Summary of Cost of Capital Data

Niels Joachim Gormsen and Kilian Huber*

January 2024

Abstract

This note summarizes the data produced by Gormsen and Huber (2024). Section 1 contains background information on the data. Section 2 contains information about the shared data. Section 3 contains answers to frequently asked questions.

*Both authors are at the University of Chicago, niels.gormsen@chicagobooth.edu and kilianhuber@uchicago.edu. The research was supported by the Becker-Friedman Institute, the Fama Faculty Fellowship, the Fama-Miller Center, and the Lee Economics Program Faculty Fellowship at the University of Chicago. We thank all research assistants and research professionals in the Cost of Capital Lab for excellent work: Alexandra Bruner, Ben Meyer, Cagdas Okay, Charlotte Wang, Chris Saroza, Daniel Marohnic, Esfandiar Rouhani, Henry Shi, Izzy Sethi, Jasmine Han, Jason Jia, Madeleine Zhou, Manhar Dixit, Meena Rakasi, Neville Nazareth, Rachel Kim, Rahul Chauhan, Raymon Yue, Rohan Mathur, Sanjna Narayan, Scarlett Li, Sean Choi, Sonali Mishra, Sungil Kim, Tony Ma.

The dataset "CoCdata_V2_0.dta" can currently be downloaded on costofcapital.org. These are data on the perceived cost of capital and discount rates for 9,500 firms between 2002 and 2021. The dataset contains around 250,000 quarterly estimates of firm-level predictions of perceived cost of capital and discount rates. The data span firms in the US, the UK, and continental Europe. We hope these data can be useful for research involving firms' cost of capital and discount rates. We ask that you please cite the relevant papers when using the data.

This note is organized as follows. Section 1 provides background on the relevant objects and the data collection. Section 2 provides an explanation for how the dataset of predicted values is constructed, along with validation of the data and information about previous versions of the shared data. Section 3 provides answers to frequently asked questions.

1 Background

Most firms make investment decisions based on discount rates and estimates of their cost of capital. Section 1.1 defines these concepts and Section 1.2 explains how we collect data on the relevant objects.

1.1 Framework and definitions

Managers often evaluate new projects based on their NPV:

$$\text{NPV} = \sum_{s=0}^{S} \delta^{-s} \mathbb{E}[\text{Revenue}_s - \text{Cost}_s].$$

The discount rate δ is set based on the *perceived cost of capital* plus a wedge κ,

$$\delta = 1 + r^{\text{per.}} + \kappa.$$

The origins of the wedge κ are discussed in Gormsen and Huber (2023).

Firms' perceived cost of capital is based on the *financial cost of capital.* Firms estimate it based on financial markets, but they do so imperfectly, leading to a wedge υ:

$$r^{\text{per.}} = r^{\text{fin.}} + \upsilon.$$

The resulting discount rates used by firms reflect two wedges:

$$\begin{aligned}\delta &= 1 + r^{\text{per.}} + \kappa \\ &= 1 + r^{\text{fin.}} + \upsilon + \kappa,\end{aligned}$$

The data in CoCdata.dta contain firm-level estimates of δ and $r^{\text{per.}}$, from which κ can be calculated.

1.2 Data collection

Gormsen and Huber (2023) measures perceived cost of capital and discount rates based on information provided by managers during corporate earnings calls, investor conferences, and similar events, which we jointly call "conference calls." The description of the data collection can be found in Section 2.1-2.3 of Gormsen and Huber (2023). A summary of the resulting data can be found in Section 3 of the same paper.

2 Predicted data

To facilitate analysis on firms' perceived cost of capital and discount rates, we construct a panel dataset of predicted values of these objects. We fit the observed values of firms' perceived cost of capital and discount rates to a large set of observed firm-level characteristics and use the resulting model to predict predict firm-level perceived cost of capital and discount rates for the universe of Compustat firms across quarters.

2.1 Methodology

We predict cross-sectional variation in the perceived cost of capital and discount rates based on firm-level characteristics. The main characteristics are the firm-level characteristics used to construct risk factors in asset pricing. According to standard theory, these risk factors should predict risk premia and therefore discount rates and the perceived cost of capital. We use the 153 characteristics collected by Jensen et al. (2023). The characteristics are all measured in cross-sectional percentiles of the universe of firms in a the given country at the given time.

We use a standard Lasso procedure to select the relevant characteristics and perform optimal shrinkage. We feed the Lasso procedure the dependent variable (the perceived cost of capital or discount rates) as well a set of candidate variables. The candidate variables are the 153 risk factors as well as a regional dummy capturing whether the firm is located in Europe or the US. A later step in our methodology will ensure that the levels are correct in US and Europe irrespectively of whether or not the Lasso procedure picks the regional dummy, but including the dummy in the set of candidate variables can help in the predictive exercise if the US and European firms have slightly different characteristics on average. We estimate the Lasso procedure based on the BIC criterion.

Based on the model estimated by the Lasso procedure, we calculate predicted values for all firms for which we observe the set of characteristics needed to calculate both perceived cost of capital and discount rates. Since we only feed the model cross-sectional predictors, there is virtually no time variation in the aggregate series. To obtain the correct time variation, we add in the estimated time variation from the full sample of discount rates and perceived cost of capital. We estimate the time variation in these objects by projecting discount rates and perceived cost of capital onto year dummies and absorbing firm fixed effects. This procedure ensures that all variation is driven by within-firm variation in the relevant estimates. We calculate time variation separately in the US and Europe. The European countries consists of both euro (or euro-pegged) countries as well as the UK, in which firms denominate in pounds. Using only one time series for euro and pound denominated series could be problematic in the presence of large divergence in inflation across the two currencies, but it helps ensure a sufficient set of firms to estimate time variation robustly. We exclude firms from other countries from our sample of predicted values as we do not have enough observations to robustly estimate the time variation.

2.2 Results

Perceived Cost of Capital The Lasso procedure picks 11 variables for predicting the perceived cost of capital. These variables are illustrated in Figure 1, which is discussed below. The in-sample R^2 of the selected model is 33%.

Figure 1 shows the slope coefficients for each of the 11 selected variables. These slope coefficients directly tell us how much the predicted value of the perceived cost of capital increases if we go from the bottom to the top of the cross-section of the given characteristics (keeping the other 10 characteristics constant). The most important characteristic is the CAPM beta. The coefficient is 1.8, which means that the perceived cost of capital is predicted to be 1.8 percentage points higher for the firms with the highest betas relative to the firms with the lowest betas. The second most important characteristic is the European dummy, which captures that the perceived cost of capital is lower in Europe than in the US. The next variable is debt-to-market, which is the main measure of leverage in Gormsen and Huber (2024). Firms with higher leverage naturally have lower perceived cost of capital due to the tax benefits of debt. The slope coefficient is around -1. We find similar slope coefficients for age and market equity (size), which capture that larger and older firms have lower perceived cost of capital.

Figure 3 plots the average predicted perceived cost of capital in the US by year. This series reflects the time-variation incorporated as explained in 2.1. The perceived cost of capital trend downwards substantially over the sample as discussed in Gormsen and Huber (2023).

Discount rates The Lasso procedure picks 13 variables for predicting discount rates. These variables are illustrated in Figure 2, which is discussed below. The in-sample R^2 of the selected model is 16%.

Figure 2 shows the slope coefficients for each of the 13 selected variables. These slope coefficients directly tell us how much the predicted value of the discount rate increases if we go from the bottom to the top of the cross-section of the given characteristics (keeping the other 12 characteristics constant). The most important characteristic is idiosyncratic volatility, which is measured over 252 days relative to the CAPM (see Jensen et al. 2023 for formal definitions). The coefficient is 3.2, which means that the perceived cost of capital is predicted to be 3.2 percentage points higher for the firms with the highest volatility relative to those with the lowest volatility. The second most important characteristic is age. The coefficient shows that the oldest firms in the economy have roughly 2 percentage points lower hurdle rates than the youngest firms. The next variable is cash-to-assets. Firms with more cash have higher hurdle rates. Firms with higher labor force efficiency and lower risk of default (higher Z-score) also have higher hurdle rates. Discount rates are lower for firms with abnormally high investment. This last finding is consistent with the idea that lower hurdle rates leads to higher investment.

Figure 3 plots the average predicted discount rates in the US by year. This series reflects the time-variation incorporated as explained in 2.1. The discount rates trend down over the sample, but by much less than the cost of capital. The predicted wedge between the two, κ, increases by more than 1.5 percentage points over the sample. This finding is discussed in detail in Gormsen and Huber (2023).

Table 1 further studies this predicted wedge by regressing the predicted value of the discount rates onto predicted values for the perceived cost of capital. Without fixed effects, we find a slope coefficient that is close to 1 and R^2 around 55%. Once including a firm-fixed effect, this slope goes down to 0.53, illustrating time variation in the wedge, κ, between discount rates and the perceived cost of capital. The slope coefficient in this regression is slightly above what we get when using the actual, observed discount rate on the left hand side. When using the actual discount rates, the slope coefficient is around 0.25 (see also Gormsen and Huber 2023). The R^2 is also higher for the predicted data than the observed data. These results reflect that there is no idiosyncratic volatility or stickiness in the variables, which increases the R^2 and slope coefficients.

2.3 Validation

We validate the predictive power of our data in an out-of-sample test. In this test, we use the predicted values to predict the perceived cost of capital and discount rates observed in the Duke-CFO survey. The seminal Duke-CFO survey is a quarterly survey of corporate managers (Graham and Harvey 2001). In some of these surveys, managers are asked about their cost of capital and their discount rates (referred to as hurdle rates in the survey). We use these data to test how well our predictive value work out of sample.[1]

The results are presented in Table 2. The first two columns shows regressions of the perceived cost of capital in the Duke-CFO data on the predicted values. The slope on the predicted values is 0.74 without year fixed effects and 0.9 with year fixed effects. These results are consistent with the notion that the time variation in the perceived cost of capital in the Duke-CFO survey differs from that of the conference call data (see Gormsen and Huber 2023 for more discussion on this result), so including year fixed effects to capture this difference increase the slope. More importantly, the finding in column 2 suggests that the cross-sectional variation in our predicted values is also

[1]We thank John Graham for generously sharing these data.

in the Duke-CFO data with the same magnitude (slope close to 1). The cross-sectional variation in the predicted values thus appears to be an unbiased predictor of the cross-sectional variation in the Duke-CFO data.

Column three and column four shows results for discount rates. Here the slope coefficients are almost exactly one, both with and without year fixed effects. The discount rates in the Duke-CFO data are around three percentage points higher than in the conference call data, as seen from the intercept. A likely driver of this difference is that our predicted data is for headquarter discount rates, which are lower than non-headquarter discount rates, whereas the the Duke-CFO data likely contain a mix of headquarter and non-headquarter discount rates. It should, however, be noted that the three percentage point difference is insignificant given the small sample of 92 observations.[2]

Overall, the results suggests that the predicted values are unbiased predictors of cross-sectional variation in the perceived cost of capital and discount rates out of sample.

2.4 Versions

The data currently shared on `costofcapital.org` represent version 2.0. The below summarizes the different versions of the data:

Version 1.0: January, 2023

- Based on conference call data as of December 2022.
- The cross-sectional predictors include firm-level implied volatility and market power, which greatly limits the number of observations
- These data are almost purely cross-sectional, with all time-variation incorporated through a time-trend
- Please email the authors for more information about, or access to, these older data.

Version 2.0: January 11, 2024

- Based on conference call data as of September 2023

[2]While the Duke-CFO data contains more than 92 observations, many of these are non-listed firms or firms that cannot be matched to firm-level identifiers.

– Only uses predictors from the Jensen et al. (2023) dataset, which greatly increases the number of observations

– Time variation is now directly incorporated through the addition estimates of the yearly averages of the given value in the region

3 FAQ

Below are answers to some of the most frequently asked questions about the data:

Q 1 *Why is the level of discount rates lower in the predicted values than in the series in Figure 2 of Gormsen and Huber (2023)?* Figure 2 in that paper plots the raw average discount rate. Some of these discount rates are headquarter level and others are not at the headquarter level. Headquarter-level discount rates are generally lower because they incorporate all overhead costs, as discussed in Section 3 of Gormsen and Huber (2023). The predicted data in "CoCdata_V2.dta" contains predictions for discount rates at the headquarter level.

Q 2 *Why is the time variation in discount rates and perceived cost of capital different in the predicted values than in the series in Figure 2 of Gormsen and Huber (2023)?* Figure 2 in that paper plots raw averages. The time variation in the predicted data, however, comes from an estimation that relies on firm-level variation. In particular, the annual averages reflected in the predicted data are pinned down by year fixed effects in a panel regression of the relevant variable (perceived cost of capital or discount rates) on the year fixed effects, absorbing firm variation. This is the same procedure as the one used in Gormsen and Huber (2023) to construct Figure 5 in that paper.

Q 3 *What is the difference between a discount rate and a hurdle rate?* We use the terminology hurdle rate and discount rate interchangeably. This is consistent with past work (see, e.g., Jagannathan et al. 2016). See also further discussion in Gormsen and Huber (2023).

References

Gormsen, N. J. and K. Huber (2023): "Corporate Discount Rates," *University of Chicago Working paper.*

——— (2024): "Firms' Perceived Cost of Capital," *University of Chicago working paper.*

Graham, J. R. and C. R. Harvey (2001): "The theory and practice of corporate finance: Evidence from the field," *Journal of Financial Economics*, 60, 187–243.

Jagannathan, R., D. Matsa, I. Meier, and V. Tarhan (2016): "Why do firms use high discount rates?" *Journal of Financial Economics*, 120, 445–463.

Jensen, T. I., B. T. Kelly, and L. H. Pedersen (2023): "Is There a Replication Crisis in Finance?" *Journal of Finance*, 78.

Figure 1
Predictors of the Perceived Cost of Capital

This figure shows slope coefficients of the predicted values of firms' perceived cost of capital onto the variables selected by the Lasso procedure. The dependent variable in the Lasso regression is firms' perceived cost of capital in a given quarter. The set of possible explanatory variables includes the firms exposure to the 153 risk factors in Jensen et al. (2023) — which risk exposure is measured by firm characteristics — as well as a dummy for the region (U.S. versus European firm). The firm-level characteristics are measured in cross-sectional percentiles of the universe of firms in a given country at the give time. The variable ranges from 0 (lowest) to 1 (highest) and the left-hand side is measured in percentage points, so a loading of 1 means that the perceived cost of capital is predicted to be 1 percentage points higher for firms with the highest characteristics relative to firms with the lowest. The data is fitted based on firms in Europe and US between 2002 and 2021.

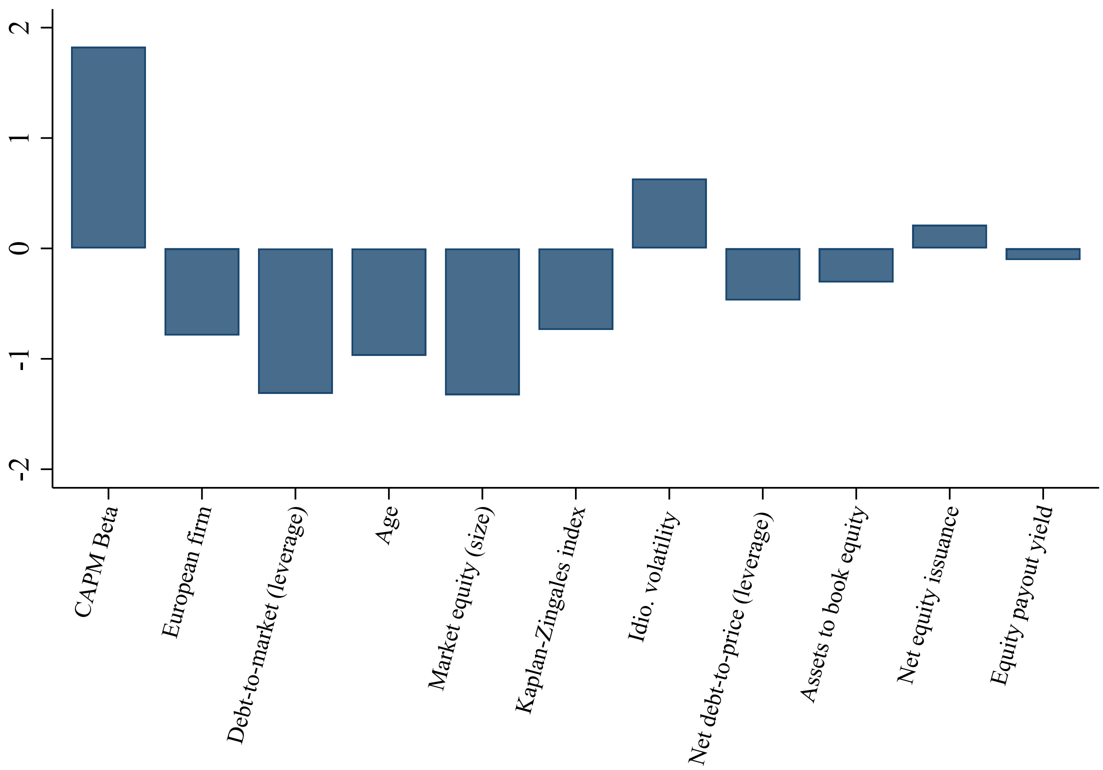

Figure 2
Predictors of Discount Rates

This figure shows slope coefficients of the predicted values of discount rates onto the variables selected by the Lasso procedure. The dependent variable in the Lasso regression is firms' discount rates in a given quarter. We correct all discount rates that are not explicitly at the head-quarter level by subtracting the mean between headquarter and non-headquarter discount rates. The set of possible explanatory variables includes the firms exposure to the 153 risk factors in Jensen et al. (2023) — which risk exposure is measured by firm characteristics — as well as a dummy for the region (U.S. versus European firm). The firm-level characteristics are measured in cross-sectional percentiles of the universe of firms in a given country at the give time. The variable ranges from 0 (lowest) to 1 (highest) and the left-hand side is measured in percentage points, so a loading of 1 means that the perceived cost of capital is predicted to be 1 percentage point higher for firms with the highest characteristics relative to firms with the lowest. The data is fitted based on firms in Europe and US between 2002 and 2021.

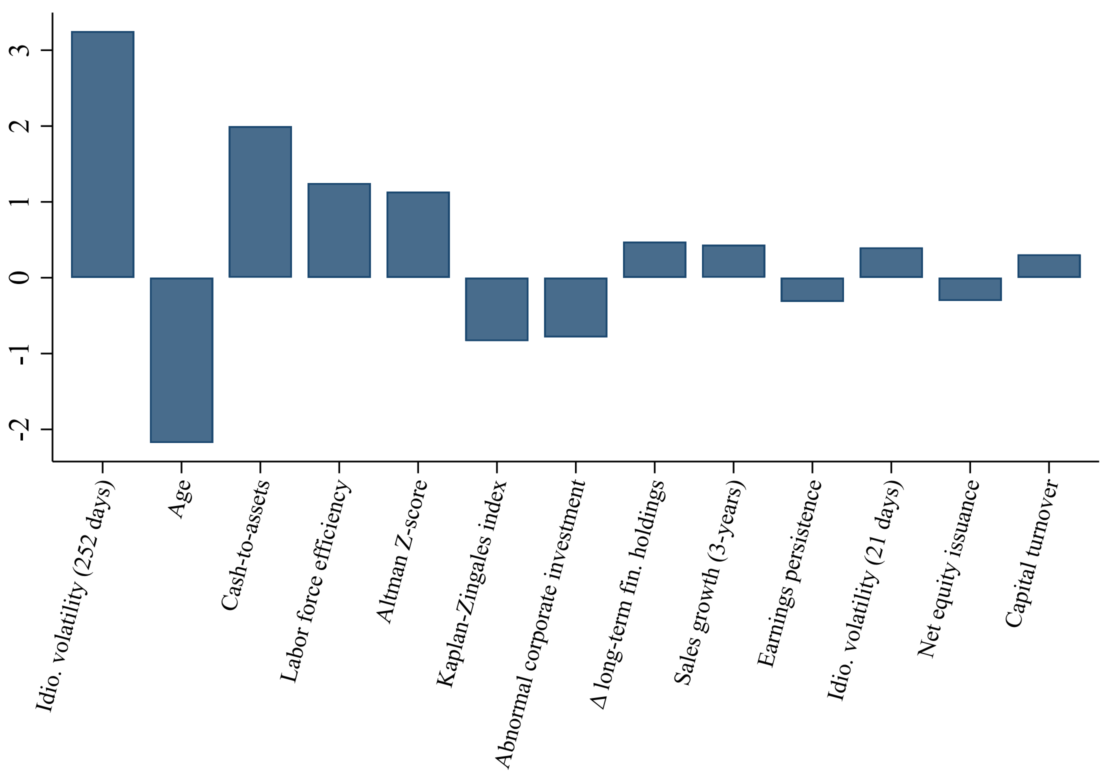

Figure 3
Time Series of Predicted Values in the US

This figure shows the average predicted values for firms' perceived cost of capital and discount rates in the US by year.

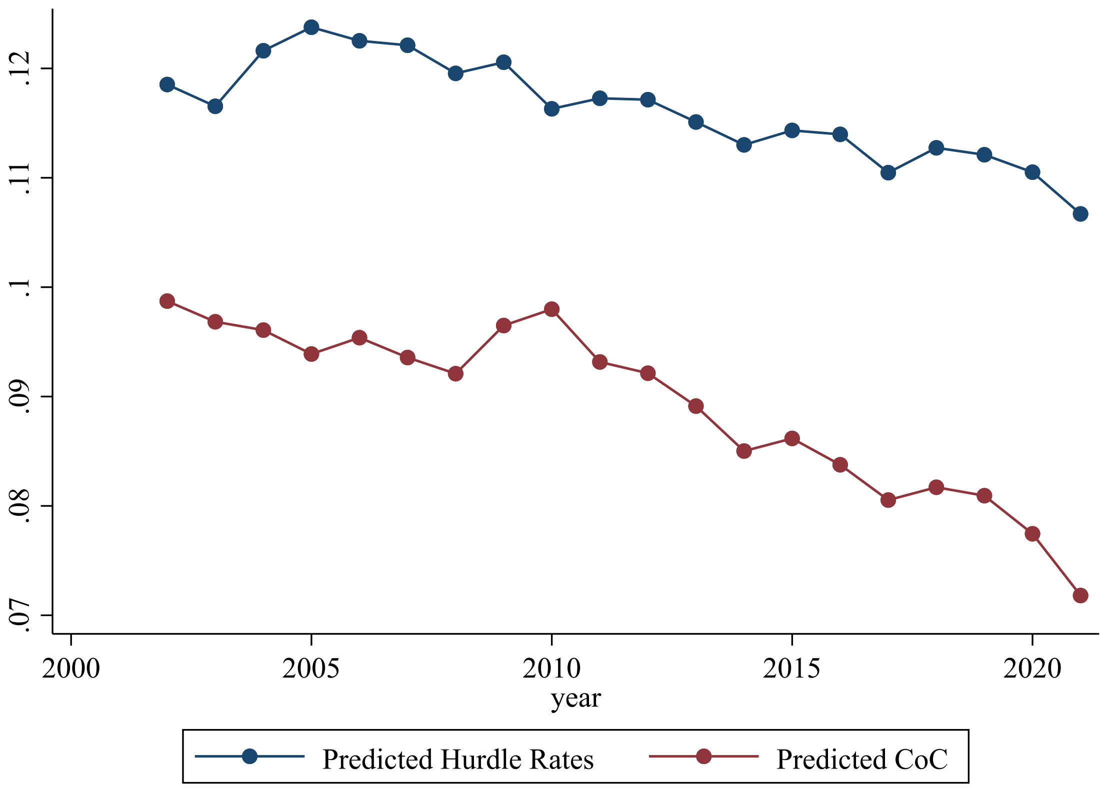

Table 1
Predicted discount rates versus predicted perceived CoC

This table reports results of regressions of predicted values of firms' discount rates on predicted values of firms' perceived cost of capital. The predicted values are calculated based on the Lasso procedure explained in the text.

	(1)	(2)
	Predicted discount rates	
Predicted Perceived CoC	0.92***	0.53***
	(0.054)	(0.035)
Observations	264,922	264,560
R-squared	0.559	0.809
FE	None	Firm
Within R^2	0.56	0.22

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2
Predicting Duke-Data

This table reports results of regressions where we use the predicted value of firms' perceived cost of capital and discount rates to predict the equivalent objects in the Duke-CFO data. The data in this regression are US only.

	(1)	(2)	(3)	(4)
	Duke-CFO Perceived CoC		Duke-CFO Hurdle rate	
Predicted Perceived CoC	0.74***	0.90***		
	(0.17)	(0.21)		
Predicted Discount Rates			1.02***	0.98***
			(0.29)	(0.30)
Constant	0.034**	0.021	0.027	0.031
	(0.014)	(0.018)	(0.030)	(0.031)
Observations	319	319	92	92
R-squared	0.057	0.067	0.118	0.136
FE	None	Year	None	Year
Within R^2	0.057	0.057	0.12	0.11

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1