**Context**

The original journal articles that used the panel dataset in question (Grunfeld 1958, Boot and de Wit 1960, and Zellner 1962) did not make clear the relationship between the real value of the firm (F) and the real value of the firm’s capital stock (C). These papers suggest that the real value of the firm reflects the market capitalization only and does not reflect the stock of plant and equipment, as captured in C.

**Variation by Firm and Year**

I started by plotting the data over time, which revealed that both investment and the value of firm’s stock of capital goods and equipment increased over time.

This also indicates that inter-year correlation exists within a given firm. However, there is also wide variation between firms, as we can see below.

The real value of the firm plotted by firm and by time yields similar results. The scatterplots suggest that firms

**Dummy Variables**

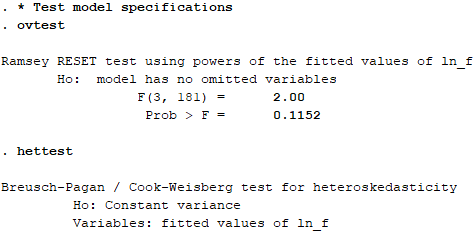
Consequently, I created the following dummy variables:

* d\_123: Includes observations for firms 1, 2, and 3
* d\_468: Includes observations for firms 4, 6, and 8
* d\_5790: Includes observations for firms 5, 7, 9, and 10
* d\_ww2: Includes all observations from 1939 through 1945 (World War II)
* d\_kw: Includes all observations from 1950 through 1953 (Korean War)
* d\_gd: Includes all observations from 1935 through 1941 (Great Depression)

These dummy variables reflect both the historical events that could introduce exogenous shocks to firm values, as well as the particular variations between firms that explain a portion of the variation.

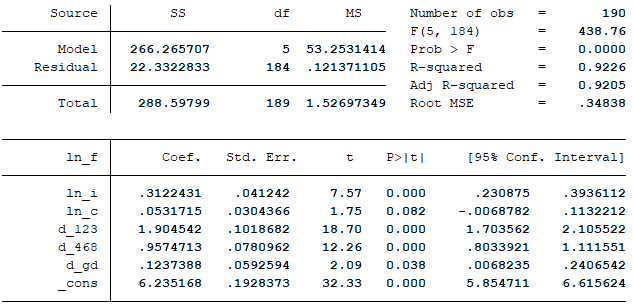
**Instrumental Variables**

My initial models failed the Breusch-Pagan and Cook-Weisberg tests for heteroscedasticity because of the influence of year-over-year trends. To de-trend the data, I created instrumental variables that represented the log of each variable divided by the square root of the year of the observation. Regressions using these instrumental variables passed the test for heteroscedasticity, as well as the omitted variables test (see below).



**Best-Fit Model**

The model that best fit the data is reflected below, with all explanatory variables significant at p < 0.1 or less and an overall R2 of 0.9205.



**Tests for Model Specificity**

I also tested model specificity in other ways. The residuals appear to follow a relatively normal, independent distribution (see below).



However, the model did not pass the tests of the normality of the distribution of the residuals (Shapiro-Wilk test and skewness/kurtosis tests). This is a potential threat to the validity of the model.

**Fixed and Random Effects Models**

Ultimately, however, the best-fit model was a random effects model in which all explanatory variables were significant at p < 0.1 or less and with an overall R2 of 0.9569 (see below). Because the panel dataset was strongly balanced (i.e., contained one observation per firm per year), it seemed appropriate to apply either a fixed effects or random effects model to the data. I performed a Hausman test, which evaluates whether the error terms and explanatory variables are correlated, to determine whether a fixed or random effects model was more appropriate. The test suggested a random effects model was more appropriate (Probability > χ2 = 0.15). It is worth noting, however, that the two models returned very similar results (R2 of 0.947 for both models; coefficients differed by approximately 0.02). Still, using random effects removes autocorrelation in the data.

