

Omitted Variable Tests

The data set nlsy.dta contains information on 252 women in work in the United States

A regression of the log of hourly pay on years of work experience (exper), years at the current firm (tenure) and years of education (educ) gives the following

```
. reg lwage exper tenure educ if female==1
```

Source	SS	df	MS			
Model	10.4983092	3	3.49943639	Number of obs =	252	
Residual	39.0353026	248	.157400414	F(3, 248) =	22.23	
				Prob > F =	0.0000	
				R-squared =	0.2119	
				Adj R-squared =	0.2024	
Total	49.5336118	251	.197345067	Root MSE =	.39674	

	lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
exper		.0022675	.0021046	1.08	0.282	-.0018776	.0064127
tenure		.0102745	.0052232	1.97	0.050	-.000013	.020562
educ		.0800361	.0104624	7.65	0.000	.0594296	.1006426
_cons		.3561163	.1410633	2.52	0.012	.0782814	.6339512

Stata's version of the Ramsey RESET test gives

```
. ovtest
```

Ramsey RESET test using powers of the fitted values of lwage

Ho: model has no omitted variables

```
F(3, 245) = 6.51
Prob > F = 0.0003
```

The equivalent manual version with 3 powers of the predicted variable

```
predict yhat
g yhat2=yhat^2
g yhat3=yhat^3
g yhat4=yhat^4
```

```
reg lwage exper tenure educ yhat2 yhat3 yhat4 if female==1
```

Source	SS	df	MS			
Model	13.3791313	6	2.22985522	Number of obs =	252	
Residual	36.1544805	245	.147569308	F(6, 245) =	15.11	
				Prob > F =	0.0000	
				R-squared =	0.2701	
				Adj R-squared =	0.2522	
Total	49.5336118	251	.197345067	Root MSE =	.38415	

	lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
exper		.0076626	.0274456	0.28	0.780	-.0463968	.0617221
tenure		.0326196	.1236523	0.26	0.792	-.2109375	.2761768
educ		.2807438	.9637558	0.29	0.771	-1.61756	2.179048
yhat2		-6.316156	16.1629	-0.39	0.696	-38.15213	25.51982
yhat3		4.110669	9.099808	0.45	0.652	-13.81317	22.03451
yhat4		-.8021463	1.838969	-0.44	0.663	-4.424352	2.820059

```
_cons | 1.877042 1.30147 1.44 0.151 -.6864544 4.440539
```

```
. test yhat2=yhat3=yhat4=0
```

- (1) yhat2 - yhat3 = 0
- (2) yhat2 - yhat4 = 0
- (3) yhat2 = 0

```
F( 3, 245) = 6.51
Prob > F = 0.0003
```

Similarly the equivalent LM Test

```
reg uhat exper tenure
predict uhat if e(sample), resid
```

```
. reg uhat exper tenure educ yhat2 yhat3 yhat4
```

Source	SS	df	MS			
Model	5.49871214	6	.916452023	Number of obs =	252	
Residual	33.5365905	245	.136884043	F(6, 245) =	6.70	
-----				Prob > F =	0.0000	
Total	39.0353027	251	.155519134	R-squared =	0.1409	
-----				Adj R-squared =	0.1198	
				Root MSE =	.36998	

uhat	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
exper	-.0022505	.0020113	-1.12	0.264	-.0062121	.001711
tenure	-.0092757	.0051921	-1.79	0.075	-.0195026	.0009511
educ	-.078739	.0163717	-4.81	0.000	-.1109862	-.0464918
yhat2	2.175833	1.807114	1.20	0.230	-1.383628	5.735294
yhat3	-1.588593	1.517461	-1.05	0.296	-4.577527	1.400341
yhat4	.3727902	.3515099	1.06	0.290	-.3195767	1.065157
_cons	-.3174041	.7014818	-0.45	0.651	-1.699109	1.0643

and $NR^2 \overset{a}{\sim} \chi^2_{(No.ofrestrictions)} = 252 * 0.1409 = 35.5 \sim \chi^2_{(3)} > 7.85$ (95% critical value)

So decide to add higher order powers of the right hand side variables (in this case the squares of experience tenure & education) [in truth the significance of the test is partly the result of the large standard errors on the yhat variables partly the result of multicollinearity]

```
. reg lwage exper tenure educ exper2 ten2 ed2 if female==1
```

Source	SS	df	MS			
Model	15.6378395	6	2.60630658	Number of obs =	252	
Residual	33.8957723	245	.138350091	F(6, 245) =	18.84	
				Prob > F	= 0.0000	
				R-squared	= 0.3157	
				Adj R-squared	= 0.2989	
Total	49.5336118	251	.197345067	Root MSE	= .37195	

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
exper	.0158065	.0064825	2.44	0.015	.003038	.0285751
tenure	.0368875	.0112897	3.27	0.001	.0146503	.0591248
educ	-.0897278	.0375597	-2.39	0.018	-.163709	-.0157466
exper2	-.000335	.0001412	-2.37	0.018	-.0006132	-.0000569
ten2	-.0012862	.0004836	-2.66	0.008	-.0022388	-.0003336
ed2	.0072476	.001617	4.48	0.000	.0040626	.0104327
_cons	1.191018	.2359601	5.05	0.000	.7262488	1.655787

This time the higher order terms are significant and the RESET test suggests no evidence of functional form misspecification

```
. ovtest
```

Ramsey RESET test using powers of the fitted values of lwage

Ho: model has no omitted variables

F(3, 242) = 1.32

Prob > F = 0.2683

However if we add a dummy variable to indicate whether the individual works in an urban area, the urban dummy variable is positive and significant (there is a wage premium to working in an urban area)

```
. reg lwage exper tenure educ exper2 ten2 ed2 urban if female==1
```

Source	SS	df	MS			
Model	16.6735777	7	2.38193967	Number of obs =	252	
Residual	32.8600341	244	.134672271	F(7, 244) =	17.69	
				Prob > F =	0.0000	
				R-squared =	0.3366	
				Adj R-squared =	0.3176	
				Root MSE =	.36698	

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
exper	.0177786	.0064352	2.76	0.006	.005103	.0304541
tenure	.0363432	.0111403	3.26	0.001	.0143997	.0582867
educ	-.101155	.0372855	-2.71	0.007	-.1745975	-.0277124
exper2	-.0003656	.0001398	-2.62	0.009	-.0006409	-.0000903
ten2	-.0012701	.0004772	-2.66	0.008	-.00221	-.0003302
ed2	.0075117	.0015982	4.70	0.000	.0043636	.0106598
urban	.150235	.0541733	2.77	0.006	.043528	.256942
_cons	1.16322	.2330184	4.99	0.000	.7042357	1.622204

Moral: the RESET test is only a test of whether the model is linear in the original variables. It cannot pick up the influence of other variables