

## Specification Error Problems

### The Model

True linear relationship  $Y_{1t} = 1.3 + 1.2X_{1t} + 3.0X_{2t} + 4.0X_{3t} + u_{1t}$

True quadratic relationship  $Y_{2t} = 0.7 + 0.9X_{1t} + 0.2X_{1t}^2 + u_{2t}$

True cubic relationship  $Y_{3t} = 0.5 + 25X_{1t} - 0.9X_{1t}^2 + 0.1X_{1t}^3 + u_{3t}$

where  $u_{1t}$ ,  $u_{2t}$ , and  $u_{3t}$  are random normally distributed error terms.

*\*Generate Data*

```
g y1=1.3+1.2*x1+3*x2+4*x3+rnormal(0,10)
g x12=x1^2
g x13=x1^3
g y2=0.7+0.9*x1+0.2*x12+rnormal(0,100)
g y3=0.5+25*x1-0.9*x12+0.1*x13+rnormal(0,1000)*x2+1.1*x3+0.6*x4+u
g x5=rnormal(7,21)
```

### Overfitting Model (Inclusion of Irrelevant Variables)

If we estimate incorrect specified model by including irrelevant variables,  $x_4$  and  $x_5$ , into the model, the estimated result will involve specification error problem.

Misspecified model  $Y_{1t} = \beta_0 + \beta_1X_{1t} + \beta_2X_{2t} + \beta_3X_{3t} + \beta_4X_{4t} + \beta_5X_{5t} + u_{1t}$

```
. reg y1 x1 x2 x3 x4 x5
```

Source	SS	df	MS	Number of obs	=	200
Model	756675.607	5	151335.121	F(5, 194)	=	1577.94
Residual	18605.9471	194	95.9069438	Prob > F	=	0.0000
				R-squared	=	0.9760
				Adj R-squared	=	0.9754
Total	775281.554	199	3895.88721	Root MSE	=	9.7932

	y1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1		1.212475	.0239136	50.70	0.000	1.165311 1.259639
x2		2.85696	.2403566	11.89	0.000	2.382913 3.331008
x3		4.113544	.1335059	30.81	0.000	3.850235 4.376854
x4		-.0245756	.0402355	-0.61	0.542	-.1039308 .0547795
x5		-.0107979	.033189	-0.33	0.745	-.0762556 .0546598
_cons		.7380569	.7807263	0.95	0.346	-.8017442 2.277858

### Redundant Variable F-test

Hypothesis  $H_0 : \beta_4 = \beta_5 = 0$

```
. test x4 x5
```

```
( 1) x4 = 0
( 2) x5 = 0
```

```
F( 2, 194) = 0.27
Prob > F = 0.7637
```

```
. est store m1
```

```
. reg y1 x1 x2 x3
```

Source	SS	df	MS	Number of obs	=	200
Model	756623.827	3	252207.942	F(3, 196)	=	2649.45
Residual	18657.7269	196	95.1924842	Prob > F	=	0.0000
				R-squared	=	0.9759
				Adj R-squared	=	0.9756
Total	775281.554	199	3895.88721	Root MSE	=	9.7567

y1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x1	1.213919	.0237389	51.14	0.000	1.167103	1.260736
x2	2.864916	.2392164	11.98	0.000	2.393147	3.336684
x3	4.107888	.1327725	30.94	0.000	3.846042	4.369734
_cons	.7893162	.7240281	1.09	0.277	-.6385695	2.217202

```
. est store m2
```

```
. est table m1 m2, star(.1 .05 .01) stat(N rss F r2 r2_a aic bic)
```

Variable	m1	m2
x1	1.2124752***	1.2139195***
x2	2.8569604***	2.8649157***
x3	4.1135443***	4.1078877***
x4	-.02457563	
x5	-.01079791	
_cons	.73805689	.78931618
N	200	200
rss	18605.947	18657.727
F	1577.9371	2649.4523
r2	.97600105	.97593426
r2_a	.97538252	.9755659
aic	1486.1592	1482.7151
bic	1505.9492	1495.9083

legend: \* p<.1; \*\* p<.05; \*\*\* p<.01

### **Underfitting Model (Omitting Relevant Variables)**

If we estimate incorrect specified model omitting relevant variables,  $x_4$  and  $x_5$ , into the model, the estimated result will involve specification error problem.

Misspecified model 
$$Y_{2t} = \beta_0 + \beta_1 X_{1t} + u_{2t}$$

True Quadratic model 
$$Y_{2t} = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{1t}^2 + u_{2t}$$

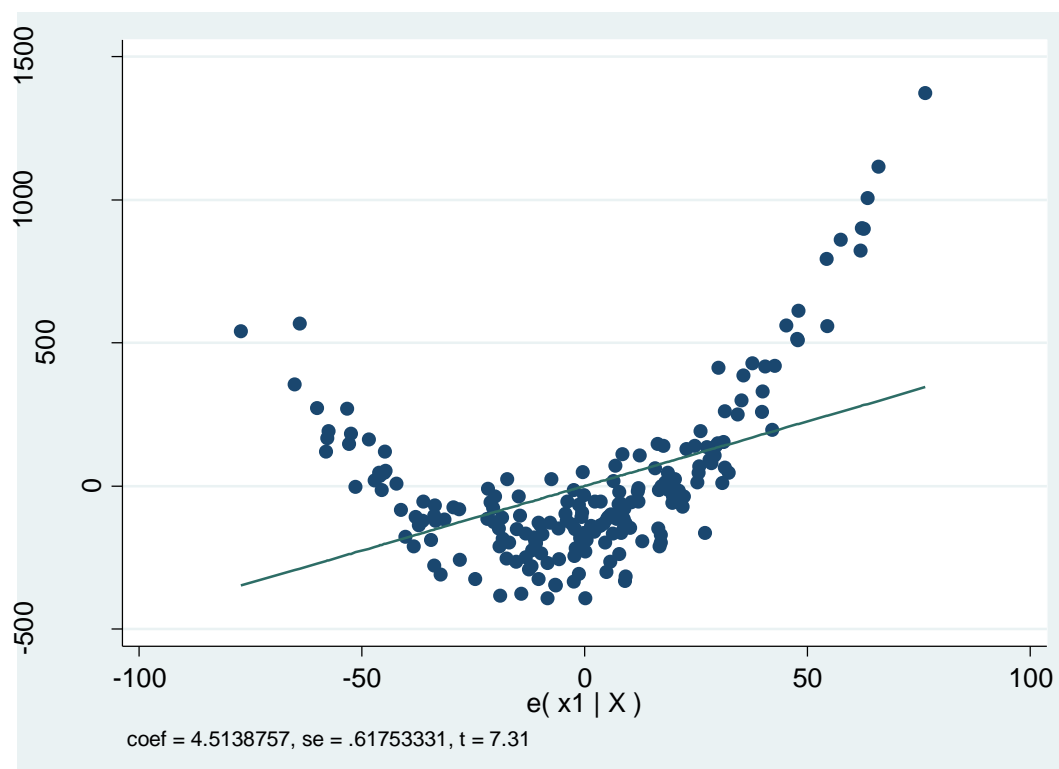
```
. reg y2 x1
```

Source	SS	df	MS	Number of obs	=	200
Model	3453089.77	1	3453089.77	F(1, 198)	=	53.43
Residual	12796601.5	198	64629.3003	Prob > F	=	0.0000
				R-squared	=	0.2125
				Adj R-squared	=	0.2085
Total	16249691.2	199	81656.7399	Root MSE	=	254.22

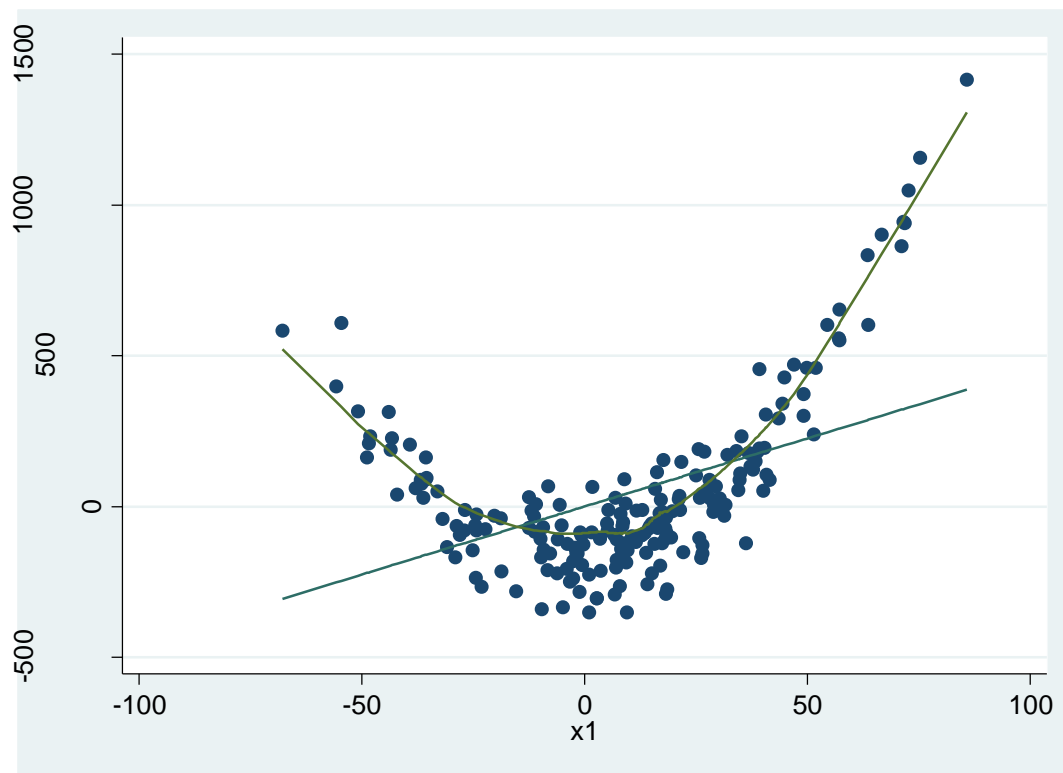
y2	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x1	4.513876	.6175333	7.31	0.000	3.296089	5.731662
_cons	150.8359	18.86267	8.00	0.000	113.6384	188.0334

```
. est store linear
```

```
. avplot x1
```



```
. cprplot x1, low
```



```
. estat ic
```

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
linear	200	-1414.314	-1390.425	2	2784.85	2791.447

Note: N=Obs used in calculating BIC; see [R] BIC note.

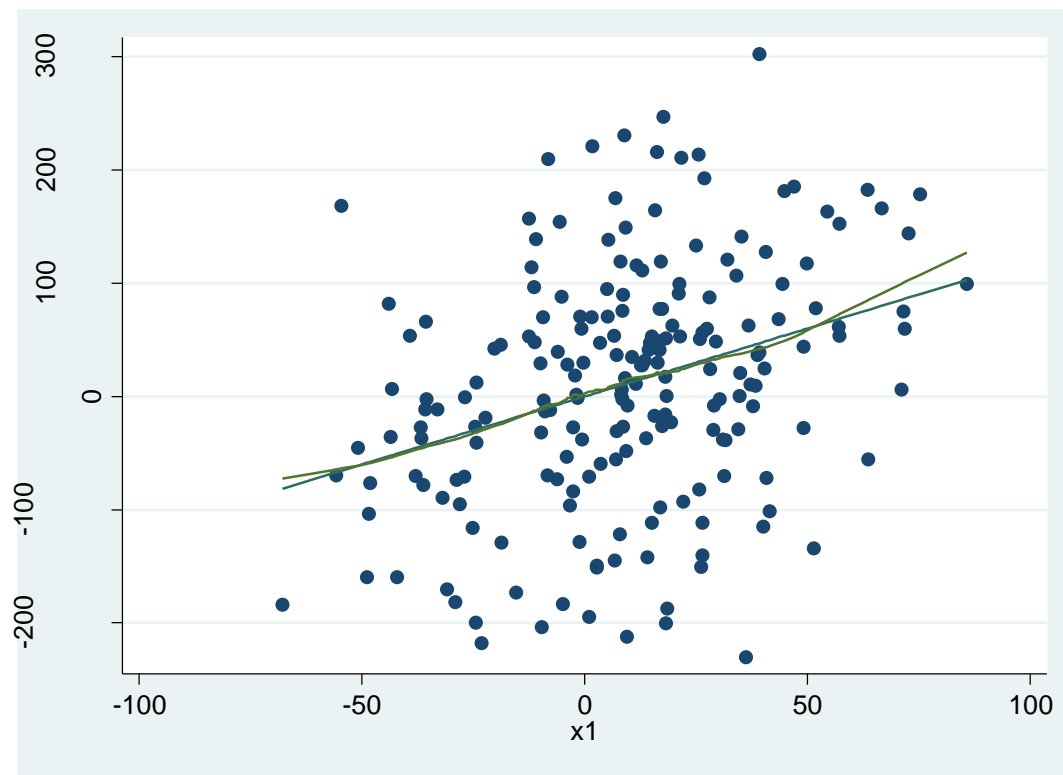
```
. reg y2 x1 x12
```

Source	SS	df	MS	Number of obs	=	200
Model	14258783.3	2	7129391.65	F(2, 197)	=	705.45
Residual	1990907.92	197	10106.1316	Prob > F	=	0.0000
Total	16249691.2	199	81656.7399	R-squared	=	0.8775
				Adj R-squared	=	0.8762
				Root MSE	=	100.53

y2	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	1.199691	.2643942	4.54	0.000	.6782851 1.721098
x12	.2001897	.0061222	32.70	0.000	.1881162 .2122632
_cons	-5.274954	8.85605	-0.60	0.552	-22.73978 12.18988

```
. est store quadratic
```

```
. cprplot x1, low
```



```
. estat ic
```

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
quadratic	200	-1414.314	-1204.366	3	2414.732	2424.627

Note: N=Obs used in calculating BIC; see [R] BIC note.

```
. est table linear quadratic, star(.1 .05 .01) stat(N rss F r2 r2_a aic bic)
```

Variable	linear	quadratic
x1	4.5138757***	1.1996913***
x12		.2001897***
_cons	150.83589***	-5.2749541
N	200	200
rss	12796601	1990907.9
F	53.429168	705.45209
r2	.21250187	.87748026
r2_a	.20852461	.87623641
aic	2784.85	2414.7322
bic	2791.4466	2424.6272

legend: \* p<.1; \*\* p<.05; \*\*\* p<.01

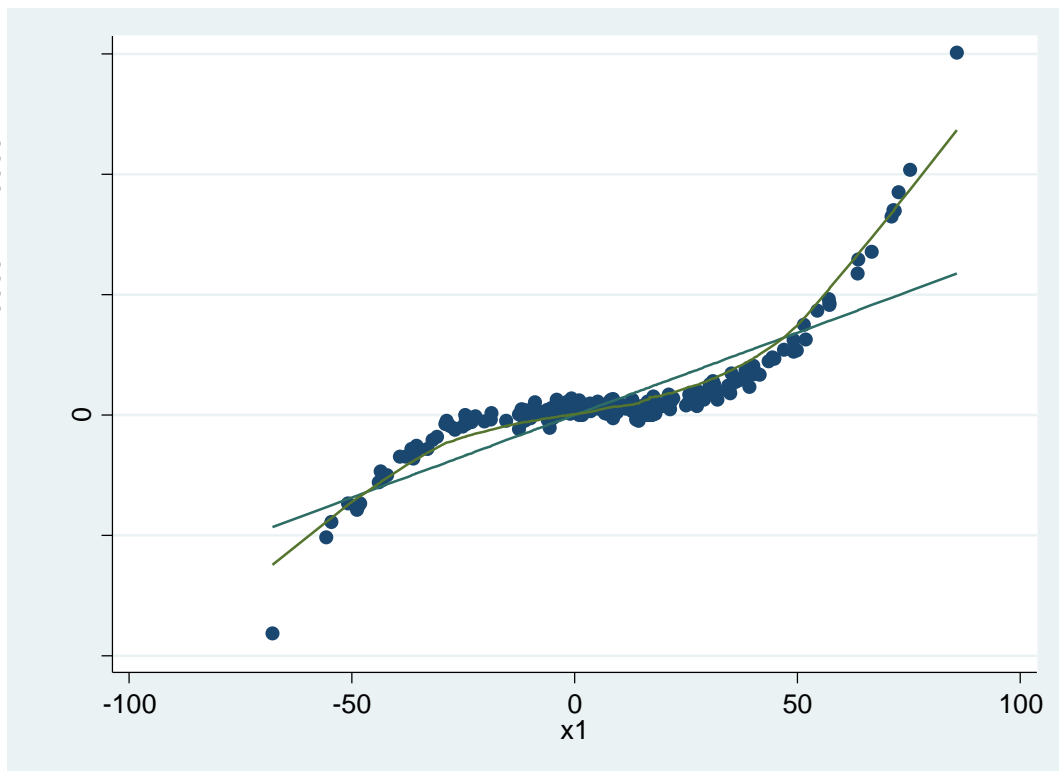
```
. reg y3 x1
```

Source	SS	df	MS	Number of obs	=	200
Model	1.2701e+10	1	1.2701e+10	F(1, 198)	=	460.87
Residual	5.4567e+09	198	27559268.5	Prob > F	=	0.0000
				R-squared	=	0.6995
				Adj R-squared	=	0.6980
Total	1.8158e+10	199	91245445.8	Root MSE	=	5249.7

y3	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	273.7578	12.75204	21.47	0.000	248.6105 298.905
_cons	-876.5405	389.5134	-2.25	0.026	-1644.668 -108.4133

```
. est store lin3
```

```
. cprplot x1, low
```



```
. estat ic
```

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
lin3	200	-2116.193	-1995.968	2	3995.935	4002.532

Note: N=Obs used in calculating BIC; see [R] BIC note.

```
. reg y3 x1 x12 x13
```

Source	SS	df	MS	Number of obs	=	200
Model	1.7971e+10	3	5.9902e+09	F(3, 196)	=	6271.35
Residual	187213565	196	955171.25	Prob > F	=	0.0000
				R-squared	=	0.9897
				Adj R-squared	=	0.9895
Total	1.8158e+10	199	91245445.8	Root MSE	=	977.33

y3	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x1	22.47765	4.193778	5.36	0.000	14.20693	30.74838
x12	-.897277	.0752069	-11.93	0.000	-1.045596	-.7489583
x13	.1006236	.0015407	65.31	0.000	.0975852	.103662
_cons	5.503314	94.37132	0.06	0.954	-180.6103	191.6169

```
. est store cubic
```

```
. estat ic
```

Akaike's information criterion and Bayesian information criterion

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
cubic	200	-2116.193	-1658.732	4	3325.464	3338.657

Note: N=Obs used in calculating BIC; see [R] BIC note.

```
. est table lin3 cubic, star(.1 .05 .01) stat(N rss F r2 r2_a aic bic)
```

Variable	lin3	cubic
x1	273.75778***	22.477655***
x12		-.89727697***
x13		.10062359***
_cons	-876.54046**	5.5033135
N	200	200
rss	5.457e+09	1.872e+08
F	460.86523	6271.3467
r2	.69948331	.98968966
r2_a	.69796554	.98953185
aic	3995.9352	3325.4641
bic	4002.5319	3338.6573

legend: \* p<.1; \*\* p<.05; \*\*\* p<.01

**Measurement Error Problems**

```
. g yls=y1+rnormal(0,5)
```

```
. g xls=x1+rnormal(0,5)
```

```
. reg yls x1 x2 x3
```

Source	SS	df	MS	Number of obs	=	200
-----						
Model	761682.715	3	253894.238	F(3, 196)	=	2500.68
Residual	19899.8696	196	101.529947	Prob > F	=	0.0000
-----						
Total	781582.585	199	3927.55068	R-squared	=	0.9745
-----						
				Adj R-squared	=	0.9741
				Root MSE	=	10.076

yls	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----						
x1	1.225836	.0245164	50.00	0.000	1.177486	1.274186
x2	2.874445	.2470511	11.64	0.000	2.387225	3.361665
x3	4.102454	.1371209	29.92	0.000	3.832032	4.372876
_cons	.255162	.747741	0.34	0.733	-1.219489	1.729813

```
. est store MisY
```

```
. reg y1 xls x2 x3
```

Source	SS	df	MS	Number of obs	=	200
-----						
Model	746691.618	3	248897.206	F(3, 196)	=	1706.33
Residual	28589.9364	196	145.867022	Prob > F	=	0.0000
-----						
Total	775281.554	199	3895.88721	R-squared	=	0.9631
-----						
				Adj R-squared	=	0.9626
				Root MSE	=	12.078

y1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----						
x1s	1.174782	.0290233	40.48	0.000	1.117544	1.23202
x2	3.246092	.2957844	10.97	0.000	2.662764	3.829421
x3	3.878197	.1642885	23.61	0.000	3.554197	4.202197
_cons	.7287071	.8984338	0.81	0.418	-1.043131	2.500545

```
. est store MisX
```

```
. reg y1 x1 x2 x3
```

Source	SS	df	MS	Number of obs	=	200
-----						
Model	756623.827	3	252207.942	F(3, 196)	=	2649.45
Residual	18657.7269	196	95.1924842	Prob > F	=	0.0000
-----						
Total	775281.554	199	3895.88721	R-squared	=	0.9759
-----						
				Adj R-squared	=	0.9756
				Root MSE	=	9.7567

y1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
-----						
x1	1.213919	.0237389	51.14	0.000	1.167103	1.260736
x2	2.864916	.2392164	11.98	0.000	2.393147	3.336684
x3	4.107888	.1327725	30.94	0.000	3.846042	4.369734
_cons	.7893162	.7240281	1.09	0.277	-.6385695	2.217202

```
. est store Correct
. est table Correct MisY MisX, star(.1 .05 .01) stat(N rss F r2 r2_a aic bic)
```

Variable	Correct	MisY	MisX
x1	1.2139195***	1.2258362***	
x2	2.8649157***	2.874445***	3.2460922***
x3	4.1078877***	4.1024538***	3.8781969***
x1s			1.1747824***
_cons	.78931618	.255162	.72870706
N	200	200	200
rss	18657.727	19899.87	28589.936
F	2649.4523	2500.6833	1706.3295
r2	.97593426	.97453901	.96312316
r2_a	.9755659	.9741493	.96255872
aic	1482.7151	1495.6056	1568.074
bic	1495.9083	1508.7989	1581.2672

legend: \* p<.1; \*\* p<.05; \*\*\* p<.01

**Incorrect Specification of Stochastic Error Term****\*Generate Data**

```

g x1e=abs(rnormal(10,100))
g x2e=abs(rnormal(-5,45))
g x3e=abs(rnormal(0,200))
g y1e=exp(1)*(x1e^1)*(x2e^1.2)*(x3e^0.7)*exp(rnormal(0,1))
g lnyle=ln(y1e)
g lnx1e=ln(x1e)
g lnx2e=ln(x2e)
g lnx3e=ln(x3e)

```

```
. reg y1e x1e x2e x3e
```

Source	SS	df	MS	Number of obs	=	200
Model	2.2602e+14	3	7.5339e+13	F(3, 196)	=	15.97
Residual	9.2440e+14	196	4.7163e+12	Prob > F	=	0.0000
Total	1.1504e+15	199	5.7810e+12	R-squared	=	0.1965
				Adj R-squared	=	0.1842
				Root MSE	=	2.2e+06

	yle	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
	x1e	10521.07	2777.478	3.79	0.000	5043.487 15998.64
	x2e	25987.99	5172.525	5.02	0.000	15787.04 36188.94
	x3e	3740.012	1228.61	3.04	0.003	1317.019 6163.004
	_cons	-1515186	379304.5	-3.99	0.000	-2263228 -767144

```
. reg lnyle lnx*
```

Source	SS	df	MS	Number of obs	=	200
Model	841.700383	3	280.566794	F(3, 196)	=	310.86
Residual	176.902138	196	.902561927	Prob > F	=	0.0000
Total	1018.60252	199	5.11860563	R-squared	=	0.8263
				Adj R-squared	=	0.8237
				Root MSE	=	.95003

	lnyle	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
	lnx1e	.9533307	.0600972	15.86	0.000	.8348105 1.071851
	lnx2e	1.235403	.0571043	21.63	0.000	1.122786 1.348021
	lnx3e	.6975647	.0593661	11.75	0.000	.5804865 .814643
	_cons	1.060631	.3962833	2.68	0.008	.2791042 1.842158