

## Description of RealData

This document describes the file RealData.mat (or RealData.csv), which contains the data you will use to estimate the model. The data is derived from an annual sample of firms from the CRSP/Compustat merged database for the 1971 to 2016 period. We do not consider in our data firms with SIC code 6 (firms with fewer than two years of data or that belong to the financial sector) and with SIC code 49 (firms that belong to the regulated sector). We also do not consider quasi-governmental firms given by SIC code 9 and the U.S. Postal Service. We drop all observations that have any missing data in any of the variables that we use. We also delete all observations in which total assets are less than 10 million dollars, or in which sales or book assets increase more than 200%. We drop firms that have less than four years of consecutive observations.

The file contains four variables:

1. (Column 1) FirmID = the firm identifier. We have created anonymous firm identifiers, so this variable does NOT equal Compustat's gvkey.
2. (Column 2) Year = the time identifier
3. (Column 3) Profitability = Operating income / Capital. This variable corresponds to the ratio  $(zK^\alpha)/K$  in our model. Dollar profits,  $zK^\alpha$ , is measured by Compustat variable OIBDP (Operating Income Before Depreciation). Capital is measured as Compustat variable Assets Total (AT) at the beginning of the year.
4. (Column 4) Investment Rate = Investment / Capital. This variable corresponds to the ratio  $I/K$  in our model. Investment is measured as Compustat variable Capital Expenditure (CAPX).

We winsorize the variables in columns 2 and 3 at the 1st and 99th percentiles. For certain moments, you may prefer to use variables demeaned at the firm level. Specifically, you might want to demean a variable at the firm level by subtracting off each firm's time-series mean of the variable, then add back the variable's full-sample mean.