean for the selected “high” group was much larger in 1972 than in 1973, which means that these people experienced regression toward the mean. We could also select instead cases that are one SD above the mean (rather than above 75%) but here that would be too few countries. We can do the same test for lower 25th percentile:

```
. for var v661972: sum X, det \ scalar Xmean1=r(mean) \ gen sample1=1 if X<r(p25)\ sum X if X<r(p25)\di r(mean)-Xmean1
```

```
-> sum v661972, det
```

```
              1972 v66
-----
Percentiles      Smallest
 1%              0          0
 5%              0          0
10%              0          0   Obs              113
25%             720          0   Sum of Wgt.      113

50%             8500
                          Largest      Mean          267416.8
75%            95600        1806401   Std. Dev.      1068867
90%           482000        2419401   Variance       1.14e+12
95%          1227901        4540001   Skewness       7.259941
99%          4540001        9900000   Kurtosis       61.94597
```

```
-> scalar v661972mean1=r(mean)
```

```
-> gen sample1=1 if v661972<r(p25)
(85 missing values generated)
```

```
-> sum v661972 if v661972<r(p25)
```

```
Variable |      Obs      Mean   Std. Dev.   Min   Max
-----+-----
v661972 |      28  109.6429   206.5498     0   650
```

```
-> di r(mean)-v661972mean1
-267307.21
```

```
. for var v661973: sum X \ scalar Xmean1=r(mean) \ sum X if sample1==1\di r(mean)-Xmean1
```

```
-> sum v661973
```

```
Variable |      Obs      Mean   Std. Dev.   Min   Max
-----+-----
v661973 |     113  199197   593450.7     0  4920001
```

```
-> scalar v661973mean1=r(mean)
```

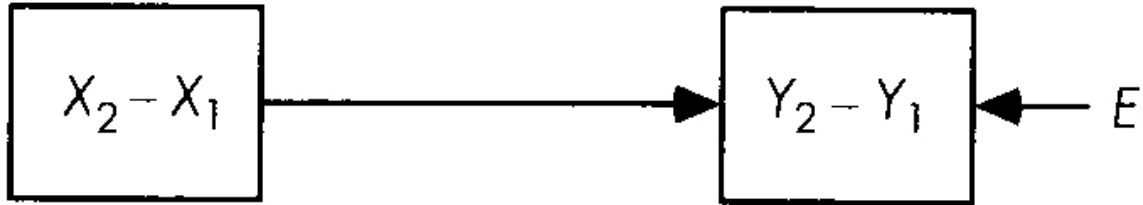
```
-> sum v661973 if sample1==1
```

```
Variable |      Obs      Mean   Std. Dev.   Min   Max
-----+-----
v661973 |      28  128.9286   239.7312     0   725
```

```
-> di r(mean)-v661973mean1
-199068.09
```

These individuals also moved closer to the mean. So we conclude that regression to the mean is a problem for our data, so LDV will be better, especially if we want to document interactions between the starting level of DV and the IVs.

First difference model



```
. for any v76 v7 v41 v36 v82 v85: gen Xdiff=X1973-X1972
-> gen v76diff=v761973-v761972
-> gen v7diff=v71973-v71972
-> gen v41diff=v411973-v411972
-> gen v36diff=v361973-v361972
-> gen v82diff=v821973-v821972
-> gen v85diff=v851973-v851972

. reg v66diff v76diff v7diff v41diff v36diff v82diff v85diff
```

Source	SS	df	MS	Number of obs =	113
Model	3.6920e+13	6	6.1533e+12	F(6, 106) =	15.55
Residual	4.1940e+13	106	3.9566e+11	Prob > F =	0.0000
Total	7.8860e+13	112	7.0411e+11	R-squared =	0.4682
				Adj R-squared =	0.4381
				Root MSE =	6.3e+05

v66diff	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
v76diff	137.7606	1480.131	0.09	0.926	-2796.744	3072.265
v7diff	-1677.299	8486.25	-0.20	0.844	-18502.11	15147.52
v41diff	23645.09	2928.318	8.07	0.000	17839.41	29450.76
v36diff	10445.98	11984.96	0.87	0.385	-13315.37	34207.33
v82diff	-6921.21	12041.42	-0.57	0.567	-30794.49	16952.07
v85diff	-3783.216	1056.735	-3.58	0.001	-5878.296	-1688.135
_cons	130643.2	101348.5	1.29	0.200	-70290.11	331576.6

This model is a fixed effects model for 2 time points, equivalent to the following fixed effects model estimated in the long format:

```
. xi: xtreg v66 v76 v7 v41 v36 v82 v85 i.v2, fe
i.v2 _Iv2_1972-1973 (naturally coded; _Iv2_1972 omitted)
```

Fixed-effects (within) regression	Number of obs =	226
Group variable: v1	Number of groups =	113
R-sq: within = 0.4717	Obs per group: min =	2
between = 0.3306	avg =	2.0
overall = 0.2625	max =	2

```
corr(u_i, Xb) = -0.9972
F(7,106) = 13.52
Prob > F = 0.0000
```

v66	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
v76	137.7606	1480.131	0.09	0.926	-2796.744 3072.265
v7	-1677.299	8486.249	-0.20	0.844	-18502.11 15147.52
v41	23645.09	2928.318	8.07	0.000	17839.41 29450.76
v36	10445.98	11984.96	0.87	0.385	-13315.37 34207.33
v82	-6921.21	12041.41	-0.57	0.567	-30794.49 16952.07
v85	-3783.215	1056.735	-3.58	0.001	-5878.296 -1688.135
_Iv2_1973	130643.2	101348.5	1.29	0.200	-70290.1 331576.5
_cons	3623289	7435302	0.49	0.627	-1.11e+07 1.84e+07

sigma_u	9072151
sigma_e	444781.25
rho	.99760211 (fraction of variance due to u_i)

```
F test that all u_i=0: F(112, 106) = 1.94 Prob > F = 0.0003
```

Once we created a first difference model, can we introduce time-invariant variables as well? We can; by doing that, we are assuming that the effect of this time-invariant variable is not stable over time.

```
. gen v4mean=(v41972+v41973)/2
```

```
. reg v66diff v76diff v7diff v41diff v36diff v82diff v85diff v4mean
```

Source	SS	df	MS	Number of obs =	113
Model	3.8577e+13	7	5.5111e+12	F(7, 105) =	14.36
Residual	4.0283e+13	105	3.8365e+11	Prob > F =	0.0000
				R-squared =	0.4892
				Adj R-squared =	0.4551
Total	7.8860e+13	112	7.0411e+11	Root MSE =	6.2e+05

v66diff	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
v76diff	76.57113	1457.781	0.05	0.958	-2813.939 2967.082
v7diff	-1760.391	8356.497	-0.21	0.834	-18329.78 14809
v41diff	22832.3	2909.909	7.85	0.000	17062.48 28602.11
v36diff	9065.059	11820.26	0.77	0.445	-14372.34 32502.46
v82diff	-6902.676	11857.17	-0.58	0.562	-30413.26 16607.91
v85diff	-3599.942	1044.296	-3.45	0.001	-5670.587 -1529.296
v4mean	-1.97364	.9495878	-2.08	0.040	-3.856498 -.0907832
_cons	172247.7	101785.5	1.69	0.094	-29574.25 374069.6

This demonstrates to us that if we make this assumption of effect changing over time, we can add time-invariant variables to fixed effects models in long data format as well:

```
. by v1: egen v4mean=mean(v4)
```

```
. xi: xtreg v66 v76 v7 v41 v36 v82 v85 v4mean i.v2, fe
i.v2          _Iv2_1972-1973      (naturally coded; _Iv2_1972 omitted)
```

```
Fixed-effects (within) regression      Number of obs = 226
Group variable: v1                    Number of groups = 113
```


Random effects models

Although it is not commonly used, we can also estimate random effects model for two time points:

```
. xi: xtreg v66 v76 v7 v41 v36 v82 v85 i.v2, re
i.v2          _Iv2_1972-1973      (naturally coded; _Iv2_1972 omitted)

Random-effects GLS regression              Number of obs   =       226
Group variable: v1                        Number of groups =       113

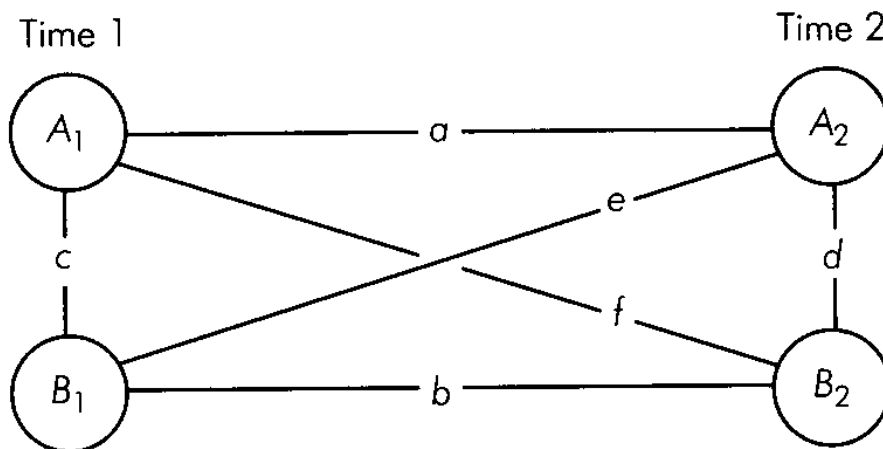
R-sq:  within = 0.0772                    Obs per group:  min =        2
        between = 0.7972                  avg =           2.0
        overall = 0.6184                  max =           2

Random effects u_i ~ Gaussian              Wald chi2(7)    =       311.00
corr(u_i, X) = 0 (assumed)                Prob > chi2     =       0.0000
```

v66	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
v76	-20.95019	75.54375	-0.28	0.782	-169.0132 127.1128	
v7	520.3216	342.5327	1.52	0.129	-151.0301 1191.673	
v41	1488.833	100.8398	14.76	0.000	1291.191 1686.475	
v36	-1974.396	3091.632	-0.64	0.523	-8033.884 4085.092	
v82	-70.02684	175.0421	-0.40	0.689	-413.103 273.0493	
v85	173.6147	50.94598	3.41	0.001	73.76243 273.467	
_Iv2_1973	-70703.78	69293.05	-1.02	0.308	-206515.7 65108.11	
_cons	-63459.06	119046.4	-0.53	0.594	-296785.8 169867.7	
sigma_u	152760.12					
sigma_e	444781.25					
rho	.10551194	(fraction of variance due to u_i)				

Cross-lagged panel model

Another type of change model, in many ways similar to LDV, is useful if you are interested in mutual effects of two variables on one another:



```

. reg v661973 v661972 v361972
-----+-----
Source |      SS      df      MS                Number of obs =      113
-----+-----+-----+-----
Model | 1.6909e+13      2  8.4544e+12           F( 2, 110) =      41.27
Residual | 2.2536e+13     110  2.0487e+11           Prob > F      =      0.0000
-----+-----+-----+-----
Total | 3.9445e+13     112  3.5218e+11           R-squared     =      0.4287
                                           Adj R-squared =      0.4183
                                           Root MSE     =      4.5e+05

-----+-----
v661973 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----+-----+-----+-----
v661972 |   .3262268   .0406362     8.03   0.000   .2456954   .4067583
v361972 |  8688.678   3116.951     2.79   0.006   2511.612   14865.74
   _cons | -117253.5   91546.09    -1.28   0.203  -298676.3   64169.38
-----+-----

```

```

. reg v361973 v361972 v661972
-----+-----
Source |      SS      df      MS                Number of obs =      113
-----+-----+-----+-----
Model | 21788.727      2 10894.3635           F( 2, 110) =     423.40
Residual | 2830.35265     110  25.7304787           Prob > F      =      0.0000
-----+-----+-----+-----
Total | 24619.0796     112  219.813211           R-squared     =      0.8850
                                           Adj R-squared =      0.8829
                                           Root MSE     =      5.0725

-----+-----
v361973 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----+-----+-----+-----
v361972 |  1.003116   .0349313    28.72   0.000   .9338906   1.072342
v661972 | -1.70e-07   4.55e-07    -0.37   0.709  -1.07e-06   7.32e-07
   _cons |   .9810228   1.025946     0.96   0.341  -1.052162   3.014207
-----+-----

```

To establish causal predominance, we can compare standardized effects:

```

. reg v661973 v661972 v361972, beta
-----+-----
Source |      SS      df      MS                Number of obs =      113
-----+-----+-----+-----
Model | 1.6909e+13      2  8.4544e+12           F( 2, 110) =      41.27
Residual | 2.2536e+13     110  2.0487e+11           Prob > F      =      0.0000
-----+-----+-----+-----
Total | 3.9445e+13     112  3.5218e+11           R-squared     =      0.4287
                                           Adj R-squared =      0.4183
                                           Root MSE     =      4.5e+05

-----+-----
v661973 |      Coef.   Std. Err.      t    P>|t|     Beta
-----+-----+-----+-----+-----
v661972 |   .3262268   .0406362     8.03   0.000   .5875689
v361972 |  8688.678   3116.951     2.79   0.006   .2040216
   _cons | -117253.5   91546.09    -1.28   0.203   .
-----+-----

```

```

. reg v361973 v361972 v661972, beta
-----+-----
Source |      SS      df      MS                Number of obs =      113
-----+-----+-----+-----
Model | 21788.727      2 10894.3635           F( 2, 110) =     423.40
Residual | 2830.35265     110  25.7304787           Prob > F      =      0.0000
-----+-----+-----+-----
Total | 24619.0796     112  219.813211           R-squared     =      0.8850
                                           Adj R-squared =      0.8829
                                           Root MSE     =      5.0725

-----+-----
v361973 |      Coef.   Std. Err.      t    P>|t|     Beta
-----+-----+-----+-----+-----
v361972 |  1.003116   .0349313    28.72   0.000   .9428256
v661972 | -1.70e-07   4.55e-07    -0.37   0.709  -.012275
   _cons |   .9810228   1.025946     0.96   0.341   .
-----+-----

```

Same analysis but with controls:

```
. reg v361973 v361972 v661972 v761972 v71972 v411972 v821972 v851972, beta
```

Source	SS	df	MS		
Model	21885.9594	7	3126.56563	Number of obs =	113
Residual	2733.12023	105	26.0297165	F(7, 105) =	120.12
				Prob > F =	0.0000
				R-squared =	0.8890
				Adj R-squared =	0.8816
Total	24619.0796	112	219.813211	Root MSE =	5.1019

v361973	Coef.	Std. Err.	t	P> t	Beta
v361972	.9747406	.0409025	23.83	0.000	.9161554
v661972	-5.62e-07	8.66e-07	-0.65	0.518	-.0405222
v761972	.0011535	.0009662	1.19	0.235	.0502164
v71972	-.0020216	.0042977	-0.47	0.639	-.0188173
v411972	.0000718	.0020884	0.03	0.973	.0019673
v821972	-.001622	.0022387	-0.72	0.470	-.0378306
v851972	.0011451	.0007383	1.55	0.124	.0790055
_cons	.3620761	1.44603	0.25	0.803	.

```
. reg v661973 v661972 v361972 v761972 v71972 v411972 v821972 v851972, beta
```

Source	SS	df	MS		
Model	2.5535e+13	7	3.6478e+12	Number of obs =	113
Residual	1.3910e+13	105	1.3247e+11	F(7, 105) =	27.54
				Prob > F =	0.0000
				R-squared =	0.6474
				Adj R-squared =	0.6238
Total	3.9445e+13	112	3.5218e+11	Root MSE =	3.6e+05

v661973	Coef.	Std. Err.	t	P> t	Beta
v661972	.0414205	.061761	0.67	0.504	.0746026
v361972	5967.985	2917.97	2.05	0.043	.1401361
v761972	-182.7192	68.92729	-2.65	0.009	-.1987215
v71972	764.4433	306.5997	2.49	0.014	.1777627
v411972	918.3877	148.9893	6.16	0.000	.6283339
v821972	378.3224	159.7066	2.37	0.020	.2204415
v851972	-60.96188	52.6678	-1.16	0.250	-.1050825
_cons	-121335.8	103159.4	-1.18	0.242	.

Simultaneous estimation with correlated residuals (unfortunately, no beta option):

```
. reg3 (v661973 v661972 v361972) (v361973 v361972 v661972), corr(unstr)
```

Three-stage least-squares regression

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
v661973	113	2	446576.9	0.4287	84.79	0.0000
v361973	113	2	5.004735	0.8850	869.90	0.0000

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
v661973					
v661972	.3262268	.0400932	8.14	0.000	.2476457 .404808
v361972	8688.678	3075.298	2.83	0.005	2661.206 14716.15
_cons	-117253.5	90322.7	-1.30	0.194	-294282.7 59775.76

```

-----
v361973
v361972 | 1.003116 .0344645 29.11 0.000 .9355671 1.070665
v661972 | -1.70e-07 4.49e-07 -0.38 0.705 -1.05e-06 7.10e-07
_cons | .9810228 1.012236 0.97 0.332 -1.002923 2.964968
-----
Endogenous variables: v661973 v361973
Exogenous variables: v661972 v361972
-----

```

We could do the same thing with the long dataset as well:

```

. reg3 (v66 l.v66 l.v36) (v36 l.v36 l.v66), corr(unstr)
Three-stage least-squares regression

```

```

-----
Equation      Obs   Parms      RMSE      "R-sq"      chi2      P
-----
v66           113     2    446576.9    0.4287     84.79    0.0000
v36           113     2     5.004735    0.8850    869.90    0.0000
-----

```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]

v66					
v66					
L1.	.3262268	.0400932	8.14	0.000	.2476457 .404808
v36					
L1.	8688.678	3075.298	2.83	0.005	2661.206 14716.15
_cons	-117253.5	90322.7	-1.30	0.194	-294282.7 59775.76

v36					
v36					
L1.	1.003116	.0344645	29.11	0.000	.9355671 1.070665
v66					
L1.	-1.70e-07	4.49e-07	-0.38	0.705	-1.05e-06 7.10e-07
_cons	.9810228	1.012236	0.97	0.332	-1.002923 2.964968

```

Endogenous variables: v66 v36
Exogenous variables: L.v66 L.v36
-----

```

Assumptions of this type of analysis:

Finite causal lag corresponding to our measurement: In such models, we are assuming that causal process happens with a specific lag, and the distance between between time points in our dataset reflects, or closely approximates that lag.

Continuity of causal process: This model assumes that the causal processes are continuous and ongoing so we can observe that at any time.

Equality of causal lags: We assume that $A \rightarrow B$ and $B \rightarrow A$ causal lag is of the same length.

Diagnostics for longitudinal data with two time points:

Since the vast majority of the models we discussed can be estimated using OLS regression, diagnostics should be conducted the same way as they are for OLS.