Online Appendices to the Corporate Propensity to Save

Appendix A: Monte Carlo Experiments

In order to allay skepticism of empirical results that have been produced by unusual estimators on fairly small samples, in Table AI we report the results of a Monte Carlo simulation using artificial data similar to our real data, both in terms of sample size and observable moments. These simulations are of particular interest because these estimators have most commonly been used on investment regressions instead of saving regressions, and because saving and investment have different statistical properties. Most important, the distribution of investment is highly skewed, whereas the distribution of saving is much more symmetric.

We do three experiments. For each we generate 10,000 simulated cross sections from

$$y_i = w_i\alpha + \chi_i\beta + u_i, \tag{A1}$$

$$x_i = \gamma + \chi_i + \varepsilon_i, \tag{A2}$$

in which χ_i is the true q of firm i, x_i is an estimate of its true q, y_i is the ratio of the change in cash to assets, and w_i is a row vector of perfectly measured regressors, whose first entry is 1, whose second entry is the ratio of cash flow to assets, and whose third entry is the natural log of total assets. The regression error, u_i, and the measurement error, ε_i, are assumed to be independent of each other and of (w_i, χ_i).

The first cross section has a sample size of 3,000, the second a sample size of 1,200, and the third a sample size of 200. These numbers correspond to the size of the largest and smallest cross sections in our data set, as well as to an intermediate size. For each simulation we set the parameters β, α, and τ^2 (the coefficient of determination of (A2)) approximately equal to the averages of the corresponding GMM estimates from Tables II and IV in the main text. Each observation is of the form (y_i, x_i, w_i), generated according to (A1)-(A2) so that (y_i, x_i, w_i) has, on average over the simulation samples, first and second moments equal to, and higher-order moments comparable with the corresponding average sample moments from our real U.S. data. χ_i and ε_i are distributed as chi-squared variables; u_i is distributed as a negative chi-squared variable; and w_i is distributed as the difference between two standard normals raised to the fourth power.

For the third-, fourth-, fifth-, and sixth-order GMM estimators, Table AI reports the mean value of the estimator of our parameter of interest, α_1, which corresponds to the coefficient on cash flow. It also reports its mean absolute deviation (MAD), the probability that an estimate is within 20% of its true value, and the actual size of a nominal 5% two-sided test of the null hypothesis that α_1 equals its true value. For the small and intermediate sample sizes Table AI shows that the fourth-order GMM estimator (GMM4) gives the best estimates in terms of expected value, MAD, and probability concentration. For the large sample size the GMM6 estimator performs best. Because the performance of the GMM4 and GMM6 estimators is similar for the large sample size, and because the J-test is most accurate for the GMM4 estimator, we therefore use the GMM4 estimator for our empirical work. Also of interest in this table are the tiny actual sizes of the test of the null hypothesis that α_1 equals its true value for the intermediate and large sample sizes. This result is the opposite of that found in Erickson and Whited (2000) for *investment* regressions.

Table AII explores the power of the J-test to detect misspecification. We examine four likely types of departures from the linear errors-in-variables model. Each is obtained by introducing one type of misspecification into the correctly specified baseline simulation described above. First, we make y_i depend nonlinearly on χ_i; second, we mismeasure the capital stock by multiplying $(y_i, x_{i,} w_i)$ from the baseline sample by an *i.i.d.* lognormal variable; third, we introduce a correlation between u_i and χ_i; and fourth, we violate the *i.i.d.* assumption by allowing different quintiles of our simulated observations to be generated by different distributions. We limit the degree of each misspecification so that the absolute biases in the GMM estimates of α_1 do not exceed 0.3, which is approximately the absolute value of the cash flow coefficient estimated in our real data. For the first three types of misspecification we find that the fourth, fifth, and sixth order GMM J-tests exhibits usefully large power for the largest sample size, ranging from 0.403 to 0.995. The test is more powerful for larger sample sizes, and all of the power figures are larger than the fractions of rejections we obtain in our empirical work. For the fourth type we find that the coefficient estimates are affected little, even though the J-test has lower power to detect non-*i.i.d.* samples. Finally, we do not combine misspecifications, which we suspect would further increase test power.

Real data contain a great deal of heterogeneity that cannot be captured by a model such as that given by (A1)-(A2). Although the fourth experiment in Table AII addresses this concern to some extent, we delve further into the issue by using our theoretical model to generate data containing heterogeneity along the lines of the cost of external finance and income uncertainty. We then use this simulated data as a laboratory in which we assess our estimators. To this end we first simulate sixteen groups of firms using our baseline simulation in Section II. We allow the linear cost of equity issuance to take the values 0, 0.04, 0.08, and 0.12, and for each of these values we allow the standard deviation of the model driving process to take the values 0.05, 0.10, 0.15, and 0.20. The upper end of the range for the cost of equity issuance comes from Hennessy and Whited (2007), and range of values for the standard deviation comes from estimates in our samples of large and small firms of the residual standard deviation of an $AR(1)$ model of the ratio of operating income to assets. To make our simulated data more closely resemble our actual data, we add an *i.i.d.* measurement error to Tobin's q that has a chi-squared distribution. Using a skewed distribution is important; otherwise observed Tobin's q often takes negative values.

We next determine whether our estimators can correctly identify different groups of firms via the estimates of the cash-flow coefficient in a regression of the ratio of saving to assets on Tobin's q, the ratio of cash flow to assets, and the natural log of assets. Because we are using simulated data we know exactly which firms are more constrained than others and exactly the amount of uncertainty each one faces. We make it difficult for our estimators to identify different groups of firms by constructing groups that are heterogeneous. Specifically, we split the sample at the median of the linear cost of equity issuance and then estimate the saving regression for each subsample. We do the same for groups of firms split by the amount of uncertainty they face. These subsamples themselves then contain simulated firms that are heterogeneous along the lines of the linear cost of equity issuance and uncertainty. It is important to note that this experiment tells us whether our estimators can distinguish groups of firms. It does not tell us whether our estimator is unbiased because the cash-flow coefficient is not a structural parameter in our model. We therefore cannot know its true value.

Our fourth-order GMM estimator produces a coefficient on cash flow of -0.17 for the group with less costly external finance and of -0.64 for the group with more costly external finance. This pattern corresponds to the prediction of our model. Next, we estimate a coefficient on cash flow of -0.26 for the group with high uncertainty and of -0.51 for the group with low uncertainty. This pattern also corresponds to the prediction of our model. We conclude that our estimators can distinguish groups of firms even when these groups contain unmodeled heterogeneity.

References

Erickson, Timothy and Toni M. Whited, 2000, Measurement error and the relationship between investment and q, *Journal of Political Economy* 108, 1027–57.

Hennessy, Christopher A. and Toni M. Whited, 2007, How costly is external financing? Evidence from a structural estimation, *Journal of Finance* 62, 1705–1745.

Table AI

Monte Carlo Performance of GMM and OLS Estimators

Indicated expectations and probabilities are estimates based on 10,000 Monte Carlo samples. The samples are generated by

$$\begin{aligned} y_i &= \chi_i \beta + w_i \alpha + u_i \\ x_i &= \gamma + \chi_i + \varepsilon_i, \end{aligned}$$

in which χ_i and ε_i are distributed as a chi-squared variables. u_i is distributed as a negative chi-squared variable. w_i is distributed as the difference between two standard normals raised to the fourth power. GMMn denotes the GMM estimator based on moments up to order $M = n$. y_i represents saving, χ_i represents true q, x_i represents an observable proxy for χ_i, w_i represents a vector of perfectly measured regressors. α is a vector of coefficients on these perfectly measured regressors. The coefficient of interest, α_1 is the coefficient on cash flow. OLS denotes estimates obtained by regressing y_i on x_i and w_i. MAD denotes mean absolute deviation. "T-Test Size" refers to the actual size of a nominal 5% test of the null hypothesis that α_1 equals its true value. "J-Test Size" refers to the actual size of a nominal 5% test of the overidentifying restrictions.

True Value: $\alpha_1 = -0.3$.

	OLS	GMM3	GMM4	GMM5	GMM6
Sample Size = 200					
$E(\hat{\alpha_1})$	0.180	-0.234	-0.223	-0.168	-0.201
MAD$(\hat{\alpha_1})$	0.481	0.409	0.299	0.326	0.342
P$(\mid \hat{\alpha_1} - \alpha_1 \mid \leq 0.2 \mid \alpha_1 \mid)$	0.001	0.085	0.130	0.150	0.130
T-test Size		0.032	0.062	0.074	0.090
J-test Size			0.235	0.329	0.529
Sample Size = 1200					
$E(\hat{\alpha_1})$	0.183	-0.349	-0.329	-0.275	-0.288
MAD$(\hat{\alpha_1})$	0.483	0.277	0.134	0.094	0.087
P$(\mid \hat{\alpha_1} - \alpha_1 \mid \leq 0.2 \mid \alpha_1 \mid)$	0.000	0.151	0.357	0.486	0.565
T-Test Size		0.001	0.008	0.007	0.011
J-test Size			0.136	0.323	0.394
Sample Size = 3000					
$E(\hat{\alpha_1})$	0.184	-0.348	-0.331	-0.292	-0.300
MAD$(\hat{\alpha_1})$	0.484	0.201	0.078	0.049	0.043
P$(\mid \hat{\alpha_1} - \alpha_1 \mid \leq 0.2 \mid \alpha_1 \mid)$	0.000	0.221	0.507	0.704	0.796
T-Test Size		0.000	0.002	0.001	0.001
J-test Size			0.096	0.317	0.418

Table AII

Monte Carlo Performance of the J-Test

The table reports the fraction of J-test rejections at a 5% nominal critical value. The samples are generated by

$$
\begin{aligned}
y_i &= \chi_i\beta + w_i\alpha + u_i \\
x_i &= \gamma + \chi_i + \varepsilon_i,
\end{aligned}
$$

in which χ_i and ε_i are distributed as a chi-squared variables. u_i is distributed as a negative chi-squared variable. w_i is distributed as the difference between two standard normals raised to the fourth power. GMMn denotes the GMM estimator based on moments up to order $M = n$. y_i represents saving, χ_i represents true q, x_i represents an observable proxy for χ_i, w_i represents a vector of perfectly measured regressors. GMMn denotes the GMM estimator based on moments up to order $M = n$.

	GMM4	GMM5	GMM6
Sample Size = 200			
Nonlinear Regression	0.275	0.289	0.661
Mismeasured Denominator	0.246	0.362	0.636
Correlated Error Regressor	0.410	0.425	0.671
Non-*i.i.d.* Sample	0.185	0.387	0.485
Sample Size = 1200			
Nonlinear Regression	0.424	0.558	0.848
Mismeasured Denominator	0.325	0.403	0.600
Correlated Error Regressor	0.542	0.743	0.918
Non-*i.i.d.* Sample	0.163	0.394	0.533
Sample Size = 3000			
Nonlinear Regression	0.504	0.725	0.960
Mismeasured Denominator	0.403	0.466	0.651
Correlated Error Regressor	0.676	0.936	0.995
Non-*i.i.d.* Sample	0.272	0.382	0.592

Appendix B: Supplementary Tables

These tables correspond to Tables II–V in the main text, except that the regression specifications here include firm fixed effects, whereas those in the main text do not. Fixed effects are removed by removing firm-level means of all variables, estimating each year of data separately, and then combining the yearly estimate using the Fama-MacBeth procedure.

Table BI

Fixed-Effects Saving Regressions: All Countries

Calculations are based on a sample of U.S. nonfinancial firms from Compustat from 1972 to 2006 and a sample of international firms from Global Vantage from 1994 to 2005. GMM estimates are from the fourth-order estimator in Erickson and Whited (2000). The dependent variable is the change in the stock of cash divided by total assets. CF stands for cash flow divided by total assets; q stands for the market-to-book ratio; and τ^2 is an index of measurement quality for the market-to-book ratio that varies between 0 and 1. Fama-MacBeth standard errors are below the estimates in parentheses. An asterisk indicates that the t-statistic exceeds the 5% bootstrapped critical value. A dagger indicates that the t-statistic exceeds the 5% asymptotic critical value.

	OLS			GMM4			
Country	q	CF	R^2	q	CF	R^2	τ^2
United States	0.021*†	0.110*†	0.093*†	0.281*†	-0.483*†	0.378*†	0.260*†
	(0.002)	(0.008)	(0.008)	(0.016)	(0.055)	(0.023)	(0.017)
Canada	0.055*†	0.108*†	0.145*†	0.195*†	-0.029	0.372*†	0.375*†
	(0.006)	(0.040)	(0.023)	(0.027)	(0.027)	(0.053)	(0.037)
United Kingdom	0.010†	0.093*†	0.049*†	0.318	-0.034*†	0.215*†	0.118*†
	(0.002)	(0.019)	(0.015)	(0.073)	(0.005)	(0.025)	(0.024)
Japan	0.023*†	0.102*†	0.033*†	0.354*†	0.055	0.195*†	0.290*†
	(0.003)	(0.011)	(0.005)	(0.092)	(0.034)	(0.082)	(0.046)
France	0.025*†	0.151*†	0.100*†	0.229*†	-0.135	0.260*†	0.226*†
	(0.006)	(0.038)	(0.024)	(0.094)	(0.130)	(0.083)	(0.060)
Germany	0.020†	0.061*†	0.075*†	0.245*†	-0.057*†	0.226*†	0.266*†
	(0.006)	(0.021)	(0.017)	(0.065)	(0.016)	(0.065)	(0.065)

Table BII

Yearly Fixed Effects Saving Regressions Summary: All Countries

Calculations are based on a sample of U.S. firms from Compustat from 1972 to 2006 and a sample of international firms from Global Vantage from 1994 to 2005. The first column contains the fraction of the yearly estimates that are negative. The second column contains the fraction of the yearly estimates that are significantly negative at the 5% level, using asymptotic critical values. The third column contains the fraction of the yearly tests of overidentifying restrictions that produce rejections at the 5% level. The fourth column contains the fraction of the yearly identification tests that produce rejections at the 5% level.

	Fraction of Negative Cash Flow Coefficients	Fraction of Significant Negative Cash Flow Coefficients	Fraction of Overidentifying Restriction Rejections	Fraction of Identification Test Rejections
United States	0.971	0.800	0.143	0.857
Canada	0.500	0.083	0.000	0.083
United Kingdom	0.500	0.083	0.167	0.417
Japan	0.250	0.000	0.167	0.500
France	0.500	0.083	0.000	0.250
Germany	0.500	0.167	0.167	0.083

Table BIII

Split Sample Fixed Effects Regressions: United States

Calculations are based on a sample of U.S. nonfinancial firms from Compustat from 1972 to 2006. GMM estimates are from the fourth-order estimator in Erickson and Whited (2000). The dependent variable is the change in the stock of cash divided by total assets. CF stands for cash flow divided by total assets; q stands for the market-to-book ratio; and τ^2 is an index of measurement quality for the market-to-book ratio that varies between 0 and 1. Serial correlation is the first-order autoregressive coefficient on the ratio of operating income to assets, and standard deviation is the standard deviation of the residual from this regression. Fama-MacBeth standard errors are below the estimates in parentheses. An asterisk indicates that the t-statistic exceeds the 5% bootstrapped critical value. A dagger indicates that the t-statistic exceeds the 5% asymptotic critical value.

	OLS			GMM4			
Subsample	q	CF	R^2	q	CF	R^2	τ^2
Small	0.033*†	0.137*†	0.134*†	0.248*†	-0.221*†	0.493*†	0.428*†
	(0.003)	(0.011)	(0.012)	(0.020)	(0.089)	(0.030)	(0.159)
Large	0.006*†	0.082*†	0.047*†	0.205*†	-0.531*†	0.186*†	0.359*†
	(0.001)	(0.009)	(0.007)	(0.062)	(0.220)	(0.027)	(0.018)
No Bond Rating	0.023*†	0.117*†	0.101*†	0.298*†	-0.481*†	0.406*†	0.255*†
	(0.002)	(0.009)	(0.009)	(0.062)	(0.198)	(0.038)	(0.017)
Bond Rating	0.012*†	0.070†	0.057*†	0.279*†	-0.974*†	0.221*†	0.379
	(0.002)	(0.010)	(0.007)	(0.057)	(0.230)	(0.041)	(0.022)
High Standard Deviation	0.028*†	0.128*†	0.126*†	0.233*†	-0.261*†	0.426*†	0.262*†
	(0.003)	(0.008)	(0.013)	(0.127)	(0.064)	(0.033)	(0.019)
Low Standard Deviation	0.012*†	0.080*†	0.055*†	0.186*†	-0.533*†	0.235*†	0.280
	(0.002)	(0.010)	(0.007)	(0.067)	(0.245)	(0.033)	(0.632)
High Serial Correlation	0.019*†	0.083*†	0.085*†	0.337*†	-0.929*†	0.393*†	0.366*†
	(0.002)	(0.007)	(0.007)	(0.058)	(0.179)	(0.047)	(0.019)
Low Serial Correlation	0.025*†	0.125*†	0.104*†	0.261*†	-0.246	0.406*†	0.186*†
	(0.003)	(0.009)	(0.008)	(0.035)	(0.074)	(0.039)	(0.050)

Table BIV

Yearly Saving Fixed Effects Regressions Summary: Split Samples

Calculations are based on a sample of U.S. nonfinancial firms from Compustat from 1972 to 2004. The first column contains the fraction of the yearly estimates that are negative. The second column contains the fraction of the yearly estimates that are significantly negative at the 5% level, using asymptotic critical values. The third column contains the fraction of the yearly tests of overidentifying restrictions that produce rejections at the 5% level. The fourth column contains the fraction of the yearly identification tests that produce rejections at the 5% level.

	Fraction of Negative Cash Flow Coefficients	Fraction of Significant Negative Cash Flow Coefficients	Fraction of Overidentifying Restriction Rejections	Fraction of Identification Test Rejections
Small	0.743	0.257	0.057	0.629
Large	0.800	0.400	0.057	0.514
No Bond Rating	0.829	0.686	0.114	0.800
Bond Rating	0.829	0.314	0.086	0.571
High Standard Deviation	0.800	0.457	0.114	0.457
Low Standard Deviation	0.686	0.486	0.229	0.457
High Serial Correlation	0.943	0.543	0.114	0.457
Low Serial Correlation	0.829	0.371	0.057	0.571