

PLSC 502 – Fall 2016

Transformations

November 29, 2016

Why Transform?

- Linearity
- Additivity
- Normality (of u_i s)
- Interpretation

This:

$$Y_i = \beta_0 X_i^{\beta_1} u_i$$

becomes this:

$$\ln(Y_i) = \ln(\beta_0) + \beta_1 X_i + \ln(u_i)$$

And this:

$$\exp(Y_i) = \beta_0 + \beta_1 X_i + u_i$$

becomes this:

$$Y_i = \ln(\beta_0) + \beta_1 \ln(X_i) + \ln(u_i)$$

Monotonic Transformations

The “Ladder of Powers”:

Transformation	p	$f(X)$	Fox's $f(X)$
Cube	3	X^3	$\frac{X^3-1}{3}$
Square	2	X^2	$\frac{X^2-1}{2}$
(None/Identity)	(1)	(X)	(X)
Square Root	$\frac{1}{2}$	\sqrt{X}	$2(\sqrt{X} - 1)$
Cube Root	$\frac{1}{3}$	$\sqrt[3]{X}$	$3(\sqrt[3]{X} - 1)$
Log	0 (sort of)	$\ln(X)$	$\ln(X)$
Inverse Cube Root	$-\frac{1}{3}$	$\frac{1}{\sqrt[3]{X}}$	$\frac{(\frac{1}{\sqrt[3]{X}}-1)}{-\frac{1}{3}}$
Inverse Square Root	$-\frac{1}{2}$	$\frac{1}{\sqrt{X}}$	$\frac{(\frac{1}{\sqrt{X}}-1)}{-\frac{1}{2}}$
Inverse	-1	$\frac{1}{X}$	$\frac{(\frac{1}{X}-1)}{-1}$
Inverse Square	-2	$\frac{1}{X^2}$	$\frac{(\frac{1}{X^2}-1)}{-2}$
Inverse Cube	-3	$\frac{1}{X^3}$	$\frac{(\frac{1}{X^3}-1)}{-3}$

A General Rule

Using higher-order power transformations (e.g. squares, cubes, etc.) “inflates” large values and “compresses” small ones; conversely, using lower-order power transformations (logs, etc.) “compresses” large values and “inflates” (or “expands”) smaller ones.

Power Transformations: Two Issues

1. X must be *positive*; so:

$$X^* = X + (|X_I| + \epsilon)$$

with (CZ's Rule of Thumb):

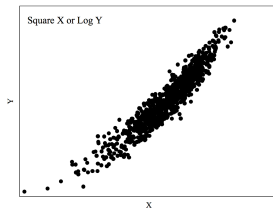
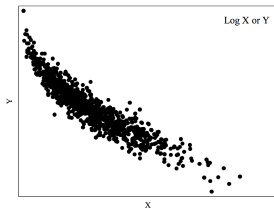
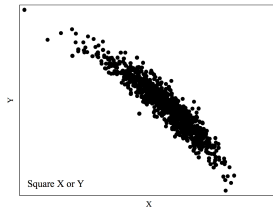
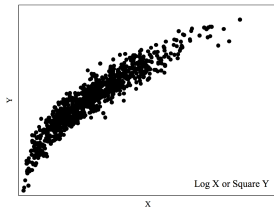
$$\epsilon = \frac{X_{I+1} - X_I}{2}$$

2. Power transformations generally require that:

$$\frac{X_h}{X_l} > 5 \text{ (or so)}$$

Which Transformation?

Mosteller and Tukey's "Bulging Rule":



Simple solution: Polynomials...

- Second-order / quadratic:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$$

- Third-order / cubic:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i$$

- p th-order:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \dots + \beta_p X_i^p + u_i$$

Transformed X s: Interpretation

For:

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i,$$

then:

$$E(Y) = \exp(\beta_0 + \beta_1 X_i)$$

and so:

$$\frac{\partial E(Y)}{\partial X} = \exp(\beta_1).$$

Transformed X s: Interpretation

Similarly, for:

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$$

we have:

$$\frac{\partial E(Y)}{\partial \ln(X)} = \beta_1.$$

So doubling X (say, from X_ℓ to $2X_\ell$):

$$\begin{aligned}\Delta E(Y) &= E(Y|X = 2X_\ell) - E(Y|X = X_\ell) \\ &= [\beta_0 + \beta_1 \ln(2X_\ell)] - [\beta_0 + \beta_1 \ln(X_\ell)] \\ &= \beta_1 [\ln(2X_\ell) - \ln(X_\ell)] \\ &= \beta_1 \ln(2)\end{aligned}$$

Log-Log Regressions

Specifying:

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + \dots + u_i$$

means:

$$\text{Elasticity}_{YX} \equiv \frac{\% \Delta Y}{\% \Delta X} = \beta_1.$$

IOW, a one-percent change in X leads to a $\hat{\beta}_1$ -percent change in Y .

An Example: Military Spending and GDP

Data are from Fordham and Walker...

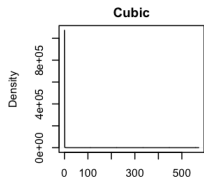
```
> with(Data, summary(milgdp))
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
0.000	0.238	0.749	2.115	2.104	136.900	4327

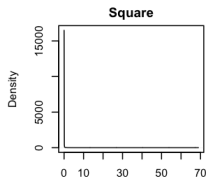
```
> with(Data, summary(gdp))
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
0.0001	0.0033	0.0047	0.0534	0.0153	8.3010	2690

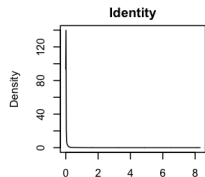
“Ladder of Powers”: GDP



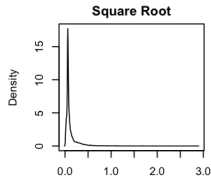
N = 11809 Bandwidth = 3.665e-07



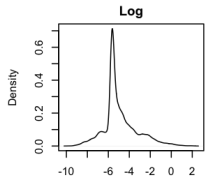
N = 11809 Bandwidth = 2.303e-05



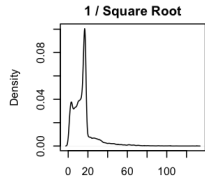
N = 11809 Bandwidth = 0.001235



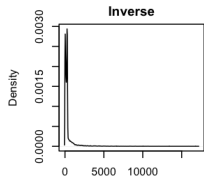
N = 11809 Bandwidth = 0.006808



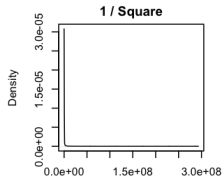
N = 11809 Bandwidth = 0.1573



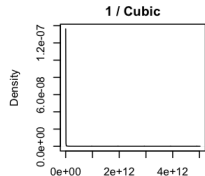
N = 11809 Bandwidth = 0.9539



N = 11809 Bandwidth = 24.25

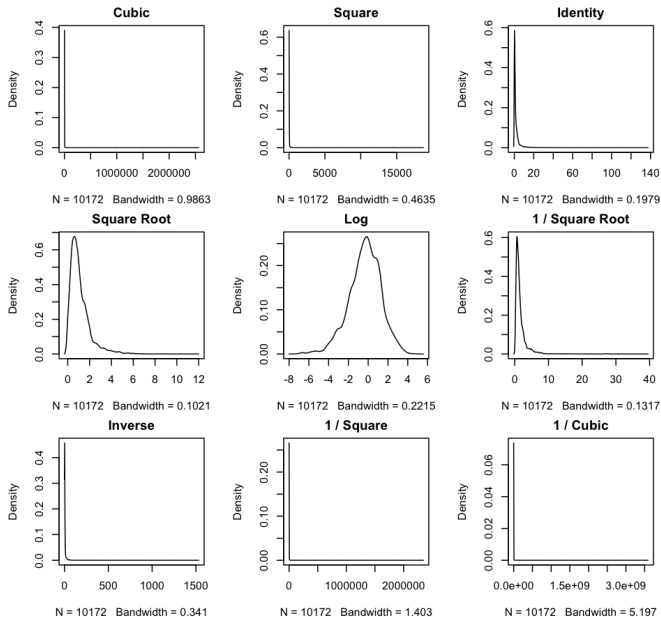


N = 11809 Bandwidth = 8876

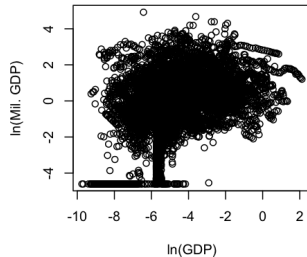
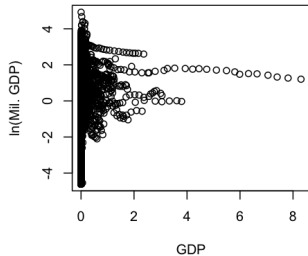
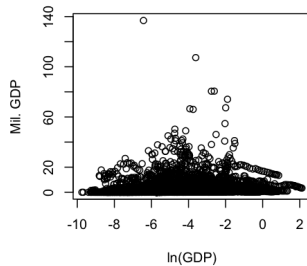
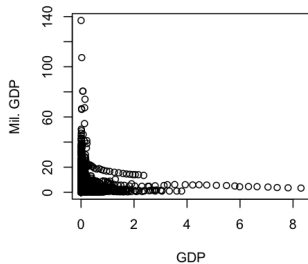


N = 11809 Bandwidth = 2.773e+06

“Ladder of Powers”: Military Spending



Scatterplots



Some Regressions

Untransformed:

```
> with(Data, summary(lm(milgdp~gdp)))
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.0538	0.0481	42.696	< 2e-16 ***
gdp	1.0038	0.1540	6.518	7.45e-11 ***

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 4.757 on 10170 degrees of freedom
(4327 observations deleted due to missingness)

Multiple R-squared: 0.00416, Adjusted R-squared: 0.004062

F-statistic: 42.49 on 1 and 10170 DF, p-value: 7.454e-11

Some Regressions

Logging X :

```
> with(Data, summary(lm(milgdp~log(gdp))))
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.60137	0.13969	32.94	<2e-16 ***
log(gdp)	0.52196	0.02766	18.87	<2e-16 ***

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 4.686 on 10170 degrees of freedom
(4327 observations deleted due to missingness)

Multiple R-squared: 0.03384, Adjusted R-squared: 0.03374

F-statistic: 356.2 on 1 and 10170 DF, p-value: < 2.2e-16

Some Regressions

Logging Y:

```
> with(Data, summary(lm(log(milgdp+0.01)~gdp)))
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.45918	0.01669	-27.51	<2e-16 ***
gdp	0.75794	0.05343	14.18	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.651 on 10170 degrees of freedom
(4327 observations deleted due to missingness)

Multiple R-squared: 0.0194, Adjusted R-squared: 0.0193

F-statistic: 201.2 on 1 and 10170 DF, p-value: < 2.2e-16

Some Regressions

Logging X and Y :

```
> with(Data, summary(lm(log(milgdp+0.01)~log(gdp))))
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.644270	0.044736	36.76	<2e-16 ***
log(gdp)	0.431875	0.008858	48.76	<2e-16 ***

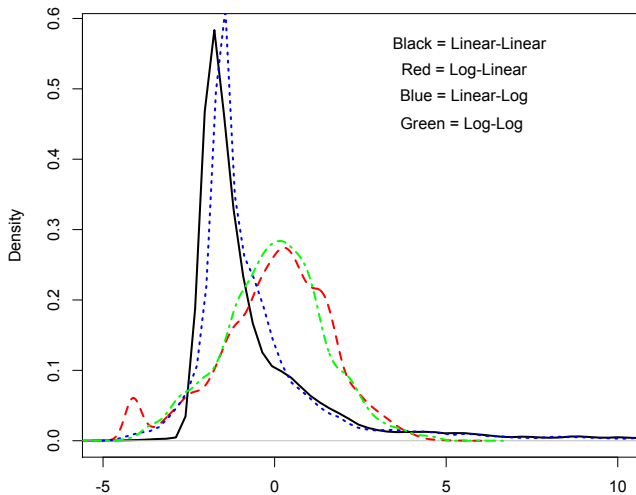
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 1.501 on 10170 degrees of freedom
(4327 observations deleted due to missingness)

Multiple R-squared: 0.1895, Adjusted R-squared: 0.1894

F-statistic: 2377 on 1 and 10170 DF, p-value: < 2.2e-16

Density Plots of \hat{u}_i s



N = 10172 Bandwidth = 0.1923

- **Theory is valuable.**
- **Try different things.**
- **Look at plots.**
- **It takes practice.**