Online Appendix for

Explaining Corporate Capital Structure: Product Markets, Leases, and Asset Similarity

Joshua D. Rauh
Amir Sufi

Section 1: Data

A. Overview of *Capital IQ*

Capital IQ (*CIQ*) is a Standard & Poor's business that collects a large amount of information on businesses throughout the world. We discovered the data available by *CIQ* through their internet interface (at http://www.capitaliq.com), which we recommend to any reader that wants to get a sense of the data available.

While the *CIQ* website contains a wealth of information, it is not suitable for downloading large amounts of data. For this purpose, we were directed to the Data Feed Team at *CIQ*, and our contact there has been John Schirripa, who is an expert in both the data available and the means by which researchers can obtain them. His email address is: jschirripa@capitaliq.com. You can also contact Alan Katz at akatz@capitaliq.com.

The data feed that we purchased from *CIQ* contains a series of text files that are linked. For example, one text file contains identifying information on companies, another contains information on balance sheet and income statement variables, and another has information on the detailed debt structure of firms. Once we downloaded these files, we had to link them to obtain our final data sets. In what follows, we describe the data we obtained.

B. Operating Lease Commitment Data

Operating lease commitment data come from a combination of both *CIQ* and *Compustat*. The relevant variables in *Compustat* are MRC1-MRC5, which represent "Rental Commitments Minimum 1st Year," "Rental Commitments Minimum 2nd Year," etc. and MRCTA, which represents "Thereafter Portion of Leases." For the variables MRC1-MRC5, the availability of these data for *Compustat* non-financial, U.S. based, parent firms increases gradually from about80% as of 1996 to 87% by 2008. However, the variable MRCTA is available for only 15% of the sample before 1999, at which point the availability jumps to about 80% by 2000. We are not sure why the availability for MRCTA was limited before 2000; we were unable to track down any reporting requirements or any other reason why *Compustat* has limited lease data before 2000.

We also have the identical variables from *CIQ*. For the observations for which we have the variables from both *CIQ* and *Compustat*, we find that the correlation is almost exactly one. For about 5% of our total sample, the MRC variables were missing from *Compustat* but included in *CIQ*. We use the *CIQ* leasing data in these cases. In other words, even if a researcher does not

have *CIQ* data, *Compustat* is sufficient to calculate leased capital for almost 95% of our final sample.

One important issue is how to handle observations for which some of the lease commitment data are missing. If any of MRC1 through MRC4 or MRCTA is missing, we do not discount the lease commitments and so leased capital is missing. If MRC5 is missing but MRC4 and MRCTA are not missing, we set MRC5 equal to zero. We make this latter change because there appears to be a large number of observations for which MRC5 is missing but no other MRC variables are missing.

C. Debt Structure Data

The debt structure data come from *CIQ*. The specific feed from the data feed group is called the "Capital Structure Feed." It includes detailed issue-level data from financial footnotes of 10-K SEC filings of firms. The exact data collection procedure used by *CIQ* seems to be quite similar to what is done in Rauh and Sufi (2010). It appears that *CIQ* has analysts record each individual issue from the debt financial footnote, the amount, the priority (i.e., senior, unsecured, or subordinated), and the type of debt.

Using these data, we are able to break down a firm's debt into one of 13 broad categories: bank revolvers, bank term loans, revenue bonds, capital leases, commercial paper, debentures, amortized discounts, mortgage debt, notes payable, smaller notes, medium term notes, convertibles, and unclassifiable debt. Our seven broad categories are bank debt (includes revolvers and term loans), arm's length non-program debt (which includes revenue bonds, debentures, and notes payable), arm's length program debt (which includes commercial paper and medium term notes), smaller notes, convertibles, and other (which is the residual).

These categories are similar to those used by Rauh and Sufi (2010) and we are able to directly compare their debt structure data with the debt structure data in *CIQ*. There is a very high correlation. In unreported results, we have replicated the Rauh and Sufi (2010) specifications using the *CIQ* data and we find very similar results.

For other researchers interested in the debt structure data, we are very happy to provide you all necessary Stata code to build the debt structure data if you obtain the data through *CIQ*.

D. Variable Construction

The core *Compustat* variables are constructed as follows:

Unadjusted Variables	
Book Leverage Ratio Without Leases	(dltt+dlc)/at
Owned PP&E$_t$ / Book Assets$_t$	ppent/at
OIBDP$_t$ / Book Assets$_{t-1}$	oibdp/at(lagged)
Market Assets$_t$ / Book Assets$_t$	(at+prcc_f*csho-ceq-txdb)/at
ln(Sales)	Ln(sale)

New Variables	
Capitalized operating lease commitments	oplease_rd (described in data section of paper
Book Leverage Ratio With Leases	(dltt+dlc+oplease_rd)/(at+oplease_rd
Total PP&E$_t$ / (Assets + Leases)$_t$	(ppent+oplease_rd)/(at+oplease_rd)
Owned PP&E$_t$ / (Assets + Leases)$_t$	(ppent)/(at+oplease_rd)
OIBDP ex Rent$_t$ / (Assets + Leases)$_{t-1}$	(oibdp+xrent)/(at(lagged)+oplease_rd(lagged))
Market Assets$_t$ / (Book Assets + Leases)$_{t-1}$	(at+prcc_f*csho-ceq-txdb+oplease_rd)/(at+oplease_rd)
Other variables	
Capital to labor ratio	(ppent+oplease_rd)/(emp*1000)
PPE in machinery	(ppegmCIQ)/(ppegtCIQ+oplease_rd)
PPE in buildings	(ppegbCIQ)/(ppegtCIQ+oplease_rd)

The last two variables come from balance sheet information from *CIQ*. They represent the gross PPE that is in machinery and buildings.

Section 2: Weighted Least Squares Estimation

Weighted least squares estimation is a specific form of generalized least squares that can improve the efficiency of estimates under certain assumptions. In our context, we have the following equation estimated via OLS:

$$Leverage_i = \alpha + \beta * \overline{Leverage}_{j,-i} + \varepsilon_i$$

where $\overline{Leverage}_{j,-i}$ is the leverage ratio of *CIQ* competitors. If there is heteroscedasticity and if there is a known variable that is a linear function of the degree of heteroscedasticity, weighted least squares with weights being the inverse square root of the known variable is a more efficient estimator than OLS. In particular, if for any *i*, $Var(\varepsilon_i) = \sigma_i^2$ and $\sigma_i^2 = \sigma^2 * 1/(\#\ competitors)$, then a WLS estimation where all variables are multiplied by the square root of the number of competitors is more efficient than OLS.

Appendix Figure 1 presents evidence that is suggestive of heteroscedasticity of the above form. To produce the figure, we first estimate the above equation via OLS to obtain predicted residuals. Appendix Figure 1 shows the standard deviation of the predicted residuals by the number of competitors over which $\overline{Leverage}_{j,-i}$ is calculated. As the figure shows, there is a strong negative relation between the standard deviation of the predicted residuals and the number of competitors. The pattern in Appendix Figure 1 strongly suggests heteroscedasticity, and that the heteroscedasticity is a function of the number of competitors. The WLS estimation down-weights firms that have fewer competitors to take into account the additional noise from mismeasurement.

The Stata command that we employ for WLS estimation is “regress” with [aweight = # of competitors]. Stata mechanically transforms the weight for any firm *i* to be equal to

$$w_i = \frac{n * (\#\ competitors_i)}{\sum_i^N \#competitors_i}$$

These weights then form the weighting matrix that is used to estimate WLS. The weighting matrix **D** is a diagonal matrix of size *nxn* with the diagonal elements being the weights above. As a robustness test, we replicate the Stata WLS command by multiplying all variables (including the constant) by the square root of the weights and find the exact same coefficient estimates.

In producing the R^2, the Stata command we employ calculates the following:

$$TSS = y'\boldsymbol{D}y - \{\frac{(1'y)^2}{n}\}$$

$$ESS = y'\boldsymbol{D}y - b'\boldsymbol{X}'\boldsymbol{D}y$$

$$R^2 = 1 - \frac{ESS}{TSS}$$

The main difference between the WLS and OLS R^2 calculations is the inclusion of the weighting matrix **D** in the WLS equations. One important note is that the R^2 of the WLS estimation is not comparable to the R^2 of OLS. In analyzing the results, we are careful to only compare the relative predictive power of variables within an OLS or WLS estimation, not across the estimations.

Section 3: Robustness Tests

Appendix Table 1 replicates the key findings of our analysis for years going back to 2004.

Appendix Table 2 replicates the key findings of our analysis using a credit rating specific discount rate to capitalize operating leases and 8X rental expense as a measure of the capitalized value of operating leases.

Section 4: Comparison with Hoberg and Phillips (2009)

We do not have access to the exact similarity scores by Hoberg and Phillips (2009). Instead, we create similarity scores using their same methodology for our sample with one important difference. Instead of extracting the full text from a firm's 10K SEC filing, we only use the short business description contained in the *Compustat* field *busdesc*. (We were unable to extract the full text from the 10K filing as in Hoberg and Phillips (2009).) Implementing the Hoberg and Phillips (2009) methodology on our sample leads to a matrix where for every firm, there is a score based on how similar the text in *busdesc* is to the text of another firm's *busdesc*.

One initial result from this exercise is that the average similarity scores from the Hoberg and Phillips (2009) methodology are much higher among *CIQ* competitors than firms in the same 3

digit SIC codes. In other words, the set of *CIQ* competitors has a higher degree of similarity in their descriptions of their business. This is yet another piece of evidence against the use of SIC codes.

There are several ways to use the Hoberg and Phillips (2009) scores to create competitors. Hoberg and Phillips (2009) use every score that is non-zero and weight competitors by how high their score is. We tried several different procedures and chose the one that makes the Hoberg and Phillips (2009) measure as strong as possible in terms of adjusted R^2. We use only the top 25 other firms based on the similarity score, and then we weight each of these 25 by how high their score is. For any outcome, we construct the HP 25 competitor average over this outcome using the 25 firms with the highest similarity scores and weighting more heavily those with higher scores.

Appendix Table 3 replicates Table 2 of the text, with the use of the HP 25 competitors instead of 3 digit SIC codes. Consistent with the evidence in Hoberg and Phillips (2009), the HP 25 classification of competitors does a great job on operating performance. In fact, it outperforms the *CIQ* competitors. The *CIQ* competitor measure does a better job of explaining variation in the standard deviation of operating income and sales growth.

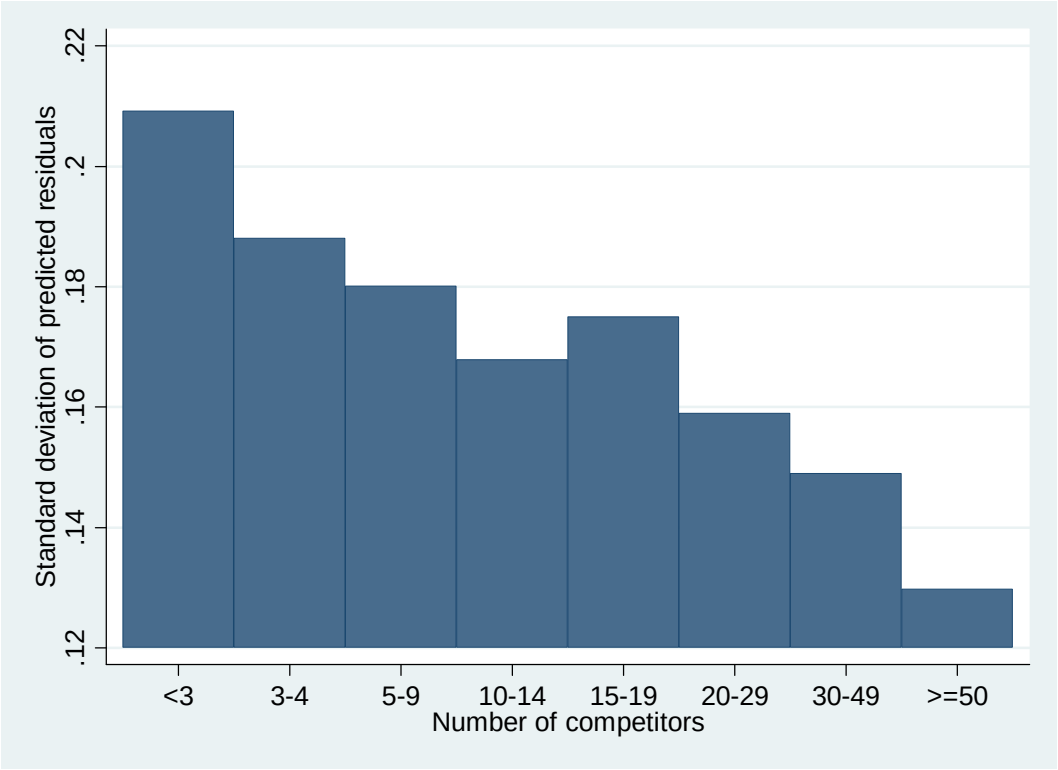
However, HP 25 competitors do a poor job of explaining variation in capital structure of a given firm. In fact, the HP 25 competitors explain less of the variation in capital structure than even firms in the same SIC3 (see Table 2). Hoberg and Phillips (2009) report this same result in their study (see in particular Table 3 of Hoberg and Phillips (2009)).

One potential reason is that the HP 25 competitors are less similar in terms of their asset composition. The R^2 when using the HP 25 competitors to explain variation in capital to labor ratios or the tangible assets to total assets ratio are much lower. In other words, the set of *CIQ* competitors is more similar in terms of their asset composition and capital to labor ratios than the set of HP 25 competitors. Consistent with the importance of asset similarity described in the text of our study, this is a likely channel through which *CIQ* competitors perform better in explaining capital structure than the Hoberg and Phillips (2009) methodology. It is worth emphasizing that Hoberg and Phillips (2011) use their alternative measure primarily to understand product market synergies, mergers, advertising, and R&D.

In Panel B, we examine the correlation of stock returns. The returns of the HP 25 competitors are more correlated with a given firm's stock returns than 3-digit SIC codes, but the correlation with *CIQ* competitors is even stronger. In other words, while the HP 25 competitors explain more of the variation in operating performance, the *CIQ* competitors explain more of the variation in stock returns. Both do substantially better than 3-digit SIC codes.

Appendix Figure 1: Justifying Weighted Least Squares

The following figure plots the standard deviation of predicted residuals from a regression of the leverage ratio with leases of a given firm on a constant and the average leverage ratio with leases of *CIQ* competitors. As the figure shows, the standard deviation of predicted residuals is much larger for firms with fewer competitors, suggesting that weighted least squares using the number of competitors as weights is more efficient than OLS. Each bin of number of competitors includes approximately 10% of the firms each.



Appendix Table 1: Main Cross-Sectional Specifications for Each Year 2004-2008

In the first column, the dependent variable is the *Leverage Ratio Without Leases* at book values. It follows extant literature and ignores the capitalized value of operating leases in both the numerator and denominator. The variable *Leverage Ratio With Leases* is defined as (Debt + Leases)$_t$ / (Assets + Leases)$_t$, where Leases are measured as the capitalized value of operating leases as described in the text. The third column in each panel presents WLS estimates where weights are given by the number of *CIQ* competitors. The explanatory variables follow accordingly. Robust standard errors are in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

	2008				
	Dependent Variable: Leverage Ratio				
	no leases	*with leases*			
Leverage Ratio of Other Firms in SIC3	0.819***	0.345***	0.159**	0.136**	0.726***
	(0.034)	(0.046)	(0.069)	(0.069)	(0.050)
Leverage Ratio of CIQ Competitors		0.653***	0.877***	0.835***	
		(0.047)	(0.066)	(0.068)	
Owned PP&E$_t$ / (Assets+Leases)$_t$				0.032	0.086***
				(0.021)	(0.021)
OIBDP ex Rent$_t$ / (Assets+Leases)$_{t-1}$, 5yr mean				-0.064*	-0.038
				(0.036)	(0.037)
OIBDP ex Rent$_t$ / (Assets+Leases)$_{t-1}$, 5yr stdev				0.063	0.074
				(0.079)	(0.082)
Market Assets$_t$ / (Assets+Leases)$_t$				-0.018***	-0.02***
				(0.006)	(0.006)
ln(Sales)				0.006**	0.007**
				(0.003)	(0.003)
Constant	0.042***	-0.002	-0.007	-0.013	0.048**
	(0.008)	(0.009)	(0.013)	(0.021)	(0.022)
Observations	2569	2569	2569	2569	2569
Adjusted R-squared	0.21	0.31	0.35	0.37	0.29

	2007				
	Dependent Variable: Leverage Ratio				
	no leases	*with leases*			
Leverage Ratio of Other Firms in SIC3	0.774***	0.379***	0.244***	0.208***	0.7***
	(0.035)	(0.043)	(0.064)	(0.063)	(0.049)
Leverage Ratio of CIQ Competitors		0.623***	0.779***	0.748***	
		(0.045)	(0.065)	(0.065)	
Owned PP&E$_t$ / (Assets+Leases)$_t$				0.01	0.048***
				(0.017)	(0.017)
OIBDP ex Rent$_t$ / (Assets+Leases)$_{t-1}$, 5yr mean				-0.05	-0.048
				(0.032)	(0.033)
OIBDP ex Rent$_t$ / (Assets+Leases)$_{t-1}$, 5yr stdev				-0.048	-0.053
				(0.059)	(0.060)
Market Assets$_t$ / (Assets+Leases)$_t$				-0.019***	-0.02***
				(0.003)	(0.004)
ln(Sales)				0.003	0.005*
				(0.002)	(0.003)
Constant	0.046***	-0.004	-0.003	0.035*	0.091***
	(0.008)	(0.009)	(0.013)	(0.020)	(0.020)
Observations	2762	2762	2762	2762	2762
Adjusted R-squared	0.16	0.27	0.31	0.33	0.27

	2006				
	Dependent Variable: Leverage Ratio				
	no leases	*with leases*			
Leverage Ratio of Other Firms in SIC3	0.843***	0.418***	0.257***	0.236***	0.741***
	(0.035)	(0.046)	(0.069)	(0.067)	(0.049)
Leverage Ratio of CIQ Competitors		0.588***	0.779***	0.748***	
		(0.048)	(0.065)	(0.064)	
Owned PP&E$_t$ / (Assets+Leases)$_t$				-0.002	0.046**
				(0.018)	(0.019)
OIBDP ex Rent$_t$ / (Assets+Leases)$_{t-1}$, 5yr mean				-0.059*	-0.063**
				(0.031)	(0.032)
OIBDP ex Rent$_t$ / (Assets+Leases)$_{t-1}$, 5yr stdev				-0.046	-0.08
				(0.066)	(0.067)
Market Assets$_t$ / (Assets+Leases)$_t$				-0.02***	-0.021***
				(0.003)	(0.004)
ln(Sales)				0.002	0.004
				(0.002)	(0.003)
Constant	0.031***	-0.002	-0.009	0.041**	0.089***
	(0.007)	(0.009)	(0.011)	(0.019)	(0.019)
Observations	2808	2808	2808	2808	2808
Adjusted R-squared	0.19	0.27	0.33	0.35	0.29

	2005				
	Dependent Variable: Leverage Ratio				
	no leases	*with leases*			
Leverage Ratio of Other Firms in SIC3	0.8***	0.425***	0.258***	0.244***	0.74***
	(0.034)	(0.044)	(0.056)	(0.054)	(0.043)
Leverage Ratio of CIQ Competitors		0.544***	0.764***	0.725***	
		(0.047)	(0.057)	(0.058)	
Owned PP&E$_t$ / (Assets+Leases)$_t$				-0.001	0.04**
				(0.017)	(0.018)
OIBDP ex Rent$_t$ / (Assets+Leases)$_{t-1}$, 5yr mean				-0.085***	-0.094***
				(0.031)	(0.032)
OIBDP ex Rent$_t$ / (Assets+Leases)$_{t-1}$, 5yr stdev				-0.117*	-0.185***
				(0.067)	(0.071)
Market Assets$_t$ / (Assets+Leases)$_t$				-0.018***	-0.019***
				(0.003)	(0.003)
ln(Sales)				0	0.001
				(0.002)	(0.002)
Constant	0.038***	0.009	-0.004	0.063***	0.114***
	(0.007)	(0.009)	(0.011)	(0.019)	(0.020)
Observations	2820	2820	2820	2820	2820
Adjusted R-squared	0.17	0.25	0.33	0.35	0.29

	2004				
	Dependent Variable: Leverage Ratio				
	no leases	*with leases*			
Leverage Ratio of Other Firms in SIC3	0.811***	0.400***	0.284***	0.253***	0.721***
	(0.032)	(0.044)	(0.057)	(0.055)	(0.044)
Leverage Ratio of CIQ Competitors		0.571***	0.746***	0.697***	
		(0.046)	(0.057)	(0.058)	
Owned PP&E$_t$ / (Assets+Leases)$_t$				-0.002	0.046**
				(0.018)	(0.018)
OIBDP ex Rent$_t$ / (Assets+Leases)$_{t-1}$, 5yr mean				-0.07**	-0.083***
				(0.029)	(0.030)
OIBDP ex Rent$_t$ / (Assets+Leases)$_{t-1}$, 5yr stdev				-0.076	-0.16***
				(0.054)	(0.056)
Market Assets$_t$ / (Assets+Leases)$_t$				-0.023***	-0.023***
				(0.003)	(0.003)
ln(Sales)				0.001	0.001
				(0.002)	(0.002)
Constant	0.033***	0.006	-0.007	0.066***	0.122***
	(0.007)	(0.008)	(0.011)	(0.020)	(0.020)
Observations	2838	2838	2838	2838	2838
Adjusted R-squared	0.19	0.26	0.34	0.37	0.32

Appendix Table 2: Alternative Measures of Capitalized Operating Leases

This table replicates the last three columns of Table 3 using alternative methods for capitalizing operating leases. The first three columns use 8X rental expense as a measure of capitalized operating leases. The second three columns use a credit-rating specific discount rate to discount operating lease commitments. Robust standard errors are in parentheses. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

	Dependent Variable					
	Leverage Ratio using 8X rental expense for leases			*Leverage Ratio using credit-rating specific discount rate for leases*		
Leverage Ratio of Other Firms in SIC3	0.811***		0.354***	0.822***		0.319 ***
	(0.030)		(0.043)	(0.031)		(0.045)
Leverage Ratio of *CIQ* Competitors		0.899***	0.636***		0.919***	0.682 ***
		(0.028)	(0.044)		(0.030)	(0.046)
Constant	0.064***	0.031***	0.002	0.050***	0.013	-0.006
	(0.010)	(0.010)	(0.010)	(0.009)	(0.008)	(0.009)
Method	OLS	OLS	OLS	OLS	OLS	OLS
Weights	—	—	—	—	—	—
Observations	2801	2801	2801	2801	2801	2801
Adjusted R-squared	0.23	0.27	0.29	0.23	0.28	0.29

Appendix Table 3: CIQ Competitors and Hoberg-Phillips Similarity Scores

Each row of Panel A shows the adjusted R-squared for three regressions: a regression of the characteristic on the average characteristic of other firms with the top 25 similarity scores based on an algorithm from Hoberg and Phillips (2010) using the short business description field in *Compustat*, a regression of the characteristic on the average characteristic of at *CIQ* competitors, and a regression of the characteristic on both. Panel B presents regressions of monthly stock returns for a given firm on the value-weighted market return and the equal-weighted portfolio return of other HP25 and *CIQ* competitors. The estimation period for Panel B is 2003 through 2008 and standard errors are clustered by year. *** significant at the 1% level, ** significant at the 5% level, * significant at the 10% level.

Panel A: Adjusted R-Squared in Regression of Characteristic on Average of Other Group Members, 2008

	Hoberg-Phillips 25	*CIQ* Competitors	HP25 & *CIQ* Competitors
$OIBDP_t$ / Book $Assets_{t-1}$	0.266	0.230	0.308
$OIBDP$ ex $Rent_t$ / $(Assets+Leases)_{t-1}$	0.273	0.239	0.316
$OIBDP$ ex $Rent_t$ / $(Assets+Leases)_{t-1}$, 5yr mean	0.325	0.301	0.384
$OIBDP$ ex $Rent_t$ / $(Assets+Leases)_{t-1}$, 5year stdev	0.110	0.123	0.191
$OIBDP_t$ / $Sales_t$	0.273	0.200	0.304
$OIBDP$ ex $Rent_t$ / $Sales_t$	0.273	0.208	0.308
Sales $Growth_t$	0.102	0.144	0.143
Leverage ratio without $leases_t$	0.175	0.246	0.253
Leverage ratio with $leases_t$	0.208	0.287	0.295
Market to book $ratio_t$	0.106	0.114	0.126
Total $PP\&E_t$ / $(Assets+Leases)_t$	0.522	0.669	0.682
Owned $PP\&E_t$ / $(Assets+Leases)_t$	0.541	0.679	0.688
Capital / $Labor_t$	0.634	0.766	0.769
$ln(Sales_t)$	0.234	0.153	0.257

Panel B: Monthly Return Regressions

	Dependent variable: return of firm i in month t					
	(1)	(2)	(3)	(4)	(5)	(6)
Value-weighted market $return_t$	0.634***	0.890***	0.467***	0.230***	0.694***	0.152***
	(0.092)	(0.223)	(0.120)	(0.030)	(0.202)	(0.049)
Portfolio return *CIQ* $Competitors_{it}$	0.546***		0.483***	0.809***		0.725***
	(0.024)		(0.034)	(0.013)		(0.035)
Portfolio return of firms in $HP25_{it}$		0.328***	0.187**		0.410***	0.142**
		(0.126)	(0.078)		(0.144)	(0.065)
Constant	0.185	0.206	0.129	0.067	0.166	0.029
	(0.195)	(0.229)	(0.153)	(0.051)	(0.152)	(0.031)
Weighted?		No		Yes, by number of CIQ competitors		
Observations	144184	144184	144184	143482	143482	143482
Adjusted R-squared	0.12	0.09	0.12	0.19	0.14	0.20