Neglected Information: What is in the Shape of the Return Distribution?*

Amin Jafarian† Maryam Farboodi‡

October 13, 2021

Preliminary and Incomplete

Abstract

We propose a novel measure of stock return dispersion, *Normal-Divergence*(χ_h), which has significant explanatory power for jumps in stock price around the dates of earning announcement. Normal-Divergence captures deviation of return distribution from a normal distribution with the same mean and variance, measured by KL-divergence. We argue that private information about the earnings is reflected in the pre-announcement return distribution through investor trading behavior and translates to deviation from normal distribution. A more informative pre-announcement return distribution means more information asymmetry prior to the announcement. To be willing to trade before the news is released the uninformed investors require a high rate of return to compensate them against the informed trades, which leads to a smaller jump in price at the announcement.

Keywords: Jumps, Entropy, Relative Entropy, Information, Return Predictability, Volatility Forecast

*We are grateful to Christopher Sims and Wei Xiong for helpful comments and suggestions.

†amin.jafarian@gmail.com

‡MIT Sloan School of Management, NBER, and CEPR; farboodi@mit.edu

1 Introduction

There is a considerable body of finance literature that studies the predictability of future stock returns using past returns. The asset pricing literature focuses on factors that influence asset prices, but abstracts away from the mechanics of asset price evolution and information aggregation. On the other hand, market microstructure literature focuses on how the trading process affect the evolution of trading prices, in particular how information is incorporated into prices.

In this paper we propose a novel approach to quantify the information embedded in the return distribution prior to earning announcement, and document that our proposed measure has significant negative predictive power for the abnormal return at the time of earning announcements. Our measure, *normal-divergence (χ_h)*, captures the information distance between the empirical return distribution and the minimum information distribution with the same volatility, the normal distribution. It is measured as the KL-divergence of return distribution and the corresponding normal distribution.

We solve an equilibrium model of asset price determination in which asset return does not follow a normal distribution. We show that investor optimal portfolio choice not only depends on the first two moments of the return distribution, as it would be the case with normal return, but also depends on the normal-divergence of the return distribution.

In equilibrium, investor behavior feeds back into equilibrium price, and a higher normal-divergence of pre-announcement return distribution leads to a lower jump in prices at the time of the announcement. The intuition is that prior to resolution of uncertainty at the announcement date, a high information content in the pre-announcement return distribution implies a high degree of asymmetric information. As such, the less informed traders require a compensation to be protected against informational trades, which in turn pushes the pre announcement price closer to true post announcement price and dampens the size of the jump.

We then test this hypothesis using high frequency data around quarterly earning announcements. We verify that indeed a higher pre-announcement normal-divergence, χ_h, is associated with a significantly lower normalized return at the announcement. This finding is very robust to controlling for different moments of return distribution, standard deviation, skewness and kurtosis. It is also robust to inclusion of price level prior to the announcement, as well as earning surprise as measured by the different between analyst expectations and true level of earning announcement. Moreover, we control for different combinations of fixed effect including stock, sector and time fixed effect, quarter fixed effect as well as an indicator for < 5$stocks, and show that normal-divergence still has a significant explanatory power for announcement return.

The natural instrument used in financial markets to learn about uncertainty or volatility surrounding the announcements are option prices. This observation motivates us to ask whether normal-divergence is fully priced in options. To answer this question, we construct a measure of jump in stock price at announcement, inferred from option prices, as suggested by Johannes and Dubinsky (2006). We then include this measure in our regressions, and find that normal-divergence retains its significant negative predictive power on price jump at announcement. This is a striking finding as it suggests that prior to earning announcement, there is information built into the return distribution which is internalized by traders and incorporated in their optimal portfolio choice. As such, it is reflected in the prices to compensate less informed traders prior to the resolution of uncertainty, and dampens the jump in price due to the public revelation of news. However, this information is not incorporated in option prices, and thus not fully priced by the broader financial markets.

We also find that a higher normal-divergence before earning announcements leads to lower ex-post volatility, and that this effect is concentrated among stocks with higher ex-ante volatility. A higher normal-divergence indicates that more information is already incorporated in the stock return distribution and thus the pre-announcement price already adjusts, which leads to less ex-post information revelation and lower ex-post volatility. The non-linearity is quite interesting as it suggests that information revelation is most important when uncertainty is higher, and that is exactly the time when it has the largest impact in investor behavior and prices.

Related Literature. This paper contributes to two different strands of literature. First, our finding of jump variation being significantly different from predictions solely based on market volatility contrasts with two existing approaches in the literature. In the no-arbitrage asset pricing models jump intensity is proportional to volatility or its factors (Bates (1991), Pan (2002)). In consumption-based equilibrium models, there is a tight affine connection between the variation in jump risk and volatility (Drechsler and Yaron (2010)). Closer to us is Andersen, Fusari, and Todorov (2017), who study short-maturity S&P 500 index options to analyze volatility and jump risks. They devise a semi-nonparametric approach to uncover variation in negative jump tail risk that helps predict future equity returns. Our evidence points to a very different, additional source of jump variation, information embedded in the return distribution.

Second, there is a literature that focuses on the role of information-based trading in affecting stock returns. Easley, Hvidkjaer, and O'hara (2002) constructs a measure of the probability of information-based trading, PIN, and incorporate it into a factor model of asset pricing. They argue that information captured by this factor affect asset prices and leads to abnormal expected returns. Using the PIN factor, as well as other measures of public information, Vega (2006) finds that the post-earnings-announcement drift of a given firm is related to the amount of private information available about that firm. Sadka (2006) uses PIN and argues that what is important in explaining asset pricing anomalies can be viewed as compensation for the unexpected variations in the aggregate ratio of informed traders to noise traders. Duarte and Young (2009) suggest that the PIN component related to asymmetric information is not priced, while the component related to illiquidity is priced. Easley, Hvidkjaer, and O'Hara (2010) argues that to the contrary, even controlling for the well-known liquidity factors, still PIN earns abnormal returns. Our paper contributes to this literature by quantifying the information in the return distribution in a novel way and linking it to announcement expected return both theoretically and empirically.

Outline The rest of the paper is organized as follows: section 2 lays out the necessary information theoretical concepts and notation. In Section 6, we introduce our proposed measure, *normal-divergence(χ_h)*, and formulate a model of stock price determination with asymmetric information to motivate this measure. Section 4 describes the data on which we base our analysis, and how normal-divergence is constructed from the data. Our main empirical findings are presented in Section 5. Section 7 concludes. All proofs and derivations, as well as further robustness checks are presented in the Appendix.

2 Preliminaries

Definition 1 (Differential Entropy.) *For a continuous random variable X with pdf $f(X)$, the differential entropy $h(X)$ is defined as*

$$h(X) = -\int_S f(x) \log f(x) dx \tag{1}$$

where S is the support of the random variable.

The concept of information entropy was introduced by Claude Shannon in his 1948 paper "A Mathematical Theory of Communication", Shannon (2001). It is defined as the average amount of information produced by a probabilistic stochastic source of data. In other words, entropy is a measure of unpredictability of the state, or equivalently, of its average information content.

The next remark shows how the differential entropy of a normal distribution is related to its standard deviation.

Remark 1 *Let $\xi = h(X)$ denote the entropy of random variable X. If $X \sim N(\mu, \sigma^2)$, then*

$$\xi = \ln(\sigma\sqrt{2\pi e})$$

Remark 1 shows that there is a unique monotonic transformation from standard deviation to differential entropy for a normal distribution.

The following definition provides a measure of information distance among two distributions.

Definition 2 (Kullback- Leibler (KL-) Divergence.) *For cumulative distribution functions $F(x)$ and $G(x)$ of a continuous random variable X, with pdf $f(x)$ and $g(x)$, respectively, and KL-divergence is defined to be*

$$D_{KL}(F||G) = \int_{-\infty}^{+\infty} f(x) \log \frac{f(x)}{g(x)} dx. \tag{2}$$

The KL-divergence, also called relative entropy, is a measure of how one probability distribution diverges from a second probability distribution. In the language of Bayesian inference, it is the amount of information lost when probability distribution $G(.)$ is used to approximate $F(.)$. Typically, $F(.)$ represents the "true" distribution of data, while $G(.)$ represents a model, or approximation of $F(.)$.

With the above definitions in mind, the following theorem is key to construction of our proposed measure, normal-divergence.

Theorem 1 (Maximum Entropy Distribution.) *With a normal distribution, differential entropy is maximized for a given variance. Put differently, the maximum entropy distribution under constraints of mean and variance is the normal distribution.*

Let $\mathbb{G}(\mu, \sigma^2)$ denote the family of distributions with mean μ and variance σ^2. The above theorem implies that among all the continuous distributions of this family, $G \in \mathbb{G}$, the corresponding normal distribution, $\hat{N} \equiv N(\mu, \sigma^2)$ has the minimum amount of information. As such, the KL-divergence of normal distribution with an alternative distribution in this family, $D_{KL}(\hat{N}||Q)$, measures the information distance between the two distributions. This observation is the key to construction of our information measure of interested, normal-divergence, which we introduce in section 6.

Now consider a portfolio choice problem, where investors make inferences based on realized return distribution to make their optimal decisions. Theorem 1 implies that any empirical distribution has more information content than the normal distribution with the same mean and variance, and the KL-divergence captures this *informational distance.*

Theorem 1 along with remark 1 implies that

$$\min_{X\sim G,\, G\in\mathbb{G}(\mu,\sigma^2)} h(X) = \ln(\sigma\sqrt{2\pi e}).$$

This implies that any distribution with mean μ and variance σ^2 conveys at least $\ln(\sigma\sqrt{2\pi e})$ information, proportional to the logarithm of its standard deviation.

3 Normal Divergence χ_h

In this section, we first provide an example to illustrate the emergence of the empirical pattern documented in the paper. We then introduce the key information measure that we suggest, Normal divergenceχ_h.

3.1 Illustrating Example

There are two dates, $t = 0, 1$, and a single asset. Assume the asset has return zero at $t = 0$, $R_0 = 0$, while date $t = 1$ its return is distributed via distribution $F(R_1)$.

$$R_1 = \begin{cases} x_1 & \text{with probability } p \\ -x_2 & \text{with probability } 1-p \end{cases}$$

For $x_1, x_2 > 0$. Moreover, assume asset return follows a martingale:

$$px_1 - (1-p)x_2 = 0 \Rightarrow x_2 = \frac{p}{1-p}x_1$$

Define the *expected jump* in price to be the expected absolute value of the change in return, which is given by

$$EJ = \mathbb{E}[|R_1 - R_0|] = px_1 + (1-p)x_2 = 2px_1. \tag{3}$$

The standard deviation of date $t = 1$ return is

$$\sigma = \sqrt{px_1^2 + (1-p)x_2^2} = \sqrt{px_1^2 + \frac{p^2}{1-p}x_1^2 =}\frac{p}{1-p} = \sqrt{\frac{p}{1-p}}x_1. \tag{4}$$

Using equations (3) and (4), the expected value of jump adjusted for the standard deviation is

$$\frac{EJ}{\sigma} = 2\sqrt{p(1-p)}.$$

Use Definition 2 to get

$$\frac{EJ}{\sigma} \equiv -\underbrace{D_{KL}(U||F)}_{\text{information}},$$

where $U(.)$ denotes the uniform distribution. Note that the uniform distribution with $\frac{1}{2}$ probability on each x_1 and $-x_2$ is the maximum entropy distribution on the same support, i.e. the distribution with minimal amount of information. As such, $D_{KL}(U||F)$ captures the information distance between the true return distribution $F(.)$ and the maximum entropy return distribution.

3.2 Information Measure

The above example, although stylized, demonstrates how the size of the jump in return distribution is negatively related to the amount of information embedded in the distribution, adjusted for the information in the minimum information (maximum entropy) distribution with the same support. To characterize the role of information more generally, the next definition introduces our main information measure, normal-divergence(χ_h).

Definition 3 *Let random variable X be distributed via continuous distribution $F(x)$, with mean μ, variance $\sigma^2 = Var(x)$ and entropy $\xi = h(x)$. Let $N(\mu, \sigma^2)$ denote the normal distribution with the same mean and variance as $F(x)$.*

Normal-Divergence(χ_h) is defined as the KL-divergence between F and N

$$\chi_h = D_{KL}(N||F) \tag{5}$$

Normal divergence measures the information distance between distribution $F(.)$ and the (minimum) information of the maximum entropy distribution, $N(.)$, with the same variance σ^2.

Why is χ_h a good measure? An immediate question is why normal distribution is a good benchmark with respect to which one can measure the information content of a distribution. Note that normal distribution is the maximum entropy distribution, among all the continuous distributions, subject to minimal constraints, i.e. with a single moment constraint on variance, allowing for $(-\infty, +\infty)$ support.[1]

Moreover, since historically return distribution has been modeled as a normal random variable, the KL-divergence of any alternative distribution with it, represented by χ_h, informs us about the missing information due to this approximation. We formalize this observation using the more general model in Section 6.

4 Data

Stock price and characteristics data. The data is obtained from Thomson Reuters I/B/E/S and the NYSE TAQ database. The main sample covers S&P500 firms for January 2005 - March 2016. There are 710 different tickers over the 12-year period, in 5250 different ticker-announcement pairs.

Event windows around the announcement dates for each ticker are computed, only including trading hours. For instance, suppose the announcement is at 13 : 00 on a Friday, and that the preceding Thursday and the subsequent Monday are both full trading days. A four-hour event window would start at 15 : 30 on the Thursday, and include the rest of the day's trading. It would include all of Friday's trading hours and 09 : 30 until 10 : 30 on the Monday. Millisecond data on trade prices and volumes are extracted from the NYSE TAQ database for each of these event windows. Any observations which includes a ticker suffix are dropped, as are any which refer to a trade that is later corrected, changed or signified as "cancel" or "error". Also dropped are any trades that had a "sale condition" indicator

[1] The maximum entropy distribution subject to a mean constraint is exponential. However it has the extra restriction that the support has to be positive, which makes it less suitable to model the return distribution.

that indicated it was not a "regular trade" or "automatic execution", and "odd lot trades" (trades with sizes of less than 100). The remaining data is then re-sampled at either second or minute frequency, taking the mean price and size. Finally, the number of trading seconds before or after the relevant announcement date is computed for each observation.

The same procedure is used to create a "midpoint" dataset, whereby the "pseudo-announcement" times were set to $09:30$ on the midpoint date between successive announcements for each ticker and event.

The unit of observation for our regressions is ticker-announcement. For each ticker-announcement pair, and for a chosen time interval T, we construct the return distribution using return data of frequency Δt. We use $\Delta t = 1$ second for intervals of $t = 30$ min, $1, 2$ and 4 trading hours before and after the announcement; and $\Delta t = 1$ minute for $6\frac{1}{2}$ and 13 trading hours (one and two days).

Normalized Announcement return (absolute value), $\hat{SR}_{i,t}$ is the dependent variable in our main regressions. For a ticker-announcement pair where the announcement has happened at time t, it is computed as the absolute value of the difference between average price within interval T after and before the announcement, divided by the standard deviation of the return distribution before the announcement, $\frac{|\mathbb{E}[p_{t+1}]-\mathbb{E}[p_t]|}{\sigma_t}$. We choose this measure to make the dependent variable unit-less per observation.

Normal-Divergence, $\chi_{h\ i,t}$ is our novel measure of information embedded in the return distribution, and is the key explanatory variable of interest. It is computed for each ticker-announcement observation as the KL-divergence between the optimal bandwidth kde of the empirical return distribution prior to announcement, and normal distribution of the same mean and variance.

Earning Surprise is measured as the absolute value of difference between the weighted average of analyst predictions, described below, and the actual level of earning per share announcement. We use its logarithm in our regressions.

Volume, Average Price are computed on interval T before announcement at frequency Δt. Volume is the total, averaged-over-Δt volume of trade.

Return Standard Deviation, Skewness, Kurtosis are computed from the kde of the pre-announcement return distribution, as $\mathbb{E}\left[\left(\frac{r_t-\hat{\mu}}{\hat{\sigma}}\right)^3\right]$ and $\mathbb{E}\left[\left(\frac{r_t-\hat{\mu}}{\hat{\sigma}}\right)^4\right]$ respectively, where $\hat{\mu}$ and

$\hat{\sigma}$ are the mean and variance of the kernel density estimate of the empirical return distribution in the relevant time interval.

Option-Implied Price Jump. There is a literature that uses option prices to predict abnormal announcement returns using option prices. We use option prices to construct the measure of stock return at announcement proposed by Johannes and Dubinsky (2006) and use this measure as an extra control variable in section 5.1.

We use implied volatility data for 91-day, at-the-money options, from 1996 onward, that we obtain from OptionMetrics. This data covers the last trading day prior to the announcement, and the next trading day after the announcement if the announcement occurred before closing time (16:00). Otherwise, it covers the day of the announcement (or the last trading day prior to it) and the subsequent trading day. The exact construction of the measure is provided in section 5.1.

Market capitalization, Industry classifications are obtained from CRSP.

Dummy Control Variables. We control for a number of fixed effects in our regressions. It includes individual stock fixed effect, sector fixed effect, time fixed effect (month), and quarter fixed effect (to remove seasonality). We aslo control for stocks with price $<$ 5$.

Analyst Data. Analyst forecasts of earnings per share (EPS) for the current fiscal year is obtained from I/B/E/S for all S&P 500 listed firms in the database. This data includes: tickers, date and time of EPS announcement, analyst forecasts of the EPS and identifiers for the analyst making the forecast. The latest forecast for each analyst is used to construct a mean forecast for each ticker - announcement pair. Three types of weighted averages were also computed:

1. The analysts are weighted according to the total share of the forecasts for that ticker - announcement pair that they had provided.

2. The most recent forecast for each analyst, for each ticker - announcement pair is given a weight of 3 if their forecast is within one month of the announcement date, 2 if within 3 months, 1 if within 12 months, and 0 otherwise.

3. The two above weights are multiplied (and re-scaled) to provide a composite weight.

	mean	std	Q1	median	Q3	$\hat{x}$
χ_h	0.022	0.028	0.001	0.010	0.031	0.766
annualized return std	1.112	1.071	0.595	0.820	1.216	1.039
earning surprise (log)	-3.133	1.988	-4.033	-3.041	-2.129	-1.576
market cap (million)	20.922	37.752	4.465	9.285	19.924	0.554
average price	48.695	43.663	25.057	40.247	60.784	1.115
volume	264.612	222.848	154.997	192.330	276.089	1.187
return kurtosis	9.402	34.813	3.415	4.509	6.205	0.270
return skewness	0.018	1.048	-0.076	-0.003	0.064	0.018
option implied jump	0.050	0.033	0.029	0.042	0.060	1.485
std adjusted announcement return	0.021	0.026	0.006	0.013	0.025	0.812

Table 1: Summary statistics for regression variables with $\Delta t = 1$ second, on $T = 4$ hour interval around announcement date.

Table 1 provides the summary statistics of the different variable in our main regression. The summary statistics include sample mean and standard deviation, as well as 0.25, 0.5 and 0.75 quantiles. The last column, $\hat{x}$, is the ratio of mean to standard deviation for the corresponding variable.

4.1 Kernel Density Estimation

We need to estimate the distribution of return for each ticker-announcement pair before and after the announcement as a number of our regressors, including our main information measure χ_h, depend on the return distribution prior to the announcement beyond its first four moments. Although we use $\Delta t = 1$ second returns, which with a $T = 4$ hour window allows for the possibility of 4×3600 observations on each side of the announcement, the number of observations varies widely across different ticker-announcement pairs. We use Kernel Density Estimation (kde) to estimate the probability density function of the return non-parametrically.

We need to pick a bandwidth for the kde. The bandwidth of the kernel is a free parameter which exhibits a strong influence on the resulting estimate. kde can be generated using a fixed or an optimally chosen bandwidth. If the bandwidth is too small the resulting estimated pdf is undersmoothed since it contains too many spurious data artifacts. If the bandwidth is too large then the resulting kde is oversmoothed since using the large bandwidth obscures much of the underlying structure.

In order to choose the optimal bandwidth, we use cross validation. In cross validation, the model is fit to part of the data, and then a quantitative metric is computed to determine how well this model fits the remaining data. The cross-validation score, i.e. the maximum likelihood, determines the optimal bandwidth. Such an empirical approach to model parameter selection is very flexible, and can be used regardless of the underlying data distribution.

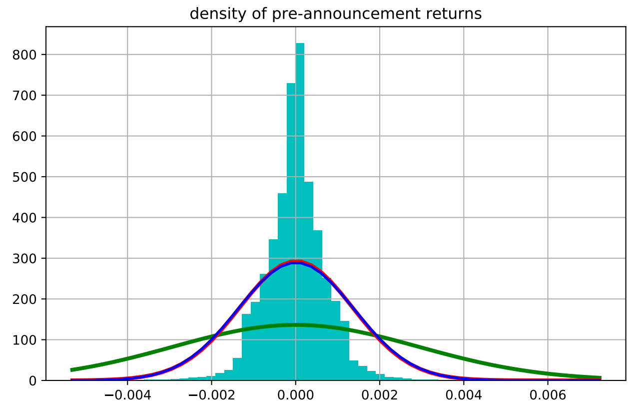
We use k-fold cross-validation. In this approach the original sample is partitioned into k equal sized subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining $k-1$ subsamples are used as training data. The cross-validation process is then repeated k time, the folds, with each of the k subsamples used exactly once as the validation data. The k results from the folds can then be averaged to produce a single estimation. The advantage of this method over repeated random sub-sampling is that all observations are used for both training and validation, and each observation is used for validation exactly once. We use 10-fold cross-validation, which is the most commonly used approach. We also restrict the number of observtions per distribution to 5000 randomly chosen observations, in order to speed up the cross validation algorithm. We also compare our estimates to constant bandwidth kde.

4.2 Test of Unimodularity

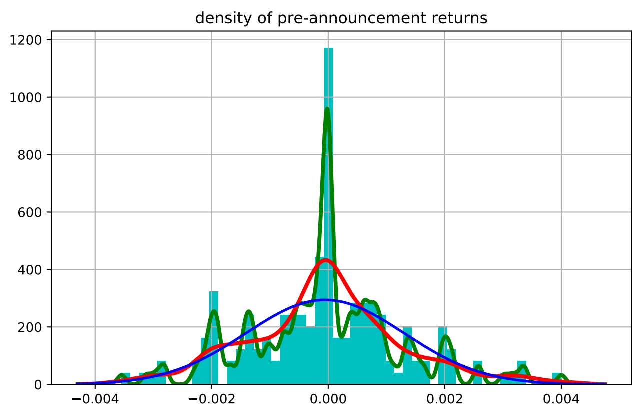
For our main regressions we focus on observations with pre-announcement return distributions that our unimodal. It turns out that the more well-known tests of unimodality, such as the DIP test developed in Hartigan and Hartigan (1985) hardly classify any distribution as unimodal. As we are working with the kde of an empirical distribution, this is very restrictive for our purposes. As such, to determine the unimodality of the distributions, we use the following heuristic approach.

For each ticker-announcement observation, we do the kde estimation, and find the maximum peak. If the ratio of the density of the two largest peaks is larger than a threshold, $\bar{f}$, we say the distribution is unimodal. We use $\bar{f} = \frac{1}{100}$ for our main estimations.

Figure 1 shows the return distribution prior to announcement for two different ticker-announcement pairs, along with the kernel density estimates, and the corresponding normal distribution with the same mean and variance as the empirical distribution. The fixed bandwidth kde clearly over-fits the empirical distribution, so we will use the optimal bandwidth kde throughout the rest of the paper. Moreover, we classify the first distribution as uni-



(a) Stock name: “Bank of America"
(‘BAC’, $2009-01-1607:00:00$); $\chi_h = 0.0005$;
Fixed Bandwidth $= 0.0028$; Optimal Bandwidth $= 0.0011$;
Number of modes $= 1$



(b) Stock name: “NewMarket Corporation"
(‘NEU’, $2014-01-3017:01:00$); $\chi_h = 0.0326$;
Fixed Bandwidth $= 7.4887e-05$; Optimal Bandwidth $= 0.0003$;
Number of modes $= 2$

Figure 1: Return distribution prior to announcement for two individual observations. $\Delta t = 1$ second return, and the time interval is 4 hours. The empirical distribution is plotted as bars. The green curve is the corresponding kde distribution with fixed bandwidth, the red curve is kde with optimal bandwidth, and the blue curve is the normal distribution with the same mean and standard deviation.
Panel (a) is classified as a unimodal distribution, while panel (b) is classified as a multi-modal distribution.

Modal, while the second one is classified as multi-Modal and is not included in the main empirical analysis. The optimal bandwidth kde visually manifests our modality classification.

5 Results

This section presents our main empirical results. We show that the normal-divergence of the pre-announcement stock return distribution, χ_h, has a significant (negative) explanatory power for the normalized return at the announcement, and explain how our model accounts for this finding. We also show that this explanatory power is robust not only to inclusion of traditional control variables, but also to inclusion of a measure of jump in prices at the announcement, implied by option prices. Furthermore, we document the forecast power of our information measure on post announcement return volatility.

Before moving to our main regressions, we start with a preliminary observation about the information content of different distributions as measure by entropy. Consider the following two normal distributions to approximate the empirical return distribution $F(.)$ with mean μ, variance σ^2, and entropy ξ. The first distribution is the Gaussian with the same mean μ, and the same variance σ^2 as the empirical distribution, $G_1(r) = N(\mu, \sigma^2)$. Using remark 1, the entropy of G_1 is given by $\xi_1 = \ln(\sigma\sqrt{2\pi e})$. The second distribution, $G_2(r)$, is the Gaussian with the same mean, μ, but the same entropy as the empirical distribution, ξ. As such, the standard deviation of the second distribution is given by $\sigma_\xi = \frac{e^\xi}{\sqrt{2\pi e}}$. From theorem 1, $\sigma_\xi^2 \leq \sigma^2$. In other words, as $G_1(r)$ is the minimum information distribution which has the same variance as $F(.)$, $\xi_1 \geq \xi$. Moreover, by construction $\xi = \xi_2$. Since there is a monotonic transformation between entropy and variance for normal distribution, it follows that the distribution with higher entropy, i.e. less information, has a higher variance too, $\sigma_2^2 = \sigma_\xi^2 \leq \sigma_1^2 = \sigma^2$.

Figure 2 verifies this intuition in the data. For each observation, we compute the ratio of entropy-implied standard deviation to standard deviation of the corresponding pre-announcement return distribution, $\frac{\sigma_\xi}{\sigma}$. Figure 2 plots the probability density of this ratio for the whole sample. As expected, $\frac{\sigma_\xi}{\sigma} < 1$ because for any given standard deviation σ, the normal distribution with the same standard deviation is the maximum entropy distribution.

Our first set of tables present the results for regressions of normalized announcement return on χ_h and multiple other relevant variables. We also gradually add more controls to ensure robustness of our findings. We pool all the observations together, and normalize each independent variables with its sample standard deviation for our main regressions.

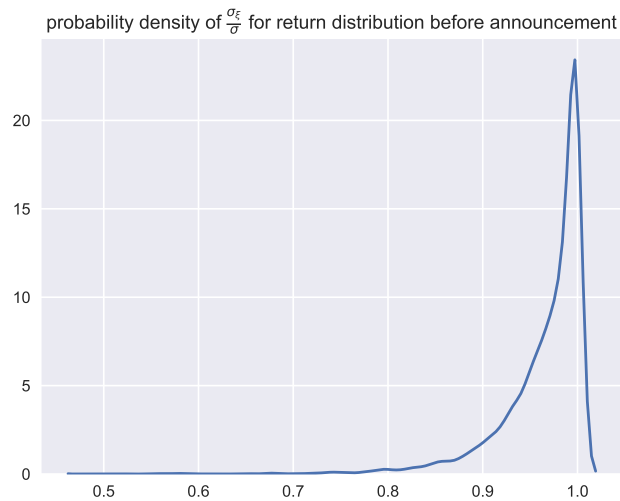


Figure 2: Sample distribution of ratio of entropy-implied standard deviation to standard deviation, $\frac{\sigma_\xi}{\sigma}$, computed for pre-announcement empirical return distribution.

Our specification is of the following form:

$$\hat{SR}_{i,t} = \alpha + \beta \chi_{h\ i,t} + \boldsymbol{X}_{i,t}\gamma + \boldsymbol{D}_{i,t}\kappa + \epsilon_{i,t} \tag{6}$$

where $\hat{SR}_{i,t}$ is the absolute value of the ratio of expected return at announcement divided by standard deviation of return distribution before announcement, and $\chi_{h\ i,t}$ is the normal-divergence of the pre-announcement return distribution. $\boldsymbol{X}_{i,t}$ are control variables, computed for interval T prior to announcement at frequency Δt, and are a subset of {earning surprise, standard deviation/skewness/kurtosis of pre-announcement return distribution, market cap, average price/total traded volume prior to the announcement}. $\boldsymbol{D}_{i,t}$ are fixed effects including stock, sector and time fixed effects, quarter fixed effects, as well as a dummy for $< 5\$$ stocks.

Tables 2 and 3 summarize our main findings. Table 2 presents the results of different variations of regression equation (6), without any fixed effects. The key observation is that β, the coefficient of regression on χ_h, is consistently negative, around -0.05, and both its level and significance are robust to inclusion of further controls. This finding is precisely in line with the predictions of our model. Right before the earning announcement is the period where asymmetric information frictions are heightened. In order for the less informed traders to trade against the more informed ones before uncertainty is resolved, they require a high

Table 2: **Announcement Return: Effect of Normal-Divergence χ_h; $T = 4$ Hour, $\Delta t = 1$ Second**

	(1)	(2)	(3)	(4)	(5)
constant	0.539	0.562	0.538	0.530	0.562
	(47.74***)	(43.60***)	(32.10***)	(21.09***)	(20.88***)
χ_h	-0.048			-0.049	-0.055
	(-5.31***)			(-5.19***)	(-5.62***)
return std		-0.058		-0.032	-0.033
		(-6.46***)		(-3.41***)	(-3.39***)
earning surprise (log)			0.023	0.027	0.027
			(2.51*)	(3.02**)	(3.05**)
market cap				0.080	0.084
				(8.68***)	(8.84***)
average price				0.037	0.032
				(3.71***)	(3.20**)
volume					-0.024
					(-2.61**)
return kurtosis					0.022
					(2.25*)
return skewness					0.020
					(2.21*)
N	4742	4742	4742	4742	4742
Adj R^2	0.57%	0.85%	0.11%	3.68%	4.03%

This table shows results from regression (6). The dependent variable is the absolute value of the ratio of expected return at announcement divided by standard deviation of return distribution before announcement, and $\chi_{h\ i,t}$ is the normal-divergence of the pre-announcement return distribution. The data covers individual stock earning announcements from January 2005- March 2016. The time interval for computing the relevant statistics is $T = 4$ hours, and the frequency is $\Delta t = 1$ second. $\boldsymbol{X}_{i,t}$ are the control variables, a subset of {earning surprise, return distribution standard deviation/skewness/kurtosis of pre-announcement return distribution, market cap, average price/total traded volume prior to the announcement}. t-stats are presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 3: **Announcement Return: Effect of Normal-Divergence χ_h; $T = 4$ Hours, $\Delta t = 1$ Second**

		(6)	(7)	(8)	(9)	(10)
constant		0.470	0.501	0.479	0.625	0.289
		(19.70***)	(21.50***)	(8.30***)	(5.08***)	(2.11*)
χ_h		-0.038	-0.041	-0.026	-0.038	-0.043
		(-4.08***)	(-3.92***)	(-2.31*)	(-3.49***)	(-3.67***)
return std		-0.046	-0.039	-0.042	-0.042	-0.043
		(-4.95***)	(-3.98***)	(-3.62***)	(-4.11***)	(-3.50***)
earning surprise (log)		0.056	0.031	0.045	0.031	0.046
		(6.23***)	(3.44***)	(4.71***)	(3.45***)	(4.89***)
market cap		0.059	0.083	0.029	0.084	0.019
		(5.91***)	(8.86***)	(1.19)	(8.91***)	(0.79)
average price		0.025	0.039	-0.001	0.029	0.018
		(2.50*)	(3.76***)	(-0.08)	(2.03*)	(1.11)
volume						-0.042
						(-3.21**)
return kurtosis						0.021
						(2.27*)
return skewness						0.021
						(2.43*)
controls	stock			✓		✓
	sectors	✓				✓
	quarter	✓	✓	✓	✓	✓
	year month		✓	✓	✓	✓
	< 5$ stocks				✓	✓
N		4742	4742	4742	4742	4742
Adj R^2		20.77%	5.65%	29.48%	5.65%	29.85%

This table continues table 2 and provides results from regression (6), controlling for different fixed effects. $\boldsymbol{D}_{i,t}$ are fixed effects including stock, sector and time fixed effects, quarter fixed effects, as well as a dummy for < 5$ stocks. t-stats are presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

compensation, which pushes the pre-announcement price closer to what the price should be after the resolution of uncertainty, which in turn decreases the size of the price jump, and thus the return normalized return at the announcement.

A second observation is that consistent with the prediction of the model, the standard deviation of return distribution before the announcement negatively predicts the return surprise.

The level effects implied by the regressions are very intuitive. A higher earning surprise, as measured by the difference between analyst expectation and the realized earning per share announcement implies a larger price change. It is also consistent with findings in the previous literature which documents earnings momentum, the positive correlation between earnings announcement surprises and cumulative abnormal returns over a longer horizon following an earnings announcement (Ball and Brown (1968)). Moreover, controlling for the information through χ_h, a larger expected price before the announcement predicts a larger change.

Lastly, column (5) clearly shows that the explanatory power of information embedded in return distribution, as measured by χ_h, is robust to inclusion of higher moments of the pre-announcement return distribution. The literature has traditionally used return kurtosis as a strong predictor of uncertainty. Our results show that once one controls for the information content of the return distribution, kurtosis becomes only marginally significant, while χ_h retains its explanatory power even in presence of kurtosis.

In table 3 we control for different fixed effects in the previous regressions to make sure our findings are robust. We control for sector fixed effects, as well as time fixed effect (year-month dummy). We control for announcement quarter to remove any potential seasonality. We also control for stock fixed effects, although this might be deemed subject to over-fitting. Finally, we include a control for $< 5\$$ stocks as they follow very different dynamics. It is clear that the observation that more information in the pre-announcement distribution implies a smaller jump in prices at the announcement is a very robust finding.

5.1 Incorporating Option Prices

We have so far argued that normal-divergence of the pre-announcement return distribution, our proposed measure of information embedded in this distribution, has significant explanatory power for the normalized announcement date return. An immediate question is whether this is prices by financial markets or not.

The natural instrument used in financial markets to learn about uncertainty or volatility surrounding the announcements are option prices. Johannes and Dubinsky (2006) proposes a method to use option prices to do inference about uncertainty embedded in earnings announcements.

To quantify the uncertainty embedded in earnings announcements, they use a sample of 20 low-dividend firms with the most actively traded options from 1996 to 2002. Based on an extension of the Black-Scholes model, they develop two estimators of the uncertainty embedded in earnings, $\sigma_t^{\mathbb{Q}}$. The first uses only ex-ante information and relies on the term structure of option implied volatility. Since this estimate can be obtained prior to the announcement, it provides an ex-ante view of investor's expectations of the uncertainty present in the earnings announcement. The second estimator, the time series estimator, uses changes in implied volatility before and after an announcement. Their results indicate that option prices are very informative about the uncertainty embedded in earnings announcements. We construct their time-series estimator using our data and add it to our main regression specification (6), to test whether normal-divergence is captured in option prices.

To construct the time series estimator of $\sigma_t^{\mathbb{Q}}$, let σ_{t,T_i}^{BS} denote the Black-Scholes implied volatility of date t of an option expiring in T_i days. If there is an earning announcement after the close on date t, then the implies volatility j days after the announcement is $\left(\sigma_{t_j,T_i-j}^{BS}\right)^2 = \sigma^2$. The change provide an estimator of the earning jump variance based on time series

$$(\sigma_{time}^{\mathbb{Q}})^2 = T_i \left((\sigma_{t,T_i}^{BS})^2 - (\sigma_{t+1,T_i-1}^{BS})^2\right)$$

We compute this volatility estimator for $j = 1$, and use it to construct their measure for jump volatility risk premium

$$J_{\tau_j+1} = \frac{r_{\tau_j+1}}{\sqrt{(\sigma_{time}^{\mathbb{Q}})^2 + \sigma^2/252}}$$

where r_{τ_j+1} is the return at the announcement. We use 90-day call options to calculate this statistic, *option-implied-jump*, and control for it in our regressions.[2]

Table 4 summarizes the results of including option-implied-jump in regression specification (6) with different sets of controls. As expected, ption-implied-jump appears significantly positive in all specifications. However, the more important observation is that the coefficient on normal-divergence changes only marginally and remains significant when we control for option-implied-jump in various specifications.

This is a striking finding, as it implies that there is some information embedded in the

[2]We do not differentiate between 90 and 91 day options.

Table 4: **Announcement Return: Effect of Normal-Divergence χ_h, Controlling for Option-Implied Jump in Prices;**
$T = 4$ **Hour,** $\Delta t = 1$ **Second**

		(11)	(12)	(13)	(14)	(15)
constant		0.358	0.395	0.322	0.492	0.303
		(22.56***)	(22.71***)	(11.01***)	(6.85***)	(2.82**)
χ_h			-0.045	-0.054	-0.022	-0.037
			(-5.07***)	(-5.69***)	(-2.01*)	(-3.26**)
return std				-0.128	-0.062	-0.068
				(-11.82***)	(-5.22***)	(-5.36***)
earning surprise (log)				0.011	0.040	0.042
				(1.25)	(4.22***)	(4.41***)
market cap				0.090	0.039	0.035
				(9.82***)	(1.65)	(1.48)
average price				0.043		
				(4.41***)		
volume				-0.008		-0.044
				(-0.82)		(-3.37***)
return kurtosis				0.038		0.024
				(4.03***)		(2.65**)
return skewness				0.020		0.020
				(2.24*)		(2.37*)
option implied jump		0.097	0.096	0.183	0.068	0.072
		(10.92***)	(10.81***)	(18.09***)	(5.91***)	(6.20***)
controls	stock				✓	✓
	sectors					✓
	quarter					✓
	year month				✓	✓
	< 5$ stocks					✓
N		4742	4742	4742	4742	4742
Adj R^2		2.43%	2.94%	10.22%	30.09%	30.51%

This table shows results from regression (6). The dependent variable is the absolute value of the ratio of expected return at announcement divided by standard deviation of return distribution before announcement, and $\chi_{h\ i,t}$ is the normal-divergence of the pre-announcement return distribution. Importantly we control for a measure of jump in prices around the announcements implied by option prices, as developed in Johannes and Dubinsky (2006). The data covers January 2005- March 2016, with interval $T = 4$ hours and frequency is $\Delta t = 1$ second. $\boldsymbol{X}_{i,t}$ are the control variables, a subset of {earning surprise, standard deviation/skewness/kurtosis of pre-announcement return distribution, market cap, average price/total traded volume prior to the announcement}. $\boldsymbol{D}_{i,t}$ are fixed effects including stock, sector, announcement time, and quarter fixed effects, as well as a dummy for < 5$ stocks. t-stats are presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

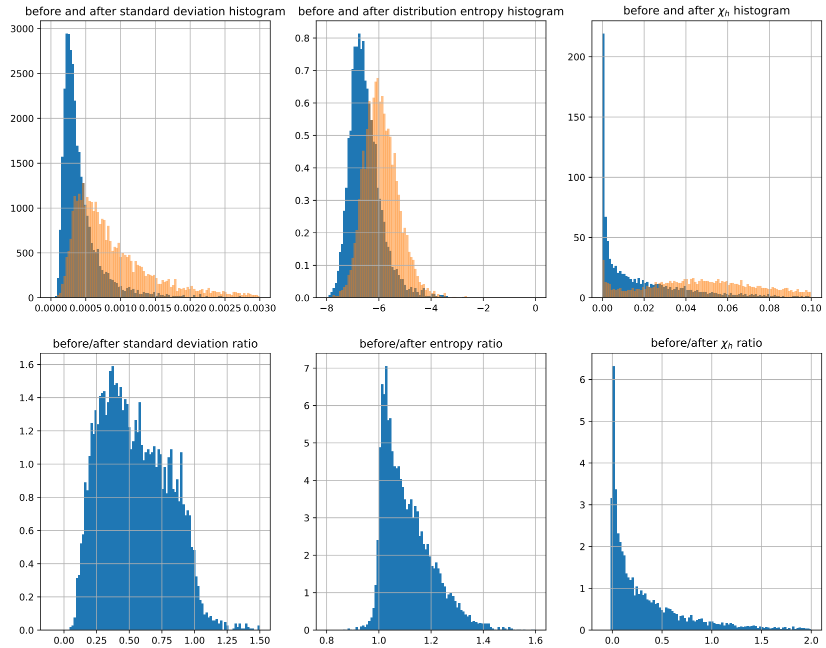


Figure 3: The first row plots the distribution of standard deviation (σ), entropy (ξ), and normal-divergence (χ_h) for the return distribution before and after the announcement, for the whole sample. The second row plot the ratio of before to after announcement return distribution.

return distribution prior to earning announcement, which is internalized by traders and reflected in their optimal demand schedule. This optimal response leads to downward adjustment in normalized announcement return, such that less informed traders are properly compensated. However, this information is not incorporated in option prices, and as such is not fully priced by the broader financial market.

5.2 Post Announcement Volatility and Information

We next turn our focus to the volatility and information in the return distribution after the announcement. Figure 3 plots the sample distribution of standard deviation (σ), entropy (ξ), and normal-divergence (χ_h) for the return distribution before and after the announcement, as well as the the ratio of the two distributions.

It is clear that all three distributions shift to the right after the announcement, which is counter-intuitive at first glance. The distribution of entropy (second column) is the most intuitive; it manifests that the information which was built in the return distribution before the announcement is all revealed, and after the announcement mainly random (Gaussian) noise remains, which has higher entropy. However, this seems inconsistent with an increase in normal-divergence post announcement (third column), which is approximately the log of standard deviation net of entropy. The key is to realize that standard deviation of the return distribution after the announce increases even more than its entropy does. The exact same build-up of information prior to announcement that leads to a distributional divergence from normal distribution, makes the return distribution also more concentrated around its mean. Once the uncertainty is resolved, this information is common knowledge, and the volatility of the post announcement return distribution is dictated by random noise, and considerably increases compared to before the announcement. This increase is sufficiently large compared to increase in entropy that the distribution of net effect, χ_h, shifts to the right.

To study the effect of information prior to announcement on post-announcement volatility and information, we focus on the following specification:

$$Y_{i,t} = \alpha + \beta \chi_{h\ i,t} + \boldsymbol{X}_{i,t}\gamma + \boldsymbol{D}_{i,t}\kappa + \epsilon_{i,t} \tag{7}$$

where $Y_{i,t}$ is the independent variable for stock announcement i at ticker t. We mainly focus on standard deviation of post-announcement distribution, $\sigma_{i,t+1}$. We also report results for normal-divergence of the post-announcement distribution, $\chi_{h\ i,t+1}$, as well as the difference between normal-divergence of pre and post distribution, $\chi_{h\ i,t+1} - \chi_{h\ i,t}$. $\boldsymbol{X}_{i,t}$ are control variables, computed for interval T prior to announcement at frequency Δt, and are a subset of {earning surprise, return distribution standard deviation/skewness/kurtosis of pre-announcement return distribution, market cap, average price/total traded volume prior to the announcement}. $\boldsymbol{D}_{i,t}$ are fixed effects including stock, sector and time fixed effects, quarter fixed effects, as well as a dummy for $< 5\$$ stocks.

Table 5 provides the results for different specifications of the regression equation (7) with $\sigma_{i,t+1}$ as the independent variable. A number of interesting findings emerge from these regressions.

First, note that when only level of χ_h is included as an explanatory variable, as in columns (1) and (2), it has a significant negative explanatory power on the ex-post volatility. This means higher information content in the pre-announcement distribution corresponds to lower ex-post volatility. This observation has an intuitive explanation: the more information is al-

Table 5: **Post Announcement Volatility: Effect of Normal-Divergence χ_h; $T = 4$ Hour, $\Delta t = 1$ Second**

		(1)	(2)	(3)	(4)	(5)
constant		1.000	1.585	0.550	0.937	0.965
		(22.59***)	(6.26***)	(9.57***)	(3.73***)	(3.80***)
χ_h		-0.046	-0.059	0.188	0.148	0.114
		(-2.83**)	(-3.29**)	(7.73***)	(5.44***)	(4.22***)
return std (σ_t)		0.853	0.568	0.988	1.100	0.819
		(52.38***)	(23.05***)	(31.04***)	(34.54***)	(18.54***)
average price		-0.088	-0.131		-0.115	-0.081
		(-5.41***)	(-4.21***)		(-3.82***)	(-2.61**)
$\chi_h \times \sigma_t$				-0.268	-0.450	-0.211
				(-7.67***)	(-12.04***)	(-5.32***)
option implied jump				0.214		0.181
				(10.72***)		(7.78***)
market cap				-0.113		-0.026
				(-7.32***)		(-0.59)
volume				-0.018		0.001
				(-1.18)		(0.06)
return kurtosis				-0.266		-0.213
				(-14.70***)		(-10.36***)
controls	sectors		✓		✓	✓
	year month		✓			✓
	quarter		✓		✓	✓
	stock		✓		✓	✓
	< 5$ stocks		✓		✓	✓
N		4290	4290	4290	4290	4290
Adj R^2		40.79%	50.35%	49.33%	48.57%	53.90%

This table shows results from regression (7). The dependent variable is the standard deviation of the individual stock return after the announcement, and $\chi_{h\ i,t}$ is the normal-divergence of the pre-announcement return distribution. The data covers individual stock earning announcements from January 2005- March 2016. The time interval for computing the relevant statistics is $T = 4$ hours, and the frequency is $\Delta t = 1$ second. $\boldsymbol{X}_{i,t}$ are the control variables, a subset of {earning surprise, return distribution standard deviation/skewness/kurtosis of pre-announcement return distribution, market cap, average price prior to the announcement, total traded volume prior to the announcement, option-implied price jump}. $\boldsymbol{D}_{i,t}$ are fixed effects including stock, sector and time fixed effects, quarter fixed effects, as well as a dummy for penny stocks. t-stats are presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 6: **Post Announcement Information in Distribution: Effect of Normal-Divergence χ_h;**
$T = 4$ **Hour,** $\Delta t = 1$ **Second**

		(1)	(2)
constant		-0.566	-0.039
		(-0.15)	(-0.15)
χ_h		0.646	-0.025
		(2.04*)	(-1.12)
return std		-0.350	-0.024
		(-1.07)	(-1.07)
average price		0.157	0.011
		(0.34)	(0.34)
controls	year month	✓	✓
	sectors	✓	✓
	< 5$ stocks	✓	✓
	quarter	✓	✓
	stock	✓	✓
N		4742	4742
Adj R^2		-2.14%	-2.29%

This table shows results from regression (7). The dependent variable is the post announcement normal-divergence of individual stock return distribution $\chi_{h\ i,t+1}$ in column (1) and $\chi_{h\ i,t+1} - \chi_{h\ i,t}$ in column (2). The data covers individual stock earning announcements from January 2005- March 2016. The time interval for computing the relevant statistics is $T = 4$ hours, and the frequency is $\Delta t = 1$ second. $\boldsymbol{X}_{i,t}$ are the control variables, a subset of {earning surprise, return distribution standard deviation/skewness/kurtosis of pre-announcement return distribution, market cap, average price/total traded volume prior to the announcement}.t-stats are presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

ready incorporated in the distribution and the pre-announcement price has already adjusted, there will be less ex-post information revelation and less volatility. More interestingly, the last three columns (3-5) reveal that this effect is non-linear and concentrated at stocks with higher ex-ante volatility. It is really when uncertainty is large that information revelation matters the most, and has a large impact on investor behavior. Columns (2), (4) and (5) show that this finding is robust to inclusion of different control variables and fixed effect. In particular, including higher moments of the pre-announcement distribution does not influence the predictive power of normal-divergence.

Moreover, note that a higher ex-ante volatility predicts a higher ex-post volatility, which is very intuitive.

Table 6 presents the results of regressions where the dependent variable is the normal-divergence for post-announcement return distribution, $\chi_{h\ i,t+1}$ (column (1)), and the difference between pre and post announcement normal divergence, $\chi_{h\ i,t+1} - \chi_{h\ i,t}$. χ_h turns out to be only marginally significant in these regressions. We are still exploring these findings.

6 Model

There are three periods, $t = 0, 1, 2$. There is a unit measure of investors who live in a large economy. Investors consume only at the last date, $t = 2$. There is one risky asset in the economy. Measure $\bar{x}$ of the asset has to be held by the investors.

Investors choose how much asset to hold at $t = 0$ at price p_0. Each investor has to sell(cover) z fraction of his portfolio at $t = 1$ at price $\tilde{p}_1$, and the complementary fraction at $t = 2$, at price p_2.[3] The investor consumes the net proceeds at $t = 2$, so his consumption is given by $c = x\left((1 - z)p_2 + z\tilde{p}_1 - p_0\right)$.

As such, there are three price, $p_0, \tilde{p}_1$, and p_2, where $\sim$ denoted that price at date $t = 1$ has a random component. p_0 and p_2 are known constants, and $\tilde{p}_1$ is determined in equilibrium. It is most useful to think of p_0 as share price sufficiently ahead of announcement, $\tilde{p}_1$ as the price close to the scheduled announcement, when some information about announcement is potentially incorporated in the price distribution, and p_2 as the price after the earning announcement has happened.

Price of the asset at $t = 1$ is given by $\tilde{p}_1 = \hat{p}_1 + r$, where r is distributed $r \sim F(r)$. $F(r)$ has mean 0 and variance σ^2, and is symmetric around its mean. It represents the de- meaned

[3] z can be endogenized using a convex inventory cost function. For clarity, in this paper we take z as the exogenous probability of market access.

return distribution entering date $t = 1$.

Distribution $F(.)$ is determined in the larger economy to reflect the existing information, and is invariant to investor decision. However, investor demand determines mean of the price distribution at $t = 1$, $\hat{p}_1$.[4]

Investors are competitive and do not internalize their influence on $\tilde{p}_1$. Each investor takes $\hat{p}_1$ and distribution $F(r)$ as given, and chooses his investment in risky asset to maximize his expected utility. In equilibrium, mean of the price of asset at $t = 1$, $\hat{p}_1$, is determined such that markets clear, and total investor demand is equal to $\bar{x}$. As such, the jump in the expected price of asset is given by $r_2 = p_2 - \mathbb{E}_F[\tilde{p}_1] = p_2 - \hat{p}_1$.[5]

6.1 Generalized Portfolio Problem

Let ρ be distributed with distribution $F(.)$ with mean μ and variance σ^2, and let $N(\mu, \sigma^2)$ denote the normal distribution with the same mean and variance. Consider the following increasing and concave utility function

$$\hat{u}(c) = 1 - e^{S(c)}$$

with $S'(c) < 0$ and $S'^2(c) + S''(c) > 0$ such that

$$\hat{u}'(c) = -S'(c)e^{S(c)} > 0 \quad \text{and} \quad \hat{u}''(c) = -\left(S'^2(c) + S''(c)\right) e^{S(c)} < 0,$$

and $c = \rho x$. Instead of utility itself, we will work with the following monotonic transformation

$$u(c) = -\left(\hat{u}(c) - 1\right) = e^{S(c)};$$

Which implies

$$\max_{c,x} \ \mathbb{E}_F[\hat{u}(c)] \equiv \min_{c,x} \ \mathbb{E}_F[u(c)]$$
$$\text{s.t. } c = \rho x$$

[4]This is equivalent to assuming investors' demand determine mean of distribution $F(r)$ but not the rest of its properties.

[5]With positive net supply of asset to the investors, it must be that $r_2 > 0$. The investor has to pay a cost at $t = 1$ with expectation 0, and bear some risk as well. So unless the return is positive he will not trade. We will focus on this case in our calculations, but all the calculations can be easily replicated for the case of $\bar{x} < 0$.

We use the Radon-Nikodym derivative $\frac{dF}{dN}$ to do a change of measure on investor objective function. The investor optimization problem is equivalent to

$$\min_x \ \mathbb{E}_N \left[\log \left(u(S(c)) \right) + \log \left(\frac{dF}{dN} \right) \right] + \frac{1}{2} Var_N \left[\log \left(u(S(c)) \right) + \log \left(\frac{dF}{dN} \right) \right].$$

This change of measure allows us to take advantage of the properties of CARA utility when the expectation is with respect to a normally distributed random variable. However, it introduces a correction term in the objective function, which depends on the Radon-Nikodym derivative. It turns out that this correction factor is closely tied to the KL-divergence of the true return distribution and the corresponding normal distribution with the same mean and variance.

Our first lemma summarizes this result. It describes the optimal portfolio for this general utility specification using the first order condition.

Lemma 1 *The investor optimal portfolio is the solution to the following equation.*

$$\mathbb{E}_N \left[\frac{dS\big(c(x;\rho)\big)}{dx} \right] (1 + D_{KL}(N||F)) + \frac{1}{2} \frac{d}{dx} Var_N \left[S\big(c(x;\rho)\big) \right] - \mathbb{E}_N \left[\frac{dS\big(c(x;\rho)\big)}{dx} \log \left(\frac{dN}{dF} \right) \right] = 0. \tag{8}$$

Lemma 1 exhibits the key mechanism of the paper. In choosing their optimal portfolio facing a general return distribution, investors not only care about the first two moments of the return distribution, but also about the information embedded in the return distribution. This information content is measured by information distance from minimum information distribution, i.e. $D_{KL}(N||F)$. We will next specialize the utility function further to make tighter predictions on how investor optimal behavior determines the equilibrium return.

6.2 Investor Optimization

In this section we specialize the utility function to piece-wise linear CARA utility. This simple extension of CARA utility has the same properties as the more general $\hat{u}(c) = 1 - e^{S(c)}$, while clearly exhibiting a case where information embedded in price distribution directly affects agents' behavior even keeping the return distribution symmetric. This is an environment where agents have different risk preferences toward good and bad outcomes, for instance one where sentiments matter for risk attitude.

$$\hat{u}(c) = 1 - e^{-(\alpha-(\alpha-\gamma)\mathbb{I}_{c>0})c}$$
$$c = (\bar{r} - z(r_2 - r))\, x.$$

with $\alpha \geq \gamma > 0$. z is the fraction of portfolio that agent knows he has to sell at $t = 1$, $\bar{r} = p_2 - p_0$ and $r_2 = p_2 - \mathbb{E}_F[\tilde{p}_1]$.

$\hat{u}(c)$ is differentiable everywhere except one point, increasing and concave. It exhibits a notion of negative sentiments towards low outcomes: the agent has a higher absolute risk aversion at low levels of consumption. Mapping this utility specification to general portfolio problem of section 6.1, we have

$$S(x;r) = -\big(\alpha - (\alpha-\gamma)\mathbb{I}_{x(\bar{r}-z(r_2-r))>0}\big)\,(\bar{r} - z(r_2 - r))\, x.$$

This function is almost everywhere differentiable with respect to x, except at $x = 0$.

$$S'(x) = \frac{dS\big(c(x;r)\big)}{dx} = -\big(\alpha - (\alpha-\gamma)\mathbb{I}_{x(\bar{r}-z(r_2-r))>0}\big)\,((\bar{r} - zr_2) + zr)) \qquad \text{iff } x \neq 0$$

Let $\Phi(z)$ denote the CDF of $N(0, \sigma^2)$ evaluated at $r = z$. Moreover, note that if z is distributed $N(0, \sigma^2)$, then $|z|$ follows a half-normal distribution with mean $\frac{\sqrt{2}\sigma}{\sqrt{\pi}}$ and variance $\sigma^2(1 - \frac{2}{\pi})$. Finally, note that $\mathrm{Sgn}(x) = \mathbb{I}_{x>0} - \mathbb{I}_{x<0}$.

We use our choice of $S(x;r)$ to simplify the general first order condition (8) as

$$\mathbb{E}_N\left[S'(x)\right]\big(1 + D_{KL}(N||F)\big) + x\left(\mathbb{E}_N\left[S'^2(x)\right] - \mathbb{E}_N^2\left[S'(x)\right]\right) - \mathbb{E}_N\left[S'(x)\log\left(\frac{dN}{dF}\right)\right] = 0.$$

And use the latter expression to derive investor optimal portfolio choice. To do so we calculate each term using the piece-wise linear CARA utility function. The algebra is cumbersome and deferred to the appendix. In order to reduce the notation, let $\mathbb{E}_F[g(r), r_1 < r < r_2] = \int_{r_1}^{r_2} g(r)dF(r)$, and define the following three functions

$$C_1(\hat{r}) = (\Phi(\frac{\hat{r}}{z}) - \frac{1}{2})\hat{r} + z\mathbb{E}_N[r, 0 < r < \frac{\hat{r}}{z}]$$
$$C_2(\hat{r}) = \mathbb{E}_N\left[(1 + zr)\log\left(\frac{dN}{dF}\right), 0 < r < \frac{\hat{r}}{z}\right]$$
$$C_3(\hat{r}) = \hat{r}^2\left(\Phi(\frac{\hat{r}}{z}) + \frac{1}{2}\right) + z^2\mathbb{E}_N[r^2, 0 < r < \frac{\hat{r}}{z}] - 2z\hat{r}\mathbb{E}_N[r, 0 < r < \frac{\hat{r}}{z}]$$

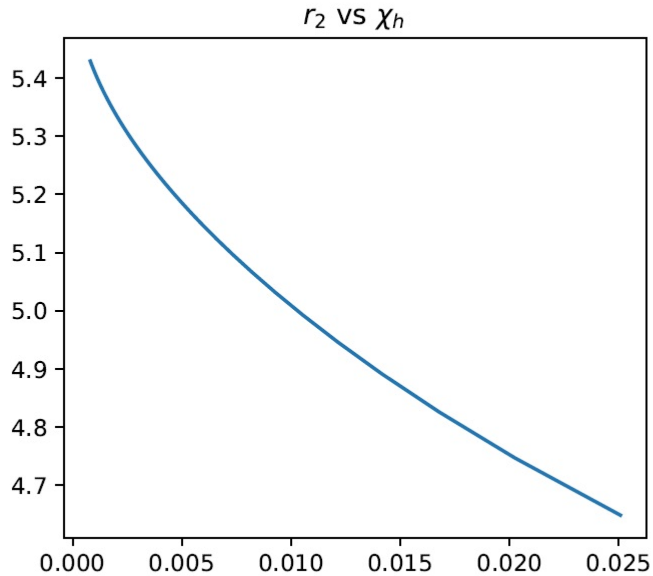


Figure 4: Jump in price at the announcement $p_2 - \mathbb{E}_F[\tilde{p}_1]$, as a function of return distribution normal-divergence, χ_h.

Using this notation, we can state our first main result, which characterizes investor's optimal portfolio.

Proposition 1 (Investor Optimal Behavior) *The optimal investor portfolio is given by*

$$\begin{aligned} x^* = & -\frac{z(\alpha+\gamma)}{a_0} r_2 - \frac{(\alpha-\gamma)\left(2C_1(\bar{r} - zr_2)Sgn(x) + \sqrt{\frac{\pi}{2}} z\sigma\right)}{a_0}(\chi_h + 1) \\ & + \frac{(\alpha-\gamma)\left(2C_2(\bar{r} - zr_2)Sgn(x) + \mathbb{E}_N\left[|r|\log\left(\frac{dN}{dF}\right)\right] + (\alpha+\gamma)\bar{r}\right)}{a_0} \end{aligned} \tag{9}$$

where

$$\begin{aligned} a_0 = & -2\left((\alpha-\gamma)\left(\frac{\sigma z}{\sqrt{2\pi}} - C_1(\bar{r} - zr_2)Sgn(x)\right) + \frac{1}{2}(\alpha+\gamma)(\bar{r} - r_2 z)\right)^2 + \frac{\sigma^2 z^2\left((2+\pi)\alpha^2 + (\pi-2)\gamma^2\right)}{\pi} \\ & + (\alpha^2 - \gamma^2)\left(2C_3(\bar{r} - zr_2)Sgn(x) - 2\sqrt{\frac{2}{\pi}}\sigma(\bar{r} - r_2 z)\right) + \left(\alpha^2 + \gamma^2\right)(\bar{r} - r_2 z)^2 \end{aligned} \tag{10}$$

The proposition shows that when $a_0 > 0$, a higher χ_h leads to a lower optimal demand submitted by the uninformed investor. Effectively, the information built in the return distribution at $t = 1$ implies that the uninformed investors require a higher rate for transacting at $t = 1$, as they lack that information and will be exploited by the informed agents in the economy. As such, more information in the current distribution leads investor to submit a smaller demand.

Moreover, r_2 is the return between $t = 1$ and $t = 2$, which is the return between dates $t = 0$ and 2, net of the return between $t = 0$ and 1. Note that the direct effect of r_2 on x^* is negative, which is quite intuitive: a higher r_2 implies a lower return for the part of portfolio sold at $t = 1$, which in turn dampens the demand. Effectively, χ_h amplifies this channel.

Since all the investors are the same, they hold the same optimal portfolio, so the aggregate demand simplifies to a representative agent. By market clearing, total demand should be equal to total supply of the asset available to investors,

$$x^*(r_2) = \bar{x} \tag{11}$$

which in turn endogenize the size of the jump.

Proposition 2 (Price Jump at Announcement) *In equilibrium expected return and the information in ex-ante return distribution, χ_h, are negatively correlated. Price jump r_2^* is characterized by*

$$\begin{aligned} r_2^* = & -\frac{a_0}{z(\alpha+\gamma)}\bar{x} - \frac{(\alpha-\gamma)\left(2C_1(\bar{r} - zr_2^*)Sgn(x) + \sqrt{\frac{\pi}{2}}z\sigma\right)}{z(\alpha+\gamma)}(\chi_h + 1) \\ & + \frac{(\alpha-\gamma)\left(2C_2(\bar{r} - zr_2^*)Sgn(x) + \mathbb{E}_N\left[|r|\log\left(\frac{dN}{dF}\right)\right] + (\alpha+\gamma)\bar{r}\right)}{z(\alpha+\gamma)}. \end{aligned} \tag{12}$$

where a_0 is defined in equation (10)*.*

Price jump is characterized as the solution to equation (9) evaluated at market clearing, (11). We solve the model for the family of student-t distributions with degree of freedom varying from 5−80, normalizing to keep the standard deviation $\sigma = 1$ across the distributions. Figure (4) plots the size of the price jump at the announcement, as a function of the χ_h. This figure reveals the main intuition of the model. In equilibrium, the more information that is built into the $4 = 0$ return distribution, a higher χ_h, the more worried the uninformed investors are that informed investors will do informational trades against them and they lose. As such, they require a higher rate of return between dates $t = 0$ and $t = 1$, which translates to a lower r_2.

Put differently, lower demand submitted by uninformed agents puts upward pressure on $t = 0$ price to get the asset market cleared, which pushes the pre-announcement price up, closer to the post announcement price, and leads to a smaller price jump. As such, a higher normal-divergence is associated with smaller returns at the announcement. We will use this intuition to guide our regressions in the empirical section.

7 Conclusion

In this paper we propose a novel measure of information embedded in return distribution, *normal-divergence* (χ_h), and document that it has significant (negative) explanatory power for jump in prices at the time of earning announcement.

To explore this finding, we take seriously the gradual information revelation prior to earnings announcements and develop a model and pricing approach to explain price jumps on earnings announcement dates. Our model suggests that a higher information distance between the return distribution prior to the announcement and the normal distribution with the same standard deviation predicts a smaller price jump at the announcement, which is captured by our novel measure, normal-divergence.

We further document that our finding is quite robust, and in particular our information measure, is not fully captured in option prices around the earning announcement. Moreover, we argue that a more informative pre-announcement return distribution leads to lower ex-post volatility, and the effect is concentrated among the stocks with more uncertain return distributions.

These findings together suggest that the return distribution prior to announcement reflects investor information which is not captured by the commonly used moments of the return distribution, and is not fully prices in options either. However, this information is internalized by the investors and affects the jump in prices at the announcement, as well as post announcement volatility.

References

Abel, A., J. Eberly, and S. Panageas (2007): "Optimal inattention to the stock market," *American Economic Review*, 97(2), 244–249.

Akins, B. K., J. Ng, and R. S. Verdi (2011): "Investor competition over information and the pricing of information asymmetry," *The Accounting Review*, 87(1), 35–58.

Andersen, T. G., N. Fusari, and V. Todorov (2017): "Short-Term Market Risks Implied by Weekly Options," *The Journal of Finance*.

Andrei, D., and M. Hasler (2014): "Investor attention and stock market volatility," *The Review of Financial Studies*, 28(1), 33–72.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006): "The cross-section of volatility and expected returns," *The Journal of Finance*, 61(1), 259–299.

Ball, R., and P. Brown (1968): "An empirical evaluation of accounting income numbers," *Journal of accounting research*, pp. 159–178.

Bates, D. S. (1991): "The Crash of '87: Was It Expected? The Evidence from Options Markets," *The journal of finance*, 46(3), 1009–1044.

Brav, A., and J. B. Heaton (2002): "Competing theories of financial anomalies," *The Review of Financial Studies*, 15(2), 575–606.

Brown, S., S. A. Hillegeist, and K. Lo (2009): "The effect of earnings surprises on information asymmetry," *Journal of Accounting and Economics*, 47(3), 208–225.

Chen, J. (2004): "Generalized entropy theory of information and market patterns," *Corporate Finance Review*, 9(3), 21–32.

Daniel, K., D. Hirshleifer, and A. Subrahmanyam (1998): "Investor psychology and security market under-and overreactions," *the Journal of Finance*, 53(6), 1839–1885.

Drechsler, I., and A. Yaron (2010): "What's vol got to do with it," *The Review of Financial Studies*, 24(1), 1–45.

Duarte, J., and L. Young (2009): "Why is PIN priced?," *Journal of Financial Economics*, 91(2), 119–138.

Easley, D., S. Hvidkjaer, and M. O'hara (2002): "Is information risk a determinant of asset returns?," *The journal of finance*, 57(5), 2185–2221.

Easley, D., S. Hvidkjaer, and M. O'Hara (2010): "Factoring information into returns," *Journal of Financial and Quantitative Analysis*, 45(2), 293–309.

Hartigan, J. A., and P. Hartigan (1985): "The dip test of unimodality," *The Annals of Statistics*, pp. 70–84.

Johannes, M., and A. Dubinsky (2006): "Earnings announcements and equity options," Working paper.

Korajczyk, R. A., and R. Sadka (2004): "Are momentum profits robust to trading costs?," *The Journal of Finance*, 59(3), 1039–1082.

Li, J., V. Todorov, and G. Tauchen (2017): "Jump regressions," *Econometrica*, 85(1), 173–195.

Pagan, A. (1984): "Econometric issues in the analysis of regressions with generated regressors," *International Economic Review*, pp. 221–247.

Pan, J. (2002): "The jump-risk premia implicit in options: Evidence from an integrated time-series study," *Journal of financial economics*, 63(1), 3–50.

Sadka, R. (2006): "Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk," *Journal of Financial Economics*, 80(2), 309–349.

Shannon, C. E. (2001): "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1), 3–55.

Tetlock, P. C. (2010): "Does public financial news resolve asymmetric information?," *The Review of Financial Studies*, 23(9), 3520–3557.

Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy (2008): "More than words: Quantifying language to measure firms' fundamentals," *The Journal of Finance*, 63(3), 1437–1467.

Vega, C. (2006): "Stock price reaction to public and private information," *Journal of Financial Economics*, 82(1), 103–133.

Veldkamp, L. (2011): *Information choice in macroeconomics and finance*. Princeton University Press.

Watts, R. L. (1978): "Systematic 'abnormal'returns after quarterly earnings announcements," *Journal of financial Economics*, 6(2-3), 127–150.

8 Appendix

8.1 Proofs

Proof of Lemma 1. The change of measure using Radon-Nikodym derivative $\frac{dF}{dN}$ implies

$$\begin{aligned}
\mathbb{E}_F[u(S(c))] =& \mathbb{E}_N\left[u(S(c))\frac{dF}{dN}\right] = \mathbb{E}_N\left[\exp\left(\log\left(u(S(c))\frac{dF}{dN}\right)\right)\right] \\
=& \mathbb{E}_N\left[\exp\left(\log\left(u(S(c))\right) + \log\left(\frac{dF}{dN}\right)\right)\right] \\
=& \exp\left(\mathbb{E}_N\left[\log\left(u(S(c))\right) + \log\left(\frac{dF}{dN}\right)\right] + \frac{1}{2}Var\left[\log\left(u(S(c))\right) + \log\left(\frac{dF}{dN}\right)\right]\right).
\end{aligned}$$

Thus the investor optimization problem is equivalent to

$$\min_x \ \mathbb{E}_N\left[\log\left(u(S(c))\right) + \log\left(\frac{dF}{dN}\right)\right] + \frac{1}{2}Var_N\left[\log\left(u(S(c))\right) + \log\left(\frac{dF}{dN}\right)\right].$$

Using the exponential transformed utility function

$$\begin{aligned}
\mathbb{E}_N\left[\log\left(u(S(c))\right) + \log\left(\frac{dF}{dN}\right)\right] =& \mathbb{E}_N\left[\log\left(u(S(c))\right)\right] - \mathbb{E}_N\left[\log\left(\frac{dN}{dF}\right)\right] \\
=& \mathbb{E}_N\left[\log\left(u(S(c))\right)\right] - D_{KL}(N||F); \\
Var_N\left[\log\left(u(S(c))\right) + \log\left(\frac{dF}{dN}\right)\right] =& Var_N\left[\log\left(u(S(c))\right)\right] + Var_N\left[\log\left(\frac{dF}{dN}\right)\right] \\
&+ 2Cov_N\left[\log\left(u(S(c))\right), \log\left(\frac{dF}{dN}\right)\right].
\end{aligned}$$

Note that

$$\begin{aligned}
Cov_N\left[\log\left(u(S(c))\right), \log\left(\frac{dF}{dN}\right)\right] =& \mathbb{E}_N\left[\log\left(u(S(c))\right)\log\left(\frac{dF}{dN}\right)\right] - \mathbb{E}_N\left[\log\left(u(S(c))\right)\right]\mathbb{E}_N\left[\log\left(\frac{dF}{dN}\right)\right] \\
=& -\mathbb{E}_N\left[\log\left(u(S(c))\right)\log\left(\frac{dN}{dF}\right)\right] + \mathbb{E}_N\left[\log\left(u(S(c))\right)\right]D_{KL}(N||F),
\end{aligned}$$

Thus the portfolio optimization problem boils down to

$$\begin{aligned}
&\min_x \mathbb{E}_N\left[\log\left(u(S(c))\right)\right] \\
&\quad +\frac{1}{2}\left(Var_N\left[\log\left(u(S(c))\right)\right]+2\left(-\mathbb{E}_N\left[\log\left(u(S(c))\right)\log\left(\frac{dN}{dF}\right)\right]+\mathbb{E}_N\left[\log\left(u(S(c))\right)\right]D_{KL}(N||F)\right)\right) \\
&\equiv \min_x \mathbb{E}_N\left[S(c)\right]\left(1+D_{KL}(N||F)\right)+\frac{1}{2}Var_N\left[S(c)\right]-\mathbb{E}_N\left[S(c)\log\left(\frac{dN}{dF}\right)\right].
\end{aligned}$$

The lemma follows directly from taking the derivative of the objective function above. ■

Proof of Proposition 1. Define, for any function $g(r)$ and constant $\hat{r} > 0$,[6]

$$M(x, \hat{r}, g(r)) = \mathbb{E}_N[g(r), r < \hat{r}]\mathbb{I}_{x>0} + \mathbb{E}_N[g(r), r > \hat{r}]\mathbb{I}_{x<0}.$$

For any any even function $g(r)$

$$M(x, \hat{r}, g(r)) = \mathbb{E}_N[g(r), r < 0] + \mathbb{E}_N[g(r), 0 < r < \hat{r}]\text{Sgn}(x),$$

and for any function $g(r) = rh(r)$ where $h(r)$ is an even function

$$M(x, \hat{r}, g(r)) = \Big[\mathbb{E}_N[g(r), r < 0] + \mathbb{E}_N[g(r), 0 < r < \hat{r}]\Big]\text{Sgn}(-x)$$

. For instance

$$\begin{aligned}
M(x, \hat{r}, 1) &= \Phi(\hat{r})\mathbb{I}_{x>0} + (1-\Phi(\hat{r}))\mathbb{I}_{x<0} = \left(\frac{1}{2} + (\Phi(\hat{r}) - \frac{1}{2})\right)\mathbb{I}_{x>0} + \left(1 - \frac{1}{2} - (\Phi(\hat{r}) - \frac{1}{2})\right)\mathbb{I}_{x<0} \\
&= \frac{1}{2} + \left(\Phi(\hat{r}) - \frac{1}{2}\right)\text{Sgn}(x)
\end{aligned}$$

[6] One can easily generalize all the expressions for $\hat{r} = \bar{r} - zr_2 < 0$, replacing $\hat{r}$ with $|\hat{r}|$ and replacing $\text{Sgn}(x)$ with $\text{Sgn}(\hat{r}x)$, appropriately, for expressions that involve taking expectations.

With this notation we get

$$\mathbb{E}_N\left[\frac{dS\big(c(x;r)\big)}{dx}\right] = -(\bar{r}-zr_2)\mathbb{E}_N\left[\alpha-(\alpha-\gamma)\mathbb{I}_{x(\bar{r}-z(r_2-r))>0}\right] - \mathbb{E}_N\left[\big(\alpha-(\alpha-\gamma)\mathbb{I}_{x(\bar{r}-z(r_2-r))>0}\big)\, zr\right]$$

$$= -\Big(\alpha-(\alpha-\gamma)M(x,\frac{(\bar{r}-zr_2)}{z},1)\Big)(\bar{r}-zr_2) + z(\alpha-\gamma)M(x,(\bar{r}-zr_2),r)$$

$$= -\frac{\alpha+\gamma}{2}(\bar{r}-zr_2) - z(\alpha-\gamma)\frac{\sigma}{\sqrt{2\pi}}$$

$$+ (\alpha-\gamma)\left((\Phi(\frac{(\bar{r}-zr_2)}{z})-\frac{1}{2})(\bar{r}-zr_2) + z\mathbb{E}_N[r, 0<r<\frac{(\bar{r}-zr_2)}{z}]\right)\mathrm{Sgn}(x)$$

$$\mathbb{E}_N\left[\frac{dS\big(c(x;r)\big)}{dx}\log\left(\frac{dN}{dF}\right)\right] = -(\bar{r}-zr_2)\mathbb{E}_N\left[\big(\alpha-(\alpha-\gamma)\mathbb{I}_{x(\bar{r}-z(r_2-r))>0}\big)\log\left(\frac{dN}{dF}\right)\right]$$

$$-\mathbb{E}_N\left[\big(\alpha-(\alpha-\gamma)\mathbb{I}_{x(\bar{r}-z(r_2-r))>0}\big)\, zr\log\left(\frac{dN}{dF}\right)\right]$$

$$= -\alpha(\bar{r}-zr_2)D_{KL}(N||F) + (\alpha-\gamma)(\bar{r}-zr_2)M\left(x,\frac{(\bar{r}-zr_2)}{z},\log\left(\frac{dN}{dF}\right)\right) - \alpha z\mathbb{E}_N\left[r\log\left(\frac{dN}{dF}\right)\right]$$

$$+ z(\alpha-\gamma)M\left(x,\frac{(\bar{r}-zr_2)}{z}, r\log\left(\frac{dN}{dF}\right)\right)$$

$$= -\frac{\alpha+\gamma}{2}(\bar{r}-zr_2)D_{KL}(N||F) - z\frac{(\alpha-\gamma)}{2}\mathbb{E}_N\left[|r|\log\left(\frac{dN}{dF}\right), r<0\right]$$

$$+ (\alpha-\gamma)\left(\mathbb{E}_N\left[\log\left(\frac{dN}{dF}\right), 0<r<\frac{(\bar{r}-zr_2)}{z}\right] + z\mathbb{E}_N\left[r\log\left(\frac{dN}{dF}\right), 0<r<\frac{(\bar{r}-zr_2)}{z}\right]\right)\mathrm{Sgn}(x)$$

where in the second expression we use $\mathbb{E}_N\left[r\log\left(\frac{dN}{dF}\right)\right] = 0$ and $\mathbb{E}_N\left[\log\left(\frac{dN}{dF}\right), r<0\right] = \frac{1}{2}D_{KL}(N||F)$ as $F(r)$ is a symmetric distribution. Moreover

$$Var_N\left[S\big(c(x;r)\big)\right] = \mathbb{E}_N\left[S\big(c(x;r)\big)^2\right] - \mathbb{E}_N\left[S\big(c(x;r)\big)\right]^2$$

$$\frac{d}{dx}\mathbb{E}_N\left[S\big(c(x;r)\big)^2\right] = \mathbb{E}_N\left[\frac{d}{dx}(\alpha-(\alpha-\gamma)\mathbb{I}_{x(\bar{r}-z(r_2-r))>0}\big)^2((\bar{r}-z(r_2-r))^2x^2\right]$$

$$=2x\mathbb{E}_N\left[\left(\alpha^2+(\alpha-\gamma)^2\mathbb{I}^2_{x(\bar{r}-z(r_2-r))>0}-2\alpha(\alpha-\gamma)\mathbb{I}_{x(\bar{r}-z(r_2-r))>0}\right)\left(\bar{r}^2+z^2(r_2^2+r^2-2r_2r)-2z\bar{r}(r_2-r)\right)\right]$$

$$=2x\mathbb{E}_N\left[\left(\alpha^2-(\alpha^2-\gamma^2)\mathbb{I}_{x(\bar{r}-z(r_2-r))>0}\right)\left((\bar{r}-zr_2)^2+z^2r^2+2z(\bar{r}-zr_2)r\right)\right]$$

$$=2x\left[\alpha^2\left((\bar{r}-zr_2)^2+z^2\sigma^2\right)-(\alpha^2-\gamma^2)\left((\bar{r}-zr_2)^2M\left(x,\frac{(\bar{r}-zr_2)}{z},1\right)\right.\right.$$
$$\left.\left.+z^2M\left(x,\frac{(\bar{r}-zr_2)}{z},r^2\right)+2z(\bar{r}-zr_2)M\left(x,\frac{(\bar{r}-zr_2)}{z},r\right)\right)\right]$$

$$=2x\left[\frac{(\alpha^2+\gamma^2)}{2}(\bar{r}-zr_2)^2+z^2\left(\frac{(\alpha^2+\gamma^2)}{2}+\frac{(\alpha^2-\gamma^2)}{\pi}\right)\sigma^2-\frac{\sqrt{2}(\alpha^2-\gamma^2)(\bar{r}-zr_2)\sigma}{\sqrt{\pi}}\right]$$
$$+2x(\alpha^2-\gamma^2)\left[(\bar{r}-zr_2)^2\left(\Phi(\frac{(\bar{r}-zr_2)}{z})-\frac{1}{2}\right)+\mathbb{E}_N[r^2,0<r<\frac{(\bar{r}-zr_2)}{z}]\right.$$
$$\left.-2z(\bar{r}-zr_2)\mathbb{E}_N[r,0<r<\frac{(\bar{r}-zr_2)}{z}]\right]\mathrm{Sgn}(x)$$

$$\frac{d}{dx}\mathbb{E}_N\left[S\big(c(x;r)\big)\right]^2=\frac{d}{dx}\mathbb{E}_N\left[(\alpha-(\alpha-\gamma)\mathbb{I}_{x(\bar{r}-z(r_2-r))>0}\big)(\bar{r}-zr_2)x\right]^2=2x\mathbb{E}_N\left[\frac{dS\big(c(x;r)\big)}{dx}\right]^2$$

Substitute in for $C_1(\hat{r}), C_2(\hat{r}), C_3(\hat{r})$ as defined in the text.

Substituting in the first order condition (8) and some further algebra leads to equation (9).

■

8.2 $\Delta t = 1$ Second Return, $T = 1h$ Before and After Announcement

This provides the results of the same regressions as in the main text for $\Delta t = 1$second return computed within $T = 1$ hour of the announcement.

Table 7: **Announcement Return: Effect of Normal-Divergence χ_h; $T = 1$ Hour, $\Delta t = 1$ Second**

	(1)	(2)	(3)	(4)	(5)
constant	0.352	0.363	0.353	0.333	0.355
	(44.64***)	(39.70***)	(29.01***)	(18.64***)	(18.87***)
χ_h	-0.026			-0.026	-0.031
	(-3.95***)			(-3.87***)	(-4.27***)
return std		-0.029		-0.012	-0.012
		(-4.41***)		(-1.74)	(-1.73)
earning surprise (log)			0.012	0.014	0.014
			(1.80)	(2.16*)	(2.21*)
market cap				0.058	0.061
				(8.65***)	(8.85***)
average price				0.020	0.016
				(2.81**)	(2.22*)
volume					-0.023
					(-3.34***)
return kurtosis					0.018
					(2.39*)
return skewness					0.012
					(1.74)
N	4613	4613	4613	4613	4613
Adj R^2	0.32%	0.40%	0.05%	2.81%	3.22%

This table shows results from regression (6). The dependent variable is the absolute value of the ratio of expected return at announcement divided by standard deviation of return distribution before announcement, and $\chi_{h\ i,t}$ is the normal-divergence of the pre-announcement return distribution. The data covers individual stock earning announcements from January 2005- March 2016. The time interval for computing the relevant statistics is $T = 1$ hours, and the frequency is $\Delta t = 1$ second. $\boldsymbol{X}_{i,t}$ are the control variables, a subset of {standard deviation/skewness/kurtosis of pre-announcement return distribution, market cap, average price/total traded volume prior to the announcement}. t-stats are presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 8: **Announcement Return: Effect of Normal-Divergence χ_h; $T = 1$ Hour, $\Delta t = 1$ Second**

		(6)	(7)	(8)	(9)	(10)
constant		0.307	0.321	0.311	0.499	0.159
		(17.93***)	(19.23***)	(7.47***)	(5.47***)	(1.57)
χ_h		-0.020	-0.013	-0.002	-0.009	-0.014
		(-3.02**)	(-1.79)	(-0.31)	(-1.22)	(-1.67)
return std		-0.026	-0.013	-0.020	-0.016	-0.018
		(-3.83***)	(-1.75)	(-2.35*)	(-2.20*)	(-2.00*)
earning surprise (log)		0.035	0.017	0.027	0.017	0.028
		(5.38***)	(2.59**)	(3.96***)	(2.62**)	(4.10***)
market cap		0.040	0.061	-0.022	0.061	-0.026
		(5.45***)	(8.84***)	(-1.31)	(8.92***)	(-1.54)
average price		0.013	0.019	-0.002	0.005	0.014
		(1.81)	(2.55*)	(-0.25)	(0.47)	(1.12)
volume						-0.031
						(-3.39***)
return kurtosis						0.013
						(1.91)
return skewness						0.006
						(0.92)
controls	stock			✓		✓
	sectors	✓				✓
	quarter	✓	✓	✓	✓	✓
	year month		✓	✓	✓	✓
	< 5$ stocks				✓	✓
N		4613	4613	4613	4613	4613
Adj R^2		21.20%	5.32%	30.71%	5.38%	30.92%

This table continues table 7 and provides results from regression (6), controlling for different fixed effects. $\boldsymbol{D}_{i,t}$ are fixed effects including stock, sector and time fixed effects, quarter fixed effects, as well as a dummy for penny stocks. t-stats are presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.
, $T = 1$ Hour, $\Delta t = 1$ Second

8.3 $\Delta t = 1$ Minute Return, $T = 2$ Trading Days Before and After Announcement

This appendix provides the results of the same regressions as in the main text for $\Delta t = 1$ Minute return computed within $T = 2$ trading days (16 hour) of the announcement.

Table 9: **Announcement Return: Effect of Normal-Divergence χ_h; $T = 2$ Days, $\Delta t = 1$ Minute**

	(1)	(2)	(3)	(4)	(5)
constant	0.248	0.305	0.176	0.303	0.294
	(11.33***)	(15.49***)	(8.31***)	(7.76***)	(6.93***)
χ_h	-0.019			-0.018	-0.022
	(-1.66)			(-1.55)	(-1.80)
return std		-0.062		-0.059	-0.063
		(-5.48***)		(-5.00***)	(-4.98***)
earning surprise (log)			-0.026	-0.017	-0.016
			(-2.28*)	(-1.44)	(-1.39)
market cap				0.003	-0.000
				(0.25)	(-0.01)
average price				-0.001	0.001
				(-0.08)	(0.05)
volume					0.009
					(0.80)
return kurtosis					0.012
					(0.90)
return skewness					-0.009
					(-0.76)
N	4290	4290	4290	4290	4290
Adj R^2	0.04%	0.67%	0.10%	0.69%	0.66%

This table shows results from regression (6). The dependent variable is the absolute value of the ratio of expected return at announcement divided by standard deviation of return distribution before announcement, and $\chi_{h\,i,t}$ is the normal-divergence of the pre-announcement return distribution. The data covers individual stock earning announcements from January 2005- March 2016. The time interval for computing the relevant statistics is $T = 2$ trading days (13 hours), and the frequency is $\Delta t = 1$ minute. $\boldsymbol{X}_{i,t}$ are the control variables, a subset of {standard deviation/skewness/kurtosis of pre-announcement return distribution, market cap, average price/total traded volume prior to the announcement}. t-stats are presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 10: **Announcement Return: Effect of Normal-Divergence χ_h; $T = 2$ Days, $\Delta t = 1$ Minute**

		(6)	(7)	(8)	(9)	(10)
constant		0.273	0.271	0.250	0.297	0.189
		(7.09***)	(7.45***)	(2.83**)	(1.99*)	(0.96)
χ_h		-0.020	-0.026	-0.038	-0.026	-0.036
		(-1.68)	(-2.17*)	(-2.85**)	(-2.17*)	(-2.54*)
return std		-0.050	-0.040	-0.020	-0.040	-0.015
		(-4.03***)	(-2.57*)	(-1.07)	(-2.57*)	(-0.71)
earning surprise (log)		-0.009	-0.013	-0.005	-0.013	-0.004
		(-0.68)	(-1.15)	(-0.33)	(-1.14)	(-0.27)
market cap		0.011	0.002	-0.024	0.002	-0.015
		(0.75)	(0.15)	(-0.70)	(0.15)	(-0.44)
average price		-0.001	-0.001	0.009	-0.003	0.012
		(-0.09)	(-0.09)	(0.50)	(-0.19)	(0.49)
volume						0.024
						(1.20)
return kurtosis						-0.008
						(-0.54)
return skewness						-0.017
						(-1.36)
controls	stock			✓		✓
	sectors	✓				✓
	quarter	✓	✓	✓	✓	✓
	year month		✓	✓	✓	✓
	< 5$ stocks				✓	✓
N		4290	4290	4290	4290	4290
Adj R^2		2.04%	2.21%	4.07%	2.19%	4.28%

This table continues table 7 and provides results from regression (6), controlling for different fixed effects. $\boldsymbol{D}_{i,t}$ are fixed effects including stock, sector and time fixed effects, quarter fixed effects, as well as a dummy for penny stocks. t-stats are presented in parenthesis. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.