UNIVERSITÀ COMMERCIALE "LUIGI BOCCONI"
PHD SCHOOL

PhD program in: Economics and Finance

Cycle: 34

Disciplinary Field (code): SECS P/03

Three essays on banking and asset pricing

Advisor: Massimiliano Mariano (Max) Croce

Co-Advisor : Elena Carletti

PhD Thesis by

Paolo Farroni

ID number: 3028023

Academic Year: 2023

Abstract

Chapter 1 of my Thesis studies in a cross-sectional analysis lenders' beliefs. The coauthor and I use a novel loan-level dataset containing borrower-specific probability of default that allows to measure accurately lenders' expectations. We found our empirical analysis on a learning model where bankers endowed with diagnostic expectations observe noisy fundamentals from firms and estimate their probability of default. We provide empirical evidence that financial institutions are subject to expectational distortions: banks tend to overreact to both micro and macro news, overestimating (underestimating) borrowers' defaults after negative (positive) signals. We also document that the degree of overreaction is quite heterogenous among banks. In addition, overreacting bankers decrease (increase) interest rates more than rational ones and the probability of issuing a new loan rises (fall) in light of positive (negative) news. We confirm these results with a structural estimation exercise departing from a model of banking competition where banks' profit function depends on borrowers' creditworthiness, driven by the level of banks' expectation distortion and firm-specific economic news.

In Chapter 2 I develop a structural model of loan demand and lender competition to study how transition risk may affect the Italian credit market. First, I show that transition risk is not currently priced by banks, nor that firms likely more exposed to this risk tend to default more frequently. Then, I use the estimated model to study the effect of policies aimed at more tightly integrating climate-related and environmental risks into banks' business planning. Modeling any such policy as in increase in the cost of lending to "brown" firms, counterfactual analyses show that if these marginal costs

were to increase by one standard deviation interest rates would on average increase by 130 basis points, while quantities would decrease by about 20k EUR.

In Chapter 3 together with coauthors I quantify the exposure of major financial markets to news shocks about global contagion risk accounting for local epidemic conditions. For a wide cross section of countries, we construct a novel dataset comprising (i) announcements related to COVID19 and (ii) high-frequency data on epidemic news diffused through Twitter. Across several financial assets, we provide novel empirical evidence about financial dynamics both around epidemic announcements and at daily/intra-daily frequency. Contagion data and social media activity about COVID19 suggest that the market price of contagion risk is significant. Hence policies that mitigate global contagion or local diffusion may be extremely valuable.

Contents

Chapter 1

Bank beliefs and firm lending: evidence from Italian loan-level data

Abstract

This paper studies in a cross-sectional analysis lenders' beliefs. We use a novel loan-level dataset containing borrower-specific probability of default that allows to measure accurately lenders' expectations. We found our empirical analysis on a learning model where bankers endowed with diagnostic expectations observe noisy fundamentals from firms and estimate their probability of default. We provide empirical evidence that financial institutions are subject to expectational distortions: banks tend to overreact to both micro and macro news, overestimating (underestimating) borrowers' defaults after negative (positive) signals. We also document that the degree of overreaction is quite heterogenous among banks. In addition, overreacting bankers decrease (increase) interest rates more than rational ones and the probability of issuing a new loan rises (fall) in light of positive (negative) news. We confirm these results with a structural estimation exercise departing from a model of banking competition where banks' profit function depends on borrowers' creditworthiness, driven by the level of banks' expectation distortion and firm-specific economic news. [*]

[*]Co-authored with Jacopo Tozzo. The views expressed in the articles are those of the authors and do not involve the responsibility of the Bank of Italy

1 Introduction

Lending decisions reflect what lenders think about borrowers' creditworthiness (Minsky (1986)). Sentiment of the credit supply side as a driving force of the credit cycle is the narrative of an influential and consistent recent literature (López-Salido et al. (2017), Baron and Xiong (2017), Bordalo et al. (2018), Greenwood et al. (2019), Krishnamurthy and Li (2020)). On the other hand, exploiting survey data, Coibion and Gorodnichenko (2015), Gennaioli et al. (2016), Bordalo et al. (2016, 2019) have shown that agents (households, managers and analysts) in many contexts exhibit expectations that deviate from the full-information-rational-expectations (FIRE) hypothesis. Fewer studies have so far focused on the empirical measurement of lenders' expectations, mostly due to the lack of granular data. Three notable exceptions are Bordalo et al. (2018), which uses analysts' forecasts on banks' credit spreads; Richter and Zimmermann (2019), which uses banks' CFOs expected earnings of their own companies to assess lenders' over-optimism about future profitability in a time series setting; Ma et al. (2021), which studies how bankers' aggregatre forecasts about macroeconomic conditions shape their lending activities.

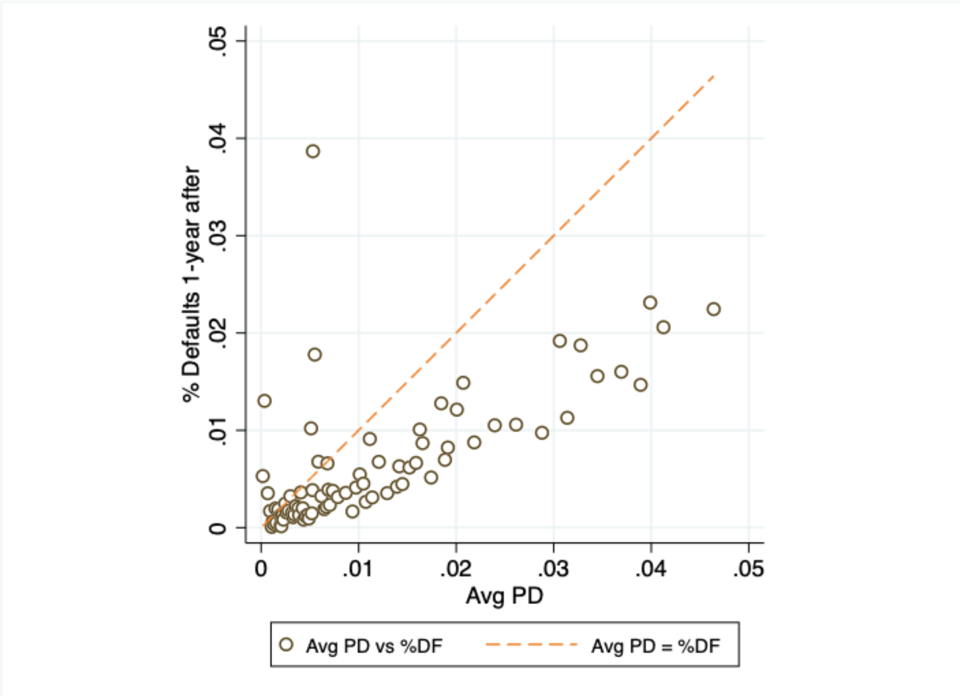
Several questions remain therefore unanswered: why banks distort debtors' creditworthiness? In which cases is this distortion greatest and why? To what extent distortions impact on interest rates and on the probability of issuing new loans?

In this paper, we use a novel granular (loan-level) dataset to measure in a cross-sectional setting lenders' expectations and estimate, for the first time to the best of our knowledge, the effects of beliefs' distortions on interest rates and on the probability of starting new bank-borrower relationships. Specifically, we rely upon the Italian section of the Analytical Credit Dataset (AnaCredit), an on-going credit registry centrally managed by the ECB. In this dataset, crucially, using the internal ratings-based approach (IRB) banks report their assessment of the one-year ahead probability of default (PD) of all firm counterparties in their credit portfolio. We are thus able not only to back out lenders' expectations at a much more granular level (bank-borrower) than the existing literature, but also to exploit a variable that is central in banks' lending decisions. We bring two further relevant contributions to the existing literature. First, we test in a cross-sectional analysis the degree of banks' forecasts distortions using a learning model of extrapolation as in Bordalo et al. (2019); in our setting, banks can overreact to news, meaning that an incoming positive (negative) news may lead lenders to underestimate (overestimate) borrowers' defaults. Second, we quantify the effects of banks' distortions on interest rates and credit allocation. Third, we estimate a structural model of banking competition to conduct counterfactual exercises on the relationship between lenders' belief distortions, interest rates and quantities.

To motivate our paper, we compare the distribution of the one-year ahead PD with that of the realized default rates one year after, by PD centiles. We find that distortions about the default rates are more dispersed for lower centiles, meaning that banks both over and underestimate the probabilities of default. As far as borrowers become riskier instead, overestimation prevails (figure 1.1). In addition, we find that the gap from expectations and realizations is heterogeneous among borrowers. For instance, banks overestimate defaults for borrowers with smaller loan size exposures, younger credit age and located in the South and Center of Italy. Moreover, we document that the autocorrelation of PD estimates is higher in the first deciles and decreases from the fourth onwards, corroborating that the PD varies across time especially for borrowers whose risk is perceived as more intense.

Distortions can be of different nature. Banks' information set about borrowers is a combination of hard data, produced by firms and observable by econometricians, the bank and the borrower, or soft information known to the bank and the borrower only. This is well reflected in a proxy variable like the PD, which is in the first place produced with validated models of credit risk, but at the same time can include borrower-specific soft components. Indeed, credit analysts can and do revise upward or downward the PD provided by credit risk models. In this paper, instead of studying how different sources affect the assessment of PD, we investigate how banks react to news, regardless of their sources, exploring possible over or under-reactions. We can therefore admit among possible explanations for the distortions we document: first, a behavioural-driven motive, which can arise because of a revision of the loan officer after model's outcome is observed and, second, a model-driven hypothesis fed by distorting factors. In either way, we document that distortions are well explained by a learning model of diagnostic expectations.

Figure 1.1. PROBABILITY OF DEFAULT AND REALIZED DEFAULT RATES BY CENTILES



Notes: The image shows the frequency of the probability of default and default rate realized one year after, by centiles. Rational expectations would require points to be on the 45-degree line. Points on the left of the 45-degree line show underestimation of the PD with respect to realized defaults, while points on the right show overestimation. Sorce: our elaborations on AnaCredit.

Following Bordalo et al. (2019), we build a learning model to which we will refer as a diagnostic Kalman filter. Our outcome of interest is a binary indicator of firms' default that occurs whenever cashflows fall below a given threshold known by the bank. Banks do not observe fundamentals directly, such as cashflows, but only a noisy signal. In forecasting defaults, banks' expectations are potentially distorted by the extrapolative nature, which makes bankers inflate the probability of future firm's cashflows whose likelihood has increased the most in light of recent news. After positive (negative) news, the probability that the firm will get an uprise (fall) of cashflows increases, becoming a representative case. The bank over-inflates the likelihood of the credit contract to be a good (bad) asset and, as a consequence, decreases (increases) the firm's probability of default more than it would have done if rational expectations had prevailed. This mechanism relies on the "kernel of truth" property, which has been proved to work quite well in social stereotypes and financial markets, as studied by Bordalo et al. (2016, 2018, 2019): the banker acts in the correct direction of news, but he does it with exaggeration.

Our model allows us to bring this mechanisms to the data and run a set of regressions that tests in a cross-sectional setting the presence of distorted banks' expectations in the assignment of the probability of default, subject to news arrival. First, we assess the average level of banks' diagnosticity, i.e. whether bankers overreact to incoming news when they have to re-calibrate the PD of their borrowers. In particular, we use two types of news: a micro-news, which is given by the quarterly difference in the PD and can be considered as a forecast revision that includes information change attributable to borrowers, and a macro-news, given by the quarterly percentage change of the sector-specific industrial production index released by the Istat, the Italian national statistical institute. In separate robustness exercises, we try alternative measures of news that qualitatively confirm our findings. In our main specification, an incoming standard deviation of micro-news makes the banker overreact on average between 20 to 250 basis points more in the determination of the PD, relative to a non-diagnostic banker[1]. Reaction to macro-news is more contained and goes from 2 to 10 basis points.

Furthermore, we investigate whether the degree of distortion differs across the PD distribution and across banks. We observe that distortions are more pronounced in the tails of the distribution, i.e. for less and more-risky borrowers. We also find that the degree of distortion is different among banks, having in the panel banks that do not exhibit over or under-reaction (that we call "rational"), and banks that do (which we name "diagnostic")[2]. We exploit this distortion heterogeneity when testing our model's predictions on interest rates and quantities and find that on average "diagnostic" banks tend to decrease (increase) interest rates following incoming positive (negative) news between 3.5 and 7 basis points more than "rational" ones. More distorted banks tend also to increase the likelihood of issuing new loans when receiving a positive micro-

[1] The higher effects on absolute value are relative to the negative news.

[2] Following Coimbra and Rey (2017), we potentially identify an additional channel of banks' heterogeneity.

news.

Finally, we extend our reduced-form findings with a structural model of imperfect competition of the banking sector. For the purpose of our paper it is useful to design a model of credit demand and supply to estimate the extent of banks' belief distortions and reproduce counterfactuals. Born for analysing asymmetric information in the loan market, the model by Crawford et al. (2018) serves as a valid starting point in the literature. The model features firms and banks, which generate demand and supply of loans, respectively. Firms ask for loans to finance their risky projects and choose a bank for their main line of credit. They decide how much credit to use and whether to repay or default. Banks compete à la Bertrand-Nash on interest rates. The banks' profit function enriches the model of Crawford et al. (2018) with risky revenues that depend directly on the borrower-specific PD, which we observe and is a function of the belief distortion parameter. We estimate individual firms' demand for loans and banks' pricing and conduct counterfactual exercises varying the level of beliefs' distortion to detect its impact on loan quantities and prices. Doubling the average level of the distortion parameter and conditional on receiving a positive news from firms, our results show that interest rates drop by 42 basis points and the probability of having a new bank-borrower relationship increases by 1.7%, on average.

Literature Review Our paper relates to three main strands of literature. First, it is directly linked to papers that explore bankers' beliefs. Fahlenbrach et al. (2018) and Richter and Zimmermann (2019) examine lenders' expectations through measures of bank's profitability and business activity, loan growth and CEO's expectations. Ma et al. (2021) uses survey data from bankers on MSA's conditions. Our contribution to this literature is measuring more granularly the expectations about the risk assessment of borrowers through the PD, instead of appealing to credit spreads, loan growth or returns on equity measures that are not bankers' direct forecasts. Loan-level data complements more standard survey information on managers expectations about macroeconomic and lending conditions since it represents actual lending decisions, and it can be used to look at how beliefs are heterogeneous across bank- and borrower-characteristics.

Second, we refer to the literature which studies departures from full information rational expectations and diagnostic expectations: Gennaioli and Shleifer (2010), Gennaioli et al. (2012), Greenwood and Shleifer (2014), Coibion and Gorodnichenko (2015), Gennaioli et al. (2016),Bordalo et al. (2016), Bordalo et al. (2019), Bordalo et al. (2020). We add to this line of research an empirical insight on lenders' beliefs using micro data. We are able to study how beliefs vary on the basis of borrowers' characteristics and show that lenders expectations overreact to news.

Third, our paper relates to the literature on credit cycle and sentiment. The importance of lenders beliefs' in credit supply has been introduced by Minsky (1977) and Kindleberger (1978), who laid the foundation of financial crisis and irrational manias. After the financial turmoil of 2008, this literature has developed extensively, with the works of Baron and Xiong (2017), López-Salido et al. (2017), Bordalo et al. (2018), Greenwood et al. (2019), Krishnamurthy and Li (2020). Our analysis does not cover an entire credit cycle, nonetheless our results are indicative through the counterfactual exercises (and conservative in estimates) of what can happen during boom and bust phases: an increase of positive/negative news would amplify the overreaction of creditors, leading to intensified distortions in loans' prices and quantities.
We refer also to a structural estimation literature, in which the main source of inspiration for our model is Crawford et al. (2018).

The paper proceeds as follows: section 2 describes data and stylized facts, section 3 presents the econometric model. Section 4 exhibits our main findings. Section 5 illustrates the results from the structural estimation exercise and section 6 presents robustness exercises.

2 Data

2.1 Anacredit

The main dataset used in this project is the Italian section of AnaCredit, which is a credit registry managed by the ECB with the aim of collecting detailed and fully harmonized monthly information on individual loans granted by euro area banks to legal entities whose total debt exposure exceeds 25,000 euros. The project to establish a euro-area credit registry was initiated in 2011 and data collection started in September 2018.

For all credit contracts banks are asked to report a wealth of information concerning, *inter alia*, the outstanding amount of loans and the interest rates charged on these loans; for each borrower banks are asked to report several characteristics among which the sector of economic activity (2-digit Ateco), the age and the geographical location and also the default status, which in our setting is a binary indicator.

Furthermore, banks that use the so-called Internal Ratings Based approach (IRB - Basel Committee (2001)) also report each month the 1-year ahead probability of default (PD) for each borrower. Since the PD is the key variable in our empirical analysis, we restrict our attention to Italian IRB banks that overall account for around 80% of total assets. Every month we have on average banks' PDs for 760,000 borrowers. Table 1.1 contains several summary statistics about the dataset.

Data ranges from June 2018 onwards. The main analysis uses data until the start of the Covid-19 in Italy (Q2 2020)[3].

Other datasets used are Italian credit registry, Cerved credit data and Istat.

2.2 Istat

From Istat we retrieve the index of industrial production in Italy. This index is released monthly at Nace 4-digit level (NACE activities B, C and D) and collects volumes of production from mining and manufacturing for firms with more than 20 employees. The measure can be considered as a macro news that banks receive from these sectors. We can only use the Nace 2-digit granularity to match the index with our bank-firm data. The measure of news is generated as the percentage quarter difference of the index for each 2-digit sub-sector:

$$News_t^s = \frac{idx_t^s - idx_{t-1}^s}{idx_{t-1}^s}$$

[3] We expanded the analysis also beyond the beginning of Covid-19. Full-sample findings can be found in section 6.

Table 1.1. ANACREDIT SUMMARY STATISTICS

	1st quartile	Median	3rd quartile
N Borrowers	748,741	762,871	781,723
N Bank-Borrowers	7,104	27,098	80,491
N Bank per Borrower	1	2	3
Def. Rate (%)	0.75%	0.87%	1.10%
PD	0.34%	0.94%	2.42%
Loan Size (EUR k)	33.19	84.12	255.25
Int. Rate (%)	1.10%	2.41%	4.98%

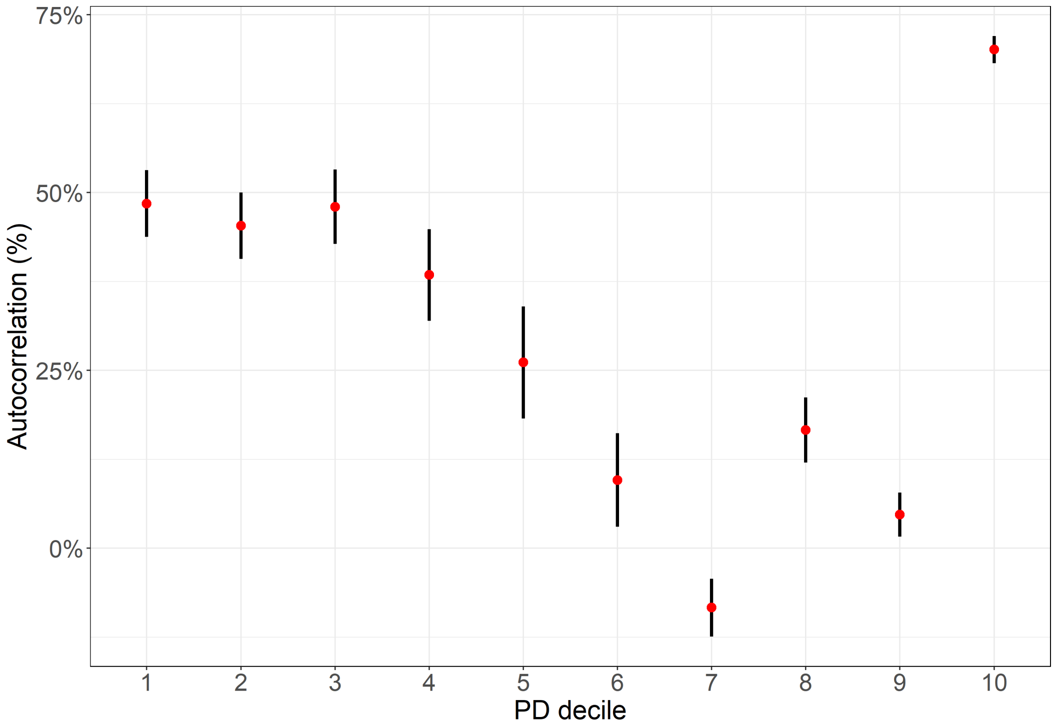
Notes: this table provides basic summary statistics of the dataset used in the paper, by quartiles. Default rate, PD and interest rate are expressed in percentages, while loan size in thousands of euro.

2.3 PD in the data

As anticipated in the introduction, the motivating evidence for the investigation in bankers' beliefs formation is given by figure 1.1. The discrepancy between the actual default rate and the probability of default makes us question about the differences across the distribution. Table 1.3 compares the lowest ad highest deciles of firms by bankers' PD forecast error. On average, bankers' forecast errors are lower on the first decile of the distribution (even if more dispersed, as shown in 1.1), while they tend to widen for the highest decile. On the top decile the average PD is around 5.9%, while bottom decile errors concern low-risky firms (average PD around 0.7%). The average loan size difference among two deciles is euro 67k, with average size higher among less risky firms. Credit age[4] is almost three years higher for bottom decile PD error, while the error is significantly more pronounced for firms located in the South and Center of Italy. Overall, bankers seem to err more on firms that are ex-ante riskier, smaller, with lower credit age, located in the Center and South and not operating in the manufacturing sector (a full list of NACE sectors in available in Appendix).

How does PD change along time and across the distribution? The autocorrelation of the PD by deciles, shows interestigly that for the first three deciles the autocorrelation coefficient is high and stable at 50%, while from the fourth decile on it starts to decrease quite monotonically (figure 1.2). The high correlation in the last decile is due to pinpointed firms as in the process of failure, who see receiving high PD until the failure is made official. This picture is instructive for our ultimate exercise on bankers' distorted belief: it shows that information retrieval plays a role in the forecasting of PD because bankers actually change it over time, on the basis of observable data.

Figure 1.2. PD Autocorrelation



Notes: the figure shows autocorrelation of borrower-specific PD on y-axis, by PD decile on x-axis.

Table 1.3. SUMMARY STATISTICS OF LOWEST AND HIGHEST DECILES BY PD FORECAST ERROR

	Bottom Decile by PD error	Top Decile by PD error	Top - Bottom
Avg. PD	0.007^{***}	0.059^{***}	0.052^{***}
	(0.000)	(0.000)	(0.000)
Avg. Def. Rate	0.006^{***}	0.038^{***}	0.032^{***}
	(0.000)	(0.003)	(0.003)
Avg. Loan Size (EUR k)	235.90^{***}	168.72^{***}	-67.18^{***}
	(0.084)	(0.122)	(0.087)
Avg. Credit Age	14.462^{***}	11.617^{***}	-2.845^{***}
	(0.369)	(0.384)	(0.328)
Agriculture Sect.	0.047	0.052	0.005
			(0.006)
Construction Sect.	0.097	0.139	0.041
			(0.038)
Manufacturig Sect.	0.268	0.175	-0.094^{***}
			(0.029)
Other Sect.	0.015	0.016	0.001
			(0.003)
Services Sect.	0.572	0.619	0.046
			(0.040)
Geo: Center	0.190	0.228	0.038^{***}
			(0.005)
Geo: North-East	0.260	0.204	-0.056^{***}
			(0.010)
Geo: North West	0.396	0.372	-0.024^{***}
			(0.007)
Geo: South	0.155	0.197	0.042^{***}
			(0.008)

Notes: this table provides summary statistics of firms in the bottom and top deciles of the bankers' PD forecast errors (given by the share of realized defaults less the average probability of default). For industry sectors and geographical locations the table reports the average frequency distribution of borrowers in the relative sector/geo. area. Standard errors are in parenthesis and are clustered at NACE 2 digit-level. Significance levels at 1%, 5%, 10% are given by (***), (**), (*) respectively.

governance, and model screening ability can use the IRB approach. The PD originating from these models is a measure upon which banks found their business and supervisory authority control capital requirements needed to ensure a valid assessment of risk. After an initial approval process, supervisory authorities (the Single Supervisory Mechanism (SSM) for Significant Institutions, and National Competent Authorities (NCAs) for Less Significant Institutions) regularly validate these models to ensure their on-going respect of prudential requirements[5].

So, the PD is a measure produced by credit risk models and can be revised judgementally by loan officers. The model of expectations described in the next section to explain forecast distortions embeds a mechanism that overweights most recent news coming from fundamentals.

3 Econometric model

We build a learning model that mimics how banks estimate borrowers' PD. If cashflows fall below a given threshold, the firm defaults. Banks do not directly observe firm's cashflows, but only a noisy signal upon which banks try to forecast default. We add representativeness in bankers' expectations on the basis of Bordalo et al. (2019), to capture how banks can produce distorted PDs. Before introducing the distorted learning process, we design a baseline Kalman filter applied to our case. Suppose the firm's cash flow follows an AR(1) process x_t but the bank cannot observe the process directly, rather only a noisy signal y_t:

$$\begin{aligned} x_{t+1} &= \rho x_t + v_t \qquad v_t \sim N(0, \sigma_v^2) \\ y_t &= x_t + w_t \qquad w_t \sim N(0, \sigma_w^2) \end{aligned} \tag{1.1}$$

where v_t and w_t are the state and measurement errors, respectively.
Standard Kalman derivation gives the following recursions in Durbin and Koopman (2012)[6]:

$$\begin{aligned} \hat{x}_{t+1|t} &= \rho \hat{x}_{t|t-1} + K_t I_t \\ \widehat{\Omega}_{t+1|t} &= \rho \widehat{\Omega}_{t|t-1}(\rho - K_t) + \sigma_v^2, \qquad K_t = \frac{\rho \widehat{\Omega}_{t|t-1}}{\widehat{\Omega}_{t|t-1} + \sigma_w^2} \end{aligned} \tag{1.2}$$

[5] For further details, we refer to Basel Committee (2001).
[6] Steps of the derivation can be found in ch.4.3, pp. 82-85

where $\hat{x}_{t|t-1} = \mathbb{E}[x_t|y^{t-1}]$, $\widehat{\Omega}_{t|t-1} = \mathbb{E}(x_t - \hat{x}_{t|t-1})^2$ and y^{t-1} is the information set available to bankers at time $t-1$ formed by all signals $y_{t-1}, y_{t-2}, \ldots$
We denote the innovation by $I_t = y_t - \mathbb{E}(y_t|y^{t-1}) = y_t - \hat{x}_{t|t-1}$ and the Kalman Gain by K_t.
Notice that K_t in (1.2) converges to a steady state value after few iterations in the model. Therefore, we assume $K_t = K$ to be a constant in the rest of the paper.

Diagnostic Expectations Diagnostic Expectations is based on the concept of representativeness heuristic of Kahneman and Tversky (1972). An element is representative (or diagnostic) in a class whenever it is diagnostic, i.e. its relative frequency in that class is much higher than in another reference class. Gennaioli and Shleifer (2010) built an analytical model describing representativeness applied to belief formation. We refer to Bordalo et al. (2018) for an analytical description of representativeness applied to time-varying economic variables.

Assume that the agent forms beliefs about an economic random variable following an AR(1) process $x_{t+1} = \rho x_t + \epsilon_t$ with $\epsilon_t \sim N(0, \sigma^2)$ and $\rho \in (0, 1)$. The agent assesses the distribution of future state $\hat{x}_{t+1}$ on the basis of realized current state $x_t = \hat{x}_t$. The rational agent predicts the future state using the true conditional distribution $f(x_{t+1}|x_t = \hat{x}_t)$. The diagnostic agent instead has the true distribution $f(x_{t+1}|x_t)$ in the back of his mind, however he selectively recovers and overweights the realizations of the state at $t+1$ that are representative in t. A given state $\hat{x}_{t+1}$ is more representative at t if it's more likely that it occurs under the realized state $(x_t = \hat{x}_t)$ than on the basis of past information $(x_t = \rho\hat{x}_{t-1})$. Hence, *representativeness* of $\hat{x}_{t+1}$ is given by:

$$R = \frac{f(\hat{x}_{t+1}|x_t = \hat{x}_t)}{f(\hat{x}_{t+1}|x_t = \rho\hat{x}_{t-1})} \tag{1.3}$$

The state is more representative the more its likelihood increases with respect to recent news. In case of absence of news, numerator and denominator coincide leading to the rational expectation case. When the news is good, states in the right tail of the distribution are made more representative, when the news is bad the opposite is true.

The overweighting states process is rationalized as if the agent uses a distorted density

$$f_t^{\theta}(\hat{x}_{t+1}) = f(\hat{x}_{t+1}|x_t = \hat{x}_t) \cdot \left[\frac{f(\hat{x}_{t+1}|x_t = \hat{x}_t)}{f(\hat{x}_{t+1}|x_t = \rho\hat{x}_{t-1})} \right]^{\theta} Z$$

The formula embeds what is defined as the "kernel of truth" property, i.e. the agent shifts its beliefs from rational expectations in the direction of the news received. Parameter θ measures the degree of *diagnosticity*, the deviation from the rational expectation case. Z is a constant ensuring that the distorted density integrates to one.

Back to our model, following Bordalo et al. (2019), we can characterize bankers' beliefs by the distorted density

$$f^{\theta}(x, I_t) = f(x, I_t)[R(x, I_t)]^{\theta} Z$$

where x represents firms' cashflows and I_t is the information received at t; $R(x, I_t)$ is the level of representativeness, as in equation (1.3). When $\theta > 0$ the agent is diagnostic and over-reacts to information with respect to previous period, if $\theta = 0$ the agent is rational. Given the linearity of the process (1.1) the rational density $f(x, I_t)$ is normal with variance $\widehat{\Omega}$ and mean $\hat{x}_{t+1|t}$. Following Bordalo et al. (2019), we can characterize the diagnostic density $f^{\theta}(x, I_t)$ as normal with the same variance $\widehat{\Omega}$ and mean

$$\begin{aligned} \hat{x}^{\theta}_{t+1|t} &= \rho\hat{x}_{t|t-1} + (1+\theta)KI_t \\ &= \hat{x}_{t+1|t} + \theta KI_t \end{aligned}$$

3.1 Kalman filter and the Probability of Default

To compute the probability of default we define z as the default status of any firm: $z_{t+1} = \mathbb{1}(x_{t+1} < a)$. The firm defaults whenever cashflows x_{t+1} are strictly lower than a given threshold $a \in \mathbb{R}$. It follows that the probability of the firm's default is given by

$$\mathbb{E}(z_{t+1}|y^t) = \mathbb{E}_t(z_{t+1}) = \mathbb{P}_t(x_{t+1} < a)$$

Given beliefs $f(x, I_t)$ and $f^{\theta}(x, I_t)$ (see proof in Appendix - Proofs) we obtain the predicted probability of default for a rational and diagnostic agent[7], respectively. Notice that Φ and ϕ stand for cumulative distribution and density function of a standard normal.

$$\begin{aligned} \mathbb{E}_t(z_{t+1}) &= \Phi\left(\frac{a - \hat{x}_{t+1}}{\Omega_t^{1/2}}\right) = \widehat{PD}_{t+1|t} \\ \mathbb{E}_t^{\theta}(z_{t+1}) &= \Phi\left(\frac{a - \hat{x}_{t+1}^{\theta}}{\Omega_t^{1/2}}\right) = \widehat{PD}_{t+1|t}^{\theta} \end{aligned} \tag{1.4}$$

From the definition of PD in 1.4, applying some algebra and approximations (see proof in Appendix - Proofs), we obtain an equation that links directly the innovation I_t to bankers' forecast error $FE_{t+1|t}^{\theta,i} = z_{t+1} - \widehat{PD}_{t+1|t}^{\theta}$ with respect to the probability of default. Then, for each firm $i = 1, \dots, N$ and bank $b = 1, \dots, B$ we have

$$FE_{t+1|t}^{\theta,i,b} \approx K\theta \frac{1}{\widehat{\Omega}^{1/2}} \phi\left(\frac{a}{\widehat{\Omega}^{1/2}}\right) I_t^{i,b} + w_{t+1}^{i,b} \tag{1.5}$$

Now, define $\beta_1 := K\theta \frac{1}{\widehat{\Omega}^{1/2}} \phi\left(\frac{a}{\widehat{\Omega}^{1/2}}\right)$. By construction $\widehat{\Omega}_t > 0$, $a > 0$, $K > 0$ and the density is strictly positive. Therefore the only term that could make $\beta_1 = 0$ is the diagnostic parameter θ. For $\theta > 0$ the agent overreacts to incoming news $I_t^{i,b}$. As a consequence, we can test the hypothesis $H_0 : (\beta_1 = 0)$ with the following linear regression

$$FE_{t+1|t}^{\theta,i,b} = \beta_0 + \beta_1 I_t^{i,b} + \epsilon_{t+1}^{i,b} \tag{1.6}$$

At each fixed point in time t, with regression (1.6) we are able to determine whether in our cross-sectional dataset banks respond to firms' news with overreaction measured through the parameter θ. Empirical results are given in section 4.

[7] As highlighted in the previous paragraph, the agent provided with diagnostic expectations perceives a process that is the distributed as $f^{\theta}(x, I_t) = f(x, I_t)[R(x, I_t)]^{\theta} Z$ with mean $\hat{x}_{t+1|t}^{\theta}$.

3.2 Learning process, representativeness and real effects

We adapt our learning model to real effects, in particular how it influences the interest rates setting for banks that are endowed with diagnostic expectations.

Consider a simple one-period loan when borrowers promise to repay tomorrow $a = L(1+r)$ for a loan today of size L. Assuming competition deprives lenders of any surplus we have:

$$\begin{aligned} L &= \mathbb{E}[a \cdot \mathbb{1}\{x_{t+1} > a\}] \\ &= a(1 - \widehat{PD}_{t+1|t}) \end{aligned}$$

We also know that the repayment at $t+1$ will be equal to the loan at $t = 0$ plus a positive interest rate r_t, such that

$$a = L(1 + r_t)$$

Combining the two equations above we get an expression for the risky interest rate, such that:

$$r_t = \frac{\widehat{PD}_{t+1|t}}{1 - \widehat{PD}_{t+1|t}}$$

This equation allows us to derive a direct relationship between the interest rate set by banks and the probability of default implied by the noisy firms' cashflow signal

$$r_t = \frac{\Phi\left(\frac{a - \hat{x}_{t+1}}{\widehat{\Omega}_t^{1/2}}\right)}{1 - \Phi\left(\frac{a - \hat{x}_{t+1}}{\widehat{\Omega}_t^{1/2}}\right)}$$

After some algebra and approximations given in Appendix - Proofs, we obtain a linearized relationship between interest rate and the probability of default, both for rational and diagnostic agents:

$$r_t \approx \Phi\left(\frac{-a}{\widehat{\Omega}^{1/2}}\right) - \frac{1}{\widehat{\Omega}^{1/2}} \frac{\phi\left(\frac{-a}{\widehat{\Omega}^{1/2}}\right)}{\Phi\left(\frac{-a}{\widehat{\Omega}^{1/2}}\right)^2} \hat{x}_{t+1|t} \tag{1.7}$$

$$r_t^\theta \approx r_t - \frac{K\theta}{\widehat{\Omega}} \frac{\phi\left(\frac{-a}{\widehat{\Omega}^{1/2}}\right)}{\Phi\left(\frac{-a}{\widehat{\Omega}^{1/2}}\right)^2} I_t \tag{1.8}$$

Equations (1.7) and (1.8) differentiate by the innovation I_t and relative multiplicative parameters. Given positive parameters $K, \Phi(\cdot), \phi(\cdot), \widehat{\Omega}$ by construction, for a positive innovation $I_t > 0$ our model predicts a lower interest rate for the diagnostic agent compared to the rational one.

4 Empirical Results

We preface that while the model forecast horizon is one time period for simplicity, given the nature of the probability of default in our dataset, in the empirical specifications we have a 12 months forecast horizon. Our sample starts in mid-2018 ending in 2019-Q2 to discard counfounding effects of the Covid-19 in the main analysis; results with the full sample are available in section 6.

For an empirical assessment of the model we adapted the equation (1.6) to our data, which brings to equation (1.9). The dependent variable is given by the banker's forecast error $FE^\theta_{t+12|t} := z_{t+12} - \widehat{PD}^\theta_{t+12|t}$, where $z_{t+12} = \mathbb{1}(x_{t+12} < a)$ is a dummy that takes value one if the firm defaults at $t+12$ and zero otherwise, and $\widehat{PD}^\theta_{t+12|t}$ is the probability of default for firm i by a banker with diagnostic expectations.

$$FE^{\theta,i,b}_{t+12|t} = \beta_0 + \beta_1 News^{i,b}_t + \mathbf{\Gamma}'\mathbf{X} + \epsilon^{i,b}_{t+12} \tag{1.9}$$

Controls and bank, sector, province, borrower, time fixed effects are contained in $\mathbf{\Gamma}'\mathbf{X}$. The main regressor $News_t$ is a measure of innovation that the bank receives about each firm i in each period t.

We remark that under rational expectations bankers' forecast errors should not be predictable using variables in the bankers' information set. At the borrower level, we choose as a proxy for the model-based news I_t the one-quarter probability of default difference at the time the forecast $\widehat{PD}^{\theta}_{t+12|t}$ is made, i.e.

$$News_t = -(\widehat{PD}^{\theta}_{t+12|t} - \widehat{PD}^{\theta}_{t+9|t-3}) = -\Delta\widehat{PD}^{\theta}_t$$

This measure captures any new information each banker has incorporated at time t with respect to $t-1$ into the valuation variable used to predict the default status. The negative sign in front of the expression makes $News_t$ a positive news, since a positive $\Delta\widehat{PD}^{\theta}_t$ means higher probability of default, hence a deterioration of credit worthiness.

To corroborate our findings, we use an alternative aggregate measure of news based on the industrial production index for Italian firms.

We also tried different variables as proxies for the innovation, left for a robustness exercise in the Robustness section. We validate our borrower-specific measure of $News_t$ in the Appendix - Proofs.

Each section of table 1.5 presents results from equation (1.9), with data selected on the basis of the sign of the news: all news in Panel A, only negative and positive news in Panel B and C respectively. The main regressor is the news coefficient, which is statistically significant and positive for the three panels that include borrower fixed effects (far-right column). In Panels A and B the effect is also robust for every other specification and the magnitude is higher when we consider only negative news in Panel B. In Panel C the coefficient flips to the right sign and becomes significant when we introduce borrower fixed effects: this is important, because it suggests that even if demand-driven components are dampened, expectational distortions by banks in the direction of over-reaction still arise. This result strengthens the motivation of using such granular dataset in studying lenders' beliefs.
A positive and significant coefficient rejects the null of $\theta = 0$ and proves that bankers overreact to both positive and negative news about their borrowers. With positive θ the agent forms forecast with diagnostic expectations: he receives a news through a noisy signal and inflates the probability of those states that became more likely in

light of recent news. When the banker gets a positive news, he tends to decrease the probability of default more than he would have done if rational. The converse happens in case of negative news.

Results in Panel A of table 1.5 suggest that for a standard deviation increase in news (so news becoming more positive), the forecast error of a diagnostic banker increases between 20 and 250 basis points more than a non-diagnostic banker. In other words, for a one s.d. more positive news, bankers forecast a default rate between 0.2% and 2.5% lower than what would have a rational forecaster.

We use loan size and credit age as controls in the regression, and time, bank and province fixed effects for specifications with no borrower fixed effects. The credit age coefficient is significant and negative, reconciling with findings of the summary statistics for bottom and top deciles by PD error in Table 1.3: bankers tend to err less with respect to firms with higher credit age. Having presumably more information on these firms, bankers tend to be more accurate when assessing their creditworthiness. With respect to loan size instead, we find that bankers overreact to incoming news irrespective of the magnitude of new firms' exposures.

We complement these main results with two alternative exercises: (1) explore if overreaction to news is different across the probability of default distribution and (2) whether it entails considerable real effects on prices. The following paragraphs are focused on these aspects.

4.1 News effect across the distribution

To complement the previous analysis we conduct a focus on the cross-sectional effects of the news. Our model (1.9), allows to test if the overreaction to news is different across the distribution, both relative to banks and borrowers heterogeneity. It is indeed likely that banks overreact to news differently on the basis of being a particular bank or observing at distinct firm characteristics, geographic locations and credit relationships. The first paragraph gives an insight on bank's, while the second one on firm's heterogeneity.

Table 1.5. Predictability on forecast errors - PD news

	$FE^{\theta,i}_{t+12\|t}$			
	Panel A: All PD News			
$News_t(all)$	0.274***	0.273***	0.274***	0.485***
	(0.0227)	(0.0226)	(0.0226)	(0.00643)
N Obs.	1036314	1036314	1036314	1034841
	Panel B: Negative PD News			
$News_t < 0$	0.562***	0.567***	0.562***	0.946***
	(0.116)	(0.0443)	(0.0442)	(0.0157)
N Obs.	239009	239008	239009	224402
	Panel C: Non-Negative PD News			
$News_t \geq 0$	-0.115***	-0.117***	-0.113***	0.0671***
	(0.0181)	(0.0187)	(0.0183)	(0.0129)
N Obs.	797305	797304	797305	794910
Time FE	Yes	Yes	Yes	Yes
Bank FE	Yes	No	Yes	No
Sector FE	No	Yes	No	No
Province FE	No	No	Yes	No
Borrower FE	No	No	No	Yes

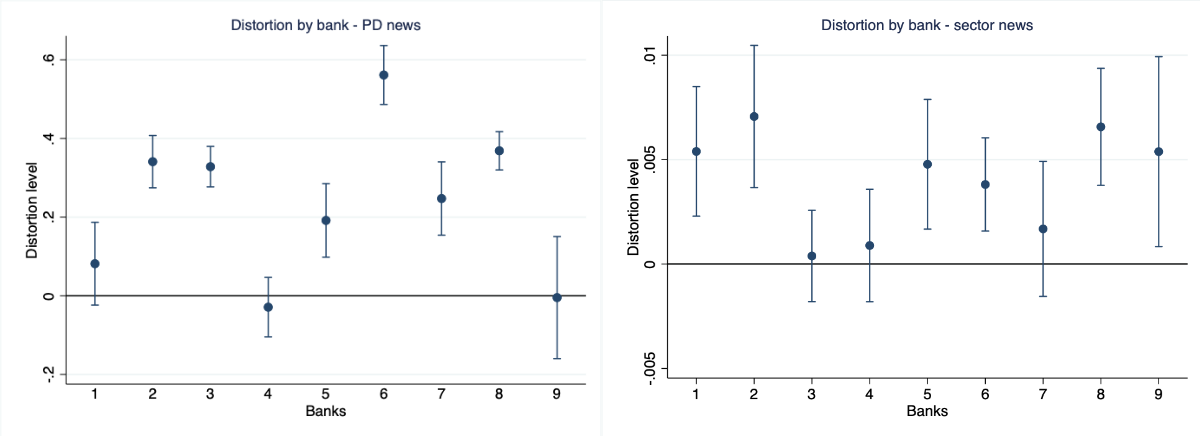
Notes: this table provides coefficient estimates of the regression $FE^{\theta,i,b}_{t+12|t} = \beta_0 + \beta_1 News^{i,b}_t + \mathbf{\Gamma}'\mathbf{X} + \epsilon^{i,b}_{t+12}$, where $\mathbf{X}$ is the controls' matrix that contains also fixed effects. Controls used are loan size, firm credit age, post-Covid-19. Main regressor *News* is borrower specific. Errors are clustered at NACE 2-digit level. Significance levels at 1%, 5%, 10% are given by (***), (**), (*) respectively.

Table 1.6. Predictability on forecast errors - Sector news

	$FE^{\theta,i}_{t+12\|t}$		
	Panel D: All Sector News		
$News_t(all)$	0.00395***	0.00449***	0.00107*
	(0.000938)	(0.00109)	(0.000403)
N Obs.	505920	505920	505330
	Panel E: Negative Sector News		
$News_t < 0$	0.0105*	0.0101*	-0.00407
	(0.00443)	(0.00433)	(0.00326)
N Obs.	291952	291952	187295
	Panel F: Non-Negative Sector News		
$News_t \geq 0$	0.00613***	0.00702	0.0000911
	(0.00140)	(0.00355)	(0.00166)
N Obs.	213968	213968	212577
Bank FE	No	Yes	No
Province FE	No	Yes	No
Borrower FE	No	No	Yes

Notes: this table provides coefficient estimates of the regression $FE^{\theta,i,b}_{t+12|t} = \beta_0 + \beta_1 News^s_t + \mathbf{\Gamma}'\mathbf{X} + \epsilon^{i,b}_{t+12}$, where $\mathbf{X}$ is the controls' matrix that contains also fixed effects. Controls used are loan size, firm credit age, post-Covid-19. Main regressor *News* is sector specific. Errors are clustered at NACE 2-digit level. Significance levels at 1%, 5%, 10% are given by (***), (**), (*) respectively.

Figure 1.3. DISTORTION COEFFICIENTS BY BANK

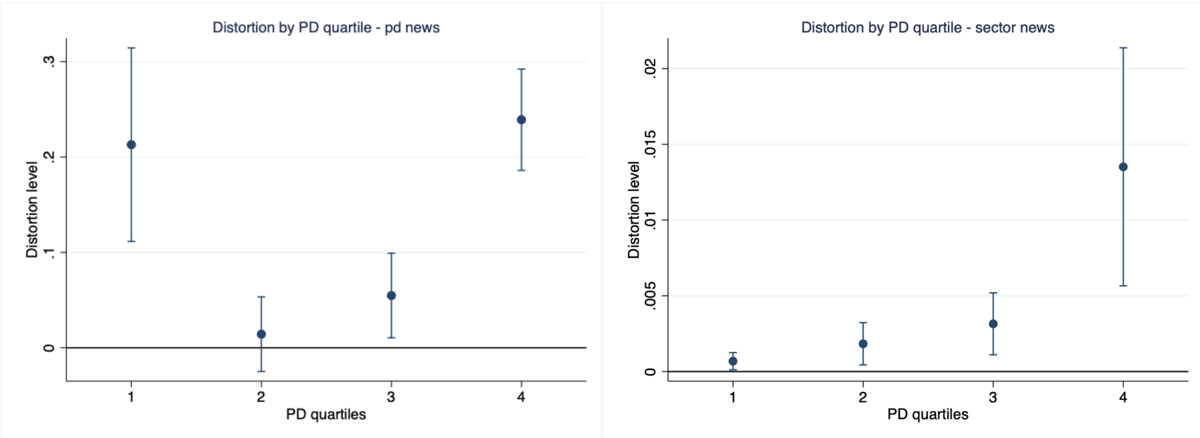


Notes: The figure plots the coefficients $\hat{\beta}_1$ with 95% confidence interval of the regression $FE^{\theta,i}_{t+12|t} = \beta_0 + \beta_1 News^i_t + \mathbf{\Gamma}'\mathbf{X} + \epsilon^i_{t+12}$, estimated by bank. The blue line represents the cutoff between high and low θ banks, i.e. banks with a diagnostic parameter above and below the median. Banks are sorted by $\hat{\theta}$. Standard errors are clustered at NACE 2 digit-level. For confidentiality reasons banks are anonymised and are assigned a cardinal identifying number.

Summary by bank diagnostic level To investigate heterogeneity among banks, we run regression (1.9) for each bank, to determine a bank-specific diagnostic level. Results are given in figure 1.3, where we sort banks by $\hat{\theta}$. Results show that six out of nine banks display a positive and significant parameter: these banks overreact when receiving positive or negative news from their customers in attributing them a new probability of default. The degree of overreaction is different, based on the nature of news received. From figure 1.3 indeed, banks non reacting to micro-news are 1, 4 and 9, while those non reacting to macro-news are 3, 4 and 7. The variability of the coefficient differs between the two sub-figures, more pronounced among the micro-news based coefficients. This is not surprising, being the micro-news borrower-level dependent. Overall, overreaction to news seems diffuse among the Italian panel of Anacredit and confirms that results of the previous section are not driven only by a singular sizable institution.

News effect (theta) by PD quartile In figure 1.4, we estimate regression (1.9) by quartile and plot the news coefficients. Coefficients of distortion based on micro-news are more significant and pronounced in the first and fourth quartiles about PD distribution, while are monotonically increasing based on macro-news. The two sub-figures share that banks over-react more with respect to riskier borrowers, independently on the news type received.

Figure 1.4. DISTORTION BY QUARTILE



Notes: The figure plots the coefficients $\hat{\beta}_1$ with 95% confidence interval of the regression $FE^{\theta,i,b}_{t+12|t} = \beta_0 + \beta_1 News^{i,b}_t + \mathbf{\Gamma}'\mathbf{X} + \epsilon^{i,b}_{t+12}$, estimated by PD decile. Standard errors are clustered at NACE 2 digit-level.

4.2 Real Effects

Interest rates A natural question about the importance of studying distortions in expectation formation mechanisms is whether they may yield considerable real effects. We try to address this point in the following exercises. First, we simply regress interest rates on the level of news, to measure how new information impacts bankers' evaluation of credit price, unconditionally. Second, we test whether interest rates set by diagnostic banks receiving news, are different from those set by rational ones.

From equations (1.7) and (1.8), we derived a regression to measure the impact of diagnostic parameter on the level of interest rates.

$$r^{i,b}_t = \beta_0 + \beta_1 D^b_t + \beta_2 News^{i,b}_t + \gamma(D^b_t \times News^{i,b}_t) + \mathbf{\Gamma}'\mathbf{X} + \epsilon^{i,b}_t \tag{1.10}$$

where $D^b_t = \mathbb{1}\{\theta^b = \text{“}high\text{''}\}$ identifies banks with high level of distortion. The idea is to test whether diagnostic expectations measured through different parameters θ

have heterogenous effects on interest rates. To pursue this test, we: (1) estimate θ^b for each bank b by means of equation (1.9), (2) sort banks by level of diagnosticity θ, (3) select banks rational and non-rational banks (θ statistically different from zero) and (4) run regression (1.10), whose coefficient of interest γ gives us the impact of innovation absorbed through diagnostic expectations on the level of interest rates. Notice that, for each date t, we select only new contracts stipulated among banks and borrowers who had already an existing credit relation. We restrict to new contracts only because it is not possible to identify news effects on prices on outstanding contracts. Therefore, the banker receives information about the borrower between $t-1$ and t and formulates an interest rate for the new contract in t.

Table 1.8 contains two sections with results on interest rates. The first column shows a simple regression between interest rates and news only (controlled by several variables), which is not derived by any model. We are interested in a first place in assessing the "unconditional" role of news on price changes. The effect of innovation on interest rate is negative, as expected, but not statistically significant: positive news make bankers more optimistic about firms outcomes' and price to new loans are reduced accordingly.

Results in columns 2-4 suggest that the interaction coefficient between news and diagnostic firms $News_t \times D_t^b$ is negative and statistically significant at the 10%, 5%, 1% levels respectively. The interpretation of this coefficient reads as follows: distorted banks compared to rational ones, conditional on the arrival of one standard deviation of positive news, tend to decrease on average the interest rates to his borrowers between 3.5 to 6.8 basis points on first contract signed [8]. In the last column, borrower-fixed effects are introduced to capture any potential unobserved demand-driven effect hidden to the econometrician.

In panel B we run the same regression, substituting the borrower-specific news with the sector-specific one. The level of significance for the coefficient of interest is lower

[8]The effect of the estimate is computed by multiplying the standard deviation of the news to the coefficient. The value of the sd(News)=0.02 in panel A, while 0.017 is the value of the interaction coefficient in panel A when borrower-fixed effects are introduced. The total effect on interest rate can be read as $0.2 * 0.017 = 3.5bp$. Standard deviation may slightly change depending on how data are selected for the ongoing exercise.

from column 2 to 4, but in the last specification, where the coefficient is significant at the 1% level, the magnitude is comparable to that in panel A: an increase of one standard deviation in news causes a 7 basis points[9] additional decrease in the interest rate offered on new loans by diagnostic banks.

To conclude, expectations-distorted firms receiving positive borrower-specific news produce an impact on interest rates that on average goes from 3.5 to 7 basis points on new signed contracts. This effect is much lower (and not significant from our specification), when the nature of the bank's expectations is not taken into account.

Quantities Similar to the exercise in the previous paragraph, we test whether the level of distortion can impact the bank's probability of issuing new contracts[10]. We derive a regression of equation (1.10) type, where the dependent variable is a *new contract*. The idea is to test whether a distorted bank receiving a positive news tends to have a different lending behaviour with respect to a rational one. In the regression we run $NC_t^{i,b} = 1$ if the contract is new and 0 otherwise, while the regressors take the same meaning of the rate regression (1.10) in a linear probability framework.

$$NC_t^{i,b} = \beta_0 + \beta_1 D_t^b + \beta_2 News_t^{i,b} + \gamma(D_t^b \times News_t^{i,b}) + \mathbf{\Gamma}'\mathbf{X} + \epsilon_t^{i,b} \tag{1.11}$$

The main coefficient of interest is the intersection between distorted bank and the level of news. Table 1.9 shows in panel A the coefficients when using the firm-specific news: unconditionally, when positive news reaches banks, the probability of a new contract increases, regardless if a bank is rational or diagnstic. Moreover, when a distorted bank receives a positive news from firms, it tends to increase the probability of signing new contracts more than their rational peers. To quantify this effect, as in the rates exercise, we multiply one standard deviation of news to the coefficient estimate. Being affiliated to an expectation-distorted bank increases (reduces) the probability of signing a new contract by 0.4% to 0.6%. Note that the first coefficient $News_t$ in the panel has not the expected sign from columns 2 to 4.

[9]Standard deviation of macro news is different from that of micro, this is why different coefficients leads to the same marginal effects on interest rates.

[10]Here we do not restrict the panel to new contracts only.

Table 1.8. EFFECTS ON INTEREST RATES

	$r^{i,b}$			
	Panel A: PD News			
$News_t$	-0.00694	0.000338	0.00556	0.00471
	(0.00450)	(0.00546)	(0.0102)	(0.00611)
D_t^b		0.00212***	0.00166***	-0.00101***
		(0.000123)	(0.000602)	(0.000264)
$News_t \times D_t^b$		-0.0279***	-0.0338**	-0.0169*
		(0.00638)	(0.0166)	(0.00946)
N Obs.	186096	190596	190596	186096
Sector FE	No	No	Yes	No
Province FE	No	No	Yes	No
Time FE	Yes	No	No	Yes
Borrower FE	Yes	No	No	Yes
	Panel B: Sector News			
$News_t$	0.00396***	0.00474***	0.00374**	0.00645***
	(0.000999)	(0.000789)	(0.00155)	(0.00128)
D_t^b		0.00591***	0.00507***	0.00420***
		(0.000179)	(0.000679)	(0.000453)
$News_t \times D_t^b$		-0.00121	-0.000395	-0.00321***
		(0.000885)	(0.00135)	(0.000910)
N Obs.	111334	112080	112080	111334
Sector FE	No	No	Yes	No
Province FE	No	No	Yes	No
Borrower FE	Yes	No	No	Yes

Notes: this table provides estimates of interest rates on news regression. First column shows results of unconditional regression. 2-4 columns exhibit estimates of regression $r_t^{i,b} = \beta_0 + \beta_1 D_t^b + \beta_2 News_t^{i,b} + \gamma(D_t^b \times News_t^{i,b}) + \mathbf{\Gamma}'\mathbf{X} + \epsilon_t^{i,b}$, where $\mathbf{X}$ is a control matrix which contains also fixed effects. Panel A uses PD news (borrower-specific), panel B sector-specific news. Errors are clustered at the NACE 2-digit level. Significance levels at 1%, 5%, 10% are given by (***), (**), (*) respectively.

Regarding panel B we do not obtain expected results when the news is aggregate at sector-level, but we have to admit the exercise is ambitious, since the measure of news is very aggregate with respect to the contract-level granularity of the dependent variable.

5 Structural estimation

We extend our reduced form findings with a model of imperfect competition of the banking sector. Designing a model of credit demand and supply is crucial to estimate the extent of expectations' distortions on real effects and to reproduce counterfactuals. We borrow the structural design from Crawford et al. (2018), developed to analyse asymmetric information in the loan market, specifically adverse selection. The model is appropriate for our goal since it allows to introduce lending imperfect competition. The empirical environment is familiar too, since the application is over the Italian banking market.

The model is composed of firms and banks. Demand of credit is represented by firms, which ask for loans to finance a risky project to a single bank for their main line of credit. They decide how much to use of the credit line and whether to repay or default. Banks compete a-lÃ Bertrand-Nash on interest rates. The banks' profit function of our model differs from the model of Crawford et al. (2018) for risky revenues, which in our case depend on borrower's specific probability of default and level of measurable information received. As outlined in the reduced form specification, the PD is in turn function of bank-specific belief distortion. Therefore, we retrieve belief distortion levels from the empirical analysis, we observe distorted riskiness associated to each contract, we use the structural estimation to obtain and quantify the real effects of these distortions on prices and quantities and conduct counterfactual exercises. In the model we adopt several important assumptions: first, we narrow the analysis on the first credit line (visible in the data) each firm opens with banks. We do this to avoid the dynamic dimension and reduce the complexity of the problem. Second, we assume both firms and banks are risk-neutral. Third, banks compete only on the interest rate. In markets with lending exclusivity bank can offer contracts that depend

Table 1.9. Effects on Quantities

	$NC^{i,b}$			
	Panel A: PD News			
$News_t$	0.112***	-0.0821***	-0.0702	-0.0759*
	(0.0104)	(0.0268)	(0.0508)	(0.0422)
D_t^b		-0.0120***	-0.00973	-0.0103*
		(0.000573)	(0.00621)	(0.00553)
$News_t \times D_t^b$		0.225***	0.210***	0.155**
		(0.0291)	(0.0695)	(0.0594)
N Obs.	2075790	2075790	2075790	2075747
Sector FE	No	No	Yes	No
Province FE	No	No	Yes	No
Time FE	Yes	Yes	Yes	Yes
Borrower FE	No	No	No	Yes
	Panel B: Sector News			
$News_t$	-0.0347***	-0.0692***	-0.0646***	-0.0454***
	(0.00168)	(0.00561)	(0.0133)	(0.0109)
D_t^b		0.0244***	0.0272***	0.0178***
		(0.00122)	(0.00328)	(0.00452)
$News_t \times D_t^b$		0.00308	0.000927	-0.0165
		(0.00626)	(0.0153)	(0.0125)
N. Obs	1206816	667225	667225	667169
Sector FE	No	No	Yes	No
Province FE	No	No	Yes	No
Time FE	Yes	Yes	Yes	Yes
Borrower FE	No	No	No	Yes

Notes: this table provides estimates of interest rates on news regression. First column shows results of unconditional regression. 2-4 columns exhibit estimates of regression $NC_t^{i,b} = \beta_0 + \beta_1 D_t^b + \beta_2 News_t^{i,b} + \gamma(D_t^b \times News_t^{i,b}) + \mathbf{\Gamma}'\mathbf{X} + \epsilon_t^{i,b}$, where $\mathbf{X}$ is a control matrix which contains also fixed effects. Panel A uses PD news (borrower-specific), panel B sector-specific news. Errors are clustered at the NACE 2-digit level. Significance levels at 1%, 5%, 10% are given by (***), (**), (*) respectively.

both on credit amount and price. Instead, with our assumption the amount of credit is exogenous and given only by the firm's project requirements. As in Crawford et al. (2018), the Italian credit market justifies this assumption, since it is not a market with lending exclusivity, as firms can open multiple credit lines with different banks. As in Chiappori and Salanié (2013), with no contract exclusivity convex price schedule cannot be enforced.

Demand Firms $i = 1, \ldots, I$ operate in markets $m = 1, \ldots, M$ representing geographical provinces, where each bank $j = 1, \ldots, J$ supply loans. Demand estimation is composed of one main equation that represents firm's utility from the credit line. It depends on loan price and market-bank characteristics.

$$U^D_{ijm} = \alpha^D_0 + X'^D_{jm}\beta^D + \xi^D_{jm} + \alpha^D P_{ijm} + Y'^D_{ijm}\eta^D + \nu_{ijm}$$

where X_{jm} is vector of bank-mkt characteristics; P_{ijm} is interest rate offered by bank j to firm i and market m; ξ are bank-market characteristics unobservables to the econometrician; Y'^D_{ijm} are firm-bank-market characteristics.

Supply On the supply side, banks compete a-lÃ Bertrand-Nash on prices and set for each market m and firm i an interest rate P_{ijm}. Bank's j expected profits from firm i is

$$\Pi_{ijm} = P_{ijm}Q_{ijm}(1 - PD(\theta_j, I_i)) - MC_{ijm}Q_{ijm}$$

Q_{ijm} represents the expected demand for loan, given by demand probability times expected amount of loan used by firm i and MC_{ijm} is the marginal cost the bank pays on issuing the loan. Probability of default $PD(\theta_j, I_i)$ depends from the bank-specific parameter of belief distortion θ_j and firm's news I_i. The first order condition of the profit function reads as

$$P_{ijm} = \frac{MC_{ijm}}{1 - PD_{ijm}(\theta_j, I_i)} + \frac{\mathcal{M}_{ijm}}{1 - PD_{ijm}(\theta_j, I_i)}$$

where $\mathcal{M}_{ijm} = -Q_{ijm}/Q'_{ijm}$ is the bank's j markup on firm i loan. The equation tells us that the interest rate is formed of an effective marginal cost and a markup components, similarly to Bertrand-Nash pricing equation. The difference is represented by the risky nature of the projects that imply the presence of the probability of default on both components.

Remind that the probability of default depends negatively (positively) on positive (negative) news and positive belief distortion. The pricing equation tells us that, conditional on having a positive news, distorted beliefs ($\theta > 0$) tend to reduce both the marginal cost and the markup components. High level of competition implies low margins, which induce the belief distortion to have an effect mainly through the marginal cost channel. On the other hand, when competition is low and markups are high, beliefs' distortion can help to mitigate the markup component in good times (positive news), but exacerbating it in bad times (negative news).

Estimation of demand requires knowledge of contract prices, which give rise to several considerations. First, the borrower-bank price observed in our dataset is the equilibrium price, but to estimate the model, prices offered from banks not chosen by firms are also needed. Second, it is likely there are unobserved characteristics to us econometricians on the demand-side. Following Crawford et al. (2018), we adopt measures to avoid the risk of incurring in inaccurate price predictions.

Loan pricing reflects borrower specific components, such as customer's riskiness, bank-specific characteristics, as the degree of expectations' distortion, and bank-borrower relationship features. The price prediction is tightly linked to how we treat information in the bank-borrower-econometrician relationship. Crawford et al. (2018) claim that the determinants of loan prices are a combination of *hard* information, those observed by firms, banks and econometricians, and *soft* information, which are unobserved by the econometrician, but known by banks and borrowers. Designing a loan pricing model bears the risk of neglecting some of the information that could be in possess of the bank, but invisible to us (*soft*).

To mitigate this concern, first note that banks in our panel follow the IRB approach and it is reasonable to believe they make predominantly use of *hard* information (even if the *soft* component cannot be removed a priori though). A large survey by Albareto et al. (2011) indeed shows how large banks in Italy tend to use the following source of information to assess the creditworthiness of new loan applicants, by order of importance: 1- financial statement data, 2- credit relations with the entire system, 3- statistical-quantitative methods, 4- qualitative information, 5- availability of guarantees, 6- first-hand information (branch-specific). Second, we include in the analysis only the first and main credit line a firm borrows, to omit any dynamic from the bank-borrower relationship. Also, we introduce firm fixed effects to absorb any borrower-specific component unobservable to the econometrician. The institutional environment favours the use of fixed effects, given that the Italian market is strongly characterized by multi-affiliated borrowers (confirmed by our data, where single borrower-bank relationships account only for around 10%). After this premise, we can now present the price prediction model: price P_{ijm} charged to firm i by firm j in market m is an OLS model as described by equation (1.12):

$$P_{ijm} = \gamma_0 + \gamma_1 T_{ijm} + \gamma_2 L_{ijm} + \lambda_{jm} + \omega_i^p + \tau_{ijm} \tag{1.12}$$

where $\omega_i^p, \lambda_{ijm}$ are firm and bank-area-time fixed effects, T_{ijm} is tenure of relationship between borrower i and the bank j in market m; L_{ijm} is loan size and τ_{ijm} are prediction errors. Using estimated coefficients of (1.12) we can predict prices $\tilde{P}_{ijm}$ offered from banks that firms decided to discard.

Another required exercise is predicting prices for non-borrowing firms. We adopted a propensity score matching, using similar characteristics between borrowing and non-borrowing firms to predict price of contracts that would have been offered to firms that have not received them. Similarly, we use the same method to retrieve information and probability of default for firms with no relations with some banks.

First stage estimation We estimate the demand for credit lines in a two-step estimation, as in Train (2009). In the first step we estimate the firm-level parameters and recover bank-market specific constants with the contraction method as in Berry et al. (1995), which represents the dependent variable of the second-step estimation, recovering the price coefficient α^D in the demand function (5).

Estimation faces two obstacles: first, endogeneity of price should be taken into account; second, as we did in the price prediction equation, we need to account for potential "soft" information, unobserved by the econometrician. Besides the prediction accuracy, it is important to account for possible *soft* information since they could give rise to omitted variable problem in the demand estimation. In what follows we try to get rid of this issue, as in Crawford et al. (2018).

The price prediction equation allows to disentangle between a bank-market and bank-market-borrower component:

$$P_{ijm} = \tilde{P}_{ijm} + \tilde{\tau}_{jm}$$
$$P_{ijm} = \tilde{P}_{jm} + \tilde{\gamma}_1 T_{ijm} + \tilde{\gamma}_2 L_{ijm} + \tilde{\omega}_i^p + \tilde{\tau}_{jm}$$

where the term $\tilde{\omega}_i^p$ is estimated firm fixed effects from pricing equation. Since "soft" information are observed by bank (and not by us), we can include them in a variable $\omega^D = \eta_4^D \omega_i^p$, dependent on the component responsible for pricing.

All of the firm level components determining the demand are then given by:

$$Y_{ijm}^D = \eta_1^D T_{ijm} + \eta_2^D L_{ijm} + \eta_3^D Y_i + \eta_4^D \tilde{\omega}_i^p$$

Including the last two equations in the demand estimation equation yields:

$$\begin{aligned}
U_{ijm}^D &= \delta_{jm}^D + \alpha^D(\tilde{P}_{jm} + \tilde{\eta}_1 T_{ijm} + \tilde{\gamma}_2 L_{ijm} + \tilde{\omega}_i^p + \tilde{\tau}_{jm}) + \\
&\quad \eta_1^D T_{ijm} + \eta_2^D L_{ijm} + \eta_3^D Y_i + \eta_4^D \tilde{\omega}_i^p + \nu_{ijm} \\
&= \underbrace{(\delta_{jm}^D + \alpha^D \tilde{P}_{jm})}_{\tilde{\delta}_{jm}^D} + \underbrace{(\eta_1^D + \alpha^D \tilde{\eta}_1)}_{\tilde{\eta}_1^D} T_{ijm} + \underbrace{(\eta_2^D + \alpha^D \tilde{\gamma}_2)}_{\tilde{\eta}_2^D} L_{ijm} + \\
&\quad \eta_3^D Y_i + \underbrace{(\eta_4^D + \alpha^D)}_{\tilde{\eta}_4^D} \tilde{\omega}_i^p + \underbrace{\alpha^D \tilde{\tau}_{jm} + \nu_{ijm}}_{\tilde{\zeta}_{ijm}} \qquad (1.13) \\
&= \tilde{\delta}_{jm}^D + \underbrace{Y_{ijm}'^D \tilde{\eta}^D}_{V_{ijm}^D} + \zeta_{ijm} \\
\Rightarrow U_{ijm}^D &= \tilde{\delta}_{jm}^D + V_{ijm}^D + \zeta_{ijm}
\end{aligned}$$

Parameters $\tilde{\eta}^D$ are a mixture of direct effect of firm and firm-bank covariates on demand and indirect effects through pricing. Differentiating these channels in step 2 of the estimation gives demand-only specific parameters η^D. In addition, as standard in the literature, we assume error ζ_{ijm} is distributed as a type I extreme value. Finally, parameter α^D must be estimated in the second step of the estimation, since not part of equation (1.13) independently. Probability that borrower i chooses bank j in market m is then given by:

$$Pr_{ijm}^D = \frac{\exp(\hat{\tilde{\delta}}_{jm}^D(X_{jm}^D, \tilde{P}_{jm}, \xi_{jm}^D, \alpha^D, \beta^D) + V_{ijm}^D(Y_{ijm}^D, \tilde{\eta}^D))}{1 + \sum_l \exp(\hat{\tilde{\delta}}_{jm}^D(X_{jm}^D, \tilde{P}_{jm}, \xi_{jm}^D, \alpha^D, \beta^D) + V_{ijm}^D(Y_{ijm}^D, \tilde{\eta}^D))} \qquad (1.14)$$

where $V_{ijm}^D = Y_{ijm}'^D \tilde{\eta}^D$ and $\hat{\tilde{\delta}}_{jm}^D$ are specific constants recovered through the contraction method from Berry et al. (1995).

Second stage estimation We use instrumental variable estimation to recover structural parameters in demand equation. In the first stage we find constants $\hat{\tilde{\delta}}_{jm}^D$, which contain bank-market-time covariates X_{jm}^D and bank-market-time specific component of predicted prices $\tilde{P}_{jm}$. We IV-regress constants on bank-market components using cost-shifters as instruments, where cost-shifters are interest rates on deposits:

$$\hat{\tilde{\delta}}^D_{jm} = \alpha^D_0 + \alpha^D \tilde{P}_{jm} + X'^D_{jm}\beta^D + \xi^D_{jm}$$

where ξ^D_{jm} is the structural error term. As indicated in Crawford et al. (2018), unobserved structural error term can be interpreted as the borrower's unobserved valuation of bank's characteristics, affecting bank's interest rates. ξ^D_{jm} can also include market specific errors. Bank and market fixed effects could solve this endogeneity concern. However, correlation between these bank-market errors can be solved through the use of an instrumental variable that represent households' deposits. Households' deposits are an important source of banks' capital and affect the lending conditions of branches[11]. The exclusion restriction is given by the fact that households' deposits respond to different market characteristics than the firm loans. Hence, as the instrumental variable for loan prices we use bank specific interest rate on households' deposits.

Estimation and results Besides estimation of demand described in the paragraphs above that accurately follows the work of Crawford et al. (2018), our estimation is characterized by a slightly different supply equation. Equation (3.3) is dependent on the borrower's creditworthiness and nests both the level of the bank specific expectations' distortion θ_j and the borrower information I_i. We can define the level of distorted probability of default as a function of the rational probability of default plus a distortion parameter that guides the reaction to firm-specific news. Note that for this equation and the estimation results the interpretation of the coefficient goes in the other direction: when news is positive, the level of PD for distorted banks decreases more than for rational ones, as a direct effect of overreaction. We are opting for this formulation because the firm-specific news and the level of belief distortion never enter independently in our economic model, rather only through the probability of default. Expressing the distorted PD as the composition of a rational PD and a theta-dependent parameter which reacts to news, allow us to include both variables in the model and estimate the coefficient of belief distortion. Equation (1.15) is mathematically derived

[11]See Albareto et al. (2011)

as equation (1.6):

$$PD^{\theta}_{ji} \approx PD^{re}_{ji} + \beta(\theta) I_i \tag{1.15}$$

Estimates of the structural model are outlined in table 1.10. Upper part contains demand parameters, including firm characteristics, while the bottom part supply ones. As expected, the average price coefficient is negative and significant meaning that higher interest rates negatively impact demand for loans. Other significant parameters are borrower unobserved characteristics, tenure of the relationship, age and sales of the firm. At the same time, increase of distortion (given by parameter *Belief Distortion*), causes an increase of loan demand though the dampening of probability of default assigned by banks.

We further conduct some counterfactual exercise where we make vary several components to the detect the response of the model; results are given by table 1.11. As a first exercise we double the level of beliefs' distortion to understand the reaction of loan quantities and prices. Results show that doubling the level of distortion, conditional on receiving a positive news from firms, interest rate tend to drop by 42 basis points and the probability of having a new bank-borrower relationship increases by 1.7%, on average.

The second exercise we run through the model consists in increasing the news by one standard deviation. Receiving a positive one standard deviation news makes diagnostic banks decrease price by 32.4 basis points and increase the likelihood of new bank-borrower relationship by 4.7%, compare to the average rational. Results for a negative news are almost symmetric. In the empirical analysis our findings display instead a higher level of asymmetry in favour of the negative news. Third, we shut down the distortion parameter for the banks identified as distorted in the reduced form analysis, and see how these banks react in prices and quantities to a median positive news. The reaction our model suggests is an increase in prices and a mild reduction in quantities. In absence of their distortion, diagnostic banks would price their loans on average 167 basis points more than a rational bank. The three exercises above strengthen the reduced form findings of section 4, confirming that expectational errors in the banks' prediction of the probability of default is a channel well identifiable

Table 1.10. STRUCTURAL ESTIMATION - RESULTS

		Prob. borr-bank relationship
Demand param.	Tenure	1.658^{***}
		(0.181)
	Previous rel.	1.403^{***}
		(0.387)
	Constant	0.940
		(15.644)
	Share branches	0.988
		(1.913)
	Avg. Price	-1.442^{***}
		(0.519)
	Borrower FE	0.899^{***}
		(0.220)
	Age	0.888^{***}
		(0.147)
	log Sales	0.890^{**}
		(0.396)
	log Asset	0.890
		(1.202)
	Debt Eq.	0.899^{***}
		(0.136)
Supply param.	Const. (Bel. dist.)	0.039^{***}
		(0.000)
	Belief distortion	$\mathbf{-0.599}^{***}$
		(0.018)
	Const. (Deposit int. rate)	1.003
		(0.873)
	Deposit int. rate	1.000
		(13.065)

This table presents estimate of the structural model.

Table 1.11. COUNTERFACTUALS - RESULTS

	ΔP	ΔQ
Exercise 1		
News	-0.419^{***}	0.017^{***}
	(0.162)	(0.003)
Bank FE	Yes	Yes
Market FE	Yes	Yes
Exercise 2		
Diagn. Bnk $\|\Delta News > 0$	-0.324^{***}	0.047^{***}
	(4.141)	(0.314)
Diagn. Bnk $\|\Delta News < 0$	0.268^{***}	-0.051^{***}
	(4.380)	(0.346)
Exercise 3		
Median News	1.671^{*}	-0.004^{*}
	(0.999)	(0.002)
Bank FE	Yes	Yes
Market FE	Yes	Yes

This table shows coefficient estimates of the structural model for three different counterfactual exercises investigating the effects on prices and quantities on diagnostic banks, keeping the rational banks as benchmark. In Exercise 1 we double the size of the average estimated expectational distortion parameter *theta* for diagnostic banks, conditional on receiving a positive news. In the Exercise 2 we perturb the model with a *News* increase of one standard deviation, both positive and negative. In Exercise 3 we shut down the coefficient *theta* for previously identified diagnostic banks and see how their lending decisions would react in absence of the expectation distortion.

through a structural model of lending imperfect competition.

6 Robustness

We conduct several robustness exercises to strengthen our main results. First, we try to mitigate the concern that PD does not deviate from realized default rates only because of banks' strategic behaviour. Second, we try an alternative measure of news with respect to the two used in the main specifications. Third, we use the entire dataset length, so including Covid-19, to investigate how results may vary. Overall we do not find significant variations and findings confirm outcomes of the main analysis.

6.1 PD and strategic behaviour

One concern when looking at IRB PDs (the PD in Anacredit, we call it in this paragraph PD^{IRB}) is that banks may systematically under-report their "true" credit risk assessment to minimize capital requirements (Behn et al. (2021)). While we cannot completely rule out banks' strategic behaviour, we take several steps to mitigate this concern.

First, looking at figure 1.1 and table 1.5, if anything, banks seem to *over estimate* the probability of default, at least in our sample period. Second, we compare our PD^{IRB} to another probability of default, which banks use to compute the expected loss of a borrower according to the IFRS 9 accounting principle, and that here we will call PD^{EL}. PD^{EL}, which is computed quarterly, is *not* used to compute capital requirements and therefore should not be subject to the same degree of strategic behaviour as PD^{IRB}. Note that the PD^{EL} is unobservable in AnaCredit. What we can observe is the "rating" class[12] Sn assigned to a specific borrower by the bank: $S1$ corresponds to borrowers with low credit risk, $S2$ to borrowers with a significant increase in credit risk but still performing, and $S3$ to defaulted borrowers. The rating class is directly linked to PD^{EL}, so we can use the observed class as a good proxy for the IFRS 9 associated probability of default. From one period to another, if the PD^{EL} changes, we are able to observe it through the corresponding change in the assigned rating class Sn.

Our test is as follows: if a bank recognizes a significant increase in credit risk of some counterparty, which corresponds to a worsening of rating from $S1$ to $S2$, and if IRB models are consistent with accounting practices, we should observe a consistent change in PD^{IRB} too. In our specification we select the subsample of borrowers that migrate from $S1$ to $S2$. We then use as a dependent variable the quarterly change of the PD^{IRB}, ΔPD^{IRB}_{t+3} and some controls as regressors. Table 1.12 shows the results: a positive and significant intercept has to be interpreted as a positive correlation between

[12]With a slight abuse of terminology we adopt the term "rating" in place of the more correct "staging". Since staging is a loan-level outcome, we pool together loans' staging for each firm to get a borrower-specific measure.

Table 1.12. TEST ON BANKS' STRATEGIC BEHAVIOUR

	ΔPD^{IRB}_{t+3}					
Intercept	3.617***	3.565***	3.829***	3.996***	3.759***	4.182***
	(0.142)	(0.174)	(0.677)	(0.708)	(0.221)	(0.794)
N Obs.	145,429	145,429	145,429	145,429	145,429	145,429
Bank FE	No	Yes	No	Yes	Yes	Yes
Time FE	No	No	Yes	Yes	No	Yes
Sector FE	No	No	No	No	Yes	Yes

Notes: This table reports the coefficients of the following regression: $\Delta PD^{IRB,i,b}_{t+3} = \beta_0 + \mathbf{\Gamma}'\mathbf{X} + \epsilon^{i,b}_t$ where $\mathbf{X}$ is a vector of controls including *total loans* and *credit age*. The regression is estimated only on the subsample with a $\Delta PD^{EL} > 0$: a positive and significant intercept means that whenever banks increase their PD^{EL} we observe a parallel increase in PD^{IRB}, too. Standard errors are clustered at 2-digit NACE sectors.

the variation in PD^{EL} and PD^{IRB}. This finding suggests to reject that banks are not overly strategic when reporting the PD^{IRB} to the supervisory authority.

6.2 News proxy with IFRS9 accounting data

As in the previous section, we use the rating class Sn given by IFRS9 accounting data for a different scope. We aim to find a measure that replaces the news measure $News_t$ for an additional robustness exercise. We look again at the subset of borrowers who flow from one rating class Sn to another as a signal of null/negative/positive news. Borrowers who pass to a more-risky rating class constitute a negative news ($D1 =$ Rating Decrease), those who pass to a less-risky rating class a positive one ($D2 =$ Rating Increase) for the bank. Borrowers who see their rating class unchanged represent the baseline case of no news. Notice that, since $D1$ signals negative news, the expected right coefficient for overreaction would be of negative sign (an overreaction to negative news induce a higher-than-due PD, hence a negative forecast error).

$$FE^{\theta,i,b}_{t+12|t} = \beta_0 + \beta_1 D1^{i,b} + \beta_2 D2^{i,b} + \mathbf{\Gamma}'\mathbf{X} + \epsilon^{i,b}_{t+12}$$

When we introduce fixed effects, the coefficients of both subgroups are statistically significant and correct in sign, as confirmed in table 1.14. The arrival of positive or negative news induced by the release of IFRS9 data makes bankers overreact.

Table 1.14. TEST ON ALTERNATIVE NEWS MEASURE

	$FE^{\theta,i,b}_{t+12\|t}$					
Rating Decrease	-0.002	-0.001	-0.002	-0.002	-0.001	-0.028^{***}
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.002)
Rating Increase	-0.002	-0.001	-0.002	-0.002	-0.000	0.004^{***}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N Obs.	1,550,735	1,550,735	1,550,735	1,550,735	1,550,735	821,889
Bank FE	No	Yes	No	No	Yes	No
Sector FE	No	No	Yes	No	Yes	No
Province FE	No	No	No	Yes	Yes	No
Borrower FE	No	No	No	No	No	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table provides coefficient estimates of the regression $FE^{\theta,i,b}_{t+12|t} = \beta_0 + \beta_1 D1^{i,b} + \beta_2 D2^{i,b} + \mathbf{\Gamma}'\mathbf{X} + \epsilon^{i,b}_{t+12}$, where $\mathbf{X}$ is the controls matrix can include *loan size* and *credit age* and bank, sector, province and/or borrower fixed effects. Standard errors are clustered at NACE 2 digit-level. Significance levels at 1%, 5%, 10% are given by (***), (**), (*) respectively.

6.3 Using the full sample

As an additional test, we replicate our main results from regressions (1.9) and (1.10) extending our sample to 2021(Q1), i.e. including also the Covid-19 shock[13]. Our results are left unaffected to those found in the pre-Covid sample. Table 1.16 confirms the overreaction of bankers' to news arrival; given an increase in the news standard deviation, the forecast error increases by 420 basis points. Table 1.18 instead, shows a very similar result to that one obtained in the main analysis.

One possible explanation for the very high degree of overreaction using the full-sample can be that banks, under the Covid-19 shock revised upward PDs while re-

[13] We believe it is reasonable to pinpoint the first data under Covid-19 with the third quarter of 2020. First partial lockdown measures in Italy started in March 2020 and we assume bankers' beliefs remained unvaried for several months thereafter.

Table 1.16. EFFECT OF NEWS ON FORECAST ERRORS - FULL SAMPLE

	$FE^{\theta,i}_{t+12\|t}$					
$News_t(all)$	0.315^{***}	0.320^{***}	0.315^{***}	0.316^{***}	0.321^{***}	0.534^{***}
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.005)
N	3,069,663	3,069,663	3,069,663	3,069,663	3,069,663	1,626,921
Bank FE	No	Yes	No	No	Yes	No
Sector FE	No	No	Yes	No	Yes	No
Province FE	No	No	No	Yes	Yes	No
Borrower FE	No	No	No	No	No	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table provides coefficient estimates of the regression $FE^{\theta,i}_{t+12|t} = \beta_0 + \beta_1 News^i_t + \mathbf{\Gamma}'\mathbf{X} + \epsilon^i_{t+12}$, where $\mathbf{X}$ is the controls matrix and includes sector, province and/or borrower fixed effects. The regression is run using the full sample period. Standard errors are clustered at the borrower level. Significance levels at 1%, 5%, 10% are given by (***), (**), (*) respectively.

alized default rates did not increase as expected because of public intervention[14]. Finally, interest rates seem to have changed homogeneously among banks and decreased on average moderately because of public intervention.

Table 1.18. EFFECTS ON INTEREST RATES - FULL SAMPLE

	$r^{i,b}_t$					
$News_t$	-0.030^{***}	-0.027^{***}	-0.029^{***}	-0.026^{***}	-0.024^{***}	-0.004
	(0.006)	(0.007)	(0.006)	(0.007)	(0.007)	(0.004)
D^b_t	0.005^{***}	0.007^{***}	0.004^{***}	0.005^{***}	0.006^{***}	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
$News_t \times D^b_t$	-0.020	-0.017	-0.020	-0.024	-0.021	-0.023^{*}
	(0.014)	(0.013)	(0.014)	(0.016)	(0.014)	(0.013)
N	204,693	204,693	204,693	204,693	204,693	108,487
Bank FE	No	Yes	No	No	Yes	No
Sector FE	No	No	Yes	No	Yes	No
Province FE	No	No	No	Yes	Yes	No
Borrower FE	No	No	No	No	No	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: In this table we report estimates of the regression $r^{i,b}_t = \beta_0 + \beta_1 D^b_t + \beta_2 News^{i,b}_t + \gamma(D^b_t \times News^{i,b}_t) + \mathbf{\Gamma}'\mathbf{X} + \epsilon^{i,b}_t$, where $\mathbf{X}$ is a control matrix which contains also fixed effects. The regression is run using the full sample period. Errors are clustered at the NACE 2-digit level. Significance levels at 1%, 5%, 10% are given by (***), (**), (*) respectively.

[14]Since the beginning of the pandemics, Italian government has put in place a moratorium on outstanding banking debts positions.

7 Conclusion

In this paper we contribute to the literature of lenders' beliefs and show that bankers overreact to news in a way that is well described adopting diagnostic expectations in the information updating process when banks adjust their borrowers' creditworthiness. We use banks' estimates of borrowers' probability of default to have the most granular measure of lenders' expectations. Cross-sectionally, we document that bankers over (under) estimate borrowers' default when receiving negative (positive) news. This is effect is stronger for negative news and riskier borrowers. We also find significant heterogeneity in lenders' diagnostic levels, which permits to quantify the effects of expectations distortions on prices and quantities. A standard deviation increase in news leads a diagnostic banker to decrease interest rates by 8 basis points more than a low diagnostic banker and induce banks to assign new contracts with higher probability. The empirical results are confirmed through a structural estimation of a banking competition model. A counterfactual exercise shows that doubling the level of distortion, subject on receiving positive (negative) news, engenders a fall (rise) in interest rate by 42 basis points and an increase (decrease) of 1.7% of probability of issuing a new contract. Overall, we provide micro level evidence that financial institutions may incorporate in their models of borrowers creditworthiness biases coming from news overweighting, that well respond to belief formation mechanisms of the diagnostic-type. Such distortions may affect also prices and quantities of loans.

Bibliography

Albareto, Giorgio, Michele Benvenuti, Sauro Mocetti, Marcello Pagnini, and Paola Rossi, "The organization of lending and the use of credit scoring techniques in Italian banks," *Journal of financial transformation*, 2011, pp. 143–57.

Baron, Matthew and Wei Xiong, "Credit expansion and neglected crash risk," *The Quarterly Journal of Economics*, 2017, *132* (2), 713–764.

Basel Committee, "The Internal Ratings-Based Approach," *BIS Publications*, 2001, p. 108.

Behn, Markus, Rainer FH Haselmann, and Vikrant Vig, "The limits of model-based regulation," *Journal of Finance, Forthcoming, LawFin Working Paper*, 2021, (20).

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile prices in market equilibrium," *Econometrica*, 1995, *63* (4), 841–90.

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer, "Stereotypes," *The Quarterly Journal of Economics*, 2016, *131* (4), 1753–1794.

, Nicola Gennaioli, and Andrei Shleifer, "Diagnostic expectations and credit cycles," *The Journal of Finance*, 2018, *73* (1), 199–227.

, , Rafael La Porta, and Andrei Shleifer, "Diagnostic expectations and stock returns," *The Journal of Finance*, 2019, *74* (6), 2839–2874.

, , Yueran Ma, and Andrei Shleifer, "Overreaction in macroeconomic expectations," *American Economic Review*, 2020, *110* (9), 2748–82.

Chiappori, Pierre-André and Bernard Salanié, "Asymmetric Information in Insurance Markets: Empirical Assessments," *Handbook of Insurance: Second Edition*, 2013, pp. 397–422.

Coibion, Olivier and Yuriy Gorodnichenko, "Information rigidity and the expectations formation process: A simple framework and new facts," *American Economic Review*, 2015, *105* (8), 2644–78.

Coimbra, Nuno and Hélene Rey, "Financial cycles with heterogeneous intermediaries," Technical Report, National Bureau of Economic Research 2017.

Crawford, Gregory, Nicola Pavanini, and Fabiano Schivardi, "Asymmetric Information and Imperfect Competition in Lending Markets," *American Economic Review*, 2018, *108* (7), 1659–1701.

Durbin, James and Siem Jan Koopman, *Time series analysis by state space methods*, Oxford university press, 2012.

Fahlenbrach, Rüdiger, Robert Prilmeier, and René M Stulz, "Why does fast loan growth predict poor performance for banks?," *The Review of Financial Studies*, 2018, *31* (3), 1014–1063.

Gennaioli, Nicola and Andrei Shleifer, "What comes to mind," *The Quarterly journal of economics*, 2010, *125* (4), 1399–1433.

, , and Robert Vishny, "Neglected risks, financial innovation, and financial fragility," *Journal of Financial Economics*, 2012, *104* (3), 452–468.

, Yueran Ma, and Andrei Shleifer, "Expectations and investment," *NBER Macroeconomics Annual*, 2016, *30* (1), 379–431.

Greenwood, Robin and Andrei Shleifer, "Expectations of returns and expected returns," *The Review of Financial Studies*, 2014, *27* (3), 714–746.

, Samuel G Hanson, and Lawrence J Jin, "Reflexivity in credit markets," Technical Report, National Bureau of Economic Research 2019.

Kahneman, Daniel and Amos Tversky, "Subjective probability: A judgment of representativeness," *Cognitive Psychology*, 1972, *3* (3), 430–454.

Kindleberger, Charles, "Manias, Panics, and Crashes," 1978.

Krishnamurthy, Arvind and Wenhao Li, "Dissecting mechanisms of financial crises: Intermediation and sentiment," Technical Report, National Bureau of Economic Research 2020.

López-Salido, David, Jeremy C Stein, and Egon Zakrajšek, "Credit-market sentiment and the business cycle," *The Quarterly Journal of Economics*, 2017, *132* (3), 1373–1426.

Ma, Yueran, Teodora Paligorova, and José-Luis Peydro, "Expectations and bank lending," *Work. Pap., Chicago Booth Sch. Bus., Chicago Google Scholar Article Location*, 2021.

Minsky, Hyman P, "The Financial Instability Hypothesis: An Interpretation of Keynes and an Alternative to "Standard" Theory," *Nebraska Journal of Economics and Business*, 1977, pp. 5–16.

, "Stabilizing an Unstable Economy: The Lessons for Industry, Finance and Government," *Hyman Minski Archives*, 1986.

Richter, Björn and Kaspar Zimmermann, "The profit-credit cycle," *Available at SSRN 3292166*, 2019.

Train, Kenneth E., "Discrete choice methods with simulation," *New York: Cambridge University Press*, 2009.

Appendix

Proofs

Model - main

1. Proof Normalizing PD (eq 8,9).

   By definition $x_{t+1} \sim N(\hat{x}_{t+1}, \Omega)$. It follows that the standardized variable for x_{t+1} is $x^s = \frac{x_{t+1}-\hat{x}_{t+1}}{\Omega^{1/2}}$. The conditional expectation of firm's default status, i.e. the probability of default, is derived as

$$\begin{aligned}
\mathbb{E}(z_{t+1}|y^t) &= \mathbb{P}(x_{t+1} < a) \\
&= \mathbb{P}(\Omega^{1/2}x^s + \hat{x}_{t+1} < a) \\
&= \mathbb{P}\left(x^s < \frac{a-\hat{x}_{t+1}}{\Omega^{1/2}}\right) \\
&= \Phi\left(\frac{a-\hat{x}_{t+1}}{\Omega^{1/2}}\right)
\end{aligned}$$

2. Taylor approximation, complete.

   From the definition of z_{t+1} and $\mathbb{E}_t(z_{t+1})$, we can decompose their sum as follows (recall that from the starting equations describing the noisy process $u_{t+1} = z_{t+1} - x_{t+1}$, which here is interpreted as the difference between z_{t+1} and $\mathbb{E}_t(z_{t+1})$.)

$$\begin{aligned}
z_{t+1} - \mathbb{E}_t^\theta(z_{t+1}) &= \underbrace{z_{t+1} - \mathbb{E}_t(z_{t+1})}_{=u_{t+1}} + \mathbb{E}_t(z_{t+1}) - \mathbb{E}_t^\theta(z_{t+1}) \\
FE^\theta_{t+1|t} &= u_{t+1} + \Phi\left(\frac{a-\hat{x}_{t+1}}{\Omega_t^{1/2}}\right) - \Phi\left(\frac{a-\hat{x}^\theta_{t+1}}{\Omega_t^{1/2}}\right) \qquad (1.16)
\end{aligned}$$

   Equation (1.16) says that the forecast error of the diagnostic bankers increases the more (1) the signal is noisy and (2) the greater is the difference between the standard and diagnostic probability of default.

   Applying a Taylor approximation to function $\Phi(\cdot)$ around $\mathbf{x_0}$, for constant A, multiplicative vector $\mathbf{B}$ and each component j of $\mathbf{x_0}$. Suppose w.l.o.g. that $\mathbf{x_0} =$

$\mathbb{E}(\hat{x}_{t+1|t} \quad I_t)' = (0 \quad 0)'$. We obtain a linear expression that reads as

$$g(\hat{x}_{t+1}, I_t) = \Phi(A + \mathbf{B}'\mathbf{x}) \approx \Phi(A + \mathbf{B}'\mathbf{x_0}) + \sum_j B_j \phi(A + \mathbf{B}'\mathbf{x_0}) \times (x - x_{0j})$$

which, applied to $\Phi\left(\frac{a-\hat{x}_{t+1}}{\Omega_t^{1/2}}\right)$ and $\Phi\left(\frac{a-\hat{x}^{\theta}_{t+1}}{\Omega_t^{1/2}}\right)$ gives:

$$\begin{aligned}
\Phi\left(\frac{a - \hat{x}_{t+1}}{\Omega_t^{1/2}}\right) &\approx \Phi\left(\frac{a}{\Omega^{1/2}} - \frac{1}{\Omega^{1/2}}\hat{x}_{0,t+1}\right) \\
&\quad + \frac{1}{\Omega^{1/2}}\phi\left(\frac{a}{\Omega^{1/2}} - \frac{1}{\Omega^{1/2}}\hat{x}_{0,t+1}\right)(\hat{x}_{t+1} - \hat{x}_{0,t+1}) \\
&= \Phi\left(\frac{a}{\Omega^{1/2}}\right) - \frac{1}{\Omega^{1/2}}\phi\left(\frac{a}{\Omega^{1/2}}\right)\hat{x}_{t+1}
\end{aligned}$$

$$\begin{aligned}
\Phi\left(\frac{a - \hat{x}^{\theta}_{t+1}}{\Omega_t^{1/2}}\right) &= \Phi\left(\frac{a - \hat{x}_{t+1} - \theta K_t I_t}{\Omega_t^{1/2}}\right) \\
&\approx \Phi\left(\frac{a}{\Omega^{1/2}} - \frac{1}{\Omega^{1/2}}\hat{x}_{0,t+1} - \frac{1}{\Omega^{1/2}}K_t\theta I_{0,t}\right) \\
&- \frac{1}{\Omega^{1/2}}\phi\left(\frac{a}{\Omega^{1/2}} - \frac{1}{\Omega^{1/2}}\hat{x}_{0,t+1}\right)(\hat{x}_{t+1} - \hat{x}_{0,t+1}) \\
&- \frac{1}{\Omega^{1/2}}K_t\theta\phi\left(\frac{a}{\Omega^{1/2}} - \frac{1}{\Omega^{1/2}}K_t\theta I_{0,t}\right)(I_t - I_{0,t}) \\
&= \Phi\left(\frac{a}{\Omega^{1/2}}\right) - \frac{1}{\Omega^{1/2}}\phi\left(\frac{a}{\Omega^{1/2}}\right)\hat{x}_{t+1} - \frac{1}{\Omega^{1/2}}K_t\theta\phi\left(\frac{a}{\Omega^{1/2}}\right)I_t
\end{aligned}$$

From the last two expressions, (1.16) becomes

$$\begin{aligned}
FE^{\theta}_{t+1|t} &\approx w_{t+1} + \Phi\left(\frac{a}{\Omega^{1/2}}\right) - \frac{1}{\Omega^{1/2}}\phi\left(\frac{a}{\Omega^{1/2}}\right)\hat{x}_{t+1} \\
&\quad - \Phi\left(\frac{a}{\Omega^{1/2}}\right) + \frac{1}{\Omega^{1/2}}\phi\left(\frac{a}{\Omega^{1/2}}\right)\hat{x}_{t+1} + \frac{1}{\Omega^{1/2}}K_t\theta\phi\left(\frac{a}{\Omega^{1/2}}\right)I_t \\
&\approx w_{t+1} + \theta\underbrace{\frac{1}{\Omega^{1/2}}}_{>0}\underbrace{K_t}_{>0}\underbrace{\phi\left(\frac{a}{\Omega^{1/2}}\right)}_{>0}I_t
\end{aligned}$$

In the last expression, the only term that can make the overall coefficient equal to zero is *theta*. Therefore, we safely derive our last form of the equation and link it to the an empirical expression as described in the main model section.

$$FE^{\theta}_{t+1|t} = K_t \theta \frac{1}{\Omega^{1/2}} \phi\left(\frac{a}{\Omega^{1/2}}\right) I_t + w_{t+1}$$

Model - Real effects

Non linear relation for interest rate looks like

$$r_t = \frac{\Phi\left(\frac{a - \hat{x}_{t+1}}{\Omega_t^{1/2}}\right)}{1 - \Phi\left(\frac{a - \hat{x}_{t+1}}{\Omega_t^{1/2}}\right)}$$

From the previous proofs we know that, linearizing the cumulative distribution function around a fixed point through a Taylor approximation, we obtain

$$\Phi(A + \mathbf{B}'\mathbf{x}) \approx \Phi(A + \mathbf{B}'\mathbf{x_0}) + \sum_j B_j \phi(A + \mathbf{B}'\mathbf{x_0}) \times (x - x_{0j})$$

If the pdf $\phi(\cdot)$ is symmetric around its mean, we obtain

$$r_t \approx \frac{\Phi\left(\frac{a}{\Omega_t^{1/2}}\right)}{1 - \Phi\left(\frac{a}{\Omega_t^{1/2}}\right)} - \frac{1}{\Omega^{1/2}} \frac{\phi\left(\frac{a}{\Omega^{1/2}}\right)}{\Phi\left(\frac{a}{\Omega^{1/2}}\right)^2} \hat{x}_{t+1|t}$$

$$r_t^{\theta} \approx r_t - \frac{\theta K_t}{\Omega^{1/2}} \frac{\phi\left(\frac{a}{\Omega^{1/2}}\right)}{\Phi\left(\frac{a}{\Omega^{1/2}}\right)^2} I_t$$

The last one can be adapted as a linear regression where the only possible term equal to zero is the parameter θ

$$r_t^{\theta} = \beta_0 + \theta \cdot \beta_1 \widehat{PD}_{t+1|t} + \beta_2 I_t + \epsilon_t$$

Innovation as PD Variation

In our empirical exercise, we define as the main measure for innovation

$$I_t = -(\widehat{PD}^{\theta}_{t+11|t-1} - \widehat{PD}^{\theta}_{t+8|t-4}) = -\Delta\widehat{PD}^{\theta}_{t+3}$$

Consider two standard OLS univariate regressions, with a common dependent variable y_i and two different regressors x_i, z_i respectively.

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x_i + \varepsilon_i \\ y_i &= \gamma_0 + \gamma_1 z_i + v_i \end{aligned}$$

where $x_i \perp \varepsilon_i, x_i \perp v_i$. Now get the coefficient of the second regression in terms of covariance and variance of the variables involved and make some substitutions

$$\begin{aligned} \gamma_1 &= \frac{Cov(y_i, z_i)}{Var(z_i)} \\ &= \frac{Cov(\beta_1 x_i + \varepsilon_i, z_i)}{Var(z_i)} \\ &= \beta_1 \frac{\sigma_{xz}}{\sigma_z^2} \\ \Rightarrow \beta_1 &= \frac{\sigma_z^2}{\sigma_{xz}} \gamma_1 \end{aligned}$$

If $\sigma_{xz} = Cov(z_i, x_i) > 0$, then between coefficients β_1 and γ_1 we have a positive relationship.

We do the same with the regressions obtained from the theoretical and empirical models, respectively:

$$\begin{aligned} FE^{\theta,i}_{t+1|t} &= \beta_0 + \beta_1 I^i_t + \varepsilon_i \\ FE^{\theta,i}_{t+1|t} &= \gamma_0 + \gamma_1 News^i_t + v_i \\ \Rightarrow \gamma_1 &= \beta_1 \frac{Cov(News^i_t, I^i_t)}{Var(News^i_t)} \end{aligned}$$

So, if $Cov(News^i_t, I^i_t) > 0$, we have a positive relationship between the main variable of theoretical and the empirical model. Recall the definition of the theoretical news in the empirical model, which can be written also as a combination of the first difference of rational PDs and innovations

$$News_t = -\Delta\widehat{PD}^{\theta}_{t+1|t} = -(B(\hat{x}_{t+1|t} - \hat{x}_{t|t-1}) + C(I_t - I_{t-1}))$$

For coefficients $A, B, C \in \mathbb{R}^+$ and K be the steady state value of the Kalman gain, we substitute the formulation of $News_t$ in the covariance between news and inovation, and get

$$\begin{aligned}
Cov(News_t, I_t) &= \mathbb{E}[Cov_{t-1}(News_t, I_t)] + Cov(\mathbb{E}_{t-1}[News_t], \underbrace{\mathbb{E}_{t-1}[I_t]}_{=0}) \\
&= \mathbb{E}[Cov_{t-1}(News_t, I_t)] \\
&= \mathbb{E}[BCov_{t-1}(-(\hat{x}_{t+1|t} - \hat{x}_{t|t-1}), I_t) - C \cdot Cov_{t-1}(I_t - I_{t-1}, I_t)] \\
&= \mathbb{E}[BCov_{t-1}(-((\rho - 1)\hat{x}_{t|t-1} + KI_t), I_t) - CVar_{t-1}(I_t)] \\
&= \mathbb{E}[-BKVar_{t-1}(I_t) - CVar_{t-1}(I_t)] \\
&= -Bk\mathbb{E}[Var_{t-1}(I_t)] - C\mathbb{E}[Var_{t-1}(I_t)] \\
Cov(News_t, I_t) &= -(BK + C)\mathbb{E}[Var_{t-1}(I_t)]
\end{aligned}$$

Recalling from equation (1.4)

$$\begin{aligned}
\widehat{PD}^{\theta}_{t+1|t} &= \Phi\left(\frac{a - \hat{x}^{\theta}_{t+1}}{\Omega_t^{1/2}}\right) \\
&\approx \underbrace{\Phi\left(\frac{a}{\Omega^{1/2}}\right)}_{=:A} \underbrace{- \frac{1}{\Omega^{1/2}}\phi\left(\frac{a}{\Omega^{1/2}}\right)}_{=:B} \hat{x}_{t+1|t} \underbrace{- K\theta\frac{1}{\Omega^{1/2}}\phi\left(\frac{a}{\Omega^{1/2}}\right)}_{=:C} I_t
\end{aligned}$$

It follows that the covariance between news and innovation is positive.

$$Cov(News_t, I_t) = \underbrace{-(BK + C)}_{>0} \underbrace{\mathbb{E}[Var_{t-1}(I_t)]}_{>0} > 0$$

This result proves that the measure $News_t = -\Delta\widehat{PD}^{\theta}_{t+1|t}$ used in the empirical exercise is a valid alternative to the innovation of the theoretical model, given that their covariance is strictly positive.

Tables

Table 1.20. NACE SECTORS

1-Digit Code	Description
A	Agricolture, forestry, fishing
B	Mining and quarrying
C	Manufacturing
D	Electricity and gas
E	Water supply and waste management
F	Construction
G	Wholesale retail
H	Transportation and storage
I	Accommodation and food service activities
J	Information and communication
K	Financial and insurance activities
L	Real estate activities
M	Professional, scientific and technical activities
N	Administrative and support activities
O	Public administration and defense
P	Education
Q	Human health and social works
R	Arts, entertainment and recreation
S	Other service activities
T	Activities of households and employers
U	Activities of extraterritorial organizations

Notes: This table shows the list of NACE differentiation of economic activity. More information can be obtained at the official page of European Commission.

Chapter 2

When the Markets Get CO.V.I.D: COntagion, Viruses, and Information Diffusion.

Abstract

We quantify the exposure of major financial markets to news shocks about global contagion risk accounting for local epidemic conditions. For a wide cross section of countries, we construct a novel dataset comprising (i) announcements related to COVID19 and (ii) high-frequency data on epidemic news diffused through Twitter. Across several financial assets, we provide novel empirical evidence about financial dynamics both around epidemic announcements and at daily/intra-daily frequency. Contagion data and social media activity about COVID19 suggest that the market price of contagion risk is significant. Hence policies that mitigate global contagion or local diffusion may be extremely valuable. [†]

1 Introduction

COVID19 has manifested itself as a very aggressive and fast epidemic that—at the time of the first draft of this paper—brought major economic countries to their knees.[1] Given the fast-increasing contagion curve of COVID19 and its global scale, this epi-

[†]Co-authored with M. J. Arteaga-Garavito, M. M. Croce, and I. Wolfskeil

[1]Our first draft is dated 3/23/2020. To assess the severity of COVID19, see the March 11, 2020 WHO Director-General's opening remarks (https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020).

demic event is challenging common economic policy interventions and depressing the global value of our assets, i.e., the wealth of millions of households all over the world.

Given that severe virus-related crises are expected to become more frequent, we find it relevant to use COVID19-related data to ask the following broad questions about financial market reactions to viral contagion risk. First, what is the average impact of medical announcements on financial returns? Equivalently, is the diffusion of this information enhancing wealth or adding risk? Second, what is the market price of risk of news related to global contagion dynamics? Third, can local contagion conditions help us predict expected returns?

Last but not least, can we use social media activity to measure the production and diffusion of information about epidemic risk? This question is important for at least two reasons. First, fast epidemic outbreaks tend to get investors off guard; hence, real-time indexes based on social media news may function as a useful predictive tool. Second, estimating multidimensional models requires many observations that we may gather by using high-frequency data, as opposed to waiting for daily medical bulletins.

In this study, we address these questions by quantifying the exposure of major financial markets to news shocks about global contagion risk accounting for local epidemic conditions. For a broad cross section of countries, we construct a novel data set comprising (i) medical announcements related to COVID19; and (ii) high-frequency data on epidemic news diffused through Twitter. Across several classes of financial assets and currencies, we provide novel empirical evidence about financial dynamics (i) around epidemic announcements, (ii) at a daily frequency, and (iii) at an intra-daily frequency. Formal estimations based on contagion data and social media activity about COVID19 confirm that the market price of epidemic risk is very significant. We conclude that prudential policies that mitigate either global contagion or local diffusion may be extremely valuable. More broadly, we offer a methodology for constructing a rich framework of information diffusion and information attention that future empiricists can adapt in order to examine future sources of global crises.

Interpretation of our findings. Before describing in more detail our findings, we must clarify how to interpret them. There is no doubt that the COVID pandemic was unprecedented in many dimensions (Baker et al. 2020b). Nevertheless, we believe that our high-frequency social-media–based approach can be informative also when adopted to future pandemic events. Tracking pandemic-related sentiment and contagion dynamics can be very important to portfolio managers in future pandemic events. In this sense, our analysis is not simply a case study of a unique rare event. Rather, it is very flexible as it shows how to gather a rich-and-reliable dataset for conducting formal statistical tests even after a few weeks from the beginning of a 'brand new' pandemic. Given our high-frequency approach, we can track the full evolution of the pandemic across multiple-and-different contagion waves. Furthermore, our methodology is broad as it allows financial economists to study many relevant dimensions of financial markets.

In addition, after having collected more than 16,000 medical announcements across many countries, our results on the positive average appreciation of equity markets can be reasonably interpreted as a statement on expected appreciation. Our novel and sizeable dataset should minimize concerns about 'peso problems'.

Even though we aim to provide a flexible set of tools for future pandemic-related studies, we acknowledge that future crises may be different from the COVID one. In this case, future research should adapt our seminal approach as already done, for example, for the analysis of rare but severe global financial crises.

Current results in detail. An important contribution of our work is the collection of a novel dataset on the COVID19 pandemic that includes (i) an extensive set of official announcements on medical conditions (more than 16,000 announcements) and (ii) news diffused on Twitter in real-time by major newspapers (based on more than 800,000 tweets). We identify major newspapers for a large cross section of countries in the spirit of Baker et al. (2016). In contrast to Baker et al. (2016), we do not analyze articles; rather we track news published on Twitter in real time to produce high frequency data when needed.

More specifically, we track tweets posted by major newspapers with key words such as 'coronavirus' and 'covid19'. For each newspaper, we use the location of its headquarters to identify its specific time zone. As a result, we gather thousands of tweets for a large cross section of countries that we can aggregate at different frequencies and across regions.

Given this data set, we document several important facts about news diffusion. First, both Twitter-based news diffusion (measured by the number of tweets) and attention (measured by the number of retweets) spike upon contagion-related announcements. Second and more broadly, the diffusion of information increases substantially in each country in our data set as soon as that country goes into an epidemic state.[2] Third, our measured increase in information diffusion is particularly pronounced precisely during the hours in which financial markets are open. All of these empirical facts suggest that tracking Twitter-diffused news can be a reliable way to characterize the information set of investors at a high frequency.

Turning our attention to financial dynamics, we look at equity returns around announcements, that is, in a ± 60 minute window. We find that cumulative equity returns have no clear pattern before the announcement, as they tend to be relatively flat and indistinguishable from zero. In the post-announcement time window, instead, cumulated returns jump upward. This result also holds when we focus only on bad news announcements and is also present in countries with relatively high contagion levels. Furthermore, the positive average effect of medical announcements on equities is present upon both local and foreign announcements.

We note that this pattern of returns is not present in the pre-epidemic state and is quite different from that documented in Lucca and Moench (2015). Lucca and Moench (2015) shows a slow and persistent accumulation of positive returns before monetary policy announcements. In our case, instead, the increase in the cumulative returns at the announcement is consistent with the Ai and Bansal (2018) model in which there is no leakage of information and no pre-announcement drift. When the representative investor cares about the timing of resolution of uncertainty, prices jump upward when uncertainty is resolved along the information cycle and then start to decline.

[2] We identify the beginning of the epidemic state with the day in which the number of confirmed COVID19 cases becomes greater than or equal to 100.

Furthermore, we conduct the same analysis by looking at the government bond market. The response of bonds is less severe than that observed in equities. In a ± 60-minute window around the announcement, there is no significant adjustment in bond returns among advanced economies (henceforth AEs). At first, this result may look surprising as bonds may be in higher demand since they are considered safer assets. Hence one may expect to find an average appreciation. On the other hand, one may expect that default concerns generate a simultaneous downward pressure on bond prices. Since we find a modest link between COVID19 news and default concerns as measured by CDS quotes, we speculate that this result is mainly driven by monetary policy (broadly consistent with the findings of Haddad et al. (2021)).

Among emerging economies (henceforth EEs), in contrast, bond prices experience a sudden positive increase around announcements, but it is less relevant than that for equities. By no-arbitrage, this observation suggests that cash-flow uncertainty is an important determinant of the market fluctuations observed during the COVID19 crisis. This high-frequency result is consistent with the results documented by Gormsen and Koijen (2020) looking at dividend futures.

We also look at equity market trading volume around announcement times and document that it exhibits an upward adjustment upon the announcement time and then a slow reversal. We show that this pattern is less severe for AEs than for EEs. When we look at bid-ask spreads of sovereign bonds, we find an immediate reduction upon announcements for AEs and a delayed one for EEs. The magnitude of the decline in the bid-ask spread is comparable across AEs and EEs. Together, these patterns suggest that investors are active with safer assets in AEs and EEs.

According to an LDA model applied to our tweets (in the spirit of Bybee, Kelly and Manela (2020a)), cases are one of the main drivers of the topics that received attention during the pandemic. Accordingly, in the last step of our analysis, we group countries into three portfolios on a daily basis according to their relative number of COVID19 cases. We do this separately for AEs and EEs. The H (L) portfolio comprises the equity returns of the top (bottom) countries in terms of COVID19 contagion cases. We then estimate a no-arbitrage based model in which we allow for time-varying betas

($\beta_{i,t}$) with respect to global contagion risk. Specifically, we allow equity returns to respond to global viral contagion news according to each portfolio's relative share of official COVID19 cases. Global contagion risk is measured either by innovations in the growth rate of global COVID19 contagion cases or by innovations in the tone of our COVID19-related tweets.

This model can potentially capture many of the features of equity returns that we document in our descriptive analysis. First, this model captures predictability through contagion-based time-varying betas. Second, this specification has the potential to capture higher negative skewness for countries that go through more severe contagion paths. Consider the case of portfolio H comprising countries receiving a sequence of relatively more severe contagion news. This portfolio will have greater exposure to adverse news ($|\beta_{H,t}|$ increases) as the relative contagion share of the portfolio grows. As the relative contagion share starts to flatten out and eventually decline, the sensitivity of this portfolio to good news is reduced ($|\beta_{H,t}|$ shrinks), meaning that returns will be less sensitive to positive news and hence the right tail of their distribution will not be very long.

Third, this model accounts for heterogeneous exposure to global contagion news, enabling us to identify the market price of risk of this global contagion component. Across all of our specifications, the market price of contagion risk is both statistically significant and extremely high. Equities are more exposed to risk than bonds. Both within advanced and emerging economies, heterogeneous exposure to contagion risk is substantial, and as a result, an equity-based HML-COVID strategy bears a high risk premium. An HML-COVID strategy that goes long in bonds of countries with a larger share of cases and short in those with a smaller share of cases, instead, provides an insurance premium. This suggests that bonds tend to become safer in countries exposed to heightened contagion risk. We find that this result is particularly sizable among EEs.

These results conform well with the data on weekly international investment flows. Countries with lower (higher) contagion levels are expected to experience equity inflows (outflows). Expected inflows are stronger in AEs than in EEs. In contrast, when looking at bonds, these findings are almost absent in AEs and reversed in EEs, meaning that in high-covid emerging economies, the flows going toward government bonds increase. This is consistent with the idea that bonds are perceived as safer assets in EEs.

In the last step of our analysis, we run intra-day regressions taking advantage of our high-frequency Twitter-based risk measure. We focus on European countries whose markets are open simultaneously, namely, ITA, ESP, UK, FRA, DEU, CHE, and SWE. Every day, we group them into three portfolios according to their relative number of COVID19 cases measured in the previous 24 hours. The H (L) portfolio comprises the equity returns of the top-2 (bottom-2) countries for COVID19 contagion cases.

Our novel high-frequency estimation confirms our main findings: policies related to the prevention and containment of contagion could be very valuable not only in terms of lives saved but also in terms of global wealth. These results also hold after controlling for the market and changes in equity volatility. Our results have been very stable over time and can be explored at `https://sites.google.com/view/when-markets-get-covid` a website that we use for the visualization of our data.[3]

Related literature. Due to its relevance, the COVID19 crisis has spurred a lot of contemporaneous research (Goldstein et al. 2021). Macroeconomic studies are focusing on both the aggregate and distributional dynamic implications of the epidemic crisis (Hagedorn and Mitman 2020; Coibion et al. 2020; Eichenbaum et al. 2020; Fornaro and Wolf 2020; Chiou and Tucker 2020; Barrot et al. 2020; Alon et al. 2020; Glover et al. 2020; Corsetti et al. 2020; Caballero and Simsek 2020; Coven and Gupta 2020; Hensvik et al. 2020).

Other analyses assess policy concerns (Acemoglu et al. 2020; Alvarez et al. 2020; Jones et al. 2020; Bahaj and Reis 2020; Elgin et al. 2020; Faria-e Castro and Louis 2020; Krueger et al. 2020; Farboodi et al. 2020). Correia et al. (2020) and Barro et al. (2020) provide evidence using data from the 1918-Flu epidemic. We differ from these studies because of our strong attention to asset prices and COVID19-driven risk.

Other studies at the intersection of macroeconomics and econometrics focus on forecasting the diffusion of both contagion cases and COVID19-implied economic ac-

[3]Our updates are schedule in October, January, and May.

tivity disruptions (Favero 2020; Ichino et al. 2020; Atkeson 2020; Ma et al. 2020; Ludvigson et al. 2020). We focus on both the cross sectional and time series implications for asset prices across different asset classes.

An important strand of the literature focuses on the measurement of both COVID19-induced uncertainty and firm-level risk exposure by utilizing textual analysis and surveys (Baker et al. 2020a; Hassan et al. 2020; Bartik et al. 2020). Giglio et al. (2020) use survey data to study investor expectations over different horizons. Lewis et al. (2020) provide a novel weekly measure of economic activity using several labor market-based timeseries. We focus on high-frequency data, Twitter-based news diffusion, epidemic announcements, and country-level asset price dynamics. Our study adds viral contagion risk considerations to the findings of Pelger (2020).

Gerding et al. (2020) look at equity market dynamics and link the epidemic risk exposure to country-level fiscal capacity. Augustin et al. (2021) looks at CDS. Bonaccolto et al. (2019) focus on currency union break up risk due to COVID19. Papanikolaou and Schmidt (2020) look at the financial implications of industry-level job disruption due to COVID19. Albuquerque et al. (2020) focus on the performance of firms with high environmental and social ratings during the COVID19 outbreak. They do not study announcements and they do not assess the market price of viral contagion risk. Ramelli and Wagner (2020) study equity returns across firms accounting for international trade, financial strength, and investor attention. They use both Google search volume and conference calls as a measure of attention, whereas we use high-frequency data on retweets of tweets issued by news provider. Pástor and Vorsatz (2020), Baker et al. (2020c), Bretscher et al. (2020b), and Kaniel and Wang (2020) study the impact of COVID19 on financial markets. We provide novel evidence about both (i) market reactions around contagion-related announcement times, and (ii) the market price of contagion risk at high frequency.

Schoenfeld (2020) examines buy-and-hold returns for many assets and finds that managers systematically underestimate their exposure to COVID19. Cororaton and Rosen (2020) look at the characteristics of firms participating to the US Paycheck Protection Program. Acharya and Steffen (2020) study firm-loan-level data to study the implications for liquidity. Carletti et al. (2020) look at Italian firms. Alfaro et al. (2020) focus on the link between aggregate equity market returns and unanticipated changes in predicted infections during the SARS and COVID19 pandemics. Bretscher et al. (2020a) look at the supply channel of uncertainty shocks. Hartley and Rebucci (2020) and Sinagl (2020) look at monetary policy announcements and cash-flow risk, respectively. Cox et al. (2020) confirm the relevance of monetary policy estimating a dynamic asset pricing model. We differ in our attention to medical announcements; our social media-based measures of information diffusion and attention; and our high frequency analysis. Our work complements the evidence in Gormsen and Koijen (2020) and Gormsen et al. (2021) who extract relevant information about expectations and risk premia from derivatives.

Within the literature that studies news coverage reaction to news, our manuscript is methodologically related to the work of, among others, Bianchi et al. (2021), Hassan et al. (2019), Manela and Moreira (2017), Garmaise et al. (2021), Tetlock (2007), Calomiris and Mamaysky (2019), Israelsen et al. (2021), Cookson et al. (2021), Bybee et al. (2020b) and Engle et al. (2020).

2 Medical Announcements

In this section, we illustrate key features of our novel data set comprising thousands of COVID19-related announcements across twenty-one countries. We then show our main results. Specifically, we document that: (i) equity markets on average appreciate upon announcements, and especially so in emerging economies (EEs); (ii) bond returns are insensitive to announcements in advanced economies (AEs), but appreciate to some extent in EEs; (iii) across both AEs and EEs, trade becomes more active after medical announcements.

2.1 Data Collection

We treat the release of each medical bulletin as an announcement. The same applies to travel limitations and lockdown policies related to COVID19. We note that we have

Table 2.1. SUMMARY STATISTICS FOR ANNOUNCEMENTS

Country	No. Announcements	Case Reports	Live Streamed	President/ Prime Minister
AR	605	33%	64%	3%
AU	678	78%	4%	1%
BR	975	64%	26%	2%
CA	791	58%	21%	18%
CH	627	78%	9%	13%
CL	896	59%	29%	3%
CN	721	82%	3%	1%
CN-HK	1,376	55%	2%	1%
CO	1,006	58%	34%	8%
DE	283	87%	1%	7%
ES	570	83%	1%	17%
FR	567	77%	16%	6%
IN	759	89%	1%	1%
IT	654	74%	17%	8%
JA	332	59%	5%	5%
KR	642	80%	1%	4%
MX	1,803	10%	45%	21%
NZ	457	61%	29%	7%
UK	711	82%	11%	7%
US	1,386	17%	54%	7%
Total	15,839	64%	18%	7%

Notes: This table shows summary statistics for COVID19-related announcements that we collect for a large cross section of countries. Our real-time data range from 1/1/2020 to the date of this manuscript. For each country, we report the total number of announcements, the fraction of announcements that report the number of positive COVID cases, that are live streamed, and that are announced by the country's President or Prime Minister.

manually tracked these policy interventions on a daily basis and hence we have constructed a novel dataset important to study real-time high frequency reactions of financial markets to epidemic risk.

Since in our sample we have also witnessed important announcements related to both monetary and fiscal policy interventions, we complement the medical announcements with major policy-related announcements as well. We note that medical announcements in our sample period are much more prominent than policy-related announcements as they represent nearly 86% of all of the announcements collected. Our data collection is very comprehensive, as documented in table 2.1, and it comprises more than 10,000 medical announcements. An example of a COVID19-related announcement follows:

> **2020-03-14 15:35:00; Vice President @Mike_Pence and members of the Coronavirus Task Force will hold a press briefing at 12:00 p.m. ET. Watch LIVE:** http://45.wh.gov/RtVRmD

In this case, we set the time of the announcement at 12:00 p.m. ET. To clarify further our methodology, we also give an example of an announcement related to a monetary policy intervention in response to COVID19:

> **2020-03-18 23:05:00; FT Breaking News; ECB to launch €750bn bond-buying programme.**

In this case, the time of the announcement is 11:05 p.m. CET.

We 'hand-collect' these announcements in several ways. First of all, for each country we look for official press statements publicly available on the webpage of the local Ministry of Health (MoH). If the press statement does not have an official time stamp, we look for it on the official Twitter account of the MoH or other related government entities (for example, the Twitter account of the Prime Minister). If this second attempt fails as well, we look at the Twitter accounts of major local newspapers and focus on news about medical reports. These steps, which we repeat multiple times during each week, are sufficient to identify the effective time of each announcements in our data set relevant for financial investors.

As an example, in figure 2.1 we report our record of the first scheduled Coronavirus Task Force Press briefing. In contrast to the following White House press meetings, this briefing took place earlier, at 3:40 p.m. EST. This example demonstrates two important aspects of our dataset construction: (i) it accounts for meetings scheduled at not-recurrent times; and (ii) it captures purely COVID-related news.[4]

2.2 Announcements and Financial Markets

Pre- and post-epidemic samples. In what follows, we study the financial dynamics around medical announcement times. In order to isolate the dynamics related solely to medical announcements, we plot the differential behavior of our variable of interest with respect to normal times, i.e., pre-epidemic times. In each country, we define the

[4]Our dataset enables researchers to easily identify each specifc announcement and hence look for the content discussed in each one of the events that we detect.

Announcements: January 31, 15:41 EST (21:41 CET)

- 9,700 cases in China, and 200 deaths
- 132 cases in 23 countries outside of China
- 6 cases in the United States
- Report from Germany affirms that asymptomatic carriers can transmit the virus
- Following the WHO the USA declared coronavirus a public health emergency
- Mandatory 14 days quarantine any U.S. citizen who has been in Hubei Province in the previous 14 days
- Temporary suspension of entry into the USA of foreign nationals who pose a risk of transmitting the 2019 novel coronavirus

Figure 2.1. ANNOUNCEMENT TIME FROM TWITTER.

Notes: This figure shows a tweet about one of the first COVID-related announcements in the US. The tweet time stamp enables us to identify the effective timing of the announcement. On the right hand side of this figure, we summarize the topics discussed during the briefing.

beginning of the epidemic period as the day in which the country experienced an official number of contagion cases greater than or equal to 100. Given this threshold, China is the first country in our sample to go in the epidemic phase, whereas New Zealand is last.

The pre-epidemic sample starts for all countries on October 1st 2019 so that the pre-epidemic period comprises at least four months of data. This subsample is long enough to run meaningful comparisons with the post-pandemic subsample. More specifically, consider, for example, an announcement on a Friday at 3:40 p.m. EST. We compare the reaction of our financial variables around this announcement to their behavior at the same time across all of the Fridays comprised in our pre-epidemic sample.

Pre- and post-announcement behavior. We run a high-frequency analysis around announcement times. In what follows, we estimate the following regression at the minute-level:

$$Z_t = (c_{pre} + c_{t>t^*}) + (\alpha_{pre} + \alpha_{t>t^*}) \cdot t + (\beta_{pre} + \beta_{t>t^*}) \cdot t^2, \quad t \in [t^* \pm K] \tag{2.1}$$

where t^* is the time of the announcement, K is equal to 60 minutes; and Z_t is the differential behavior of our variable of interest across the pre- and post-epidemic sample. This specification is a quadratic function of time that includes dummy variables to account for post-announcement jumps in both the level and the slope. We test the null assumption that there is no difference post-announcement, $H_0 : c_{t>t^*} = \alpha_{t>t^*} = \beta_{t>t^*} = 0$, and if we fail to reject the null we depict the resulting smooth quadratic fit. Standard errors are always HAC-adjusted.

Information Diffusion. Our novel social media-based data set enables us to measure the diffusion of information at a very high frequency. For each announcement in our data set, we compile all COVID-related tweets issued in a ± 60-minute window around announcement time by major newspapers in each country. We provide a detailed description of our data collection procedure in the next section. For the sake of statistical power, we aggregate all of these tweets across all of our countries and we call the resulting aggregate 'World'.

In the left panel of figure 2.2, we show per-country per-minute average number of tweets around announcement times during epidemic periods in excess of the same average measured in the pre-epidemic samples (dots). This procedure enables us to capture news diffusion patterns specific to the epidemic period. The mid panel refers to retweets, that is, our measure of attention to the news. The right panel, instead, shows a measure of the tone of the content of the tweets.

Formal tests reject the null assumption of a common time-behavior before and after the announcement for information diffusion. In figure 2.2, the solid line denotes our estimate whereas the shaded area refers to our confidence intervals. Importantly, both information diffusion and attention to the news increase significantly in the hour after announcements. Similar conclusions apply when we focus only on the post-epidemic sample (see figure .1 in the Appendix). In addition, when looking at cumulative numbers of tweets only in the post-epidemic sample, it is evident that most of the twitting activity takes place in the time-window [-90 -30] ('preview' tweets about the announcement) and [0 +90] ('rehash' tweets).

Since we focus solely on announcements related to medical bulletins and policy

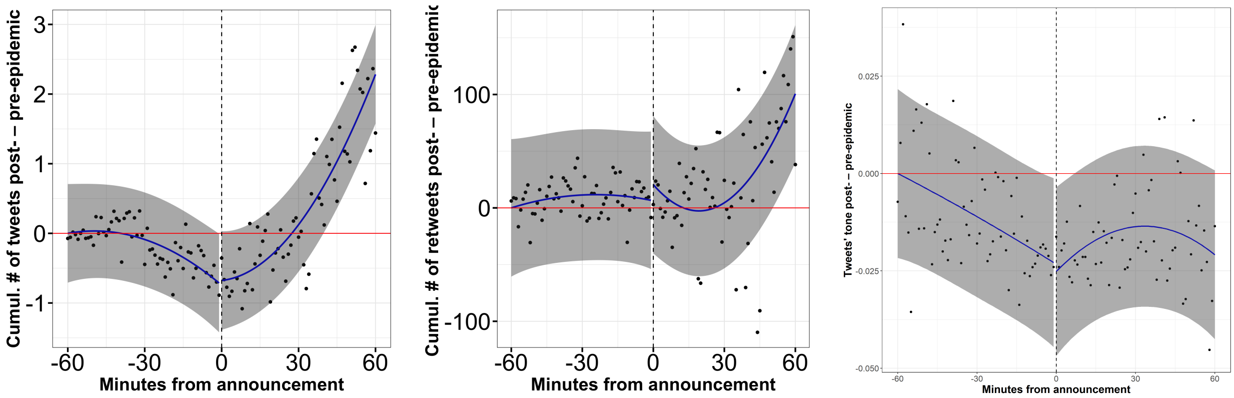


Figure 2.2. **INFORMATION DIFFUSION AND ATTENTION AROUND ANNOUNCEMENTS**

Notes: The left (mid) panel of this figure shows the average per-minute and per-country number of tweets (retweets) around announcement times in excess of the same average in the pre-epidemic period. The right panel refers to a measure of the tone of the tweets. In each country, the epidemic period starts when there are more than 100 cases of COVID19. Solid line and shaded areas are based on the estimation of equation (2.1). The sample starts on October 1st 2019 and ends on the date of this draft.

measures to fight the epidemic, our results refer to both sources and topics distinct from those studied in the previous papers about economic announcements. Our results confirm that medical announcements gather special attention and hence it is important to understand whether they have a significant impact on financial markets.

Financial data sources. All data are from Thomson Reuters and Bloomberg. Equity, bond and currency data are obtained at the minute frequency and then aggregated at lower frequencies when necessary. For each country, we collect data on its major equity index and 10-year maturity treasury bond index. We measure the risk-free rate by focusing on the yield of 3-month government bills. Due to data availability CDS data are collected at the daily frequency. All details about our data can be found in table .3 (see 4).

Equity markets. In figure 2.3, we show the average cumulative returns obtained from buying country-specific equities 60 minutes before a country-specific announcement and holding them for 120 minutes. Our results are averaged across both countries and announcements. Countries are divided in two groups, advanced and emerging economies, according to the IMF classification.[5] In the Appendix, we report our results in the post-epidemic sample (see figure .2, 4).

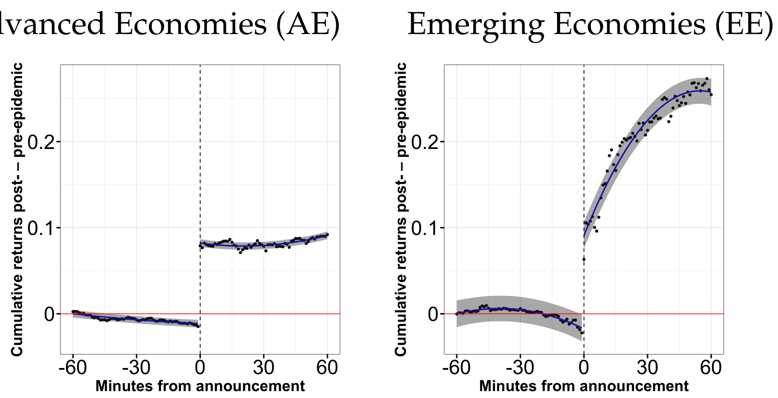
The top panels show what happens when we consider all countries and all announcements. Namely, in AEs (EEs) equity values tend to slightly decline (stay flat) before the announcement and then appreciate substantially upon the announcement. This appreciation is persistent, as it remains almost constant during the next hour in AEs and it gets amplified in EEs. This observation suggests that the release of covid-related news helps equities. Since we are considering both announcements conveying positive news and announcements conveying negative news, we think of this jump in equity valuation as a measure of the value of the pure release of information about epidemic risk.

More specifically, we note that this figure shows a time varying behavior of returns that is quite different from that documented in Lucca and Moench (2015). Lucca and Moench (2015) show a slow and persistent accumulation of positive returns before monetary policy announcements. This drift may be explained by information leakage. In our case, instead, the increase in the cumulative returns at the announcement is consistent with the basic Ai and Bansal (2018) model with no information leakage. When the representative investor cares about the timing of a resolution of uncertainty, prices jump upward when uncertainty is resolved along the information cycle, and then they eventually start to decline.

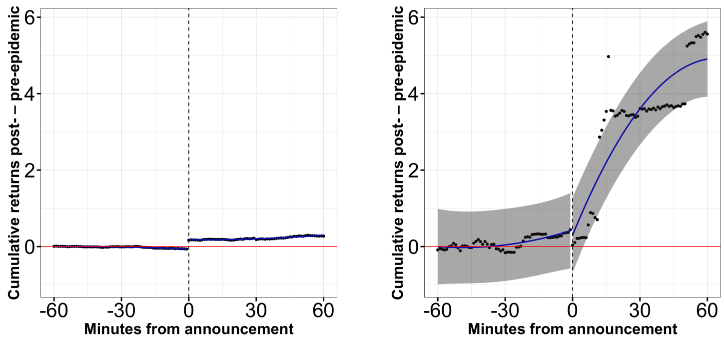
In figure 2.3(b), left panel, we show that the same phenomenon is present to a similar extent when we focus on the subset of announcements associated to bad news within the group of AEs.[6] We measure bad news as an unexpected increase in the growth rate of contagion cases on the day of the announcement. We explain in detail our construction of the news in the next section when we price them using the cross section of equity and bond returns. More broadly, this result is consistent with the pattern of the tone of our covid-related tweets (figure 2.2, right panel): even though

[5]If a country-specific announcement happens when the exchange of the country is closed, we consider the 60 minutes prior to the closing time of the previous day and the first 60 minutes after the opening of the exchange in the next day.

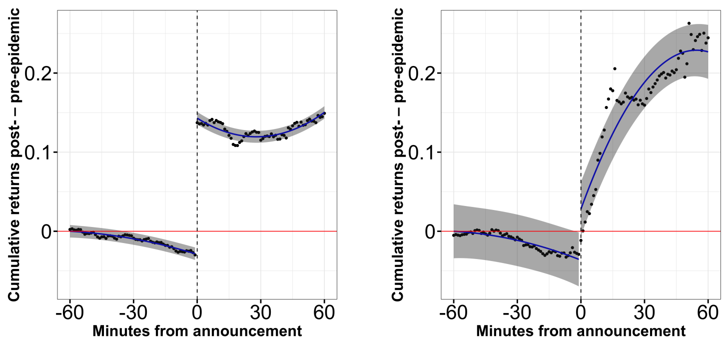
[6]Note that the scale for this panel is one order of magnitude greater than that in figure 2.3(a). Hence the announcement jump has the same magnitude as in panel a even though it looks smaller.



(a) All Countries



(b) Only Bad News



(c) High-COVID Countries

Figure 2.3. EQUITY RETURNS AROUND ANNOUNCEMENTS

Notes: In each panel, dots denote the difference across subsamples of the cross-country-cross-announcement average cumulative returns obtained from buying equities 60 minutes before an announcement and holding them for 120 minutes. Panel a (c) comprises announcements from all countries (top-50% countries in terms of contagion cases) in each group. Panel b excludes announcements conveying good news. Returns are in log units and multiplied by 100. Solid line and shaded areas are based on the estimation of equation (2.1). Our sample starts on October 1st 2019 and ends on the date of this draft.

announcements are associated with an average decline of the tone of the news, equities appreciate.

Turning our attention to EEs, we note that there still exists a positive jump in equity

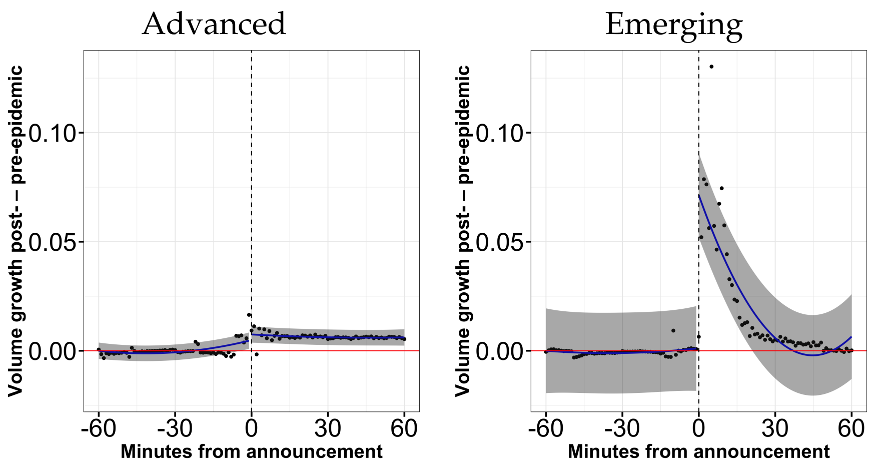


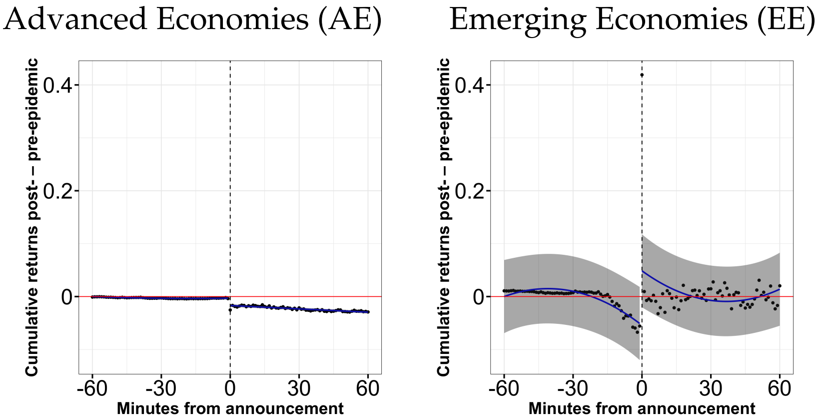
Figure 2.4. EQUITY VOLUME AROUND ANNOUNCEMENTS

Notes: The left (right) panel shows the average equity log-volume growth for all (above median of contagion cases) countries around announcement times. We depict the difference across pre- and post-epidemic samples. In each country, the epidemic period starts when there are more than 100 cases of COVID19. Solid line and shaded areas are based on the estimation of equation (2.1). Our sample starts on October 1st 2019 and ends on the date of this draft.

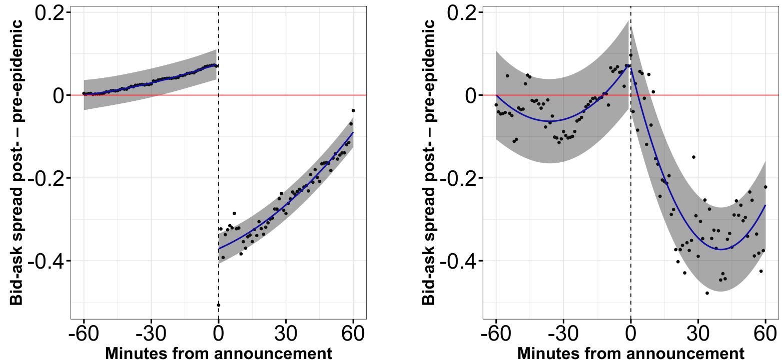
valuations, but it happens with about a 15-minute delay with respect to our announcement time stamps. Given our quadratic specification, this phenomenon is captured through a significant increase in the slope of our cumulative returns time series. We also point out that in this case the jump is one order of magnitude greater than under the case in which we consider all announcements, implying that in these countries the value of resolution of uncertainty may be extremely high even when we condition on bad news.

In figure 2.3(c), we consider all of our announcements but we limit our attention to countries that are above median in terms of total contagion cases. The scale in these panels is identical to that used in figure 2.3(a). Not surprisingly, the smaller sample that we use produces estimates surrounded by higher estimation uncertainty. Taking this into account, the value of the information disclosed during these announcements is higher among high-COVID AEs and remains almost unchanged among high-COVID EEs. More broadly, when we look at the entire cross section of our 21 countries, low-COVID countries appear to be less sensitive to contagion-risk news. This is consistent with the results of the no-arbitrage factor model that we estimate in the second part of our study.

The equity returns patterns that we document may also be consistent with models



(a) Bond Returns



(b) Bonds Bid-Ask Spread

Figure 2.5. **SOVEREIGN BONDS AROUND ANNOUNCEMENTS**

Notes: In the top panels, dots denote the difference across subsamples of the cross-country-cross-announcement average cumulative returns obtained from buying 10-year sovereign bonds 60 minutes before an announcement and holding them for 120 minutes. In the bottom panels, dots refer to the difference across subsamples of the cross-country-cross-announcement average of the bid-ask spread of the bonds. Returns are in log units. All series are multiplied by 100. Solid line and shaded areas are based on the estimation of equation (2.1). Our sample starts on October 1st 2019 and ends on the date of this draft.

featuring behavioral attributes and micro-frictions. In order to provide more data to distinguish across theories, we also look at equity volume. In figure 2.4, we directly depict the difference in volume log-growth across normal and epidemic subsamples. We find that both in AEs and in EEs trade volume features no change before the announcements. Consistent with previous studies (see, among others, Han (2020)), trade volume increases right after the announcement. This upward adjustment is more pronounced in EEs. In the next part of this study, we focus on sovereign bonds and document that liquidity seems to increase in the bond markets as well.

Bond markets. Figure 2.5(a) shows our results for bonds returns. The construction of the depicted data is identical to that used for equities. We note that the dynamics in the bond markets are less severe than those observed from equities. In a ± 60-minute window around the announcement, there is no significant adjustment in bonds returns for AEs. This observation is important as, by no-arbitrage, it suggests that cash-flow uncertainty is an important determinant of the market fluctuations observed during the COVID19 crisis. This high-frequency result is consistent with the results documented by Gormsen and Koijen (2020) looking at dividend futures.

Focusing on EEs, however, we note that sovereign bonds loose value ahead of announcements and then appreciate at the time of announcement like equities. Over our $\pm 60-$minute window, however, the cumulative return is nearly zero both across AEs and EEs, suggesting that bonds are an important hedge against contagion risk announcements.

In order to further investigate the role of sovereign bonds, we also look at the behavior of their bid-ask spread. Absent high-frequency data on bonds trading volume, we think of this spread as a measure of illiquidity in the market. We note an immediate decline in the bid-ask spread in AEs and a delayed one in EEs. This observation, paired with the decline in equity volume depicted in figure 2.4, suggests that investors may tilt their trade toward bonds right after announcements. In AE countries, we should not be surprised that such a reallocation of investment flows comes with almost no adjustment in bond prices since it may be the result of their monetary policy (Haddad et al. (2021)).

An alternative explanation for this muted response is that bond markets are subject to two offsetting forces. Specifically, flight to safety may promote bond appreciation but, simultaneously, sovereign default risk may increase and push bond prices downward. In order to study the plausibility of this hypothesis, we collect daily country-level data on CDS spreads and link their daily variation to daily news on contagion cases. We explain in detail how we measure news in the next section. Given that different countries entered this crisis with different levels of fiscal capacity, exploring country-level heterogeneity is important. For this reason, in our empirical analysis we

Table 2.2. CDS SPREADS AND CONTAGION NEWS

	A.E.		E.E.	
Contagion cases - news	6.138***	7.747**	27.669***	27.223***
	(1.984)	(3.792)	(8.226)	(8.355)
Adj. R2	0.02%	4.58%	0.18%	14.22%
Adj. R2 w/o	0.02%	4.58%	0.18%	14.22%
Country FE	Yes	Yes	Yes	Yes
Week FE	No	Yes	No	Yes

Notes: this table reports the results of the following regression:

$$\Delta S_t^i = d_0^i + d_t^i \cdot D_t^{Week} + \beta^g \cdot news_{t-1} + \epsilon_t^i, \quad \forall i \in g$$

where ΔS_t^i refers to the daily change of the CDS spread in country i; g refers to either the group of Advanced Economies (AEs) or that of Emerging Economies (EEs); d_0^i is a country-level fixed effect and D_t^{Week} is a weekly time fixed effect. 'Contagion cases - news' refers to the innovation in the growth of the global number of contagion cases as measured in section 3. 'Adj. R2 w/o' refers to the adjusted R squred from the same regression in which we omit the contagion news. Standard Errors are clustered at the country-level. Our sample starts on October 1st 2019 and ends on the date of this draft.

include both country-level fixed effects and week-level time fix effects.

In table 2.2, we show that that adverse contagion news tend to increase CDS spreads in a statistically significant way. This effect is three times stronger in EEs. Simultaneously, we document that these news produce a very modest increase in the adjusted R-squared of our regression, implying that for AEs, default concerns have been a second-order issue.

The role of domestic announcements. Recall that our cross section comprises 21 countries. We can think about the previous results about equity (bond) returns as the equal-weighted cumulative returns that an investor could obtain by trading ahead of each announcement across 21 sources of announcements (one per country) and in 21 equity (bond) markets, for a total of 21^2 possible trade combinations.

In order to disentangle the effects of local announcements on local markets, we also consider the average cumulative return of an investor that trades only in the domestic market ahead of domestic announcements. In figure 2.6, we focus on the average cumulative returns across 21 trade strategies that involve neither foreign news nor foreign assets. Our data confirms that bonds have a muted response to announcements. Equities, in contrast, tend to depreciate ahead of the announcement and then suddenly

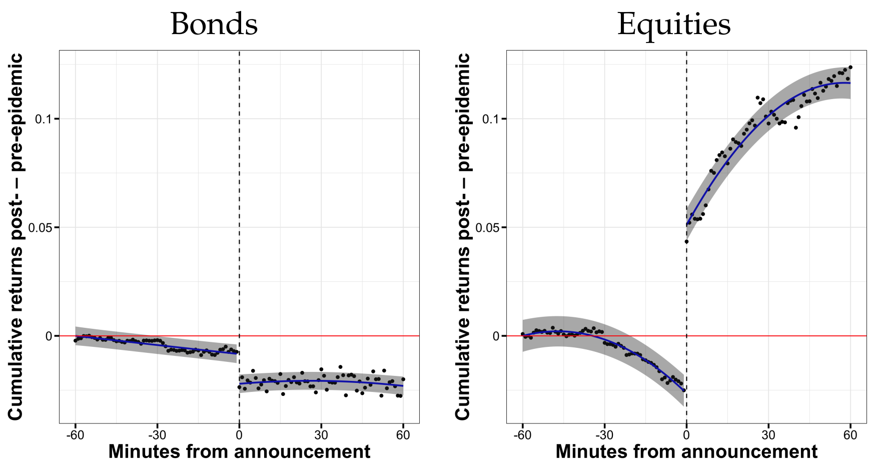


Figure 2.6. **LOCAL RETURNS AROUND DOMESTIC ANNOUNCEMENTS**

Notes: In each panel, dots denote the difference across subsamples of the cross-country average cumulative returns obtained from buying domestic equities 60 minutes before a domestic announcement and holding for 120 minutes. Returns are in log units. All series are multiplied by 100. Solid line and shaded areas are based on the estimation of equation (2.1). Our sample starts on October 1st 2019 and ends on the date of this draft.

appreciate afterward. This pattern resembles that derived by Ai and Bansal (2018) in a model in which the timing of information matters.

Covid vs macroeconomic announcements. In order to further isolate the role of medical announcements, we have created a dataset comprising the dates in which either inflation, industrial production, or GDP data are released in each country in our cross section. Our results continue to hold also when we exclude these days from our dataset (see, for example, figure .3 in 4).

3 Contagion News

In this section, we attempt to price news about pandemic risk. We do it using two fundamental measures, namely, unexpected changes in the number of contagion cases and unexpected changes in the tone of the news about contagion. The first measure is based on an objective count of COVID19 positive cases. Nevertheless, across different months or contagion waves, the same variation in the number of cases may be associated with different assessments of risk. For this reason, we find it necessary also to study a media-based measure of news tone.

Our analysis confirms that global epidemic news have a significant market price of risk. In April 2020, at the peak of the first COVID contagion wave in AE, daily equity risk premia may have increased by 28% in AEs and by 13% in EEs compared to the median risk premia in our sample.[7]

3.1 Data Collection

Twitter-based news. In the spirit of Baker et al. (2016), we identify major newspapers for a large cross section of countries (see table .1 in the appendix). In contrast to Baker et al. (2016), we do not analyze articles, rather we track news published on Twitter in real time, so that we can produce high frequency data when needed. More specifically, we track the news related to the COVID19 pandemic posted by major newspapers on Twitter. We do so by searching for keywords such as 'coronavirus' and 'covid19'. For each newspaper, we identify the location of its headquarter so that we can identify its specific time-zone.

In table 2.3, we report a summary of our social media–based dataset. It is very comprehensive and it features several dimensions that enable us to study both information production and diffusion. Specifically, our ability to track retweets and likes gives us a high-frequency measure of attention. Google searches are often used to measure attention (Da et al. 2011; Ramelli and Wagner 2020), but to the best of our knowledge they are not provided minute-by-minute and they do not account for the timing of initial production of the news, an aspect that is very important when analyzing capital market reactions.

The time series behavior of our news indicator is depicted in figure 2.7. For each country, we also depict the beginning of the epidemic period which we identify on the day in which the number of confirmed cases of COVID19 becomes greater than 100. We note several interesting patterns. First of all, there is significant heterogeneity across countries in the timing of the information diffusion. Across several countries, information diffusion becomes more intense after the beginning of the local epidemic

[7] These numbers are annualized according to the number of annual trading days and are net of the median risk premium in our full sample.

Table 2.3. NEWSPAPERS DATASET

Country	No. News Providers	Tweets	Retweets	Likes	Topics			
					Mortality	Quarant.	Med. Supply	Vaccines
Argentina	4	77,407	1,205,844	3,155,405	13%	10%	14%	63%
Australia	4	17,680	144,940	348,606	20%	39%	12%	29%
Brazil	4	32,596	1,332,180	8,710,524	45%	8%	15%	32%
Canada	5	48,716	443,544	863,678	33%	10%	17%	40%
Chile	4	34,061	408,725	631,767	56%	6%	10%	28%
China	3	32,879	948,862	2,582,197	39%	14%	19%	28%
Colombia	4	32,942	475,007	1,451,463	17%	12%	25%	45%
France	4	47,095	1,426,120	2,388,336	25%	26%	27%	22%
Germany	4	12,240	148,118	332,098	20%	24%	20%	35%
Hong Kong	3	21,339	420,614	607,725	17%	32%	21%	31%
India	4	103,814	937,109	5,610,418	32%	23%	16%	29%
Italy	3	33,721	265,694	715,064	10%	32%	29%	28%
Japan	4	19,051	157,250	278,263	18%	13%	30%	39%
Korea	4	13,550	82,916	144,299	45%	10%	26%	20%
Mexico	4	79,338	1,626,362	4,265,100	14%	11%	25%	50%
New Zealand	3	28,103	73,736	302,778	12%	38%	18%	32%
Spain	4	38,856	2,669,028	4,796,419	30%	20%	14%	36%
Switzerland	4	8,394	37,183	47,194	22%	20%	25%	33%
UK	4	25,366	1,145,886	2,287,563	27%	30%	15%	29%
USA	11	116,644	7,274,708	17,294,236	29%	7%	23%	41%
Total	85	823,792	21,223,826	56,813,133	26%	19%	20%	34%

Notes: This table shows summary statistics of COVID19-related news data that we collect for a large cross section of countries. Our real-time data range from January 1st 2020 to the date of this manuscript. For each country, we report number of news providers and number of tweets collected. We also report the total number of retweets and likes as measures of attention. The last four columns report the share of tweets mentioning number of deaths, quarantine measures, medical supply, and vaccines, respectively.

period. We note that both the diffusion of news, that is, number of tweets, and the attention to the news, that is, number of retweets, increase rapidly after the beginning of the local epidemic period.

Figure 2.8 shows both diffusion and attention to the news at the global level, that is, when we aggregate all of our tweets and retweets across countries. In figure 2.8(a), the right panel of this figure provides a breakdown of the most prominent topics addressed in the COVID19 tweets, namely, vaccines, death risk, quarantine measures, and availability of medical supply. The attention to all of them increased substantially, with vaccines becoming prominent in the fall 2020. In figure 2.8(b), we document similar results for high-attention tweets, i.e., tweets ranked top-1% by number of retweets within each one of our countries. For this subset of tweets, we collected also their retweets with 'quote', i.e., with text written by the retwitters in order to study their tone. We discuss this task in what follows.

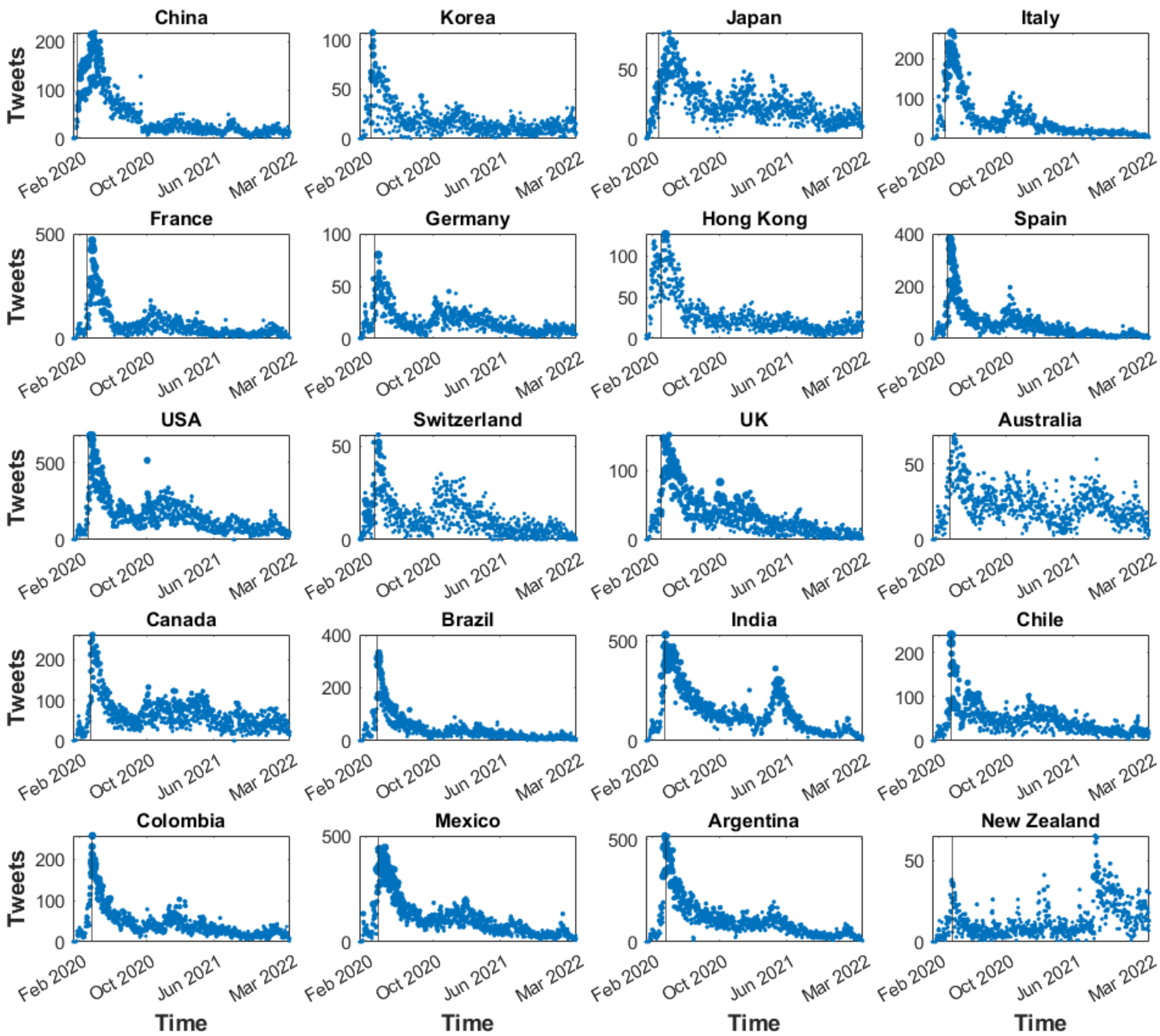
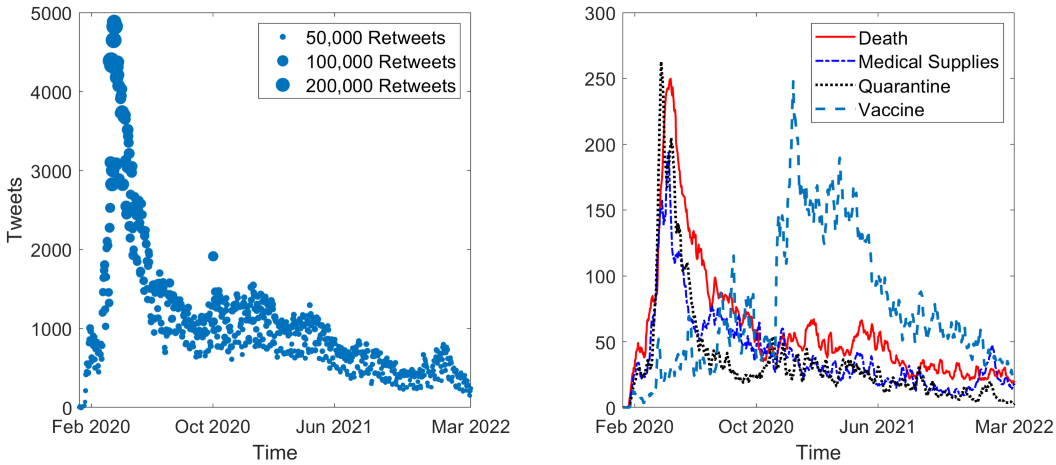


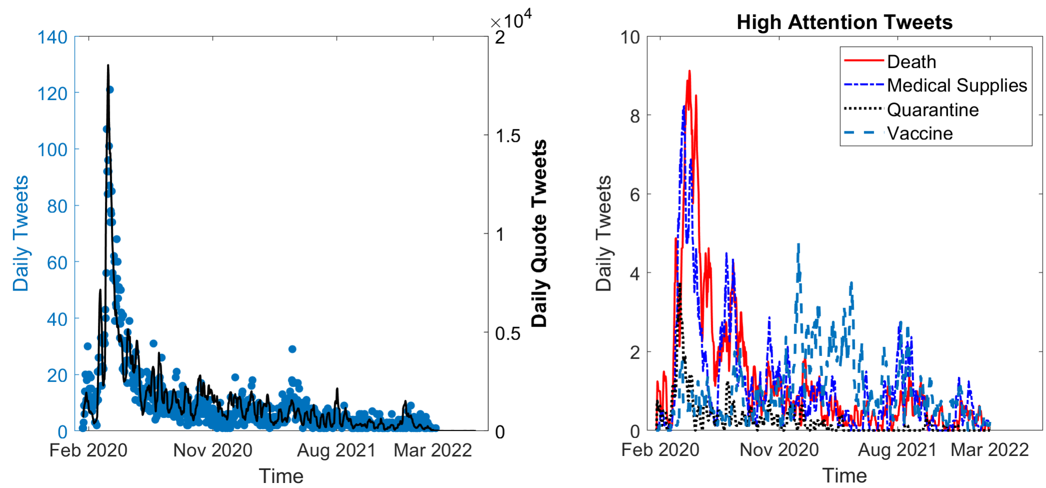
Figure 2.7. **INFORMATION DIFFUSION AND ATTENTION ACROSS COUNTRIES**

Notes: This figure shows the daily number of tweets posted in each country by major newspapers. The vertical axis shows the daily number of tweets. The size of each data point represents the number of retweets scaled by the maximum daily number of retweets for each country. The sample starts on January 8th 2020 and ends on the date of this draft. The vertical line depicts the date that each country had more than 100 confirmed cases of COVID19. More details on the data collection are reported in the Appendix.

Figure 2.9 shows the intraday pattern of the diffusion of COVID19 news for each country. This figure is not based on universal time, rather it accounts for country-specific time. In each country, we consider two country-specific subsamples, that is, the pre-epidemic and epidemic period. There are two main takeaways from this picture: (i) the diffusion of COVID19-related news increases significantly with local epi-



(a) Full Sample



(b) High-Attention Tweets

Figure 2.8. Global Information Diffusion

Notes: In panel a, the left panel of this figure shows the daily total number of tweets posted across countries by major newspapers. The vertical axis shows the daily number of tweets. The size of each data point represents the number of retweets scaled by the maximum daily number of retweets. The right panel shows the daily number of tweets related to death-risk, (scarcity of) medical supplies, quarantine, and vaccines. The tweets were identified using a multilingual bag-of-words approach. In panel b, we focus on high-attention tweets, i.e., top-1% by number of quote (re)tweets. The sample starts on January 8th 2020 and ends on the date of this draft. More details on the data collection are reported in the Appendix.

demic conditions; and (ii) a significant share of the diffusion takes place while the local capital markets are open. Hence monitoring media activity can be a very useful tool to track in real-time the information set of financial market participants.

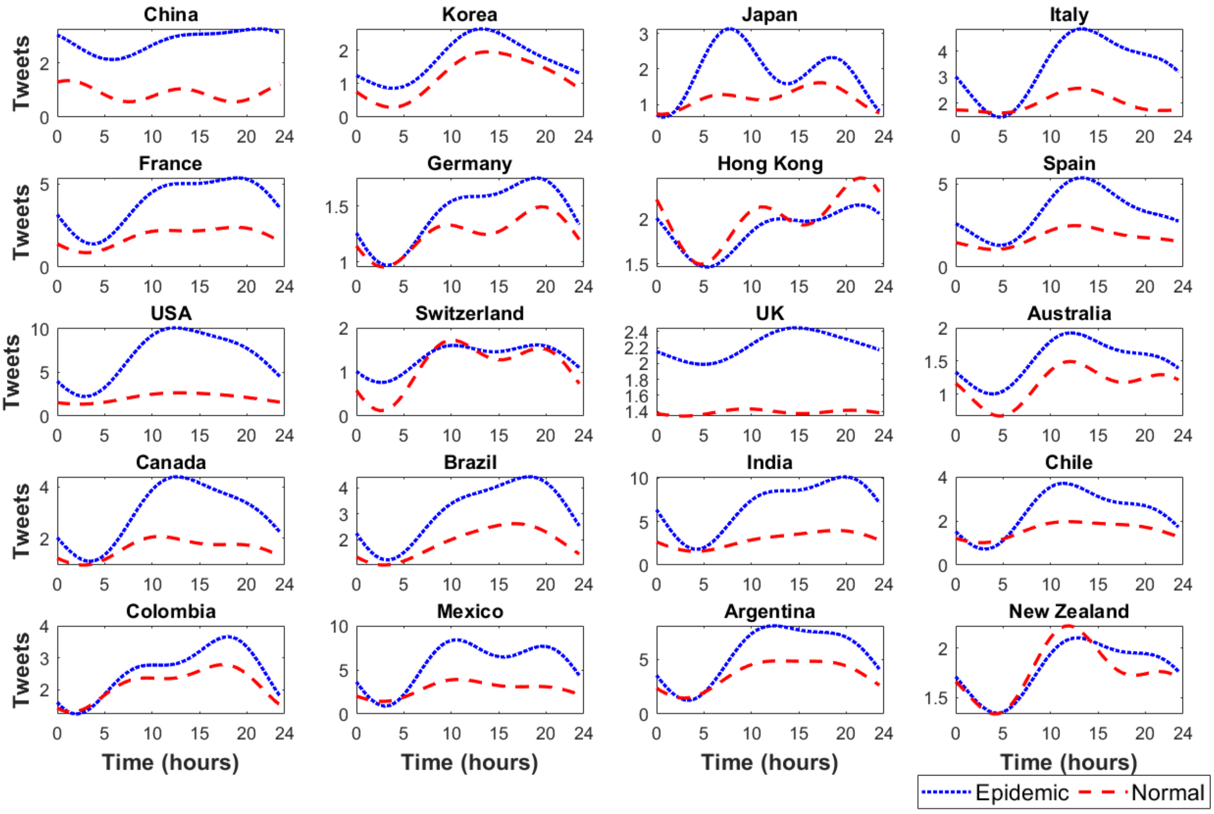


Figure 2.9. INTRADAY INFORMATION DIFFUSION

Notes: This figure shows the intra-day trend of the number of tweets posted every 30 minutes across several countries in our dataset. The dotted line represents the intra-day trend in the epidemic period, identified when a country has more than 100 cases of COVID19. The dashed line represents the intra-day trend in the pre-epidemic period. The sample starts on January 8th 2020 and ends on the date of this draft. Time refers to local time zone of each newspaper. More details on the data collection are reported in the Appendix.

Tweet Tone. Since we use Twitter activity to form a high-frequency risk factor, we need to identify the tone of the tweets, that is, we need to know whether they relate to either good or bad news. Given (i) the high volume of tweets that we collect, and (ii) the fact that our tweets are written in different languages, we use Polyglot (available at `https://pypi.org/project/polyglot/`), i.e., a natural language pipeline that supports multilingual applications with polarity lexicons for 136 languages. This computer-based mapping algorithm reads our text and classifies the words into three degrees of polarity: +1 for positive words, -1 for negatives words and 0 for neutral words. We provide two examples in table .2 (see our appendix).

Our measure of the tone of the tweets is based on the count of positive words minus the count of negative words, divided by the sum of positive and negative word counts (Twedt and Rees, 2012). We compute this measure at the country level at both the hourly and the daily frequency. We then aggregate this measure across countries in order to obtain a global measure.

We depict our global tone factor in figure 2.10, left panel. Its time-pattern is consistent with the observed contagion dynamics. Specifically, the tone became very negative by the end of January as the conditions in China started to precipitate. It improved in early February, when there was still no sign of massive contagion in Europe, and it declined again when the epidemic started in Italy. The slow improvement of the tone of our tweets observed after the beginning of March pairs well with the observed flattening of the contagion curves in many of the countries in our dataset. We find these results reassuring as they confirm that our text analysis algorithm tracks the contagion dynamics in a reliable manner.

In addition, we note that collecting both all original tweets and all of their retweets is computationally impossible for us. In table .4 (see 4), we show that there is a positive and significative correlation between the tone of the original tweets and that of the top-1% quote (re)tweets, meaning that our methodology captures a relevant-and-consistent partition of tweets.

For the sake of our asset pricing analysis, we focus on the innovations to the tone of our tweets. One simple way to extract these innovations is to consider the difference in the tone at day t and its 5-day backward looking moving average assessed at time $t-1$. We depict this time series in the right panel of figure 2.10 and note that it is nearly serially uncorrelated.

Contagion and financial data. Contagion data are from official medical bulletins. Our primary source is CSSE at Johns Hopkins University.[8] News to the contagion factor are obtained by computing the difference between the daily growth rate of contagion cases at time t and its backward-looking time $t-1$ moving average computed over the previous 5 days. We choose a 5-day window because it matches the number

[8] https://github.com/CSSEGISandData/COVID19/tree/master/csse_covid_19_data/csse_covid_19_time_series

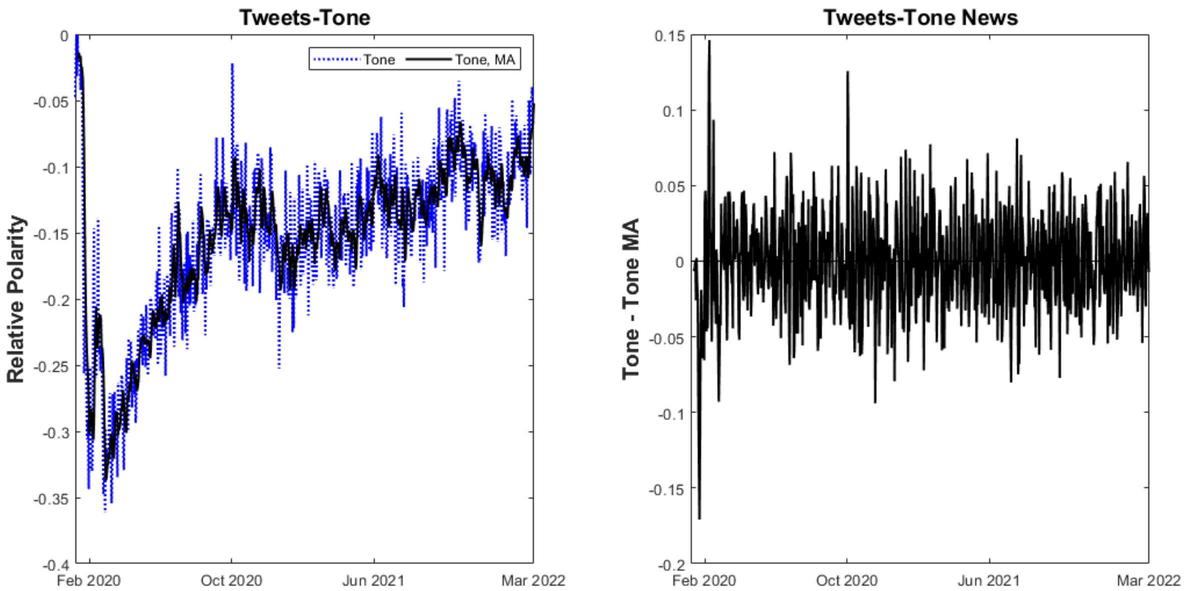


Figure 2.10. TWITTER-BASED COVID19 FACTOR

Notes: This figure shows our daily global Twitter-based COVID19 factor. We use Polygot to measure the polarity of our tweets and compute the tone of each tweet according to Twedt and Rees (2012). We aggregate the tones at a daily frequency and across countries. MA refers to a backward looking 5-day moving average. The news at time t is computed as the difference between the tweets-tone at time t and their MA at time $t-1$. The sample starts in early January 2020 and ends on the date of this draft.

of days of a typical trading week.

Since our contagion-based factor spans a 7-day week, we assign to Friday the average growth rate of global contagion cases that occurred on Friday, Saturday, and Sunday.[9] Our financial data sources are detailed in table .3 (see 4).

We note that equity returns have been much more volatile in the epidemic period. Most importunately, the intra-day patterns have become much more correlated once all countries have gone into an epidemic state. This result suggests that we can think of the epidemic as a slowly diffusing common factor. Our empirical asset pricing analysis is based on this observation.

When we turn our attention to bonds in the epidemic period, we see more volatile patterns than in the pre-epidemic period. In contrast to equities, we see no substantial change in their commonalities across countries. We see this as consistent with COVID19 being a global risk factor that affects countries at different times and with

[9]For the Easter Holiday, we assign to Thr the average daily growth rate of global cases from Thr to the following Mon.

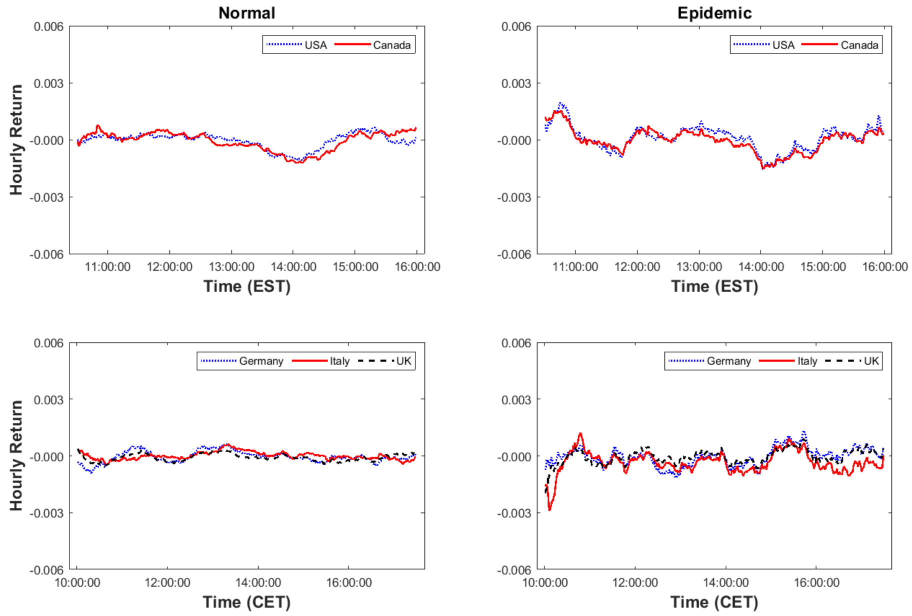
different intensities. Our empirical asset pricing analysis takes into consideration the hypothesis that our countries may feature heterogeneous exposure to global contagion risk.

In order to show the relevance of local epidemic conditions, in figure 2.11 we show the intra-day behavior of returns pre- and post-epidemic for equities, bonds, and currencies. We focus on two groups of countries with similar stock exchange timing, namely US and Canada (EST timezone), and Italy, UK, and Germany (CET timezone). The countries in the second group are interesting because they have experienced very different exposures to COVID19. Italy has been affected first and in an intensive way. Germany has been able to mitigate the contagion during the first contagion wave and has seen a pick up in contagion numbers as soon as it lessened the lockdown measures. The UK has changed its strategic response to the crisis in the middle of the epidemic period.

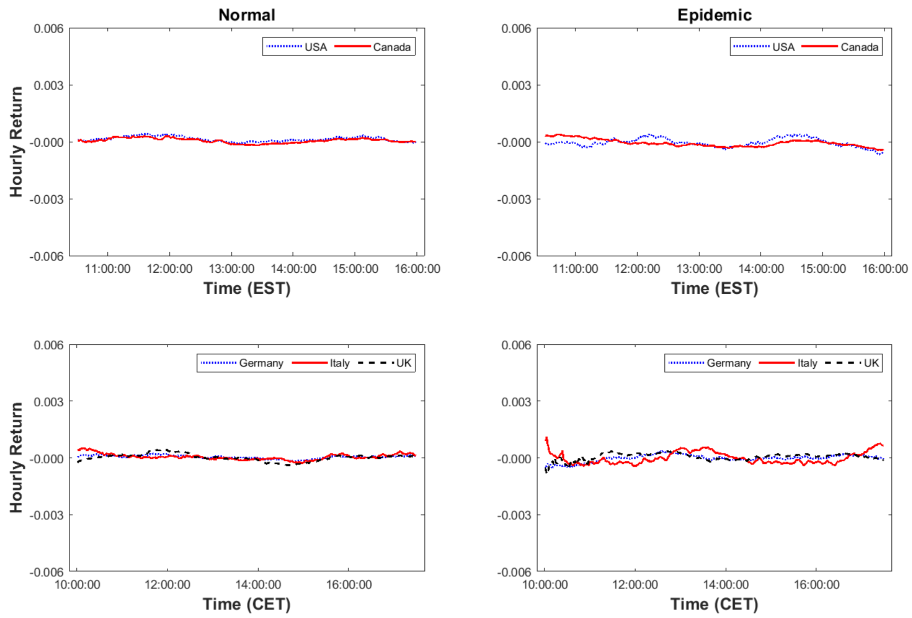
We note that equity returns have been much more volatile in the epidemic period. Most importunately, the intra-day patterns have become much more correlated once all countries have gone into an epidemic state. This result suggests that we can think of the epidemic as a slowly diffusing common factor. Our empirical asset pricing analysis is based on this observation.

When we turn our attention to bonds in the epidemic period, we see more volatile patterns than in the pre-epidemic period. In contrast to equities, we see no substantial change in their commonalities across countries. We see this as consistent with COVID19 being a global risk factor that affects countries at different times and with different intensities. Our empirical asset pricing analysis takes into consideration the hypothesis that our countries may feature heterogeneous exposure to global contagion risk.

Additional relevance of cases. One additional reason to focus on the number of cases as a relevant determinant of risk-premia is that many tweets in our sample focus on this topic. Specifically, we apply the Sievert and Shirley (2014) Latent Dirichlet Allocation (LDA) topic model to our covid-related tweets from English-written newspapers. Tweets are preprocessed, i.e., we remove stopwords and symbols such as #,



Equities



Bonds

Figure 2.11. INTRA-DAY RETURNS BEHAVIOR AND EPIDEMIC CONDITIONS

Notes: For each asset class, we depict per- and post-pandemic intra-day return patterns. Data are averaged across days. In each country, the epidemic period starts when there are more than 100 cases of COVID19. The sample starts in October 2019 and it ends October 2020. Bond and stock hourly returns start one hour after the opening of the markets. All returns are in raw units.

we 'stem' our words, and account for both unigrams and bigrams. We apply the unsupervised machine learning model to our data at the country-level.[10] When λ is set to the canonical value of 0.5, in most of our English-speaking countries, the top unigrams and bigrams from the main topics include 'covid cases' or related terms. We report an example in figure .4 (see 4). Our data visualization webpage lets the interested reader choose different values of λ.

3.2 The Market Price of Viral Contagion News

Daily news. Every day, we group countries into three portfolios according to their relative number of COVID19 cases measured the previous day. We do this separately for AEs and EEs. The H (L) portfolio comprises the top (bottom) countries in terms of COVID19 cases. We also consider an investment strategy long in the H portfolio and short in the L portfolio. We refer to the returns of this portfolio as HML-COVID19.

We report common summary statistics for these portfolios in table 2.4. The turnover in each portfolio is moderate. The in-sample average of the returns in all portfolios is not different from zero, which is not surprising given our short sample which comprises both contagion waves and their temporary disappearing. All portfolio returns have substantial volatility and negative skewness. Focusing on the first quartile of the distribution of returns, we see that the portfolio comprising the more exposed countries tends to have more severe negative downside risk. This is an aspect that we capture in our conditional no-arbitrage model.

Given these preliminary observations, we consider the following conditional asset pricing model,

$$r^{ex}_{f,t+1} = \bar{r}^{ex}_{f,t} + \beta_{f,t} \cdot news^{glob}_{t+1}, \quad f \in \{H, M, L\}, \tag{2.2}$$

$$\beta_{f,t} = \beta_0 + \beta_{f,1} X_{f,t}, \tag{2.3}$$

$$\frac{\partial \bar{r}^{ex}_{f,t}}{\partial X_{f,t}} = \lambda \beta_{f,1}, \tag{2.4}$$

[10]Topics are indexed by $k = 1, ..., 5$. λ determines the weight given to the probability of term w under topic k relative to its lift (measuring both on the log scale). Setting $\lambda = 1$ results in the familiar ranking of terms in decreasing order of their topic-specific probability, and setting $\lambda = 0$ ranks terms solely by their lift.

Table 2.4. SUMMARY STATISTICS FOR PORTFOLIOS

	Low	**Medium**	**High**	**HML$_{COVID19}$**
Panel A: Advanced economies				
Mean	0.014	0.039	0.020	0.006
	(0.060)	(0.060)	(0.072)	(0.031)
StDev	1.159	1.331	1.433	1.044
Skewness	-1.222	-0.762	-1.648	-0.108
First Quartile	-0.471	-0.496	-0.497	-0.545
Avg. N. Countries	5.004	4.004	4.992	-
Turnover (%)	0.5	1.3	0.6	-
Panel B: Emerging economies				
Mean	0.009	0.044	0.096**	0.086
	(0.087)	(0.094)	(0.049)	(0.063)
StDev	1.69	1.855	1.75	1.605
Skewness	-2.106	-1.255	-0.752	0.316
First Quartile	-0.662	-0.862	-0.774	-0.942
Avg. N. Countries	3.003	1.997	2	-
Turnover (%)	0.4	0.8	0.5	-

Notes: This table shows summary statistics for the equity excess returns of portfolios formed on a daily basis according to the relative share of country-specific COVID19 cases measured the day before formation. Hourly excess returns are in log units and multiplied by 100. Portfolios are obtained from equity indexes. Our real-time data range from February 2020 to the date of this manuscript. Turnover measures the number of countries entering or exiting a portfolio relative to the total number of countries in a specific portfolio $\times$ number of days in our sample. Numbers in parenthesis are HAC-adjusted standard errors.

where X_t is the share of contagion cases associated to portfolio f at time t, and λ is the market price of risk (MPR) of the global news factor $news^{glob}_{t+1}$.

This model can potentially capture many of the features of returns seen so far. First, it captures predictability through contagion-based time-varying betas. Second, it has the potential to capture higher negative skewness for countries that go through more severe contagion paths. Consider the case of portfolio H comprising countries receiving a sequence of relatively more severe adverse contagion news. This portfolio will have severe exposure to adverse news as the relative contagion share of the portfolio grows. When the relative contagion share starts to flatten out and decline, the sensitivity of this portfolio to good news is reduced ($|\beta_{H,t}|$ shrinks). This means that returns become less sensitive to positive news and hence the right tail of the returns distribution is shortened.

Third, consistent with our previous descriptive returns, it accounts for heteroge-

Table 2.5. SUMMARY OF MPR ESTIMATION

	Covid Cases		Twitter News	
	A.E.	E.E.	A.E.	E.E.
Local units				
coef	-0.003^{***}	-0.006^{***}	0.013^{***}	0.007^{***}
se	(0.001)	(0.001)	(0.003)	(0.001)
USD units				
coef	-0.005^{***}	-0.005^{***}	0.011^{***}	0.006^{***}
se	(0.001)	(0.002)	(0.003)	(0.001)
Controlling for MKT				
coef	-0.002^{***}	-0.007^{***}	0.008^{***}	0.008^{***}
se	(0.001)	(0.001)	(0.002)	(0.001)

Notes: This table shows the results of the conditional linear factor model described in equations (2.2)–(2.4). Portfolios are formed on a daily basis according to the relative share of country-specific COVID19 cases measured the day before formation (X_t). On the left (right), the COVID19 factor is measured as the news to global COVID cases growth (tone of COVID-related tweets). When we measure the COVID19 news as unexpected number of contagion cases (unexpected improvement in COVID19-related tweets), we expect a negative (positive) market price of risk (MPR). Both daily excess returns and market prices of risk are in log units. Our cross section of test assets comprises both equity and bond portfolios. Our real-time data range from 2/1/2020 to the date of this manuscript. Estimates and HAC-adjusted standard errors are obtained through GMM.

neous exposure to global contagion news. Last but not least, it enables us to identify the market price of risk of this global contagion component, λ. By no-arbitrage, the extent of time-series predictability of our excess returns must equal $\lambda\beta_{f,1}$, and $\beta_{f,1}$ can be easily estimated in the time-series by considering the multiplicative factor $X_{f,t} \cdot news^{glob}_{t+1}$.

We report our main results obtained from daily data in table 2.5. The first two columns are based on unexpected changes in the growth of global contagion cases. The right-most columns are based on unexpected changes in the global tone of tweets. Note that the set of countries that we consider provide daily updates about contagion cases at the end of the day. In order to properly represent the information set of investors, in our asset pricing model we lag the news by one day, i.e., we assume that day-t returns respond to news released in the evening of day $t-1$.

We estimate our asset pricing model through GMM and notice that all portfolios have an untabulated significant exposure to our contagion-based news, $\beta_{f,t}$.[11] In our

[11]The share of contagion cases across our three portfolios have very different scales and variability.

sample, the portfolio of countries with the highest share of COVID19 cases tends to be more exposed to contagion news. This sign is consistent with our expectations since positive (negative) news about global contagion growth (tone of tweets) refers to an adverse shock to equity returns. Most importantly, the implied daily market price of risk is negative (positive) and significant with respect to contagion (tone of tweets) news. This means that the relative share of contagion cases forecasts an increase in expected future returns across all portfolios ($\lambda\beta_{f,1} > 0$). Equivalently, the share of contagion cases is a relevant positive predictor of future cost of capital.

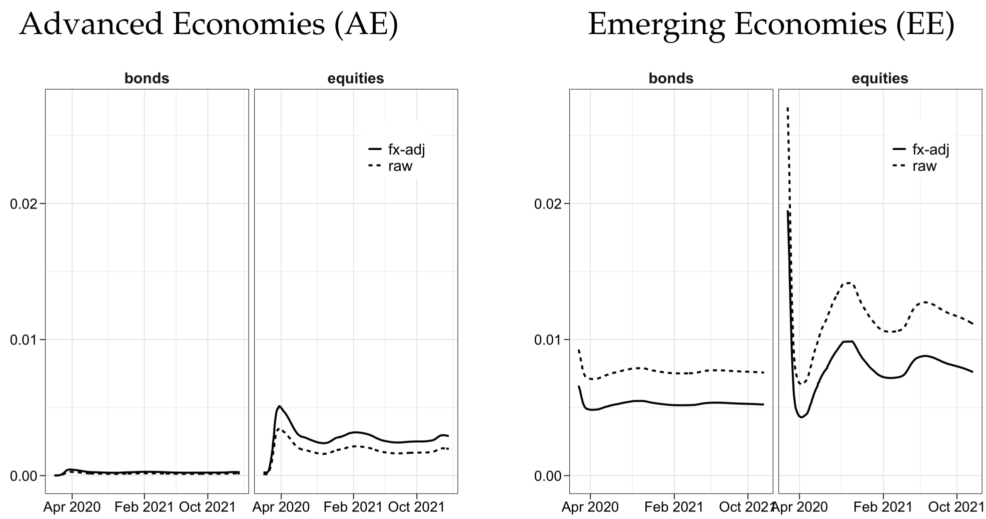
Our results hold regardless of whether we run our model using local-currency returns or returns in USD. Furthermore, our results remain significant when we estimate a two-factor version of our model which controls by global market risk as measured by the MSCI Global Index.[12] Looking at the output of our specifications and accounting for estimation uncertainty, we conclude that 0.3% is a reasonable lower bound on the daily market price of risk of daily contagion news. We consider this estimate as very significant, consistent with the great contraction experienced in equity markets during the first waive of the epidemic period.

Simultaneously, we note that this value is very plausible once we account for two observations. First, this is not the MPR of a financial factor and the associated estimated betas are very small. Second, contagion risk follows waves with a relatively short half-life. Equivalently, the exposure of our assets to this risk are small and relatively quick in reverting to zero. This phenomenon is depicted in figure 2.12(a). Our results confirm that sovereign bonds issued by AEs are not sensitive to contagion risk. Equities, instead, experienced a more pronounced increase in their required risk premium among High-COVID countries. In contrast, in EEs both bonds and equities feature a much more pronounced increase in their riskiness. Bonds' exposure, however, has been smaller than that of equities', confirming that also EEs' bonds are safer with respect to contagion risk.

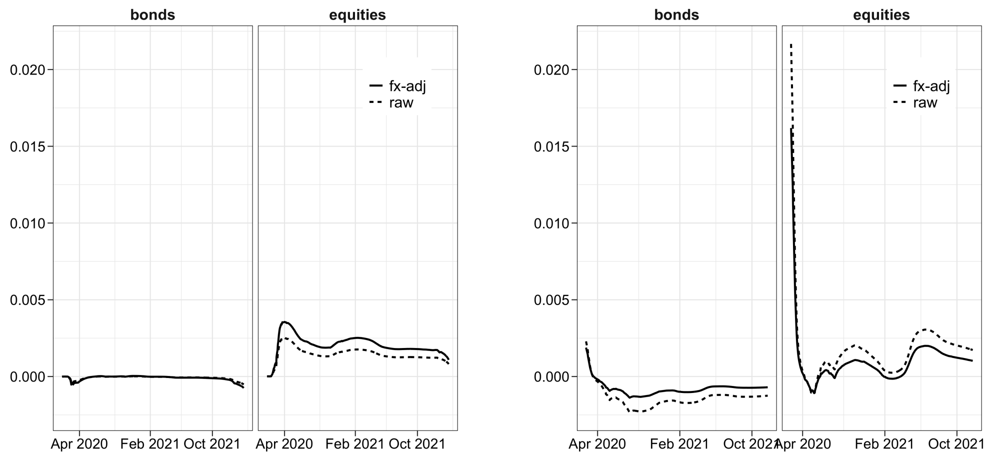
In figure 2.12(b), we show the estimated risk premium on an HML-COVID19 strat-

As a result, the coefficients $\beta_{f,1}$ are not revealing of the sorting of $\beta_{f,t}$ across portfolios. For this reason, we report only estimated MPRs.

[12]Throughout our study, when we consider the MSCI index to control for the market we use returns in USD.



(a) Expected returns for H_{COVID} portfolios



(b) Expected HML_{COVID}

Figure 2.12. EXPECTED RISK PREMIA

Notes: The left (right) panels refer to portfolios of countries within the AE (EE) group. The top panels show the estimated risk premium on a portfolio of countries with a share of High-COVID19 cases on bond and equity portfolios. The bottom panels refer to the HML-COVID strategy. These results are based on the specifications reported in the last two columns of table 2.5. The solid line refers to exchange rate–adjusted returns, i.e., returns expressed in USD.

egy on either bond or equity portfolios across AEs and EEs. Focusing on this strategy helps us to highlight the role played by heterogeneous exposure to contagion risk. We document several novel empirical results. First of all, we note that the riskiness of bonds has increased less in High-COVID countries than in Low-COVID countries. Equivalently, in High-COVID countries, bonds are relatively safer assets. As a result, an HML-COVID strategy on bonds provides an insurance premium. In AEs, this premium is very moderate, consistent with our prior empirical evidence on the muted response of bonds around medical announcements time. In EEs, instead, the insurance premium is quantitatively relevant both in local units and in USD. Hence this HML strategy may be of interest to international investors seeking a strong hedge against contagion risk.

Second, we notice that the equity-based HML strategy in AEs features a required premium similar to that estimated for the High-COVID portfolio. Equivalently, Low-COVID countries have experienced nearly zero change in their risk premium. This result is important because it implies that containment policies that keep contagion cases relatively low may be very valuable both in terms of lives saved and in terms of preventing severe financial wealth losses.

Turning our attention to equities in the EEs, we notice that the required premium on the associated HML strategy has increased dramatically at the beginning of the pandemic and it has followed the contagion waives that we have observed over the last 24 months. The initial jump should not be surprising as both China and India are in the High-COVID portfolio. It is interesting, however, that the response to global news of High- and Low-COVID EEs quickly became less heterogeneous by the end of April. At the time we are writing this manuscript, our estimation suggests that the HML-COVID is quantitatively very similar across AE and EE equity markets.

Additional results with daily data. In table .1 (see 4), we show that replacing covid-related news with market returns in our conditional model delivers no positive and statistically significant market price of risk. This result confirms that (i) a conditional CAPM model fails in capturing viral contagion risk; and (ii) our measures are informative about viral risk.

K. French provides the FF5 factors at daily frequency for developed countries (Fama and French (2017)). Given our limited cross section, estimating our model with time-varying betas for both our covid factor and the FF3/FF5 factors is not feasible. We take an hybrid approach and estimate a model in which the betas of our covid factor are time-varying, whereas the betas of the additional FF3/FF5 factors are constant. We report our estimated MPRs in table .2 and confirm our main results.

So far, we have estimated a model with heterogeneous and time-varying exposure to a common risk factor related to global contagion news. Our dataset enables us also to construct AE- and EE-specific measures of both COVID19 case growth and Twitter tone. See, for example, figure .1 in the Appendix.

We identify purely AE- and EE-specific components by regressing these fundamental measures on their global counterpart. The residuals of these two separate regressions represent for us AE- and EE-specific news. In 4, table .3, we show mixed results. Specifically, when we use only equity-based test assets, local contagion news (panel A) are priced negatively in AEs and positively in EEs. Twitter-based local news (panel B) have a market price of risk statistically not different from zero. Only when we use both bond and equity indices as test assets, local news are priced. Given these considerations, we consider our specification with heterogeneous and time-varying exposure to global contagion risk news as more robust.

Intra-day news. An important advantage of our Twitter-based risk-factor is that we can measure it at very high frequencies, in contrast to daily contagion cases. Using higher frequency data may help sharpen the estimate of the market price of risk because it provides an increased number of observations.

In this section, we focus only on European countries whose markets are open simultaneously. Specifically, we focus on ITA, ESP, UK, FRA, DEU, CHE, and SWE. Every day, we group them into three portfolios according to their relative number of COVID19 cases. In table 2.6, we show our estimation results when we link hourly equity and bond excess returns to hourly Twitter-based news.

As for daily data, we consider multiple specifications of our no-arbitrage model. In this case, we also report our estimated beta coefficients. The implied market price of risk is positive, well identified, and sizable. Our implied betas continue to be positive, i.e., viral contagion is priced as a source of risk. Consistent with the failure of the international-CAPM documented in table .1, our the implied market price of risk is still positive and sizable when we control for the market and use a broader cross section of test assets.

Table 2.6. Hourly Conditional Linear Factor Model

	β_0	$\beta_{L,1}$	$\beta_{M,1}$	$\beta_{H,1}$	*MPR*	N.Obs	N. Assets
			Panel A: equities and bonds, equities betas				
Hourly log returns							
coef	-0.090^{***}	9.879^{***}	4.043^{***}	2.853^{***}	0.014^{***}	4190	6
se	(0.007)	(0.712)	(0.294)	(0.207)	(0.003)	4190	6
Hourly log EUR returns (adjusting for FX)							
coef	-0.083^{***}	9.164^{***}	3.773^{***}	2.673^{***}	0.017^{***}	4190	6
se	(0.006)	(0.598)	(0.249)	(0.177)	(0.003)	4190	6
Hourly log returns controlling for the Market							
coef	-0.158^{***}	16.892^{***}	6.980^{***}	4.968^{***}	0.009^{***}	3951	6
se	(0.014)	(1.549)	(0.643)	(0.457)	(0.003)	3951	6
			Panel B: equities and bonds, bond betas				
Hourly log returns							
coef	-0.062^{***}	6.872^{***}	2.780^{***}	1.966^{***}	0.014^{***}	4190	6
se	(0.005)	(0.496)	(0.201)	(0.144)	(0.003)	4190	6
Hourly log EUR returns (adjusting for FX)							
coef	-0.058^{***}	6.385^{***}	2.609^{***}	1.851^{***}	0.017^{***}	4190	6
se	(0.004)	(0.421)	(0.174)	(0.124)	(0.003)	4190	6
Hourly log returns controlling for the Market							
coef	-0.109^{***}	11.743^{***}	4.831^{***}	3.439^{***}	0.009^{***}	3951	6
se	(0.010)	(1.072)	(0.442)	(0.315)	(0.003)	3951	6

Notes: This table shows the results of the conditional linear factor model described in equations (2.2)–(2.4). Portfolios are formed on a daily basis according to the relative share of country-specific COVID19 cases measured the day before formation (X_t). The coefficient $\beta_{f,t} = \beta_0 + \beta_f X_{f,t}$ refers to the exposure of the equity portfolio $f \in \{H, M, L\}$ to the COVID19 factor. We measure hourly COVID19 news as unexpected improvement in the hourly tone of COVID19-related tweets. Both hourly excess returns and market prices of risk are in log units. When we control for the market, returns are in USD, the market is measured by the MSCI Global Index and our factor model comprises a total of two factors. Our real-time data range from February 2020 to the date of this manuscript. Estimates and HAC-adjusted standard errors are obtained through GMM.

In 4, we show that these results are robust to weighting the share of covid cases by population (table .4). More broadly, there may be several country-specific characteristics (fiscal conditions, competition, ...) that could make our portfolios differently exposed to pandemic risk. In order to address this concern, we replace β_0 in equation (2.3) with β_0^f, i.e., a country-specific fixed effect to its exposure. Thanks to the hourly frequency, we have enough observations to estimate this richer model. Our results continue to hold and are reported in table .5.

Table 2.7. VOL-ADJUSTED CONDITIONAL LINEAR FACTOR MODEL

	β_0	$\beta_{L,1}$	$\beta_{M,1}$	$\beta_{H,1}$	MPR	N.Obs	N. Assets
			Panel A: equities, news from Twitter				
Daily log returns							
coef	-0.112^{**}	17.422^{***}	8.557^{***}	2.510^{*}	0.011^{*}	520	3
se	(0.046)	(4.664)	(1.926)	(1.381)	(0.006)	520	3
			Panel B: equities, news from Twitter				
Hourly log returns							
coef	-0.006^{***}	0.505^{***}	0.199^{***}	0.153^{***}	0.044^{**}	4298	6
se	(0.001)	(0.133)	(0.057)	(0.041)	(0.022)	4298	6

Notes: This table shows the results of the conditional linear factor model described in equations (2.2)–(2.4). Portfolios are formed on a daily basis according to the relative share of country-specific COVID19 cases measured the day before formation (X_t). The coefficient $\beta_{f,t} = \beta_0 + \beta_f X_{f,t}$ refers to the exposure of the equity portfolio $f \in \{H, M, L\}$ to the COVID19 factor. We measure hourly (daily) COVID19 news as unexpected improvement in the hourly (daily) tone of COVID19-related tweets. We project this factor on realized market volatility and use the implied residual in our estimation. Both excess returns and market prices of risk are in log units and are expressed in USD. The market is measured by the MSCI Global Index. Our real-time data range from February 2020 to the date of this manuscript. Estimates and HAC-adjusted standard errors are obtained through GMM.

Controlling for Volatility. In this last step of our research, we project our Twitter-based COVID factor on realized market volatility and use the implied residual to redo our analysis. Equivalently, we look at COVID news that are orthogonal to pure volatility shocks. We measure realized volatility as the standard deviation of the MSCI Global Index at the daily (hourly) frequency using a rolling window of a trading week (a single trading day). We report our results in table 2.7. Both daily data and intra-day data confirm that contagion news have an extremely high MPR, even after controlling for volatility.

International Flows. In order to further validate our results, we study international investment flows related to the countries in our cross section. Weekly net flows are from EPFR and they are rescaled by country-level GDP so that our results are not driven by country size. In this step, we exclude the US given its special role played in international markets (among others, see Maggiori 2017). After forming portfolios according to relative contagion levels, we forecast one-week ahead flows using the (lagged) weekly share of portfolio-level COVID19 cases.

Table 2.8. INTERNATIONAL FLOWS AND NEWS

	Bonds		Equities	
	AE	EE	AE	EE
β_0	0.247***	−0.171***	0.589***	0.135***
	(0.026)	(0.032)	(0.051)	(0.036)
β_1	−0.921***	0.311***	−4.500***	−0.996***
	(0.073)	(0.070)	(0.274)	(0.123)
J-stat	11.234	11.825	7.069	11.676
N	75	71	75	71

*Notes:*This table reports the results of the following linear system:

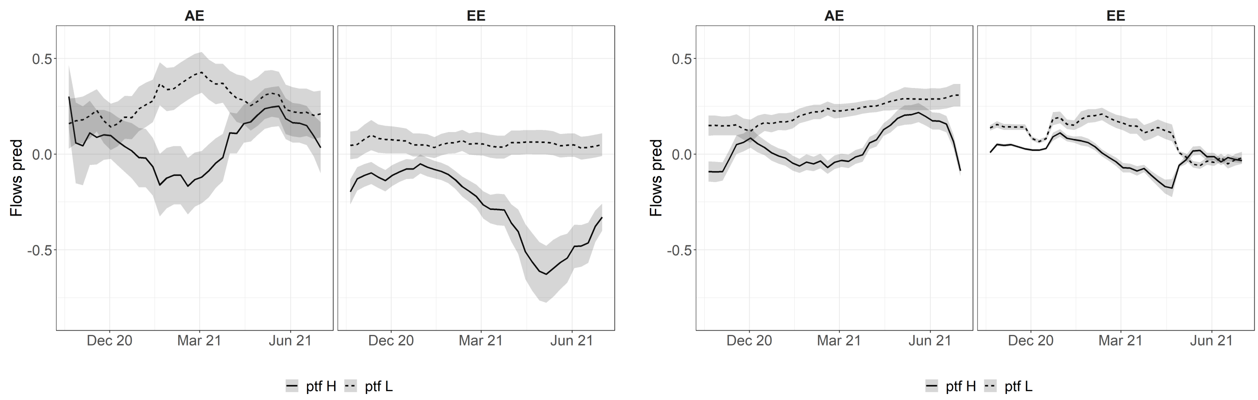
$$FL_t^f = \beta_0 + \beta_1 X_{t-1}^f + \epsilon_t^f$$

where FL_t^f is the flow to funds that invest in portfolio $f \in \{H, M, L\}$ during week t rescaled by portfolio$-f$ 2019 GDP; X_{t-1}^f refers to the weekly share of portfolio-specific COVID19 cases. Portfolios are formed on a weekly basis according to the relative share of country-specific COVID19 cases measured the week before formation. Fund flows-to-GDP is expressed in basis points (bps). Our data range from February 2020 to the date of this manuscript at a weekly frequency. Estimates and HAC-adjusted standard errors are obtained through GMM.

As reported in table 2.8, countries that start the week with a higher level of relative contagion are expected to receive lower net inflows ($\beta_1 < 0$)). This effect is reversed ($\beta_1 > 0$)) when we focus on net bond flows in EE, consistent with the idea that they may be perceived as safer assets and hence their demand may actually increase due to flight to safety. As shown in figure 2.13, low-COVID countries tend to receive a higher net inflow than high-COVID countries. This statement, however, does not apply to bonds in EEs. During the summer 2021, high-covid EEs have experienced higher inflows for their sovereign bonds.

4 Conclusion

In this study, we quantify the exposure of major financial markets to news shocks about global contagion risk while accounting for local epidemic conditions. We construct a novel data set comprising (i) medical announcements related to COVID19 for a broad cross section of countries; and (ii) high-frequency data on epidemic news diffused through Twitter. Across several classes of financial assets and currencies, we provide novel empirical evidence about financial dynamics (i) around epidemic an-



Equities *Bonds*

Figure 2.13. EXPECTED INVESTMENT FLOWS

Notes: For each asset class, we depict forecasted net investment flows. 'ptf H' ('ptf L') refers to a portfolio of countries with relatively high (low) contagion cases. We split our sample across advanced and emerging economies (AE and EE, respectively). The estimates are based on the following linear system:

$$FL_t^f = \beta_0 + \beta_1 X_{t-1}^f + \epsilon_t^f$$

where FL_t^f is the flow to funds that invest in portfolio $f \in \{H, M, L\}$ during week t rescaled by portfolio$-f$ 2019 GDP; X_{t-1}^f refers to the weekly share of portfolio-specific COVID19 cases. Portfolios are formed on a weekly basis according to the relative share of country-specific COVID19 cases measured the week before formation. Fund flows-to-GDP is expressed in basis points (bps). Our data range from February 2020 to the end of summer 2021 at a weekly frequency. Estimates and HAC-adjusted standard errors are obtained through GMM.

nouncements, (ii) at a daily frequency, and (iii) at an intra-daily frequency. Formal estimations based on both contagion data and social media activity about COVID19 confirm that the market price of epidemic risk is very significant. In the spirit of Mulligan (2020), we conclude that policies related to the prevention and containment of contagion could be precious not only in terms of lives saved but also in terms of global wealth.

Bibliography

Acemoglu, Daron, Victor Chernozhukov, Ivan Werning, and Michael D. Whinston, "TA Multi-Risk SIR Model with Optimally Targeted Lockdown," *NBER Working Paper No. 27102*, 2020.

Acharya, Viral V and Sascha Steffen, "The Risk of Being a Fallen Angel and the Corporate Dash for Cash in the Midst of COVID," *CEPR's Covid Economics Review No. 10*, 2020.

Ai, Hengjie and Ravi Bansal, "Risk Preferences and the Macroeconomic Announcement Premium," *Econometrica*, 2018, *86* (4).

Albuquerque, Rui A., Yrjo J Koskinen, Shuai Yang, and Chendi Zhang, "Love in the Time of COVID-19: The Resiliency of Environmental and Social Stocks," *CEPR's Covid Economics Review No. 11*, 2020.

Alfaro, Laura, Anusha Chari, Andrew N Greenland, and Peter K Schott, "Aggregate and Firm-Level Stock Returns During Pandemics, in Real Time," *CEPR's Covid Economics Review No. 4*, 2020.

Alon, Titan M, Matthias Doepke, Jane Olmstead-Rumsey, and Michèle Tertilt, "The Impact of COVID-19 on Gender Equality," *CEPR's Covid Economics Review No. 4*, 2020.

Alvarez, Fernando E, David Argente, and Francesco Lippi, "A Simple Planning Problem for COVID-19 Lockdown," Working Paper 26981, National Bureau of Economic Research April 2020.

Atkeson, Andrew, "How Deadly Is COVID-19? Understanding The Difficulties With Estimation Of Its Fatality Rate," Working Paper 26965, National Bureau of Economic Research April 2020.

Augustin, Patrick, Valeri Sokolovski, Marti G Subrahmanyam, and Davide Tomio, "In sickness and in debt: The COVID-19 impact on sovereign credit risk," *Journal of Financial Economics*, 2021.

Bahaj, Saleem and Ricardo Reis, "Central bank swap lines during the Covid-19 pandemic," *CEPR's Covid Economics Review No. 4*, 2020.

Baker, Scott R., Nicholas Bloom, and Steven J. Davis, "Measuring Economic Policy Uncertainty*," *The Quarterly Journal of Economics*, 2016, *131* (4), 1593.

Baker, Scott R, Nicholas Bloom, Steven J Davis, and Stephen J Terry, "COVID-Induced Economic Uncertainty," Working Paper 26983, National Bureau of Economic Research April 2020.

, , , Kyle Kost, Marco Sammon, and Tasaneeya Viratyosin, "The unprecedented stock market reaction to COVID-19," *Review Asset Pricing Studies*, 2020, (10).

, , , , , and , "The unprecedented stock market reaction to COVID-19," *The Review of Asset Pricing Studies*, 2020, *10* (4).

Barro, Robert J, José F Ursúa, and Joanna Weng, "The Coronavirus and the Great Influenza Pandemic: Lessons from the "Spanish Flu" for the Coronavirus's Potential Effects on Mortality and Economic Activity," Working Paper 26866, National Bureau of Economic Research March 2020.

Barrot, Jean-Noel, Basile Grassi, and Julien Sauvagnat, "Sectoral effects of social distancing," *CEPR's Covid Economics Review No. 3*, 2020.

Bartik, Alexander W, Marianne Bertrand, Zoë B Cullen, Edward L Glaeser, Michael Luca, and Christopher T Stanton, "How Are Small Businesses Adjusting to COVID-19? Early Evidence from a Survey," Working Paper 26989, National Bureau of Economic Research April 2020.

Bianchi, F, R Cram, and F Kung, "Using Social Media to Identify the Effects of Congressional Viewpoints on Asset Prices," *Working Paper*, 2021.

Bonaccolto, Giovanni, Nicola Borri, and Andrea Consiglio, "Breakup and Default Risks in the Eurozone," *Available at SSRN 3487453*, 2019.

Bretscher, Lorenzo, Alex Hsu, and Andrea Tamoni, "The Supply Channel of Uncertainty Shocks and the Cross-Section of Returns: Evidence From the COVID-19 Crisis," *Available at SSRN 3588418*, 2020.

, , Peter Simasek, and Andrea Tamoni, "COVID-19 and the cross-section of equity returns: Impact and transmission," *The Review of Asset Pricing Studies*, 2020, *10* (4).

Bybee, L, Bryan Kelly, and A Manela, "Business News and Business Cycles," *Working paper*, 2020.

Bybee, Leland, Bryan T Kelly, Asaf Manela, and Dacheng Xiu, "The structure of economic news," 2020.

Caballero, Ricardo J and Alp Simsek, “A Model of Asset Price Spirals and Aggregate Demand Amplification of a “COVID-19" Shock,” *CEPR Discussion Paper*, 2020.

Calomiris, Charles W and Harry Mamaysky, “How news and its context drive risk and returns around the world,” *Journal of Financial Economics*, 2019, *133* (2), 299–336.

Carletti, Elena, Tommaso Oliviero, Marco Pagano, Loriana Pelizzon, and Marti G Subrahmanyam, “The COVID-19 Shock and Equity Shortfall: Firm-level Evidence from Italy,” 2020.

Chiou, Lesley and Catherine Tucker, “Social Distancing, Internet Access and Inequality,” Working Paper 26982, National Bureau of Economic Research April 2020.

Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber, “The Cost of the COVID-19 Crisis: Lockdowns, Macroeconomic Expectations, and Consumer Spending,” *BFI Working Paper*, 2020.

Cookson, J. A., J Engelberg, and W. Mullins, “Echo Chambers,” *Working paper*, 2021.

Cororaton, Anna and Samuel Rosen, “Public Firm Borrowers of the US Paycheck Protection Program,” *CEPR’s Covid Economics Review No. 15*, 2020.

Correia, Sergio, Stephan Luck, and Emil Verner, “Pandemics Depress the Economy, Public Health Interventions Do Not: Evidence from the 1918 Flu,” *CEPR working paper*, 2020.

Corsetti, Giancarlo, Martin Bodenstein, and Luca Guerrieri, “Social Distancing and Supply Disruptions in a Pandemic,” *CEPR Discussion Paper*, 2020.

Coven, Joshua and Arpit Gupta, “Disparities in Mobility Response to COVID-19,” *Working paper.*, 2020.

Cox, J, , D. L. Greenwald, and S C Ludvigson, “What Explains the COVID-19 Stock Market?,” *Working Paper*, 2020.

Da, ZHI, JOSEPH Engelberg, and PENGJIE Gao, “In Search of Attention,” *The Journal of Finance*, 2011, *66* (5), 1461–1499.

e Castro, Miguel Faria and FRB St Louis, “Fiscal Policy during a Pandemic,” *Federal Reserve Bank of St. Louis*, 2020. Working paper.

Eichenbaum, Martin S, Sergio Rebelo, and Mathias Trabandt, “The Macroeconomics of Epidemics,” Working Paper 26882, National Bureau of Economic Research March

2020.

Elgin, Ceyhun, Gokce Basbug, and Abdullah Yalaman, "Economic Policy Responses to a Pandemic: Developing the COVID-19 Economic Stimulus Index," *CEPR's Covid Economics Review No. 3*, 2020.

Engle, Robert F, Stefano Giglio, Bryan Kelly, Heebum Lee, and Johannes Stroebel, "Hedging Climate Change News," *The Review of Financial Studies*, 2020, *33* (3).

Fama, Eugene F. and Kenneth R. French, "International tests of a five-factor asset pricing model," *Journal of Financial Economics*, 2017, *123* (3).

Farboodi, Maryam, Gregor Jarosch, and Robert Shimer, "Internal and External Effects of Social Distancing in a Pandemic," *CEPR's Covid Economics Review No. 9*, 2020.

Favero, Carlo A, "Why is COVID-19 Mortality in Lombardy so High? Evidence from the Simulation of a SEIHCR Model," *CEPR's Covid Economics Review No. 4*, 2020.

Fornaro, Luca and Martin Wolf, "Covid-19 Coronavirus and Macroeconomic Policy," *CEPR working paper*, 2020.

Garmaise, M., Y. Levi, and Hanno Lustig, "Spending Less After (Seemingly) Bad News," *Working paper*, 2021.

Gerding, Felix, Thorsten Martin, and Florian Nagler, "The Value of Fiscal Capacity in the Face of a Rare Disaster," *Working paper.*, April 2020.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus, "Inside the Mind of a Stock Market Crash," *arXiv preprint arXiv:2004.01831*, 2020.

Glover, Andrew, Jonathan Heathcote, Dirk Krueger, and José-Víctor Rios-Rull, "Health versus Wealth: On the Distributional Effects of Controlling a Pandemic," *CEPR's Covid Economics Review No. 6*, 2020.

Goldstein, Itay, Ralph Koijen, and Holger Mueller, "COVID-19 and Its Impact on Financial Markets and the Real Economy," *The Review of Financial Studies*, 2021.

Gormsen, Niels Joachim and Ralph S J Koijen, "Coronavirus: Impact on Stock Prices and Growth Expectations," *The Review of Asset Pricing Studies*, 2020, *10* (4), 574–597.

, Ralph Koijen, and Ian Martin, "Implied Dividend Volatility and Expected Growth," *Working Paper*, 2021.

Haddad, Valentin, Alan Moreira, and Tyler Muir, "When Selling Becomes Viral: Disruptions in Debt Markets in the COVID-19 Crisis and the Fed's Response," *The Review of Financial Studies*, 2021, *34* (11).

Hagedorn, M and K Mitman, "Corona Policy According to HANK," *CEPR Discussion Paper*, 2020.

Han, Leyla Jianyu, "IAnnouncements, Expectations, and Stock Returns with Asymmetric Information," *Working Paper*, 2020.

Hartley, Jonathan S and Alessandro Rebucci, "An Event Study of COVID-19 Central Bank Quantitative Easing in Advanced and Emerging Economies," 2020.

Hassan, Tarek A, Stephan Hollander, Laurence Van Lent, and Ahmed Tahoun, "Firm-level political risk: Measurement and effects," *The Quarterly Journal of Economics*, 2019, *134* (4), 2135–2202.

Hassan, Tarek Alexander, Stephan Hollander, Laurence van Lent, and Ahmed Tahoun, "Firm-level Exposure to Epidemic Diseases: Covid-19, SARS, and H1N1," *Working Paper Series*, 2020.

Hensvik, Lena, Thomas Le Barbanchon, and Roland Rathelot, "Which jobs are done from home? Evidence from the American Time Use Survey," *CEPR Discussion Papers*, 2020.

Ichino, Andrea, Carlo A Favero, and Aldo Rustichini, "Restarting the economy while saving lives under Covid-19," *CEPR Discussion Papers*, 2020.

Israelsen, Ryan, Nanadini Gupta, and Goldman Eitan, "Political Polarization in Financial News," *Working paper*, 2021.

Jones, Callum J, Thomas Philippon, and Venky Venkateswaran, "Optimal Mitigation Policies in a Pandemic: Social Distancing and Working from Home," Working Paper 26984, National Bureau of Economic Research April 2020.

Kaniel, Ron and Pingle Wang, "Unmasking Mutual Fund Derivative Use During the COVID-19 Crisis," *CEPR's Covid Economics Review No. 49*, 2020.

Krueger, Dirk, Harald Uhlig, and Taojun Xie, "Macroeconomic Dynamics and Reallocation in an Epidemic," *CEPR's Covid Economics Review No. 5*, 2020.

Lewis, Daniel, Karel Mertens, and James H Stock, "U.S. Economic Activity During the Early Weeks of the SARS-Cov-2 Outbreak," *CEPR's Covid Economics Review No. 6*, 2020.

Lucca, DAVID O. and EMANUEL Moench, "The Pre-FOMC Announcement Drift," *The Journal of Finance*, 2015, *70* (1).

Ludvigson, Sydney C, Sai Ma, and Serena Ng, "Covid19 and the Macroeconomic Effects of Costly Disasters," *CEPR's Covid Economics Review No. 9*, 2020.

Ma, Chang, John H Rogers, and Sili Zhou, "Global Economic and Financial Effects of 21st Century Pandemics and Epidemics," *CEPR's Covid Economics Review No.5*, 2020.

Maggiori, Matteo, "Financial Intermediation, International Risk Sharing, and Reserve Currencies," *American Economic Review*, 2017, *107* (10).

Manela, Asaf and Alan Moreira, "News implied volatility and disaster concerns," *Journal of Financial Economics*, 2017, *123* (1), 137–162.

Mulligan, Casey B, "Economic Activity and the Value of Medical Innovation during a Pandemic," Working Paper 27060, National Bureau of Economic Research April 2020.

Papanikolaou, Dimitris and Lawrence D. W. Schmidt, "Working Remotely and the Supply-side Impact of Covid-19," *Working Paper*, 2020.

Pástor, Luboš and Blair Vorsatz, "Mutual Fund Performance and Flows During the COVID-19 Crisis," *The Review of Asset Pricing Studies*, 2020, *10* (4).

Pelger, Markus, "Understanding Systematic Risk - A High-Frequency Approach," *The Journal of Finance*, 2020.

Ramelli, Stefano and Alexander Wagner, "Feverish Stock Price Reactions to COVID-19," *Forthcoming, Review of Corporate Finance Studies; Swiss Finance Institute Research Paper No. 20-12*, 2020.

Schoenfeld, Jordan, "The Invisible Risk: Pandemics and the Financial Markets," *Available at SSRN 3567249*, 2020.

Sievert, Carson and Kenneth Shirley, "LDAvis: A method for visualizing and interpreting topics," *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, June 2014.

Sinagl, Petra, "Cash-flow Risk and Industry Response to COVID-19 Outbreak," *Available at SSRN 3566511*, 2020.

Tetlock, Paul C, "Giving content to investor sentiment: The role of media in the stock market," *The Journal of finance*, 2007, *62* (3), 1139–1168.

Twedt, Brady and Lynn Rees, "Reading between the lines: An empirical examination of qualitative attributes of financial analysts' reports," *Journal of Accounting and Public Policy*, 2012, *31* (1), 1–21.

Data Sources

Table .1: **News Papers**

Country	Newspaper	Twitter Account	BBD	Language
Argentina	La Nacion	@LANACION		Spanish
Argentina	Clarin	@clarincom		Spanish
Argentina	Diario Cronica	@cronica		Spanish
Argentina	Infobae	@infobae		Spanish
Australia	The Age	@theage		English
Australia	The Australian	@australian		English
Australia	The Daily Telegraph	@dailytelegraph		English
Australia	Financial Review	@FinancialReview		English
Brazil	O Globo	@JornalOGlobo		Portuguese
Brazil	O Estado de Sao Paulo	@Estadao		Portuguese
Brazil	Folha de S.Paulo	@folha		Portuguese
Brazil	Gaucha ZH	@GauchaZH		Portuguese
Canada	Gazette	@mtlgazette	Yes	English
Canada	Globe and Mail	@globeandmail	Yes	English
Canada	Ottawa Citizen	@OttawaCitizen	Yes	English
Canada	Toronto Star	@TorontoStar	Yes	English
Canada	Vancouver Sun	@VancouverSun	Yes	English
Chile	La Tercera	@latercera		Spanish
Chile	BioBioChile	@biobio		Spanish
Chile	El Mostrador	@elmostrador		Spanish
Chile	The Clinic	@theclinicci		Spanish
China	People's Daily, China	@PDChina		English
China	China Xinhua News	@XHNews		English
China	China Daily	@ChinaDaily		English
Colombia	El Espectador	@elespectador		Spanish
Colombia	El Colombiano	@elcolombiano		Spanish
Colombia	El Heraldo	@elheraldoco		Spanish

(To be continued)

Country	Newspaper	Twitter Account	BBD	Language
Colombia	El Tiempo	@ELTIEMPO		Spanish
France	Le Monde	@lemondefr	Yes	French
France	Le Figaro	@Le_Figaro		French
France	Liberation	@libe		French
France	Le Parisien	@le_Parisien		French
Germany	Handelsblatt	@handelsblatt	Yes	German
Germany	Frankfurter Allgemeine Zeitun	@faznet	Yes	German
Germany	BILD	@BILD		German
Germany	Zeit Online	@zeitonline		German
Hong Kong	South China Morning Post	@SCMPNews	Yes	English
Hong Kong	Hong Kong Free Press	@HongKongFP		English
Hong Kong	RTHK English News	@rthk_enews		English
India	Economic Times	@EconomicTimes	Yes	English
India	Times of India	@timesofindia	Yes	English
India	Hindustan Times	@htTweets	Yes	English
India	The Hindu	@the_hindu	Yes	English
Italy	Corriere Della Sera	@Corriere	Yes	Italian
Italy	La Repubblica	@repubblica	Yes	Italian
Italy	Il Sole 24 ORE	@sole24ore		Italian
Japan	Asahi Shimbun AJW	@AJWasahi	Yes	English
Japan	The Japan News by Yomiuri	@The_Japan_News	Yes	English
Japan	The Japan Times	@japantimes		English
Japan	Japan Today News	@JapanToday		English
Korea	Korea JoongAng Daily	@JoongAngDaily		English
Korea	The Korea Herald	@TheKoreaHerald		English
Korea	Yonhap News Agency	@YonhapNews		Korean
Korea	The Korea Times	@koreatimescokr		Korean
Mexico	La Jornada	@lajornadaonline		Spanish
Mexico	Reforma	@Reforma		Spanish
Mexico	El Universal	@El_Universal_Mx		Spanish

(To be continued)

Country	Newspaper	Twitter Account	BBD	Language
Mexico	Milenio	@Milenio		Spanish
New Zealand	The New Zealand Herald	@nzherald		English
New Zealand	The Dominion Post	@DomPost		English
New Zealand	The National Business Review	@TheNBR		English
Spain	EL MUNDO	@elmundoes	Yes	Spanish
Spain	EL PAIS	@el_pais	Yes	Spanish
Spain	ABC.es	@abc_es		Spanish
Spain	La Vanguardia	@LaVanguardia		Spanish
Switzerland	Neue Zurcher Zeitung	@NZZ		German
Switzerland	20 Minuten	@20min		German
Switzerland	24heures	@24heuresch		French
Switzerland	Le Temps	@LeTemps		French
USA	LA Times	@latimes	Yes	English
USA	USA Today	@USATODAY	Yes	English
USA	Chicago Tribune	@chicagotribune	Yes	English
USA	Washinton Post	@washingtonpost	Yes	English
USA	Boston Globe	@BostonGlobe	Yes	English
USA	Wall Street Journal	@WSJ	Yes	English
USA	Miami Herald	@MiamiHerald	Yes	English
USA	Dallas Morning News	@dallasnews	Yes	English
USA	Houston Chronicle	@HoustonChron	Yes	English
USA	San Fransisco Chronicle	@sfchronicle	Yes	English
USA	New York Times	@nytimes	Yes	English
UK	The Times	@thetimes	Yes	English
UK	Financial Times	@FinancialTimes	Yes	English
UK	BBC News (UK)	@BBCNews		English
UK	Guardian news	@guardiannews		English

Notes: This table reports our newspaper sources. For each newspaper, we specify headquarter location, original language, and twitter account. A 'Yes' under the column BBD denotes a newspaper used also in Baker et al. (2016).

Table .2. COMPUTING TONE OF TWEETS: TWO EXAMPLES

Tweet Text	Negative Words	Positive Words	Tone
The coronavirus pandemic has been particularly devastating to the United States's biggest cities. It comes as the country's major urban centers were already losing their appeal for many Americans.	"devastating", "losing"	"appeal"	$\frac{1-2}{3} = -0.33$
A shortage of test kits and technical flaws in the U.S. significantly delayed widespread coronavirus testing. This is how testing has increased since the beginning of March — and how far it still needs to go, according to the Harvard estimates	"shortag", "flaws", "delayed"		$\frac{-3}{3} = -1$

Notes: This table shows two examples of the computation of the tone of a tweet using Polyglot.

Table .3. DATA SOURCES

Country	Equity Index	Equity Volume Index	Long Term Bond Index	Sovereign CDS	Short Term Bond Index	Currency
Canada	SPTSX Composite Index	TSXVOL Index	GCAN10YR INDEX	CAGV5YUSAC	CA 3M benchmark rate	USDCAD
China	SHSZ300 INDEX	SHSZ300V INDEX	GCNY10YR INDEX	CNGV5YUSAC	CN 1Y benchmark rate	USDCNY
France	CAC Index	CACVOLC Index	GECU10YR INDEX	FRGV5YUSAC	FR 3M benchmark rate	EURUSD
Germany	DAX Index	DAXVOLC Index	GDBR10 INDEX	DEGV5YUSA	DE 3M benchmark rate	EURUSD
Hong Kong	HSI INDEX	HSIVOLC INDEX	HKGG10Y Index	HKGV5YUSAC	HK 3M benchmark rate	USDHKD
Italy	FTSE MIB Index	FTMIBVOL Index	GBTPGR10 INDEX	ITGV5YUSAC	IT 3M benchmark rate	EURUSD
India	SENSEX INDEX	SNSXVOLC INDEX	GIND10YR INDEX	INGV5YUSAC	ES 3M benchmark rate	USDINR
Japan	NKY INDEX	NKYVOLC INDEX	GJGB10 INDEX	JPGV5YUSAC	JP 3M benchmark rate	USDJPY
Korea	KOSPI Index	KOSPIVOLC INDEX	GVSK10YR INDEX	KRGV5YUSAC	KR 1Y benchmark rate	USDKRW
New Zealand	NZSE50FG INDEX	NZ50VOL Index	GNZGB10 INDEX	NZGV5YUSAQ	NZ 3M benchmark rate	NZDUSD
Spain	IBEX 35	IBEXVOLC INDEX	GSPG10YR INDEX	ESGV5YUSAC	ES 3M benchmark rate	EURUSD
Switzerland	SMI Index	SMIVOLC Index	GSWISS10 INDE	CHGV5YUSAC	CH 3M benchmark rate	USDCHF
Sweden	OMXS30 Index	OMXVOLC Index	GSGB10YR INDEX	SEGV5YUSAC	SE 3M benchmark rate	USDSEK
USA	SPX Index	SPXVOLC Index	USGG10YR INDEX	USGV5YEUAC	US 3M benchmark rate	USD
UK	UKX INDEX	UKXVOLC INDEX	GUKG10 INDEX	GBGV5YUSAC	GB 3M benchmark rate	GBPUSD
Source	Bloomberg	Bloomberg	Bloomberg	Thomson Reuters	Bloomberg	Bloomberg
Frequency	Minute	Minute	Minute	Day	Minute	Minute

Notes: This table shows our data sources.

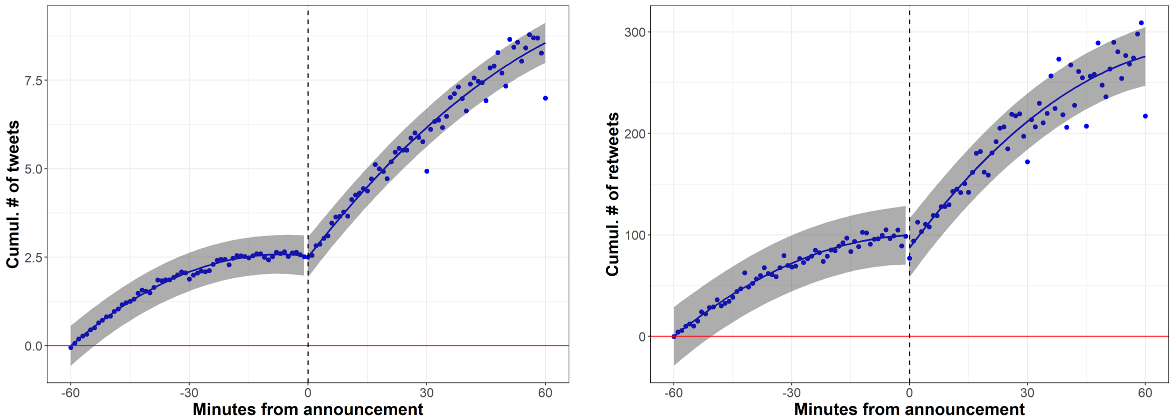
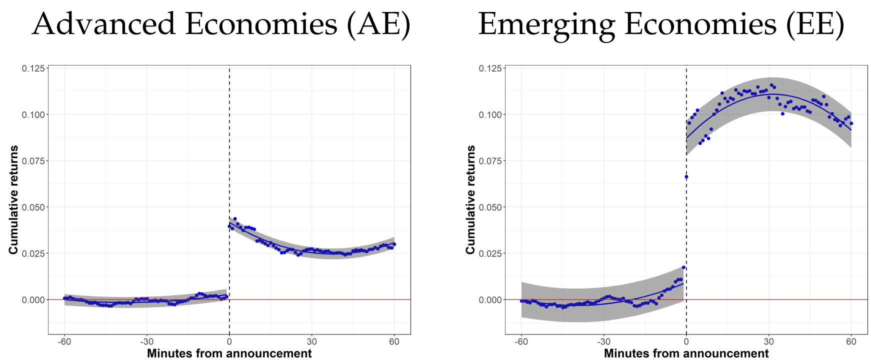
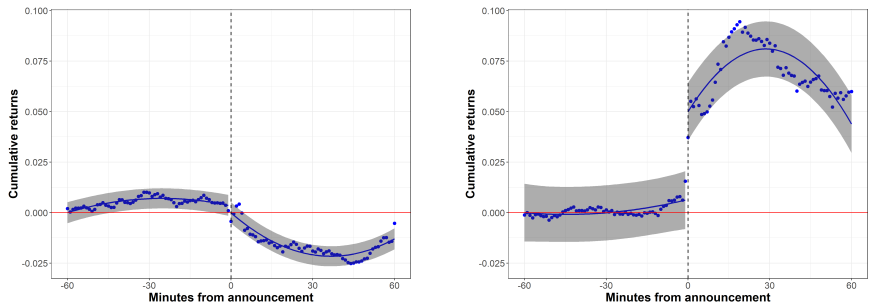


Figure .1. INFORMATION DIFFUSION AND ATTENTION AROUND ANNOUNCEMENTS

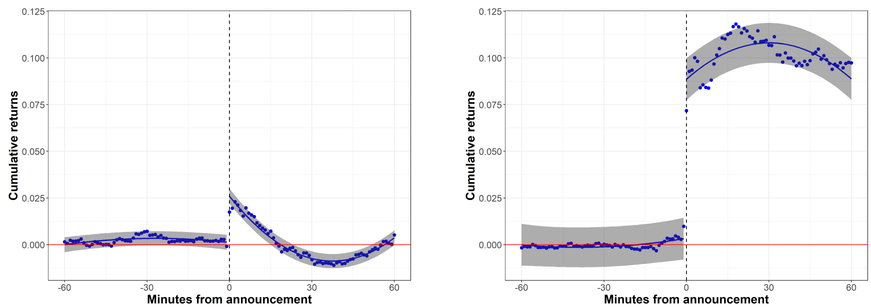
Notes: The left (right) panel of this figure shows the average per-minute and per-country number of tweets (retweets) around announcement times in the post-epidemic period. In each country, the epidemic period starts when there are more than 100 cases of COVID19. Solid line and shaded areas are based on the estimation of equation (2.1). The sample starts on October 1st 2019 and ends on the date of this draft.



(a) All Countries



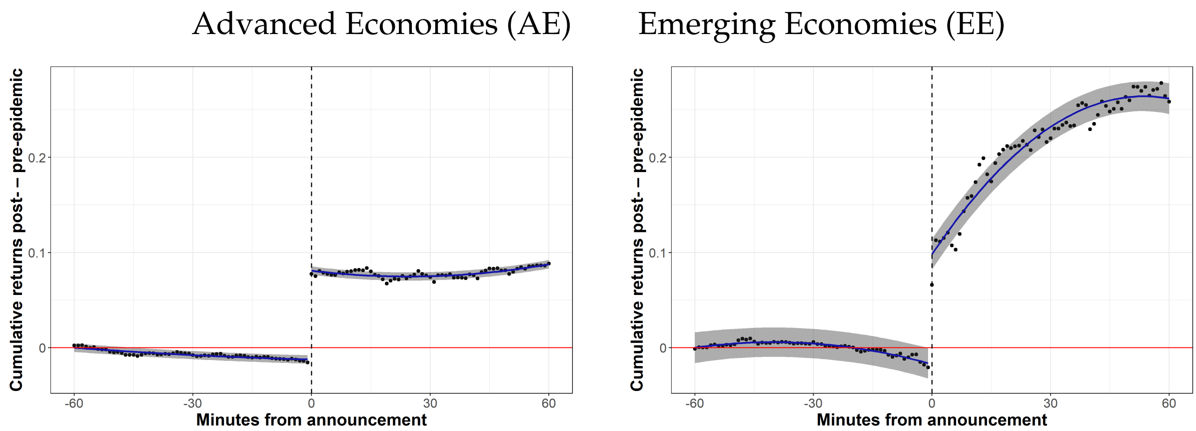
(b) Only Bad News



(c) High-COVID Countries

Figure .2. EQUITY RETURNS AROUND ANNOUNCEMENTS (POST-EPIDEMIC)

Notes: In each panel, dots denote the cross-country-cross-announcement average cumulative returns obtained from buying equities 60 minutes before an announcement and holding them for 120 minutes. Panel a (c) comprises announcements from all countries (top-50% countries in terms of contagion cases) in each group. Panel b excludes announcements conveying good news. Returns are in log units and multiplied by 100. Solid line and shaded areas are based on the estimation of equation (2.1). Our sample starts on October 1st 2019 and ends on the date of this draft. We consider only post-epidemic observations.



(a) All Countries

Figure .3. EQUITY RETURNS AROUND ANNOUNCEMENTS: NO MACRO-NEWS

Notes: In each panel, dots denote the cross-country-cross-announcement average cumulative returns obtained from buying equities 60 minutes before an announcement and holding them for 120 minutes. Panel a comprises announcements from all countries in each group. Returns are in log units and multiplied by 100. Solid line and shaded areas are based on the estimation of equation (2.1). Our sample starts on October 1st 2019 and ends on the date of this draft. We exclude days in which new data for inflation, GDP and industrial productions have been released.

Table .4. TONE OF TWEETS AND RETWEETS

Region	Correlation
AE	0.210***
	(0.035)
EE	0.290***
	(0.033)
World	0.334***
	(0.037)

Notes: this table reports the correlation between our benchmark tone indicator (computed using all of our tweets) and the tone of the quote (re)tweets associated to the top-1% of our tweets ranked by their retweets within each one of our countries. Both tone measures are smoothed by using a 5-day moving average. Our sample starts on October 1st 2019 and ends on the date of this draft.

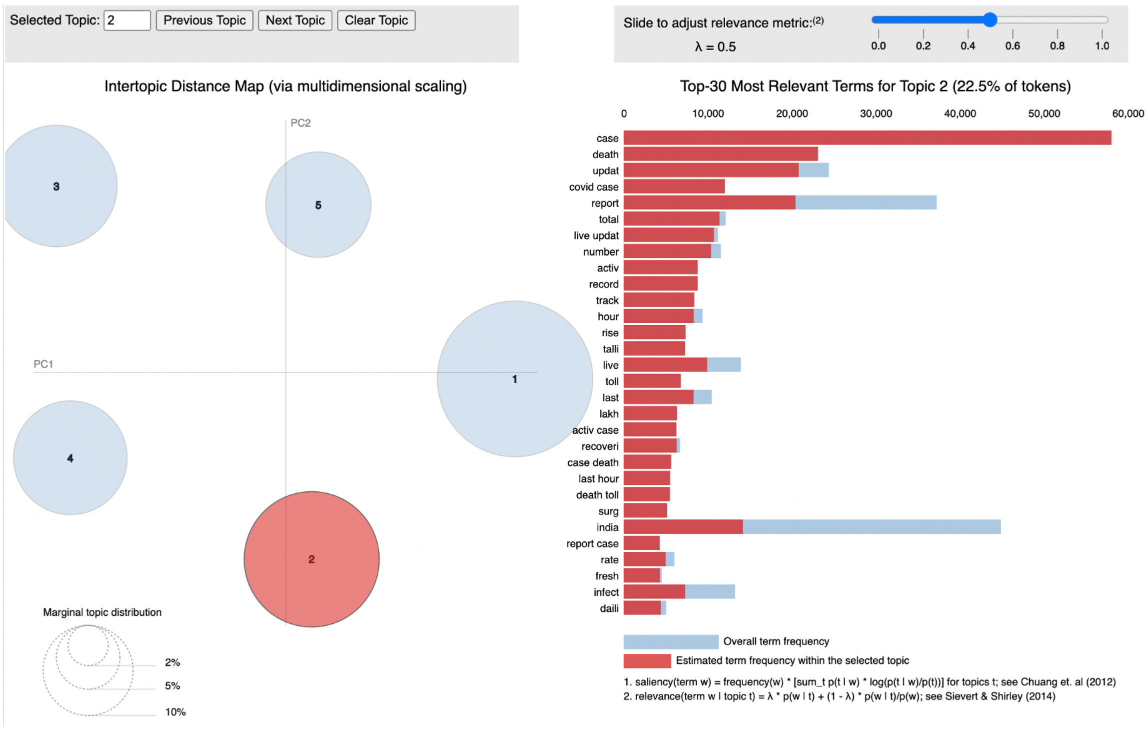


Figure .4. LDA MODEL APPLIED TO INDIA

Notes: This figure shows the output of a Latent Dirichlet Allocation (LDA) topic model applied to our covid-related tweets written in English by newspapers based in India. Results for other countries can be found at https://sites.google.com/view/when-markets-get-covid/. On the left-hand side of the figure, the topics' prevalence is represented by the area of the circles and the center of each circle relates to the distance between topics. We set the number of topics to 5 and show the results for topic 2 (highlighted in red). On the right-hand side, the bars represent the unigrams and bigrams that are most useful for interpreting the selected topic.

Additional Estimation Results

Table .1. SUMMARY OF MPR ESTIMATION: CONDITIONAL CAPM

	Equity		Bonds & Equity	
	A.E.	E.E.	A.E.	E.E.
Local units				
coef	-0.009^{**}	0.003	-0.016^{***}	0.003
se	(0.004)	(0.005)	(0.004)	(0.002)
USD units				
coef	-0.012^{**}	0.007	-0.017^{***}	0.002
se	(0.005)	(0.005)	(0.003)	(0.003)

Notes: This table shows the results of the conditional linear factor model described in equations (2.2)–(2.4) where the risk factor is measured by the news in the MSCI Global Index. Portfolios are formed on a daily basis according to the relative share of country-specific COVID19 cases measured the day before formation (X_t). Both daily excess returns and market prices of risk are in log units and expressed in USD. Our real-time data range from 2/1/2020 to the date of this manuscript. Estimates and HAC-adjusted standard errors are obtained through GMM.

Table .2. SUMMARY OF MPR ESTIMATION (CONTROLLING FOR FF3/FF5)

	Covid Cases		Twitter News	
	Local	USD	Local	USD
FF3				
coef	-0.006^{***}	-0.008^{***}	0.013^{***}	0.011^{***}
se	(0.002)	(0.002)	(0.002)	(0.002)
FF5				
coef	-0.004^{***}	-0.004^{***}	0.012^{***}	0.010^{***}
se	(0.001)	(0.001)	(0.002)	(0.002)

Notes: This table shows the results of the conditional linear factor model described in equations (2.2)–(2.4) augmented with constant betas for the additional FF3/FF5 international factors for developed countries. Portfolios are formed on a daily basis according to the relative share of country-specific COVID19 cases measured the day before formation (X_t). On the left (right), the COVID19 factor is measured as the news to global COVID cases growth (tone of COVID-related tweets). When we measure the COVID19 news as unexpected number of contagion cases (unexpected improvement in COVID19-related tweets), we expect a negative (positive) market price of risk (MPR). Both daily excess returns and market prices of risk are in log units. Our cross section of test assets comprises both equity and bond portfolios. Our real-time data range from 2/1/2020 to the date of this manuscript. Estimates and HAC-adjusted standard errors are obtained through GMM.

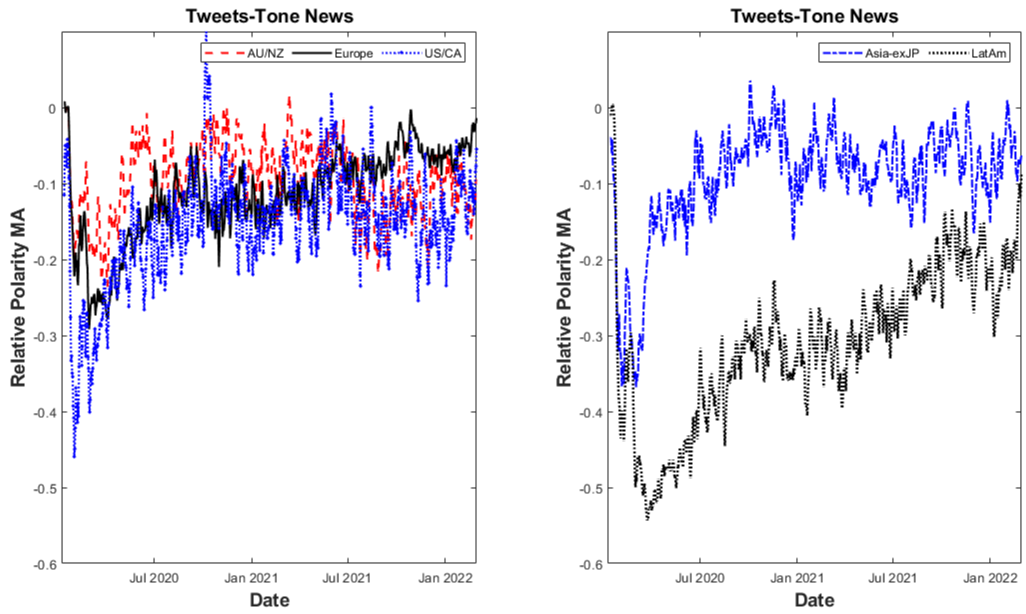


Figure .1. REGIONAL TWITTER-BASED TONE

Notes: This figure shows our daily Twitter-based tone for different countries. We use Polygot to measure the polarity of our tweets and compute the tone of each tweet according to Twedt and Rees (2012). We aggregate the tones at a daily frequency and across regions. MA refers to a backward looking 5-day moving average.

Table .3. SUMMARY OF MPR ESTIMATION: LOCAL NEWS

	Equity		Bonds & Equity	
	A.E.	E.E.	A.E.	E.E.
		Panel A: Local News about Covid cases		
Local units				
coef	-0.004^{**}	0.008^{***}	-0.003^{***}	-0.006^{***}
se	(0.002)	(0.001)	(0.001)	(0.001)
USD units				
coef	-0.005^{**}	0.006^{***}	-0.003^{***}	-0.005^{***}
se	(0.002)	(0.001)	(0.001)	(0.001)
Controlling for MKT				
coef	0.001	0.002	-0.002^{***}	-0.007^{***}
se	(0.001)	(0.002)	(0.001)	(0.001)
		Panel B: Local News from Twitter		
Local units				
coef	0.029	-3.076	0.012^{***}	0.007^{***}
se	(0.028)	(12.864)	(0.004)	(0.001)
USD units				
coef	0.035	-6.337	0.011^{***}	0.006^{***}
se	(0.033)	(84.315)	(0.003)	(0.001)
Controlling for MKT				
coef	-0.000	0.005^{*}	0.008^{***}	0.009^{***}
se	(0.004)	(0.002)	(0.002)	(0.001)

Notes: This table shows the results of the conditional linear factor model described in equations (2.2)–(2.4) applied to AE- and EE-specific news. Portfolios are formed on a daily basis according to the relative share of country-specific COVID19 cases measured the day before formation (X_t). In panel A (panel B), the COVID19 factor is measured as the news to local COVID cases growth (tone of COVID-related tweets). When we measure the COVID19 news as unexpected number of contagion cases (unexpected improvement in COVID19-related tweets), we expect a negative (positive) market price of risk (MPR). Both daily excess returns and market prices of risk are in log units. The last two columns are based on a broader cross section of test assets comprising both equity and bond portfolios. When we control for the market, returns are in USD, the market is measured by the MSCI Global Index, and our factor model comprises a total of two factors. Our real-time data range from 2/1/2020 to the date of this manuscript. Estimates and HAC-adjusted standard errors are obtained through GMM.

Table .4. SUMMARY OF MPR ESTIMATION (CONTROLLING FOR POPULATION)

	$\hat{X}^i_t = X^i_t \frac{Pop^i}{Pop^W}$		$\hat{X}^i_t = X^i_t \frac{Pop^W}{Pop^i}$	
	MPR	N.Obs	MPR	N.Obs
Local units	0.020***	4190	0.010***	4187
	(0.003)	4190	(0.003)	4187
EUR units	0.018***	4190	0.029***	4187
	(0.003)	4190	(0.002)	4187
Controlling for MKT	0.013***	3951	0.014***	3949
	(0.003)	3951	(0.003)	3949

Notes: This table shows the results of the conditional linear factor model described in equations (2.2)–(2.4) in which the share of covid cases is either multiplied (left) or divided (right) by the relative population of each country. Portfolios are formed on a daily basis according to $\hat{X}_t$ measured the day before formation. The COVID19 factor is measured as the news to the tone of COVID-related tweets. We expect a positive market price of risk (*MPR*). Both hourly excess returns and market prices of risk are in log units. Our cross section of test assets comprises both equity and bond portfolios. Our real-time data range from 2/1/2020 to the date of this manuscript. Estimates and HAC-adjusted standard errors are obtained through GMM.

Table .5. HOURLY CONDITIONAL LINEAR FACTOR MODEL (II)

	MPR	N.Obs
Hourly log returns		
coef	0.033***	4190
se	(0.003)	4190
Hourly log EUR returns (adjusting for FX)		
coef	0.015***	4190
se	(0.003)	4190
Hourly log returns controlling for the Market		
coef	0.041***	3951
se	(0.005)	3951

Notes: This table shows the results of the conditional linear factor model described in equations (2.2)–(2.4) with β_0 replaced by β_0^f. Portfolios are formed on a daily basis according to the relative share of country-specific COVID19 cases measured the day before formation (X_t). The coefficient $\beta_{f,t} = \beta_0^f + \beta_f X_{f,t}$ refers to the exposure of the equity portfolio $f \in \{H, M, L\}$ to the COVID19 factor. We measure hourly COVID19 news as unexpected improvement in the hourly tone of COVID19-related tweets. Both hourly excess returns and market prices of risk are in log units. Our real-time data range from February 2020 to the date of this manuscript. Estimates and HAC-adjusted standard errors are obtained through GMM.

Chapter 3

Transition risk and imperfect competition: evidence from a structural model of the Italian credit market

Abstract

I develop a structural model of loan demand and lender competition to study how transition risk may affect the Italian credit market. First, I show that transition risk is not currently priced by banks, nor that firms likely more exposed to this risk tend to default more frequently. Then, I use the estimated model to study the effect of policies aimed at more tightly integrating climate-related and environmental risks into banks' business planning. Modeling any such policy as in increase in the cost of lending to "brown" firms, counterfactual analyses show that if these marginal costs were to increase by one standard deviation interest rates would on average increase by 130 basis points, while quantities would decrease by about 20k EUR.[‡]

1 Introduction

Climate-related financial risks are now part of the agenda of central bankers and financial regulators around the world (Bolton et al. (2020); ECB (2020)). There is growing evidence that the transition to a low-carbon economy may entail severe losses from both material physical risk (Hsiang et al. (2017); Meucci and Rinaldi (2022)) and tran-

[‡]The views expressed in the articles are those of the author and do not involve the responsibility of the Bank of Italy

sition risk (Faiella and Lavecchia, 2020). Physical risk refers to the financial impact of a changing climate, including more frequent extreme weather events, as well as of environmental degradation; transition risk refers to an institution's financial loss that can result, directly or indirectly, from the process of disorderly adjustment towards a lower-carbon and more environmentally sustainable economy.

While there is a large consensus that some combination of physical and transition risk *will* materialize in the future (NFGS, 2019), there is a high degree of uncertainty on the timing and nature of impacts of climate change. In this context, traditional approaches to risk management consisting in extrapolating historical data, which largely does not reflect the chance of occurrence of extreme phenomena (Weitzman, 2009), are not well-suited to assess future climate-related risks. The failure of traditional risk assessment tools may in turn hinder the transition to a low-carbon economy since financial institutions could be overly exposed to markets, sectors or geographic areas that may be particularly affected by physical and transition risks. This issue is likely more severe for retail credit markets for which reliable data on greenhouse gas (GHG) emissions is particularly scarce. For these reasons, banks are "expected to integrate climate-related and environmental risks that impact their business environment in the short, medium or long term" (ECB, 2020).

In this paper, I develop a structural model of the Italian credit market to quantify the cost of policies aimed at incorporating transition risk in financial institutions' business plans and their equilibrium impact on interest rates and the quantity of credit to non-financial firms. I try to bridge the data gap on GHG emissions by relying on the recently approved Taxonomy Regulation and its Delegated Acts (EU Parliament (2020); EU Commission (2021b); EU Commission (2022)). Compared to the multitude of data-providers of GHG emissions, the EU Taxonomy *(i)* establishes a core set of homogeneous principles applicable to the entire EU to classify an economic activity as (taxonomy-)eligible or aligned, and *(ii)* by building on a granular sectoral approach overcomes the scarcity of information that often hampers a firm-level approach. Activities that are taxonomy-eligible/aligned should be *less* exposed to transition risk.

As a first step, I provide descriptive evidence on the distribution of eligible and aligned retail firms geographically and across economic sectors, where I identify a firm with its main economic activity[1]. According to EU Commission (2021b) an economic activity is "eligible" if it belongs to one of the NACE sectors (2d, 3d or 4d) that satisfy one of the six environmental objectives (currently, only climate mitigation or climate adaption) set forward in the Taxonomy Regulation; an economic activity is "aligned" if it is eligible and satisfy additional technical screening criteria. Since it is currently not yet feasible to assess whether an eligible firm complies with the aforementioned technical criteria, I use the Taxonomy Alignment Tool of Alessi et al. (2019) to estimate the probability of alignment.

Then, I show that the distribution of risk and realized firms' defaults in the Italian credit market do not differentiate between aligned/eligible firms and non-aligned/non-eligible ones. This evidence is consistent with recent findings from EBA (2021) and ECB (2020) which show that most of the institutions do not have the tools to assess the impact of climate-related and environmental risks on their balance sheets. When looking at actual credit market outcomes in terms of pricing and quantities, I also do not find any statistically significant difference for aligned/eligible firms.

As a next step, I build a structural model of loan demand and loan supply along the lines of Crawford et al. (2018). I estimate the model on detailed microdata starting from 2018 and covering individual loans between firms and banks. I use two main datasets. First, I use the Italian section of the Analytic Credit Registry (AnaCredit), which provides detailed information on all individual loans extended by the largest Italian banks, including the identity of the borrower, interest rate charged and an assessment of borrowers' creditworthiness performed by each bank and measured as a 1-year probability of default (PD). Second, I use the Cerved/Cebil database which provides granular information on borrowers' balance sheets and income statements. Using the estimated model I recover the *unobservable* firm-specific marginal cost of lending that banks face. Similarly to the distribution of risk and realized defaults, and further strengthening this evidence, I find no statistical difference in the cost of lending to firms that are eligible and to firms that are not.

[1]This assumption should be innocuous since retail firms typically are active in a single, clearly identifiable sector.

To estimate the effect of a tighter integration of climate-related risks into banks' business planning I run two counterfactuals. In the first experiment, I increase in the cost of lending to firms that are *not* eligible and thus are likely more exposed to transition risk; in the second experiment I increase the cost of lending to firms that are *not* aligned according to the Taxonomy Alignment Tool of Alessi et al. (2019). The channel I have in mind is that eventually, perhaps through increased regulatory scrutiny (ECB, 2020), lending to firms more exposed to climate-related risks will become *more expensive* because banks will have to put aside more capital to cover *unexpected* losses. Alternative ways to model transition risk could be through either *(i)* an increase of (non-aligned/non-eligible) borrowers' probability of default (PD), or *(ii)* through "green capital requirements", i.e. a reduction (or an increase) in capital requirements when lending to "green" ("brown") firms (Oehmke and Opp, 2022). Regarding an increase in the PDs, given the high uncertainty on the timing and nature of transition risk, it is not clear how to incorporate climate considerations in a 1-year probability of default estimated on past data. On the other hand, green capital requirements are a highly controversial topic since their introduction would require a substantial revision of the Basel Accords. Hence, an increase in marginal costs for non-aligned/non-eligible firms can be seen as just a reduced-form way of capturing much more flexible supervisory instruments (ECB, 2020) that act on the governance, risk appetite and risk management.

The main results of the counterfactual exercises are the following: if the marginal cost of lending to non-aligned (non-eligible) firms were to increase by a standard deviation, I estimate that on average interest rates for these firms would increase by 134 (71) basis points, and that quantities would decrease by about 20k (13k) EUR. I stress that an increase of marginal costs by one standard deviation is just *one* of the possible scenarios. Other scenarios of different severity could be calibrated based e.g. on NFGS data[2]. If and when more granular data on emissions and alignments becomes available, the methodology is also flexible enough to incorporate firm-specific variation in marginal costs.

All in all, my results suggest that policies aimed at more tightly integrating tran-

[2] https://www.ngfs.net/ngfs-scenarios-portal/

sition risks into banks' business planning could have a material impact in the Italian credit market.

2 Data and Institutional Details

2.1 EU Taxonomy

My main source of data to assess the exposure of an economic activity to transition risk comes from Regulation (EU) 2020/852 (the Taxonomy Regulation; EU Parliament (2020)). This regulation aims at defining environmentally sustainable activities on the basis of technical screening criteria set out in specific delegated acts. The first act was adopted on 4 June 2021 (EU Commission, 2021a) and dealt with criteria for economic activities that make a substantial contribution to climate change mitigation and adaptation; the remaining delegated regulations concerning the criteria for the last four environmental objectives will be adopted at a later stage.

EU Commission (2021b) defines an activity as "taxonomy-eligible" (hereafter, "eligible") if it is included in the delegated acts of the European Commission issued pursuant to Articles 10(3), 11(3), 12(2), 13(2), 14(2), and 15(2) of the Taxonomy Regulation, irrespective of whether that economic activity meets any or all of the technical screening criteria laid down in those delegated acts. An activity is "taxonomy-aligned" (hereafter, "aligned") when, in addition to being included in the delegated acts above, it complies with all the technical criteria set out in these delegated acts[3].

In this paper, I classify a (non-financial) firm as eligible if its main economic activity is eligible according to EU Commission (2021a). This categorization could be biased for large firms that are active in multiple sectors but these companies are subject to different reporting requirements that could be exploited (see next paragraph). On average, about 37% of firms are eligible on an equally weighted basis and about 47%

[3] For example, "manufacture of aluminium", that could be associated with NACE code C24.42 or C24.53, is an eligible activity since it is included in a delegate act (EU Commission, 2021a). However, the activity is "aligned" if either *(a)* it manufactures secondary aluminium or *(b)* manufactures primary aluminium complying (from 2025) with all of the following criteria: *(i)* the greenhouse gas (GHG) emissions do not exceed 1.484 tCO_2e per ton of aluminium manufactured; *(ii)* the average carbon intensity for the indirect GHG emissions does not exceed 100 g CO_2e/kWh; *(iii)* the electricity consumption for the manufacturing process does not exceed 15.5 MWh/t Al.

Table 3.1. Distribution of eligible firms

	No. of firms (%)	Total # firms	Assets (%)	Total asset (bln EUR)
A	1.59%	61921	1.00%	23.85
B	0.00%	1057	0.00%	1.11
C	57.21%	139838	56.46%	253.48
D	86.21%	3424	81.63%	20.43
E	89.31%	4155	95.26%	9.7
F	100.00%	103106	100.00%	55.21
G	5.73%	218409	8.45%	125.09
H	85.20%	30794	79.36%	33.65
I	0.00%	75679	0.00%	19.15
J	78.43%	18922	74.08%	18.39
K	0.00%	4140	0.00%	0.41
L	100.00%	49094	100.00%	44.96
M	23.80%	55599	6.77%	54.78
N	10.18%	28538	30.28%	22.84
O	0.00%	64	0.00%	0.18
P	0.00%	3813	0.00%	1.34
Q	0.00%	19816	0.00%	8.48
R	0.00%	10247	0.00%	2.73
S	2.55%	19269	1.58%	2.68
all	36.57%	847885	47.39%	698.44

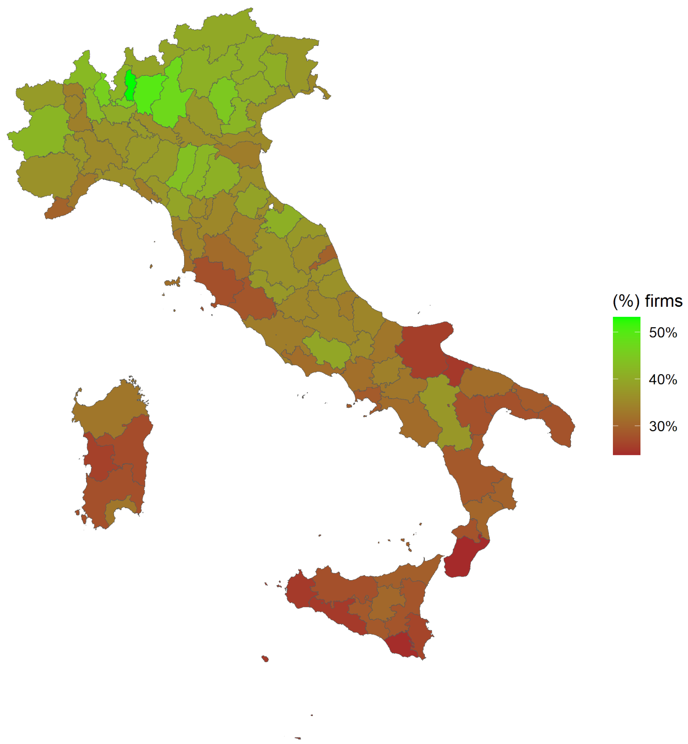
Notes: This table reports the percentage of firms, equally-weighted and value-weighted, that are eligible according to the EU Taxonomy aggregated at the NACE 1d level.

on a value-weighted basis. This evidence is broadly in line with the findings of the EBA pilot exercise on climate risk (EBA, 2021) that found about 42% of *large* corporate exposures to EU obligors in sectors that might be sensitive to transition risk. Table 3.1 shows complete summary statistics by NACE sector[4]. Figure 3.1 plots the geographical dispersion of eligible firms; the south of Italy appears to be the region with the lowest concentration of eligible firms.

With regards to disclosure requirements, according to Article 8(1) of the Taxonomy Regulation only large companies which are required to publish non-financial information under the Non-Financial Reporting Directive (NFRD, EU Parliament and EU Council (2014)) shall include in their non-financial statement further information on

[4]For the complete list of NACE 1d sectors see Table 3.13 in the Appendix

Figure 3.1. GEOGRAPHICAL DISTRIBUTION OF ELIGIBLE FIRMS



Notes: This figure plots the percentage of firms in each province that are eligible according to the EU Taxonomy.

how and to what extent they are associated with environmentally sustainable economic activities. There are no *comparable reporting requirements for smaller firms*. To give a rough estimate of the coverage of the NFRD, in Italy only 219 firms, of which financial institutions account for about 25%, disclosed some non-financial information according to the NFRD for the reference year 2021[5]. Moreover, until 2022 large firms could report vastly heterogeneous climate-related metrics since a specific reporting obligation did not yet exist.

Indeed, to operationalize the EU Taxonomy, starting from 2023 large non-financial firms under the NFRD shall disclose several standardized metrics: the proportion of taxonomy-eligible and taxonomy non-eligible economic activities in their total turnover, capital and operational expenditure (Capex and Opex), and related qualitative information (EU Commission, 2021b). With regards to credit institutions, starting from Jan-

[5]Source: https://www.osservatoriodnf.it/en/home/.

uary 2024 banks shall disclose the proportion of their assets financing and invested in taxonomy-aligned economic activities as a proportion of total assets (so called “Green Asset Ratio”, (GAR)); the percentage of financial guarantees supporting debt instruments financing taxonomy-aligned economic activities; the assets under management from undertakings financing taxonomy-aligned economic activities in percentage of their total assets under management. When available, researchers could use the GAR to indirectly assess the exposure of financial institutions to transition risk. However, since no specific reporting requirements exist for small and medium-sized enterprises (SMEs), which constitute a very sizable portion of loan books, it will still be hard to conduct scenario analysis and run counterfactuals.

To assess existing and newly developed climate risk assessment and classification, in 2021 the European Banking Authority (EBA) has launched a pilot exercise on climate risk (EBA, 2021). The pilot was run on a sample of 29 volunteer banks, which provided raw data on *non-SME corporate exposures* to EU countries for which some data on GHG emissions was more readily available. Banks were also asked to participate in a questionnaire on the application of the EU taxonomy. The exercise mapped exposures in sectors more exposed to transition risk using a sectoral based approach or using GHG emissions[6]. To classify an activity as (taxonomy-)aligned the EBA used the Taxonomy Alignment Tool by Alessi et al. (2019).

The main findings of the exercise were that the classification methods based on the sectors of the counterparty allow higher coverage and to compute estimates more easily with current available information. However, measuring the climate relevance of a counterparty based on its main activity does not give a comprehensive and precise picture of its level of environmental sustainability. This was confirmed by the presence of relatively high emitters in sectors considered to be not relevant from a climate perspective. On the other hand, using methods that rely on the carbon emission of the counterparty allows banks’ environmental profiles to be analysed more accurately but at a cost of a much reduced coverage and consistency of the data. The ECB climate stress test run in 2022 (ECB, 2022) further reinforced this evidence.

Given the outlined constraints, in this paper I follow the approach of EBA (2021)

[6]The source used for GHG emission is Trucost (S&P Global): `https://www.trucost.com`.

Table 3.3. Distribution of aligned firms

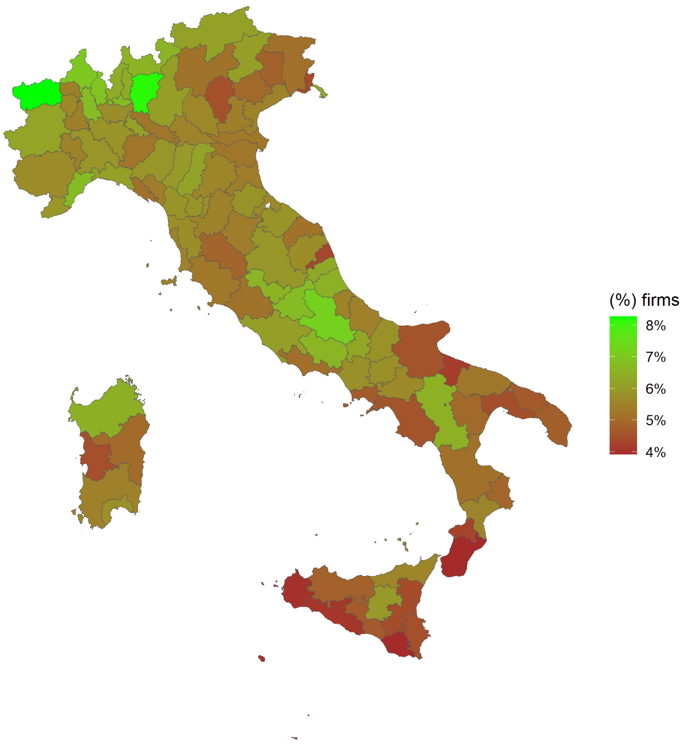
	No. of firms (%)	Total # firms	Assets (%)	Total asset (bln EUR)
A	0.00%	61921	0.00%	23.85
B	0.00%	1057	0.00%	1.11
C	0.06%	139838	0.26%	253.48
D	17.80%	3424	32.65%	20.43
E	0.00%	4155	0.00%	9.7
F	38.68%	103106	38.44%	55.21
G	0.10%	218409	0.15%	125.09
H	0.86%	30794	5.01%	33.65
I	0.00%	75679	0.00%	19.15
J	0.00%	18922	0.00%	18.39
K	0.00%	4140	0.00%	0.41
L	15.00%	49094	15.00%	44.96
M	0.00%	55599	0.00%	54.78
N	0.06%	28538	0.46%	22.84
O	0.00%	64	0.00%	0.18
P	0.00%	3813	0.00%	1.34
Q	0.00%	19816	0.00%	8.48
R	0.00%	10247	0.00%	2.73
S	0.00%	19269	0.00%	2.68
all	5.71%	847885	5.34%	698.44

Notes: This table reports the percentage of firms, equally-weighted and value-weighted, that are aligned according to the Taxonomy Alignment Tool of Alessi et al. (2019) aggregated at the NACE 1d level.

and estimate the probability that a firm is aligned using the Taxonomy Alignment Tool of Alessi et al. (2019). Using a granularity up to NACE 4 digits, the tool estimates for each (sub-)sector that contributes to the objective of climate change mitigation a Taxonomy Alignment Coefficient (TAC) that represents the share of activities that satisfy the relevant technical screening criteria and characteristics of the sector as a whole. For each firm I therefore define the probability of alignment as the TAC of its main activity at the highest level of granularity.

On average, the probability of alignment (or, alternatively, the average share of aligned firms) is about 6% on an equally weighted basis and 5% on a value-weighted basis. These findings are consistent with the evidence from (EBA, 2021) that estimates an aggregate EU green asset ratio of about 8%. Table 3.3 shows complete summary statistics by NACE sector[7]. Figure 3.2 plots the geographical dispersion of aligned firms; as with eligible firms the south of Italy appears to be the region with the lowest concentration of aligned firms.

Figure 3.2. GEOGRAPHICAL DISTRIBUTION OF ALIGNED FIRMS



Notes: This figure plots the percentage of firms in each province that are aligned according to the Taxonom Alignment Tool of Alessi et al. (2019).

2.2 Loan Data

The main dataset used in this project is the Italian section of AnaCredit. AnaCredit is a new European credit registry centrally managed by the ECB with the aim of collecting detailed information on individual bank loans from euro area banks for supervisory and research purposes. In AnaCredit banks report outstanding amounts, interest rates types, periodicity, guarantees, default information, etc. for all credit facilities granted to all legal entities in their credit portfolio. The reporting frequency is monthly and is

[7]For the complete list of NACE 1d sectors see Table 3.13 in the Appendix

Table 3.5. SUMMARY STATISTICS

	N	avg	p25	p50	p75
Panel A: Loan level					
loan size ('000)	841380	260.1	25	34	150
r (%)	841380	2.41	1.06	1.66	3.38
probability of default (PD, %)	841380	3.22	0.46	1.12	3.13
Loan-to-value (LTV)	841380	0.76	0.56	0.87	1
maturity (years)	841380	6.81	5.01	6.01	7.07
Panel B: Firm level					
age	603019	16.87	7	13	23
D/E	603019	12.12	1.8	4.01	9.35
log(asset)	603019	6.78	5.72	6.66	7.68
log(sales)	603019	6.45	5.41	6.42	7.47
Panel C: Bank level					
log(assets)	32	25.89	25.23	25.62	26.35
cet1 ratio	32	0.14	0.12	0.14	0.15
cost-income	32	0.72	0.67	0.72	0.78
leverage	32	16.39	14.17	15.21	17.96
loans/deposits	32	0.81	0.71	0.8	0.91
npl ratio	32	0.06	0.04	0.05	0.08
roe	32	0.05	0.04	0.05	0.08

Notes: This table reports summary statistics for the raw data. The sample runs from 2018 Q3 to 2022 Q2 at quarterly frequency.

subject to a minimum materiality threshold of EUR 25,000 computed across all exposures for of any given single borrower. The first available reporting date is June 2018. Banks that use the Internal Ratings-Based approach (IRB - Basel Committee (2001)) also report their estimation of borrowers' one-year probability of default (PD).

My sample comprises all loans to non-financial corporations for which IRB banks report a valid PD, and ranges from September 2018 to June 2022 at quarterly frequency. While my time series dimension is limited, I have available a fairly large cross-section of borrowers from the largest Italian IRB banks. These lenders account for around 80% of the Italian banking sector in terms of total assets.

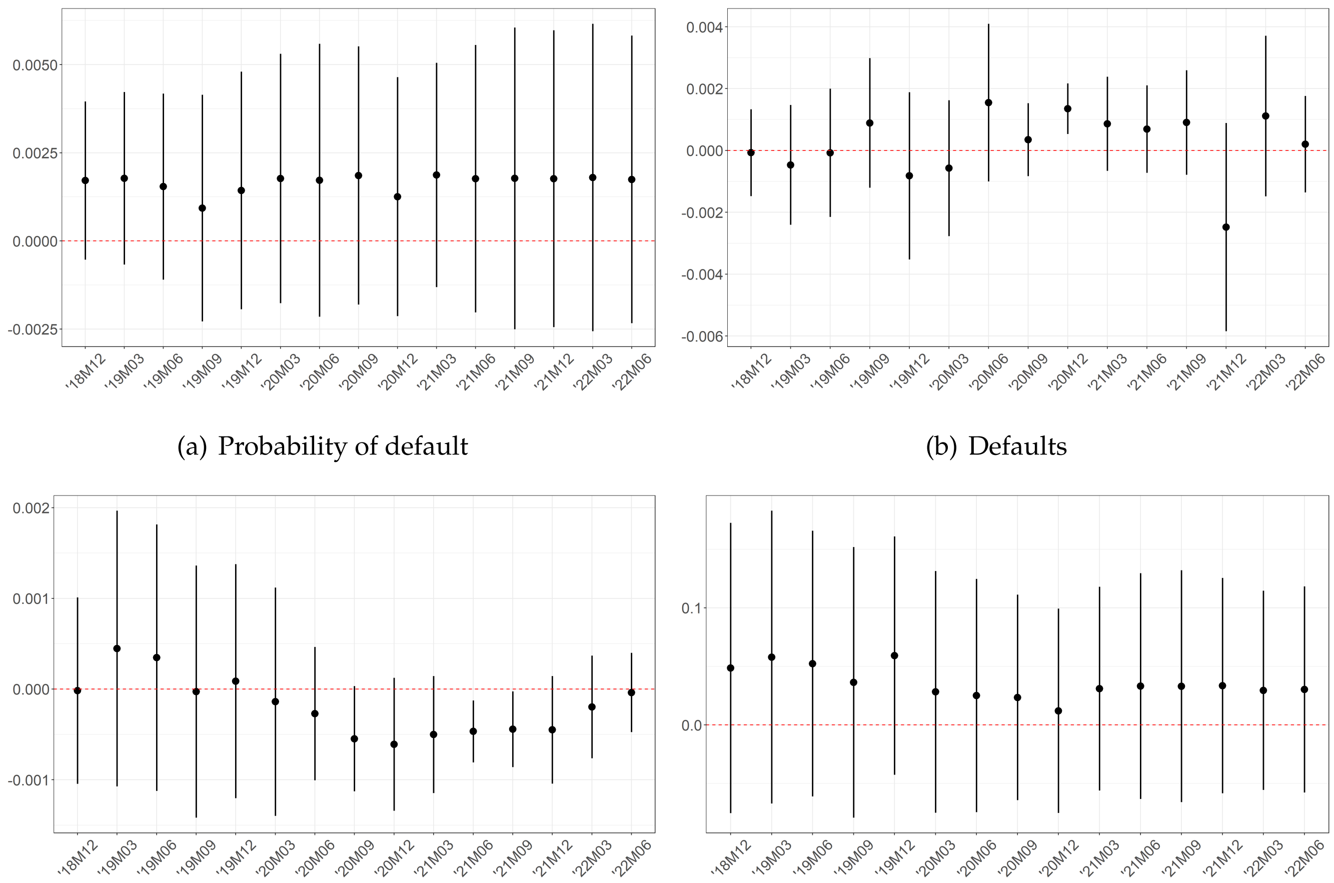
Firm balance sheet and income statement data comes the Cerved/Cebil (CV) dataset. CV collects yearly data on the balance sheets and income statements of the universe of

Italian limited companies (about 800,000 firms). The information is typically collected and standardized from balance sheets deposited with local chambers of commerce, where limited liability companies are obliged to file. The unique feature of the CV dataset is that, differently from other widely used datasets on individual companies, it has wide coverage of small and medium enterprises, almost all of which are unlisted. I then match the CV database with loan level data from AnaCredit reducing the sample size to about 110,000 distinct firms per year. Table 3.5 reports some summary statistics pooled across quarters.

Using my matched firm-bank-loan dataset in figure 3.3 I test for equality of means of various credit market outcomes between eligible and non-eligible borrowers. My main variable of interests are: firms' probability of default (*panel a)*), actual defaults (*panel b)*), interest rates (*panel c)*), and loan size (*panel d)*). Controlling for various firm-level observables such as size, revenues, debt-equity ratio and including fixed effects at the bank, province and sector level, I find no systematic difference between eligible and non-eligible borrowers at standard confidence levels. I also partition my sample to include only firms belonging to NACE 2d sectors that include both eligible and not eligible NACE 3d or 4d sub-sectors, so that I am effectively comparing firms that belong to the same "industry" and that should be as similar as possible along observable characteristics. These findings are consistent with evidence from EBA (2021) and ECB (2022) which reports that "although some progress has been made ... bank do not yet sufficiently incorporate climate risk into their stress-testing frameworks and internal models."

In the structural estimation I will focus on all newly secured loans with maturity greater than 1 year and without a specific liquidity, import/export, or working capital financing purpose. If a firm has more than one loan outstanding with two or more banks, I restrict the sample to the loan providing the most financing. Finally, I randomly sample a subset of the population (about 100k observations) since estimation would not be computationally feasible otherwise.

Figure 3.3. **CREDIT MARKET OUTCOMES FOR ELIGIBLE VS NOT-ELIGIBLE BORROWERS**



(a) Probability of default

(b) Defaults

(c) Interest rate r

(d) Loan size q

Notes: This figure plots the coefficients $\{\beta_1^t\}_t$ with 95% confidence interval of the regression: $Y_{i,b,t} = \beta_0^t + \beta_1^t Eligib_{i,t} + \Gamma X_{i,t} + FE^t + \epsilon_{i,b,t}$ where $Eligib_{i,t}$ is an indicator variable equal to 1 if firm i at time t belongs to a (sub-)sector eligible according to the EU Taxonomy. The dependent variable $Y_{i,t,b}$ is the probability of default (panel *a)*), an indicator variable equal to 1 if firm i is in default at time t (panel *b)*), the interest rate r (panel *c)*), and size of the loan q (panel *d)*). Controls in the regression are fixed effects FE^t at the bank, province and sector (NACE 2d) level, and a vector of firm-level observables $X_{i,t}$ containing size, revenues, debt-equity ratio etc. My sample includes all firms belonging to NACE 2d sectors that include both eligible and not eligible NACE 3d or 4d sub-sectors. I estimate the regression separately for each available

3 Model

I build a model of imperfect competition following Crawford et al. (2018). My main objective is to recover the unobservable marginal costs of lending and run counterfactuals using the estimated equilibrium conditions. In the model each of the firms $i = 1, 2, \ldots, N_m$ in market m at time t demands credit to finance a (long-term) risky project to banks $j = 1, 2, \ldots, J$ active in market m. Among the products offered by banks, a firm chooses the one that gives the highest "utility", and, conditional on taking a loan, decides how much to borrow. Banks are risk neutral and compete a-là Bertrand-Nash on interest rates. Differently from Crawford et al. (2018), in my setting banks' profit function depends explicitly on the bank-specific assessment of borrowers' likelihood of default that credit institutions directly report in my dataset.

In the model I adopt several important assumptions similar to Crawford et al. (2018): first, I focus only on all newly secured loans with maturity greater than 1 year and without a specific liquidity, import/export, or working capital financing purpose. I do this to avoid the dynamic dimension and reduce the complexity of the problem. If a firm has more than one loan outstanding with two or more banks, I restrict the sample to the loan providing the most financing. Second, I assume both firms and banks are risk-neutral. Third, in the model banks compete only on the interest rate while lenders could offer contracts that depend *both* on the price and on the credit amount to extract some private information from borrowers. However, since in the Italian market firms can (and do) engage in multiple relationship with different lenders, borrowers could obtain the desired funding with multiple loans from different banks. Hence, the assumption that the amount of credit is exogenous (from a bank's perspective) appears plausible.

Finally, motivated by the preliminary evidence from figure 3.3, in the model I do not allow for any differentiation between eligible/aligned firms and non-eligible/non-aligned ones other than what is already encoded in the data e.g., in firms' probability of default as assessed by banks, in the soft information component or in other observable firm characteristics. As mentioned in the introduction, marginal costs for firms likely more exposed to transition risk will ultimately vary only in the counterfactual analysis.

Demand A firm i in market m at time t demands for a loan to bank j according to (3.1). The firm has access to $j = 0, 1, 2, \ldots, J$ banks in each market, with $j = 0$ denoting the outside-option of not borrowing, and selects the product offered by the bank that delivers the highest utility U^D_{ijmt}

$$U^D_{ijmt} = \alpha^D_0 + X^D_{jmt}\beta^D + \alpha^D P_{ijmt} + Y^D_{ijmt}\eta^D + \xi^D_{jmt} + u_i + \epsilon^D_{ijmt} \tag{3.1}$$

where X^D_{jmt} is a vector of bank-market characteristics, P_{ijmt} is the price (interest rate), Y^D_{ijmt} is a vector of non-price firm-bank-market characteristics, and ξ^D_{jmt} represents bank-market unobservables (to the econometrician) that may be correlated with prices. Finally, ϵ^D_{ijmt} is an idiosyncratic shock to firm demand and u_i is the firm propensity to demand a loan that is know to the firm but not to the bank.

Conditional on borrowing, each firm chooses the amount of credit to use to maximize the following utility

$$U^L_{ijmt} = \alpha^L_0 + X^L_{jmt}\beta^L + \alpha^L P_{ijmt} + Y^L_{ijmt}\eta^L + u_i \tag{3.2}$$

where ξ^D_{jmt} represents bank-market unobservables (to the econometrician) that may be correlated with prices, and $u_i \sim N(0, \sigma^2)$ is the firm propensity to demand a loan that is known to the firm but not to the bank.

Supply The supply side is standard. Banks compete à la Bertrand-Nash and for each borrower i maximize the following (expected) profit function

$$\Pi_{ijmt}(MC_{ijmt}) = P_{ijmt}Q_{ijmt}(1 - PD_{ijmt}) - MC_{ijmt}Q_{ijmt} \tag{3.3}$$

where Q_{ijmt} is the expected loan size and is equal to the probability that firm i demands new credit from bank j times the average loan size, $PD_{ijmt} \in [0, 1)$ is the probability that firm i will default as assessed by bank j, and MC_{ijmt} is the marginal cost of lending to firm i by lender j that is unobserved by me as econometrician.

The first order condition of (3.3) reads as

$$
\begin{aligned}
& 0 = P_{ijmt}\frac{\partial Q_{ijmt}}{\partial P_{ijmt}}(1 - PD_{ijmt}) + Q_{ijmt}(1 - PD_{ijmt}) - MC_{ijmt}\frac{\partial Q_{ijmt}}{\partial P_{ijmt}} \qquad (3.4)\\
\iff \; & P_{ijm} = \frac{MC_{ijmt}}{1 - PD_{ijmt}} + \frac{\mathcal{M}_{ijmt}}{1 - PD_{ijmt}}
\end{aligned}
$$

where $\mathcal{M}_{ijmt} = -Q_{ijmt}/Q'_{ijmt}$ is bank's j markup on firm i loan. The equation decomposes interest rates as an effective marginal cost plus a markup, similarly to standard Bertrand-Nash pricing. The difference is represented by the risky nature of the projects that deflate both components by the probability of default.

Estimation requires knowledge of the full set of contract prices, which give rise to several considerations. First, the borrower-bank price observed in our dataset is the equilibrium price, but to estimate the model I also need prices offered from banks not chosen by firms. Second, it is likely there are unobserved characteristics that affect pricing but are unobserved by me as econometrician.

Following Crawford et al. (2018) I address these challenges using a unique feature of my data, multi-bank borrowing, to estimate a price prediction model with firm fixed effects. This allows for the prediction of interest rates considering for any price-relevant firm characteristic that is common across banks and that they observe and I do not. This reduces the likelihood that my estimates are driven by informational differences between me as econometrician and banks.

When designing a loan pricing model I must consider both *hard* and *soft* information (Liberti and Petersen, 2019) that is available to banks but may be unobservable in the data. To mitigate this concern, first note that banks in my panel follow the Internal Ratings-Based (IRB) approach and it is reasonable to believe they make predominantly use of *hard* information (even if the *soft* component cannot be removed a priori though). Indeed, a large survey by Albareto et al. (2011) shows the order in which large banks in Italy tend to acquire and use information to assess the creditworthiness of loan applicants: 1- financial statement data, 2- credit relations with the entire system, 3- statistical-quantitative methods, 4- qualitative information, 5- availability of guarantees, 6- first-hand information (branch-specific). Second, for each firm

I include in the analysis only the first main source of long-term financing, to omit any dynamics from the bank-borrower relationship. Third, exploiting the fact that a large share (about 85%) of firms borrow from multiple banks, I include firm fixed effects that absorb any borrower-specific feature that is common across banks, including soft information, but that is otherwise unobservable in the data.

Motivated by these facts, the best specification for price prediction that I estimate on the sample of firms with multiple bank relationships is

$$\begin{aligned} P_{ijmt} &= \tilde{P}_{ijmt} + \tilde{\tau}_{ijmt} \\ &= \tilde{\lambda}_{jmt} + \gamma_1 T_{ijm} + \gamma_2 L_{ijmt} + \tilde{\omega}_i^P + \tilde{\tau}_{ijmt} \\ &= \tilde{P}_{jmt} + X_{jmt}^D \tilde{\beta}^P + Y_{ijmt}^D \tilde{\eta}^P + \tilde{\omega}_i^P + \tilde{\tau}_{ijmt} \end{aligned} \tag{3.5}$$

where I pad the coefficients $\tilde{\beta}^P, \tilde{\eta}^P$ to be zero everywhere except when the component of the vectors Y_{ijmt}^D and X_{jmt}^D is, respectively, L_{ijmt} and T_{ijm}. The variables $\tilde{\omega}_i^P, \tilde{\lambda}_{jmt}$ are firm and bank-market-time fixed effects, T_{ijm} is the tenure of the relationship between borrower i and the bank j in market m, L_{ijmt} is loan size and $\tilde{\tau}_{ijmt}$ are the predictions errors. The bank-market-time fixed effects $\tilde{P}_{jmt} := \tilde{\lambda}_{jmt}$ represent the average price offered at time t in market m by bank j. For the sample of firms not borrowing or borrowing by a single bank, I use a propensity score matching to assign the firm fixed effect, loan size and PD of "closest[8]" firm in the characteristic space.

4 Estimation and Results

I estimate eqs. (3.1)-(3.2) using a GMM estimator with simulated moments (Train, 2009) and the contraction of Berry et al. (1995) (BLP). First, I rewrite the utility of bor-

[8] as measured by the euclidean distance

rowing (3.1) using the price prediction (3.5) as

$$
\begin{aligned}
U^D_{ijmt} &= \alpha^D_0 + X^D_{jmt}\beta^D + \alpha^D(\tilde{P}_{jmt} + X^D_{jmt}\tilde{\beta}^P + Y^D_{ijmt}\tilde{\eta}^P + \tilde{\omega}^P_i + \tilde{\tau}_{jm}) \\
&\quad + Y^D_{ijmt}\eta^D + \xi^D_{jmt} + u_i + \epsilon^D_{ijmt} \\
&= (\alpha^D_0 + \alpha^D\tilde{P}_{jmt} + \tilde{\beta}^D X^D_{jmt} + \xi^D_{jmt}) + (Y^D_{ijmt}\tilde{\eta}^D + u_i + \alpha^D\tilde{\omega}^P_i) + (\alpha^D\tilde{\tau}_{jm} + \epsilon^D_{ijmt}) \\
&= \delta_{jmt}(\theta_1) + V_{ijmt}(\theta_2, u_i) + \tilde{\epsilon}^D_{ijmt}
\end{aligned}
$$

Assuming that $\tilde{\epsilon}^D_{ijmt}$ is double exponential I have the familiar expressions

$$
\begin{aligned}
\mathbb{P}(D_{ijmt} = 1|u_i) &= \frac{\exp(V_{ijmt}(u_i) + \delta_{jmt})}{1 + \sum_{k\geq 1}\exp(V_{ikmt}(u_i) + \delta_{kmt})} \\
\mathbb{P}(D_{ijmt} = 1) &= \int \mathbb{P}(D_{ijmt} = 1|u_i)p(u_i)\,\mathrm{d}u_i
\end{aligned} \tag{3.6}
$$

where $u_i \sim N(0, \sigma^2)$ with density $p(u_i)$ is the firm propensity to demand a loan that is known to the firm but not to the bank.

The main issue in the estimation is that the demand equation features unobservables ξ^D_{jmt} that may be correlated with prices and that vary for each market, bank and time in my sample. Since the expression for the probability of new credit $\mathbb{P}(D_{ijmt} = 1)$ is non-linear, the inclusion of dummy variables for each ξ^D_{jmt} would entail a very high dimensional parameter space making estimation unfeasible. Hence, I follow standard practice in the literature and use the two-step procedure of Berry et al. (1995) in a system GMM to reduce the dimensionality of the problem.

In the first step, I recover bank-market-time specific constants $\hat{\tilde{\delta}}^D_{jm}$ with the contraction of BLP. These constants in turn represent the dependent variable of the second-step estimation

$$
\hat{\tilde{\delta}}^D_{jm} = \alpha^D_0 + \alpha^D\tilde{P}_{jmt} + \tilde{\beta}^D X^D_{jmt} + \xi^D_{jmt} \tag{3.7}
$$

where the unobservable ξ^D_{jmt} act as a residual that can be interpreted as the borrower's unobserved valuation of bank's characteristics, affecting bank's interest rates. To address the endogeneity of prices, as in Crawford et al. (2018) I use as instrument the

Table 3.7. STRUCTURAL ESTIMATES

	Prob(new loan)	Amount new loan
intercept	-0.229	1.629^{***}
	(2.501)	(0.136)
price	-0.609^{***}	-0.434^{***}
	(0.134)	(0.009)
D/E	0.059^{***}	0.011^{***}
	(0.005)	(0.001)
log(sales)	-0.040	0.296^{***}
	(0.083)	(0.011)
log(assets)	-0.242	1.102^{***}
	(0.232)	(0.017)
tenure	0.430^{***}	-0.246^{***}
	(0.022)	(0.009)
loan-to-value (LTV)		0.783^{***}
		(0.039)
Number of bank branches	0.791^{***}	
	(0.028)	
Bank branch density	-0.044	
	(0.231)	
Nobs	104848	104848

Notes: This table reports selected coefficients from the structural estimation. GMM standard errors are in parenthesis.

interest rates on households' deposits. These deposits are a relevant instrument since they are an important source of banks' capital and thus affect the lending conditions of branches[9]. The exclusion restriction rests on the fact that households' deposits respond to different market characteristics than firm loans.

The estimation of the loan usage (3.2) conditional on borrowing $D_{ijmt} = 1$ is more

[9]See Albareto et al. (2011).

standard and is based on

$$\begin{aligned}
\mathbb{E}[U^L_{ijmt}|D_{ijmt}=1] &= \mathbb{E}[LoanSize_{ijmt}|D_{ijmt}=1] \\
&= \alpha^L_0 + X^L_{jm}\beta^L + \alpha^L P_{ijmt} + Y^L_{ijmt}\eta^L + \mathbb{E}[u_i|D_{ijmt}=1] \\
&= \alpha^L_0 + X^L_{jm}\beta^L + \alpha^L P_{ijmt} + Y^L_{ijmt}\eta^L \\
&\quad + \int \frac{u_i \mathbb{P}(D_{ijmt}=1|u_i)p(u_i)}{\int \mathbb{P}(D_{ijmt}=1|x)p(x)\,\mathrm{d}x}\,\mathrm{d}u_i
\end{aligned} \tag{3.8}$$

As before, I need to address potential endogeneity of prices since they might be influenced by soft information that is observed by the banks but not by me as econometrician. I follow Hausman (1996) and use prices in other markets as instrument. The rationale is that banks face cost shocks that are common across markets and are reflected in their interest rates.

Finally, I stack moment conditions from equations (3.6), (3.7) and (3.8) in a joint GMM. To speed up the estimation I hand-coded analytical gradients using the Implicit function theorem for the contraction. I solve integrals numerically through Montecarlo simulation.

Results and counterfactuals Table 3.7 outlines the estimates of the structural model. As expected, the average price coefficient is negative and significant for both the loan demand and loan amount equations implying that higher interest rates negatively impact demand for loans. Other significant parameters for the demand of credit are the tenure of the relationship and the debt-equity ratio of the firm. When looking at the loan amount, larger firms (as measure by sales and total assets) tend to demand larger loans; also the loan-to-value seem to impact positively on the size of the loan.

Using the estimated coefficients I recover the marginal costs for each bank-firm pair by inverting the FOC (3.4). These marginal costs, which are otherwise unobservable to me as econometrician, constitute the key ingredient that allows me to estimate the effect of policies aimed at more tightly integrating climate-related risks into banks' business planning. Recall that the model does *not* reflect any differentiation between eligible/aligned firms and non-eligible/non-aligned ones other than what is already

Table 3.9. MARGINAL COSTS

	(1)	(2)	(3)	(4)
green	0.058	0.008	0.058	0.008
	(0.075)	(0.072)	(0.075)	(0.072)
D/E	0.011^{***}	0.011^{***}	0.011^{***}	0.011^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
log(asset)	-0.209^{***}	-0.163^{***}	-0.209^{***}	-0.163^{***}
	(0.042)	(0.041)	(0.042)	(0.041)
log(sales)	-0.084	-0.128^{**}	-0.084	-0.128^{**}
	(0.054)	(0.058)	(0.054)	(0.058)
age (years)	-0.001	-0.002	-0.001	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)
prob of def (PD)	-1.065^{***}	-1.081^{***}	-1.047^{***}	-1.064^{***}
	(0.125)	(0.126)	(0.133)	(0.135)
N	43344	43344	43344	43344
Province FE	Yes	Yes	Yes	Yes
Sector FE	No	Yes	No	Yes
Bank FE	No	No	No	Yes
Date FE	Yes	Yes	Yes	Yes

Notes: This table reports the coefficients of the regression $\widehat{mc}_{ijmt} = \beta_0 + \beta_1 Eligib_{i,t} + \Gamma' X_{ijmt} + \epsilon_{ijmt}$ where $\widehat{mc}_{ijmt}$ denotes the (estimated) marginal costs recovered by the structural estimation and FOC inversion, and $Eligib_{i,t}$ is an indicator variable equal to 1 if firm i at time t belongs to a (sub-)sector eligible according to the EU Taxonomy. My sample includes all firms belonging to NACE 2d sectors that include both eligible and not eligible NACE 3d or 4d sub-sectors. Standard errors are clustered at NACE 2 digits level.

encoded in my dataset e.g., in firms' probability of default as assessed by banks, in the soft information component or in other observable firm characteristics, consistently with the preliminary evidence from figure 3.3. The channel I have in mind is that eventually, perhaps through increased regulatory scrutiny (ECB, 2020), lending to firms more exposed to climate-related risks will become *more expensive* because banks will have to put aside more capital to cover *unexpected* losses.

Before running any counterfactual, I test if indeed the estimated model shows any systematic variation in the cost of lending between firms that belong to eligible (sub-)sectors and firms that do not. Further strengthening my initial evidence, I see no difference in marginal costs between these two groups of firms. Table 3.9 presents the results[10] and figure 3.4 in the Appendix shows the distribution of the (normalized[11]) marginal costs for all main economic sectors.

Motivated by these facts, in my first counterfactual I model any policy aimed at integrating environmental risks into banks' business planning as a standard deviation increase in the cost of lending to *non-aligned* firms according to Alessi et al. (2019). In this case, the counterfactual profit function for bank j when lending to a firm i in sector s is given by

$$\begin{aligned} \Pi^{na}_{ijmt} &= \Pi_{ijmt}(MC^{na}_{ijmt}) \\ MC^{na}_{ijmt} &= MC_{ijmt} + \sigma_{MC_s}(1 - p^a_{i,s}) \end{aligned}$$

where $p^a_{i,s}$ is the probability that firm i in sector s is aligned, and σ_{MC_s} is the standard deviation of (estimated) marginal costs for all firms in sector s. In my second counterfactual I focus on *non-eligible* firms and look at the same standard deviation increase in their cost of lending. The counterfactual profit function is thus

$$\begin{aligned} \Pi^{ne}_{ijmt} &= \Pi_{ijmt}(MC^{ne}_{ijmt}) \\ MC^{ne}_{ijmt} &= MC_{ijmt} + \sigma_{MC_s}(1 - \mathbb{I}^e_{i,s}) \end{aligned}$$

where $\mathbb{I}^e_{i,s}$ is an indicator function equal to 1 if firm i in sector s is eligible. This exercise is less broad (less severe) than the first one since the number of affected firms in the non-eligible case is smaller compared to the non-alignment one.

For both counterfactuals I use the modified marginal costs $MC^{na}_{ijmt}, MC^{ne}_{ijmt}$ to recompute price and quantities using the equilibrium conditions from the estimated structural model. I then compare the new equilibrium outcomes with the fitted ones. To speed up the computations of the changes in prices dP_{ijmt} and quantities dQ_{ijmt} following the counterfactual increase in marginal costs dMC_{ijmt}, for each firm i, in market m at time t I use the Implicit function theorem applied to the function $G :=$

[10] To recover marginal costs for the moment I ignore any estimation uncertainty from the GMM.

[11] I normalize marginal costs as $\frac{MC_{ijmt} - \mathbb{E}[MC_{ijmt}]}{\mathrm{SD}(MC_{ijmt})}$ where the expectation $\mathbb{E}[\cdot]$ and standard deviation $\mathrm{SD}(\cdot)$ are taken cross-sectionally

Table 3.11. AVERAGE CHANGE IN CREDIT CONDITIONS

	Taxonomy aligned		Taxonomy eligible	
	Δp	$\Delta \log(q)$	Δp	$\Delta \log(q)$
(Intercept)	134.277***	−0.524***	71.120***	−0.277***
	(4.948)	(0.020)	(17.346)	(0.068)
N	104848	104848	104848	104848
Province FE	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes

Notes: This table reports the average change between counterfactual and fitted interest rates (Δp, basis points) and log loan size (Δq). I obtain counterfactual (equilibrium) log loan size and interest rates increasing the marginal cost of lending to non-aligned or non-eligible firms by one standard deviation. Standard errors are in parenthesis and are clustered at NACE 2 digits level.

$(G_1, G_2, \dots, G_J) : \mathbb{R}^{J+J} \to \mathbb{R}^J$ where $G_j(P_1, P_2, \dots, P_J, MC_1, MC_2, \dots, MC_J)$ is the LHS of (3.4) for bank j.

Table 3.11 reports the main results for aggregate prices and (log) quantities[12]. Following a standard deviation increase in marginal cost for non-aligned (non-eligible) firms, I estimate that on average interest rates for these firms would increase by 134 (71) basis points, and that quantities would decrease by about 20k (13k) EUR. Figures 3.5 - 3.8 in the Appendix show the sectoral distribution of changes in price and (loq) quantities at NACE 1d level.

5 Conclusion

In this paper, I quantify the potential exposure of the Italian credit market to transition risk arising from disorderly mitigation strategies to a low-carbon economy. I complement a rich loan-level dataset that provides detailed information about credit contracts between firms and the main Italian banks, with a new granular measure of exposure to transition risk based on taxonomy-eligible and taxonomy-aligned activities. I estimate a structural model of firms' demand for credit, and use it to recover the unobservable

[12]In the table I ignore any uncertainty in the coefficients coming from the structural estimation.

firm-specific marginal cost of lending that banks face.

Using the estimated marginal costs I run two counterfactuals. In the first (second) exercise I show that if marginal costs for aligned (eligible) firms were to raise as a result of policies aimed at integrating environmental risks into banks' business planning prices would rise and quantities would fall. These findings suggest that the costs of a disorderly transition may be sizable and that a careful incorporation of climate risk into banks' stress testing frameworks and internal models could be very valuable.

Bibliography

Albareto, Giorgio, Michele Benvenuti, Sauro Mocetti, Marcello Pagnini, and Paola Rossi, "The organization of lending and the use of credit scoring techniques in Italian banks," *Journal of financial transformation*, 2011, pp. 143–57.

Alessi, Lucia, Stefano Battiston, Ana Sofia Melo, and Alan Roncoroni, "The EU Sustainability Taxonomy: a financial impact assessment," *European Commission, available at: https://ec. europa. eu/jrc/en/publication/eusustainability-taxonomy-financial-impact-assessment*, 2019.

Basel Committee, "The Internal Ratings-Based Approach," *BIS Publications*, 2001, p. 108.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile prices in market equilibrium," *Econometrica: Journal of the Econometric Society*, 1995, pp. 841–890. Publisher: JSTOR.

Bolton, Patrick, Morgan Després, L Pereira da Silva, Frédéric Samama, and Romain Svartzman, "'Green Swans': central banks in the age of climate-related risks," *Banque de France Bulletin*, 2020, *229* (8).

Crawford, Gregory S., Nicola Pavanini, and Fabiano Schivardi, "Asymmetric Information and Imperfect Competition in Lending Markets," *American Economic Review*, July 2018, *108* (7), 1659–1701.

EBA, "Mapping climate risk: Main findings from the EU-wide pilot exercise," 2021.

ECB, "Guide on climate-related and environmental risks: supervisory expectations relating to risk management and disclosure," November 2020.

, "2022 climate risk stress test," July 2022.

EU Commission, "Commission Delegated Regulation (EU) 2021/2139," July 2021.

, "Commission Delegated Regulation (EU) 2021/2178," July 2021.

, "Commission Delegated Regulation (EU) 2022/1214," March 2022.

EU Parliament, "Regulation (EU) 2020/852," July 2020.

and EU Council, "Directive 2014/95/EU," November 2014.

Faiella, Ivan and Luciano Lavecchia, "The carbon footprint of Italian loans," *Bank of Italy Occasional Paper*, 2020, (557).

Hausman, Jerry A, "Valuation of new goods under perfect and imperfect competition," in "The economics of new goods," University of Chicago Press, 1996, pp. 207–248.

Hsiang, Solomon, Robert Kopp, Amir Jina, James Rising, Michael Delgado, Shashank Mohan, DJ Rasmussen, Robert Muir-Wood, Paul Wilson, Michael Oppenheimer, and others, "Estimating economic damage from climate change in the United States," *Science*, 2017, *356* (6345), 1362–1369. Publisher: American Association for the Advancement of Science.

Liberti, José María and Mitchell A Petersen, "Information: Hard and soft," *Review of Corporate Finance Studies*, 2019, *8* (1), 1–41. Publisher: Oxford University Press.

Meucci, Giorgio and Francesca Rinaldi, "Bank exposure to climate-related physical risk In Italy: an assessment based on AnaCredit data on loans to non-financial corporations," *Bank of Italy Occasional Paper*, 2022, (706).

NFGS, "A call for action: climate change as a source of financial risk," 2019.

Oehmke, Martin and Marcus M Opp, "Green Capital Requirements," *Available at SSRN*, 2022.

Train, Kenneth E, *Discrete choice methods with simulation*, Cambridge university press, 2009.

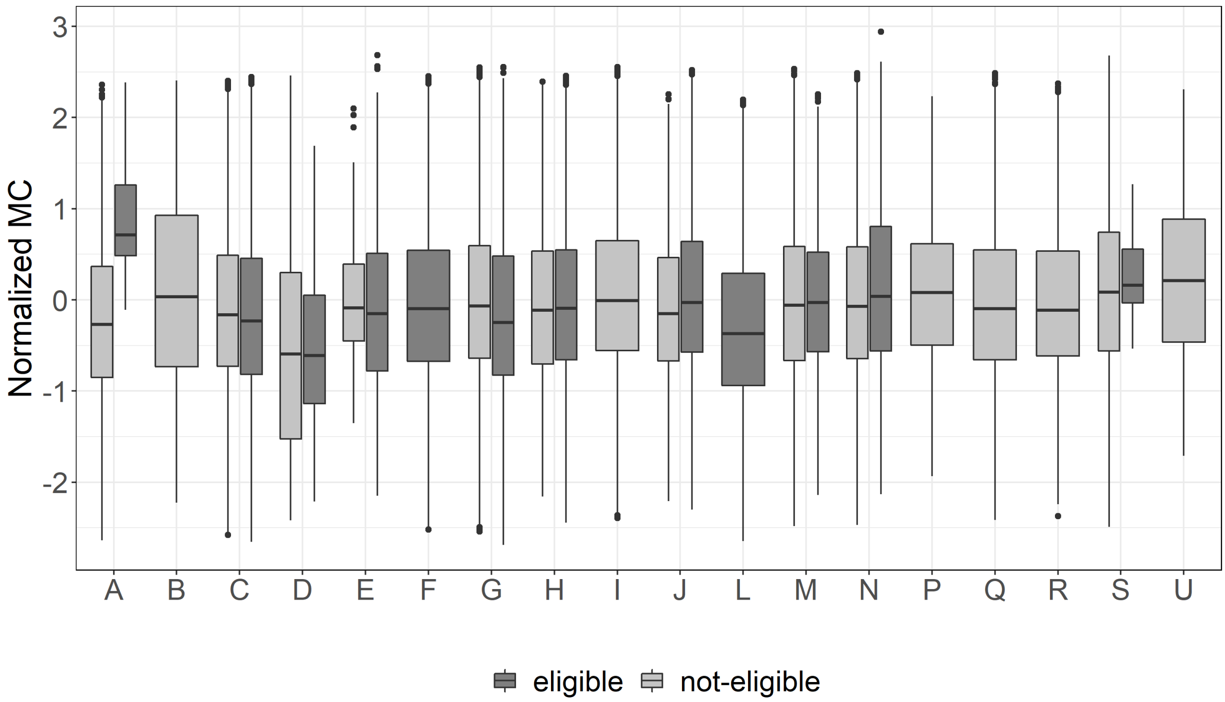
Weitzman, Martin L, "On modeling and interpreting the economics of catastrophic climate change," *The review of economics and statistics*, 2009, *91* (1), 1–19. Publisher: The MIT Press.

Appendix

Table 3.13. NACE LEVEL 1 CODES

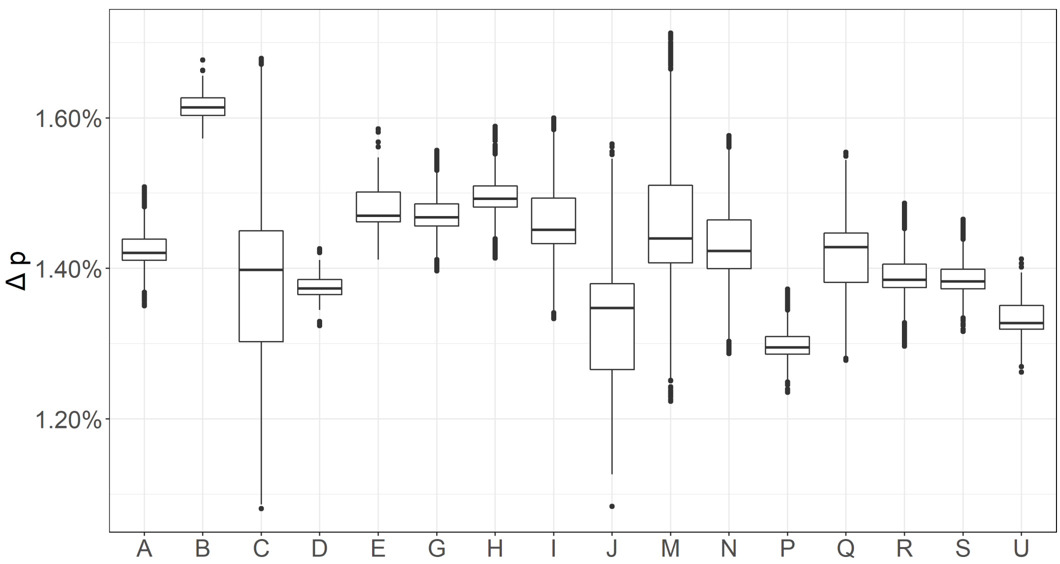
NACE Lev 1 Code	Economic Area
A	Agriculture, Forestry and Fishing
B	Mining and Quarrying
C	Manufacturing
D	Electricity, Gas, Steam and Air Conditioning Supply
E	Water Supply; Sewerage, Waste Management and Remediation Activities
F	Construction
G	Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles
H	Transportation and Storage
I	Accommodation and Food Service Activities
J	Information and Communication
K	Financial and Insurance Activities
L	Real Estate Activities
M	Professional, Scientific and Technical Activities
N	Administrative and Support Service Activities
O	Public Administration and Defence; Compulsory Social Security
P	Education
Q	Human Health and Social Work Activities
R	Arts, Entertainment and Recreation
S	Other Service Activities
T	Activities of Households as Employers
U	Activities of Extraterritorial Organisations and Bodies

Figure 3.4. Normalized marginal costs



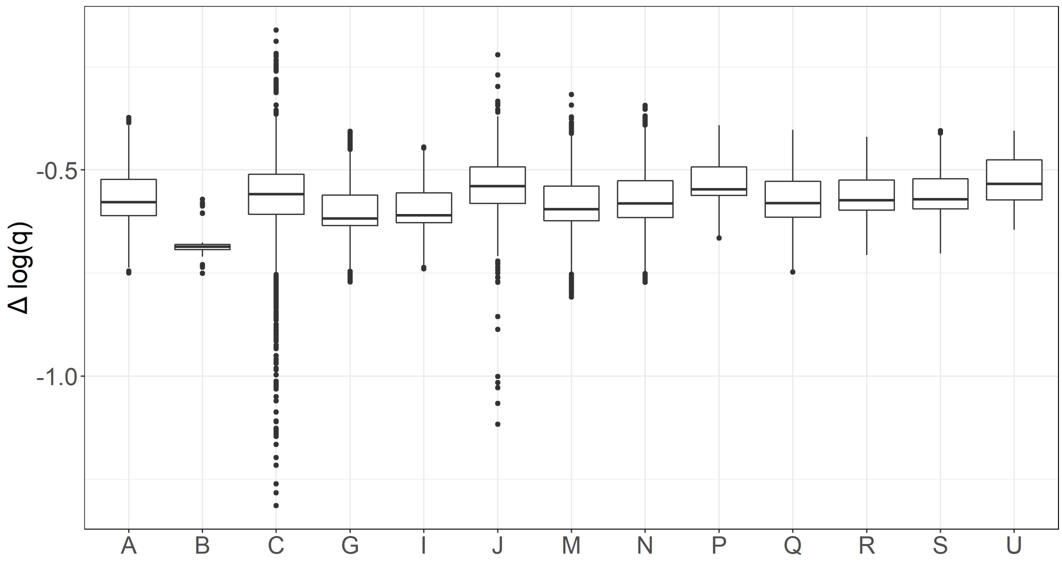
Notes: This figure plots the distribution of normalized marginal costs $\sigma_{MC}^{-1}(MC - \mu_{MC})$ obtained from the structural estimation by inverting the FOC.

Figure 3.5. Impact on prices for non-eligible firms



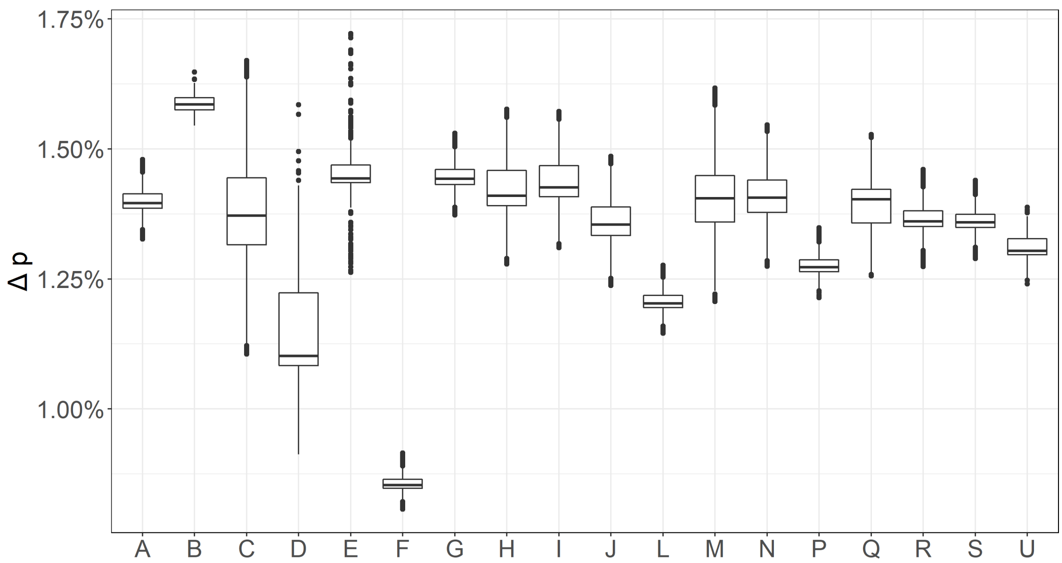
Notes: This figure plots the distribution of differences (Δp) between current and counterfactual interest rates for non-eligible firms. I obtain (equilibrium) counterfactual interest rates increasing the marginal cost of lending to non-eligible firms by one standard deviation.

Figure 3.6. IMPACT ON QUANTITIES FOR NON-ELIGIBLE FIRMS



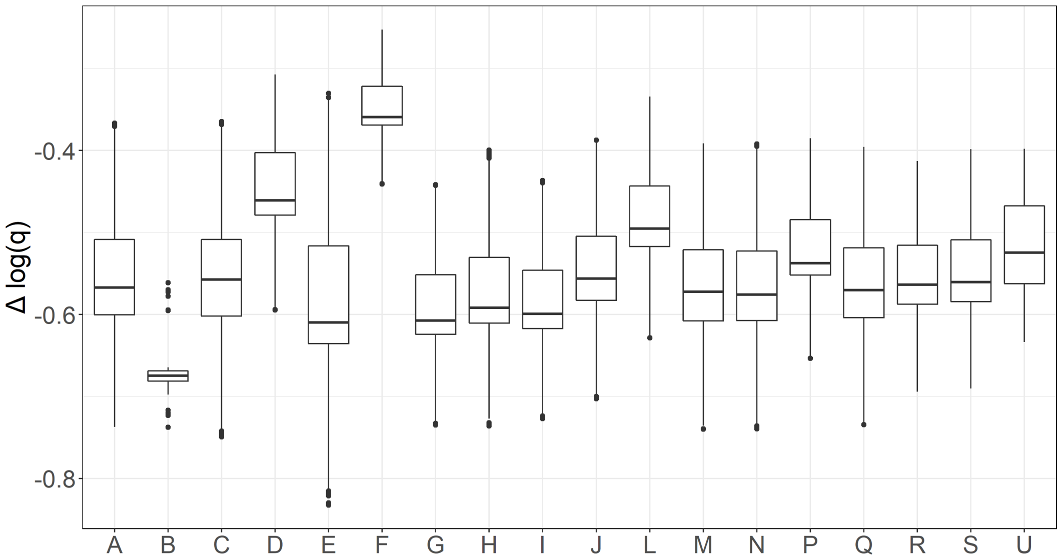
Notes: This figure plots the distribution of differences ($\Delta log(q)$) between current and counterfactual log loan size for non-eligible firms. I obtain (equilibrium) log loan size increasing the marginal cost of lending to non-eligible firms by one standard deviation.

Figure 3.7. IMPACT ON PRICES FOR NON-ALIGNED FIRMS



Notes: This figure plots the distribution of differences (Δp) between current and counterfactual interest rates for non-aligned firms. I obtain (equilibrium) counterfactual interest rates increasing the marginal cost of lending to non-aligned firms by one standard deviation.

Figure 3.8. IMPACT ON QUANTITIES FOR NON-ALIGNED FIRMS



Notes: This figure plots the distribution of differences ($\Delta log(q)$) between current and counterfactual log loan size for non-aligned firms. I obtain (equilibrium) log loan size increasing the marginal cost of lending to non-aligned firms by one standard deviation.