Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches

Mitchell A. Petersen
Kellogg School of Management, Northwestern University
and NBER

Supplementary Tables

Table of Contents

Supplementary Table 1: Estimating Probit Coefficients and Standard Errors with a Firm Effect
Standard ("OLS") and Clustered Standard Errors

Avg(β_{OLS}) Std(β_{OLS}) Avg(SE_{OLS}) % Sig(T_{OLS}) Avg(SE_C) % Sig(T_C)	Source of Independent Variable Volatility			
Source of Residual Volatility	0%	25%	50%	75%
0%	1.0018	1.0008	0.9996	1.0008
	0.0404	0.0408	0.0396	0.0408
	0.0403	0.0403	0.0403	0.0404
	[0.0110]	[0.0090]	[0.0102]	[0.0126]
	0.0403	0.0403	0.0403	0.0403
	[0.0112]	[0.0094]	[0.0104]	[0.0134]
25%	1.0013	0.9999	1.0009	1.0017
	0.0408	0.0461	0.0517	0.0565
	0.0403	0.0403	0.0404	0.0404
	[0.0114]	[0.0250]	[0.0462]	[0.0670]
	0.0410	0.0462	0.0510	0.0557
	[0.0100]	[0.0086]	[0.0100]	[0.0132]
50%	1.0021	1.0002	1.0009	1.0022
	0.0435	0.0544	0.0619	0.0689
	0.0403	0.0403	0.0404	0.0404
	[0.0180]	[0.0574]	[0.0938]	[0.1296]
	0.0434	0.0529	0.0614	0.0694
	[0.0112]	[0.0116]	[0.0126]	[0.0102]
75%	1.0010	1.0021	1.0023	1.0040
	0.0489	0.0623	0.0733	0.0856
	0.0403	0.0404	0.0404	0.0404
	[0.0346]	[0.0946]	[0.1558]	[0.2232]
	0.0482	0.0611	0.0726	0.0837
	[0.0122]	[0.0112]	[0.0114]	[0.0140]

Notes:

The table contains estimates of the coefficient and standard errors from a probit model. As in Table 1, the results are based on 5,000 simulated panel data set each of which contains 500 firms and 10 years per firm. The independent variable and the residual are simulated as described in the paper (see equations 4 and 5). The latent variable, y, is created as described in equation 1. The dependent variable is coded as equal to one if y is greater than zero and zero otherwise. Since probit coefficients are only estimated up to a scaling factor, I scaled the data so that the true coefficient is one. The fraction of X's variance which is due to a firm specific component varies across the columns of the table from 0 percent (no firm effect) to 75 percent and the fraction of the residual variance which is due to a firm specific component varies across the rows of the table from 0 percent (no firm effect) to 75 percent. Each cell contains the average slope coefficient estimated by maximum likelihood and the standard deviation of this estimate. This is the true standard error of the estimated coefficient. The third entry is the unadjusted standard error estimated by maximum likelihood under the assumption that the residuals are independent. The percent of t-statistics based on this standard error which are significant at the one percent level (e.g. $|t|>2.58$) is reported in square brackets. The fifth entry is the average standard error clustered by firm (i.e. accounts for the possible correlation between observations of the same firm in different years). The percent of clustered t-statistics which are significant at the one percent level is reported in square brackets.

Supplementary Table 2: Estimating Tobit Coefficients and Standard Errors with a Firm Effect
Standard (“OLS”) and Clustered Standard Errors

Avg(β_{OLS}) Std(β_{OLS}) Avg(SE_{OLS}) % Sig(T_{OLS}) Avg(SE_C) % Sig(T_C)	Source of Independent Variable Volatility			
Source of Residual Volatility	0%	25%	50%	75%
0%	1.0006	1.0001	0.9997	1.0002
	0.0303	0.0304	0.0303	0.0306
	0.0302	0.0302	0.0303	0.0303
	[0.0088]	[0.0114]	[0.0106]	[0.0116]
	0.0302	0.0302	0.0302	0.0302
	[0.0074]	[0.0112]	[0.0114]	[0.0112]
25%	1.0008	0.9995	0.9993	1.0004
	0.0301	0.0367	0.0427	0.0484
	0.0302	0.0302	0.0302	0.0303
	[0.0082]	[0.0328]	[0.0608]	[0.1102]
	0.0302	0.0370	0.0427	0.0479
	[0.0080]	[0.0108]	[0.0080]	[0.0118]
50%	1.0005	1.0000	0.9992	0.9991
	0.0299	0.0440	0.0530	0.0605
	0.0302	0.0302	0.0302	0.0302
	[0.0092]	[0.0746]	[0.1346]	[0.2008]
	0.0303	0.0429	0.0526	0.0608
	[0.0098]	[0.0114]	[0.0134]	[0.0092]
75%	0.9999	1.0001	1.0001	1.0001
	0.0310	0.0492	0.0623	0.0730
	0.0302	0.0302	0.0302	0.0302
	[0.0130]	[0.1204]	[0.2038]	[0.2882]
	0.0306	0.0486	0.0613	0.0722
	[0.0122]	[0.0098]	[0.0122]	[0.0118]

Notes:

The table contains estimates of the coefficient and standard errors from a tobit model. As in Table 1, the results are based on 5,000 simulated panel data set each of which contains 500 firms and 10 years per firm. The independent variable and the residual are simulated as described in the paper (see equations 4 and 5). The latent variable, y, is created as described in equation 1. The dependent variable is equal to y if y is greater than -1.5 and equal to -1.5 otherwise. Given the variance of y is 5, this means that on average 25 percent of the sample will be truncated. The fraction of X's variance which is due to a firm specific component varies across the columns of the table from 0 percent (no firm effect) to 75 percent and the fraction of the residual variance which is due to a firm specific component varies across the rows of the table from 0 percent (no firm effect) to 75 percent. Each cell contains the average slope coefficient estimated by maximum likelihood and the standard deviation of this estimate. This is the true standard error of the estimated coefficient. The third entry is the unadjusted standard error estimated by maximum likelihood under the assumption that the residuals are independent. The percent of t-statistics based on this standard error which are significant at the one percent level (e.g. $|t|>2.58$) is reported in square brackets. The fifth entry is the average standard error clustered by firm (i.e. accounts for the possible correlation between observations of the same firm in different years). The percent of clustered t-statistics which are significant at the one percent level is reported in square brackets.

Supplementary Table 3: Estimating Logit, Probit, and Tobit Models with Standard ("OLS") and Clustered Standard Errors
(data contains a firm and time effect)

	Type of Estimation		
	Logit	Probit	Tobit
Avg(β_{OLS})	1.0315	1.0067	1.0012
Std(β_{OLS})	0.1092	0.1024	0.0957
Avg(SE_{OLS})	0.0308	0.0287	0.0214
% Sig(T_{OLS})	[0.4823]	[0.4679]	[0.5747]
Avg(SE_{F})	0.0794	0.0745	0.0679
% Sig(T_{F})	[0.0765]	[0.0687]	[0.0763]
Avg(SE_{T})	0.0794	0.0744	0.0678
% Sig(T_{T})	[0.0780]	[0.0697]	[0.0755]
Avg(SE_{FT})	0.1085	0.1017	0.0970
% Sig(T_{FT})	[0.0141]	[0.0124]	[0.0119]

Notes:

The table contains estimates of the coefficient and standard errors from a logit, a probit, and a tobit model. The results are based on 5,000 simulated panel data set each of which contains 100 firms and 100 years per firm (similar to the intermediate example in Figure 7). The simulated data contains both a firm effect and a time effect. One third of the variability of the residual and the independent variable is due to the firm effect and one third of the variability is due to the time effect [e.g. $V(\gamma)=V(\delta)=Var(\eta)$ and $Var(\mu)= V(\zeta)=V(\nu)$, see equation (17) in the paper]. The latent variable, y, is created as described in equation (1). In the case of the tobit model, the dependent variable is equal to y if y is greater than -1.5 and equal to -1.5 otherwise. Given the variance of y is 5, this means that on average 25 percent of the sample will be truncated. In the case of the logit and probit models, the dependent variable is equal to 1 if y is greater than zero and zero otherwise. Since probit and logit coefficients are only estimated up to a scaling factor, I scaled the data so that the true coefficient is one.

Each cell contains the average slope coefficient estimated by maximum likelihood and the standard deviation of this estimate. This is the true standard error of the estimated coefficient. The third entry is the unadjusted ("OLS") standard error estimated by maximum likelihood under the assumption that the residuals are independent. The percent of t-statistics based on this standard error which are significant at the one percent level (e.g. $|t|>2.58$) is reported in square brackets. The fifth and sixth entry are the average standard error clustered by firm (i.e. accounts for the possible correlation between observations of the same firm in different years) and the percent of t-statistics (clustered by firm) which are significant at the one percent level (reported in square brackets). The seventh and eighth entry are the average standard error clustered by time (i.e. accounts for the possible correlation between observations of the same year in different firms) and the percent of t-statistics (clustered by time) which are significant at the one percent level (reported in square brackets). The ninth and tenth entry are the average standard error clustered by firm and time (i.e. accounts for the possible correlation between observations of the same firm in different years and possible correlation between observations of the same year in different firms) and the percent of t-statistics (clustered by firm and time) which are significant at the one percent level (reported in square brackets).

Since the data has both a firm and time effect, the standard errors clustered by only firm or time are biased, and the unclustered standard errors are even more biased. The standard errors clustered by firm and time are unbiased and produce p-values which are correct.

Supplementary Table 4: Estimating Standard Errors with a Firm Effect
OLS, White, and Clustered Standard Errors

Avg(β_{OLS}) Std(β_{OLS}) Avg(SE_{OLS}) % Sig(T_{OLS}) Avg(SE_W) % Sig(T_W) Avg(SE_C) % Sig(T_C)	Source of Independent Variable Volatility			
	0%	25%	50%	75%
Source of Residual Volatility: 0%	1.0004	1.0003	1.0001	0.9997
	0.0286	0.0284	0.0278	0.0285
	0.0283	0.0283	0.0283	0.0283
	[0.0098]	[0.0092]	[0.0076]	[0.0128]
	0.0283	0.0283	0.0283	0.0283
	[0.0098]	[0.0088]	[0.0078]	[0.0122]
	0.0283	0.0282	0.0282	0.0283
	[0.0108]	[0.0096]	[0.0082]	[0.0126]
25%	1.0003	1.0001	1.0000	1.0004
	0.0291	0.0351	0.0414	0.0456
	0.0283	0.0283	0.0283	0.0283
	[0.0122]	[0.0396]	[0.0764]	[0.1068]
	0.0283	0.0283	0.0283	0.0283
	[0.0124]	[0.0384]	[0.0752]	[0.1074]
	0.0282	0.0353	0.0411	0.0462
	[0.0122]	[0.0112]	[0.0110]	[0.0126]
50%	0.9999	1.0011	1.0004	1.0014
	0.0283	0.0415	0.0514	0.0598
	0.0283	0.0283	0.0283	0.0283
	[0.0092]	[0.0784]	[0.1544]	[0.2176]
	0.0283	0.0283	0.0283	0.0283
	[0.0090]	[0.0786]	[0.1552]	[0.2202]
	0.0283	0.0411	0.0508	0.0590
	[0.0096]	[0.0096]	[0.0116]	[0.0130]
75%	1.0005	1.0006	1.0000	0.9999
	0.0283	0.0460	0.0598	0.0690
	0.0283	0.0283	0.0283	0.0283
	[0.0096]	[0.1094]	[0.2218]	[0.2914]
	0.0283	0.0283	0.0282	0.0283
	[0.0098]	[0.1106]	[0.2232]	[0.2934]
	0.0283	0.0462	0.0589	0.0695
	[0.0096]	[0.0106]	[0.0124]	[0.0096]

Notes:
The table contains estimates of the coefficient and standard errors based on 5,000 simulated panel data set each of which contains 500 firms and 10 years per firm. The true slope coefficient is 1, the standard deviation of the independent variable is 1 and the standard deviation of the error term is 2 (see equation 1). The independent variable and the residual are simulated as described in the paper (see equations 4 and 5). The fraction of X's variance which is due to a firm specific component [Var(μ)/Var(X)] varies across the columns of the table from 0 percent (no firm effect) to 75 percent and the fraction of the residual variance which is due to a firm specific component [Var(γ)/Var(ε)] varies across the rows of the table from 0 percent (no firm effect) to 75 percent. Each cell contains the average slope coefficient estimated by OLS and the standard deviation of this estimate. This is the true standard error of the estimated coefficient. The third entry is the average standard error estimated by OLS. The percent of OLS t-statistics which are significant at the one percent level (e.g. |t|>2.58) is reported in square brackets. The fifth entry is the average white error clustered by firm (i.e. standard errors which are robust to heteroscedasticty, but which are not robust to within cluster correlation). The percent of clustered t-statistics which are significant at the one percent level is reported in square brackets. The seventh entry is the average standard error clustered by firm (i.e. accounts for the possible correlation between observations of the same firm in different years). The percent of clustered t-statistics which are significant at the one percent level is reported in square brackets.

White standard errors, like OLS standard errors, are biased when there is a firm effect in the data. In this case, the White standard errors are identical to the OLS standard errors since the simulated data is homoscedastic.

Supplementary Table 5: Estimating the Intercept and its Standard Errors with a Firm Effect
OLS and Clustered Standard Errors

Avg(β_{OLS}) Std(β_{OLS}) Avg(SE_{OLS}) % Sig(T_{OLS}) Avg(SE_C) % Sig(T_C)		Source of Independent Variable Volatility			
		0%	25%	50%	75%
Source of Residual Volatility	0%	-0.0005	-0.0005	-0.0000	-0.0001
		0.0283	0.0283	0.0287	0.0287
		0.0283	0.0283	0.0283	0.0283
		[0.0086]	[0.0114]	[0.0104]	[0.0106]
		0.0283	0.0283	0.0283	0.0283
		[0.0092]	[0.0114]	[0.0106]	[0.0116]
	25%	-0.0004	-0.0006	-0.0004	-0.0001
		0.0504	0.0511	0.0506	0.0507
		0.0283	0.0283	0.0283	0.0283
		[0.1474]	[0.1532]	[0.1488]	[0.1516]
		0.0510	0.0510	0.0510	0.0510
		[0.0100]	[0.0100]	[0.0106]	[0.0108]
	50%	-0.0002	0.0003	0.0012	-0.0004
		0.0670	0.0662	0.0662	0.0664
		0.0283	0.0283	0.0283	0.0283
		[0.2796]	[0.2722]	[0.2704]	[0.2682]
		0.0663	0.0663	0.0663	0.0663
		[0.0114]	[0.0104]	[0.0090]	[0.0114]
	75%	0.0011	-0.0011	0.0022	0.0002
		0.0795	0.0785	0.0796	0.0786
		0.0283	0.0283	0.0283	0.0283
		[0.3664]	[0.3554]	[0.3620]	[0.3588]
		0.0787	0.0787	0.0787	0.0787
		[0.0110]	[0.0098]	[0.0102]	[0.0092]

Notes:

The table contains estimates of the intercept and standard errors based on 5,000 simulated panel data set each of which contains 500 firms and 10 years per firm. The independent variable and the residual are simulated as described in the paper (see equations 4 and 5). The true intercept is zero. The fraction of X's variance which is due to a firm specific component varies across the columns of the table from 0 percent (no firm effect) to 75 percent and the fraction of the residual variance which is due to a firm specific component varies across the rows of the table from 0 percent (no firm effect) to 75 percent. Each cell contains the average intercept coefficient estimated by OLS and the standard deviation of this estimate. This is the true standard error of the estimated coefficient. The third entry is the average standard error estimated by OLS. The percent of OLS t-statistics which are significant at the one percent level (e.g. $|t|>2.58$) is reported in square brackets. The fifth entry is the average standard error clustered by firm (i.e. accounts for the possible correlation between observations of the same firm in different years). The percent of clustered t-statistics which are significant at the one percent level is reported in square brackets.

The results in this table differ from the results and intuition in Table 1 of the paper, even though the data structure is the same. In this table, the OLS standard errors are biased when the residuals are correlated, independent of the correlation of the independent variable (X in equation 1). This table, however, reports the standard errors of the intercept, not the slope. The intercept variable is always one for every firm, and thus the persistence of this variable is very high. This is why the bias in the standard error of the intercept is a function only of the serial correlation in the residual.

Supplementary Table 6: Mean Squared Error of Standard Error Estimation with a Firm Effect
Panel A: OLS Standard Errors

MSE Variance Bias^2 99% CI		Source of Independent Variable Volatility			
		0%	25%	50%	75%
Source of Residual Volatility	0%	0.0000 0.0000 0.0000 [0.9902]	0.0000 0.0000 0.0000 [0.9912]	0.0000 0.0000 0.0000 [0.9906]	0.0001 0.0001 0.0000 [0.9906]
	25%	0.0000 0.0000 0.0000 [0.9884]	0.0050 0.0000 0.0050 [0.9652]	0.0168 0.0000 0.0167 [0.9322]	0.0326 0.0001 0.0325 [0.8826]
	50%	0.0000 0.0000 0.0000 [0.9876]	0.0168 0.0000 0.0168 [0.9230]	0.0516 0.0001 0.0516 [0.8466]	0.0953 0.0001 0.0952 [0.7924]
	75%	0.0001 0.0001 0.0000 [0.9872]	0.0328 0.0001 0.0328 [0.8910]	0.0954 0.0001 0.0954 [0.7770]	0.1712 0.0001 0.1711 [0.7094]

Notes: The first entry of the table contains the mean squared error (MSE) of the OLS standard errors. The MSE is the sum of the variance of the estimate and the bias squared, which are reported in the second and third entries in the table. The MSE, the variance of the estimate and the bias squared are all multiplied by 100. The true value of the standard error is the reported in equation (7) of the paper and depends upon the fraction of the variance of the independent and the residual which is due to the firm effect (e.g. ρ_x and ρ_r). The underlying data structure is the same as in Table 1.

$$\begin{aligned} MSE &= E\left[(\hat{\sigma}-\sigma)^2\right] \\ &= E\left[(\hat{\sigma}-\bar{\sigma})^2-(\bar{\sigma}-\sigma)^2\right] \\ &= Var(\hat{\sigma})-Bias(\hat{\sigma})^2 \end{aligned} \qquad \textbf{(1)}$$

The fourth entry in the tables is the coverage ratio of the 99% confidence interval. This is the percent of t-statistics (based on the estimated coefficients and standard error) which are less than 2.58 in absolute value. If the standard errors are correct, this should be 0.9900.

Supplementary Table 6: Mean Squared Error of Standard Error Estimation with a Firm Effect
Panel B: Clustered Standard Errors

MSE Variance $Bias^2$ 99% CI		Source of Independent Variable Volatility			
		0%	25%	50%	75%
Source of Residual Volatility	0%	0.0001 0.0001 0.0000 [0.9892]	0.0001 0.0001 0.0000 [0.9908]	0.0002 0.0002 0.0000 [0.9904]	0.0002 0.0002 0.0000 [0.9902]
	25%	0.0001 0.0001 0.0000 [0.9880]	0.0003 0.0003 0.0000 [0.9936]	0.0004 0.0004 0.0000 [0.9888]	0.0006 0.0006 0.0000 [0.9882]
	50%	0.0002 0.0002 0.0000 [0.9872]	0.0004 0.0004 0.0000 [0.9886]	0.0007 0.0007 0.0000 [0.9912]	0.0012 0.0012 0.0000 [0.9898]
	75%	0.0002 0.0002 0.0000 [0.9872]	0.0007 0.0007 0.0000 [0.9886]	0.0010 0.0010 0.0000 [0.9906]	0.0016 0.0016 0.0000 [0.9888]

Notes: The first entry of the table contains the mean squared error (MSE) of the standard errors clustered by firm. The MSE is the sum of the variance of the estimate and the bias squared, which are reported in the second and third entries in the table. The MSE, the variance of the estimate and the bias squared are all multiplied by 100. The true value of the standard error is the reported in equation (7) of the paper and depends upon the fraction of the variance of the independent and the residual which is due to the firm effect (e.g. ρ_x and ρ_r). The underlying data structure is the same as in Table 1. The fourth entry in the tables is the coverage ratio of the 99% confidence interval. This is the percent of t-statistics (based on the estimated coefficients and standard error) which are less than 2.58 in absolute value. If the standard errors are correct, this should be 0.9900.

Supplementary Table 6: Mean Squared Error of Standard Error Estimation with a Firm Effect
Panel C: Fama-MacBeth Standard Errors

MSE Variance Bias^2 99% CI		Source of Independent Variable Volatility			
		0%	25%	50%	75%
Source of Residual Volatility	0%	0.0044 0.0044 0.0000 [0.9712]	0.0044 0.0043 0.0000 [0.9696]	0.0044 0.0044 0.0000 [0.9764]	0.0046 0.0046 0.0001 [0.9706]
	25%	0.0044 0.0043 0.0001 [0.9664]	0.0112 0.0040 0.0073 [0.9242]	0.0273 0.0038 0.0235 [0.8798]	0.0492 0.0035 0.0456 [0.8082]
	50%	0.0044 0.0044 0.0000 [0.9670]	0.0273 0.0039 0.0234 [0.8736]	0.0773 0.0031 0.0742 [0.7540]	0.1414 0.0028 0.1386 [0.6612]
	75%	0.0044 0.0044 0.0000 [0.9690]	0.0502 0.0037 0.0465 [0.8222]	0.1424 0.0029 0.1395 [0.6346]	0.2656 0.0020 0.2636 [0.5006]

Notes: The first entry of the table contains the mean squared error (MSE) of the Fama-MacBeth standard errors. The MSE is the sum of the variance of the estimate and the bias squared, which are reported in the second and third entries in the table. The MSE, the variance of the estimate and the bias squared are all multiplied by 100. The true value of the standard error is the reported in equation (7) of the paper and depends upon the fraction of the variance of the independent and the residual which is due to the firm effect (e.g. ρ_x and ρ_r). The underlying data structure is the same as Table 2 in the paper. The fourth entry in the tables is the coverage ratio of the 99% confidence interval. This is the percent of t-statistics (based on the estimated coefficients and standard error) which are less than 2.58 in absolute value. If the standard errors are correct, this should be 0.9900.

Supplementary Table 7: Bootstrapped Standard Errors Estimates

	100 Samples Drawn	1000 Samples Drawn
Avg(β_{OLS})	1.0002	1.0000
Std(β_{OLS})	0.0513	0.0518
Avg(SE_{OLS})	0.0283	0.0283
Avg(SE_{C})	0.0508	0.0508
Avg($SE_{BT\text{-}NC}$)	0.0282	0.0282
Avg($SE_{BT\text{-}C}$)	0.0505	0.0505

Notes:

The table contains estimates of the coefficient and standard errors based on 5,000 simulated panel data set each of which contains 500 firms and 10 years per firm. The independent variable and the residual are simulated as described in the paper (see equations 4 and 5). The firm effect in both the independent variable and the residual are assumed to The fraction of independent variable's and the residual's variance which is due to a firm specific component is assumed to be 50 percent. The first four entries are similar to Table 1. The first two entries are the average slope coefficient estimated by OLS and the standard deviation of this estimate. The third and fourth entry are the average OLS standard errors and the average standard error clustered by firm.

The last two entries contain bootstrapped standard errors. For each iteration, I drew 100 (column I) or 1000 (column II) additional samples, with replacement, from the 5,000 observations in the data set and then re-estimated the β. The boot strap standard error is the standard deviation across these 100 or 1,000 estimated βs. When drawing observations, they can be drawn one observation at a time (which assumes the observations are independent of each other) or one cluster (firm) at a time (which allows for within cluster correlation if it exists in the data). The fifth entry is the bootstrap standard error, when observations are drawn independently, and these are very close to the OLS standard errors. The sixth entry is the bootstrap standard error, when observations are drawn as a cluster (firm) and these are close to the standard errors clustered by firm.

Supplementary Table 8: Estimating Standard Errors with a Firm and Time Effect
OLS and Clustered Standard Errors with Time Dummies

Avg(β_{OLS}) Std(β_{OLS}) Avg(SE_{OLS}) % Sig(T_{OLS}) Avg(SE_C) % Sig(T_C)		Independent Variable Volatility from Firm Effect		
		25%	25%	50%
		Independent Variable Volatility from Time Effect		
Residual Volatility from Firm Effect	**Residual Volatility from Time Effect**	25%	50%	25%
25%	25%	0.9997 0.0407 0.0283 [0.0754] 0.0400 [0.0112]	1.0004 0.0547 0.0347 [0.0996] 0.0548 [0.0100]	1.0004 0.0489 0.0283 [0.1312] 0.0489 [0.0084]
25%	50%	1.0005 0.0362 0.0231 [0.1032] 0.0364 [0.0104]	1.0015 0.0515 0.0283 [0.1542] 0.0508 [0.0122]	0.9993 0.0468 0.0231 [0.2116] 0.0461 [0.0078]
50%	25%	1.0002 0.0493 0.0283 [0.1382] 0.0490 [0.0108]	1.0008 0.0690 0.0347 [0.1930] 0.0692 [0.0120]	0.9994 0.0631 0.0283 [0.2418] 0.0630 [0.0128]

Notes:

The table contains estimates of the coefficient and standard errors based on 5,000 simulation of a panel data set with 5,000 observations. The true slope is 1, the standard deviation of the independent variable is 1 and the standard deviation of the error term is 2. Each data set contains both a fixed firm and a time effect. Each regression includes a full set of time dummies. The proportion of the variance of the independent variable and the residual which is due to the firm effect is either 25 or 50 percent. The proportion which is due to the time effect is also 25 or 50 percent. For example, in the bottom left cell 25 percent of the variability in the independent variable is from the firm effect and 25 percent is from the time effect. 50 percent of the variability of the residual is from the firm effect and 25 percent is from the time effect. Each cell contains the average estimated slope coefficient from OLS and the standard deviation of this estimate. This is the true standard error of the estimated coefficient. The third entry is the average standard error estimated from OLS. The percent of OLS t-statistics which are significant at the one percent level (e.g. t>2.58) is reported in square brackets. The fifth entry is the average standard error clustered by firm (i.e. accounts for the possible correlation between observations of the same firm in different years). The percent of clustered t-statistics which are significant at the one percent level (e.g. t>2.58) is reported in square brackets. Each regression includes nine year dummies.

Supplementary Table 9: Estimating Standard Errors with a Firm and Time Effect
Fama-MacBeth Standard Errors

Avg(β_{FM}) Std(β_{FM}) Avg(SE_{FM}) % Sig(T_{FM})				Independent Variable Volatility from Firm Effect		
				25%	25%	50%
				Independent Variable Volatility from Time Effect		
				25%	50%	25%
Residual Volatility from Firm Effect	25%	Residual Volatility from Time Effect	25%	0.9997 0.0407 0.0258 [0.1290]	1.0004 0.0547 0.0309 [0.1698]	1.0004 0.0489 0.0243 [0.2210]
	25%		50%	1.0005 0.0362 0.0206 [0.1626]	1.0015 0.0515 0.0239 [0.2502]	0.9993 0.0469 0.0185 [0.3346]
	50%		25%	1.0002 0.0493 0.0244 [0.2160]	1.0008 0.0691 0.0275 [0.3138]	0.9994 0.0632 0.0206 [0.4032]

Notes:

The table contains estimates of the coefficient and standard errors based on 5,000 simulation of a panel data set with 5,000 observations (500 firms and 10 years per firm). The true slope coefficient is 1, the standard deviation of the independent variable is 1 and the standard deviation of the error term is 2. Each data set contains both a fixed firm and a time effect. The proportion of the variance of the independent variable and the residual which is due to the firm effect is either 25 or 50 percent. The proportion which is due to the time effect is also 25 or 50%. For example, in the bottom left cell 25 percent of the variability in the independent variable is from the firm effect and 25 percent is from the time effect. 50 percent of the variability of the residual is from the firm effect and 25 percent is from the time effect. The first entry in each cell is the average estimated slope coefficient based on a Fama-MacBeth estimation (e.g. the regression is run for each of the 10 years and the estimate is the average of the 10 estimated slope coefficients). The second entry is the standard deviation of this coefficient. This is the true standard error of the Fama-MacBeth coefficient. The third entry is the average standard error estimated by the Fama-MacBeth procedure (e.g. equation 9). The percent of Fama-MacBeth t-statistics which are significant at the one percent level (e.g. t>2.58) is reported in square brackets.

Supplementary Table 10: Fama-MacBeth Standard Errors by Year
(Average of Time Series Regression Coefficients)

Avg(β_{FM}) Std(β_{FM}) Avg(SE_{FM}) % Sig(T_{FM})		Source of Independent Variable Volatility			
		0%	25%	50%	75%
Source of Residual Volatility	0%	1.0011 0.0335 0.0338 [0.0094]	1.0001 0.0396 0.0390 [0.0116]	1.0004 0.0479 0.0477 [0.0096]	1.0004 0.0679 0.0676 [0.0082]
	25%	1.0009 0.0294 0.0293 [0.0096]	0.9997 0.0332 0.0338 [0.0072]	1.0001 0.0413 0.0414 [0.0100]	1.0004 0.0580 0.0585 [0.0080]
	50%	1.0006 0.0235 0.0239 [0.0080]	1.0003 0.0275 0.0276 [0.0096]	0.9996 0.0345 0.0338 [0.0124]	0.9989 0.0474 0.0477 [0.0106]
	75%	1.0002 0.0167 0.0169 [0.0088]	1.0002 0.0193 0.0195 [0.0100]	1.0003 0.0243 0.0239 [0.0108]	1.0005 0.0336 0.0337 [0.0102]

Notes:

The table contains estimates of the Fama-MacBeth coefficient and standard errors based on 5000 simulation of a panel data set (10 years per firm and 500 firms). The true slope coefficient is 1, the standard deviation of the independent variable is 1 and the standard deviation of the error term is 2. The fraction of the residual variance which is due to a firm specific component is varied across the rows of the table and varies from 0% (no firm effect) to 75%. The fraction of the independent variable's variance which is due to a firm specific component is varied across the columns of the table and varies from 0% (no firm effect) to 75%. The first entry is the average estimated slope coefficient based on a Fama-MacBeth estimation. However these are not traditional Fama-MacBeth estimates. To calculate the coefficient and standard error estimates, I ran a time series regression for each of the 500 firms and then took the average of the 500 slope coefficient as an estimate of the slope. The second entry is the standard deviation of this coefficient estimated. This is the true standard error of the Fama-MacBeth coefficient. The third entry is the standard error estimated by the Fama-MacBeth procedure. This is the standard deviation of the 500 time series slope coefficients. The percent of Fama-MacBeth t-statistics which are significant at the one percent level (e.g.|t|>2.58) is reported in square brackets.

There are two things to note. First, in the presence of a firm effect, Fama-MacBeth regressions by firm produces correct standard errors and correctly sized confidence intervals. Second, the true standard error (efficiency of Fama-MacBeth) changes across the cells of the table. As you increase the residual volatility due to the firm effect and/or decrease the independent variability due to the firm effect, the volatility of the Fama-MacBeth slope coefficients increases.

Supplementary Table 11: Estimated Standard Errors with a Non-Fixed Time Effect
Panel A: OLS and Clustered Standard Errors

Avg(β_{OLS}) Std(β_{OLS}) Avg(SE_{OLS}) % Sig(T_{OLS}) Avg(SE_C) % Sig(T_C)	I	II	III	IV
ρ_X / ρ_ε	0.25 / 0.25	0.00 / 0.00	0.25 / 0.25	0.50 / 0.50
φ_X / φ_ε	0.00 / 0.00	0.75 / 0.75	0.75 / 0.75	0.50 / 0.50
OLS	1.0002 0.1098 0.0283 [0.5092] 0.1035 [0.0228]	1.0002 0.0526 0.0283 [0.1634] 0.0524 [0.0190]	0.9978 0.1253 0.0283 [0.5646] 0.1157 [0.0280]	1.0042 0.2224 0.0282 [0.7446] 0.2028 [0.0334]
OLS with time dummies	1.0008 0.0284 0.0283 [0.0100] 0.0279 [0.0214]	1.0001 0.0529 0.0284 [0.1686] 0.0525 [0.0182]	0.9999 0.0531 0.0283 [0.1682] 0.0525 [0.0170]	1.0000 0.0365 0.0283 [0.0466] 0.0360 [0.0190]

Panel B: Fama-MacBeth Standard Errors

Avg(β_{FM}) Std(β_{FM}) Avg(SE_{FM}) % Sig(T_{FM}) Avg($SE_{FM\text{-}AR1}$) % Sig($T_{FM\text{-}AR1}$)	I	II	III	IV
ρ_X / ρ_ε	0.25 / 0.25	0.00 / 0.00	0.25 / 0.25	0.50 / 0.50
φ_X / φ_ε	0.00 / 0.00	0.75 / 0.75	0.75 / 0.75	0.50 / 0.50
Fama-MacBeth	1.0008 0.0285 0.0280 [0.0218] 0.0271 [0.0424]	1.0001 0.0531 0.0526 [0.0192] 0.0510 [0.0348]	0.9999 0.0532 0.0526 [0.0184] 0.0513 [0.0380]	1.0000 0.0366 0.0361 [0.0184] 0.0349 [0.0338]
Avg(1st order auto-correlation)	-0.0558	-0.0546	-0.0504	-0.0558

Notes:

This table has the same structure as Table 5 in the paper. However, this data is generated with a time effect which is potentially non-fixed. Each simulation contains 500 firms and 20 years of data per firm. Column I is a fixed time effect, which is the same data structure as I used in Table 3 and 4. Column II allows the residuals (and independent variable) to be correlated for observations of the same firm, but the correlation dies off as the firms become more distant from each other (e.g., Corr[$\varepsilon_{i\,t}$, $\varepsilon_{k\,t}$] = ρ^{i-k}). This is what I call a non-fixed or temporary time effect. Column III and IV are combinations of the fixed and temporary time effect.

Panel A: Since the dependence is in the time dimension, the standard errors are clustered by year. The intuition is the same as in the original Table 5. Time dummies fully account for the dependence when the time effect is fixed (column I). When it is not (columns II-IV), clustering is necessary in addition to inclusion of the time dummies. The efficiency of the OLS estimates improves when there is a fixed time component (columns I, III, and IV), although this is not always true.

Panel B: Fama-MacBeth was designed to account for a time effect, whether fixed or not. does well with this kind of data structure. Consistent with this intuition, the Fama-MacBeth standard errors are unbiased and produced correctly sized confidence intervals. The first-order correction creates a small bias in the standard error since the first order auto-correlation (within firm) should be zero but is estimated to be negative in sample due to a small sample bias.

Supplementary Table 12: Estimating Standard Errors with an Unbalanced Panel
OLS and Clustered Standard Errors

Avg(β_{OLS}) Std(β_{OLS}) Avg(SE_{OLS}) % Sig(T_{OLS}) Avg(SE_C) % Sig(T_C)		Source of Independent Variable Volatility			
		0%	25%	50%	75%
Source of Residual Volatility	0%	1.0004	1.0002	1.0000	0.9991
		0.0286	0.0281	0.0280	0.0281
		0.0283	0.0283	0.0283	0.0284
		[0.0098]	[0.0078]	[0.0090]	[0.0094]
		0.0282	0.0282	0.0282	0.0282
		[0.0098]	[0.0078]	[0.0104]	[0.0116]
	25%	1.0004	0.9993	1.0012	1.0009
		0.0289	0.0400	0.0491	0.0554
		0.0283	0.0283	0.0283	0.0284
		[0.0116]	[0.0698]	[0.1386]	[0.1894]
		0.0282	0.0394	0.0480	0.0553
		[0.0118]	[0.0118]	[0.0134]	[0.0132]
	50%	0.9998	1.0005	1.0009	1.0001
		0.0280	0.0483	0.0628	0.0739
		0.0283	0.0283	0.0282	0.0283
		[0.0070]	[0.1336]	[0.2456]	[0.3216]
		0.0281	0.0480	0.0617	0.0729
		[0.0084]	[0.0104]	[0.0124]	[0.0138]
	75%	1.0003	1.0010	0.9998	0.9996
		0.0280	0.0559	0.0725	0.0890
		0.0282	0.0283	0.0283	0.0283
		[0.0094]	[0.1880]	[0.3110]	[0.4170]
		0.0281	0.0553	0.0729	0.0873
		[0.0096]	[0.0116]	[0.0110]	[0.0108]

Notes:

The table contains estimates of the coefficient and standard errors based on 5,000 simulated panel data set each of which contains 500 firms, but in this case the panel is unbalanced. Half of the firms have 2 years of data and half of the firms have 18 years of data (10 years per firm on average). The independent variable and the residual are simulated as described in the paper (see equations 4 and 5). The fraction of X's variance which is due to a firm specific component [Var(μ)/Var(X)] varies across the columns of the table from 0 percent (no firm effect) to 75 percent and the fraction of the residual variance which is due to a firm specific component [Var(γ)/Var(ε)] varies across the rows of the table from 0 percent (no firm effect) to 75 percent. Each cell contains the average slope coefficient estimated by OLS and the standard deviation of this estimate. This is the true standard error of the estimated coefficient. The third entry is the average standard error estimated by OLS. The percent of OLS t-statistics which are significant at the one percent level (e.g. $|t|>2.58$) is reported in square brackets. The fifth entry is the average standard error clustered by firm (i.e. accounts for the possible correlation between observations of the same firm in different years). The percent of clustered t-statistics which are significant at the one percent level is reported in square brackets.

The results in this table are broadly similar to the results in Table 1, with one notable exception. The true standard deviation and the bias in the standard deviation are larger when the panel is unbalanced (Table S12) than when the panel is balanced (Table 1). This is because the true standard error and the standard error estimated by clustering is a function of the variance of the number of years per firm (see Moulton, 1986).

$$AVar\left[\hat{\beta}_{OLS}-\beta\right]=\frac{\sigma_{\varepsilon}^{2}}{\sigma_{X}^{2}NT}\left(1+\left[\frac{Var\left(T\right)}{Avg\left(T\right)}+Avg\left(T\right)-1\right]\rho_{X}\ \rho_{\varepsilon}\right) \qquad \textbf{(2)}$$

Supplementary Table 13: Estimating Standard Errors with an Unbalanced Panel
Fama-MacBeth Standard Errors

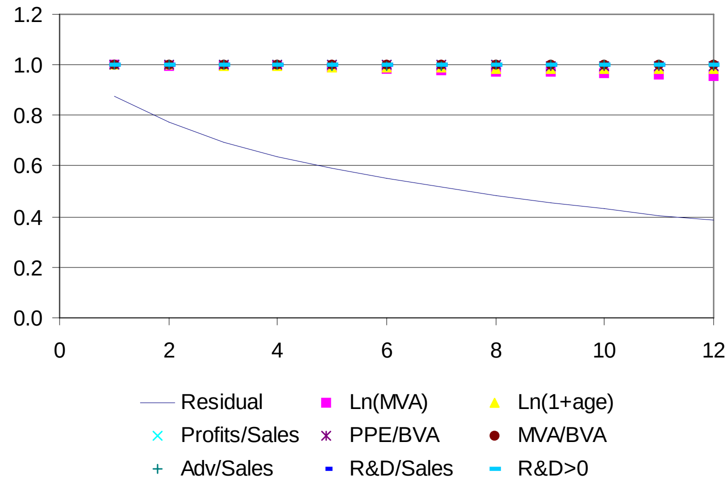
Source of Residual Volatility	Avg(β_{FM}) Std(β_{FM}) Avg(SE_{FM}) % Sig(T_{FM})	Source of Independent Variable Volatility 0%	25%	50%	75%
	0%	1.0003 0.0295 0.0288 [0.0174]	1.0000 0.0290 0.0287 [0.0160]	1.0000 0.0289 0.0287 [0.0170]	0.9990 0.0288 0.0287 [0.0192]
	25%	1.0006 0.0296 0.0287 [0.0210]	0.9999 0.0415 0.0280 [0.0986]	1.0012 0.0512 0.0270 [0.1876]	1.0010 0.0579 0.0260 [0.2538]
	50%	1.0000 0.0289 0.0286 [0.0196]	1.0005 0.0503 0.0270 [0.1782]	1.0010 0.0659 0.0251 [0.3326]	1.0001 0.0773 0.0230 [0.4524]
	75%	1.0002 0.0290 0.0286 [0.0210]	1.0012 0.0584 0.0260 [0.2578]	0.9996 0.0759 0.0230 [0.4412]	0.9995 0.0930 0.0197 [0.5874]

Notes:

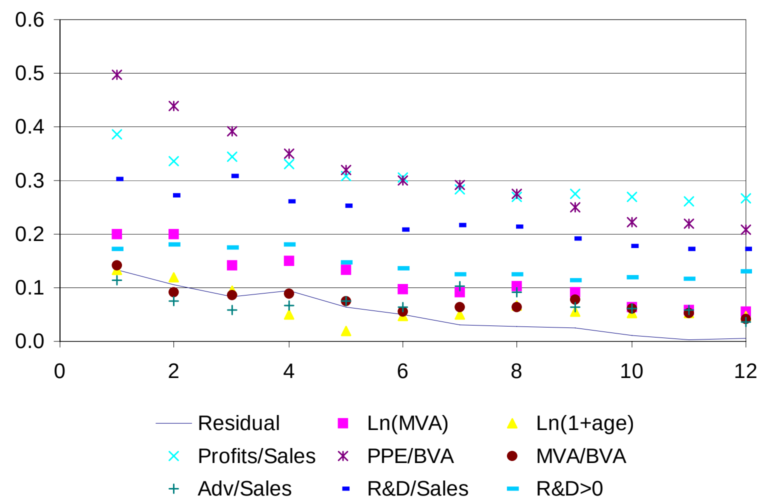
The table contains estimates of the coefficient and standard errors based on 5,000 simulated panel data set each of which contains 500 firms but the panel is unbalanced. Half of the firms have 2 years of data and half of the firms have 18 years of data (10 years per firm on average). The independent variable and the residual are simulated as described in the paper (see equations 4 and 5). The fraction of X's variance which is due to a firm specific component [Var(μ)/Var(X)] varies across the columns of the table from 0 percent (no firm effect) to 75 percent and the fraction of the residual variance which is due to a firm specific component [Var(γ)/Var(ε)] varies across the rows of the table from 0 percent (no firm effect) to 75 percent. The first entry is the average slope coefficient based on a Fama-MacBeth estimation (e.g. a regression is run for each of the 10 years and the estimate is the average of the 10 estimated slope coefficients). The second entry is the standard deviation of the coefficient estimated by Fama-MacBeth. This is the true standard error of the Fama-MacBeth coefficient. The third entry is the average standard error estimated by Fama-MacBeth (see equation 9). The percent of Fama-MacBeth t-statistics which are significant at the one percent level (e.g.$|t|>2.58$) is reported in square brackets.

The Fama-MacBeth standard errors are biased in the presence of a firm effect, but when the panel is unbalanced the bias is greater. Compare the results in this table (unbalanced panel) to Table 2 (balance panel). For example when 75 percent of the variability in both the residual and the independent variable comes from the firm effect, 50 percent of the t-statistics are statistically significant at the one percent level when the panel is balanced, but 59 percent of the t-statistics are statistically significant in this table.

Supplementary Figure 1: Residuals and Independent Variables Auto-Correlation: Corporate Finance Example
Panel A: Within Firm



Panel B: Within Month



Notes:

The auto-correlations of the residual and the eight independent variables are graphed for one to twelve lags. In Panel A, the correlations are within firm and are only calculated for observations of the same firm [i.e. Corr ($\varepsilon_{i\,t}\,\varepsilon_{i\,t\text{-}k}$) for k equal one to twelve]. In Panel B, the correlations are within year and are only calculated for observations of the same year [i.e. Corr ($\varepsilon_{i\,t}\,\varepsilon_{i\text{-}k\,t}$) for k equal one to twelve]. The data was sorted by month and then industry (4 digit) in Panel B. The independent variables are described in Appendix I.