As this example demonstrates, even when the errors in the DGP are normally, identically, and independently distributed, using the wrong transformation of the dependent variable as the regressand yields, in general, a regression with error terms that are neither homoskedastic nor symmetric. Thus, when we encounter heteroskedasticity and skewness in the residuals of a regression, one possible way to eliminate them is to estimate a different regression model in which the dependent variable has been subjected to some sort of nonlinear transformation.

Comparing Alternative Models

It is perfectly easy to subject the dependent variable to various nonlinear transformations and estimate one or more regression models for each of them. However, least-squares estimation does not provide any way to compare the fits of competing models that involve different transformations. But maximum likelihood estimation under the assumption that the error terms are normally distributed does provide a straightforward way to do so. The idea is to compare the loglikelihoods of the alternative models considered as models for the same dependent variable.

For Model 1, in which y_t is the regressand, the concentrated loglikelihood function is simply

$$-\frac{n}{2}\log 2\pi - \frac{n}{2} - \frac{n}{2}\log\biggl(\frac{1}{n}\sum_{t=1}^{n}(y_t - \boldsymbol{X}_{t1}\boldsymbol{\beta}_1)^2\biggr). \tag{10.92}$$

Expression (10.92) is just expression (10.11) specialized to Model 1. Most regression packages report the value of (10.92) evaluated at the OLS estimates as the maximized value of the loglikelihood function.

In order to construct the loglikelihood function for the loglinear Model 2, interpreted as a model for y_t rather than for $\log y_t$, we need the density of y_t as a function of the model parameters. This requires us to use a standard result about **transformations of variables**. Suppose that we wish to know the CDF of a random variable X, but that what we actually know is the CDF of a random variable Z defined as $Z = h(X)$, where $h(\cdot)$ is a strictly increasing deterministic function. Denote this known CDF by F_Z. Then we can obtain the CDF F_X of X as follows.

$$\begin{aligned} F_X(x) &= \Pr(X \le x) = \Pr\bigl(h(X) \le h(x)\bigr) \\ &= \Pr\bigl(Z \le h(x)\bigr) = F_Z\bigl(h(x)\bigr). \end{aligned} \tag{10.93}$$

The second equality above follows because $h(\cdot)$ is strictly increasing. The relation between the densities, or PDFs, of the variables X and Z is obtained by differentiating the leftmost and rightmost quantities in (10.93) with respect to x. Denoting the PDFs by $f_X(\cdot)$ and $f_Z(\cdot)$, we obtain

$$f_X(x) = F'_X(x) = F'_Z\bigl(h(x)\bigr)h'(x) = f_Z\bigl(h(x)\bigr)h'(x).$$

If h is strictly decreasing, the above result must be modified so as to use the absolute value of the derivative. As readers are asked to show in Exercise 10.23, the result then becomes

$$f_X(x) = f_Z\big(h(x)\big)|h'(x)|. \tag{10.94}$$

It is not difficult to see that (10.94) is a perfectly general result which holds for any strictly monotonic function h.

The factor by which $f_Z(z)$ is multiplied in order to produce $f_X(x)$ is the absolute value of what is called the **Jacobian** of the transformation. For Model 2, X is replaced by y_t, and the transformation h is the logarithm, so that Z becomes $\log y_t$. The density of y_t is then given by (10.94) in terms of that of $\log y_t$:

$$f(y_t) = f(\log y_t)\left|\frac{d\log y_t}{dy_t}\right| = \frac{f(\log y_t)}{y_t}, \tag{10.95}$$

where we drop subscripts and denote the PDFs of y_t and $\log y_t$ by $f(y_t)$ and $f(\log y_t)$, respectively.

We can now compute the loglikelihood for Model 2 thought of as a model for the y_t. The concentrated loglikelihood for the $\log y_t$ is given by (10.11):

$$-\frac{n}{2}\log 2\pi - \frac{n}{2} - \frac{n}{2}\log\bigg(\frac{1}{n}\sum_{t=1}^{n}(\log y_t - \boldsymbol{X}_{t2}\boldsymbol{\beta}_2)^2\bigg). \tag{10.96}$$

This expression is the log of the product of the densities of the $\log y_t$. Since the density of y_t, by (10.95), is equal to $1/y_t$ times the density of $\log y_t$, the loglikelihood function we are seeking is

$$-\frac{n}{2}\log 2\pi - \frac{n}{2} - \frac{n}{2}\log\bigg(\frac{1}{n}\sum_{t=1}^{n}(\log y_t - \boldsymbol{X}_{t2}\boldsymbol{\beta}_2)^2\bigg) - \sum_{t=1}^{n}\log y_t. \tag{10.97}$$

The last term here is a **Jacobian term**. It is the sum over all t of the logarithm of the **Jacobian factor** $1/y_t$ in the density of y_t. This Jacobian term is absolutely critical. If it were omitted, Model 2 would be a model for $\log y_t$, and it would make no sense to compare the value of the loglikelihood for (10.96) with the value for Model 1, which is a model for y_t. But when the Jacobian term is included, the loglikelihoods for both models are expressed in terms of y_t, and it is perfectly valid to compare their values. We can say with confidence that the model corresponding to whichever of (10.92) and (10.97) has the largest value is the model that better fits the data.

Most regression packages evaluate expression (10.96) at the OLS estimates for the loglinear model and report that as the maximized value of the loglikelihood function. In order to compute the loglikelihood (10.97), which is what we need if we are to compare the fits of the linear and loglinear models, we have to add the Jacobian term to the value reported by the package.