

NBER WORKING PAPER SERIES

PRICE DISCOVERY AND LIQUIDITY RECOVERY:
FOREX MARKET REACTIONS TO MACRO ANNOUNCEMENTS

Masahiro Yamada
Takatoshi Ito

Working Paper 27036
<http://www.nber.org/papers/w27036>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2020

The authors acknowledge financial support from JSPS Grant Number 17H00995. The authors are grateful to comments on an earlier version by Martin Evans, Yushi Yoshida, Takamitsu Kurita, Yoshihiro Ohashi and participants of the 4th annual international conference on “High Frequency Exchange Rate Dynamics: Econophysics and Econometric/Theoretical Analysis,” December 2019 in Tokyo. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at <http://www.nber.org/papers/w27036.ack>

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by Masahiro Yamada and Takatoshi Ito. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Price Discovery and Liquidity Recovery: Forex Market Reactions to Macro Announcements
Masahiro Yamada and Takatoshi Ito
NBER Working Paper No. 27036
April 2020
JEL No. E44,F31,G14,G15

ABSTRACT

We examine whether the forex market quality, measured by the speed of price discovery and liquidity recovery after macro statistics announcements, has improved using the EBS high-frequency data for 20 years. Considering the recent rise of computer-based trading, a popular conjecture is that the market quality has improved. Our empirical analysis, however, suggests that an improving trend is only observed in price discovery. Moreover, two measures are negatively correlated because an increasing number of traders improves liquidity but slows down price discovery. Theoretically, the latter finding implies that “fast” traders have a poor interpretation of how the news will impact prices.

Masahiro Yamada
Department of Economics
Osaka University
1-7, Machikaneyama
Toyonaka, Osaka, 560-0043
Japan
m.yamada@econ.osaka-u.ac.jp

Takatoshi Ito
Columbia University
School of International and Public Affairs
Room 927 IAB (MC 3333)
420 West 118th Street
New York, NY 10027
and NBER
ti2164@columbia.edu

An appendix is available at <http://www.nber.org/data-appendix/w27036>

1 Introduction

The foreign exchange market has transformed from a human-dominated market to a machine-dominated market over the last two decades. Computers loaded with algorithms receive and process information and send limit orders and market orders to the order-matching machine, operated by EBS, Reuters, or dark pools. The order-matching machines process orders mostly on a first-come, first-served basis. Bank computers are colocated with the order-matching machine to minimize the loss in transmission time. The entire process of orders from market takers being matched with orders of market makers is completed within a fraction of a second.

In the world of high-frequency traders – machines armed with algorithms – new information should be integrated into prices much quicker than before. Although it is a near-consensus in the literature that price discovery is faster, whether high-frequency traders provide or consume liquidity is still controversial.¹

New information can be a macroeconomic news announcement at scheduled time, an unexpected result of known events like an election or a referendum, or some economic and political development at an unexpected time.

After a macro news announcement, the foreign exchange rate (price) has to move to a new equilibrium that reflects this new information. Let \tilde{v} denote a change in an efficient price after the announcement, or the true market reaction to the news. Informed traders have the correct

¹ Much academic research claims that algorithmic high-frequency traders help price discovery and liquidity provision during regular trading hours (e.g., Brogaard, Hendershott, and Riordan (2014), Chordia, Green, and Kottimukkalur (2018)). On the other hand, in situations of extremely volatile markets such as a Flash crash, high-frequency traders can consume rather than provide liquidity (Kirilenko et al. (2017)). As a study directly related to ours, Chordia, Green, and Kottimukkalur (2018) investigate equity markets and report that price discovery after macro announcements has become faster during these years. Scholtus, Van Dijk, and Frijns (2014) investigated the U.S. equity market and emphasize that the speed of trading is important for its profitability from the news trading. In line with the finding on market quality by Scholtus, Van Dijk, and Frijns (2014), Jiang, Lo, and Valente (2012) report that high-frequency trading lowers the depth on the limit order book during the post-announcement period, but increases price efficiency through the trade.

expectation \hat{y} , which predicts \tilde{v} . Upon the arrival of news y , the surprise component of new information $y - \hat{y}$ is integrated into the market price through trades among market participants, achieving $\tilde{v} = y - \hat{y}$. Some of the market participants are informed and others are uninformed. Not knowing \hat{y} , uninformed traders try to learn the correct contents and interpretation of news through trading and price and volume movements.

After new information arrives, the price level moves to a new equilibrium level. This is called price discovery. When traders do not share the same information and interpretation, they extract information from prices during the trading process toward price discovery. It is expected that the volatility becomes higher upon the arrival of news, and remains high for a period of time. This is a phenomenon called volatility clustering.

Interesting and important questions are, first, whether widespread high-frequency traders have made the price discovery process faster; and second, whether they provide more liquidity so that volatility is less than that during previous times without machines. With tick-by-tick data, the price discovery process can be examined with very high frequency. The impact of the widespread use of machines can only be analyzed with tick-by-tick data.

In cases when there is a macroeconomic statistics announcement at a scheduled time, liquidity dries up prior to the announcement time, as traders prefer to stay on the sidelines given a strong amount of uncertainty. After the number of macro statistics is announced and news contents are fully digested, the liquidity recovers to the pre-announcement time.

We use 20 years of high-frequency data to examine the changes in the market quality defined by the speed of price discovery and liquidity recovery. The data have been collected in the ICAP EBS order-matching system, which makes it possible for us to examine long-term changes. We focus on price and liquidity reactions to macro announcements. We expect the market quality has changed over time as machines have gradually replaced humans.

We define the speed of price discovery as a variance ratio, i.e., the variance of return in $[u, u + k]$ over the variance of return in $[u, u + T]$, where u is the time of news arrival; k and T are free parameters satisfying $k \leq T$.² Liquidity recovery is measured by the shrinkage of effective bid-ask spreads in a similar manner; the liquidity dries up at the moment of announcements and then gradually recovers in some minutes.

Figure 1 describes the developments of the speed of price discovery and liquidity recovery for EUR/USD in these 20 years.

[Figure 1 is inserted here]

Contrary to our expectations, the changes in the speed of price discovery are only slight over the years, with large fluctuations. We do not observe a trend of increasing speed in liquidity recovery. In fact, the liquidity recovery was fast enough even in 1999, and the room for improvements was small. Moreover, these two measures fluctuate to the opposite directions and do not necessarily improve simultaneously. The correlation coefficient is -0.355 . These findings call for more investigation into the underlying factors that drive the market quality.

Based on the observations from Figure 1, the next question is what drives the fluctuation of price discovery and liquidity. We set two key variables that are theoretically relevant and empirically available: explanatory power of the news for predicting after-the-news returns, and the number of traders who submit limit orders. The first variable, which we call “R-squared measure,” is obtained by an explanatory power, or R-squared, of the surprise of macro statistics announcements against after-the-news returns.³ Because the surprise component is calculated from the expectation of news

² Chordia, Green, and Kottimukkalur (2018) employ a similar measure and report that price discovery after announcements becomes faster in recent years in the U.S. equity market. They employ two second for T and 0.1 second for k . Considering the market structure and trade frequency, we employ the longer time interval.

³ For defining the surprise of news, we use pre-expectations of macro statistics provided by Bloomberg.

provided by Bloomberg’s survey, its explanatory power can be close to that of market participants. The R-squared is proportionally related to the precision of information held by traders. Based on this idea, we use this R-squared measure as a proxy for the precision of information held by traders.

The second variable is “quote counts,” or the number of traders who hold limit orders, which is available in our high-frequency dataset. The number of traders is an important component to determine the price discovery and liquidity, but it depends on whether they are informed or uninformed traders.

Our regression analysis shows that the price discovery is faster when the R-squared measure is higher, but slower when the quote counts are higher. Conversely, the liquidity recovery is slower when the R-squared is higher, but it becomes faster when the quote counts are higher. Figure 2 graphically illustrates these main empirical findings.

[Figure 2 is inserted here]

Theoretically, our empirical findings are consistent with a situation in which market participants behave like uninformed traders, as implied by classic noisy rational expectation models (e.g., Grossman and Stiglitz (1980)). When traders behave like uninformed traders, they provide liquidity but cannot add any information to the market price.⁴ Therefore, an increasing number of uninformed traders deteriorates the price discovery. In reality, however, trading after macro announcements may be more like informed trading, because they hold their own interpretation of

⁴ Uninformed traders can contribute to price discovery when noise traders’ temporary move the market price away from its efficient price. But this situation fits to regular trading hours rather than to the timing of news releases. In the literature, Anand, Tanggaard, and Weaver (2009) investigated data on the Stockholm Stock Exchange and reported that liquidity provision leads to an improvement in price discovery. Scholtus, Van Dijk, and Frijns (2014) show that the speed of HFTs is an important determinant of the profitability of their news-driven trading strategy, and the activity of HFTs positively affect market quality defined as depth, bid-ask spread, and price resiliency.

the news as private information. This is an idea of Kim and Verrecchia (1997) to investigate the trading after news announcements.

In the theoretical model of Kim and Verrecchia (1997), each trader is endowed with a private expectation \tilde{y}_i involving “interpretation error of news” $\tilde{\epsilon}_i$: $\tilde{y}_i = \hat{y} + \tilde{\epsilon}_i$. Each trader’s objective is to make a prediction for $\tilde{v} = y - \hat{y}$. In this situation, a trader’s own surprise regarding news, $y - \tilde{y}_i = \tilde{v} + \tilde{\epsilon}_i$, is related with \tilde{v} , and it is associated with the interpretation error. When the error is large, the trader only has noisy information and trades like an uninformed trader. In this way, a problem with public information is transformed into a problem with the standard microstructure model of private information.

Thus, a problem is whether we can reproduce the reality with a model where there are only informed traders. To handle this situation, we rely on the Kyle (1989) model, which provides a general microstructure framework. This model is regarded as a general one, because (i) the model is inhabited with N risk-averse informed, M risk-averse uninformed, and random noise traders; (ii) each informed trader has their own private information, and other informed and uninformed traders extract signals from equilibrium prices, i.e., the information structure is not nested; (iii) both informed and uninformed traders have market power; and (iv) it is a model for an order-driven market, the same as the structure of the EBS interbank market. Because of the property (ii), an informed trader has an aspect of both an informed and uninformed trader, and we can consider a situation in which there are only informed traders.

Our empirical measures of market quality, i.e., the variance ratio and liquidity, can be formulated in this theoretical model to pin down their connections with exogenous parameters. We show comparative statics analysis for these market quality measures with respect to the number of traders and the precision of their private information. A key additional assumption for replicating the empirical results is that the sum of precision of trader’s private information is fixed. In this situation, price discovery is not faster even if there are many traders, since such traders only

improve liquidity, like an uninformed trader. As a result, better liquidity and fast price discovery are not necessarily achieved at the same time.

After controlling for these determinants, a time trend of market quality can be empirically obtained. We find that the exogenous quality improvements in the Forex market are mixed: price discovery becomes faster, but liquidity recovery becomes slower. These findings are not only applicable to the specific currency pair, but are observable across a wide range of currency pairs.

The outline of this paper is as follows. Section 2 provides a theoretical model that suggests the empirical measures of price discovery and liquidity, as well as proxies for exogenous parameters. Section 3 explains the data and defines the empirical measures and their overview, and then we conduct regression analyses to find determinants of market quality and any exogenous time-trend of improvements. Section 4 further conducts related empirical analyses in order to check their robustness. We also demonstrate numerical simulations on the model by feeding parameters that match the reality. Section 5 concludes.

2 A Stylized Model

In this section, we present a model that provides a guide to constructing the possible measures of price discovery and liquidity recovery, their determinants, and an interpretation of empirical analysis. Based on the classical Kyle (1989) model, we consider an environment where the sum of precision of trader's private information is fixed. In this set up, despite that each informed trader has exclusive private information, their marginal contribution to the price informativeness can be small, so that they behave like uninformed traders.

Our model features one trading stage, one risky and risk-free asset, and N informed and random noise traders. At $t = 0$, informed trader n is endowed with an initial belief about the news statistics $\tilde{y}_n = \hat{y} + \tilde{\epsilon}_n$, $n = 1, 2, \dots, N$ as well as a public announcement y . Following Kim and Verrecchia

(1994), $\tilde{\epsilon}_n$, $n = 1, 2, \dots, N$ is regarded as an interpretation error for each trader. After-the-news price \tilde{v} is realized in a form of $\tilde{v} = y - \hat{y}$. Thus, we assume that the after-the-news price is a function of the “true” unobserved surprise of the announcement. Based on this structure, trader k can generate their own signal $\tilde{t}_n \equiv y - \tilde{y}_n = \tilde{v} + \tilde{\epsilon}_n$ for predicting \tilde{v} .

Both informed and noise traders trade the risky asset at $t = 1$, and the asset value is realized at $t = 2$. We assume the normal distribution of random variables: $\tilde{v} \sim \mathbf{N}(0, \tau_v^{-1})$ and $\tilde{\epsilon}_n \sim \mathbf{N}(0, \tau_e^{-1})$ which are mutually independent. Note that the unconditional expectation of \tilde{v} is assumed to be zero, so that the price of risk asset at $t = 0$ is zero as well. This structure allows us to interpret the price of risky asset at $t = 1$ as one-period return of the asset and \tilde{v} as the after-the-news return.⁵ The risk-free asset earns zero interest.

Informed traders have a negative exponential utility function, and they maximize a terminal wealth by choosing an order submission strategy. An informed trader’s profit maximization problem transforms into a mean-variance optimizing problem:

$$\max_{x_n} \mathbf{E}[(\tilde{v} - p(x_n))x_n | i_n, p] - \frac{\rho}{2} \text{Var}[(\tilde{v} - p(x_n))x_n | i_n, p],$$

where ρ is a risk aversion parameter. At $t = 1$, informed traders submit their demand/supply functions to an auctioneer and simultaneously observe an aggregate demand/supply curve $p(x_n)$. Noise traders randomly submit market orders $\tilde{z} \sim \mathbf{N}(0, \sigma_z^2)$, which is inelastic to the price and independent of \tilde{v} and $\tilde{\epsilon}_n$. The price at $t = 1$ clears the demand/supply of informed and noise traders; i.e., the market-clearing condition $\sum_n x_n(p) + \tilde{z} = 0$ holds.

This model involves signal extraction and imperfect competition. From the assumptions on information structure, each informed trader receives a different news signal; that is, the information is not nested, which enables informed traders to extract additional information from the aggregate

⁵ In our empirical analysis, we measure the surprise as a deviation of released statistics from the Bloomberg survey results. Although such survey results are also publicly available for market participants, for simplicity, the theoretical structure does not account them explicitly. Alternatively, we can assume that the unconditional expectation of \tilde{v} already reflects the survey results and is set as zero.

demand/supply curve – namely, $E(\tilde{v}|i_n, p) \neq E(\tilde{v}|i_n)$ and $\text{Var}(\tilde{v}|i_n, p) \leq \text{Var}(\tilde{v}|i_n)$. In contrast to the competitive model of Grossman and Stiglitz (1980), informed traders utilize their monopoly power, which is inherited from their informational advantage. Therefore, compared to a competitive case, traders are less aggressive in informed trading; the incorporation of information into prices can be delayed, and the liquidity can decrease.

In the original model of Kyle (1989), there are uninformed traders who do not receive private signals, but optimally characterize their demand/supply functions. In our model, they are muted for making the characterization of equilibrium easier. In fact, it is reasonable to assume there are only informed traders after an announcement, because each trader observes public news and trades based on their interpretation of news. Even without uninformed traders, however, informed traders learn from prices and act like uninformed traders depending on the precision of their information.

Readers familiar with the literature might notice that Kyle (1989) is a model for private information, not for public information. Microstructure models regarding public information release are studied by, for example, Kim and Verrecchia (2006), Kim and Verrecchia (1994), Wang (1994), Holden and Subrahmanyam (2002), and Llorente and Michaely (2002). Although these papers investigate a market structure after the release of public information, their model structures allow private information to facilitate informed trading. The information structure of our model can be regarded as a variation of Kim and Verrecchia (1994): a source of private information is individual’s information production that relates the public news and terminal asset values.⁶ Another source of private information can be idiosyncratic liquidity shocks during pre-announcement, which is studied by Llorente and Michaely (2002) and Tetlock (2010). But this reasoning is more suitable to modeling longer-time horizon situations, not the intraday tick-by-tick transactions.

⁶ Another important difference with Kim and Verrecchia (1994) is that they employ a quote-driven market (Kyle (1985) type model) while we employ an order-driven one (Kyle (1989) type model). The difference is discussed by Bernhardt and Taub (2006), but it is not in the scope of this paper.

Overall, our theoretical framework is not quite original. We use the Kyle (1989) model to derive reasonable predictions of intraday order-driven market transactions in a tractable and general manner. However, this model is rich enough to provide theoretical predictions that can be compared with our empirical analyses.

2.1 Characterization of equilibrium⁷

We focus on a symmetric linear equilibrium that determines the strategy of informed traders. We start from a conjecture of the strategy that is written by constants β, γ, μ ;

$$x_n(p, i_n) = \mu + \beta i_n - \gamma p, \quad n = 1, 2, \dots, N \quad (1)$$

The market-clearing condition gives the equilibrium price as

$$p = \lambda \left(N\beta\tilde{v} + \beta \sum_n \tilde{e}_n + \tilde{z} + N\mu \right), \quad \lambda \equiv \frac{1}{N\gamma} \quad (2)$$

λ is price impact, a price change per unit of unexpected order flow, and it is interpreted as a measure of illiquidity. To derive the expression of β, γ, μ , we solve the profit maximization problem as well as a signal extraction problem for informed traders. Profit maximization gives the following demand function

$$x_n(i_n, p) = \frac{E(\tilde{v}|i_n, p) - p}{\lambda_I + \rho \text{Var}(\tilde{v}|i_n, p)}, \quad \lambda_I \equiv \frac{N}{N-1} \lambda, \quad (3)$$

and a second order condition $2\lambda_I + \rho \text{Var}(\tilde{v}|i_n, p) > 0$. The signal extraction problem gives a form of conditional expectation and conditional variance⁸

$$E(\tilde{v}|i_n, p) = \frac{(1-\varphi)\tau_e}{\tau_I} i_n + \frac{\varphi\tau_e}{\beta\lambda\tau_I} p,$$

$$\tau_I^{-1} \equiv \text{Var}(\tilde{v}|i_n, p) = (\tau_v + \tau_e + (N-1)\varphi\tau_e)^{-1},$$

⁷ The derivation of the equilibrium follows Theorem 4.1, 5.1 and 5.2 of Kyle (1989). The proofs of these theorems are provided by Kyle (1989), and we do not repeat the details.

⁸ Each trader receives their own signal and prices, and they can extract that is correlated with the future return but independent of the own signal. Applying normal-normal update, we can obtain the results. See Theorem 4.1 in Kyle (1989) for more detailed derivation.

$$\varphi \equiv \frac{(N-1)\beta^2}{(N-1)\beta^2 + \sigma_z^2 \tau_e}. \quad (4)$$

φ is a measure of informational efficiency of the price. $E(\tilde{v}) = 0$ leads $\mu = 0$. Thus, φ and τ_I are a function of β . Plugging these expressions into the demand function and matching the strategic constants β and γ with the conjecture (1), we obtain

$$\beta = \frac{(1-\varphi)\tau_e}{\lambda_I \tau_I + \rho}, \quad \gamma = \frac{\beta \lambda \tau_I - \varphi \tau_e}{\beta \lambda (\lambda_I \tau_I + \rho)}. \quad (5)$$

The second order condition implies that the denominator of β is positive even if λ_I were to be negative. φ is less than 1 and τ_e is positive by their definitions. Thus $\beta > 0$. Substituting out $\lambda, \lambda_I, \varphi$, these two equations are combined to yield a cubic equation for the endogenous variable β ,

$$\rho \beta^3 + \frac{N\tau_e}{N-1} \beta^2 + \frac{\sigma_z^2 \tau_e}{N-1} \rho \beta - \frac{(N-2)\sigma_z^2 \tau_e^2}{(N-1)^2} = 0. \quad (6)$$

By definition, all the coefficients on β are positive, and the constant term is negative. Thus, the left-hand side of equation (6) is monotonically increasing in β if $\beta > 0$, and a unique solution exists for $N \geq 3$. By determining β , we obtain the solutions of other endogenous variables. From equation (5), λ is a function of β and exogenous parameters. Rewriting (5) with the definition of τ_I yields

$$\lambda = \frac{1}{N\beta} \left(1 - \frac{\tau_v}{\tau_I} \right) \quad (7)$$

The signal extraction problem (4) implies that $1 - \tau_v/\tau_I$ means the goodness of fit of the regression for \tilde{v} on i_n and p . Substituting out λ and λ_I from equation (5), we find that γ is a function of β and exogenous parameters.

Monopolistic competition. Based on this baseline model, if we further assume $\tau_e = \tau_E/N$ for some constant τ_E , then, for large N , an informed speculator's signal becomes so noisy that it contains only a small amount of information. Even with this situation, each informed trader still executes their tiny monopolistic power, and in this sense, this is a monopolistic competition model.

We regard this specification fits more to reality: since each trader has similar views on the announcement results, an increasing number of informed traders does not necessarily lead to more information in the market. We set this monopolistic competition model as a benchmark. Assuming $\tau_e = \tau_E/N$, equation (6) is now rewritten:

$$\rho\beta^3 + \frac{\tau_E}{N-1}\beta^2 + \frac{\sigma_z^2\tau_E}{N(N-1)}\rho\beta - \frac{(N-2)\sigma_z^2\tau_E^2}{N^2(N-1)^2} = 0. \quad (8)$$

Still, we can find a unique solution for β .

2.2 Implication of the model

Measures of market quality. Two endogenous variables are of our interest regarding market quality: liquidity and price discovery. Liquidity is defined as the inverse of price impact λ per unit of unexpected order flow; it is easy to find its empirical equivalent. We use effective bid-ask spread as an empirical proxy of illiquidity. Although this model does not explicitly define bid-ask spread, $p(z=1) - p(z=-1) = 2\lambda$ can be regarded as the equivalent of an effective bid-ask spread.⁹ Note that λ is a measure of *illiquidity* rather than liquidity. We use the inverse of it for the definition of the measure of *liquidity*.

Price discovery is measured as

$$V \equiv \frac{\text{Var}(p - \lambda\tilde{z})}{\text{Var}(\tilde{v})} \quad (9)$$

Here we employ $p - \lambda\tilde{z}$ instead of p , in order to omit the effect of noise trading; we can interpret $p - \lambda\tilde{z}$ a mid-price rather than a transaction price.¹⁰ Note that, substituting out p with (2), taking a variance and using equation (7), we obtain the expression of V as a function of $R^2 \equiv$

⁹ This follows the standard treatment which can be found in, for example, Bondarenko (2001).

¹⁰ This definition of price discovery appears that it is defined on price levels. But p and \tilde{v} are interpreted as the deviation from their unconditional expectation so that we can regard the measure as being based on returns: $\frac{\text{Var}(p - \lambda\tilde{z} - E[p - \lambda\tilde{z}])}{\text{Var}(\tilde{v} - E[\tilde{v}])}$.

$1 - \tau_v/\tau_I$: $V = (1 + \tau_v/\tau_E)(R^2)^2$. Thus, the price discovery measure reflects each trader's power to predict the future return.

Determinants of market quality. Two exogenous parameters are of interest regarding the determinants of market quality: total precision of informed traders' information τ_E , and the number of informed traders N .

Empirically, we use the R-squared measure, a goodness of fit of the surprise of news for predicting the after-the-news return, as a proxy for τ_E . We explore comparative statics for $\tau_E = N\tau_e$ instead of directly investigating R^2 , because R^2 is endogenously obtained in the model. The relation between these two elements are obtained as follows. Knowing the average expectation of the informed traders, $\bar{v} \equiv \tilde{v} + \sum_n \tilde{e}_n/N$, we can calculate the predictive power as $R_{-p,N}^2 \equiv 1 - \tau_v/\text{Var}(\tilde{v}|\bar{v})^{-1} = 1 - \tau_v/(\tau_v + N\tau_e)$. Thus, keeping other exogenous variables constant, an exogenous increase in $R_{-p,N}^2$ translates into an increase of $\tau_E = N\tau_e$. Our conjecture is that the predictive power of the news surprise to after-the-news return is parallel with $R_{-p,N}^2$, allowing us to proxy the sum of informed trader's precision τ_E .

Traders are assumed to trade according to their interpretation of the news just after the announcements; the number of informed traders N is regarded as the number of market participants. We use quote counts as a proxy for N .

Regarding these variables, implications of the model are summarized in the following propositions.

Proposition 1. Comparative statics for the number of informed traders

In the monopolistic competition model,

- (i) *Price impact:* $\frac{d\lambda}{dN} < 0$ if $\tau_E < \frac{(N-1)\sigma_z^2\rho}{2(N-2)}$.
- (ii) *Price discovery:* $\frac{dV}{dN} < 0$ if $\tau_E < \frac{(N-1)\sigma_z^2\rho}{2(N-2)}$.

Proof. Corollary 1 and 2.

The first statement is straightforward; liquidity improves as N becomes larger. This is ascribed to first, the direct effect of thicker demand/supply curve, and second, a reduction of asymmetric information. The first effect can be obtained because $N\beta$, the aggregate order aggressiveness, increases according to N . The second effect is obtained from the second statement.

The second statement means that an increase in N delay the price discover. This is because τ_I , the total information obtained from own signal and the signal extraction from price, decreases of N (Lemma 3). Since $V = (1 + \tau_v/\tau_E)(R^2)^2 = (1 + \tau_v/\tau_E)(1 - \tau_v/\tau_I)^2$, V changes in proportion to τ_I . The information obtained from own signal τ_e becomes small because of the assumption of fixed total information, $\tau_e = \tau_E/N$. Signal extraction improves as N increases (Lemma 2), but it does not overcome this first effect. Therefore, V can be increasing in N when we do not assume monopolistic competition model. We test this implication in section 4.2.

In short, in the monopolistic competition model, informed traders are more like uninformed traders because their private information becomes trivial. As a result, an increase in the number of informed traders results in the similar outcome to an increase of uninformed traders.

Proposition 2. Comparative statics for the precision of signal τ_E

In the monopolistic competition model,

(i) *Price impact: For any finite N, τ_v, σ_z^2 , there is sufficiently large τ_E where $\frac{d\lambda}{d\tau_E} < 0$.*

Also, for any finite τ_E , there is sufficiently large τ_v where $\frac{d\lambda}{d\tau_E} > 0$.

(ii) *Price discovery: $\frac{dV}{d\tau_E} > 0$.*

Proof. Lemma 7 and 8.

The second statement is straightforward; more precise signal makes traders more aggressive in their information-based trading, which leads to more informative prices, followed by better price discoveries.

For obtaining the intuition for the first statement, it is useful to consider two extreme cases;¹¹ when R^2 is close to one and when R^2 is much smaller. When R^2 is close to one (and τ_E is very large), traders are more certain about the true fundamental and they feel less risky to trade. As a result, they tend to provide liquidity (i.e., λ becomes small). When R^2 is much smaller, traders are more sensitive to the other trader's private information and they feel riskier to trade. As a result, they tend to provide less liquidity (i.e., λ becomes large). Thus, an increase of τ_E (and R^2) has two opposite effects on liquidity depending on the levels of exogenous parameters and τ_E itself. It is an empirical question whether the liquidity is increasing in τ_E or not, which will be examined numerically in section 4.2.

In summary, price discovery will be facilitated by the informativeness of the private signal (Proposition 2-(ii)). The effect of the number of traders, the implications of Proposition 1-(i) and 1-(ii), can depend on parameters level but they are valid in a fairly wide range of parameters. Effect on liquidity, or Proposition 2-(i), may be sensitive to the value of exogenous parameters. In section 4.2, we will match exogenous parameters to data, and demonstrate the comparative statics numerically.

3 Empirical Analysis

Our empirical analysis is implemented as follows. First, we estimate the impact of news surprise on after-the-announcement Forex returns. A key finding on this estimation is that the predictive power of news to future Forex returns is time-varying. As we mentioned in section 2.2, the

¹¹ We gratefully acknowledge that this intuition is suggested by Martin Evans.

predictive power can be interpreted as a proxy of precision of trader’s private information, which is an important determinant of price discovery. For second-step regressions, we regress the price discovery measure and liquidity recovery measures on such theoretically motivated determinants, time trend, and control variables. We can interpret the results based on the theoretical model, and the coefficients on the time trend can be interpreted as market quality improvements that cannot be explained by the factors considered in the regression model.

3.1 High-Frequency data

In this section, we describe the dataset and its handling. Our main dataset includes firm (ready to transact) quotes and actual deal prices and volumes within the trading platform of ICAP EBS.¹² The data cover the quotes and transactions from January 1998 to December 2017, but the extent of details depend on the technical development of the platform. From January 1998 to December 2005, EBS provides “EBS Ticker Historical Data,” which records prices of deals and best quotes; and from January 2006 to December 2017, it provides EBS Level-5 data, which additionally record deal volumes as well as information regarding limit order book. The entire dataset allows us to examine long-term changes in Forex market quality, and recent detailed data are utilized to examine the market quality determinants.

ICAP EBS has broadened the universe of currency pairs over years, but we focus on EUR/USD for this analysis. Unless otherwise noted, all the empirical results are in EUR/USD. As a robustness check, we will use seven other currency pairs to execute panel regressions; AUD/USD, GBP/USD, NZD/USD, USD/CAD, USD/CHF, USD/CNH, and USD/JPY.

¹² Market participants in the interbank market are mostly financial institutions. In the literature of Forex market microstructure (e.g., Evans and Lyons (1999)), order flows sometimes refer to “customer” order flows which is submitted to such financial institutions from non-financial companies. In our dataset, such orders are not available. All the transactions are among traders who has a contract to participate in the EBS interbank market.

Data on deals. Order submission and matching in the actual trading environment is in real-time, but the recording omits certain deals that show multiple transactions within a certain time slice. The observation at time-stamp t contains the deals that occur between $t - 1$ and t . The minimum time slice, the difference between $t - 1$ and t , varies from a second (in earlier years in the sample) to 100 milliseconds (recent years in the sample) depending on the period.¹³ The transaction prices are the most extreme ones (highest paid and lowest given) during the time slice. The minimum tick size also varies depending on the period.¹⁴

Data on quotes. The quote information contains the limit order prices and volumes up to ten steps (tenth best) of the limit order book. The observation is a snapshot of the limit order book, which is recorded every time when any change occurs in the book. The bid-ask spread is calculated based on the best bid and ask prices. The dataset also contains quote counts, the number of traders who are submitting limit orders at each step of the book. Note that, the data acquisition occurs for each side of the book separately; for some time, only one-side of the book is updated. In this analysis, for making correct filtration for each observation, we use the sample when the both sides of the book update simultaneously.

There are some caveats regarding the dataset. First, EBS allows negative spreads: the best ask price can be lower than the best bid price. This situation happens when the two entities in the book do not have credit lines between them. An arbitrage opportunity for this negative spread is discussed in Ito et al. (2012). Second, when an observation has both a deal and a quote in the same time slice, the dataset does not specify the order of each transaction. We need to estimate the order of transactions. Lastly, the minimum time slice in the EBS market is 100 milliseconds, which

¹³ The grid of time-slices has changed during the following periods: “one second” before January 22, 2008, “a quarter-second” from January 22, 2008 to August 31, 2009, and “a 100 milliseconds” from August 31, 2009 to present.

¹⁴ The minimum tick size was traditionally four decimal places for EUR/USD (or 0.1 cent), which is called one pip. It was decimalized (i.e., five decimals, 0.01 cent) on March 7, 2011 and then rolled back to half pips after September 24, 2012.

is coarser than that of the U.S. equity market. With this wide time grid, it is difficult to construct high-frequency trading activity. Chordia, Green, and Kottimukkalur (2018) examine the price discovery of 100-millisecond returns against two-second returns. For studying the influence of high-frequency traders in the Forex market, we need a more detailed dataset. We leave it for future research.

Table 1 shows the descriptive statistics of the data around macro announcements.

[Table 1 is inserted here]

In recent years, the frequency of quote revision has increased, but it is partly due to technical developments in data acquisition that occurred in 2008. After 2008, the frequency is mostly flat. Quote counts and depth dropped in 2011, but again, this is ascribed to the changes in minimum tick size (from 0.1 cent to 0.01 cent, effective from March 2011). Subsequent recovery is also explained by the rollback of minimum tick size (from 0.01 cent to 0.5 cent, effective from September 2012). Such discontinuous changes in data potentially affect our analysis. For its treatment, we divided the quote counts and depth by the max offer limit order price minus the min of the bid limit order price.

Another notable change is that the deal volume decreases in recent years. This is not only in the ICAP EBS market, but also in a BIS Central Bank Survey that shows that the OTC forex exchange turnover of EUR/USD also dropped globally. BIS survey points heightened activity in Japanese yen against the background of monetary policy developments (BIS Triennial Central Bank Survey 2016).¹⁵ The level of liquidity, measured by bid-ask spread and price impact, barely shows time trends, but a relative imbalance of order flows rose after 2014.

¹⁵ <https://www.bis.org/publ/rpfx16.htm>

3.2 News announcements

Government agencies in the U.S., Japan, and the EU announce macroeconomic statistics (e.g. GDP, unemployment, inflation rate and others) at a pre-announced day and time. Some are monthly; some are quarterly. While some announcements are published when the US equity markets are closed, Forex markets are open 24-hour a day, and we can obtain a rich sample of the market reaction to the news.

For gauging the market reaction, we define a “surprise” part of the announcement. For that, a pre-announcement expectation to each news is needed. Several days before the announcement, Bloomberg compiles “forecasts” by market participants and discloses the median. In accordance with these forecasts, we define the news surprise as

$$\frac{X_{t,\text{actual}} - X_{t,\text{forecast}}}{\text{Std}_{t-1}(X_{t,\text{actual}} - X_{t,\text{forecast}})}, \quad (10)$$

where X_t is a certain announced statistic made in public at time t .¹⁶ The denominator is a standard deviation of the surprise estimated by using observations up to $t - 1$. This is implemented by a rolling estimate with 12-period windows.

Impact of surprise on returns. First, we regress 10-minute after-the-news returns on the surprises of news:¹⁷

$$\Delta S(t, u + 600) = \sum_{i=1}^n \alpha_i N_i(t, u) + \epsilon(t, u). \quad (11)$$

- (t, u) pins down each announcement timing indexed by date t and its intraday time u . u is in intraday seconds from 00:00 New York time.

¹⁶ In this paper, timing of each announcement is indexed by t . Roughly, t corresponds to announcement day, but sometimes there are multiple news announcements in the same day. Furthermore, there is more than one news at the same t . We regard multiple news at the same time as one sample chunk and do not duplicate the sample by the number of simultaneous news.

¹⁷ A similar regression analysis can be found in literature: Andersen et al. (2003), Hashimoto and Ito (2010), Fatum, Hutchison, and Wu (2012), Cheung, Fatum, and Yamamoto (2019), Andersen et al. (2007).

- $\Delta S(t, u + 600)$: exchange rate return (log-difference) in basis points, from time u to $u + 600$ on day t . The rate is defined by mid-quotes.
- $N_i(t, u)$: surprise of i -th macroeconomic news statistics on day t at time u , defined as (10). If there are no news on i -th statistics, it takes zero.
- n : the number of different indicators. We consider total $n = 48$ indicators, most of them are consistently available since 1998.¹⁸
- The coefficients are estimated by OLS with Newey-West standard errors (lag of 10).

In this estimation, we extract 10-minute after-the-news returns to define the dependent variable.¹⁹ Another way to estimate the impact is to use all the transaction data and control for intraday seasonality, as Andersen et al. (2007) implemented. In our study, we limit the sample around the announcements because the sample size of high-frequency data tends to be too large.

After the regression, using the absolute value of the regression residuals, Andersen et al. (2007) estimate a time-series model for the disturbance volatility for each asset (i.e., two-step weighted least squares procedure). Similarly, we utilize the residual component for defining price discovery, which will be described in section 3.4. The statistical model in Andersen et al. (2007) explains the volatility with its own lags, news surprise, and calendar time patterns. In contrast, we focus on the dynamics of volatility, and investigate economic determinants for the dynamics and its development over time.

The regression results are summarized in Table 2, which is associated with the median of absolute after-the-news returns and the number of quote revisions.

¹⁸ Note that Chicago PMI and U. of Michigan Sentiment data are systematically released early to subscribers. In this research, however, omitting these indicators has little effect on the results, and particular treatments are not made for overall estimates.

¹⁹ We check the average path of prices after the announcement (Figure 7) and determined that 10 minutes is appropriate. At this point, Evans and Lyons (2008) argues that it takes more than days for the Forex price to integrate the news contents. In our paper, we limit the duration for focusing on microstructure of the market.

[Table 2 is inserted here]

The qualitative results are consistent with existing literature; the most influential statistics include labor statistics (changes in non-farm payroll and unemployment rate), GDP statistics (advance estimate), and some leading and sentiment indicators such as ISM Manufacturing and consumer confidence, etc. After the announcement of such influential statistics, the magnitude of jumps in the forex rates and quote revisions tend to become large.

As Fatum, Hutchison, and Wu (2012) report, the impact of the news on forex rates depends on the macro-financial environment. So does the explanatory power of the news on forex rates. To show this, we executed a rolling regression of (11) and presented the results in the last column in Table 1. After the global financial crisis, predictability was sluggish until 2013.

3.3 Market quality measures

Price discovery and liquidity recovery. As introduced in equation (9), we measure the price discovery as a variance ratio:

$$PD_{k,T} = \sqrt{\frac{\text{Var}(\Delta S(t, u + k))}{\text{Var}(\Delta S(t, u + T))}}. \quad (12)$$

k and T maintain $k \leq T$.²⁰ This definition is essentially the same as the one used in Chordia, Green, and Kottimukkalur (2018), where they use a ratio of absolute return for measuring the speed of market reaction. We take a square root of the variance ratio to match the definition based on absolute returns.

²⁰ Alternatively, we can use the weighted price contribution (WPC), and the R-squared of unbiasedness regressions. The statistical property of each measure is examined by van Bommel (2011). We use variance ratio because its theoretical equivalent is easily available. In fact, the autocorrelation of return seems small (as described in Figure 7 and the potential bias of using WPC may not be large.

This variance ratio gauges the convergence of volatility clustering. In the literature, Andersen et al. (2007) investigate the residual of the regression (11) rather than using the sequence of raw returns. Following this, we utilize the prediction by (11) for defining price discovery for eliminating potential confusion of interpretations, which will be specified in next section.

We measure liquidity recovery as an inverse of bid-ask spreads (BAS):

$$LR_{k,T} = E \left[\left\{ \frac{BAS(t, u + k)}{BAS(t, u + T)} \right\}^{-1} \right]. \quad (13)$$

We take the inverse because a bid-ask spread is a measure of illiquidity rather than liquidity. Among many different liquidity measures, we employ bid-ask spread because (i) it needs only quotes information and constructing long-term statistics is easy; the other measures such as depth and price impacts need limit order book or transaction volume information, which is not available before 2006. And (ii) other popular descriptive liquidity measures such as depth are not stable after announcements. Bid-ask spread, however, may represent only a part of liquidity. In the regression analysis in later sections, we employ volume-weighted average of limit prices to reflect the information of limit order book.

Both measures approach one as $k \rightarrow T$; each path shows the speed of price discovery and liquidity recovery. Note that price discovery corresponds to the settlement of volatility clustering, but it does not mean return predictability.

Figure 3 plots the year-by-year path of each.²¹ We can observe that liquidity recovery is much faster than price discovery.

[Figure 3 is inserted here]

²¹ Parameterization: for price discovery, we set $T=600$; for liquidity recovery, we set $T=300$. For a comparison with Chordia, Green, and Kottimukkalur (2018), we show the time-series change of price discovery using raw returns rather than residual. For liquidity, we use effective bid-ask spread rather than volume-weighted bid-ask spread, this is due to data availability. Limit order book information is only available after 2006.

Figure 3 shows that both measures do not necessarily improve over time. For price discovery, the slowest year is 2011 and the fastest year is 2003 and 2017. For liquidity recovery, most years are slower than the year 2011. To emphasize this point, we provide Figure 1, which shows the year-over-year changes in both measures. In fact, there is little evidence that both measures improve over time.

3.4 Regression analysis

Benchmark specification. Based on the observation above, our next question is (i) what the determinant of such market quality is, and (ii) whether the market quality has improved, after taking such elements into account. As we discussed theoretically, we consider two key drivers as determinants: the number of traders and the precision of the trader’s private information. The coefficients on time trend are considered to be market quality developments.

For the proxy of the number of traders, we use *quote counts*: the number of traders submitting limit orders. As Proposition 1 suggests, they provide liquidity, but whether it contributes to price discovery depends on the quality of private information. The empirical specification is a simple linear regression model, with its LHS including the price discovery measure or the liquidity recovery measure. Estimated coefficients can be compared with Proposition 1-(i) and 1-(ii). The liquidity should be positively related to the number of traders. If traders behave like an uninformed trader, price discovery is negatively related to the number of traders.

The precision of the trader’s private information is proxied by the R-squared measure. If traders have rich information that helps in predicting the impact on Forex rates, price discovery would be fast. As we have mentioned in section 3.2, the explanatory power of the news surprise for the Forex return depends on the monetary and macroeconomic situation, and it is time-varying. In fact, a rolling regression of equation (11) produces time-varying R-squared (the result is available in

Table 1), which can be regarded as a proxy for the precision of the trader’s private information.²² The precision of information is positively correlated with the speed of price discovery after the announcement (Proposition 2-(ii)), but it’s effect on liquidity depends on other variables (Proposition 2-(i)).

Corrections on the market quality measures. A possible criticism to use the price discovery measure (12) is that the Bloomberg survey is considered as public information and the speed of convergence can be explained by a simple market efficiency argument. Also, for implementing regression analysis, the measure must be realized at each macro announcement. In order to circumvent these problems, we introduce a modified definition of price discovery measure as follows.

$$PD_{t,k,T}^R = \frac{|\Delta S(t, u + k) - \widehat{\Delta S}(t, u + T)|}{|\Delta S(t, u + T) - \widehat{\Delta S}(t, u + T)|},$$

where $\widehat{\Delta S}$ is the prediction by the equation (11); we estimated (11) by rolling regression with 1,000 overlapping windows. The measure is calculated for each announcement time; it is a realized measure for previous definition (12). Here we set $T = 600$, because of the observation that convergence of price ceases in 10 minutes on average (see Figure 7 for the average price path after the announcement).

For liquidity recovery measure, we employ volume-weighted bid-ask spread (VWBAS) instead of bid-ask spread. VWBAS is defined as a spread of bid and ask prices where each bid (ask) price is an average of limit bid (ask) prices weighted by their limit order volumes. With VWBAS, we can exploit the information of limit order book. Now, the liquidity recovery measure is defined as

$$LR_{t,k,T}^R = \left\{ \frac{VWBAS(t, u + k)}{VWBAS(t, u + T)} \right\}^{-1}$$

We set $T = 300$. The bid-ask spread is affected by the tick size changes implemented in 2011 and 2012, but the effects can be offset by taking the ratio of spreads at different times.

²² R-squared measure is calculated for news-by-news basis to avoid non-observatory cases. Please refer to the caption of Table 4 for the detailed construction.

Time trend. In the regression equation, we assume a linear time trend which is scaled from zero to one. The coefficient on the trend can be interpreted as improvements of market quality that cannot be ascribed to the other independent variables.

Control variables. For independent variables, other than the two proxies and time trend introduced above, we also add control variables: information on the limit order books (the depth and frequency of quote revisions are both normalized by tick size²³), the magnitude of news surprises, and the year dummies for 2009, 2010, 2011, and 2012,²⁴ order imbalance and cumulative volumes. The time-gap of the tick-by-tick data is not equal. For constructing these variables, we interpolated the data by replacing the missing variables with their last observances. Sampling timing is as follows. The quote counts and depth are a snapshot variable at $(t, u + k)$, the quote revision is the sum of $[(t, u - 15\text{min}), (t, u + 15\text{min})]$, and the order imbalance and cumulative volumes are the sum of $[(t, u), (t, u + k)]$. The correlation coefficients of independent variables are presented in Table 3.²⁵

[Table 3 is inserted here]

Regression results and interpretations. The regression results are presented in Table 4 and Table 5. For a robustness check, we tried different k , the duration defining the snapshot of price discovery and liquidity recovery measure. As Scholtus, Van Dijk, and Frijns (2014) stress, the speed matters for trading based on news, and the demography of market participants may change in different time slice k . In the last column in each table, we also presented a result without controls.

²³ See section 3.1 for this treatment.

²⁴ These years had volatile macroeconomic conditions in the wake of the Global Financial Crisis.

²⁵ We changed the timing of the sampling of quote counts and depth, and obtain qualitatively the same results. These variables vary symmetrically between before and after the announcement; they gradually decrease toward the announcements, and gradually recover after the announcements.

[Table 4 and Table 5 are inserted here]

In Table 4, the regression for price discovery, the coefficients on R-squared measures are significantly positive, but those of quote counts are negative. Conversely, in Table 5, the regression for liquidity recovery, the coefficients on R-squared measures tend to be negative, but those of quote counts are positive significant. These results hold true regardless of the inclusion of controls, but the significance may depend on time slice k ; typically, the coefficients tend to be less significant for large k . This tendency is natural because price movement approaches to more like random walk as k becomes large. Even if we adjust this by discounting the denominator of the measures by k , we can obtain qualitatively the same results.

As we discussed in section 2.2, the R-squared measure represents asymmetric information in the market. When it is large, informed traders have a greater informational advantage regarding the true reaction of price to the news. Hence, the incorporation of information into prices is more rapid (or $dV/d\tau_E > 0$), but uninformed traders are reluctant to trade, and the liquidity decreases (or $d\lambda/d\tau_E < 0$). If each informed trader has a distinguished interpretation of news, the increase in traders would facilitate price discovery; but this is not the case. Conversely to this conjecture, the increase in the number of traders delays price discovery (or $dV/dN < 0$), implying that informed traders have poor information and act more like uninformed traders.

The trend coefficient is positive and statistically significant for price discovery but is negative significant for liquidity recovery. The trend variable is normalized by the total time period, and it lets us know the percentage changes in price discovery/liquidity recovery that is not ascribed to the variables considered in this regression. From 2006 to 2017, the improvement in the speed of price discovery is up by about 24% points at maximum, but liquidity recovery is down by about 28% points at maximum. Note that each independent variable has its own trend component, as shown in Table 3. The regression coefficient on trend is strongly affected by the other independent

variables. For an alternative analysis, not reported here, we implemented the regression with de-trended independent variable. We can obtain the similar results for each independent variable and overall qualitative results do not change, but the magnitude of trend coefficients becomes less than ten.

For control variables, the pattern of coefficients is not very consistent between Table 4 and Table 5. The number of quote revisions negatively affects both measures. This is somewhat surprising because it implies that an active trading environment does not necessarily improve market quality. The order-imbalance per total deal volumes, usually regarded as another measure of informed trading,²⁶ positively affects price discovery, which is consistent with theoretical intuition. But its influence on liquidity is weak.

Overall, faster price discovery does not necessarily imply faster liquidity recovery, and vice versa. Since the R-squared measure and quote counts affect each measure in the opposite direction, a negative correlation between price discovery and liquidity is a possibility. In fact, time series plots in Table 1 consistently support this observation.

Our result may be contrasted to the existing literature in that high-frequency algorithmic traders are considered to enhance both price discovery and liquidity provision (Brogaard, Hendershott, and Riordan (2014), Chordia, Green, and Kottimukkalur (2018)). Because our dataset does not contain clear identification of high-frequency traders, our results, unfortunately, do not necessarily directly describe the behavior of high-frequency traders. But the implication of our empirical result suggests that the informational advantage of fast-moving traders should not be taken for granted.

4 Robustness Analysis

²⁶ Kaul, Lei, and Stoffman (2008) and Easley, O'Hara, and Lopez de Prado (2012).

In this section, we conduct supplemental analyses that support the robustness of the results in section 3. First, we show extended regression analyses on price discovery. Second, we examine numerical experiments of the theoretical model. Lastly, we check whether there is market inefficiency after news announcements.

4.1 Extended analyses on price discovery

Some arguments should be discussed to justify the robustness of the results in section 3.4. First, it is questioned whether the technological development of the EBS matching system can improve price discovery or not. The time trend in Table 4 is positive, implying that such an association is likely. But this can be a natural consequence of the increase in transaction activity. In the regression, we add the number of quote revisions to control activity effects. A more direct way to establish a control is to use tick-time instead of calendar time for constructing the dataset.²⁷

Second, we check whether news types affect the regression results. Some news announcements, such as labor statistics, are far more influential to the forex rates, as market participants pay more attention. It is possible that market reactions may also differ.

Third, we check whether these empirical results are applicable to other currencies than EUR/USD. To investigate this, we expand the universe of currency pairs to include seven other USD currency pairs and attempt a panel regression analysis.²⁸

Price discovery in tick-time. The setup for a regression model is identical to that in section 3.4 except that the price discovery measure is defined based on tick-time: $\frac{|\Delta S(t,u+n) - \widehat{\Delta S}(t,u+T)|}{|\Delta S(t,u+T) - \widehat{\Delta S}(t,u+T)|}$, where the numerator is now indexed by tick-time n from the announcement time (t, u) . Thus, we measure the speed of price discovery not by the calendar time but by the number of quote revisions. With

²⁷ A similar idea can be found in Barclay and Hendershott (2003). They extend the weighted-price contribution, a standard measure for price discovery, by normalizing it via the number of transactions.

²⁸ We stick to USD pairs because US macro news affects the market the most.

this formulation, we can measure the ability of price discovery on a per-transaction basis. The regression results are shown in Table 6.

[Table 6 is inserted here]

The results are similar to that of Table 4, the calendar-time regression. R-squared measures have a positive impact on the price discovery process, and quote counts have a negative impact. The coefficients on time trends are slightly smaller in their magnitude but still positive and significant, indicating that the capacity of improving price discovery has moderately increased.

News types. Do the speed of price discovery and liquidity recovery differ for different news types? We implement regressions with news-type dummy which is interacted with three main determinants. To this aim, two major news announcements are selected: labor statistics announcement and GDP (advance estimate) which are the most influential to the changes in forex rates (see Table 2).

[Table 7 is inserted here]

The regression results are presented in Table 7. For three determinants, R-squared measures, quote count, and time trend, implications for quote count and time trend are unchanged from the previous regression results. In fact, they tend to be more exaggerated in these news announcements: larger quote counts make price discovery much slower and make liquidity recovery much quicker; and in recent years, price discovery becomes much faster but liquidity recovery much slower.

Coefficients on R-squared measures, however, show opposite signs to the non-dummy coefficients. GDP dummies are not significant, but Labor statistics dummies are significant. This

suggests that higher predictive power leads to slower price discovery and better liquidity as if traders are uninformed.

Price discovery: A panel regression analysis. We apply the same analysis to other seven currency pairs: AUD/USD, GBP/USD, NZD/USD, USD/CAD, USD/CHF, USD/CNH, and USD/JPY.

[Figure 4 is inserted here]

Figure 4 corresponds to the analysis in Figure 1 and Table 4. The bottom panel is the result of a fixed effect panel regression for the speed of price discovery. For readability, we only present a plot of the t-statistics of independent variables in different k-sec samples.

Similar to the results in Table 4, R-squared measures have large positive coefficients, and quote counts have large negative ones. The time trend variable is positive. Thus, we can confirm that major findings are still valid. Contrary to the results in EUR/USD, the number of quote revisions is positive in this panel regression. This indicates that the cross-sectional differences for price discovery in major currency pairs are faster than those for minor currency pairs.

The top two panels plot the time-series development of price discovery and liquidity recovery for each currency pair. We can observe that the development of market qualities in EUR/USD is in line with other currency pairs: price discovery has been improving but experienced a sluggish movement between 2009 and 2012. Liquidity recovery tends to be slow in recent years for all pairs except for an outlier: USD/CNH. As indicated by the regression results, the major currency pairs tend to show faster price discovery, but EUR/USD does not necessarily show the best market quality. As observed in EUR/USD, the negative correlation between price discovery and liquidity recovery is also observed in these panels.

4.2 Numerical experiments

Our empirical findings are summarized as follows. After macro announcements,

- the increase in the number of traders does not improve price discovery but increases liquidity.
- the higher explanatory power of the news for after-the-announcement returns helps price discovery but reduces liquidity.

But the theoretical analysis shows that some of the results depend on the level of parameter values and the assumption of monopolistic competition. To check whether the theory correctly predicts the empirical results, we conduct numerical experiments on these findings.

Numerical experiments on price discovery. First, we explore how the assumption of monopolistic competition changes the relationships between the number of traders and price discovery. For a numerical experiment, we set up a smooth transition from the monopolistic competition case to the imperfect competition case. The precision of interpretation noise τ_e is now replaced as

$$\tau_e(N, k) = w(k) \frac{\tau_E}{N_{\text{fix}}} + (1 - w(k)) \frac{\tau_E}{N}, \quad w(k) = \frac{1}{1 + \exp(-(k - k_c))}.$$

Here, $N_{\text{fix}} = 4$ and $k_c = 5$ are assumed to be fixed. k is a parameter that shifts the information structure from a monopolistic competition case (i.e., $\tau_e(N) = \tau_E/N$ when $k = -\infty$) toward an imperfect competition case (i.e., $\tau_e = \tau_E/N_{\text{fix}}$ when $k = \infty$). Other model parameters are specified as follows.

- $\rho = 1$ for normalization.
- $\tau_v = 25^{-2}$, which accounts for the standard deviation of the total return of 25bp.
- $\tau_E = \tau_v \times 0.5$, which is set to achieve $R^2 = 1 - \tau_v/(\tau_v + \tau_E) = 1/3$.

- $\sigma_z^2 = 0.001 \times \sigma_v^2 + 0.00005 \times \tau_E$, which is set for keeping the informed trading ratio $\frac{N\beta\sigma_v + \beta \sum_n \tilde{e}_n}{N\beta\sigma_v + \beta \sum_n \tilde{e}_n + \bar{z}}$ around 10%.

These parameters are set to simulate the reality that is observed in Table 1 and Table 2. Under this model calibration, Figure 5 plots the variance ratio, or price discovery measure, $V(N) \equiv \text{Var}(p(N) - \lambda(N)\bar{z})/\text{Var}(\bar{v})$ as a function of N .

[Figure 5 is inserted here]

In the imperfect competition region, or when k is large, the variance ratio initially decreases, but then increases when N becomes large. In the monopolistic competition region, or when k is small, the variance ratio is decreasing in N , corresponding to our empirical finding. Thus, we can confirm that the assumption of monopolistic competition is necessary to replicate the reality.

Numerical experiments on liquidity. Next, we check the theoretical results on liquidity. Proposition 1 states that the liquidity is increasing in N (or price impact λ is decreasing in N) if τ_E is sufficiently small compared with σ_z^2 . On the other hand, as in proposition 2, the changes in price impact with respect to τ_E can depend on parameter values. Figure 6 shows the simulation result with the same parameter setting as above except for τ_E ; it plots the inverse of λ , or a measure of liquidity, as a function of N with different parameter values of τ_E . Now that, $\tau_E = \tau_v \times 0.5 \times k$, and $k = 0.5, 1.0, \dots, 5.0$.

[Figure 6 is inserted here]

Liquidity increases in N as the theory suggests, but it is not monotonic with respect to the changes in the precision of noise τ_E . When N is small, the increase of τ_E results in the improvement of

liquidity; when N is large, the effect reverses. Empirically, we observe the latter effect, namely, it applies to the upper-right region, but their significances were not very sounding. In fact, the numerical analysis shows that the effect of τ_E on liquidity might be sensitive to the parameter values.

4.3 Predictability of after-the-news return

Post-earnings announcement drift is a well-known anomaly in stock markets (Bernard and Thomas (1989)). Also, a significant autocorrelation of return potentially biases the estimation of variance ratio (van Bommel (2011)). In this section, we examine whether such a drift, or a positive autocorrelation of returns, is observed after the macro announcement in the forex market. This is used to determine the time interval to calculate after-the-news return.

Testing methodology is straightforward: calculating cumulative returns and autocorrelation of returns. Following the standard methodology for showing the post-announcement drift, we can calculate the cumulative return conditional on news surprise. Each announcement sample is stratified into seven bins by their surprise, and then the cumulative return from -20 minutes to +3 hours from the announcement is calculated. An alternative method is to calculate the simple autocorrelation of return after announcements as $\text{Corr}_t(S(t, u + k) - S(t, u), S(t, u + 3\text{hours}) - S(t, u + k))$, $k = 100\text{sec}, 200\text{sec}, \dots$. The autocorrelation is calculated by each year sample. Figure 7 describes the two results.

[Figure 7 is inserted here]

The results indicate that there is little evidence of post-announcement drift: the price path conditional on news surprise does not have a clear drift. Rather, there is a reversal after the announcements when the magnitude of news surprise is small. In fact, the autocorrelation shows a

mild negative correlation. For the overall sample, the autocorrelation is around -5%. This result, however, claims the absence of a drift *on average*, and does not state the absence of drift *conditional* on the arrival rate of informed traders (Vega (2006)). Thus, we can confirm the absence of apparent inefficiency after the news announcement; a more precise investigation is left for future research.

5 Conclusion

In the paper, we investigate the forex market reaction to macro announcements, both in pricing and liquidity, and examine whether the market quality has improved over time. Novelties of our research are three-fold: first, data are long-term (from 1999 to 2017) as well as high-frequency (tick-by-tick transaction data); second, the market impact of news is studied with particular focus on the issues of market quality; and third, the empirical methodology and interpretation are fully grounded in a classical market microstructure theory.

Overall, empirical evidence shows that market quality in the forex market has not necessarily improved in recent years: the speed of price discovery has an improving trend, but liquidity recovery does not. Liquidity recovery was fast enough in 1999, leaving only small room for improvement. Moreover, these two measures are negatively correlated. When some traders have private information (or precise interpretation of the news statistics), their trades increase price discovery but other traders are reluctant to provide liquidity to avoid risks of trading against the informed. When most of the traders do not have private information, their trades do not contribute to price discovery but provide liquidity. In particular, our regression analysis shows that an increase in the number of traders does not improve price discovery just after the announcement; fast-moving traders are likely to be uninformed. This result suggests that high-frequency trader's informational advantage in interpreting macro statistics announcement should not be taken for granted.

Appendix: Proof of propositions

Lemma 1.

$$\frac{d\beta}{dN} \frac{N}{\beta} > -1. \text{ If } N \geq 4, \quad \frac{d\beta}{dN} \frac{N}{\beta} < 0.$$

Proof. Multiplying the cubic equation (8), $\rho\beta^3 + \frac{\tau_E}{N-1}\beta^2 + \frac{\sigma_z^2\tau_E}{N(N-1)}\rho\beta - \frac{(N-2)\sigma_z^2\tau_E^2}{N^2(N-1)^2} = 0$, by N^3 , we obtain

$$\rho(N\beta)^3 + \frac{N\tau_E}{N-1}(N\beta)^2 + \frac{N\sigma_z^2\tau_E}{(N-1)}\rho N\beta - \frac{N(N-2)\sigma_z^2\tau_E^2}{(N-1)^2} = 0. \quad (14)$$

As N increases, the coefficients on the first three terms weakly decrease, while the last term increases.

Therefore $N\beta$ should be increasing in N . Since $\beta > 0$, we obtain $\frac{d\beta}{dN} \frac{N}{\beta} > -1$.

Next, applying the implicit function theorem to $F(N, \beta) \equiv \rho\beta^3 + \frac{\tau_E}{N-1}\beta^2 + \frac{\sigma_z^2\tau_E}{N(N-1)}\rho\beta - \frac{(N-2)\sigma_z^2\tau_E^2}{N^2(N-1)^2} = 0$, we have

$$\frac{d\beta}{dN} = -\frac{\partial F/\partial N}{\partial F/\partial \beta} = -\left(\frac{\partial F}{\partial \beta}\right)^{-1} \left(-\frac{\tau_E}{(N-1)^2}\beta^2 - \frac{(2N-1)\sigma_z^2\tau_E\rho}{N^2(N-1)^2}\beta - \frac{(-3N^2+9N-4)\sigma_z^2\tau_E^2}{N^3(N-1)^3} \right), \quad (15)$$

where $\frac{\partial F}{\partial \beta} = 3\rho\beta^2 + \frac{2\tau_E}{N-1}\beta + \frac{\sigma_z^2\tau_E}{N(N-1)}\rho > 0$ is straightforwardly obtained because $\beta > 0$. Using $F(N, \beta) = 0$, the last term of $\partial F/\partial N$ is substituted out. The resulting expression is a polynomial of β without constant terms, and each coefficient of β is greater than zero if $N \geq 4$. Thus, we have $\partial F/\partial N > 0$ if $N \geq 4$, and $\frac{d\beta}{dN} \frac{N}{\beta} < 0$ can be obtained. ■

Regarding Lemma 1, the equation for β in (5) indicates why the increase of N results in a decrease of β . There are two channels through which N affects the behavior of traders; its effect on liquidity (which is represented by λ in β) and price informativeness (which is represented by τ_I and φ). The liquidity channel is straightforward. Because the aggregate demand/supply function is linear in N , the increase of N makes the price inelastic with respect to the unit order flow: improvement of liquidity. In the imperfect competition model, when the market is more liquid (or the price is less elastic to the order flow), the traders become more eager to trade, i.e., β increases (this is implied by equation (5)).

The price informativeness channel is more complicated because it can increase as well as decrease β . If price becomes more informationally efficient and revealing, each trader becomes reluctant to trade on their signals, and β decreases. Because informational efficiency is increasing in N (i.e., $d\varphi/dN > 0$), the larger N reduces β . On the other hand, if total available information increases in N , the precision of signal for traders τ_I increases

accordingly. Then, trading on price information becomes less risky, traders become more aggressive, and β increases.

If we assume a monopolistic competition model, the second effect of informational channel is shut down, because total available information is fixed: $N\tau_e = \tau_E$. Therefore, information channel only reduces β , which can overcome the liquidity channel and an increase in N result in decreasing β .

Lemma 2.

$$\frac{d\varphi}{dN} > 0.$$

Proof. Differentiating the informational efficiency parameter $\varphi = \frac{(N-1)\beta^2}{(N-1)\beta^2 + \sigma_z^2\tau_e}$ with respect to N , we have

$$\frac{d\varphi}{dN} = \left((N-1)\beta^2 + \frac{\sigma_z^2\tau_E}{N} \right)^{-2} \left\{ \beta^2 + 2(N-1)\frac{d\beta}{dN}\beta - (N-1)\beta^2 \left(2(N-1)\frac{d\beta}{dN} - \frac{\sigma_z^2\tau_E}{N^2} \right) \right\}. \quad (16)$$

The first parenthesis is clearly positive. The second parenthesis reduces to $(2 - \frac{1}{N})\beta^2 + 2(N-1)\frac{d\beta}{dN}\beta$. After rearranging terms, we have

$$\frac{d\varphi}{dN} > 0 \Leftrightarrow \frac{d\beta/\beta}{dN/N} > -\frac{2N-1}{2(N-1)}.$$

Lemma 1 claims that this is satisfied in $N \geq 4$. ■

Lemma 3.

$$\frac{d\tau_I}{dN} < 0 \text{ if } \tau_E < \frac{(N-1)\sigma_z^2\rho}{2(N-2)}.$$

Proof. By definition (4), $\tau_I \equiv \tau_v + \tau_E/N + (N-1)\varphi\tau_E/N$. Differentiating with respect to N , we have

$$\frac{d\tau_I}{dN} = -\frac{\tau_E}{N^2} + \frac{d\varphi}{dN}(N-1)\frac{\tau_E}{N} + \varphi\frac{\tau_E}{N^2}.$$

Rearranging terms, we have

$$\frac{d\tau_I}{dN} < 0 \Leftrightarrow \frac{d\varphi}{dN} < \frac{1-\varphi}{N(N-1)}.$$

Substituting out $d\varphi/dN$ with the equation (16), this is rewritten:

$$\frac{d\tau_I}{dN} < 0 \Leftrightarrow \frac{d\varphi}{dN} < \frac{1-\varphi}{N(N-1)} \Leftrightarrow \frac{d\beta/\beta}{dN/N} < -1 + \frac{\sigma_z^2\tau_E}{2N(N-1)^2\beta^2}.$$

We have the expression for $d\beta/dN$ with the equation (15). Rearranging it yields

$$\frac{d\beta/\beta}{dN/N} = -\frac{-\frac{N\tau_E}{(N-1)^2}\beta^2 - \frac{(2N-1)\sigma_z^2\tau_E\rho}{N(N-1)^2}\beta - \frac{(-3N^2+9N-4)\sigma_z^2\tau_E^2}{N^2(N-1)^3}}{-\frac{\tau_E}{(N-1)}\beta^2 - \frac{2\sigma_z^2\tau_E\rho}{N(N-1)}\beta + \frac{3(N-2)\sigma_z^2\tau_E^2}{N^2(N-1)^2}}.$$

Here, the denominator is obtained by substituting $3\rho\beta^3$ out by $F(N, \beta) = 0$. Using this expression, after some calculations, we have

$$\begin{aligned} \frac{d\beta/\beta}{dN/N} &< -1 + \frac{\sigma_z^2\tau_E}{2N(N-1)^2\beta^2} \\ \Leftrightarrow \frac{\tau_E}{(N-1)}\beta^3 + \left(1 - \frac{3}{2}\right)\frac{\sigma_z^2\tau_E\rho}{N(N-1)}\beta^2 + \left(\frac{2}{N} - 1\right)\frac{\sigma_z^2\tau_E^2}{N(N-1)^2}\beta - \frac{\rho\sigma_z^4\tau_E^2}{2N^2(N-1)^2} &< 0. \end{aligned}$$

Again, we substitute out the last term by using $F(N, \beta) = 0$, and make sure each coefficient is negative. In the resulting expression, the coefficients on β^2 and β are clearly negative. The coefficients on β^3 is $\frac{\tau_E}{(N-1)} - \frac{\sigma_z^2\rho}{2(N-2)}$, which can be negative if $\tau_E < \frac{(N-1)\sigma_z^2\rho}{2(N-2)}$. ■

Corollary 1.

$$\frac{dR^2}{dN} < 0 \text{ and } \frac{dV}{dN} > 0 \text{ if } \tau_E < \frac{(N-1)\sigma_z^2\rho}{2(N-2)}.$$

Proof. $R^2 \equiv 1 - \tau_v/\tau_I$ is a definition. The sign of dR^2/dN coincides with $d\tau_I/dN$ because $\frac{dR^2}{dN} = \frac{\tau_v}{\tau_I^2} \frac{d\tau_I}{dN}$. $V \equiv (1 + \tau_v/\tau_E)(R^2)^2$ and $dV/dN > 0$ immediately follow. ■

Corollary 2.

$$\frac{d\lambda}{dN} < 0 \text{ if } \tau_E < \frac{(N-1)\sigma_z^2\rho}{2(N-2)}.$$

Proof. As equation (7) indicates, the price impact is $\lambda = \frac{1}{N\beta} \left(1 - \frac{\tau_v}{\tau_I}\right)$. Lemma 1 suggests that $1/N\beta$ is decreasing in N , and Corollary 1 suggests $1 - \tau_v/\tau_I$ is also decreasing in N . Thus λ is decreasing in N . ■

Lemma 4.

$$\frac{d\beta}{d\tau_E} > 0.$$

Proof. Applying the implicit function theorem to $F(N, \beta) \equiv \rho\beta^3 + \frac{\tau_E}{N-1}\beta^2 + \frac{\sigma_z^2\tau_E}{N(N-1)}\rho\beta - \frac{(N-2)\sigma_z^2\tau_E^2}{N^2(N-1)^2} = 0$, we have

$$\frac{d\beta}{d\tau_E} = -\frac{\partial F/\partial\tau_E}{\partial F/\partial\beta} = -\left(\frac{\partial F}{\partial\beta}\right)^{-1} \left(\frac{1}{(N-1)}\beta^2 + \frac{\sigma_z^2\rho}{N(N-1)}\beta - \frac{2(N-2)\sigma_z^2\tau_E}{N^2(N-1)^2} \right). \quad (17)$$

We can show $\partial F/\partial\beta > 0$ like Lemma 1. The last term is substitute out by $F(N, \beta) = 0$, and we have

$$\frac{\partial F}{\partial\tau_E} = -\frac{2\rho\beta^3}{\tau_E} + \left(\frac{1}{N-1} - \frac{2}{N-1} \right)\beta^2 + \left(\frac{1}{N(N-1)} - \frac{2}{N(N-1)} \right)\sigma_z^2\rho\beta < 0,$$

because each term in parentheses is negative and $\beta > 0$. Combining these, we obtain $d\beta/d\tau_E > 0$. ■

Lemma 5.

$$\frac{d\varphi}{d\tau_E} > 0.$$

Proof. By definition (4), $\varphi \equiv \frac{(N-1)\beta^2}{(N-1)\beta^2 + \sigma_z^2\tau_E/N}$. Differentiating with respect to by τ_E and rearranging terms, we have

$$\frac{d\varphi}{d\tau_E} = \left((N-1)\beta^2 + \frac{\sigma_z^2\tau_E}{N} \right)^{-2} \left\{ 2(N-1)\frac{d\beta}{d\tau_E}\beta\frac{\sigma_z^2\tau_E}{N} - (N-1)\beta^2\frac{\sigma_z^2}{N} \right\}.$$

The denominator is clearly positive, and it is enough to show the numerator is positive as well. Rearranging terms, we can show that the numerator is positive if and only if

$$\frac{d\beta}{d\tau_E}\frac{\tau_E}{\beta} > \frac{1}{2}.$$

In the LHS, substituting out $d\beta/d\tau_E$ by (17), we have

$$\frac{d\beta}{d\tau_E}\frac{\tau_E}{\beta} = -\frac{\frac{\tau_E}{(N-1)}\beta^2 + \frac{\sigma_z^2\tau_E\rho}{N(N-1)}\beta - \frac{2(N-2)\sigma_z^2\tau_E^2}{N^2(N-1)^2}}{-\frac{\tau_E}{(N-1)}\beta^2 - \frac{2\sigma_z^2\tau_E\rho}{N(N-1)}\beta + \frac{3(N-2)\sigma_z^2\tau_E^2}{N^2(N-1)^2}}. \quad (18)$$

Note that the denominator is positive. Rearranging terms, we have

$$\frac{d\beta}{d\tau_E}\frac{\tau_E}{\beta} > \frac{1}{2} \Leftrightarrow \frac{-\tau_E}{(N-1)}\beta^2 + \frac{(N-2)\sigma_z^2\tau_E^2}{N^2(N-1)^2} > 0 \Leftrightarrow \rho\beta^3 + \frac{\sigma_z^2\tau_E\rho}{N(N-1)}\beta > 0.$$

Last equivalence holds from $F(N, \beta) = 0$. ■

Note that, evaluating $\frac{d\varphi}{d\tau_E}$, we have $\frac{d\varphi}{d\tau_E}\frac{\tau_E}{\varphi} = (1-\varphi)\left(2\frac{d\beta}{d\tau_E}\frac{\tau_E}{\beta} - 1\right)$. We can also argue the limiting case of $\tau_E \rightarrow \infty$ and $N \rightarrow \infty$. From equation 14), with $N \rightarrow \infty$, we have,

$$\rho(N\beta)^3 + \tau_E(N\beta)^2 + \rho\sigma_z^2\tau_EN\beta - \sigma_z^2\tau_E^2 = 0.$$

Here $N\beta$ should be finite to satisfy the equality. Taking the limit for τ_E , we obtain $(N\beta)^2 = \sigma_z^2 \tau_E$. Plugging this into the definition of φ , we find that $\lim_{\tau_E \rightarrow \infty, N \rightarrow \infty} \varphi = 1/2$.²⁹

Lemma 6.

$$\frac{d\tau_I}{d\tau_E} > 0.$$

Proof. Differentiating $\tau_I \equiv \tau_v + \frac{\tau_E}{N} + \frac{(N-1)\tau_E}{N} \varphi(\tau_E)$ with respect to τ_E , we have $\frac{d\tau_I}{d\tau_E} = \frac{1}{N} + \frac{(N-1)}{N} \varphi(\tau_E) + \frac{(N-1)\tau_E}{N} \frac{d\varphi}{d\tau_E}$. Applying Lemma 5, with $\varphi > 0$ by definition, we have the desired result. ■

Corollary 3.

$$\frac{dR^2}{d\tau_E} > 0.$$

Proof. Differentiating of $R^2 = \left(1 - \frac{\tau_v}{\tau_I(\tau_E)}\right)$ by τ_E , we have $\frac{dR^2}{d\tau_E} = \frac{\tau_v}{\tau_I^2} \frac{d\tau_I}{d\tau_E}$. Applying Lemma 6, we have the desired result.

Lemma 7.

For any finite exogenous parameters, there is sufficiently large τ_E that satisfies $\frac{d\lambda}{d\tau_E} < 0$.

Also, for any finite τ_E , there is sufficiently large τ_v that satisfies $\frac{d\lambda}{d\tau_E} > 0$.

Proof. Since $\lambda = R^2/N\beta$, taking a derivative of $\log \lambda$ with respect to τ_E , we have $\frac{d\lambda}{d\tau_E} \frac{\tau_E}{\lambda} = \frac{dR^2}{d\tau_E} \frac{\tau_E}{R^2} - \frac{d\beta}{d\tau_E} \frac{\tau_E}{\beta}$.

Since both λ and τ_E are positive, we examine the sign of $\frac{dR^2}{d\tau_E} \frac{\tau_E}{R^2} - \frac{d\beta}{d\tau_E} \frac{\tau_E}{\beta}$ to explore the sign of $\frac{d\lambda}{d\tau_E}$.

Because of the definition of τ_I and φ is bounded, as τ_E goes to infinity, τ_v/τ_I approach to 0 and $R^2 \rightarrow 1$. After some calculation, ignoring exogenous constants, we have $\frac{dR^2}{d\tau_E} \frac{\tau_E}{R^2} = \frac{\tau_v}{\tau_I^2} \frac{d\tau_I}{d\tau_E} \frac{\tau_E}{R^2} \sim \frac{1}{\tau_E^2} \left(\tau_E + \frac{d\varphi}{d\tau_E} \tau_E^2 \right)$, which goes to zero as τ_E goes to infinity because $\frac{d\varphi}{d\tau_E} \rightarrow 0$ (φ is increasing but bounded). Since $\frac{d\beta}{d\tau_E} \frac{\tau_E}{\beta} > 1/2$, $\frac{d\lambda}{d\tau_E} \frac{\tau_E}{\lambda}$ becomes negative if τ_E is sufficiently large. The cutoff value depends on other exogenous parameters.

Next, we show the second statement. First, we show that $\frac{d\beta}{d\tau_E} \frac{\tau_E}{\beta} < 1$. From equation (18), a direct calculation leads $\frac{d\beta}{d\tau_E} \frac{\tau_E}{\beta} < 1 \Leftrightarrow \rho\beta - \frac{N-2}{N(N-1)} \tau_E < 0$. Plugging $\beta^* = \frac{N-2}{\rho N(N-1)} \tau_E$ into the cubic equation (8), we find that $F(N, \beta^*) > 0$. Since $F(N, \beta)$ is an increasing function of β , β that satisfies $F(N, \beta) = 0$ is less than β^* . Thus,

²⁹ The result $0 < \varphi < \frac{1}{2}$ is also stated in Kyle (1989).

we can show that $\rho\beta - \frac{N-2}{N(N-1)}\tau_E < 0$. Second, we show that we can find τ_v that satisfies $\frac{dR^2}{d\tau_E} \frac{\tau_E}{R^2} > \frac{d\beta}{d\tau_E} \frac{\tau_E}{\beta}$.

Expanding $\frac{dR^2}{d\tau_E} \frac{\tau_E}{R^2}$, we have

$$\frac{dR^2}{d\tau_E} \frac{\tau_E}{R^2} = \frac{\tau_v}{\tau_I} \left(1 + (1 - \varphi) \left(2 \frac{d\beta}{d\tau_E} \frac{\tau_E}{\beta} - 1 \right) \varphi \frac{N-1}{1 + \varphi(N-1)} \right) \quad (19)$$

Thus, collecting terms on τ_v for $\frac{dR^2}{d\tau_E} \frac{\tau_E}{R^2} > \frac{d\beta}{d\tau_E} \frac{\tau_E}{\beta}$, we can find the condition for τ_v . Note that $\beta, \varphi, \frac{d\beta}{d\tau_E} \frac{\tau_E}{\beta}$ are not a function of τ_v and $\frac{1}{2} < \frac{d\beta}{d\tau_E} \frac{\tau_E}{\beta} < 1$ and $0 < \varphi < \frac{1}{2}$ ensure the existence of such τ_v . For a limiting case, the condition for τ_v reduces to $\tau_v > \frac{\varphi}{1-\varphi} \tau_E \rightarrow \tau_E$ when $N \rightarrow \infty$. ■

Note that we can also show that $\frac{d\lambda}{d\tau_E} \frac{\tau_E}{\lambda}$ approaches to zero when τ_E goes to zero (from equation (8), (18)).

Lemma 8.

$$\frac{dV}{d\tau_E} > 0.$$

Proof. Note that $V = \left(1 + \frac{\tau_v}{\tau_E}\right) (R^2)^2 = \left(1 + \frac{\tau_v}{\tau_E}\right) \left(1 - \frac{\tau_v}{\tau_I}\right)^2$. Substituting out τ_I , we obtain

$$V = \left(1 + \frac{\tau_v}{\tau_E}\right) \left(\frac{1 + (N-1)\varphi}{1 + (N-1)\varphi + N\tau_v/\tau_E}\right)^2.$$

Since $V > 0$, for obtaining $\frac{dV}{d\tau_E} > 0$, it is enough to show $\frac{d\ln V}{d\tau_E} > 0$;

$$\frac{d\ln V}{d\tau_E} = \frac{d\varphi}{d\tau_E} \left(\frac{2(N-1)}{1 + (N-1)\varphi} - \frac{2(N-1)}{1 + (N-1)\varphi + N\tau_v/\tau_E} \right) - \frac{N\tau_v}{\tau_E^2} \left(\frac{1}{N + N\tau_v/\tau_E} - \frac{2}{1 + (N-1)\varphi + N\tau_v/\tau_E} \right).$$

From Lemma 5 we have $d\varphi/d\tau_E > 0$, and the first parenthesis is positive because $\varphi > 0$. Also, we can show the second parenthesis is negative. This results in $d\ln V/d\tau_E > 0$. ■

References

- Anand, Amber, Carsten Tanggaard, and Daniel G. Weaver, 2009, Paying for Market Quality, *Journal of Financial and Quantitative Analysis* 44, 1427–1457.
- Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Clara Vega, 2003, Micro effects of macro announcements: Real-time price discovery in foreign exchange, *The American Economic Review* 93, 38–62.
- Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Clara Vega, 2007, Real-time price discovery in global stock, bond and foreign exchange markets, *Journal of International Economics* 73, 251–277.
- Barclay, Michael J., and Terrence Hendershott, 2003, Price Discovery and Trading After Hours, *Review of Financial Studies* 16, 1041–1073.
- Bernard, Victor, and Jacob Thomas, 1989, Post-Earnings-Announcement Drift : Delayed Price Response or Risk Premium ?, *Journal of Accounting Research* 27, 1–36.
- Bernhardt, Dan, and Bart Taub, 2006, Kyle v.s. Kyle ('85 v. '89), *Annals of Finance* 2, 23--38.
- Bondarenko, Oleg, 2001, Competing market makers, liquidity provision, and bid-ask spreads, *Journal of Financial Markets* 4, 269–308.
- Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan, 2014, High-Frequency Trading and Price Discovery, *Review of Financial Studies* 27, 2267–2306.
- Cheung, Yin-wong, Rasmus Fatum, and Yohei Yamamoto, 2019, The exchange rate effects of macro news after the global financial crisis, *Journal of International Money and Finance* 95, 424--443.
- Chordia, Tarun, T Clifton Green, and Badrinath Kottimukkalur, 2018, Rent seeking by low-latency traders: Evidence from trading on macroeconomic announcements, *Review of Financial Studies* 31, 4650–4687.
- Easley, David, Maureen O'Hara, and Marcos M. Lopez de Prado, 2012, Flow toxicity and liquidity in a high-frequency world, *Review of Financial Studies* 25, 1457–1493.
- Evans, Martin D. D., and Richard K. Lyons, 1999, Order flow and exchange rate dynamics, *Journal of Political Economy* 110, 170–180.

- Evans, Martin D.D., and Richard K. Lyons, 2008, How is macro news transmitted to exchange rates?, *Journal of Financial Economics* 88, 26–50.
- Fatum, Rasmus, Michael Hutchison, and Thomas Wu, 2012, Asymmetries and State Dependence : The Impact of Macro Surprises on Intraday Exchange Rates, *Journal of the Japanese and International Economies* 26, 542--560.
- Grossman, Sanford J., and Joseph E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *The American Economic Review* 70, 393–408.
- Hashimoto, Yuko, and Takatoshi Ito, 2010, Effects of Japanese macroeconomic statistic announcements on the dollar/yen exchange rate: High-resolution picture, *Journal of the Japanese and International Economies* 24, 334–354.
- Holden, Craig W., and Avanidhar Subrahmanyam, 2002, News Events, Information Acquisition, and Serial Correlation, *The Journal of Business* 75, 1–32.
- Ito, Takatoshi, Kenta Yamada, Misako Takayasu, and Hideki Takayasu, 2012, Free Lunch! Arbitrage Opportunities in the Foreign Exchange Markets, NBER working paper (18541).
- Jiang, George, Ingrid Lo, and Giorgio Valente, 2012, High-Frequency Trading in the U.S. Treasury Market Around Macroeconomic News Announcements, Working Paper, Washington State University.
- Kaul, Gautam, Qin Lei, and Noah Stoffman, 2008, Aiming at Pin: Order Flow, Information, and Liquidity, Working paper.
- Kim, Oliver, and Robert E. Verrecchia, 1994, Market liquidity and volume around earnings announcements, *Journal of Accounting and Economics* 17, 41–67.
- Kim, Oliver, and Robert E. Verrecchia, 1997, Pre-announcement and event-period private information, *Journal of Accounting and Economics* 24, 395–419.
- Kim, Oliver, and Robert E Verrecchia, 2006, Trading Volume and Price Reactions to Public Announcements, *Journal of Accounting Research* 29, 302.
- Kirilenko, Andrei, Albert S Kyle, Mehrdad Samadi, and Tugkan Tuzun, 2017, The Flash Crash: High Frequency Trading in an Electronic Market, *The Journal of Finance* 72, 967–998.
- Kyle, Albert S, 1989, Informed speculation with imperfect competition, *The Review of Economic Studies* 56, 317–355.

- Llorente, Guillermo, and R Michaely, 2002, Dynamic volume-return relation of individual stocks, *Review of Financial Studies* 15, 1005–1047.
- Scholtus, Martin, Dick Van Dijk, and Bart Frijns, 2014, Speed, algorithmic trading, and market quality around macroeconomic news announcements, *Journal of Banking and Finance* 38, 89–105.
- Tetlock, Paul C., 2010, Does public financial news resolve asymmetric information?, *Review of Financial Studies* 23, 3520–3557.
- van Bommel, Jos, 2011, Measuring price discovery: The variance ratio, the R2, and the weighted price contribution, *Finance Research Letters* 8, 112–119.
- Vega, Clara, 2006, Stock price reaction to public and private information, *Journal of Financial Economics* 82, 103–133.
- Wang, Jiang, 1994, A Model of Competitive Stock Trading Volume, *Journal of Political Economy* 102, 127.

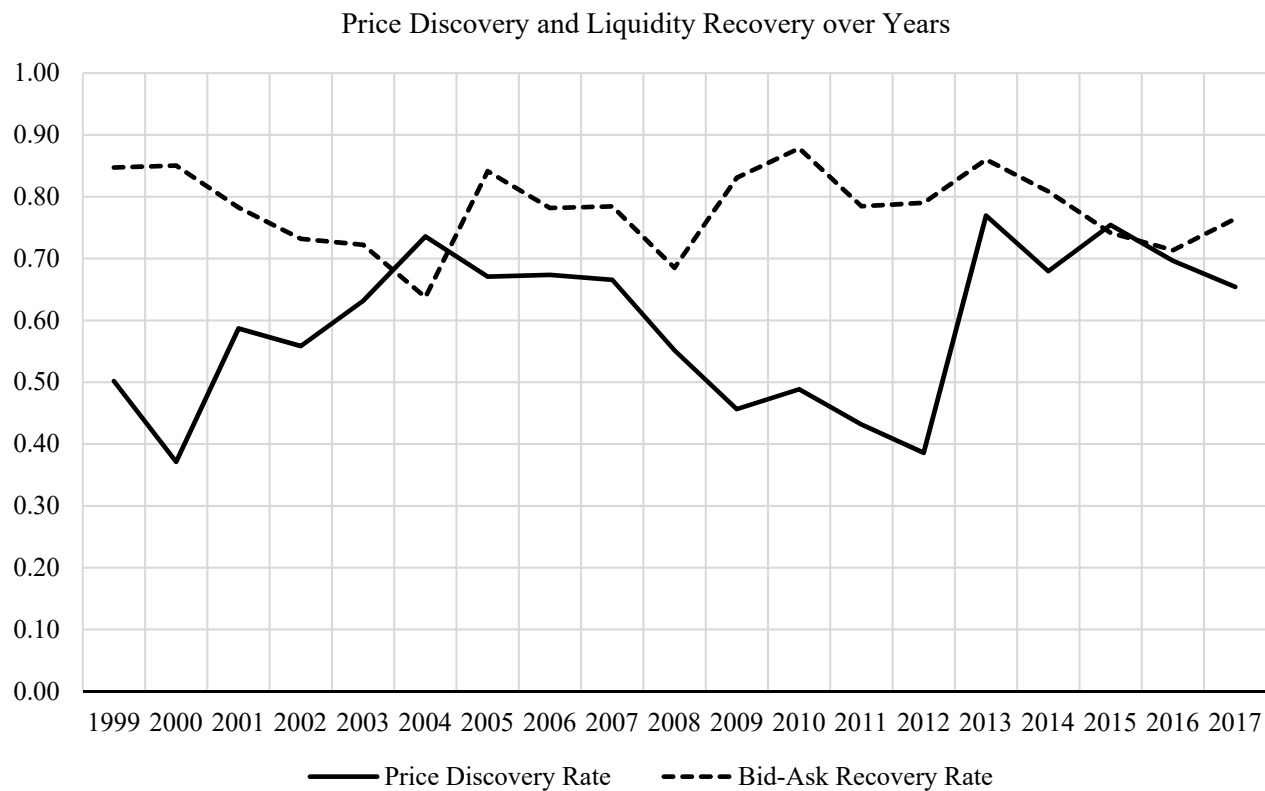


Figure 1: Yearly changes in the speed of price discovery and Liquidity recovery

The definition of price discovery and liquidity recovery are provided at equation (12) and (13). We calculate $PD_{k=40, T=600}$, $LR_{k=5, T=300}$ for each yearly sub-sample. Both measures do not show evident improvement over years; the market quality does not necessarily become better in recent years.

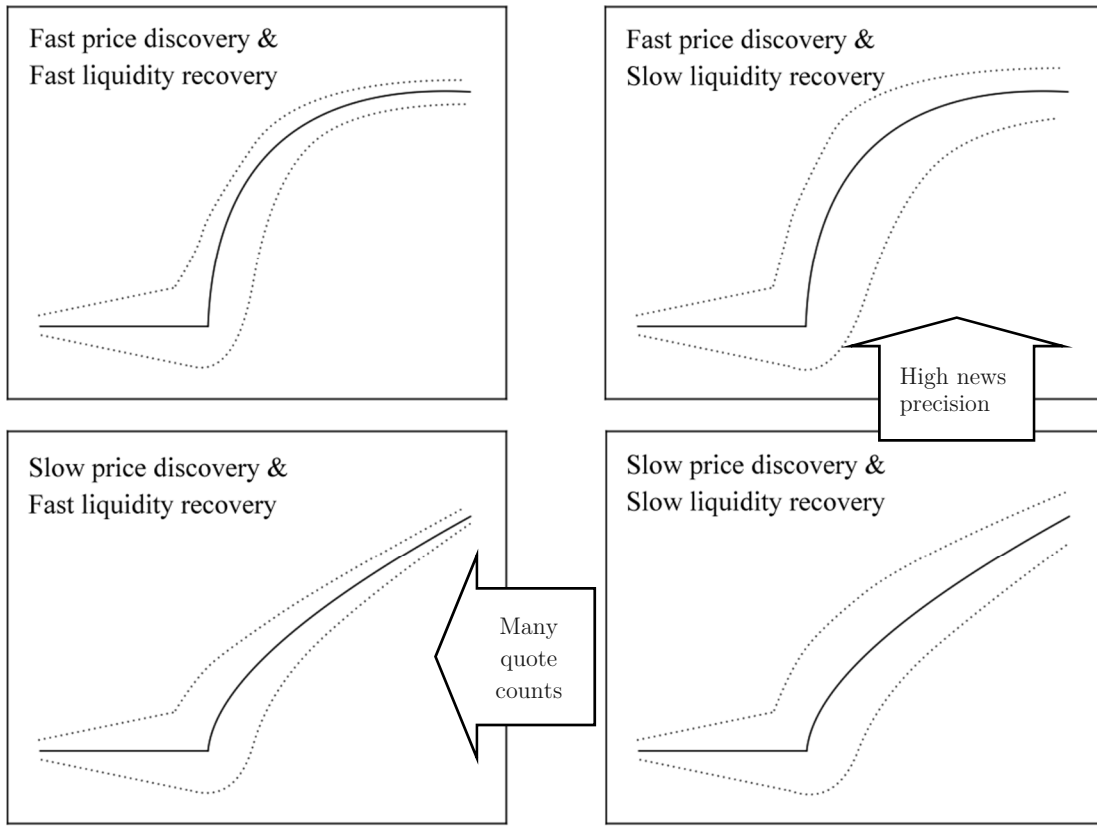


Figure 2: Diagram of empirical findings

In each panel, the solid line describes a sample path of mid-quotes, which is associated with the bid prices (the line below) and ask prices (the line above). When the price discovery is fast, the price path converges to the stable level quickly. When the liquidity recovery is fast, bid-ask spreads become narrow quickly. Our interest is what effects two key variables, the explanatory power of news for after-the-news return, and the number of traders have on these.

Table 1: Selected descriptive statistics after macro announcements

Descriptive statistics between (u,t) to (u,t+30sec)	# of quote revision	quote counts	depth (n<=5)	depth (n>=5)	depth all	bid-ask spread	return	# of trade	volume	OIB /VOL	ret/OIB	R ²
1999	9	-	-	-	-	0.95	1.32	-	-	-	-	-
	6	-	-	-	-	8.12	3.66	-	-	-	-	-
2000	11	-	-	-	-	1.13	1.58	-	-	-	-	-
	6	-	-	-	-	3.89	3.78	-	-	-	-	-
2001	14	-	-	-	-	1.14	2.36	-	-	-	-	0.31
	6	-	-	-	-	0.95	6.52	-	-	-	-	0.09
2002	14	-	-	-	-	1.10	2.51	-	-	-	-	0.34
	6	-	-	-	-	0.93	4.37	-	-	-	-	0.03
2003	17	-	-	-	-	0.91	2.60	-	-	-	-	0.27
	6	-	-	-	-	1.01	6.65	-	-	-	-	0.03
2004	19	-	-	-	-	0.82	2.48	-	-	-	-	0.37
	7	-	-	-	-	1.22	12.25	-	-	-	-	0.08
2005	17	-	-	-	-	0.83	2.07	-	-	-	-	0.48
	7	-	-	-	-	0.64	7.25	-	-	-	-	0.02
2006	28	84	146	96	252	0.80	2.94	26	202	0.15	0.05	0.31
	7	27	63	53	95	0.39	7.23	9	153	0.22	2.21	0.05
2007	28	108	175	145	335	0.74	1.81	23	124	0.18	0.02	0.34
	6	38	81	84	137	0.43	4.78	9	141	0.21	0.31	0.01
2008	78	74	91	75	173	0.74	2.77	36	104	0.20	0.05	0.34
	28	36	50	63	97	0.79	5.35	27	159	0.18	0.56	0.06
2009	96	103	88	66	156	1.35	2.25	36	87	0.17	0.07	0.16
	47	44	47	41	78	0.72	4.03	31	103	0.19	0.96	0.03
2010	130	165	125	99	233	0.79	1.96	47	102	0.15	0.04	0.13
	61	43	51	49	82	0.56	3.63	42	116	0.18	0.90	0.01
2011	96	36	21	23	45	0.95	2.19	44	83	0.18	0.05	0.16
	58	47	38	42	76	0.44	4.11	44	115	0.19	0.56	0.02
2012	78	39	23	27	51	0.77	1.33	31	53	0.16	0.02	0.13
	57	39	38	38	64	0.31	2.63	40	90	0.20	0.61	0.02
2013	104	86	49	57	108	0.76	2.11	43	78	0.18	0.06	0.20
	68	24	23	25	41	0.33	6.51	54	137	0.23	0.59	0.07
2014	82	85	54	57	116	0.77	1.63	32	69	0.25	0.04	0.35
	59	19	39	67	93	0.37	6.53	46	155	0.24	0.51	0.02
2015	100	74	39	48	89	1.33	3.36	36	72	0.28	0.07	0.35
	61	18	30	40	62	1.04	13.44	50	171	0.25	1.67	0.03
2016	71	81	43	51	97	1.33	2.00	20	37	0.36	0.03	0.41
	54	28	31	32	51	0.73	7.25	35	115	0.28	1.10	0.04
2017	69	103	56	61	115	0.89	1.78	16	29	0.35	0.04	0.41
	53	34	25	42	57	0.84	5.18	29	93	0.29	0.85	0.04

Note: Shaded rows are yearly median, non-shaded rows are standard deviation. Quote counts, depth, bid-ask spread are a snapshot at $(t, u + 30\text{sec})$. Other indicators are the sum between (t, u) to $(t, u + 30\text{sec})$. “depth (n<=5)” is the sum of limit order up to 5 steps. Each statistic is the median in each year. “OIB” stands for order imbalance, or the sum of deals initiated by buyers minus deals initiated by sellers. Returns are defined as log return in basis points. Before 2006, detailed limit order book data (quote counts and depth) and deal data are not available. Note that the minimum tick size is revised in 2011 from 1 pip to 1/10 pip, and in 2012 from 1/10 pip to 1/2 pip, showing discrete jumps in quote counts and depth. R-squared is calculated from the 500period window rolling regression of (11).

Table 2: Impact of News Surprise, t-values, and Average return magnitudes

$y = (\log(S(t,u+600\text{sec}) - \log(S(t,u))) \times 10,000$	coefficient	t-stat	Median of abs(y) 10-min after the news	Median quote revision in 10-min after the news	Frequency (monthly UON)	Availability (from 1999 UON)	Significance reported in Andersen et al. (2003)
Chg. Nonfarm Payrolls	-20.187	-6.460	22.9	1995			†
GDP_A	-11.073	-4.446	13.4	1252	Quarterly		†
ISM Manufacturing	-6.839	-4.446	10.0	784			†
Retail Sales	-6.054	-4.431	9.4	751			†
Trade Balance	-6.454	-4.298	9.8	589			†
Empire Manufacturing	-3.873	-4.121	6.4	954		from 2004	
Consumer Confidence	-4.642	-3.443	8.9	556			†
New Home Sales	-2.781	-3.311	7.0	595			
Durable Goods Orders	-3.054	-3.096	7.5	587			†
ADP	-4.984	-2.962	6.2	1461		from 2007	
Chicago Purchasing Manager	-2.599	-2.899	8.1	577			
Avg. Hourly Earnings	-7.963	-2.839	22.9	1995			
Current Account	-3.683	-2.798	6.5	549	Quarterly		
Factory Orders	-2.423	-2.724	8.4	590			†
Phil. Fed Outlook	-3.228	-2.233	7.0	594			
CPI	-2.347	-2.148	9.6	1415			
Housing Starts	-1.706	-1.917	6.9	579			
Industrial Production	-2.021	-1.837	6.6	544			†
PPI. Ex. Food Energy	-1.660	-1.501	8.0	691			
U of Mich Sentiment_F	-1.245	-1.413	8.8	648		from 2001	
GDP_S	-2.367	-1.403	8.8	830	Quarterly		
U of Mich Sentiment_P	-1.421	-1.357	8.7	852		from 2001	
Personal Consumption_T	-3.193	-1.318	7.3	847	Quarterly	from 2004	
Leading Index	-0.949	-1.254	7.1	629			
NAHB Housing Market Index	-1.313	-1.203	4.7	656		from 2004	
Personal Consumption_S	-4.416	-1.022	8.1	1164	Quarterly	from 2004	
Existing Home Sales	-1.116	-0.916	8.1	576			
PCE Core	-0.863	-0.821	6.4	868		from 2005	
GDP Def_A	-1.753	-0.793	13.4	1332	Quarterly	from 2001	
Import Price	-1.181	-0.739	7.6	897		from 2005	
GDP Def_T	-1.056	-0.704	7.3	570	Quarterly	from 2001	
Chg. Manufact. Payrolls	-1.715	-0.625	22.9	1995			
GDP Def_S	-0.723	-0.530	8.8	830	Quarterly	from 2001	
Personal Consumption_A	-1.278	-0.391	13.4	1631	Quarterly	from 2004	
Budget Statement	-0.141	-0.318	3.6	143			
Whole Sale Inventories	-0.207	-0.234	7.0	571			
Personal Spending	-0.174	-0.217	6.4	538			
GDP_T	-0.079	-0.041	7.0	561	Quarterly		
Unit Labor Cost_F	-0.053	-0.030	7.6	708	Quarterly	from 2002	
Employment Cost Index	-0.054	-0.017	9.4	615	Quarterly		
Construction Spending	0.186	0.150	9.4	633			
Building Permits	0.466	0.471	6.9	846		from 2004	
Consumer Credit	0.287	0.773	2.5	127			
Unit Labor Cost_P	2.360	0.855	7.5	904	Quarterly	from 2002	
Business Inventories	0.748	0.862	7.7	566			
Capacity Utilization	1.367	1.291	6.6	548			
Personal Income	1.251	1.621	6.4	537			
Unemployment Rate	7.819	3.661	22.9	1995			
Obs.		5497					
R-squared		0.105					

Note: Suffix A, S, T, F, P denotes 'Advance', 'Second', 'Third', 'Final', and 'Preliminary'

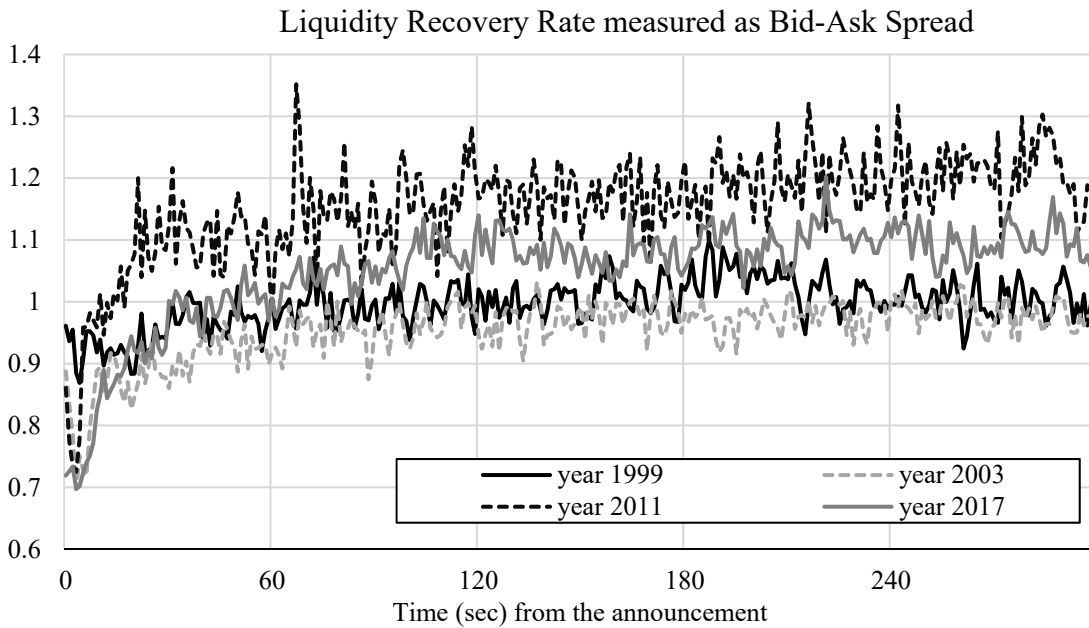
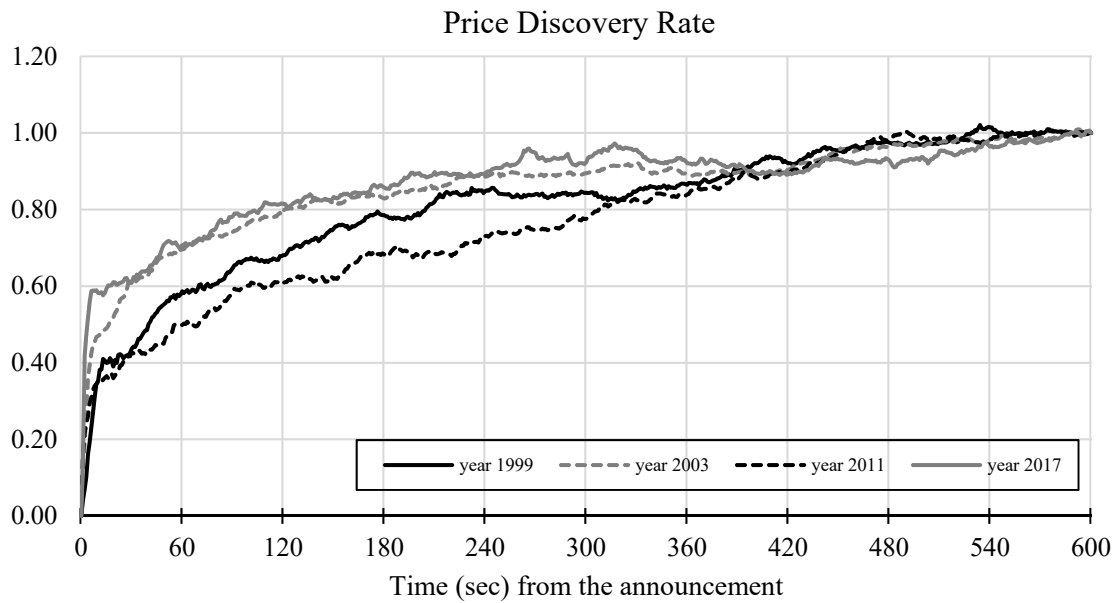


Figure 3: Price discovery and liquidity recovery after the news

The definition of price discovery and liquidity recovery are presented in equation (12) and (13). The panel above shows the $PD_{k,T=600}$ and the one below shows $LR_{k,T=300}$.

Table 3: Correlation coefficients of independent variables

	R^2	Quote Count	Liquidity recovery	Depth	Surprise of news	# of quote revision	Time trend	Y2009	Y2010	Y2011	Y2012	OIB /VOL	VOL
R^2 measure	1.00	-0.12	-0.03	0.04	0.00	0.04	-0.04	-0.02	-0.07	-0.06	-0.06	-0.07	0.18
Quote Count at $(t, u+k)$	-0.12	1.00	0.17	0.22	-0.02	0.07	0.46	-0.16	0.17	-0.02	0.13	0.12	-0.27
Liq. recovery rate at $(t, u+k) \times 100$	-0.03	0.17	1.00	0.12	0.01	-0.07	-0.07	0.06	0.03	-0.05	-0.07	0.02	-0.03
Depth at $(t, u+k)$	0.04	0.22	0.12	1.00	0.01	-0.23	-0.34	0.02	0.24	-0.34	-0.32	-0.02	0.12
Surprise of news	0.00	-0.02	0.01	0.01	1.00	0.03	-0.01	0.05	-0.05	-0.02	-0.01	-0.04	0.10
# of quote revision	0.04	0.07	-0.07	-0.23	0.03	1.00	0.17	0.19	0.24	0.20	0.10	-0.27	0.40
Linear time trend	-0.04	0.46	-0.07	-0.34	-0.01	0.17	1.00	-0.21	-0.14	-0.04	0.05	0.23	-0.32
Y2009	-0.02	-0.16	0.06	0.02	0.05	0.19	-0.21	1.00	-0.09	-0.09	-0.09	-0.07	0.04
Y2010	-0.07	0.17	0.03	0.24	-0.05	0.24	-0.14	-0.09	1.00	-0.09	-0.09	-0.11	0.08
Y2011	-0.06	-0.02	-0.05	-0.34	-0.02	0.20	-0.04	-0.09	-0.09	1.00	-0.09	-0.08	0.03
Y2012	-0.06	0.13	-0.07	-0.32	-0.01	0.10	0.05	-0.09	-0.09	-0.09	1.00	-0.04	-0.09
OIB /cumulative volume	-0.07	0.12	0.02	-0.02	-0.04	-0.27	0.23	-0.07	-0.11	-0.08	-0.04	1.00	-0.43
Cumulative volume	0.18	-0.27	-0.03	0.12	0.10	0.40	-0.32	0.04	0.08	0.03	-0.09	-0.43	1.00

Table 4: Regression results for the determinants of speed of price discovery

$y = PD_{t,k,T=600}^R$		k=5sec	k=40sec	k=120sec	k=180sec	k=300sec	k=40sec
(1)	R^2 measure	2.365 *** [0.6646]	1.595 ** [0.7476]	1.845 *** [0.7781]	0.7169 [0.7958]	1.123 * [0.7562]	2.554 *** [0.7282]
(2)	Quote Count at $(t,u+k)$	-3.763 ** [1.673]	-4.407 ** [2.533]	-4.673 * [3.33]	-1.944 [3.566]	4.727 * [3.042]	-5.222 *** [1.965]
(3)	Liq. recovery rate at $(t,u+k) \times 100$	-14.39 *** [3.826]	-1.005 [5.059]	10.1 ** [5.868]	15.74 *** [5.88]	1.365 [6.108]	
(4)	Depth at $(t,u+k)$	0.2582 [1.542]	2.091 [1.972]	0.379 [2.309]	0.887 [2.43]	-2.361 [2.135]	
(5)	Surprise of news	12.64 *** [1.591]	6.155 *** [1.575]	4.814 *** [1.553]	0.3209 [1.46]	0.3869 [1.551]	
(6)	# of quote revision	-6.094 *** [1.465]	-7.389 *** [1.653]	-10.18 *** [1.877]	-11.1 *** [2.035]	-6.932 *** [2.158]	
(7)	Linear time trend	11.85 *** [3.811]	24.14 *** [4.786]	23.13 *** [5.554]	24.54 *** [5.779]	14.2 *** [5.723]	7.445 ** [3.341]
(8)	Y2009	1.266 [3.153]	5.358 * [3.472]	4.099 [3.49]	10.66 *** [3.912]	4.11 [3.613]	
(9)	Y2010	3.852 [3.073]	4.222 [3.414]	7.699 ** [3.511]	10.16 *** [3.679]	2.08 [3.551]	
(10)	Y2011	2.644 [3.402]	8.049 ** [3.893]	3.535 [4.063]	5.49 [4.597]	1.506 [4.381]	
(11)	Y2012	3.995 [3.367]	4.541 [3.86]	2.661 [3.971]	3.461 [4.056]	1.536 [3.935]	
(12)	OIB /cumulative volume x100	-0.03651 * [0.02729]	0.04118 [0.04413]	0.1752 *** [0.06303]	0.2429 *** [0.07624]	0.3077 *** [0.08481]	
(13)	cumulative volume	3.794 *** [0.781]	8.398 *** [0.9528]	9.74 *** [1.115]	11.78 *** [1.197]	11.15 *** [1.372]	
	constant	89.85 *** [14.81]	62.4 *** [15.48]	76.62 *** [16.61]	59.14 *** [17.17]	53.04 *** [17.06]	63.28 *** [3.45]
	N	3007	3013	2925	2882	2880	3013
	R-squared	0.070	0.042	0.039	0.039	0.030	0.0074

Note: Independent variable (1) is logit transformation of R^2 , variables (2), (4), (6), (13) are taken log. Standard errors are reported in parentheses. *, **, *** indicate significance at 10%, 5%, 1% level

The dependent variable is a measure of price discovery for each announcement time t , defined as

$$PD_{t,k,T=600}^R = \frac{|\Delta S(t, u+k) - \widehat{\Delta S}(t, u+T)|}{|\Delta S(t, u+T) - \widehat{\Delta S}(t, u+T)|}$$

This expression is a realized version of the price discovery measure (12). Overshooting of PD and LR by 100% is omitted from the sample. Independent variable (1) is a R-squared of following regression over t :

$$\Delta S(t, u+600) = constant + \alpha_i N_i(t, u) + \epsilon(t, u), i = 1, 2, \dots, \# \text{ of different news.}$$

This is a news-by-news version of regression equation (11). LHS takes a value only when each news i is observed. This measures the information held by informed traders. We estimate the regression equation by a rolling regression with window of 12 periods. This procedure generates R-squared for each i at each t . For each t , we employ a max of such R-squared over i for the independent variable (1). (2) quote count $(t,u+k)$ is the number of traders submitting limit order at $(t,u+k)$. This is a proxy of the number of traders inhabited with the market at each time. (6) the number of quote revision plus and minus 10 minutes around the announcement. (7) time trend is the linear trend from 2006 to 2017, scaled by N. Year-by-year dummies are introduced for four years after the global financial crisis.

Table 5: Regression results for the determinants of speed of liquidity recovery

$y = LR_{t,k,T=300}^R$	k=5sec	k=10sec	k=40sec	k=60sec	k=150sec	k=10sec
(1) R^2 measure	-0.2853 [0.3309]	-0.4636 * [0.2918]	-0.3507 [0.2768]	-0.08903 [0.2658]	-0.3879 * [0.2662]	-0.4555 * [0.3105]
(2) Quote count at $(t,u+k)$	25.37 *** [0.6526]	22.1 *** [0.7076]	10.84 *** [1.331]	7.801 *** [1.585]	3.315 *** [1.256]	18.66 *** [0.6356]
(3) PD rate at $(t,u+k)$ x100	-1.526 ** [0.8158]	-1.036 * [0.7842]	1.289 ** [0.6722]	1.534 *** [0.6386]	1.671 *** [0.6244]	
(4) Depth at $(t,u+k)$	-6.928 *** [0.7568]	-6.166 *** [0.7405]	-2.799 *** [0.8108]	-2.491 *** [0.9813]	-1.776 ** [0.8347]	
(5) Surprise of news	0.1552 [0.6903]	1.444 *** [0.5944]	0.7945 * [0.5491]	0.9929 ** [0.5382]	-0.08586 [0.4777]	
(6) # of quote revision	-0.8276 [0.7949]	-0.3127 [0.6988]	0.08382 [0.7217]	0.539 [0.7085]	0.1385 [0.6961]	
(7) Linear time trend	-28.63 *** [1.826]	-27.1 *** [1.799]	-16.11 *** [2.05]	-12.1 *** [2.198]	-5.093 *** [1.967]	-20.08 *** [1.357]
(8) Y2009	6.466 *** [1.648]	3.833 *** [1.399]	1.017 [1.325]	0.01929 [1.278]	-0.3303 [1.177]	
(9) Y2010	11.36 *** [1.567]	5.554 *** [1.337]	-2.481 ** [1.171]	-1.33 [1.116]	-0.8933 [1.04]	
(10) Y2011	-19.2 *** [2.071]	-17.92 *** [1.977]	-9.778 *** [1.765]	-7.607 *** [1.874]	-4.862 *** [1.638]	
(11) Y2012	-19.78 *** [1.546]	-19.48 *** [1.463]	-9.304 *** [1.526]	-6.802 *** [1.694]	-1.043 [1.516]	
(12) OIB /cumulative volume x100	0.003099 [0.01351]	0.005166 [0.01303]	0.01515 [0.01658]	0.01219 [0.01913]	0.02284 [0.02334]	
(13) cumulative volume	-0.4807 [0.406]	-0.1474 [0.3617]	-0.2942 [0.373]	-0.5785 * [0.39]	0.2105 [0.4277]	
constant	101.2 *** [7.442]	97.45 *** [6.596]	96.47 *** [6.073]	96.75 *** [6.112]	100.1 *** [5.31]	69.25 *** [1.039]
N	2986	3043	2997	2953	2964	3043
R-squared	0.447	0.330	0.075	0.039	0.010	0.248

Note: Independent variable (1) is logit transformation of R^2 , variables (2), (4), (6), (13) are taken log. Standard errors are reported in parentheses. *, **, *** indicate significance at 10%, 5%, 1% level

The dependent variable is a measure of liquidity recovery for each announcement time t, defined as

$$LR_{t,k,T=300}^R = \left\{ \frac{VWBAS(t, u+k)}{VWBAS(t, u+T)} \right\}^{-1},$$

where VWBAS is volume weighted bid and ask spread. In fact, this is an element of liquidity recovery measure (12). Overshooting of PD and LR by 100% is omitted from the sample. The definition of independent variables is the same as Table 4.

Table 6: Regression results for the determinants of speed of price discovery (tick-time)

	$y = PD_{t,n,T=600}^R$	n=5tick	n=50tick	n=100tick	n=200tick	n=100tick
(1)	R^2 measure	2.588 *** [0.6566]	1.398 ** [0.7849]	1.198 * [0.8309]	2.557 *** [0.7908]	2.554 *** [0.7282]
(2)	Quote Count at (t,n)	-8.96 *** [1.376]	-7.174 *** [1.932]	-9.015 *** [2.505]	-10.63 *** [3.018]	-5.222 *** [1.965]
(3)	Liq. recovery rate at (t,n) x 100	-0.0265 * [0.01923]	0.002167 [0.02245]	0.01563 [0.02217]	-0.00527 [0.02275]	
(4)	Depth at (t,n)	3.171 ** [1.473]	1.687 [1.992]	4.416 ** [2.058]	2.033 [2.342]	
(5)	Surprise of news	15.15 *** [1.561]	7.03 *** [1.644]	5.736 *** [1.635]	3.845 *** [1.649]	
(6)	# of quote revision	-8.084 *** [1.547]	-7.466 *** [1.757]	-9.955 *** [1.815]	-14.28 *** [1.753]	
(7)	Linear time trend	15.06 *** [3.88]	20.75 *** [5.005]	29.88 *** [5.005]	24.23 *** [5.695]	7.445 ** [3.341]
(8)	Y2009	-4.049 * [3.061]	3.331 [3.449]	1.543 [3.722]	8.284 ** [3.981]	
(9)	Y2010	2.301 [2.982]	2.727 [3.298]	3.739 [3.568]	5.686 * [3.798]	
(10)	Y2011	7.005 ** [3.181]	11.43 *** [4.366]	9.605 *** [3.94]	12.8 *** [4.417]	
(11)	Y2012	12.81 *** [3.362]	9.761 *** [3.986]	6.299 * [4.109]	4.502 [4.084]	
(12)	OIB /cumulative volume	0.1909 ** [0.09718]	0.04276 *** [0.009086]	0.0219 *** [0.004988]	0.01419 *** [0.002845]	
(13)	cumulative volume(t,u+k)	-0.04921 ** [0.0292]	0.03596 [0.5967]	-0.8715 [0.9264]	5.925 * [4.35]	
	constant	92.04 *** [14.54]	100 *** [17.24]	116.9 *** [18.51]	179.9 *** [17.92]	63.28 *** [3.45]
	N	2770	2753	2688	2620	3013
	R-squared	0.084	0.036	0.038	0.051	0.0074

Note: Independent variable (1) is logit transformation of R^2 , variables (2), (4), (6), (13) are taken log. Standard errors are reported in parentheses. *, **, *** indicate significance at 10%, 5%, 1% level.

The dependent variable is a measure of price discovery for each announcement time t , defined as

$$PD_{t,n,T=600}^R = \frac{|\Delta S(t, u+n) - \widehat{\Delta S}(t, u+T)|}{|\Delta S(t, u+T) - \widehat{\Delta S}(t, u+T)|}$$

Contrary to the specification in Table 4, we use tick-time to define the speed of price discovery. The numerator of the measure is indexed by the tick-time n , instead of the physical time k . Other specifications are identical to that of Table 4.

Table 7: Regression analysis with news type dummies

Independent variables	Interaction dummies	$y = PD_{t,k,T=600}^R$		$y = LR_{t,k,T=300}^R$	
		k=40sec	k=40sec	k=10sec	k=10sec
With controls		No	Yes	No	Yes
(1) R^2 measure	None	2.425 *** [0.7446]	1.723 ** [0.7525]	-0.5952 ** [0.2948]	-0.4787 * [0.2921]
(2) Quote Count at (t,n)	None	-2.715 [2.207]	-3.183 [2.59]	18.16 *** [0.7449]	20.58 *** [0.8409]
(3) Time trend	None	3.892 [3.722]	22.13 *** [4.948]	-16.79 *** [1.475]	-24.55 *** [1.912]
(4) R^2 measure	GDP Dummy	-19.85 [18.05]	-21.35 [18.48]	4.708 [5.358]	4.177 [5.361]
(5) Quote Count at (t,n)	GDP Dummy	-8.163 [12.61]	-9.513 [12.88]	15.26 *** [4.597]	11.5 *** [4.367]
(6) Time trend	GDP Dummy	29.44 [37.44]	22.57 [38.26]	-32.19 *** [7.983]	-23.61 *** [7.615]
(7) R^2 measure	Labor Stat. Dummy	-9.219 * [5.907]	-8.412 * [5.938]	5.449 *** [2.201]	4.414 ** [2.156]
(8) Quote Count at (t,n)	Labor Stat. Dummy	0.1862 [6.796]	-2.949 [6.491]	5.723 *** [2.253]	2.909 [2.393]
(9) Time trend	Labor Stat. Dummy	20.08 * [15.31]	15.37 [14.65]	-23.46 *** [3.341]	-17.68 *** [3.422]
N		3013	3013	3043	3043
R-squared		0.012	0.044	0.321	0.336

Note: Independent variable (1) is logit transformation of R^2 , variables (2), (4), (6) are taken log. *, **, *** indicate significance at 10%, 5%, 1% level. Standard errors are reported in parentheses. Control variables include depth at (t,n) , surprise of news, number of quote revisions, year dummies, $|OIB|$ /cumulative volume, and cumulative volume $(t,u+k)$. Each news type dummy takes one if the news type is Labor statistics (or GDP advance) and takes zero otherwise.

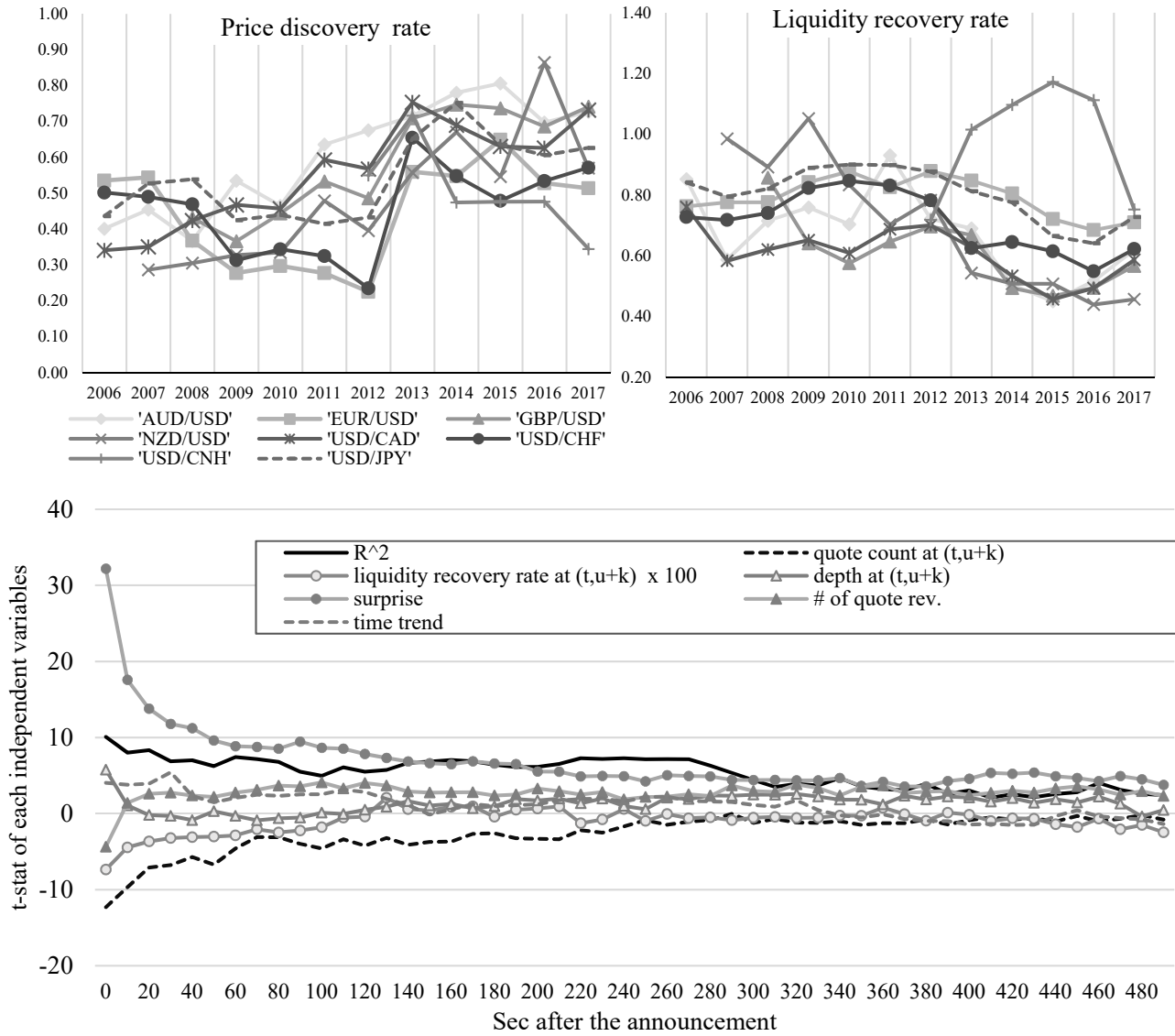


Figure 4: Panel regression (t-stat in different k) and yearly changes in market quality

Top panels. Yearly changes in price discovery rate ($k=40$, $T=1800$) and liquidity recovery rate ($k=5$, $T=60$) for each eight currency pairs. The way of construction is the same as Table 1.

Bottom panel. In addition to EUR/USD, other seven currency pairs are taken account: AUD/USD, GBP/USD, NZD/USD, USD/CAD, USD/CHF, USD/CNH, USD/JPY. Fixed effect panel regression is employed. The dependent variable is a measure of price discovery for each announcement time t , defined as $\frac{|\Delta S(t,u+k) - \Delta S(t,u+T)|}{|\Delta S(t,u+T) - \Delta S(t,u+T)|}$ where $T=600$.

The set of independent variables is the same as Table 4. R-squared measure for EUR/USD and surprise of news are common, and other four variables are calculated for each pair. The panel plots the t-statistics of independent variables in different k .

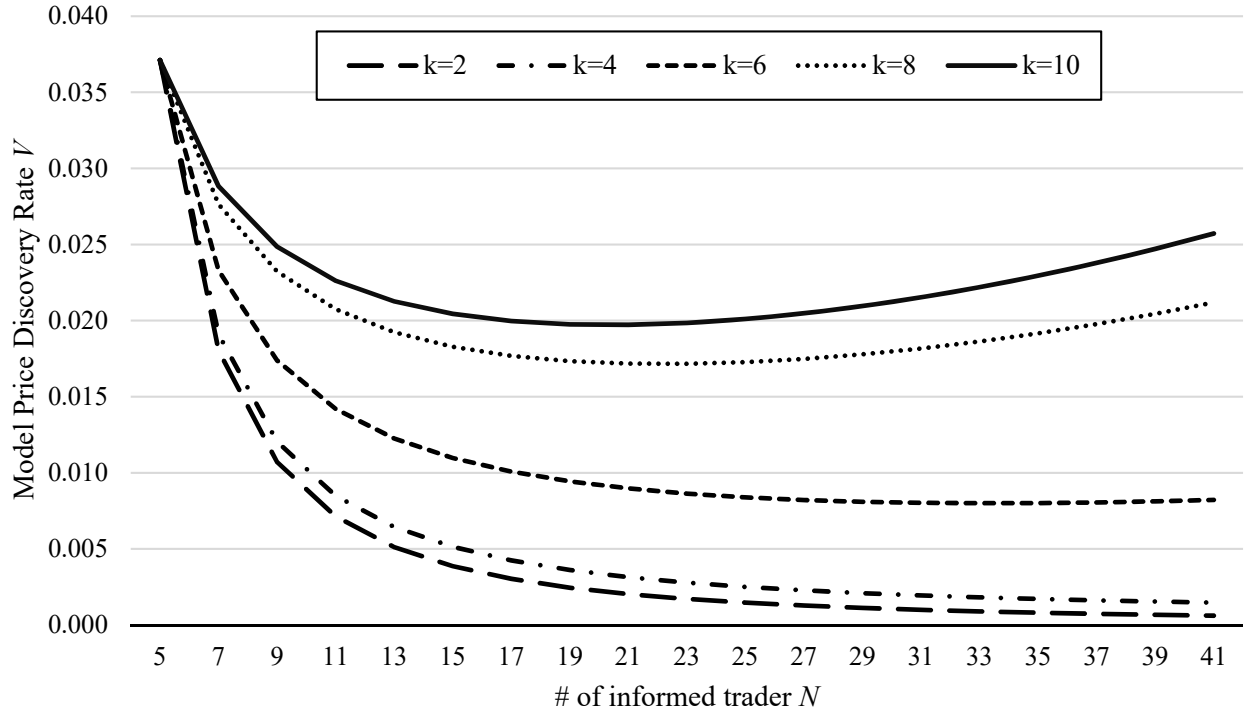


Figure 5: Model Simulation of Price Discovery as a Function of the number of Informed Trader

This plot is the variance ratio $V(N) \equiv \text{Var}(p(N) - \lambda(N)\tilde{z})/\text{Var}(\tilde{v})$, a measure of price discovery, with the informed trader's signal precision τ_e shifting from the monopolistic competition (when k is small) toward informed competition (when k is large). Specifically,

$$\tau_e(N, k) = w(k) \frac{\tau_E}{N_{\text{fix}}} + (1 - w(k)) \frac{\tau_E}{N}, w(k) = \frac{1}{1 + \exp(-(k - k_c))},$$

where $N_{\text{fix}} = 4$ and $k_c = 5$ are assumed. The downward sloping curve is consistent with our empirical findings.

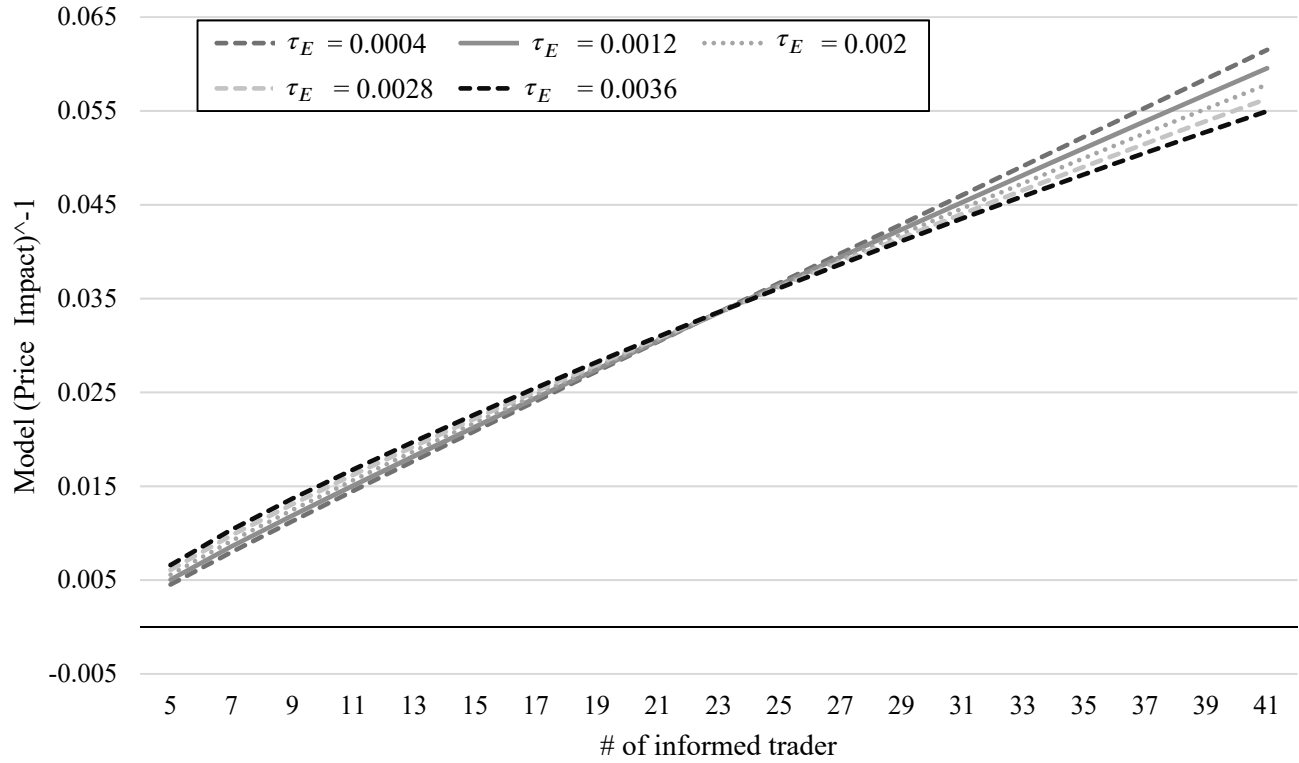


Figure 6: Model Simulation of Liquidity as a Function of the Number of Informed Traders

This plot is the inverse of price impact $\lambda(N)$, a measure of liquidity, as a function of N , with different τ_E . Upward sloping curve is consistent with our empirical findings. Increasing τ_E of reducing liquidity, our empirical findings, is observed in the region of sufficiently high number of informed traders.

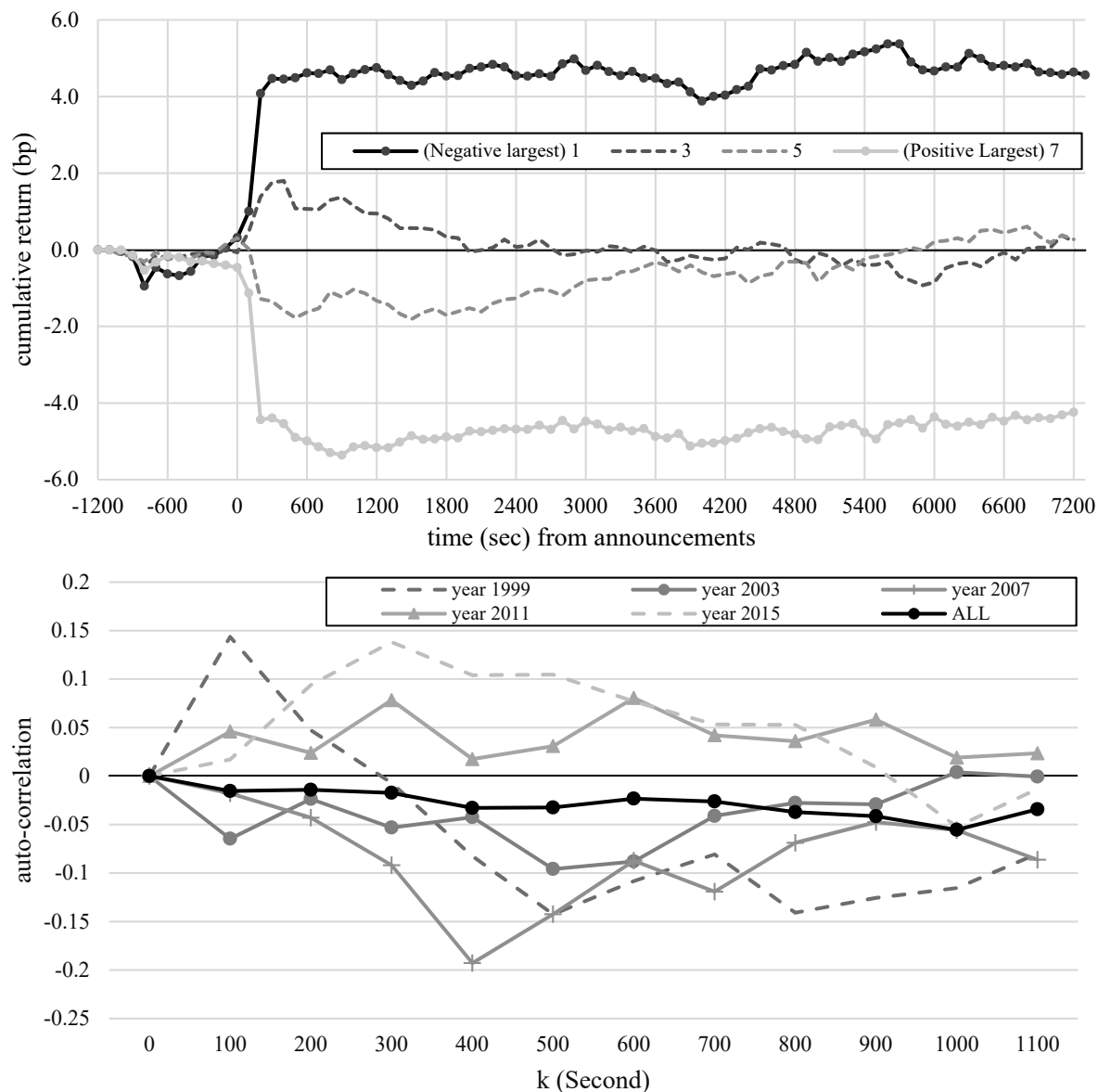


Figure 7: Predictability of return after the announcements

Top panel. Top panel plots the average cumulative return over two hours (+ 20 minutes) surrounding macro announcements. Each plot is an average of sample stratified by the news surprise defined as (10). Bin 1 represents the largest negative surprise (i.e., depreciates USD and earns positive returns on EUR/USD) and Bin 7 represents the largest positive surprise (i.e., appreciates USD and earns negative returns on EUR/USD). Bin 4 consists of zero surprise sample.

Bottom panel. Bottom panel plots the predictability of return after announcements, measured by the autocorrelation of returns defined as $\text{Corr}_t(s(t, u+k) - s(t, u), s(t, u+3\text{hours}) - s(t, u+k))$, $k = 0, 100\text{sec}, 200\text{sec}, \dots$, where t denotes announcement days and u denotes intraday announcement time. Each plot is grouped by years.