5-2020

Three essays on financial economics

Jiangyuan LI
Singapore Management University, jiangyuanli.2015@pbs.smu.edu.sg

Three Essays on Financial Economics

Jiangyuan LI

LKCSB

Supervised by Dashan Huang

Send correspondence to Jiangyuan LI, Lee Kong Chian School of Business, Singapore Management University,

Singapore 178899. e-mail: jiangyuanli.2015@pbs.smu.edu.sg.

Are Disagreements Agreeable?
Evidence from Information Aggregation*

Dashan Huang **Jiangyuan Li** **Liyao Wang**
Singapore Management University

April 2020

*We are grateful to G. William Schwert (the editor) and an anonymous referee for very insightful and helpful comments that significantly improved the paper. We thank Adem Atmaz, Suleyman Basak, Bong-Geun Choi, Liyuan Cui, Zhi Da, Stefano Giglio, Shiyang Huang, Tao Li, Weikai Li, Ye Li, Francis Longstaff, Sungjune Pyun, Guofu Zhou, seminar and conference participants at St. Louis University, Washington University in St. Louis, 2018 Asian Bureau of Finance and Economic Research (ABFER) Annual Conference, 2018 AsianFA Annual Meeting, 2018 City University of Hong Kong International Finance Conference on Corporate Finance and Financial Markets, 2018 Research in Behavioral Finance Conference, and 2018 SMU Finance Summer Camp for insightful comments.

Send correspondence to Dashan Huang, Lee Kong Chian School of Business, Singapore Management University, Singapore 178899. Phone: (65) 6808–5476 and e-mail: dashanhuang@smu.edu.sg.

Are Disagreements Agreeable? Evidence from Information Aggregation

Abstract

Disagreement measures are known to predict cross-sectional stock returns but fail to predict market returns. This paper proposes a partial least squares disagreement index by aggregating information across individual disagreement measures and shows that this index significantly predicts market returns both in- and out-of-sample. Consistent with the theory in Atmaz and Basak (2018), the disagreement index asymmetrically predicts market returns with greater power in high sentiment periods, is positively associated with investor expectations of market returns, predicts market returns through a cash flow channel, and can explain the positive volume-volatility relationship.

Keywords: Disagreement, Return predictability, PLS, PCA, LASSO, Machine learning

JEL Classification: G12, G14, G17

1. Introduction

Researchers in economics and finance have long been interested in studying the effects of expectations across investors. Investor disagreement, usually measured by the second moment of investor expectations, plays an important role in explaining stock returns, volatility, and trading volume. Due to its wide impacts, Hong and Stein (2007) conclude that disagreement represents "the best horse" for behavioral finance to obtain as many insights as classical asset pricing theories. However, unlike Baker and Wurgler's (2006) sentiment index that has been widely used to capture the first moment of investor expectations (see, e.g., Yu and Yuan, 2011; Stambaugh, Yu, and Yuan, 2012), investor disagreement has only been approximated through various proxies in the literature.[1] To date, there is a lack of research that examines disagreement measures collectively and it is unclear as to whether they are able to predict (excess) market returns in real time.

This paper examines whether extant disagreement measures can become agreeable. If extant measures capture disagreement, they should display commonality and have a common factor. To aggregate information across 24 individual measures, we propose a disagreement index by using the partial least squares (PLS) method in Kelly and Pruitt (2013, 2015). Empirically, we show that the 24 individual measures do have a common factor and the disagreement index significantly predicts market returns up to 12 months. Over the sample period of 1969:12–2018:12, a one-standard deviation increase in the disagreement index is associated with a 0.83% decrease in the next one-month market return and a 7.04% decrease in the next 12-month market return, where the latter is comparable to 6.6% in Yu (2011) who measures investor disagreement with analyst forecast dispersion. The in- and out-of-sample R^2s are 2.52% and 1.56% at the one-month horizon and 13.88% and 13.26% at the 12-month horizon. In contrast, there are only four individual disagreement measures that are significant at the one-month horizon and four others significant at the 12-month horizon for in-sample forecasting, but none of the 24 individual measures exhibits any out-of-sample forecasting power.

[1] Professional forecast dispersions (Anderson, Ghysels, and Juergens, 2009; Li, 2016; Bordalo et al., 2020), analyst forecast dispersions (Diether, Malloy, and Scherbina, 2002; Hong and Sraer, 2016), household forecast dispersions (Li and Li, 2017), unexplained trading volume (Garfinkel, 2009), and stock idiosyncratic volatility (Boehme, Danielsen, and Sorescu, 2006) are some of the most prominent disagreement measures to date.

PLS is chosen for information aggregation due to its simplicity and efficacy. PLS is initially proposed by Wold (1966) and further developed by Kelly and Pruitt (2013, 2015), which extracts the disagreement index with a three-pass regression filter to reduce common noises in the individual disagreement measures. Theoretically, PLS outperforms PCA in extracting factors for prediction if individual predictors contain a common (noise) component that is unrelated with future market returns. The intuition is that, as a supervised learning technique, PLS incorporates the target information—market returns—in the factor extracting procedure and teases out any common component that is uncorrelated with future market returns. Empirically, Kelly and Pruitt (2013), Lyle and Wang (2015), Huang et al. (2015), Giglio, Kelly, and Pruitt (2016), Light, Maslov, and Rytchkov (2017), and Gu, Kelly, and Xiu (2020), among others, show that PLS is effective in extracting factors for predicting stock returns and economic activities in the time series and cross-section.

The forecasting power of the disagreement index is not subsumed by economic predictors and uncertainty measures. It remains significant after controlling for the 14 economic predictors in Welch and Goyal (2008), output gap in Cooper and Priestley (2009), and aggregate short interest in Rapach, Ringgenberg, and Zhou (2016). Also, while the disagreement index represents one type of uncertainty (see, e.g., Anderson, Ghysels, and Juergens, 2009; Atmaz and Basak, 2018), it is distinct from extant uncertainty measures, such as economic uncertainty (Bali, Brown, and Caglayan, 2014), treasury implied volatility (Choi, Mueller, and Vedolin, 2017), financial uncertainty and macro uncertainty (Jurado, Ludvigson, and Ng, 2015), economic policy uncertainty (Baker, Bloom, and Davis, 2016), news implied volatility (Manela and Moreira, 2017), sample variance (Welch and Goyal, 2008), and the Chicago Board Options Exchange (CBOE) volatility index (VIX).

The ability of disagreement in predicting market returns is robust to alternative econometric methods. In addition to PLS, we explore six LASSO-related machine learning methods (see, e.g., Rapach, Strauss, and Zhou, 2013; Chinco, Clark-Joseph, and Ye, 2019; Diebold and Shin, 2019; Han, He, Rapach, and Zhou, 2019; Freyberger, Neuhierl, and Weber, 2020; Kozak, Nagel, and Santosh, 2020), and find that all of them generate significant out-of-sample R^2s, although the magnitudes are slightly smaller than that with the PLS disagreement index.[2] For example, the out-of-sample R^2 by using the elastic net is 1.36% at the one-month horizon and 8.43% at the 12-month horizon, respectively, which are both significant at the 5% level. These results suggest genuine predictability of extant disagreement measures on market returns.

After providing evidence on the forecasting power of the disagreement index, we show that it is indeed consistent with the theory in Atmaz and Basak (2018). In their equilibrium model with infinite heterogeneous investors, Atmaz and Basak show that the overall effect of belief heterogeneity depends on two sufficient statistics, average bias and disagreement, which can be intuitively defined as the mean and cross-sectional standard deviation of investor expectation biases. Suppose investors are risk averse and exhibit a wealth effect that endogenously limits their risk taking. Atmaz and Basak (2018) show that, in equilibrium, investor disagreement affects stock returns via two channels. The first channel is a direct effect: disagreement represents uncertainty and investors require a higher expected return to hold a stock when disagreement on the stock increases, suggesting a positive disagreement-return relation. The second channel is an indirect effect: investor disagreement affects stock returns via an amplification effect on the average bias. That is, higher disagreement leads to higher average bias and more overvaluation, thereby suggesting a negative disagreement-return relation. With these two channels, Atmaz and Basak (2018) reconcile the mixed disagreement-return relation documented in the empirical finance literature (see, e.g., Chen, Hong, and Stein, 2002; Diether, Malloy, and Scherbina, 2002; Yu, 2011; Carlin, Longstaff, and Matoba, 2014). Since investors, regardless of whether they are sophisticated or not, are generally upward biased (see, e.g., Barber and Odean, 2008; Edelen, Ince, and Kadlec, 2016; DeVault, Sias, and Starks, 2019; Engelberg, McLean, and Pontiff, 2020), the second channel is more likely to dominate the first channel, thereby explaining why the disagreement index negatively predicts market returns in this paper.

In the following, we test four implications raised by Atmaz and Basak (2018). The first, and most important, implication is that the forecasting power of disagreement is asymmetric: it is stronger when investors are optimistic or among stocks with optimistic investor expectations, and weaker or insignificant otherwise. The intuition is that when investors are relatively pessimistic, the first and second channels have different forecasting signs and are likely to offset each other, making the disagreement-return

[2] The reason why the PLS disagreement index performs the best is that the PLS forecast is asymptotically consistent and can generate the minimum mean squared forecast error (MSFE) so long as the consistency condition is satisfied (Kelly and Pruitt, 2015).

relation insignificant. In contrast, when investors are overly optimistic, the second channel dominates the first channel, and as a consequence, disagreement negatively predicts future stock returns. To test this implication, we perform two tests. First, in time series, we show that the forecasting power of the disagreement index is concentrated in high investor sentiment periods and nonexistent in low sentiment periods. Second, cross-sectionally, we form ten decile portfolios based on firm level investor expectation, which is measured by the analyst long-term growth rate (LTG) forecast (Bordalo et al., 2019), and find that the disagreement index displays much stronger power in predicting portfolios with higher LTG forecasts, especially in high sentiment periods.

The second implication is that disagreement should be linked to investor optimism about market returns and ex post forecast errors. To capture investor expectations of market returns, we consider four measures, including aggregate analysts' return forecast (Engelberg, McLean, and Pontiff, 2020), Michigan survey of consumers attitudes (Das, Kuhnen, and Nagel, 2019), Graham-Harvey's survey of CFOs and Shiller's survey of individual investor confidence (Greenwood and Shleifer, 2014). We find that all these four measures positively correlate with the disagreement index. For example, a one-standard deviation increase in disagreement is associated with a 3.26% increase in the analysts' return forecast about the following 12-month market return. Since investor expectations are upward biased, the disagreement index negatively predicts ex post return forecast errors.

The third implication is that the predictive ability of disagreement on market returns is more likely to operate via a cash flow channel in the sense of Campbell (1991). According to Atmaz and Basak (2018), after positive cash flow news, investors whose beliefs are supported by the cash flow news become relatively wealthier, which makes them more optimistic about future cash flows or discount rates or both, and consequently, increases investor disagreement. For this reason, both the cash flow news and discount rate news can have a positive effect on disagreement. Empirically, we find that the cash flow news-based disagreement index displays strong forecasting power, while the discount rate news-based disagreement index does not.

The fourth, and last, implication is that disagreement plays an important role for the positive relationship between trading volume and market volatility. In Atmaz and Basak (2018), disagreement is the only driver of trading volume and market volatility. In the absence of disagreement, there is no trade and the market volatility is constant. In the presence of disagreement, however, both trading volume and market volatility increase as disagreement increases. Empirically, we find that the disagreement index is positively related to the volume-volatility elasticity. Intuitively, a one-standard deviation increase in disagreement predicts a 5.22% increase in the volume-volatility correlation in the following month. Overall, our empirical results are consistent with the theoretical implications of Atmaz and Basak (2018).

This paper contributes to the disagreement literature by showing that disagreement predicts market returns in- and out-of-sample. While many papers have explored the relationship between disagreement and stock returns at the firm level, studies at the market level are relatively rare. There are two exceptions, Yu (2011) on the stock market and Carlin, Longstaff, and Matoba (2014) on the mortgage market, but they do not investigate the out-of-sample forecasting power and the economic value for a real time investor. Also, Yu (2011) documents a negative forecasting sign whereas Carlin, Longstaff, and Matoba (2014) find a positive forecasting sign, and therefore, they interpret their results with different theories. This paper reconciles the seemingly conflicting results by using the unified theory of Atmaz and Basak (2018).

This paper is also related to the broad literature on return predictability. Since Welch and Goyal (2008), a large number of variables have been identified to significantly predict market returns in- and out-of-sample, such as the output gap (Cooper and Priestley, 2009), 52-week high and historical high (Li and Yu, 2012), aggregate implied cost of capital (Li, Ng, and Swaminathan, 2013), disaggregate book-to-market ratio (Kelly and Pruitt, 2013), aggregate short interest (Rapach, Ringgenberg, and Zhou, 2016), aggregate liquidity (Chen, Eaton, and Paye, 2018), fourth quarter consumption (Møller and Rangvid, 2015), metal prices (Jacobsen, Marshall, and Visaltanachoti, 2019), dividend-price ratio (Golez and Koudijs, 2018), variance risk premium (Pyun, 2019), gold to platinum price ratio (Huang and Kilic, 2019), aggregate skewness (Jondeau, Zhang, and Zhu, 2019), and many others. This paper does not aim at identifying a new variable to predict market returns, but proposes to aggregate predictive information from extant individual disagreement measures.

The rest of the paper is organized as follows. Section 2 considers 24 extant disagreement measures and shows that they fail to predict the stock market at the one- to 12-month horizons. Section 3 proposes a PLS disagreement index by aggregating information across individual measures and shows that it significantly predicts market returns in- and out-of-sample. Section 4 shows that the predictability of disagreement on market returns is consistent with the theoretical implications of Atmaz and Basak (2018), which is followed by Section 5 with a brief conclusion.

2. Forecasting Power of Extant Disagreement Measures

At the one- to 12-month horizons, we show in this section that most of the extant disagreement measures fail to predict market returns in-sample and none of them displays significant out-of-sample forecasting power.

2.1. Extant disagreement measures

We consider 24 disagreement measures, among which, 13 are based on professional forecasts on eight macro variables, two based on analyst forecasts, six based on household forecasts on macroeconomic conditions, and three based on market information. While these measures originate from different time periods, dating as early as December 1968, all of them conclude by December 2018.

2.1.1. 13 disagreement measures based on professional forecasts

The disagreements between professional forecasts on macro variables are based on the oldest quarterly survey of professional forecasters (SPF) in the US. The survey begins in 1968Q4 and is typically released in the mid-to-late second month of each quarter.[3] However, the accurate release dates before 1990Q2 are unavailable, and therefore, to be conservative, we assume that all surveys are made known in the last month of each quarter in our analysis. Also, because most of our analyses are performed on a monthly frequency, we convert the quarterly measures into monthly frequencies by assigning the most recent quarterly value to each month. For example, the observation in the first quarter of 2018 is assigned to the months of March, April, and May, respectively.

[3] Three exceptions with delayed releases are 1990Q2, 1996Q3, and 2013Q4, respectively.

We consider professional forecasts on eight macro variables, including gross domestic production (GDP), industrial production (IP), consumption (CON), investment (INV), housing starts (HSG), unemployment (UEP), consumer price index (CPI), and the 3-month Treasury bill rate (TBL). As the forecasts on GDP, IP, CON, INV, and HSG include both level and growth rate, we therefore have 13 disagreement measures in total. In each quarter, the forecasters predict macro variables for horizons ranging from the current up to four quarters ahead. Following Li (2016) and documents from the SPF, we define disagreement on each macro variable as the difference between the 75th percentile and 25th percentile forecasts for each horizon, taking the average across all horizons as the disagreement measure of that macro variable. In the literature, Anderson, Ghysels, and Juergens (2009), Bali, Brown, and Tang (2020), and many others use the SPF in a similar fashion in constructing aggregate uncertainty and disagreement measures, and find significant power for pricing the cross-section of stock returns.

2.1.2. Two disagreement measures based on analyst forecasts

Numerous studies have employed analyst forecast dispersion as the measure of investor disagreement. Following Yu and Yuan (2011) and Hong and Sraer (2016), we adopt the "bottom-up" approach by defining disagreement in month t as:

$$D_t^{\mathrm{Yu}} = \frac{\sum_i \mathrm{MKTCAP}_{i,t} \cdot D_{i,t}}{\sum_i \mathrm{MKTCAP}_{i,t}}, \tag{1}$$

and

$$D_t^{\mathrm{HS}} = \frac{\sum_i \beta_{i,t} \cdot D_{i,t}}{\sum_i \beta_{i,t}}, \tag{2}$$

where $D_{i,t}$ is the analyst forecast dispersion on the earnings per share (EPS) long-term growth rate (LTG) of firm i, and MKTCAP and $\beta_{i,t}$ are firm i's market cap and market beta. We only include common stocks (with CRSP item SHRCD = 10 or 11) listed on the NYSE, NASDAQ, and AMEX. As explained in Yu (2011), the LTG forecast features prominently in valuation models and is less affected by a firm's earnings guidance than the short-term forecast. When constructing D_t^{HS}, we follow Hong and Sraer (2016) and focus on all-but-micro stocks, stocks that are larger than the 20th percentile of the market cap of NYSE stocks. For each firm i in month t, we regress the daily returns of the past one year on contemporaneous and one to five lagged market returns, and use the sum of the slopes as the estimate of $\beta_{i,t}$.

2.1.3. Six disagreement measures based on household forecasts

Empirical studies often focus on how the trading of securities is affected by disagreement among institutional investors (see, e.g., Diether, Malloy, and Scherbina, 2002; Chen, Hong, and Stein, 2002; Jiang and Sun, 2014), but seldom explore the disagreement effect of households or retail investors on the stock market. From the Flow of Funds Accounts, households own about 60% of outstanding equities in the US (about 40% direct holding and additional 20% indirect holding through mutual funds), and therefore, their opinions should play a similarly important role as those of institutional investors. Li and Li (2017) show that the effect of household disagreement remains significant after controlling for professional forecast dispersions and even dominates the professional forecast dispersion measures.

We construct household disagreement based on the Michigan survey of consumers attitudes (SCA). The SCA starts conducting monthly surveys on a minimum of 500 households in January 1978, with accurate release dates available after January 1991. In each survey, the SCA collects responses to 50 core questions that are generally related to households opinions on current economic conditions and their expectations about future economic conditions. In this paper, we construct our disagreement measures from six questions. The first question is about households' realized opinions on current personal financial condition compared with those of the prior year, while the other five are about households' expectations about the following year, consisting of the expected personal financial condition, business condition, unemployment condition, interest rate condition, and house purchase condition.

For each question, the surveyed households' replies are classified into three categories, better (good), same (depends), and worse (bad). In a consistent way, we rename the categories as positive, neutral, and negative, respectively, and define the proportion of each category as P_{positive}, P_{neutral}, and P_{negative}. We follow

Li and Li (2017) and define disagreement as the unevenly weighted negative Herfindahl index,

$$D = -\sum w_i P_i^2, \quad i = \text{positive, neutral, negative}, \tag{3}$$

where w_i is the weight of each category as $w_{\text{positive}} = 1$, $w_{\text{neutral}} = 2$, and $w_{\text{negative}} = 1$. We assign a higher weight to the neutral category to avoid the unfavourable feature of the evenly weighted Herfindahl index. For example, if 50% of households indicate the positive response and 50% indicate the negative response, the evenly weighted Herfindahl index would be the same as if the responses are 50% positive and 50% neutral. However, the disagreement in the former situation is obviously more dispersed than in the latter.

2.1.4. *Disagreement based on unexplained stock trading volume*

Ajinkya, Atiase, and Gift (1991) find that high trading volume is associated with an increase in the analyst forecast dispersion, suggesting that trading volume may measure investor disagreement. We follow Garfinkel (2009) and construct a disagreement measure with the standardized unexplained volume. Specifically, we obtain the monthly aggregate trading volume data of the NYSE from Pinnacle and define volume as the residual of applying an AR(4) to the log turnover with the past 120-month observations (Hamilton, 2018).[4] Then, we run the following time series regression with data from the past 120-month period at the end of each month on a rolling basis as

$$\text{Volume}_t = \alpha + \beta_1 R_t^+ + \beta_2 R_t^- + \varepsilon_t, \tag{4}$$

and use the last value of the residuals as the estimate of unexpected volume. In Eq. (4), the plus and minus signs in the superscript indicate that market returns can be either positive or negative, and capture the empirical fact that positive and negative returns generate different levels of trading volume. Thus, investor disagreement can be defined by the standardized unexplained volume:

$$D_t^{\text{SUV}} = \frac{\varepsilon_t}{\sigma_{\varepsilon,t}}, \tag{5}$$

[4]The results are quantitatively similar with alternative specifications, such as using AR(12) or using the past 60-month observations.

where $\sigma_{\varepsilon,t}$ is the standard deviation of the regression residuals.

2.1.5. Disagreement based on idiosyncratic volatility

Inspired by theoretical studies that construct a close relation between belief dispersion and volatility, Boehme, Danielsen, and Sorescu (2006) and Berkman et al. (2009) propose idiosyncratic volatility as a disagreement measure at the firm level. We extend this measure to the market level. Specifically, following Ang et al. (2006), we regress daily stock returns on the Fama and French (1993) three factors with a 12-month rolling window and estimate the firm level idiosyncratic volatility at the end of each month. We then define investor disagreement as the value-weighted idiosyncratic volatility.

2.1.6. Disagreement based on option open interest

Disagreement can also be constructed from the option market. Investors who hold call options have a bullish view, whereas investors who hold put options have a bearish view. Following Ge, Lin, and Pearson (2016), we define disagreement as one minus the scaled difference between the OEX call and put open interests:

$$D_t^{\text{OID}} = 1 - \frac{|\text{COI}_t - \text{POI}_t|}{|\text{COI}_t + \text{POI}_t|}, \tag{6}$$

where COI_t (POI_t) is the call (put) option open interest. The scaled call and put option open interest difference $|\text{COI}_t - \text{POI}_t|/|\text{COI}_t + \text{POI}_t|$ ranges from zero to one. The explanation is that when disagreement is low, investors' beliefs polarize into bullish or bearish extremes. The difference between the call and put option open interests diverges and the scaled difference approaches one. As a result, one minus this scaled difference is accordingly low. When disagreement is high, the opinions between optimists and pessimists diverge. The call and put option open interests should be commensurable. The scaled difference between the call and put option open interests approaches zero. Hence, one minus the scaled difference is accordingly large.

2.2. *Summary statistics*

Table 1 presents summary statistics of the 24 disagreement measures, including the sample period, mean, standard deviation, minimum, maximum, skewness, and kurtosis. It is apparent that the scales across disagreement measures vary dramatically due to the nature of macro variables. For instance, the mean of disagreement on GDP is 61.32 billion, while the mean of disagreement on TBL is only 0.46%. Thus, to make them comparable and to avoid forward-looking bias, we standardize each disagreement measure in month t by its last six-year mean and standard deviation, with a requirement of at least one year data. For this reason, the analyses in all other tables start from December 1969. To remove possible fundamental information, we measure disagreement as the residuals from the regression of each individual disagreement measure on the six macro variables in Baker and Wurgler (2006), consisting of the growth of industrial production, the growth of durable consumption, the growth of nondurable consumption, the growth of service consumption, the growth of employment, and a dummy variable for NBER dated recessions (we recursively do so when performing out-of-sample tests).

Table A1 in the Online Appendix presents pairwise correlations between individual disagreement measures. Most of the measures are positively correlated, with several exceptions of negative values. For example, professional forecast dispersions are generally positively correlated, and they are also positively correlated with the two analyst forecast dispersion measures. Business condition forecast dispersion is an exception, and it is negatively correlated with other measures in general. Overall, this table indicates that extant measures capture both the common and different aspects of individual disagreement measures across the whole economy, and an individual measure is unlikely to completely capture the aggregate effect of disagreement on the stock market.

2.3. *Forecasting market returns with extant disagreement measures*

We explore the forecasting power of disagreement on market returns with the following predictive regression,

$$R_{t+1} = \alpha + \beta D_t + \varepsilon_{t+1}, \tag{7}$$

where R_{t+1} is the log excess return of the S&P 500 index in month $t+1$ and D_t is one of the 24 individual disagreement measures.[5] When the forecast horizon is h months, we denote the cumulative market return as $R_{t,t+h} = \sum_{j=1}^{h} R_{t+j}$.

The predictive power is assessed based on the regression slope β or the R^2 statistic. If β is significantly different from zero or if the R^2 is significantly larger than zero, it means that D_t is a predictor of the market returns. The out-of-sample forecast of the next one-month market return is recursively computed as

$$\hat{R}_{t+1} = \hat{\alpha}_t + \hat{\beta}_t D_t, \tag{8}$$

where $\hat{\alpha}_t$ and $\hat{\beta}_t$ are the ordinary least squares estimates of α and β based on data from the start of the available sample through month t. The in-sample forecast is computed similarly as before, except that $\hat{\alpha}_t$ and $\hat{\beta}_t$ are replaced by those estimated by using the entire sample. For ease of exposition, we always normalize the time series of disagreement in all the in-sample predictive regressions, so that the regression slope measures the change in response to a one-standard deviation increase in disagreement.

We use the out-of-sample R^2 statistic in Campbell and Thompson (2008) as the out-of-sample performance evaluation criterion and define it as:

$$R^2_{OS} = 1 - \frac{\sum_{t=M+1}^{T} (R_t - \hat{R}_t)^2}{\sum_{t=M+1}^{T} (R_t - \bar{R}_t)^2}, \tag{9}$$

where M is the size for in-sample parameter training and $T-M$ is the number of out-of-sample observations. $\hat{R}_t$ is the market return forecast with Eq. (8), and $\bar{R}_t$ is the historical return mean, both of which are estimated using data up to month $t-1$. If D_t is a valid predictor, its MSFE is lower than the MSFE with the historical return mean and the R^2_{OS} will be positive. Campbell and Thompson (2008) show that a monthly R^2_{OS} of 0.5% can generate a significant economic value. The null hypothesis of interest is therefore $R^2_{OS} \leq 0$ against the alternative hypothesis that $R^2_{OS} > 0$. We test this hypothesis by using the MSFE-adjusted statistic as proposed by Clark and West (2007).

Panel A of Table 2 presents the regression slope β, Newey-West t-value, in-sample R^2, and out-of-

[5]For brevity, returns in this paper always refer to excess returns except for Section 4.3, where we follow Campbell (1991) to decompose the total market returns into three subcomponents.

sample R^2_{OS}. Throughout this paper, the out-of-sample period is from February 1991 to December 2018 because the accurate release dates of household dispersion measures are only available as of January 1991. 20 out of 24 disagreement measures have a negative forecasting sign, among which, however, only four measures reveal significant in-sample predictive power at the 5% level, which are the housing starts forecast dispersion, CPI forecast dispersion, TBL forecast dispersion, and business condition forecast dispersion. The out-of-sample performance is more dismal, with all R^2_{OS} values being negative. For instance, the TBL forecast dispersion exhibits the highest in-sample R^2 of 1.94%, but generates a -4.21% out-of-sample R^2_{OS}. These results suggest that none of the extant individual disagreement measures can predict market returns in real time at the one-month forecast horizon.

Panels B and C of Table 2 present similar results as Panel A when the forecast horizon is extended to three months or 12 months. The in-sample regression slopes are seldom significant and the R^2_{OS} values are all negative. For in-sample prediction over the 1981:12–2005:12 sample period, Yu (2011) shows that analyst forecast dispersion exhibits insignificant forecasting power at the one-month horizon but significant forecasting power at the 12-month or longer horizons. Panel C suggests that when we extend the sample to the most recent period, analyst forecast dispersion becomes insignificant. Yu (2011) does not show out-of-sample forecasting performance and our results suggest that analyst forecast dispersion cannot generate meaningful real time forecasting value either.

Overall, Table 2 shows that while all of the extant disagreement measures may have cross-sectional forecasting power, they are unable to predict market turns in general, especially for out-of-sample forecasting.

3. PLS disagreement index

In this section, we construct a disagreement index by aggregating information across individual disagreement measures and show that it significantly predicts market returns in- and out-of-sample.

3.1. Methodology

The method we choose for information aggregation is PLS, which consists of three steps. In the first step, we run a time series regression of each individual disagreement measure on the realized subsequent market returns (as a proxy of expected return) with the full sample, denoted as:

$$D^k_{t-1} = \pi_{k,0} + \pi_k R_t + u_{k,t-1}, \quad k = \text{GDP}, \cdots, \text{OID}, \tag{10}$$

where π_k captures the sensitivity of proxy D^k_{t-1} to the expected market return. In the second step, we run a cross-sectional regression of D^k_t on $\hat{\pi}_k$ at the end of each month:

$$D^k_t = a_t + D_t \hat{\pi}_k + \nu_{k,t}, \tag{11}$$

where the regression slope D_t is the PLS disagreement index in month t. In the last and third step, to predict R_{t+1}, we run the following predictive regression:

$$R_{t+1} = \alpha + \beta D_t + \varepsilon_{t+1}. \tag{12}$$

The above three steps are for in-sample analysis. For out-of-sample forecasting, the standard approach is to repeat the three steps by truncating the observations that are not known at month $t+1$. Specifically, consider a forecast for return R_{t+1} that is realized in month $t+1$. A properly constructed forecast can only use information known through month t. In the first step, the latest return that can be used on the right-hand side is R_t and the last observation of disagreement on the left-hand side is, therefore, D^k_{t-1}. In the second step, the cross-sectional regressions are run from months 1 through t. In the last step, the latest return on the left-hand side entering the predictive regression is R_t and the forecast for R_{t+1} is $\hat{\alpha}_t + \hat{\beta}_t D_t$, where $\hat{\alpha}_t$ and $\hat{\beta}_t$ are the estimates using information up to month t. In summary, for out-of-sample forecasting, all inputs to the forecast are constructed using data that are observed no later than month t.

To iron out extreme outliers, we smooth the disagreement index with its six-month moving average values and plot the time series in Fig. 1. There are two interesting observations. First, the disagreement index is time-varying and does not diminish over time, which is consistent with the finding in Hong and Stein (2007) and Cookson and Niessner (2020) that permanent disagreement can arise even when investors have common priors and observe the same time series of public information, so long as they interpret information differently. Second, the disagreement index value can be large in bad times, such as the recessions of 1981 to 1982 and 2007 to 2008, and also in good times, such as the dot-com boom of the late 1990s. This evidence is consistent with the beta-weighted analyst forecast dispersion in Hong and Sraer (2016).

3.2. *Forecasting performance*

This section explores the in- and out-of-sample forecasting performance of the disagreement index. For comparison, we consider two alternative disagreement indexes as benchmarks. The first alternative disagreement index is constructed based on PCA, which extracts the first principal component of the 24 individual disagreement measures as the aggregate index. This method has been widely used in finance, such as Baker and Wurgler (2006) who construct an investor sentiment index as the first principal component of six individual sentiment proxies. The second alternative disagreement index is constructed by simply equal-weighting the 24 (standardized) individual disagreement measures. The intuition is that if each individual measure is unbiased, equal-weighting will efficiently reduce the idiosyncratic errors.

Panel A of Table 3 reports the results of predicting market returns with the three disagreement indexes. At the one-month horizon, a one-standard deviation increase in disagreement leads to a 0.38% decrease in the next one-month market return with the PCA disagreement index (t-value $= -1.96$), a 0.60% decrease with the equal-weight disagreement index (t-value $= -2.87$), and a 0.83% decrease with the PLS disagreement index (t-value $= -3.96$). When turning to out-of-sample forecasting, the R^2_{OS} with the PCA disagreement index is 0.20% and not significant. In contrast, the R^2_{OS} is 0.90% with the equal-weight disagreement index and 1.56% with the PLS disagreement index, which are both significant at the 5% level.

Panels B and C of Table 3 report the results when the forecast horizons are three and 12 months, respectively. In these two cases, all the three disagreement indexes display significant in- and out-of-sample forecasting power. For example, a one-standard deviation increase in disagreement leads to 2.92%, 4.93%, and 7.04% decreases in the next 12-month market returns with the three disagreement indexes, respectively. The R^2 and R^2_{OS} of the PLS disagreement index, 13.88% and 13.26%, are comparable with the most powerful predictor to date, the aggregate short interest in Rapach, Ringgenberg, and Zhou (2016), whose corresponding values are 12.89% and 13.24%, respectively.

Why does the PLS disagreement index have stronger forecasting power than the two alternative disagreement indexes? The reason is that while the PCA and equal-weight disagreement indexes can efficiently reduce the idiosyncratic measurement and observation errors in the individual disagreement measures, they cannot tease out the common errors that are unrelated to expected market returns. In contrast, as a supervised learning technique, the PLS aggregates information relevant to expected market returns and is supposed to perform the best.

To better understand their differences in forecasting power, Fig. 2 depicts the forecasted 3-month market returns based on the PCA, equal-weight, and PLS disagreement indexes for the 1991:02–2018:12 out-of-sample period (the results with other forecasting horizons are similar and omitted for brevity). The PLS disagreement index generates more volatile forecasts than the other two and naturally does a better job in capturing the variation of expected market returns. To explore the dominant variables in constructing the PLS index, Fig. 3 exhibits the top five individual disagreement measures at each point in time when conducting the out-of-sample forecasting. In general, consumption growth forecast dispersion, TBL forecast dispersion, realized personal financial improvement dispersion, business condition forecast dispersion, and house purchase condition forecast dispersion are more likely to be chosen.

In this paper, we measure disagreement with the first PLS factor. One natural question is how many PLS factors we should use in our setting. Following Kelly and Pruitt (2015), we calculate the Bayesian Information Criterion (BIC) via the Krylov representation method and find that only one factor is chosen statistically. To see this is true, Table A2 reports the R^2s and R^2_{OS}s with the first to sixth moment PLS factors in predicting market returns, where the PLS factors are extracted by using the automatic proxy-selection algorithm in Kelly and Pruitt (2015). The results show that the second to sixth PLS factors do not have any in- and out-of-sample forecasting power, thereby supporting our choice of focusing on the first PLS factor.

In summary, extant disagreement measures do have a common component that is able to predict market returns, and the forecasting power depends on how we aggregate information across individual measures.

3.3. *Controlling for economic predictors*

This section examines whether the forecasting power of the disagreement index on market returns remains significant after controlling for extant economic predictors. In so doing, we consider the 14 economic predictors in Welch and Goyal (2008), output gap in Cooper and Priestley (2009), and aggregate short interest in Rapach, Ringgenberg, and Zhou (2016), and run the following regression:

$$R_{t+1} = \alpha + \beta D_t + \psi Z_t + \varepsilon_{t+1}, \tag{13}$$

where Z_t is one of the 16 economic predictors.

Table 4 reports the results. For comparison, Panel A considers the predictive power of the 16 economic predictors and shows that only four variables are able to significantly predict market returns, including the long-term bond return, term spread, output gap, and aggregate short interest. Panel B shows that controlling for extant economic predictors does not reduce the forecasting power of the disagreement index. For example, when controlling for output gap, the corresponding slope slightly decreases to -0.75 in absolute value and is significant at the 1% level. When the disagreement index and aggregate short interest are jointly used as predictors, the regression slope on the disagreement index remains at a value of -0.84, which is almost the same as without controlling for the aggregate short interest. In the last row, we consider a kitchen sink regression by including all the economic predictors. To handle highly correlated predictors, we estimate the regression slopes with the elastic net method, which has been successfully used in Rapach, Strauss, and Zhou (2013) and Kozak, Nagel, and Santosh (2020) for time series and cross-sectional predictability. The result shows that the forecasting power of the disagreement index remains quantitatively the same as the case of using the disagreement index alone. Therefore, the predictive ability of the disagreement index is not subsumed by extant economic predictors and it contains independent information beyond these economic predictors.

3.4. Controlling for uncertainty measures

In the literature, disagreement has two alternative interpretations: investor heterogeneity and uncertainty. For example, Anderson, Ghysels, and Juergens (2005) show theoretically and empirically that investor heterogeneity matters for asset pricing and measure it with analyst forecast dispersion. In contrast, Wang, Yan, and Yu (2017) proxy analyst forecast dispersion for uncertainty. While these two alternative explanations can be reconciled by the theory of Atmaz and Basak (2018), it remains empirically interesting to explore whether the disagreement index is different from extant uncertainty measures. Specifically, we employ eight uncertainty measures, including economic uncertainty (Bali, Brown, and Caglayan, 2014), treasury implied volatility (Choi, Mueller, and Vedolin, 2017), financial uncertainty and macro uncertainty (Jurado, Ludvigson, and Ng, 2015), economic policy uncertainty (Baker, Bloom, and Davis, 2016), news implied volatility (Manela and Moreira, 2017), sample variance (Welch and Goyal, 2008), and VIX.

Panel A of Table 5 reports correlations of the disagreement index with the eight uncertainty measures. Consistent with Anderson, Ghysels, and Juergens (2009), the disagreement index does positively correlate with uncertainty except for economic policy uncertainty. For example, the correlation of the disagreement index is 0.33 with Choi, Mueller, and Vedolin's (2017) treasury implied volatility and 0.24 with Jurado, Ludvigson, and Ng's (2015) macro uncertainty.

Then, we investigate the forecasting power of disagreement by controlling for macro uncertainty as:

$$R_{t+1} = \alpha + \beta D_t + \psi U_t + \varepsilon_{t+1}, \tag{14}$$

where U_t is one of the eight uncertainty measures. As a benchmark, Panel B of Table 5 shows that extant uncertainty measures cannot significantly predict market returns with one exception, namely financial uncertainty in Jurado, Ludvigson, and Ng (2015). However, the forecasting sign of financial uncertainty seems inconsistent with asset pricing theories that higher uncertainty implies higher risk premium. Panel C shows that the disagreement index remains significant in predicting market returns after controlling for extant uncertainty measures. For example, in the kitchen sink regression that includes all the eight uncertainty measures, the slope on the disagreement index is still −0.89, close to the case without any controls (-0.83%). Overall, while the disagreement index is positively correlated with extant uncertainty measures in general, it contains different information for future market returns.

3.5. *Economic value with disagreement prediction*

In this section, we examine the economic value of forecasting market returns with the disagreement index from the perspective of investing. Following Ferreira and Santa-Clara (2011) and many others, we explore the certainty equivalent return (CER) gain and Sharpe ratio. The higher the CER gain and Sharpe ratio, the larger the risk-rewarded returns by using the disagreement index.

Suppose a mean-variance investor invests her wealth between the stock market and the one-month T-bill rate. At the start of each month, she allocates a proportion of w_t to the stock market to maximize her next one-month expected utility

$$U(R_p) = \mathrm{E}(R_p) - \frac{\gamma}{2}\mathrm{Var}(R_p), \quad (15)$$

where R_p is the return of the investor's portfolio, $\mathrm{E}(R_p)$ and $\mathrm{Var}(R_p)$ are the mean and variance of the market returns, and γ is the investor's risk aversion.

Let R_{t+1} and $R_{f,t+1}$ be the market return and T-bill rate. The investor's portfolio return at the end of each month is

$$R_{p,t+1} = w_t R_{t+1} + R_{f,t+1}, \quad (16)$$

where $R_{f,t+1}$ is known at t. With a simple calculation, the optimal portfolio weight is

$$w_t = \frac{1}{\gamma}\frac{\hat{R}_{t+1}}{\hat{\sigma}^2_{t+1}}, \quad (17)$$

where $\hat{R}_{t+1}$ and $\hat{\sigma}^2_{t+1}$ are the investor's estimates on the mean and variance of the market returns based on information up to time t.

The CER of the portfolio is

$$\text{CER} = \hat{\mu}_p - \frac{\gamma}{2}\hat{\sigma}_p^2, \tag{18}$$

where $\hat{\mu}_p$ and $\hat{\sigma}_p^2$ are the mean and variance of the investor's portfolio over the out-of-sample evaluation period. The CER can be interpreted as the compensation to the investor for holding the stock market. The difference between the CERs for the investor using the predictive regression based on disagreement and the historical return mean is naturally an economic measure of predictability significance.

Table 6 presents the economic value generated by optimally trading on the disagreement index for the investor with a risk aversion of 3 and 5, respectively. That is, we report the CER difference between the strategy using the disagreement forecast and the strategy using the historical return mean. We annualize the CER by multiplying it by 1,200 so that the CER difference denotes the percentage gain per year for the investor to use the disagreement index forecast instead of the historical return mean. Following Campbell and Thompson (2008), we assume that the investor uses a ten-year moving window of past monthly returns to estimate the variance of market returns, and constraints w_t to lie between 0 and 1 in order to exclude extreme cases.

For comparison, we also consider the alternative PCA and equal-weight disagreement indexes. The results show that among the three disagreement indexes, the PLS disagreement index performs the best and the PCA index performs the worst, which is consistent with the results in Table 3 that both the PLS and equal-weight disagreement indexes can generate significant R_{OS}^2s at the one-month horizon. In Panel A, when there is no transaction cost, the annualized CER gain by using the PLS disagreement index is 2.50%, suggesting that investing with the PLS disagreement index forecast can generate 2.50% more risk-adjusted return relative to the historical return mean. The monthly Sharpe ratio is 0.18 and much higher than the market Sharpe ratio of 0.10 in our sample period. When there is a transaction cost of 50 basis points, the CER gain by using the PLS disagreement index is 1.92%, which is still economically sizeable. The corresponding Sharpe ratio is 0.16. Panel B shows similar results when the investor's risk aversion is 5. For example, the CER gain is 2.68% without transaction costs and is 1.88% with atransaction cost of 50 basis points. In summary, the PLS disagreement index is able to deliver considerable economic value for a mean-variance investor.

3.6. *Alternative econometric methods*

In the previous sections, we have shown that market returns can be significantly predicted by the PLS disagreement index. This section examines whether the result is robust to alternative econometric methods. Particularly, we consider six LASSO-related machine learning methods: equal-weight LASSO, combination LASSO (Han, He, Rapach, and Zhou, 2019), encompassing LASSO (Han, He, Rapach, and Zhou, 2019), adaptive LASSO (Freyberger, Neuhierl, and Weber, 2020), egalitarian LASSO (Diebold and Shin, 2019), elastic net (Kozak, Nagel, and Santosh, 2020). These six methods are introduced in detail in the Online Appendix.

Table 7 reports the results. There are three observations. First, the out-of-sample R^2_{OS}s are all significant at the one- to 12-month horizons, which confirms the predictability of the PLS disagreement index on market returns. For example, with the elastic net method, the R^2_{OS}s are 1.36% and 8.43% at the one- and 12-month horizons, respectively, and significant at the 5% level. Second, the disagreement index by using the equal-weight LASSO method significantly improves the forecasting power of the equal-weight disagreement index in Section 3.2. The R^2_{OS} increases from 0.90% to 1.26% at the one-month horizon and from 9.41% to 12.08% at the 12-month horizon, thereby suggesting that machine learning techniques are useful for return predictability. Finally, while these six alternative methods work well for predicting market returns, they underperform the PLS. This finding lends empirical support to Kelly and Pruitt (2015) that the PLS forecast is asymptotically consistent and will generate the minimum MSFE so long as the consistency condition is satisfied.

To explore which individual disagreement measures are important in predicting market returns, Figs. A1 and A2 plot the selected measures and their frequencies according to the six LASSO-related methods at each point in time when conducting out-of-sample forecasting. Over the 1991:02–2018:12 out-of-sample period, some measures are commonly and frequently selected by all the methods. For example, the housing starts forecast dispersion and business condition forecast dispersion are the two most important individual measures and selected by all the six methods with a probability of 100%. The next three important measures are the CPI forecast dispersion, TBL forecast dispersion, and value-weighted analyst forecast dispersion, which are commonly selected with a probability of around 50%. In contrast, the disagreement measures based on the standardized unexplained volume, idiosyncratic volatility, and option open interest are rarely selected by any of the six LASSO-related methods. These results are generally consistent with Fig. 3 and suggest that disagreement measures that are based on professional and household forecasts are equally important in predicting market returns, whereas measures that are based on market information are not.

4. Economic implications

This section shows that the predictability of the disagreement index is consistent with the theory in Atmaz and Basak (2018). In particular, we test four implications. The disagreement index 1) predicts market returns asymmetrically, with stronger power in high sentiment periods, 2) negatively predicts investors' ex post return forecast errors, 3) predicts market returns via a cash flow channel in the sense of Campbell (1991), and 4) can explain the positive relationship between trading volume and market volatility.

4.1. *Asymmetric forecasting power*

One key implication in Atmaz and Basak (2018) is that disagreement should display an asymmetric forecasting pattern in different market states. The reason is that when investors are relatively pessimistic, the first and second channels have different forecasting signs and are likely to offset each other, making the disagreement-return relation insignificant. In contrast, when investors are overly optimistic, the second channel dominates the first channel and hence disagreement should negatively predict future stock returns. In the following, we use the investor sentiment index of Baker and Wurgler (2006) to test whether the forecasting power of the disagreement index is asymmetric over the high and low sentiment periods, where a month is defined as high if the past 18-month average sentiment index is positive, and low otherwise. The results by using the PLS sentiment index in Huang et al. (2015) are quantitatively similar and omitted for brevity.

4.1.1. *Time series evidence*

Following Rapach, Strauss, and Zhou (2010), we calculate the in-sample R^2s in high and low sentiment periods as:

$$R_c^2 = 1 - \frac{\sum_{t=1}^{T} S_t^c (\hat{\varepsilon}_t)^2}{\sum_{t=1}^{T} S_t^c (R_t - \bar{R})^2}, \quad c = \text{high, low}, \tag{19}$$

where $S_t^{\text{high,}}$ (S_t^{low}) is an indicator that takes a value of one when month t is in a high (low) sentiment period and zero otherwise, $\hat{\varepsilon}_t$ is the fitted residual based on the in-sample estimates, $\bar{R}$ is the full-sample mean of R_t, and T is the number of observations for the full sample. Note that, unlike the full-sample R^2 statistic, the R^2_{high} and R^2_{low} statistics can be either positive or negative. Similarly, we can also calculate the R^2_{OS} in high and low sentiment periods separately. Another way to test the forecasting asymmetry is to run the following state-dependent regression:

$$R_{t+1} = \alpha + \beta_{\text{high}} S_t^{\text{high}} D_t + \beta_{\text{low}} S_t^{\text{low}} D_t + \varepsilon_{t+1}. \tag{20}$$

Table 8 shows that the forecasting power of the disagreement index is concentrated in high sentiment periods. In Panel A, the R^2 and R^2_{OS} are 5.28% and 3.69% in high sentiment periods, and 0.80% and −0.55% in low sentiment periods, respectively. In Panel B, the regression slope of the disagreement index in high sentiment periods is −1.12 with a t-value of −4.71, but it is only −0.42 with an insignificant t-value of −1.21 in low sentiment periods. Therefore, the predictability of disagreement on market returns is asymmetric and concentrated in high sentiment periods.

4.1.2. *Cross-sectional evidence*

In a multiple-stock economy, Atmaz and Basak (2018) suggest that the forecasting power of disagreement should be asymmetric across stocks, stronger among stocks with optimistic investor expectations and weaker or insignificant among stocks with pessimistic investor expectations. Different from Miller (1977), this implication holds even in the absence of short-sale constraints, so long as there are infinite risk averse investors. For this reason, we test the implication based on portfolios sorted by expectation directly.

Following Bordalo et al. (2019), we proxy the analyst LTG forecast for investor expectation at the firm level and construct 10 decile portfolios at the end of December for each year. The portfolios are subsequently held for one year. In the 1982–2018 period, the portfolio with low LTG forecast earns an annual return of 13.58% and the portfolio with high LTG forecast earns an annual return of 7.89%, with the difference between the high and low LTG forecast portfolios equal to 5.69% per year. Panel A of Fig. 4 plots the regression slopes of predicting the ten decile portfolio returns with the disagreement index. Apparently, the slope increases in magnitude from −0.61 for the portfolio with low LTG forecast to −1.09% for the portfolio with high LTG forecast.

Also, to explore the time-varying effect of the average bias of investor expectations, we run the following state-dependent regression:

$$R_{i,t+1} = \alpha_i + \beta_{\text{high},i} S_t^{\text{high}} D_t + \beta_{\text{low},i} S_t^{\text{low}} D_t + \varepsilon_{i,t+1}. \tag{21}$$

Panels B and C of Fig. 4 plot the regression slopes in high and low sentiment periods, respectively. As expected, the forecasting power of the disagreement index is concentrated in high sentiment periods. β_{high} monotonically increases in magnitude from −0.57 for the low LTG forecast portfolio to −1.96 for the high LTG forecast portfolio. In contrast, β_{low} is flat and displays a slightly upward trend. Also, Fig. A3 in the Online Appendix shows that portfolios with lower institutional ownership, higher beta, or higher IVOL earn lower average returns among high disagreement periods and confirms the argument that disagreement and arbitrage costs have an interaction effect (see, e.g., Hong and Sraer, 2016). In general, the predictability of disagreement is both time series and cross-sectionally asymmetric, stronger among stocks with optimistic cash flow expectation in high sentiment periods.

4.2. Disagreement and expectations of market returns

In the previous section we have linked disagreement with investor expectation (measured by investor sentiment) in an indirect manner. In this section we examine the relation of disagreement with investor expectations of market returns directly. According to Atmaz and Basak (2018), since disagreement amplifies investor optimism, it should be negatively related to ex post return forecast errors.[6]

Specifically, we consider four measures of investor expectations of 12-month ahead market returns. The first measure is the value-weighted aggregate analysts' return forecast, where the analysts' return forecast of an individual stock is defined as the mean of 12-month ahead analysts' target prices divided by current price (Engelberg, McLean, and Pontiff, 2020), and the target prices are restricted to those reported in the past one month. The second measure is Michigan survey of consumers attitudes. Following Das, Kuhnen, and Nagel (2019), we use the responses to "Suppose that tomorrow someone were to invest one thousand dollars in a type of mutual fund known as a diversified stock fund. What do you think is the percent chance that this one thousand dollar investment will increase in value in the year ahead, so that it is worth more than one thousand dollars one year from now?" The third and last measures are Graham-Harvey's survey of CFOs and Robert Shiller's survey of individual investor confidence in the stock market, which are constructed strictly following Greenwood and Shleifer (2014). The first three expectation measures have the same measurement unit as the realized market returns, whereas the last one, Shiller's survey, is based on binary variables. As such, when calculating the ex post return forecast errors, we project the aggregate analysts return forecast on Shiller's survey, so that the projected time series has the same measurement unit as the realized returns.

Table 9 reports the results. In Panel A, the disagreement index is positively related to the four investor expectation measures, with significant correlations ranging from 0.24 to 0.35. In terms of economic magnitude, a one-standard deviation increase in disagreement is associated with 3.26%, 2.71%, 1.57%, and 2.16% increases in investor expectations of 12-month ahead market returns with the analyst's return forecast, Michigan survey, Graham-Harvey's survey, and Shiller's survey, respectively. Untabulated results also confirm Greenwood and Shleifer (2014) that these investor expectation measures negatively predict 12-month ahead market returns. In Panel B, when regressing ex post return forecast errors on the disagreement index, we find that all the regression coefficients are significantly negative. For example, a one-standard deviation increase in disagreement is associated with a 7.34% increase in the analysts' return forecast error (i.e., the deviation of analysts' return forecast from the realized return increases by 7.34%), and the disagreement index explains about one quarter of variations of analysts' return forecast errors (i.e., $R^2 = 23.26\%$). Overall, Table 9 suggests that disagreement is closely linked with investor expectations of

[6] We thank the anonymous referee for this and many other intriguing suggestions

market returns, for both sophisticated and retailed investors.

4.3. *Relation of disagreement with cash flow news and discount rate news*

The section examines the contemporaneous relation of disagreement with cash flow news and discount rate news, so that we can disentangle the forecasting channel in the sense of Campbell (1991).

Following Campbell (1991), the log total market return can be decomposed into three components,

$$\tilde{R}_t \approx \mathrm{E}_{t-1}(R_t) + \mathrm{CF}_t - \mathrm{DR}_t, \tag{22}$$

where CF_t and DR_t are cash flow news and discount rate news, and they are defined as

$$\mathrm{CF}_t = (\mathrm{E}_t - \mathrm{E}_{t-1}) \sum_{j=0}^{\infty} \kappa^j \Delta d_{t+j} = (\Delta d_t - \mathrm{E}_{t-1} \Delta d_t) + (\mathrm{E}_t - \mathrm{E}_{t-1}) \sum_{j=1}^{\infty} \kappa^j \Delta d_{t+j}, \tag{23}$$

$$\mathrm{DR}_t = (\mathrm{E}_t - \mathrm{E}_{t-1}) \sum_{j=1}^{\infty} \kappa^j \tilde{R}_{t+j}, \tag{24}$$

where Δd_{t+j} and $\tilde{R}_{t+j}$ are the log dividend growth and log total market return at time $t+j$, and κ is a log-linearization constant slightly less than one. In another word, CF_t and DR_t are return innovations due to updates in expectations of current and future cash flows and future expected returns, respectively.

Atmaz and Basak (2018) posit that, after positive cash flow news, say $\Delta d_t - \mathrm{E}_{t-1}\Delta d_t > 0$, investors whose beliefs are supported by the cash flow news become relatively wealthier, which makes them more optimistic about future cash flows or discount rates or both, and consequently, increases disagreement. In contrast, after negative cash flow news, investors who have been optimistic become relatively poorer and pessimistic, thereby shrinking disagreement. This suggests that both CF_t and DR_t can positively affect disagreement and drive its movements.

To explore which component is the main driver of disagreement, we use contemporaneous CF_t and DR_t as the targets in Eq. (10) to extract a cash flow news-based PLS disagreement index and a discount rate news-based PLS disagreement index, and then examine their power in predicting future market returns. The results are reported in Table 10, where the cash flow news and discount rate news are estimated based on individual VARs comprising the total market return, dividend-price ratio, and one of the rest 15 economic predictors explored in Table 4. We always include the dividend-price ratio in the VARs because Engsted, Pedersen, and Tanggaard (2012) show that it is important to include this variable to properly estimate the cash flow and discount news components. In the last row of Table 10, we also consider the decomposition based on a VAR comprising the total market return, log dividend-price ratio, and the first three principal components extracted from the 15 economic predictors.

Table 10 shows that only the cash flow news-based disagreement index has forecasting power on market returns. For example, when the cash flow news and discount rate news are estimated with the VAR comprising the total market return and dividend-price ratio, a one-standard deviation increase in the cash flow news-based disagreement index predicts a 0.65% decrease in the next one-month market return, while the discount rate news-based disagreement index displays nil power. This finding echoes Section 3.2 in which we show statistically that there is only one PLS factor exhibiting forecasting power on market returns. Thus, we conclude that the ability of disagreement in predicting market returns is more likely to operate via a cash flow channel in the sense of Campbell (1991).

4.4. Relation of disagreement with trading volume and market volatility

In Atmaz and Basak (2018), in the absence of disagreement, trading volume is zero and market volatility is constant. In the presence of disagreement, however, higher disagreement leads to both higher trading volume and higher market volatility, thereby suggesting that disagreement is the driver of the positive volume-volatility relationship.

To test the implication, we estimate the volume-volatility elasticity in month t as the slope of regressing the daily change in market turnover on the daily change in volatility within month t, and then regress the monthly elasticity on the lagged disagreement index. For robustness, we consider four daily volatility measures, including realized volatility, realized semi-volatility, and median realized volatility based on the S&P 500 index returns from 5-minute intervals from Andersen, Dobrev, and Schaumburg (2012), and realized volatility of the S&P 500 index futures contract returns from 5-minute intervals from Johnson (2019).

Panel A of Table 11 shows that the disagreement index positively predicts the volume-volatility elasticity. The intuition is that increased disagreement increases the average bias of investor expectations, which in turn increases both the fluctuation of stock price and the trading demand (due to the increased weight of investors with relatively different beliefs), thereby increasing the volume-volatility elasticity. In a more intuitive way, we show in Table A3 in the Online Appendix that the disagreement index positively predicts the correlation between trading volume and market volatility. For example, a one-standard deviation increase in disagreement predicts a 5.22% increase in the volume-volatility correlation of next month when market volatility is estimated with the realized volatility.

To corroborate Panel A, we decompose market volatility into two components: one is contemporaneously related to disagreement and extracted via the PLS method and the other is unrelated to disagreement. Then we regress the one-month ahead trading volume on these two volatility components and report the results in Panel B of Table 11. As expected, the disagreement-related volatility significantly positively predicts future trading volume, whereas the disagreement-unrelated volatility does not have any predictive power. Similarly, when decomposing trading volume into disagreement-related and unrelated components, we find that the disagreement-related volume predicts future market volatility but the disagreement-unrelated volume fails to do so, which is reported in Panel B of Table A3.

In sum, this section provides empirical support to Atmaz and Basak (2018) that disagreement seems a key driver of the positive volume-volatility relationship.

5. Conclusion

This paper examines whether extant individual disagreement measures are agreeable and proposes a disagreement index by using the PLS methodology in Kelly and Pruitt (2013, 2015). We show that this PLS disagreement index significantly predicts market returns both in- and out-of-sample. Consistent with the theory in Atmaz and Basak (2018), the disagreement index asymmetrically predicts market returns with greater power in high sentiment periods, is negatively related to investors' ex post return forecast errors, predicts market returns through a cash flow channel, and is able to explain the positive volume-volatility relation.

There are some open issues for future research. First, it will be valuable to apply the disagreement index to other markets, such as bonds, commodities, and currencies, to see whether the forecasting power remains significant. Second, it will be of interest to construct aggregate disagreement indexes at different frequencies, such as daily or weekly, so that investors can use them for real time investing. Finally, as Hong and Stein (2007) posit that there are two main sources of disagreement—differences in information sets and differences in models that investors use to interpret information, it will be interesting to disentangle them.

References

Ajinkya, B. B., Atiase, R. K., Gift, M. J., 1991. Volume of trading and the dispersion in financial analysts' earnings forecasts. Accounting Review 66, 389–401.

Allen, L., Bali, T. G., Tang, Y., 2012. Does systemic risk in the financial sector predict future economic downturns? Review of Financial Studies 25, 3000–3036.

Andersen, T., Dobrev, D., Schaumburg, E., 2012. Jump-robust volatility estimation using nearest neighbor truncation. Journal of Econometrics 169, 75–93.

Anderson, E. W., Ghysels, E., Juergens, J. L., 2005. Do heterogeneous beliefs matter for asset pricing? Review of Financial Studies 18, 875–924.

Anderson, E. W., Ghysels, E., Juergens, J. L., 2009. The impact of risk and uncertainty on expected returns. Journal of Financial Economics 94, 233–263.

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

Atmaz, A., Basak, S., 2018. Belief dispersion in the stock market. Journal of Finance 73, 1225–1279.

Baker, M., Wurgler, J., 2000. The equity share in new issues and aggregate stock returns. Journal of Finance 55, 2219–2257.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. Journal of Finance 61, 1645–1680.

Baker, S. R., Bloom, N., Davis, S. J., 2016. Measuring economic policy uncertainty. Quarterly Journal of Economics 131, 1593–1636.

Bali, T. G., Brown, S., Tang, Y., 2020. Disagreement in economic forecasts and expected stock returns: Risk or mispricing? Unpublished working paper. Georgetown University.

Bali, T. G., Brown, S. J., Caglayan, M. O., 2014. Macroeconomic risk and hedge fund returns. Journal of Financial Economics 114, 1–19.

Barber, B. M., Odean, T., 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. Review of Financial Studies 21, 785–818.

Berkman, H., Dimitrov, V., Jain, P. C., Koch, P. D., Tice, S., 2009. Sell on the news: differences of opinion, short-sales constraints, and returns around earnings announcements. Journal of Financial Economics 92, 376–399.

Boehme, R. D., Danielsen, B. R., Sorescu, S. M., 2006. Short-sale constraints, differences of opinion, and overvaluation. Journal of Financial and Quantitative Analysis 41, 455–487.

Bordalo, P., Gennaioli, N., La Porta, R., Shleifer, A., 2019. Diagnostic expectations and stock returns. Journal of Finance 74, 2839–2874.

Bordalo, P., Gennaioli, N., Ma, Y., Shleifer, A., 2020. Over-reaction in macroeconomic expectations. American Economic Reivew, forthcoming.

Campbell, J. Y., 1991. A variance decomposition for stock returns. Economic Journal 101, 157–179.

Campbell, J. Y., Thompson, S. B., 2008. Predicting excess stock returns out of sample: Can anything beat the historical average? Review of Financial Studies 21, 1509–1531.

Carlin, B. I., Longstaff, F. A., Matoba, K., 2014. Disagreement and asset prices. Journal of Financial Economics 114, 226–238.

Chen, J., Hong, H., Stein, J. C., 2002. Breadth of ownership and stock returns. Journal of Financial Economics 66, 171–205.

Chen, Y., Eaton, G. W., Paye, B. S., 2018. Micro (structure) before macro? The predictive power of aggregate illiquidity for stock returns and economic activity. Journal of Financial Economics 130, 48–73.

Chinco, A., Clark-Joseph, A. D., Ye, M., 2019. Sparse signals in the cross-section of returns. Journal of Finance 74, 449–492.

Choi, H., Mueller, P., Vedolin, A., 2017. Bond variance risk premiums. Review of Finance 21, 987–1022.

Clark, T. E., West, K. D., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291–311.

Cookson, J. A., Niessner, M., 2020. Why don't we agree? Evidence from a social network of investors. Journal of Finance 75, 173–228.

Cooper, I., Priestley, R., 2009. Time-varying risk premiums and the output gap. Review of Financial Studies 22, 2801–2833.

Das, S., Kuhnen, C. M., Nagel, S., 2019. Socioeconomic status and macroeconomic expectations. Review of Financial Studies 33, 395–432.

DeVault, L., Sias, R. W., Starks, L. T., 2019. Sentiment metrics and investor demand. Journal of Finance 74, 985–1024.

Diebold, F. X., Shin, M., 2019. Machine learning for regularized survey forecast combination: Partially-egalitarian LASSO and its derivatives. International Journal of Forecasting 35, 1679–1691.

Diether, K. B., Malloy, C. J., Scherbina, A., 2002. Differences of opinion and the cross section of stock returns. Journal of Finance 57, 2113–2141.

Edelen, R. M., Ince, O. S., Kadlec, G. B., 2016. Institutional investors and stock return anomalies. Journal of Financial Economics 119, 472–488.

Engelberg, J., McLean, R. D., Pontiff, J., 2020. Analysts and anomalies. Journal of Accounting and Economics 69, 1–13.

Engsted, T., Pedersen, T. Q., Tanggaard, C., 2012. Pitfalls in VAR based return decompositions: A clarification. Journal of Banking & Finance 36, 1255–1265.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Ferreira, M. A., Santa-Clara, P., 2011. Forecasting stock market returns: The sum of the parts is more than the whole. Journal of Financial Economics 100, 514–537.

Freyberger, J., Neuhierl, A., Weber, M., 2020. Dissecting characteristics nonparametrically. Review of Financial Studies 33, 2326–2377.

Garfinkel, J. A., 2009. Measuring investors' opinion divergence. Journal of Accounting Research 47, 1317–1348.

Ge, L., Lin, T.-C., Pearson, N. D., 2016. Why does the option to stock volume ratio predict stock returns? Journal of Financial Economics 120, 601–622.

Giglio, S., Kelly, B., Pruitt, S., 2016. Systemic risk and the macroeconomy: An empirical evaluation. Journal of Financial Economics 119, 457–471.

Golez, B., Koudijs, P., 2018. Four centuries of return predictability. Journal of Financial Economics 127, 248–263.

Greenwood, R., Shleifer, A., 2014. Expectations of returns and expected returns. Review of Financial Studies 27, 714–746.

Gu, S., Kelly, B. T., Xiu, D., 2020. Empirical asset pricing via machine learning. Review of Financial Studies 33, 2223–2273.

Hamilton, J. D., 2018. Why you should never use the hodrick-prescott filter. Review of Economics and Statistics 100, 831–843.

Han, Y., He, A., Rapach, D., Zhou, G., 2019. Firm characteristics and expected stock returns. Unpublished working paper. Washington University in St. Louis.

Harvey, D. I., Leybourne, S. J., Newbold, P., 1998. Tests for forecast encompassing. Journal of Business & Economic Statistics 16, 254–259.

Hong, H., Sraer, D. A., 2016. Speculative betas. Journal of Finance 71, 2095–2144.

Hong, H., Stein, J. C., 2007. Disagreement and the stock market. Journal of Economic Perspectives 21, 109–128.

Huang, D., Jiang, F., Tu, J., Zhou, G., 2015. Investor sentiment aligned: A powerful predictor of stock returns. Review of Financial Studies 28, 791–837.

Huang, D., Kilic, M., 2019. Gold, platinum, and expected stock returns. Journal of Financial Economics 132, 50–75.

Jacobsen, B., Marshall, B. R., Visaltanachoti, N., 2019. Stock market predictability and industrial metal returns. Management Science 65, 2947–3448.

Jiang, H., Sun, Z., 2014. Dispersion in beliefs among active mutual funds and the cross-section of stock returns. Journal of Financial Economics 114, 341–365.

Johnson, T. L., 2019. A fresh look at return predictability using a more efficient estimator. Review of Asset Pricing Studies 9, 1–46.

Jondeau, E., Zhang, Q., Zhu, X., 2019. Average skewness matters! Journal of Financial Economics 134, 29–47.

Jurado, K., Ludvigson, S. C., Ng, S., 2015. Measuring uncertainty. American Economic Review 105, 1177–1216.

Kelly, B., Pruitt, S., 2013. Market expectations in the cross-section of present values. Journal of Finance 68, 1721–1756.

Kelly, B., Pruitt, S., 2015. The three-pass regression filter: A new approach to forecasting using many predictors. Journal of Econometrics 186, 294–316.

Kozak, S., Nagel, S., Santosh, S., 2020. Shrinking the cross section. Journal of Financial Economics 135, 271–292.

Lee, J. D., Sun, D. L., Sun, Y., Taylor, J. E., 2016. Exact post-selection inference, with application to the lasso. Annals of Statistics 44, 907–927.

Li, D., Li, G., 2017. Household investors' belief dispersion, trading volume,and excess returns. Unpublished working paper. Federal Reserve Board.

Li, F. W., 2016. Macro disagreement and the cross-section of stock returns. Review of Asset Pricing Studies 6, 1–45.

Li, J., Yu, J., 2012. Investor attention, psychological anchors, and stock return predictability. Journal of Financial Economics 104, 401–419.

Li, Y., Ng, D. T., Swaminathan, B., 2013. Predicting market returns using aggregate implied cost of capital. Journal of Financial Economics 110, 419–436.

Light, N., Maslov, D., Rytchkov, O., 2017. Aggregation of information about the cross section of stock returns: A latent variable approach. Review of Financial Studies 30, 1339–1381.

Lyle, M. R., Wang, C. C., 2015. The cross section of expected holding period returns and their dynamics: A present value approach. Journal of Financial Economics 116, 505–525.

Manela, A., Moreira, A., 2017. News implied volatility and disaster concerns. Journal of Financial Economics 123, 137–162.

Miller, E. M., 1977. Risk, uncertainty, and divergence of opinion. Journal of Finance 32, 1151–1168.

Møller, S. V., Rangvid, J., 2015. End-of-the-year economic growth and time-varying expected returns. Journal of Financial Economics 115, 136–154.

Pyun, S., 2019. Variance risk in aggregate stock returns and time-varying return predictability. Journal of Financial Economics 132, 150–174.

Rapach, D. E., Ringgenberg, M. C., Zhou, G., 2016. Short interest and aggregate stock returns. Journal of Financial Economics 121, 46–65.

Rapach, D. E., Strauss, J. K., Zhou, G., 2010. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. Review of Financial Studies 23, 821–862.

Rapach, D. E., Strauss, J. K., Zhou, G., 2013. International stock return predictability: What is the role of the United States? Journal of Finance 68, 1633–1662.

Stambaugh, R. F., Yu, J., Yuan, Y., 2012. The short of it: Investor sentiment and anomalies. Journal of Financial Economics 104, 288–302.

Tibshirani, R. J., Taylor, J., Lockhart, R., Tibshirani, R., 2016. Exact post-selection inference for sequential regression procedures. Journal of the American Statistical Association 111, 600–620.

Wang, H., Yan, J., Yu, J., 2017. Reference-dependent preferences and the riskreturn trade-off. Journal of Financial Economics 123, 395–414.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21, 1455–1508.

Wold, H., 1966. Estimation of principal components and related models by iterative least squares. Multivariate Analysis. New York: Academic Press.

Yu, J., 2011. Disagreement and return predictability of stock portfolios. Journal of Financial Economics 99, 162–183.

Yu, J., Yuan, Y., 2011. Investor sentiment and the mean–variance relation. Journal of Financial Economics 100, 367–381.

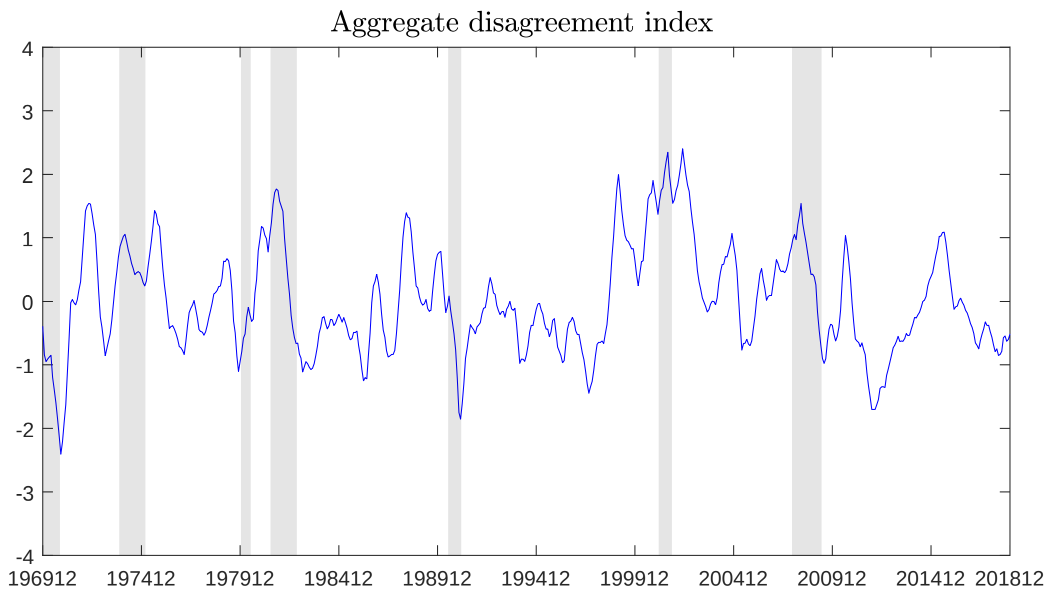


Fig. 1. This figure plots the time series dynamics of the disagreement index constructed by the PLS method in Kelly and Pruitt (2013, 2015). Grey shadow bars denote NBER recessions. The sample period is 1969:12–2018:12.

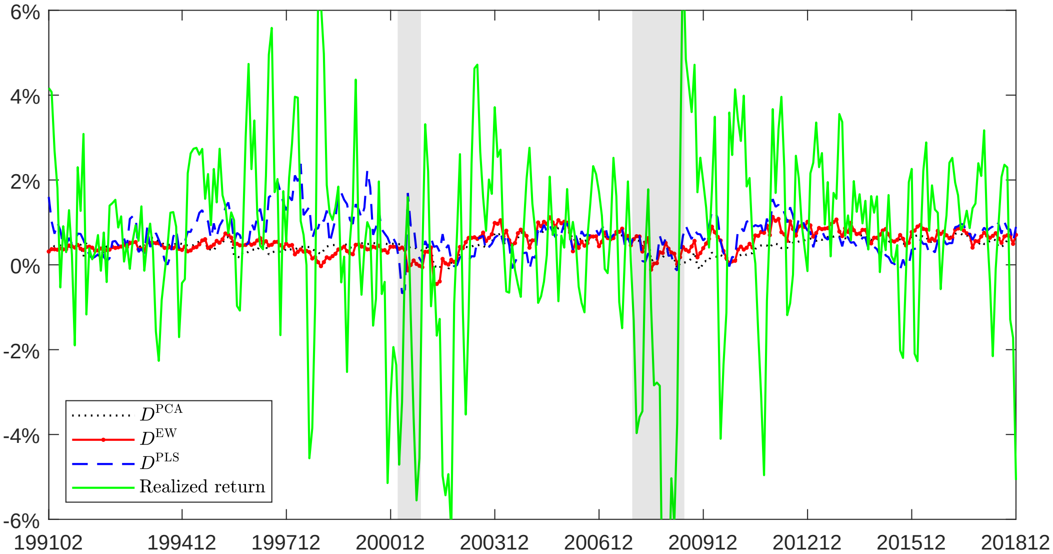


Fig. 2. This figure plots the out-of-sample 3-month market return forecasts with the PCA, equal-weight, and PLS disagreement indexes, respectively. For comparison, the figure also plots the realized 3-month market returns. Grey shadow bars denote NBER recessions. The out-of-sample period is 1991:02–2018:12.

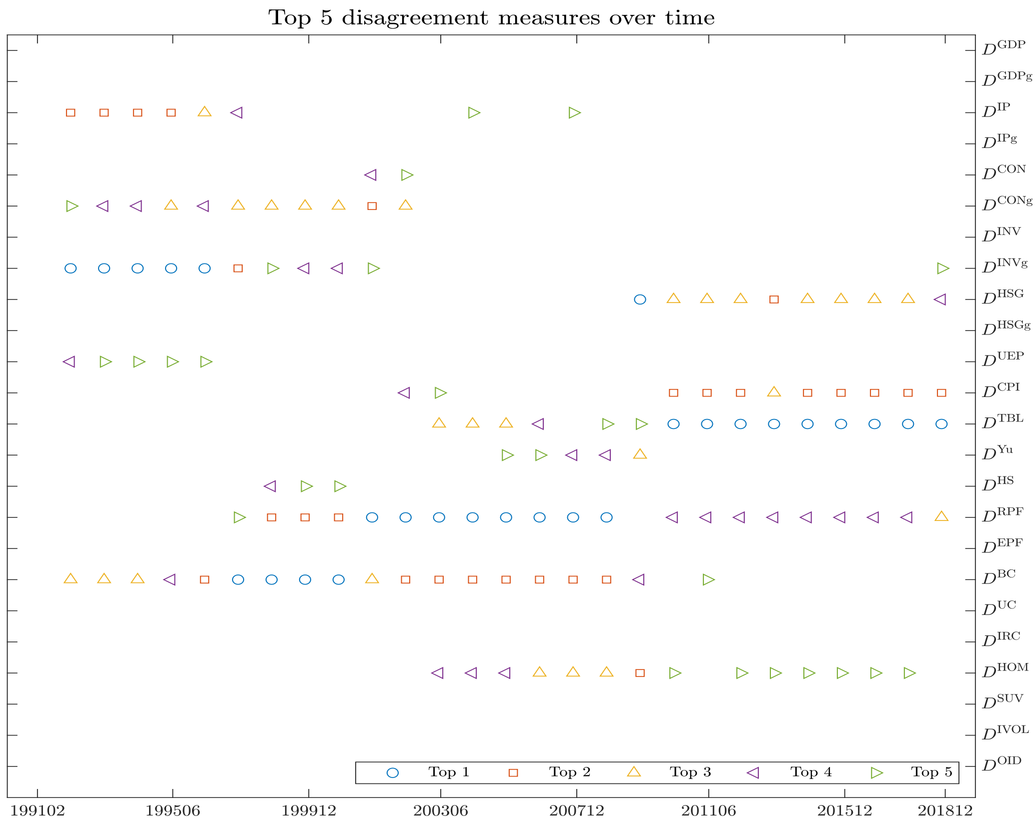


Fig. 3. This figure plots top five individual disagreement measures in the PLS disagreement index at each point in time when conducting out-of-sample forecasting.

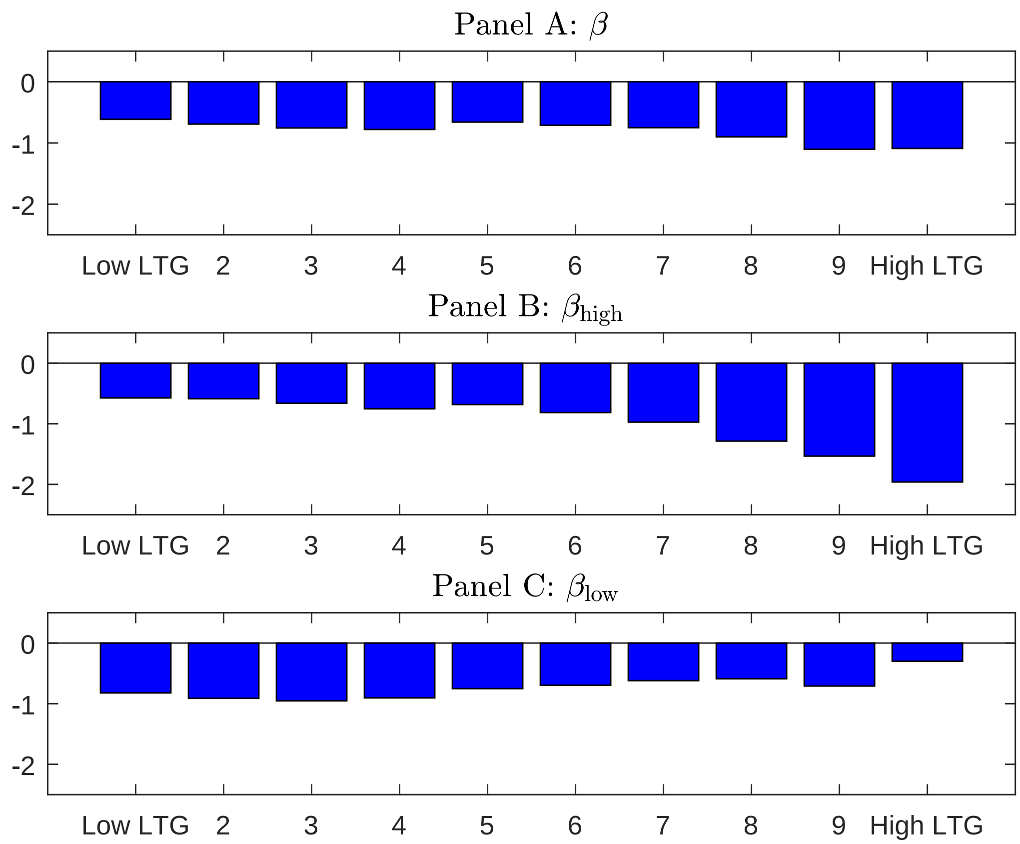


Fig. 4. This figure plots the regression slopes of predicting portfolio excess returns sorted by analyst long-term growth rate (LTG) forecast with the disagreement index as

$$R_{i,t+1} = \alpha_i + \beta_i D_t + \varepsilon_{i,t+1}$$

for Panel A, and

$$R_{i,t+1} = \alpha_i + \beta_{\text{high},i} S_t^{\text{high}} D_t + \beta_{\text{low},i} S_t^{\text{low}} D_t + \varepsilon_{i,t+1}$$

for Panels B and C. Low (high) LTG refers to the portfolio with low (high) analyst LTG forecast and is constructed the same as Bordalo et al. (2019). S_t^{high} (S_t^{low}) is a dummy variable that equals 1 if month t is in high (low) sentiment periods and 0 if month t is in low (high) sentiment periods (Baker and Wurgler, 2006). The sample period is 1982:01–2018:12.

Table 1 Summary statistics of individual disagreement measures

This table reports the summary statistics of 24 individual disagreement measures used in this paper. The first 13 measures are obtained from the survey of professional forecasters (SPF) at a quarterly frequency, each of which is defined by the level or growth difference between the 75th and 25th percentiles of the forecasts. D^{Yu} and D^{HS} are value- and beta-weighted analyst forecast dispersions (Yu, 2011; Hong and Sraer, 2016). The next six are household belief dispersions on macroeconomic conditions from the Michigan survey of consumers attitudes. D^{SUV} is a disagreement measure based on the standardized unexplained trading volume of NYSE stocks (Garfinkel, 2009). D^{IVOL} is the value-weighted idiosyncratic volatility proposed by Boehme, Danielsen, and Sorescu (2006) for measuring investor disagreement. D^{OID} is a disagreement measure defined by the open interest difference of OEX call and put options (Ge, Lin, and Pearson, 2016).

Disagreement measure	Sample Period	Obs	Avg	Std	Min	Max	Skew	Kurt
GDP forecast dispersion (D^{GDP})	1968Q4–2018Q4	201	61.32	41.13	6.80	248.60	1.28	2.53
GDP growth forecast dispersion (D^{GDPg})	1968Q4–2018Q4	201	1.65	0.71	0.71	4.25	1.07	0.81
Industrial production forecast dispersion (D^{IP})	1968Q4–2018Q4	201	1.97	1.03	0.52	6.10	1.14	1.30
Industrial production growth forecast dispersion (D^{IPg})	1968Q4–2018Q4	201	2.73	1.43	0.84	8.04	1.14	0.99
Consumption forecast dispersion (D^{CON})	1968Q4–2018Q4	150	31.61	18.31	5.00	101.87	1.08	1.91
Consumption growth forecast dispersion (D^{CONg})	1968Q4–2018Q4	150	0.97	0.40	0.39	2.79	1.33	2.07
Investment forecast dispersion (D^{INV})	1981Q3–2018Q4	150	22.46	12.47	3.40	57.92	0.48	−0.39
Investment growth forecast dispersion (D^{INVg})	1981Q3–2018Q4	150	3.63	1.19	1.43	8.62	0.71	1.33
Housing starts forecast dispersion (D^{HSG})	1968Q4–2018Q4	201	0.12	0.04	0.05	0.27	0.90	0.53
Housing starts growth forecast dispersion (D^{HSGg})	1968Q4–2018Q4	201	18.70	10.00	6.46	57.34	1.38	1.52
Unemployment rate forecast dispersion (D^{UEP})	1968Q4–2018Q4	201	0.32	0.13	0.15	1.04	1.75	5.10
CPI forecast dispersion (D^{CPI})	1981Q3–2018Q4	150	0.83	0.30	0.38	2.02	1.35	1.89
TBL forecast dispersion (D^{TBL})	1981Q3–2018Q4	150	0.46	0.36	0.04	2.96	3.36	17.15
Value-weighted analyst forecast dispersion (D^{Yu})	1981:12–2018:12	445	3.67	0.61	2.64	5.79	1.04	0.60
Beta-weighted analyst forecast dispersion (D^{HS})	1981:12–2018:12	445	5.15	1.29	3.41	9.62	1.39	1.87
Realized personal financial improvement dispersion (D^{RPF})	1978:01–2018:12	492	−0.44	0.02	−0.50	−0.39	−0.42	−0.37
Expected personal financial improvement forecast dispersion (D^{EPF})	1978:01–2018:12	492	−0.64	0.05	−0.80	−0.50	−0.22	0.24
Business condition forecast dispersion (D^{BC})	1978:01–2018:12	492	−0.42	0.07	−0.69	−0.28	−0.96	1.04
Unemployment condition forecast dispersion (D^{UC})	1978:01–2018:12	492	−0.63	0.08	−0.95	−0.47	−0.60	0.14
Interest rate condition forecast dispersion (D^{IRC})	1978:01–2018:12	492	−0.53	0.08	−0.77	−0.35	−0.17	−0.59
House purchase condition forecast dispersion (D^{HOM})	1978:01–2018:12	492	−0.59	0.08	−0.80	−0.42	−0.02	−0.64
Standardized unexplained volume (D^{SUV})	1968:12–2018:12	589	0.14	1.25	−3.45	3.17	−0.15	−0.68
Idiosyncratic volatility (D^{IVOL})	1968:12–2018:12	589	0.02	0.00	0.01	0.03	1.72	3.25
OEX call/put open interest difference (D^{OID})	1984:02–2018:12	419	0.86	0.09	0.55	1.00	−0.99	0.73

Table 2 Forecasting market returns with individual disagreement measures

This table presents the regression slope, Newey-West t-value, in-sample R^2, and out-of-sample R^2_{OS} of predicting market returns with individual disagreement measures:

$$R_{t,t+h} = \alpha + \beta D_t + \varepsilon_{t,t+h},$$

where $R_{t,t+h}$ is the cumulative market return between months t and $t+h$ ($h = 1,3$, or 12), and D_t is one of the 24 individual disagreement measures. The in-sample period is 1969:12–2018:12 and the out-of-sample period is 1991:02–2018:12 (because the accurate release dates of the Michigan survey of consumers attitudes are only available as of January 1991). Statistical significance for R^2_{OS} is based on the p-value of the Clark and West (2007) MSFE-adjusted statistic for testing $H_0 : R^2_{OS} \leq 0$ against $H_A : R^2_{OS} > 0$. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	Panel A: $h=1$				Panel B: $h=3$				Panel C: $h=12$			
Disagreement	β	t-value	R^2	R^2_{OS}	β	t-value	R^2	R^2_{OS}	β	t-value	R^2	R^2_{OS}
D^{GDP}	−0.11	−0.50	0.06	−1.88	−0.22	−1.35	0.72	−5.40	−0.36***	−3.54	7.43	−13.61
D^{GDPg}	−0.27	−1.53	0.38	−3.44	−0.24	−1.62	0.89	−7.18	−0.24**	−2.56	3.21	−16.47
D^{IP}	−0.06	−0.24	0.02	−2.38	−0.07	−0.42	0.08	−5.04	−0.04	−0.31	0.09	−12.18
D^{IPg}	0.04	0.20	0.01	−2.23	−0.21	−1.46	0.65	−9.50	−0.14	−1.09	1.13	−20.03
D^{CON}	−0.11	−0.49	0.07	−1.79	−0.24	−1.21	0.88	−4.34	−0.17	−1.27	1.72	−15.22
D^{CONg}	−0.10	−0.41	0.05	−2.38	−0.19	−0.92	0.59	−5.73	−0.17	−1.29	1.73	−19.59
D^{INV}	−0.24	−1.36	0.31	−2.73	−0.27*	−1.75	1.12	−8.11	−0.14	−0.88	1.15	−12.17
D^{INVg}	0.19	1.31	0.21	−1.60	0.02	0.17	0.01	−4.60	0.02	0.14	0.04	−8.25
D^{HSG}	−0.40**	−2.11	0.85	−5.39	−0.26*	−1.70	1.04	−12.29	−0.21	−1.57	2.48	−23.20
D^{HSGg}	−0.21	−0.99	0.23	−6.49	−0.34**	−2.08	1.70	−24.59	−0.33*	−1.90	5.73	−28.78
D^{UEP}	0.16	0.76	0.13	−0.73	0.13	0.82	0.27	−2.64	0.16	1.47	1.48	−3.38
D^{CPI}	−0.36*	−1.90	0.73	−6.39	−0.27**	−2.26	1.18	−27.39	−0.11	−1.32	0.74	−19.33
D^{TBL}	−0.60**	−2.23	1.94	−4.21	−0.48**	−2.12	3.60	−9.36	−0.25	−1.53	3.84	−13.52
D^{Yu}	−0.30	−1.23	0.35	−2.72	−0.30	−1.32	1.05	−4.07	−0.27	−1.25	3.06	−25.93
D^{HS}	−0.12	−0.42	0.06	−3.67	−0.16	−0.55	0.30	−7.92	−0.23	−0.85	2.52	−11.16
D^{RPF}	−0.22	−1.26	0.26	−3.35	−0.09	−0.55	0.11	−6.13	−0.16	−1.57	1.52	−17.10
D^{EPF}	−0.22	−1.03	0.26	−4.49	−0.16	−1.11	0.40	−9.37	−0.05	−0.49	0.17	−17.36
D^{BC}	−0.44**	−2.48	1.05	−5.77	−0.23	−1.56	0.86	−10.44	−0.08	−0.66	0.37	−19.99
D^{UC}	−0.06	−0.27	0.02	−3.40	−0.03	−0.19	0.02	−6.35	−0.05	−0.37	0.17	−19.27
D^{IRC}	−0.17	−0.72	0.16	−2.64	−0.41**	−2.12	2.66	−12.05	−0.44**	−2.52	11.72	−25.38
D^{HOM}	0.11	0.65	0.07	−0.86	0.01	0.04	0.00	−3.71	−0.19	−1.12	1.94	−20.53
D^{SUV}	−0.30	−1.24	0.23	−2.80	−0.34	−1.48	0.86	−6.86	−0.43*	−1.86	4.81	−18.40
D^{IVOL}	−0.03	−0.15	0.00	−5.46	−0.02	−0.07	0.00	−14.61	0.00	−0.02	0.00	−14.82
D^{OID}	−0.20	−0.73	0.10	−2.73	−0.15	−0.57	0.15	−6.09	−0.24	−0.93	1.47	−15.53

Table 3 Forecasting market returns with aggregate disagreement indexes

This table presents the regression slope, Newey-West t-value, in-sample R^2, and out-of-sample R^2_{OS} of predicting market returns with disagreement as

$$R_{t,t+h} = \alpha + \beta D_t + \varepsilon_{t,t+h},$$

where $R_{t,t+h}$ is the cumulative market return between months t and $t+h$ ($h = 1, 3$, and 12), and D_t is the PCA, equal-weight (individual measures), or PLS disagreement index. Statistical significance for R^2_{OS} is based on the p-value of the Clark and West (2007) MSFE-adjusted statistic for testing $H_0 : R^2_{OS} \leq 0$ against $H_A : R^2_{OS} > 0$. The in- and out-of-sample periods are 1969:12–2018:12 and 1991:02–2018:12, respectively. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Method	β	t-value	R^2	R^2_{OS}
Panel A: $h = 1$				
PCA	−0.38**	−1.96	0.61	0.20
Equal-weight	−0.60***	−2.87	1.46	0.90**
PLS	−0.83***	−3.96	2.52	1.56**
Panel B: $h = 3$				
PCA	−1.13**	−2.14	1.74	1.71***
Equal-weight	−1.73***	−2.99	3.88	3.74***
PLS	−2.24***	−3.82	5.98	7.68***
Panel C: $h = 12$				
PCA	−2.92**	−1.99	2.78	3.04***
Equal-weight	−4.93***	−3.06	7.49	9.41***
PLS	−7.04***	−4.16	13.88	13.26***

Table 4 Controlling for economic variables

Panel A presents the results of predicting market returns as

$$R_{t+1} = \alpha + \psi Z_t + \varepsilon_{t+1},$$

where Z_t is one of the 14 economic predictors in Welch and Goyal (2008), output gap in Cooper and Priestley (2009), or aggregate short interest in Rapach, Ringgenberg, and Zhou (2016). Panel B reports the results of forecasting market returns with the disagreement index and one economic predictor as

$$R_{t+1} = \alpha + \beta D_t + \psi Z_t + \varepsilon_{t+1}.$$

The last row reports the slope of the disagreement index from an elastic net regression by including all the economic predictors, where the t-value is calculated following Tibshirani et al. (2016) and Lee et al. (2016). The sample period is 1969:12–2018:12. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	Panel A: Univariate		Panel B: Bivariate		
Economic predictor	ψ	R^2	β	ψ	R^2
Dividend-price ratio (DP)	0.15	0.11	−0.82***	0.02	2.52
Dividend yield (DY)	0.17	0.15	−0.82***	0.03	2.52
Earning-price ratio (EP)	0.08	0.03	−0.83***	−0.01	2.52
Dividend payout ratio (DE)	0.08	0.03	−0.82***	0.03	2.53
Sample variance (SVAR)	−0.22	0.26	−0.81***	−0.07	2.54
Book-to-market ratio (BM)	0.00	0.00	−0.83***	−0.03	2.53
Net equity expansion (NTIS)	−0.06	0.02	−0.83***	−0.07	2.55
Treasury bill rate (TBL)	−0.26	0.36	−0.81***	−0.21	2.76
Long-term bond yield (LTY)	−0.15	0.11	−0.82***	−0.12	2.60
Long-term bond return (LTR)	0.42**	0.93	−0.82***	0.41**	3.41
Term spread (TMS)	−0.41**	0.89	−0.81***	−0.39**	3.31
Default yield spread (DFY)	−0.16	0.13	−0.84***	−0.21	2.74
Default return spread (DFR)	0.36	0.68	−0.80***	0.32	3.06
Inflation rate (INFL)	0.01	0.00	−0.83***	0.05	2.54
Output gap (OG)	−0.46***	1.08	−0.75***	−0.33**	3.07
Short interest (SI)	−0.55**	1.48	−0.84***	−0.46*	3.85
Kitchen sink (via elastic net)	–	–	−0.72***	–	5.50

Table 5 Controlling for uncertainty measures

Panel A presents the correlations between the disagreement index and eight uncertainty measures, including economic uncertainty (Bali, Brown, and Caglayan, 2014), treasury implied volatility (Choi, Mueller, and Vedolin, 2017), financial uncertainty and macro uncertainty (Jurado, Ludvigson, and Ng, 2015), economic policy uncertainty (Baker, Bloom, and Davis, 2016), news implied volatility (Manela and Moreira, 2017), sample variance (Welch and Goyal, 2008), and the Chicago Board Options Exchange (CBOE) volatility index (VIX). Panel B presents the results of predicting market returns with one uncertainty measure as

$$R_{t+1} = \alpha + \psi U_t + \varepsilon_{t+1}.$$

Panels C presents the results of predicting market returns with the disagreement index and one uncertainty measure as

$$R_{t+1} = \alpha + \beta D_t + \psi U_t + \varepsilon_{t+1}.$$

The last row reports the slope of the disagreement index from an elastic net regression by including all the uncertainty measures, where the t-value is calculated following Tibshirani et al. (2016) and Lee et al. (2016). ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	Panel A: Correlation	Panel B: Univariate		Panel C: Bivariate		
Uncertainty	Corr(Uncertainty$_t$, D_t)	ψ	R^2	β	ψ	R^2
Economic uncertainty	0.09	−0.13	0.09	−1.02***	−0.04	4.88
Treasury implied volatility	0.33***	−0.37	0.70	−1.01***	−0.06	4.72
Financial uncertainty	0.23***	−0.62**	2.01	−0.69***	−0.49*	3.68
Macro uncertainty	0.24***	−0.45	1.06	−0.74***	−0.30	2.96
Economic policy uncertainty	−0.20***	0.25	0.32	−0.86***	0.10	3.02
News implied volatility	0.07	0.09	0.04	−0.85***	0.14	2.74
Sample variance	0.23***	−0.22	0.26	−0.81***	−0.07	2.54
VIX	0.26***	0.00	0.00	−1.06***	0.24	4.66
Kitchen sink (via elastic net)	–	–	–	−0.89**	–	5.09

Table 6 Asset allocation results

This table reports portfolio gains of a mean-variance investor with risk-aversion $\gamma = 3$ or 5 for predicting market returns with the PCA, equal-weight (individual measure), and PLS disagreement indexes, respectively. The investor allocates her wealth monthly among the stock market and the risk-free asset by applying the out-of-sample forecasts based on one of the three disagreement indexes. CER gain is the annualized certainty equivalent return difference between applying a disagreement index forecast and applying the historical return mean forecast. Sharpe ratio is the monthly average portfolio excess return divided by its standard deviation. The portfolio weight is estimated recursively using data available at the forecast formation month t. The investment period is 1991:02–2018:12. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	No transaction cost		50 bps transaction costs	
	CER gain (%)	Sharpe ratio	CER gain (%)	Sharpe ratio
Panel A: Risk aversion $\gamma = 3$				
PCA	0.71	0.14**	0.52	0.13
Equal-weight	1.70**	0.16**	1.33*	0.15**
PLS	2.50***	0.18***	1.92**	0.16**
Panel B: Risk aversion $\gamma = 5$				
PCA	0.96**	0.12**	0.81**	0.11**
Equal-weight	2.10***	0.16***	1.69**	0.14**
PLS	2.68***	0.17***	1.88**	0.14**

Table 7 Out-of-sample R^2_{OS}s of forecasting market returns with alternative methods

This table presents the out-of-sample R^2_{OS}s of forecasting h-month ahead market returns with six alternative information aggregation methods: equal-weight LASSO, combination LASSO, encompassing LASSO, adaptive LASSO, egalitarian LASSO, and elastic net. Statistical significance for R^2_{OS} is based on the p-value of the Clark and West (2007) MSFE-adjusted statistic for testing $H_0 : R^2_{OS} \leq 0$ against $H_A : R^2_{OS} > 0$. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Method	$h = 1$	$h = 3$	$h = 12$
Equal-weight LASSO	1.26^{**}	6.09^{***}	12.08^{***}
Combination LASSO	1.08^{**}	2.87^{***}	9.67^{***}
Encompassing LASSO	1.09^{**}	2.92^{***}	9.41^{***}
Adaptive LASSO	0.71^{*}	2.34^{***}	7.47^{***}
Egalitarian LASSO	1.30^{*}	2.69^{***}	8.42^{***}
Elastic net	1.36^{**}	2.69^{***}	8.43^{***}

Table 8 Asymmetric forecasting power of disagreement

Panel A reports the in- and out-of-sample R^2s of predicting market returns with the disagreement index in different time periods, which are calculated as Eq. (19). Panel B presents the results of predicting market returns with a state-dependent regression:

$$R_{t+1} = \alpha + \beta_{\text{high}} S_t^{\text{high}} D_t + \beta_{\text{low}} S_t^{\text{low}} D_t + \varepsilon_{t+1},$$

where S_t^{high} (S_t^{low}) is a dummy variable that equals 1 if month t is in high (low) sentiment periods and 0 if month t is in low (high) sentiment periods (Baker and Wurgler, 2006). Statistical significance for R^2_{OS} is based on the p-value of the Clark and West (2007) MSFE-adjusted statistic for testing $H_0 : R^2_{OS} \leq 0$ against $H_A : R^2_{OS} > 0$. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Forecasting performance in different periods

In-sample R^2		Out-of-sample R^2_{OS}	
High sentiment	Low sentiment	High sentiment	Low sentiment
5.28	0.80	3.69**	−0.55

Panel B: State-dependent regression

	β_{high}	t-value	β_{low}	t-value	R^2
Sentiment-based state	−1.12***	−4.71	−0.42	−1.21	2.96

Table 9 Disagreement and expectations of market returns

Panel A reports the results of regressing expectations of market returns on the disagreement index:

$$\text{Expectation}_{t:t+12} = \alpha + \beta D_t + \varepsilon_t,$$

where $\text{Expectation}_{t:t+12}$ is investor expectation of 12-month ahead market return at time t, which is measured by (value-weighted) aggregate analysts' return forecast (Engelberg, McLean, and Pontiff, 2020), Michigan survey of consumers attitudes, Graham-Harvey survey of CFOs, or Robert Shiller's survey of individual investor confidence. Panel B reports the results of regressing return forecast errors on the PLS disagreement index:

$$\text{Realized return}_{t:t+12} - \text{Expectation}_{t:t+12} = \alpha + \beta D_t + \varepsilon_t,$$

where the $\text{Expectation}_{t:t+12}$ in the case of Shiller's survey is the projection of analysts' return forecast on Shiller's survey, so that the projection has the same measurement unit as the realized return. The sample periods all end in 2018:12, but start differently, from 1999:04 for analysts return forecast, 2002:06 for Michigan's survey, 2000:10 for Graham-Harvey's survey, and 2001:07 for Shiller's survey, respectively. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Expectations of market returns	$\text{Corr}(\text{Expectation}_{t:t+12}, D_t)$	β	t-value	R^2
Analysts' return forecast	0.35***	3.26***	2.61	12.45
Michigan survey	0.24***	2.71*	1.69	5.55
Graham-Harvey's survey	0.26**	1.57**	2.40	6.58
Shiller's survey	0.25***	2.16**	2.31	6.50

Panel B: Market return forecast errors	β	t-value	R^2
Analysts' return forecast	−7.34***	−2.73	23.26
Michigan survey	−8.42***	−3.36	21.17
Graham-Harvey's survey	−9.65***	−4.41	31.61
Shiller's survey	−9.18***	−4.45	30.34

Table 10 Forecasting market returns with cash flow news- and discount rate news-based disagreement indexes

This table reports the slopes and Newey-West t-values from the regression of

$$R_{t+1} = \alpha + \beta_{\text{CF}} D_t^{\text{CF}} + \beta_{\text{DR}} D_t^{\text{DR}} + \varepsilon_{t+1},$$

where D_t^{CF} (D_t^{DR}) is the PLS disagreement index that uses the contemporaneous cash flow news (discount rate news) as the regressor in Eq. (10). The cash flow news and discount rate news are estimated by using the Campbell (1991) VAR approach. In the leftmost column, "$\tilde{R}$" represents the total market return, economic variables are defined in Table 4, and "PC" represents the first three principal components extracted from all the economic variables (except for DP). ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

VAR variables	β_{CF}	t-value	β_{DR}	t-value	R^2
$\tilde{R}$, DP	-0.65^{***}	-2.92	-0.01	-0.08	2.15
$\tilde{R}$, DP, DY	-0.65^{***}	-3.08	-0.01	-0.03	2.16
$\tilde{R}$, DP, EP	-0.67^{***}	-2.73	-0.11	-0.51	2.06
$\tilde{R}$, DP, DE	-0.67^{***}	-2.73	-0.11	-0.51	2.06
$\tilde{R}$, DP, RVOL	-0.70^{**}	-2.52	-0.06	-0.23	2.23
$\tilde{R}$, DP, BM	-0.65^{***}	-3.29	0.11	0.61	2.16
$\tilde{R}$, DP, NTIS	-0.66^{***}	-3.02	-0.02	-0.11	2.21
$\tilde{R}$, DP, TBL	-0.55^{***}	-2.86	0.10	0.59	1.62
$\tilde{R}$, DP, LTY	-0.64^{***}	-3.23	0.09	0.53	2.13
$\tilde{R}$, DP, LTR	-0.65^{***}	-3.13	0.02	0.09	2.24
$\tilde{R}$, DP, TMS	-0.57^{***}	-2.87	0.07	0.39	1.81
$\tilde{R}$, DP, DFY	-0.65^{***}	-2.82	-0.02	-0.11	2.16
$\tilde{R}$, DP, DFR	-0.65^{***}	-2.70	-0.03	-0.13	2.10
$\tilde{R}$, DP, INFL	-0.65^{***}	-3.29	0.03	0.16	2.20
$\tilde{R}$, DP, OG	-0.57^{***}	-2.83	-0.06	-0.32	1.57
$\tilde{R}$, DP, SI	-0.47^{**}	-2.08	0.27	1.28	1.93
$\tilde{R}$, DP, PC	-0.66^{***}	-2.78	0.04	0.16	2.26

Table 11 Relation of disagreement with market volatility and trading volume

Panel A presents the results of predicting the volume-volatility elasticity with the disagreement index:

$$\text{Elasticity}_{t+1} = \alpha + \beta D_t + \varepsilon_{t+1},$$

where the elasticity in month $t+1$ is the slope of regressing the daily change in turnover of NYSE stocks on the daily change in volatility within month $t+1$. Realized volatility, realized semi-volatility, and median realized volatility are estimated based on the S&P 500 index returns from 5-minute intervals (Andersen, Dobrev, and Schaumburg, 2012), and futures realized volatility is estimated based on the S&P 500 index futures contract returns from 5-minute intervals (Johnson, 2019). Panel B presents the results of the following regression:

$$\text{Volume}_{t+1} = \alpha + \beta_1 \text{D\_Volatility}_t + \beta_2 \text{Volatility}_t^{\circ} + \varepsilon_{t+1}.$$

D_Volatility is the disagreement-related volatility and extracted with the PLS method, and Volatility$^{\circ}$ is the residual of regressing volatility on D_Volatility. Following Hamilton (2018), we apply AR(4) to both trading volume and market volatility to remove potential trends and expected information. Reported are regression coefficient, Newey-West t-value, and R^2. The sample period is 2000:01–2018:12 for the first three volatility measures and 1990:01–2015:12 for the last one. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Predicting volatility-volume elasticity			
Volatility measure	β	t-value	R^2
Realized volatility	4.01***	4.41	4.94
Realized semi-volatility	1.85**	2.08	1.83
Median realized volatility	1.99**	1.98	1.25
Futures realized volatility	2.24**	2.17	1.49

Panel B: Predicting trading volume					
Volatility measure	β_1	t-value	β_2	t-value	R^2
Realized volatility	2.38***	3.32	−1.30	−1.36	5.31
Realized semi-volatility	2.53***	3.44	−1.16	−1.32	5.58
Median realized volatility	2.60***	3.31	−1.44*	−1.77	6.36
Futures realized volatility	1.13**	2.15	−0.46	−0.75	1.18

Online Appendix

Are Disagreements Agreeable? Evidence from Information Aggregation

A. Six LASSO Methods

In this section, for each method we explain how to construct the out-of-sample forecast in month t for the return in month $t+1$.

Equal-weight LASSO In month t, we choose J out of K individual disagreement measures via the following LASSO optimization problem:

$$\max_{\beta} \sum_{j=1}^{t-1} \left(R_{t+1} - \sum_{k=1}^{K} \beta_k D_t^k\right)^2 + \lambda \sum_{k=1}^{K} |\beta_k|, \tag{A1}$$

where D_t^k is the observation of individual disagreement measure k ($k = 1, \cdots, K$) in month t. Then we construct an equal-weight disagreement index as

$$D_t^{\text{EW}} = \sum_{j=1}^{J} \tilde{D}_t^j, \tag{A2}$$

where $\tilde{D}_t^1$ through $\tilde{D}_t^J$ are the selected individual disagreement measures in month t. Based on the predictive regression (7), we estimate the expected market return as

$$\hat{R}_{t+1}^{\text{EW-LASSO}} = \hat{\alpha}_t + \hat{\beta}_t D_t^{\text{EW}}. \tag{A3}$$

Empirically, Chinco, Clark-Joseph, and Ye (2019) find that the LASSO performs well in identifying sparse and high-frequency return predictors in a cross-sectional framework.

Combination LASSO To reduce model instability and uncertainty, Han et al. (2019) propose a combination LASSO method to improve the forecasting power of individual stock return predictors, which directly combines individual stock return forecasts. In this paper, suppose $\hat{R}_{t+1}^k$ is the market return forecast based on disagreement measure D_t^k and M is the initial sample size for parameter training. In month t, the combination LASSO estimates the expected market return as

$$\hat{R}_{t+1}^{\text{C-LASSO}} = \sum_{k=1}^{K} \hat{\beta}_k \hat{R}_{t+1}^k, \tag{A4}$$

where $\hat{\beta} = (\hat{\beta}_1, \cdots, \hat{\beta}_K)$ is the estimate via the following LASSO optimization problem,

$$\max_{\beta} \sum_{j=M+1}^{t} \left(R_{t+1} - \sum_{k=1}^{K} \beta_k \hat{R}_{t+1}^k\right)^2 + \lambda \sum_{k=1}^{K} |\beta_k|. \tag{A5}$$

Encompassing LASSO Suppose $\hat{R}_{t+1}$ is the market return forecast based on all the individual disagreement measures via a multivariate predictive regression. Han et al. (2019) propose an encompassing LASSO method as

$$\hat{R}_{t+1}^{\text{E-LASSO}} = \theta_t \hat{R}_{t+1} + (1-\theta_t)\hat{R}_{t+1}^{\text{C-LASSO}}, \tag{A6}$$

where θ_t is estimated with the Harvey, Leybourne, and Newbold (1998) forecast encompassing test.

Adaptive LASSO As in Freyberger, Neuhierl, and Weber (2020), the adaptive LASSO weights the terms in the penalty of (A5) to encourage small first-round coefficient estimates to be set to zero,

$$\max_{\beta} \sum_{j=M+1}^{t} \big(R_{t+1} - \sum_{k=1}^{K} \beta_k \hat{R}_{t+1}^{k}\big)^2 + \lambda \sum_{k=1}^{K} w_k |\beta_k|. \tag{A7}$$

and estimate the expected market return as

$$\hat{R}_{t+1}^{\text{A-LASSO}} = \sum_{k=1}^{K} \hat{\beta}_k \hat{R}_{t+1}^{k}, \tag{A8}$$

where $w_i = 1/|\hat{\beta}_k|^{\nu}$, $\hat{\beta}_k$ is the univariate predictive regression estimate, and $\nu > 0$.

Egalitarian LASSO Instead of shrinking the coefficient to zero, Diebold and Shin (2019) propose to shrink it to the simple average,

$$\max_{\beta} \sum_{j=M+1}^{t} \big(R_{t+1} - \sum_{k=1}^{K} \beta_k \hat{R}_{t+1}^{k}\big)^2 + \lambda \sum_{k=1}^{K} |\beta_k - \frac{1}{K}|. \tag{A9}$$

Then the expected market return can be estimated as

$$\hat{R}_{t+1}^{\text{Eg-LASSO}} = \sum_{k=1}^{K} \hat{\beta}_k \hat{R}_{t+1}^{k}. \tag{A10}$$

Elastic net To handle the potential highly correlated return forecasts, one may solve for the following optimization problem,

$$\max_{\beta} \sum_{j=1}^{t} \big(R_t - \sum_{i=1}^{N} \beta_i \hat{R}_{i,t}\big)^2 + \lambda_1 \sum_{i=1}^{N} |\beta_i| + \lambda_2 \sum_{i=1}^{N} \beta_i^2, \tag{A11}$$

and estimate the expected market return as

$$\hat{R}_{t+1}^{\text{EN-LASSO}} = \sum_{k=1}^{K} \hat{\beta}_k \hat{R}_{t+1}^{k}. \tag{A12}$$

Empirically, Kozak, Nagel, and Santosh (2020) show that the elastic net is powerful in predicting stock returns in a cross-sectional framework.

In all the six LASSO-related methods, the tuning parameter λ is chosen via the corrected version of the Akaike information criterion (AICc). Han et al. (2019) show that the AICc performs quantitatively similar as alternative cross validation criteria.

B. Forecasting economic activities

This section shows that the disagreement index negatively predicts future economic activities. Specifically, we consider six macro variables as the proxy of economic activities, including the CFNAI (Allen, Bali, and Tang, 2012), industrial production growth, unemployment rate, aggregate equity issuance (Baker and Wurgler, 2000), total business inventory, and capacity utilization.

The macro variables are adjusted for seasonality and annualized for ease of exposition. To control for the autocorrelations, we follow Allen, Bali, and Tang (2012) and run the following regression:

$$y_{t+1} = \alpha + \beta D_t + \sum_{i=1}^{12} \lambda_i y_{t-i+1} + \varepsilon_{t+1}, \tag{A13}$$

where y_{t+1} is one of the macro variables.

Table A4 shows that the disagreement index negatively predicts future economic activities. For instance, a one-standard deviation increase in the disagreement index predicts a 0.93% decrease in the CFNAI and a 0.22% increase in unemployment, respectively.

Fig. A1. This figure plots the individual disagreement measures selected by the LASSO-related techniques at each point in time when conducting out-of-sample forecasting over 1991:02–2018:12.

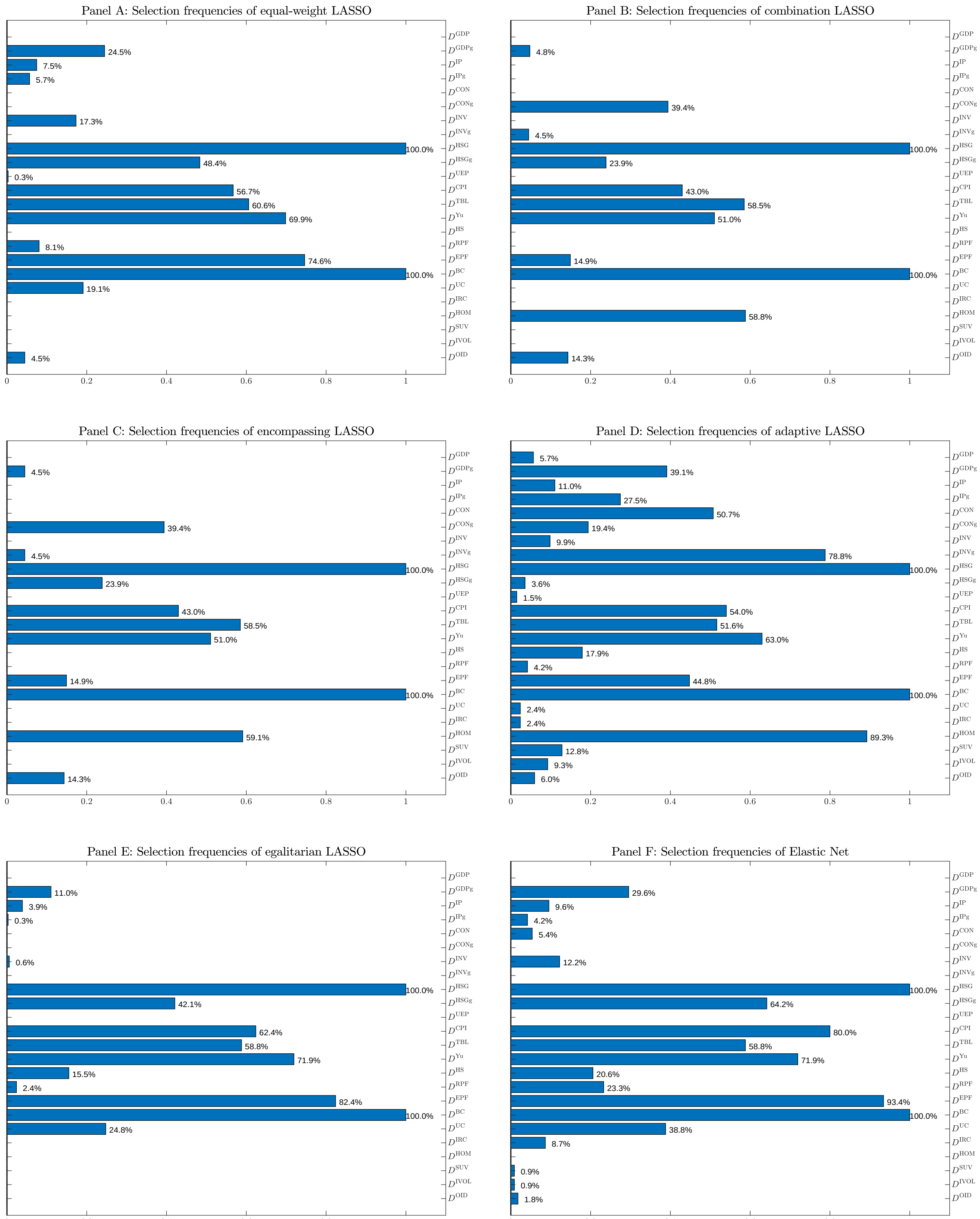


Fig. A2. This figure plots the selection frequency of each individual disagreement measure over the 1991:02–2018:12 out-of-sample period.

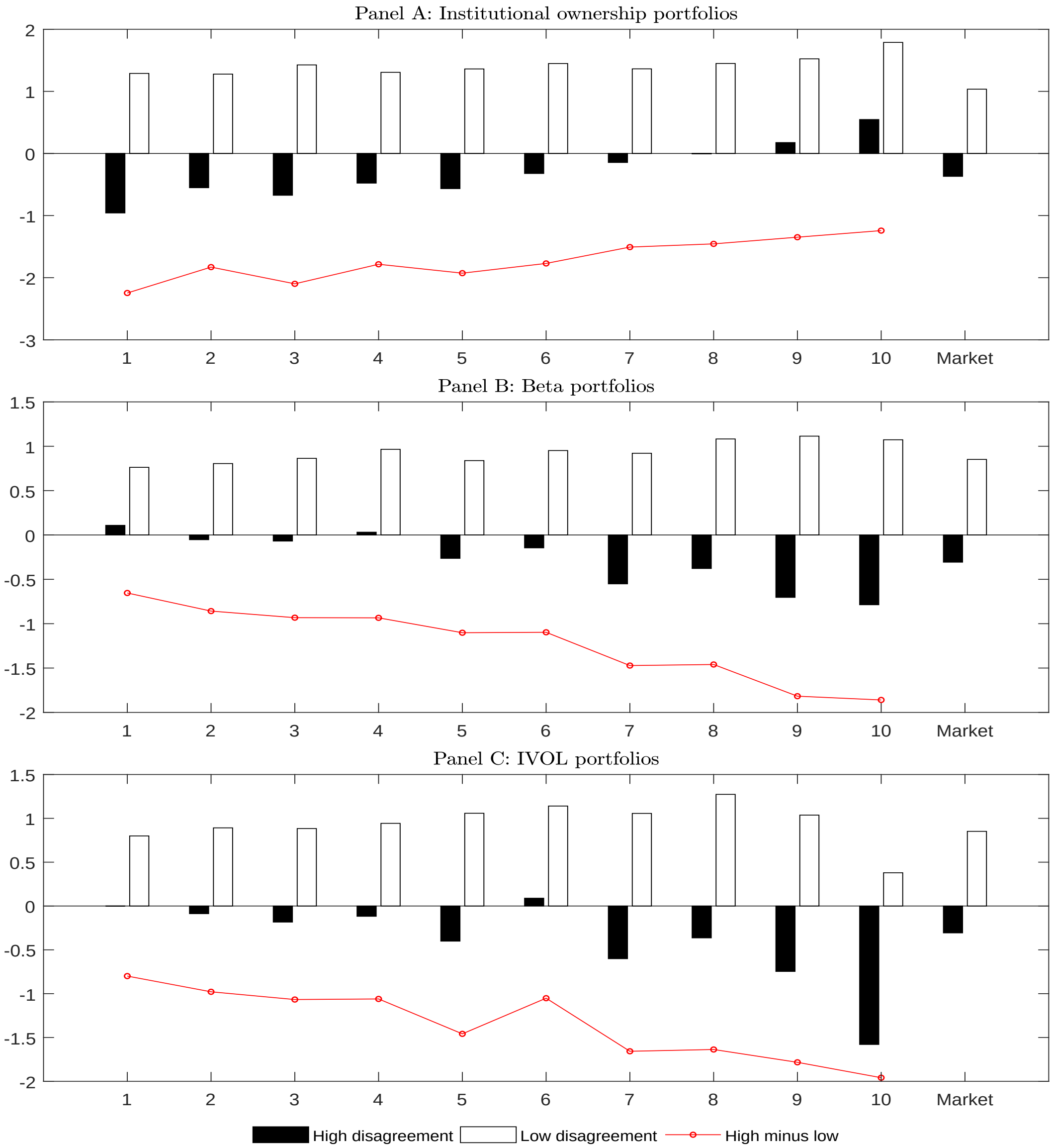


Fig. A3. This figure plots the average monthly excess returns of decile portfolios in high and low disagreement periods, where a month is in a high disagreement period if D^{PLS} in month $t-1$ is above its previous 24-month moving average, and otherwise in a low disagreement period.

Table A1 Correlations between individual disagreement measures

This table reports the pairwise correlations of 24 individual disagreement measures used in this paper. The first 13 measures are obtained from the survey of professional forecasters (SPF) at a quarterly frequency, each of which is defined by the level or growth difference between the 75th and 25th percentiles of the forecasts. D^{Yu} and D^{HS} are value- and beta-weighted analyst forecast dispersions (Yu, 2011; Hong and Sraer, 2016). The next six are household belief dispersions on macroeconomic conditions from the Michigan survey of consumers attitudes. D^{SUV} is a disagreement measure based on the standardized unexplained trading volume of NYSE stocks (Garfinkel, 2009). D^{IVOL} is the value-weighted idiosyncratic volatility proposed by Boehme, Danielsen, and Sorescu (2006) as a measure of investor disagreement. D^{OID} is a disagreement measure defined by the open interest difference of OEX call and put options (Ge, Lin, and Pearson, 2016).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1 D^{GDP}	1.00																							
2 D^{GDPg}	0.80	1.00																						
3 D^{IP}	0.54	0.56	1.00																					
4 D^{IPg}	0.56	0.67	0.83	1.00																				
5 D^{CON}	0.62	0.61	0.49	0.51	1.00																			
6 D^{CONg}	0.66	0.71	0.54	0.60	0.82	1.00																		
7 D^{INV}	0.45	0.42	0.38	0.34	0.49	0.42	1.00																	
8 D^{INVg}	0.42	0.52	0.50	0.58	0.32	0.49	0.68	1.00																
9 D^{HSG}	0.38	0.35	0.22	0.31	0.24	0.26	−0.06	0.04	1.00															
10 D^{HSGg}	0.50	0.54	0.32	0.40	0.28	0.37	0.16	0.35	0.51	1.00														
11 D^{UEP}	0.47	0.50	0.61	0.54	0.42	0.54	0.26	0.41	0.19	0.43	1.00													
12 D^{CPI}	0.43	0.39	0.32	0.45	0.26	0.36	0.14	0.26	0.43	0.44	0.31	1.00												
13 D^{TBL}	0.38	0.35	0.37	0.29	0.13	0.19	0.10	0.12	0.29	0.15	0.16	0.35	1.00											
14 D^{Yu}	0.26	0.39	0.35	0.26	0.26	0.33	0.32	0.33	0.16	0.21	0.22	0.15	0.08	1.00										
15 D^{HS}	0.23	0.25	0.42	0.30	0.20	0.18	0.21	0.27	0.08	0.01	0.16	0.07	0.06	0.64	1.00									
16 D^{RPF}	0.01	0.00	−0.03	0.07	−0.09	−0.05	−0.07	0.00	0.08	0.04	−0.14	0.05	0.16	−0.30	−0.13	1.00								
17 D^{EPF}	0.10	0.13	0.20	0.16	−0.01	0.07	−0.01	0.13	0.05	0.03	0.07	−0.03	0.19	−0.02	0.12	0.16	1.00							
18 D^{BC}	−0.22	−0.23	−0.28	−0.19	−0.22	−0.24	−0.08	−0.11	−0.28	−0.19	−0.41	−0.36	−0.14	−0.31	−0.25	0.23	−0.02	1.00						
19 D^{UC}	0.14	0.18	0.24	0.32	−0.10	0.09	−0.08	0.22	0.05	0.23	0.15	0.21	0.17	−0.13	0.06	0.27	0.45	0.01	1.00					
20 D^{IRC}	0.36	0.28	0.34	0.33	0.24	0.25	0.09	0.08	0.30	0.34	0.28	0.17	0.31	0.04	0.14	0.07	0.20	−0.14	0.26	1.00				
21 D^{HOM}	0.06	0.03	−0.04	−0.03	−0.05	0.07	−0.11	−0.10	0.13	−0.05	−0.04	0.25	0.34	0.17	0.09	−0.07	−0.03	−0.14	−0.07	−0.01	1.00			
22 D^{SUV}	0.00	0.00	−0.17	0.00	0.10	0.10	0.03	0.05	0.18	0.09	−0.11	0.14	−0.10	−0.19	−0.19	0.19	−0.12	0.19	−0.10	−0.09	0.06	1.00		
23 D^{IVOL}	0.32	0.34	0.36	0.29	0.37	0.31	0.50	0.41	0.20	0.19	0.27	0.00	−0.03	0.53	0.69	−0.23	0.03	−0.24	−0.11	0.15	−0.07	−0.09	1.00	
24 D^{OID}	0.22	0.19	0.13	0.17	0.15	0.17	0.24	0.31	0.11	0.21	0.17	0.00	−0.01	0.12	0.16	−0.05	0.13	−0.04	0.21	0.00	0.07	0.08	0.20	1.00

Table A2 Forecasting market returns with different moment PLS disagreement indexes

This table presents the regression slopes, Newey-West t-values, in-sample R^2s, and out-of-sample R^2_{OS}s of predicting market returns with the first to sixth moment PLS disagreement indexes, respectively. Statistical significance for R^2_{OS} is based on the p-value of the Clark and West (2007) MSFE-adjusted statistic for testing $H_0 : R^2_{OS} \leq 0$ against $H_A : R^2_{OS} > 0$. The in- and out-of-sample periods are 1969:12–2018:12 and 1991:02–2018:12, respectively. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Moment	β	t-value	R^2	R^2_{OS}
1st	−0.83***	−3.96	2.52	1.56**
2nd	−0.49	−1.06	0.21	−0.87
3rd	−0.20	−0.56	0.05	−0.29
4th	−0.01	−0.02	0.00	−0.24
5th	−0.10	−0.56	0.06	−0.16
6th	−0.12	−1.27	0.29	−0.08

Table A3 Relation of disagreement with market volatility and trading volume: Robustness check

Panel A presents the results of predicting the volume-volatility correlation with the disagreement index:

$$\text{Correlation}_{t+1} = \alpha + \beta D_t + \varepsilon_{t+1},$$

where the correlation in month $t+1$ refers to the correlation between the daily change in turnover of NYSE stocks and the daily change in volatility within month $t+1$. Realized volatility, realized semi-volatility, and median realized volatility are estimated based on the S&P 500 index returns from 5-minute intervals (Andersen, Dobrev, and Schaumburg, 2012), and futures realized volatility is estimated based on the S&P 500 index futures contract returns from 5-minute intervals (Johnson, 2019). Panel B presents the results of the following regression:

$$\text{Volatility}_{t+1} = \alpha + \beta_1 \text{D\_Volume}_t + \beta_2 \text{Volume}^\circ_t + \varepsilon_{t+1}.$$

D_Volume is the disagreement-related volume and extracted with the PLS method, and Volume$^\circ$ is the residual of regressing volume on D_Volume. D_Volatility is the disagreement-related volatility. Following Hamilton (2018), we apply AR(4) to both trading volume and market volatility to remove potential trends and expected information. Reported are regression coefficient, Newey-West t-value, and R^2. The sample period is 2000:01–2018:12 for the first three volatility measures and 1990:01–2015:12 for the last one. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Predicting volatility-volume correlation			
	β	t-value	R^2
Realized volatility	5.22***	3.36	4.02
Realized semi-volatility	3.25	1.51	1.27
Median realized volatility	3.21**	2.05	1.57
Futures realized volatility	5.30***	3.97	4.68

Panel B: Predicting market volatility						
	β_1	t-value	β_2	t-value	R^2	Corr(D_Volume, D_Volatility)
Realized volatility	3.17*	1.80	0.74	0.60	1.45	0.45
Realized semi-volatility	3.29*	1.79	0.71	0.50	1.32	0.44
Median realized volatility	4.19**	2.33	1.15	0.77	2.20	0.59
Futures realized volatility	5.84***	4.90	1.22	1.13	5.43	0.62

Table A4 Forecasting economic activities with disagreement

The table presents the regression slope, Newey-West t-value, and R^2 of predicting economic activities with the disagreement index as

$$y_{t+1} = \alpha + \beta D_t + \sum_{i=1}^{12} \lambda_i y_{t-i+1} + \varepsilon_{t+1}.$$

Economic activities include Chicago Fed National Activity Index (CFNAI), industrial production growth, unemployment, aggregate equity issuance (Baker and Wurgler, 2000), business inventory, and capacity utilization. ***, **, * indicate significance at the 1%, 5%, and 10% levels, respectively.

Economic activity	β	t-value	R^2
CFNAI	-0.93^{**}	-2.05	27.13
Industrial production	-1.04^{***}	-2.68	20.92
Unemployment	0.22^{**}	2.22	18.06
Equity issuance	-4.73^{**}	-2.46	29.35
Business inventory	-0.57^{***}	-3.49	58.52
Capacity utilization	-0.72^{***}	-2.30	20.00

Time series momentum: Is it there?[☆]

Dashan Huang[a,*], Jiangyuan Li[a], Liyao Wang[b], Guofu Zhou[c,d]

[a] *Singapore Management University, Lee Kong Chian School of Business, 50 Stamford Road, Singapore 178899*

[b] *Singapore Management University, School of Economics, 90 Stamford Road, Singapore 178903*

[c] *Washington University in St. Louis, Olin School of Business, 1 Brookings Drive, St. Louis, MO 63130, USA*

[d] *China Academy of Financial Research (CAFR), 211 West Huaihai Road, Shanghai 200030, China*

Abstract

Time series momentum (TSM) refers to the predictability of the past 12-month return on the next one-month return and is the focus of several recent influential studies. This paper shows that asset-by-asset time series regressions reveal little evidence of TSM, both in- and out-of-sample. While the t-statistic in a pooled regression appears large, it is not statistically reliable as it is less than the critical values of parametric and nonparametric bootstraps. From an investment perspective, the TSM strategy is profitable, but its performance is virtually the same as that of a similar strategy that is based on historical sample mean and does not require predictability. Overall, the evidence on TSM is weak, particularly for the large cross section of assets.

Keywords: Time series momentum, Risk premium, Return predictability, Pooled regression

JEL classification: G12, G14, G15, F37

[☆] We are grateful to William Schwert (the editor) and an anonymous referee for very insightful and helpful comments that significantly improved the paper. We also thank Lauren H. Cohen, Eugene Fama, Amit Goyal, Wesley R. Gray, Campbell R. Harvey, Johan Hombert, Narasimhan Jegadeesh, Raymond Kan, Yan Liu, Roger Loh, David Rapach, Yiuman Tse, John Wald, Dacheng Xiu, and seminar and conference participants at Baruch College, Rutgers University, Washington University in St. Louis, 2017 Conference on Financial Predictability and Data Science, 2017 Workshop on Advanced Econometrics at the University of Kansas, and 2018 National Bureau of Economic Research-National Science Foundation (NBER-NSF) Time Series Conference for valuable comments. Dashan Huang acknowledges that this study was partially funded at the Singapore Management University through a research grant (C207MSS17B009) from the Ministry of Education Academic Research Fund Tier 1.

[*] Corresponding author.

Email addresses: dashanhuang@smu.edu.sg (Dashan Huang), jiangyuanli.2015@pbs.smu.edu.sg (Jiangyuan Li), liyao.wang.2015@phdecons.smu.edu.sg (Liyao Wang), zhou@wustl.edu (Guofu Zhou)

1. Introduction

Whether past returns predict future returns is a central topic in finance. Fama (1965), French and Roll (1986), Lo and MacKinlay (1988), and Conrad and Kaul (1988), among others, show that past returns can positively predict future returns at a short horizon, but the magnitude is too small to be exploitable. Moskowitz, Ooi and Pedersen (MOP, 2012) conclude with a much greater degree of predictability that time series momentum (TSM) is everywhere: The past 12-month return positively predicts the next one- to 12-month return for a comprehensive set of approximately 55 assets. In addition, MOP show that a TSM trading strategy, which buys assets if their past 12-month returns are positive and sells them otherwise, earns significant average and risk-adjusted returns. Hurst, Ooi and Pedersen (2017) and Georgopoulou and Wang (2017) examine the TSM on a broader range of asset classes and longer sample periods, and they find similar results. Koijen, Moskowitz, Pedersen and Vrugt (2018) use TSM portfolio returns as a risk factor to analyze carry trade. Kim, Tse and Wald (2016) find that the profits of the TSM strategy are driven by volatility scaling and that its performance without volatility scaling is no better than that of a buy-and-hold strategy. Moreover, in a concurrent study, Goyal and Jegadeesh (2018) show that the traditional cross-sectional momentum strategy is more profitable than the TSM once the leverage ratio is properly adjusted. Despite an improved understanding, whether time series predictability is present at the 12-month frequency remains an open question.[1]

In this paper, using the same data set as MOP (2012), we reexamine the statistical and economic evidence of TSM. From both time series and cross-sectional analyses, we find that the evidence on TSM is weak. Hence, concluding that TSM exists across the global asset classes appears questionable.

We conduct our study in three stages. In the first stage, we run a time series regression of monthly return for each asset on its past 12-month return. At the 10% significance level, 47 of the 55 assets have a t-statistic of less than 1.65, suggesting that the in-sample evidence of TSM

[1] In this paper, the terms "TSM" and "time series predictability" are used interchangeably.

is weak.[2] Since Welch and Goyal (2008), studies on return predictability have shifted the focus to out-of-sample performance. We compute the standard Campbell and Thompson (2008) out-of-sample R^2_{OS} for each asset and find that only three assets deliver significant R^2_{OS} at the 10% level. Univariate time series regressions thus indicate that the evidence on time series predictability is weak among the assets.

In the second stage, we follow MOP's approach and run a pooled regression by stacking all asset returns together. Consistent with MOP, we find a t-statistic of 4.34 in the regression of predicting the next one-month return using the past 12-month return. At the conventional critical level of 2, one could interpret this t-statistic of 4.34 as strong evidence against no predictability. We argue that the pooled regression is likely to over-reject the null hypothesis for three reasons. First, if assets have different mean returns, the slope estimate from the pooled regression without controlling for fixed effects tends to be biased upward (Hjalmarsson, 2010). Second, as a predictor, the past 12-month return is persistent and can generate substantial size distortions (Hodrick, 1992; Stambaugh, 1999; Campbell and Yogo, 2006; Ang and Bekaert, 2007; Boudoukh, Richardson and Whitelaw, 2008; Li and Yu, 2012). Third, because volatility varies dramatically across assets, volatility scaling in the pooled regression without controlling for fixed effects further exacerbates the upward bias.

To assess the degree of over-rejection, we use two bootstrap methods. The first is a parametric wild bootstrap that simulates samples based on the fitted pooled regression residuals, and the second is a nonparametric pairs bootstrap that resamples the predictor and the dependent variable simultaneously. Both methods accommodate conditional heteroskedasticity, but the latter allows for more general data-generating processes. We find that the 5% critical values of the bootstraps are 12.53 and 4.83, respectively. They are larger than 4.34, the t-statistic from the pool regression with real data. This finding is robust to all alternative cases, such as within each asset class, with different sample periods, and without volatility scaling. Hence, a high t-statistic found by MOP is

[2]The multiple test framework of Harvey, Liu and Zhu (2016) could suggest even weaker evidence on TSM.

not statistically significant in supporting the existence of TSM.

In the third stage, we examine why the TSM strategy is profitable even though the statistical evidence on time series predictability is weak. In their study, MOP (2012) construct the TSM strategy by buying assets with positive past 12-month return and selling assets with negative past 12-month return. At the same time, MOP assign a portfolio weight equal to 40% divided by the asset volatility, so that for each asset the ex ante annualized volatility is 40%. This volatility scaling can make the attribution of the performance of the TSM strategy complex. To separate the volatility effect, we follow Kim, Tse and Wald (2016) and Goyal and Jegadeesh (2018) and focus on the TSM strategy with the simple equal weighting without volatility scaling as the benchmark. Volatility scaling is not an issue if all strategies are based on volatility-scaled returns when comparing their performances.

We examine the performance of the TSM strategy in four ways. First, we investigate the performance of an alternative trading strategy that does not require predictability. Based on the observation that a high mean asset is more likely to have a positive past 12-month return and therefore is more likely to be bought by the TSM strategy, we propose a times series history (TSH) strategy that buys assets if their historical mean returns are positive and sells them otherwise, which is theoretically profitable even if asset returns are independent over time, but some have significantly higher means than others. We find that the TSM and TSH strategies perform virtually the same and their differences in average returns, as well as in risk-adjusted ones, are indifferent from zero. Also, we find that the performances of the TSM and TSH strategies mainly stem from their long legs, and their short legs have insignificant average and risk-adjusted returns. This result suggests that both the TSM and TSH tend to long assets with greater mean returns and short those with lower means. Because the TSH strategy is defined without requiring time series predictability, it seems questionable to attribute the performance of the TSM strategy to predictability.

Second, we report the results with alternative portfolio weighting schemes, such as volatility weighting as in MOP (2012), past 12-month return weighting, and equal weighting with a zero-investment constraint as in Goyal and Jegadeesh (2018). The results are similar, and the alpha differential between the TSM and TSH strategies is always indifferent from zero. In short, the profitable performance of the TSM strategy is similar to that of the TSH strategy that requires no predictability, suggesting that the performance of the TSM does not necessarily support predictability at the 12-month frequency across asset classes.

Third, based on the predictive slope of Lewellen (2015), we examine the overall predictability of TSM across assets. The slope measures how realized returns are explained by predicted returns. If the past 12-month return perfectly predicts the next one-month return, the slope should have a value of one. We find that for the TSM forecasts the slope has a value close to zero, suggesting that the TSM forecasts have little predictive power.[3] When regressing the TSM forecasts on the TSH forecasts, the slope is very close to one, irrespective of whether the TSM forecasts use volatility scaling or not. This result indicates little difference in predictability between the two forecasts, suggesting, again, no evidence of TSM across the assets.

Fourth, of interest is examining under what conditions the TSM is a better trading strategy than the TSH. Based on one thousand simulated samples by using pooled regression with varying assumed degrees of time series momentum (i.e., the slope varies from 0.1 to 0.4), we find that the TSM and TSH strategies perform similarly when the slope is 0.1 (with real data, the slope of the pooled regression is 0.08, controlling for fixed effects). When the slope is 0.2, the TSM outperforms the TSH, but the difference is statistically insignificant. This means that if there is genuine time series predictability, the advantage of the TSM strategy is not apparent as long as the slope is small. When the slope is 0.4, the TSM dominates the TSH in the sense that it does better in almost all the simulated samples. Because the two strategies generate similar performance using real data, our simulation indicates that the evidence of TSM is likely weak if it exists. Combined with other results, the TSM is unlikely to be statistically significant for all the assets. In short, a lack of empirical evidence exists to support the hypothesis that the TSM is everywhere.

[3] Lewellen (2015) shows in his footnote 3 that a slope of less than 0.5 under-performs a naive forecast. Han, He, Rapach and Zhou (2019) discuss more properties of the slope.

As a final remark, our results do not claim in any way that there is no predictability in the asset classes, but that the predictability, if it exists, is not as simple as a constant 12-month return rule. The best example could be the stock market, with ample evidence that its risk premium can be predicted by a wide range of predictors such as macroeconomic variables and investor sentiment [see, e.g., Jiang, Lee, Martin and Zhou (2019) for the latest literature]. Cross-sectionally, since Jegadeesh and Titman's (1993) discovery of momentum, there have been hundreds of potential anomalies [see, e.g., Hou, Xue and Zhang (2019) for the replication]. Recently, Gu, Kelly and Xiu (2018) and Freyberger, Neuhierl and Weber (2019), among others, use machine learning tools to find even stronger predictability.[4] However, none of them is related to the TSM.

The rest of this paper is organized as follows. Section 2 introduces data we use in this paper. Section 3 shows that asset-by-asset regressions suggest that the evidence of TSM, if any, is weak. Section 4 finds that the pooled regression overstates the presence of TSM and that bootstrap-corrected t-statistics cannot reject the null hypothesis of no predictability. Section 5 shows that the TSM strategy performs the same as an alternative trading strategy that does not require predictability. Section 6 concludes.

2. Data

We collect futures prices for 24 commodities, nine developed country equity indexes, 13 developed government bonds, and nine currency forwards from the same data sources as MOP (2012). These 55 instruments are the same as those in MOP's Table 1.[5] The sample period is from January 1985 to December 2015. For each day, we calculate the daily excess return of each futures contract with the nearest- or next-nearest-to-delivery contract and compound the daily returns to a cumulative month return index. For brevity, returns in this paper always refer to excess returns,

[4] Rapach, Strauss and Zhou (2013) use LASSO, a major machine learning tool, to forecast international equity markets, which is the earliest study that we find in the finance literature applying LASSO to predict stock returns.

[5] MOP use 12 cross-currency pairs in trading strategies but report only nine underlying currencies in their summary statistics table. To maximally replicate the results, we focus on the nine underlying currencies. As a consequence, we examine a total of 55 assets, not 58.

unless otherwise stated.

Table 1 reports the sample mean (arithmetic), volatility (standard deviation), and first-order autocorrelation of the returns on the 55 futures contracts. The mean and volatility are annualized and represented in percentage. Significant variations exist in mean return and volatility across different contracts. Within the commodity asset class, 24 contracts yield positive, zero, and negative mean returns, from -11.33% for natural gas to 12.31% for unleaded gasoline. The volatility ranges from 13.56% for cattle to 50.79% for natural gas. On average, the mean return and volatility are 2.59% and 28.50%, respectively. The nine equity index futures contracts are more homogenous, with mean return from 3.09% for TOPIX (Tokyo Stock Price Index) to 9.69% for DAX (German Stock Index) and volatility from 15.87% for FTSE 100 (Financial Times Stock Exchange 100 Index) to 23.21% for FTSE/MIB (Italian National Stock Exchange Index). On average, the return and volatility are 7.24% and 19.26%, respectively. Finally, bond futures and currency forwards earn lower mean returns with lower volatilities. Within each asset class, the average mean and volatility are 4.54% and 7.85% for bond futures and 1.22% and 10.90% for currency forwards, respectively. Table 1 also highlights a well-known fact that the past one-month return cannot predict the next one-month return, because the first-order correlation is generally close to zero.

3. Univariate time series regression

In this section, we run univariate time series regressions to explore the predictability of the past 12-month return for individual assets. These regressions clearly tell which asset return can be predicted by its past 12-mopnth return and which cannot, thereby providing direct evidence on whether the finding in MOP (2012) is common across asset classes.

3.1. In-sample performance

Time series regressions are standard for identifying return predictability, but the pooled regression is seldom used in the literature. Ang and Bekaert (2007) and Hjalmarsson (2010) are exceptions, showing a heterogeneous pattern in predictability. For example, Ang and Bekaert (2007, p.663) show that, in contrast to the US, the UK and Japan return predictability disappears when expected returns are constrained to be non-negative and conclude that "none of the [return predictability] patterns in other countries resembles the US pattern."

For each asset, we run the predictive regression

$$r^i_{t+1} = \alpha + \beta r^i_{t-12,t} + \varepsilon^i_{t+1}, \tag{1}$$

where r^i_{t+1} is the return of asset i in month $t+1$ and $r^i_{t-12,t}$ is its past 12-month return (i.e., the return between months $t-12$ and t). The predictive power is based on either the regression slope β or the R^2 statistic. If the regression R^2 statistic is significantly larger than zero with a positive β, then asset i displays TSM, i.e., its past 12-month return predicts the next one-month return.

Table 2 reports the regression slope, the Newey-West t-statistic, and R^2. We have four observations. First, the presence of TSM is not prevalent. Of the 55 assets, only eight display significant regression slopes at the 10% level, representing 15% of assets (only three significant at the 5% level). Second, the significance is not concentrated but disperse among the four asset classes, including three commodities, two equity indexes, two government bonds, and one currency. Third, although not significant, 17 assets deliver negative slopes, amounting to 31% of the assets. Fourth, the R^2s are small, with an average of 0.39%, and only five assets generate an R^2 larger than 1%. To have an intuitive understanding of the predictive performance, Panel A of Fig. 1 plots the R^2 statistic for each asset. Only two assets have R^2s that stand out above 2%.

3.2. *Out-of-sample performance*

Due to concerns of data mining and structural breaks, studies on return predictability have shifted the focus to out-of-sample performance since Welch and Goyal (2008). To investigate the out-of-sample performance of TSM, we use the Campbell and Thompson (2008) out-of-sample

R^2_{OS} statistic as the assessment criterion, which is defined as

$$R^2_{OS} = 1 - \frac{\sum_{t=K}^{T-1}(r^i_{t+1} - \hat{r}^i_{t+1})^2}{\sum_{t=K}^{T-1}(r^i_{t+1} - \bar{r}^i_{t+1})^2}, \tag{2}$$

where K is the initial sample size for parameters training, $\hat{r}^i_{t+1}$ is the expected return estimated with information up to month t and calculated as $\hat{r}^i_{t+1} = \hat{\alpha}_t + \hat{\beta}_t r^i_{t-12,t}$, $\hat{\alpha}_t$ and $\hat{\beta}_t$ are the coefficients of the time series regression Eq. (1), and $\bar{r}^i_{t+1}$ is the sample mean of asset i with data up to month t. The choice of K is ad hoc in the literature, which depends on the nature of the possible model instability and the timing of the possible breaks. Hansen and Timmermann (2012) theoretically show that a large K is preferable if the data-generating process is stationary, but it comes at the cost of low power as there are fewer observations for out-of-sample evaluation. A small K can give the out-of-sample test more desirable size properties, but it perhaps does not provide precise estimation. For these reasons, we select the first 15 years of data for in-sample training and the remaining 16 years of data for out-of-sample evaluation. That is, the full sample period is from January 1985 to December 2015 and the out-of-sample period is from January 2000 to December 2015.

Welch and Goyal (2008) show that the sample mean is a very stringent out-of-sample benchmark. If $R^2_{OS} > 0$, the forecast $\hat{r}^i_{t+1}$ outperforms the sample mean in terms of mean squared forecast error (MSFE). Empirically, they show that the in-sample forecasting abilities of a variety of return predictors generally do not hold in out-of-sample tests. To ascertain whether a forecast delivers a statistically significant improvement in MSFE relative to the sample mean, we use the Clark and West (2007) statistic to test the null hypothesis that the MSFE of the sample mean forecast is less than or equal to the MSFE of the forecasted expected return, corresponding to H_0: $R^2_{OS} \leq 0$ against H_A: $R^2_{OS} > 0$.

Although no strict relation exists between the in-sample and out-of-sample performance (Inoue and Kilian, 2005), the last column of Table 2 and Panel B of Fig. 1 show that the R^2_{OS} is smaller than the in-sample R^2 on average. Of the 55 assets, 45 have negative R^2_{OS}, indicating no out-of-sample predictability. Of the remaining ten assets with positive R^2_{OS}, only three are significant at the 10% level, which are the two-year European bond, the two-year US bond, and the JPY/USD (Japanese yen/US dollar) forward. As a result, the average R^2_{OS} across the 55 assets is -0.67%, suggesting that there is no TSM out of sample.

To further explore the robustness, Fig. 2 plots the R^2s and R^2_{OS}s of TSM by regressing the next one-month return on the past one-, three-, and six-month return, respectively. The results are similar to the case with the past 12-month return in Fig. 1. In addition, we consider volatility scaling when running the asset-by-asset time series regressions. The results are still quantitatively the same: Only the same three assets have significant R^2_{OS}s. In sum, based on the typical univariate time series regression, the evidence of TSM across all the assets is very weak.

4. Pooled regression

In this section, we first replicate the results in MOP (2012) and then show that the pooled regression tends to overstate the presence of TSM.

4.1. The t-statistic

By stacking all futures contracts' returns and dates, MOP (2012) run a pooled predictive regression of monthly returns scaled by volatility on the scaled returns lagged h months,

$$r^i_{t+1}/\sigma^i_t = \alpha + \beta r^i_{t-h+1}/\sigma^i_{t-h} + \varepsilon^i_{t+1}, \tag{3}$$

where r^i_{t+1} is asset i's return in month $t+1$ and σ^i_t is the ex ante annualized volatility estimated by its exponentially weighted lagged squared daily returns:

$$(\sigma^i_t)^2 = 261 \sum_{j=0}^{\infty} (1-\delta)\delta^j (r^i_{t-1-j} - \bar{r}^i_t)^2, \tag{4}$$

where $\bar{r}_t^i$ is the exponentially weighted average return and δ is chosen so that $\sum_{j=0}^{\infty}(1-\delta)\delta^j = 60$ days.

MOP (2012) also use an alternative specification with the sign of lagged returns as the regressor to examine the robustness of TSM,

$$r_{t+1}^i/\sigma_t^i = \alpha + \beta sign(r_{t-h+1}^i) + \varepsilon_{t+1}^i, \tag{5}$$

where *sign* is the sign function that equals $+1$ when $r_{t-h+1}^i \geq 0$ and -1 when $r_{t-h+1}^i < 0$.

As in MOP (2012), we calculate the t-statistics by clustering the standard errors by time (month) and plot the t-statistics of the pooled regression slopes with lagged returns from one month to 60 months in Fig. 3.[6] Qualitatively, we confirm MOP (2012) that the past 12-month return of each asset is a positive predictor of its future returns for one month to 12 months in the pooled regression. After 12 months, the forecasting sign changes and the forecasting power decays. In Fig. 3, Panel A shows the results with Eq. (3) over all asset classes, and Panel B replaces the lagged return with its sign as Eq. (5). Both have a sizable t-statistic of about 4 at the 12-month horizon.

Panels C to F of Fig. 3 plot the t-statistics of Eq. (5) within each asset class and exhibit a similar pattern. The t-statistics appear to show a strong return continuation for the first 12 months and weak reversal for the following 48 months. Overall, Fig. 3 appears to provide strong evidence on TSM. However, the t-statistics at the conventional level tend to overstate the predictability of the past 12-month return.

4.2. Estimation bias

In Eq. (3), MOP (2012) make an implicit assumption that the mean returns of all assets are the same by imposing a common intercept. From Table 1, the sample means of individual assets

[6] The t-statistics that double-cluster the standard errors by time and asset are quantitatively similar.

vary dramatically across asset classes. In the literature, Jorion and Goetzmann (1999) show strong evidence that the equity premium varies across countries. Ang and Bekaert (2007) investigate return predictability with pooled regression but explicitly consider the variation in average returns. Menzly, Santos and Veronesi (2004) analyze cross-sectional differences in time series return predictability.

To highlight fixed effects, a possible specification is

$$r^i_{t+1}/\sigma^i_t = \alpha + \beta r^i_{t-h+1}/\sigma^i_{t-h} + \mu_i/\sigma_i + \varepsilon^i_{t+1}, \tag{6}$$

where μ_i and σ_i are the unconditional mean and volatility of asset i. Hence, the estimate of β from Eq. (3) should be

$$\hat{\beta} = \beta + \frac{\mathrm{Cov}(r^i_{t-h+1}/\sigma^i_{t-h}, \mu_i/\sigma_i)}{\mathrm{Var}(r^i_{t-h+1}/\sigma^i_{t-h})}. \tag{7}$$

If all assets have the same Sharpe ratio (or mean, if volatilities are the same), the second term is zero. Otherwise, it would be significantly positive when the number of assets is large, as the correlation between realized returns and their means is mechanically positive. As a result, the slope estimate of Eq. (3) is biased upward.

The question then is whether the 55 assets have the same mean or Sharpe ratio. We perform four tests for this hypothesis. The first is analysis of variance (ANOVA), which was proposed by Ronald A. Fisher in 1918 (Fisher, 1918) with two assumptions: normality and homoskedasticity. The second is B.L. Welch's ANOVA (Welch, 1951), which allows the variance to be Heteroskedastic. The third is the Kruskal-Wallis test, which relaxes both the normality and homoskedasticity assumptions (Kruskal and Wallis, 1952), and the fourth is a bootstrap test. When applying these four tests to real data, Table 3 shows that the null hypothesis that all 55 assets have the same mean is strongly rejected. In addition, we reject the null that they have the same Sharpe ratio. Therefore, the evidence of TSM shown in MOP (2012) is at least partially driven by the fixed effects.

The fixed effects are not easily corrected in the predictive regression framework. Statistically, Hjalmarsson (2010) shows that when different assets have different average returns, the pooled regression, after controlling for fixed effects, suffers from a looking-forward bias because the time series demeaning of the data requires information after month t, which induces a correlation between the lagged value of the demeaned regressor and the error term in the predictive regression.

In addition to the fixed effects, two more reasons can lead to overstating the evidence on time series predictability. First, as a predictor, the past 12-month cumulative return is persistent and can generate substantial size distortions (Hodrick, 1992; Stambaugh, 1999; Valkanov, 2003; Campbell and Yogo, 2006; Ang and Bekaert, 2007; Boudoukh, Richardson and Whitelaw, 2008; Li and Yu, 2012). For example, Ang and Bekaert (2007) show substantial size distortions with the Newey-West t-statistic when predicting stock returns with persistent variables. Second, because volatility varies dramatically across assets, volatility scaling in the pooled regression without controlling for fixed effects can further exacerbate the upward bias. For example, in Eq. (6), even when all assets have the same mean, volatility scaling generates the fixed effects as σ_i varies dramatically across assets. MOP (2012) also explore the pooled regression in Eq. (5) by using the sign of the past 12-month return as the predictor, which can distort and change the true statistical significance as the sign of the past 12-month return is highly skewed.

4.3. Bootstrap tests

Due to the concerns discussed above, the t-statistic from the pooled regression is questionable. To correctly evaluate the statistic significance, we use bootstrap to simulate the distribution of the t-statistic and define its 97.5% quantile as the simulated t-statistic for significance at the 5% level. If the t-statistic from the real data is larger than 1.96 but smaller than the simulated t-statistic, we can conclude that the pooled regression tends to overreject the null hypothesis and no significant evidence supports TSM.

We use two standard bootstrap approaches. The first is a more restrictive parametric wild bootstrap that samples data based on the pooled regression residuals, and the second is a more general nonparametric pairs bootstrap that resamples the predictor and the dependent variable simultaneously. Both approaches accommodate conditional heteroskedasticity, but the second allows for a wider range of data-generating processes. Pairs bootstrap is considered the most general and applicable method of bootstrapping.

Wild bootstrap. Suppose the true data-generating process is as Eq. (3). Let $\hat{\alpha}$ and $\hat{\beta}$ be the estimates from the full sample of real data. Then, the residuals are

$$\hat{\varepsilon}^i_{t+1} = r^i_{t+1}/\sigma^i_t - \hat{\alpha} - \hat{\beta} r^i_{t-h+1}/\sigma^i_{t-h}. \tag{8}$$

We simulate a pseudo sample path with T observations as

$$r^{i*}_{t+1}/\sigma^{i*}_t = \hat{\alpha} + \hat{\beta} r^i_{t-h+1}/\sigma^i_{t-h} + \hat{\varepsilon}^i_{t+1} v^i_{t+1}, \tag{9}$$

where $*$ indicates that the value is a bootstrapped observation and v^i_t is a random draw from a two-point Rademacher distribution with mean 0 and variance 1:

$$v^i_t = \begin{cases} 1 & \text{with probability } 1/2, \\ -1 & \text{with probability } 1/2. \end{cases} \tag{10}$$

This distribution has an appealing property that the error-in-rejection probability is minimal when the sample size is small, and it is robust to other distributions such as the Mammen distribution and standard normal distribution (Davidson and Flachaire, 2008). After constructing a pseudo sample path, we run pooled regression Eq. (3). We repeat this procedure one thousand times to calculate the simulated t-statistic of $\hat{\beta}$.

Pairs bootstrap. Pairs bootstrap resamples T pairs of $(r^i_{t+1}/\sigma^i_t, r^i_{t-h+1}/\sigma^i_{t-h})$ with replacement from the real data and uses these pairs to run the pooled regression

$$r^{i*}_{t+1}/\sigma^{i*}_t = \alpha_h + \beta_h r^{i*}_{t-h+1}/\sigma^{i*}_{t-h} + \varepsilon^i_{t+1}. \tag{11}$$

Hence, the simulated t-statistic of $\hat{\beta}$ can be calculated after repeating the procedure one thousand times. This bootstrap allows not only more general data-generating processes, but also potential model misspecification [e.g., $\mathrm{E}(r)$ is not a linear function of r_{t-h+1}].

Table 4 reports the t-statistics with real data and the bootstrapped t-statistics. Consistent with Ang and Bekaert (2007), the t-statistics that cluster by time tend to overreject the null hypothesis. For example, when forecasting the next one-month return with the past one-month return, the t-statistic from the real data is 3.11, suggesting strong evidence of TSM. However, this is not the case because the bootstrapped t-statistics are 9.26 and 3.63, respectively. Similarly, when forecasting the next one-month return with the past 12-month return, the t-statistic from the real data is 4.34, and the simulated t-statistics are 12.53 and 4.83, respectively, suggesting that the evidence is weak in support of TSM. Moreover, forecasting with the sign of lagged returns does not support TSM either.

Table 5 presents the simulated t-statistics within each asset class. For brevity, we report the results of predicting the next one-month return with the past one-, three-, six-, and 12-month return, respectively. Consistent with Tables 2 and 4, the results reveal that TSM is unlikely to be present in any of the four asset classes.

Does volatility scaling play a role in estimating the regression slopes and detecting the existence of TSM because MOP (2012) run Eqs. (3) and (5) with volatility scaling while volatility varies across assets and is predictable by its lagged values (Paye, 2012)? Table 6 reports the t-statistics with real data and boostrapped t-statistics from Eqs.(3) and (5) without volatility scaling. The results display two empirical facts. First, the t-statistics without volatility scaling are much smaller than those with volatility scaling. For example, when predicting the next one-month return with the past one- and 12-month return without volatility scaling, the t-statistics of the regression slope are 1.80 and 1.68, respectively, which are much smaller than the values with volatility scaling in Table 4 (3.11 and 4.34), lending little support to TSM. Second, when predicting the next one-month return with the signs of the past one- and 12-month return, the t-statistics are 2.20 and 3.72, respectively, and they are smaller than that with volatility scaling. Therefore, volatility scaling plays a role. In fact, it seems at least partially responsible for the performance of the TSM trading strategy.

This paper extends the sample ending period of MOP (2012) from 2009 to 2015 and raises a possibility that TSM exists before 2009 and disappears thereafter. Table 7 reports the results for the 1985 to 2009 sample period and rules out the possibility. The t-statistics from real data are still smaller than the simulated t-statistics, regardless of which bootstrap approach is employed. For example, when forecasting the next one-month return with the past 12-month return, the t-statistic is 4.48 with real data, but it is 12.76 and 4.96 with the two bootstrap approaches, respectively. The TSM is also insignificant for each asset class. The results are reported in the Online Appendix.

4.4. Controlling for fixed effects

Earlier evidence shows that the assets do not have the same mean, implying that fixed effects should be controlled in the pooled regression. In so doing, one can run the pooled regression by removing the asset means. The bootstrap procedures can also make a similar modification. The question is whether controlling for fixed effects can alter substantially the evidence on TSM.

Following Gonçalves and Kaffo (2015), we now compute the t-statistic from the pooled regression

$$r^i_{t+1}/\sigma^i_t - \overline{r^i/\sigma^i} = \beta(r^i_{t-h+1}/\sigma^i_{t-h} - \overline{r^i_{-h+1}/\sigma^i_{-h}}) + \varepsilon^i_{t+1}, \tag{12}$$

where $\overline{r^i/\sigma^i}$ and $\overline{r^i_{-h+1}/\sigma^i_{-h}}$ denote the time series averages of r^i_{t+1}/σ^i_t and $r^i_{t-h+1}/\sigma^i_{t-h}$, respectively. Suppose the estimate of β is $\hat{\beta}_{FE}$, then the residual $\hat{\varepsilon}^i_{t+1}$ can be calculated as

$$\hat{\varepsilon}^i_{t+1} = r^i_{t+1}/\sigma^i_t - \overline{r^i/\sigma^i} - \hat{\beta}_{FE}(r^i_{t-h+1}/\sigma^i_{t-h} - \overline{r^i_{-h+1}/\sigma^i_{-h}}). \tag{13}$$

Then, we simulate a pseudo sample path with T observations as

$$(r_{t+1}^i/\sigma_t^i - \overline{r^i/\sigma^i})^* = \hat{\beta}_{FE}(r_{t-h+1}^i/\sigma_{t-h}^i - \overline{r_{-h+1}^i/\sigma_{-h}^i}) + \hat{\varepsilon}_{t+1}^i v_{t+1}^i, \tag{14}$$

where v_{t+1}^i follows the Rademacher distribution. We can then estimate the model with the simulated samples and repeat the procedure one thousand times, to obtain the critical value of the wild bootstrap t-statistic.

Regarding the pairs bootstrap, we resample T pairs of the predictor and the dependent variable in Eq. (12) with replacement from real data after de-meaning and then use these pairs to rerun the pooled regression. The pairs bootstrap t-statistic is naturally obtained after repeating the procedure one thousand times. Both the wild and pairs bootstraps do not suffer from the incidental parameter bias emphasized in Gonçalves and Kaffo (2015), because the sample size is relatively large here.

Table 8 reports the t-statistics from the pooled regression and the bootstrapped t-statistics. Compared with Table 4, after controlling for the fixed effects, the t-statistic is smaller than that without controlling for fixed effects. For example, when predicting the next one-month return with the past 12-month return, the t-statistic is 4.34 in Table 4 and 3.37 in Table 8. The most important result is that the t-statistic of 3.37 when controlling for fixed effects does not affect the conclusion that insufficient evidence exists in support of TSM. In the Online Appendix, we show that this finding is robust to the cases within each asset class and without volatility scaling.

4.5. Out-of-sample performance

A further implicit assumption of a pooled regression is that all 55 futures contracts are homogenous with the same slope in Eqs. (3) and (5). If the individual slopes are all identical, the pooled estimate converges to the common slope, and pooling data leads to a more precise estimate than the individual time series regression estimate. Whether the slopes of all individual assets are identical or not, there is no guarantee that pooling the data will help. Nevertheless, whether pooling the data improves out-of-sample performance is of interest to examine empirically.

Fig. 4 plots the out-of-sample R^2_{OS} for each individual asset. In Panel A, we present the results with volatility scaling when running the pooled regression. To predict returns in month $t+1$ at the end of month t, we run pooled regression Eq. (3) with returns up to month t as MOP (2012). Let $\hat{\alpha}_t$ and $\hat{\beta}_t$ be the estimated intercept and slope. We calculate the expected return of asset i for month $t+1$ as

$$\mathrm{E}_t(r^i_{t+1}) = \hat{\alpha}_t \sigma^i_t + \hat{\beta}_t \frac{r^i_{t-12,t}}{\sigma^i_{t-1}} \sigma^i_t, \tag{15}$$

which can be plugged into Eq. (2) to calculate the R^2_{OS} accordingly.

In comparison with earlier univariate regressions (Table 2), the pooled regression does improve the out-of-sample forecasting performance in some of the markets. The R^2_{OS} is significantly positive for three commodity futures contracts: cocoa, copper, and gold. Of the nine international equity markets, six are significant at the 10% level, with the remaining three positive but not significant. The average R^2_{OS} in the equity markets is 2.08%, indicating potential economic significance as well (Campbell and Thompson, 2008). The R^2_{OS}s in the bond and currency markets are generally negative or slightly positive. The two exceptions are the two-year European bond and two-year US bond. Overall, if there is any TSM, it appears to show up in the international equity markets only, not present in the entire cross section of assets.

Panel B of Fig. 4 plots the R^2_{OS} for each asset without volatility scaling when running regression Eq. (3). The out-of-sample performance does not change significantly. Untabulated results show that the value of R^2_{OS} with volatility scaling is generally similar to that without volatility scaling. Two exceptions are the two-year European bond and two-year US bond, which have extreme positive R^2_{OS} in the case with volatility scaling (15.18% and 3.48%) but extreme negative R^2_{OS} in the case without volatility scaling (-19.54% and -16.71%). As such, the average R^2_{OS} using volatility scaling is -0.06%, which is larger than -0.35% without volatility scaling.

Overall, to a certain extent, a pooled regression can improve the out-of-sample forecasting power relative to the asset-by-asset time series regression, but such improvement is restricted to some specific assets. For the entire cross section of assets, it does little to improve their out-of-sample forecasting performances, and it cannot provide significant support for TSM either.

5. Trading strategy

In this section, we examine the source of profitability of the TSM strategy proposed by MOP (2012). We show in various ways that its performance does not necessarily indicate that TSM exists across assets.

5.1. TSM versus TSH at asset level

The early univariate regressions show that time series predictability is not a common feature across assets, which suggests that the performance of the TSM strategy perhaps is not attributed to predictability, at least not entirely. Furthermore, it raises the possibility that some strategies that do not require predictability can perform as well as the TSM strategy. As it turns out, this is the case.

Suppose the return of asset i follows an independent and identically distributed normal distribution with mean μ^i and volatility σ^i. Then, the probability of the past 12-month return being positive is

$$\Pr(r^i_{t-12,t} > 0) = 1 - \Pr\left(\frac{r^i_{t-12,t} - 12\mu^i}{\sqrt{12}\sigma^i} \leq -\sqrt{12}\frac{\mu^i}{\sigma^i}\right) = \Phi(\sqrt{12}\mu^i/\sigma^i), \tag{16}$$

where $\Phi(\cdot)$ is the $N(0,1)$ cumulative distribution function. Hence, without time series predictability, the TSM strategy tends to buy an asset with high mean return (i.e., Sharpe ratio). Based on this observation, we consider an alternative strategy based on the time series history of asset i's return

$$r^{\text{TSH},i}_{t+1} = sign(r^i_{1,t})r^i_{t+1}, \tag{17}$$

where $r^i_{1,t}$ is the accumulative return of asset i from month 1 to month t or the historical sample mean multiplied by t.

Volatility scaling on the TSH strategy is unnecessary because we compare the TSM and TSH strategies at the asset level in this section. Without volatility scaling, the corresponding return of the TSM strategy in month $t+1$ for asset i is

$$r_{t+1}^{\text{TSM},i} = sign(r_{t-12,t}^{i})r_{t+1}^{i}. \tag{18}$$

Comparing Eqs. (17) and (18), the TSM strategy attempts to exploit possible predictability of the past 12-month return, and the TSH does not rely on any predictability at all. Our goal here is to examine their performances across assets.

Table 9 reports the average returns and Sharpe ratios, and their differences, of the TSM and TSH strategies based on Eqs. (18) and (17), respectively. The results show that the TSM strategy generally performs the same as the TSH strategy. Of the 55 assets, only five show that the TSM strategy generates a higher average return than the TSH strategy. When we use the Sharpe ratio as the performance measure, the results remain unchanged. Thus, the TSM strategy does not significantly outperform at the asset level the TSH strategy that does not require predictability.

5.2. *TSM versus TSH at portfolio level*

Even though no time series predictability exists, the TSM strategy could still be profitable. Conrad and Kaul (1988), Jegadeesh (1990), and Jegadeesh and Titman (1993) note that if there are differences in mean returns, a strategy that buys high-mean assets using the proceeds from selling low-mean assets has a natural tilt toward high-mean assets. To see this, consider just two assets. If the mean return of the first asset far exceeds that of the second, buying the first and shorting the second is profitable. Because the past 12-month return can be viewed as an estimate of the mean return, the TSM strategy could profit from the differences in mean returns. If this is the case, it will unlikely outperform the TSH strategy at the portfolio level.

To make the two strategies comparable, we consider the following equal-weighting scheme for the TSM and TSH:

$$r_{t+1}^{\mathrm{TSM}} = \frac{1}{N_t} \sum_{i=1}^{N_t} sign(r_{t-12,t}^i) r_{t+1}^i \tag{19}$$

and

$$r_{t+1}^{\mathrm{TSH}} = \frac{1}{N_t} \sum_{i=1}^{N_t} sign(r_{1,t}^i) r_{t+1}^i, \tag{20}$$

where N_t is the number of assets investable at time t.[7] By doing so, the two strategies differ only in how the past information is used to select the assets. So, differences in performance stem from the differences in asset selection, not from differences in scaling the portfolio weights. Because our goal here is not to improve the performance, the concern that the TSM strategy tilts weighting toward low volatility assets is not an issue, because the TSH strategy has the same tilts. Nevertheless, we will examine volatility weighting.

Panel A of Table 10 presents the average and risk-adjusted returns of the two strategies, the second of which is computed from two benchmark asset pricing models as in MOP (2012). The first model is the Fama-French four-factor model that uses the MSCI World Index as the market factor, and the second is the Asness, Moskowitz and Pedersen (2013) three-factor model with the MSCI World Index, the value everywhere factor, and the momentum everywhere factor. Over the 1986 to 2015 investment period, the average return differential between the two strategies is as small as 0.14%, not significant with a p-value of 0.19. The results suggest that the TSM strategy generates virtually the same average portfolio return as the TSH strategy, consistent with earlier comparison at the asset level.

When turning to the risk-adjusted returns, the TSM and TSH alphas are 0.15% (t-statistic = 1.94) and 0.05% (t-statistic = 0.80) with the Fama-French four-factor model and 0.07% (t-

[7] TSH and TSM have similar expressions here, in parallel with the asset-level comparison. At the portfolio level, we also explore two alternative TSH strategies that buy (sell) half or one-third of the assets with high (low) historical sample means and find that their performances are quantitatively similar.

statistic $= 1.01$) and 0.09% (t-statistic $= 1.14$) with the Asness, Moskowitz and Pedersen (2013) three-factor model, respectively. As for the case with average return, the two strategies' alpha differentials with the two models are 0.10% (p-value $= 0.29$) and -0.02% (p-value $= 0.84$) and, therefore, are not statistically significant from zero. Thus, the TSM and TSH strategies do not generate sizable abnormal returns. Moreover, their difference in alpha is even smaller than the difference in average return and is also insignificant.

Also reported in Panel A of Table 10 are the average and risk-adjusted returns of the long- and short-leg portfolios of the TSM and TSH strategies. The results show that the performance of the two strategies mainly stems from the long legs and that the performance of their short legs is always indifferent from zero. This new finding, not shown by MOP (2012) or Goyal and Jegadeesh (2018), is consistent with our argument that the strong TSM performance is due to the difference in mean returns.

For robustness, we also consider three alternative portfolio weighting schemes: volatility weighting as in MOP (2012), past-12-month-return weighting, and equal weighting with a zero-investment constraint as in Goyal and Jegadeesh (2018). The results do not change qualitatively, and the alpha differential between the TSM and TSH strategies is always indifferent from zero. The Online Appendix considers the case of constructing the TSM with the past six-month return, instead of the past 12-month return, and the results remain unchanged. In sum, the performance of the TSM strategy in MOP (2012) seems mainly stemming from the difference in mean returns, not from the times series predictability of the past 12-month return.

5.3. TSM and TSH forecast comparison: predictive slope

Lewellen (2015) proposes an interesting predictive slope that assesses the degree of predictability of cross-sectional forecasts in an elegant way, in which one simply runs a cross-sectional regression of the realized returns on the forecasts. If the forecasts are perfect, the slope should be one. Generally, a value less than 0.5 indicates no predictability as the forecasts under-perform naive forecasts.

At the end of month t, we calculate the expected return of asset i as $\hat{r}_{t+1}^{\text{TSM},i}$, which is estimated by pooled regression Eq. (3) with data up to month t, and then we regress month $t+1$ return r_{t+1}^{i} on $\hat{r}_{t+1}^{\text{TSM},i}$. The first three columns of Table 11 report the results. Consistent with the earlier no predictability results, the regression slope is close to zero and its t-statistic is less than one standard deviation, suggesting that the in-sample performance with pooled regression Eq. (3) is not reliable and that the TSM estimates do not line up with the true expected returns out of sample.

We also explore whether the TSM and TSH forecasts have the same mean. The last three columns of Table 11 report the results of regressing the TSM forecasts of expected returns on the TSH forecasts. Because these two estimates are potentially unbiased, we do not include an intercept to improve the estimate efficiency. Consistent with earlier results indicating little difference between the two strategies, the slope is close to one and the t-statistic is much larger than two standard deviations. For robustness, we also perform the tests for each asset class, and the results are the same as the overall case. In short, the TSM strategy has little predictive power and behaves in a very similar manner to the TSH strategy.

5.4. When does the TSM outperform the TSH?

In the previous sections, we have shown that the predictability of the past 12-month return is weak with real data and that the TSM strategy performs similarly as the TSH strategy that does not require any predictability. This section attempts to answer another question: If the predictability is strong, to what extent does the TSM strategy outperform the TSH strategy?[8]

Consistent with the pooled regression of MOP, we assume the following data-generating process:

$$r_{t+1}^{i} = \alpha^{i} + \beta \frac{r_{t-12,t}^{i}}{12} + \varepsilon_{t+1}^{i}, \tag{21}$$

where $\beta = 0.1,\ 0.2$, and 0.4 (we divide the past 12-month return $r_{t-12,t}^{i}$ by 12 to make β easy to

[8] We thank the anonymous referee for this intriguing research question.

interpret; β is 0.08 for the real data). There are two ways to draw the residuals. The first is to ignore the off-diagonal elements and draw the residuals asset by asset, thereby assuming no cross-asset predictability, and the second is to keep the covariance matrix structure and draw the residuals across assets. For a given specification and beta, we simulate a path for the 55 assets with $T = 372$ observations and construct the TSM and TSH strategies accordingly. We repeat this procedure one thousand times to test whether these two strategies generate the same mean returns. Empirically, we find that the two specifications generate almost the same results and therefore focus on the first specification.

Fig. 5 reports the results. When the slope is 0.1, the two strategies perform almost the same. When the slope is 0.2, the TSM outperforms the TSH, but the difference is not significant. Thus, if there is genuine time series predictability, the advantage of the TSM strategy is not apparent as long as the slope is small. When the slope is 0.4, the TSM dominates the TSH in the sense that it does better in almost all the simulated data sets. Because the two strategies generate similar performance using the real data, our simulation indicates that the evidence of time series predictability is weak if it exists. Combined with other results, the TSM is unlikely to be statistically significant for all the assets. In short, a lack of empirical evidence exists to support that the TSM is everywhere.

Because the TSM and TSH use overlapping data at the beginning of the investment period, one could expect that this explains the statistically indistinguishable difference even when the slope is 0.2. In fact, it is not the case. To see this, we simulate a path of $T+240$ observations for the 55 assets, construct the TSM and TSH strategies starting from the 241th observation (i.e., the TSM is based on the past 12-month return and the TSH is based on the historical sample mean), and calculate their mean returns. We repeat this procedure one thousand times to test whether these two strategies generate the same mean returns. Fig. A1 of the Online Appendix summarizes the results. Generally, the basic patterns are similar to Fig. 5; that is, the TSM strategy performs similarly as the TSH when the data exhibit weak or intermediate time series predictability, and it outperforms the TSH when the data exhibit strong time series predictability. One observation is that with longer historical data in calculating the historical mean, the TSH generates 1% more annualized return relative to the case of Fig. 5, which is simply due to the fact that observations over a longer time horizon generate a better estimate of the sample mean (Merton, 1980).

6. Conclusion

In their influential study, MOP (2012) assert a surprising time series momentum that the past 12-month return can positively predict the next one-month to 12-month return everywhere. They also show that a trading strategy based on TSM generates significant average and risk-adjusted returns. As TSM is in stark contrast to the previous literature and strongly challenges the weak form market efficiency hypothesis, we revisit TSM in this paper. Employing the same data as MOP but extending the sample period to 2015, we show that, statistically, the evidence for TSM is weak in asset-by-asset time series regressions and a pooled regression accounting for size distortions. Economically, we show that the performance of the TSM strategy is likely driven by differences in mean returns, not predictability. A predictive slope analysis following the approach of Lewellen (2015) further confirms weak evidence on TSM.

A number of topics are of interest for future research. First, while the TSM strategy focuses on the 12-month return predictability, examining such a strategy at other time horizons and an optimal combination of all would be worthwhile. Second, the predictability horizon can be time-varying and could be different across assets (instead of the same horizon), and developing a test and a new trading strategy for this possibility would be important. Finally, considering the conditions under which time series predictability can exist in an equilibrium model and developing an empirical test for the implications would be desirable.

References

Ang, A., Bekaert, G., 2007. Stock return predictability: is it there? Review of Financial Studies 20, 651–707.

Asness, C., Moskowitz, T., Pedersen, L., 2013. Value and momentum everywhere. Review of Financial Studies 20, 651–707.

Boudoukh, J., Richardson, M., Whitelaw, R., 2008. The myth of long-horizon predictability. Review of Financial Studies 21, 1577–1605.

Campbell, J., Thompson, S., 2008. Predicting excess stock returns out of sample: can anything beat the historical average? Review of Financial Studies 21, 1509–1531.

Campbell, J., Yogo, M., 2006. Efficient tests of stock return predictability. Journal of Financial Economics 81, 27–60.

Clark, T., West, K., 2007. Approximately normal tests for equal predictive accuracy in nested models. Journal of Econometrics 138, 291–311.

Conrad, J., Kaul, G., 1988. An anatomy of trading strategies. Review of Financial Studies 11, 489–519.

Davidson, R., Flachaire, E., 2008. The wild bootstrap, tamed at last. Journal of Econometrics 146, 162–169.

Fama, E., 1965. The behavior of stock market prices. Journal of Business 38, 34–105.

Fisher, R.A., 1918. The correlation between relatives on the supposition of Mendelian inheritance. Philosophical Transactions of the Royal Society of Edinburgh 52, 399–433.

French, K., Roll, R., 1986. Stock return variances: the arrival of information and the reaction of traders. Journal of Financial Economics 17, 5–26.

Freyberger, J., Neuhierl, A., Weber, M., 2019. Dissecting characteristics nonparametrically. Review of Financial Studies, forthcoming.

Georgopoulou, A., Wang, J., 2017. The trend is your friend: time series momentum strategies across equity and commodity markets. Review of Finance 21, 1557–1592.

Gonçalves, S., Kaffo, M., 2015. Bootstrap inference for linear dynamic panel data models with individual fixed effects. Journal of Econometrics 186, 407–426.

Goyal, A., Jegadeesh, N., 2018. Cross-sectional and time series tests of return predictability: what is the difference? Review of Financial Studies 31, 1784–1824.

Gu, S., Kelly, B.T., Xiu, D., 2018. Empirical asset pricing via machine learning. Unpublished working paper. University of Chicago.

Han, Y., He, A., Rapach, D., Zhou, G., 2019. What firm characteristics drive US stock returns? Unpublished working paper. Washington University in St. Louis.

Hansen, P., Timmermann, A., 2012. Choice of sample split in out-of-sample forecast evaluation. Unpublished working paper. University of California, San Diego.

Harvey, C.R., Liu, Y., Zhu, H., 2016. ... and the cross section of expected returns. Review of Financial Studies 29, 5–68.

Hjalmarsson, E., 2010. Predicting global stock returns. Journal of Financial and Quantitative Analysis 45, 49–80.

Hodrick, J., 1992. Dividend yields and expected stock returns: alternative procedures for inferecne and measurement. Review of Financial Studies 5, 357–386.

Hou, K., Xue, C., Zhang, L., 2019. Replicating anomalies. Review of Financial Studies, forthcoming.

Hurst, B., Ooi, Y., Pedersen, L., 2017. A century of evidence on trend-following investing. Journal of Portfolio Management 44, 15–29.

Inoue, A., Kilian, L., 2005. In-sample or out-of-sample tests of predictability: which one should we use? Econometric Reviews 23, 371–402.

Jegadeesh, N., 1990. Evidence of predictable behavior of security returns. Journal of Finance 45, 881–898.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. Journal of Finance 48, 65–91.

Jiang, F., Lee, J., Martin, X., Zhou, G., 2019. Manager sentiment and stock returns. Journal of Financial Economics 132, 126–149.

Jorion, P., Goetzmann, W., 1999. Global stock markets in the 20th century. Journal of Finance 54, 953–980.

Kim, A., Tse, Y., Wald, J., 2016. Time series momentum and volatility scaling. Journal of Financial Markets 30, 102–124.

Koijen, R., Moskowitz, T., Pedersen, L., Vrugt, E., 2018. Carry. Journal of Financial Economics 127, 197–225.

Kruskal, W.H., Wallis, W.A., 1952. Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association 47, 583–621.

Lewellen, J., 2015. The cross section of expected stock returns. Critical Finance Review 4, 1–44.

Li, J., Yu, J., 2012. Investor attention, psychological anchors, and stock return predictability. Journal of Financial Economics 104, 401–419.

Lo, A., MacKinlay, A., 1988. Stock market prices do not follow random walks: evidence from a simple specification test. Review of Financial Studies 1, 41–66.

Menzly, L., Santos, T., Veronesi, P., 2004. Understanding predictability. Journal of Political Economy 112, 1–47.

Merton, R.C., 1980. On estimating the expected return on the market: an exploratory investigation. Journal of Financial Economics 8, 323–361.

Moskowitz, T., Ooi, Y., Pedersen, L., 2012. Time series momentum. Journal of Financial Economics 104, 228–250.

Paye, B.S., 2012. 'déjà vol': predictive regressions for aggregate stock market volatility using macroeconomic variables. Journal of Financial Economics 106, 527–546.

Rapach, D., Strauss, J., Zhou, G., 2013. International stock return predictability: what is the role of the united states? Journal of Finance 68, 1633–1662.

Stambaugh, R.F., 1999. Predictive regressions. Journal of Financial Economics 54, 375–421.

Valkanov, R., 2003. Long-horizon regressions: theoretical results and applications. Journal of Financial Economics 68, 201–232.

Welch, B.L., 1951. On the comparison of several mean values: an alternative approach. Biometrika 38, 330–336.

Welch, I., Goyal, A., 2008. A comprehensive look at the empirical performance of equity premium performance. Review of Financial Studies 21, 1455–1508.

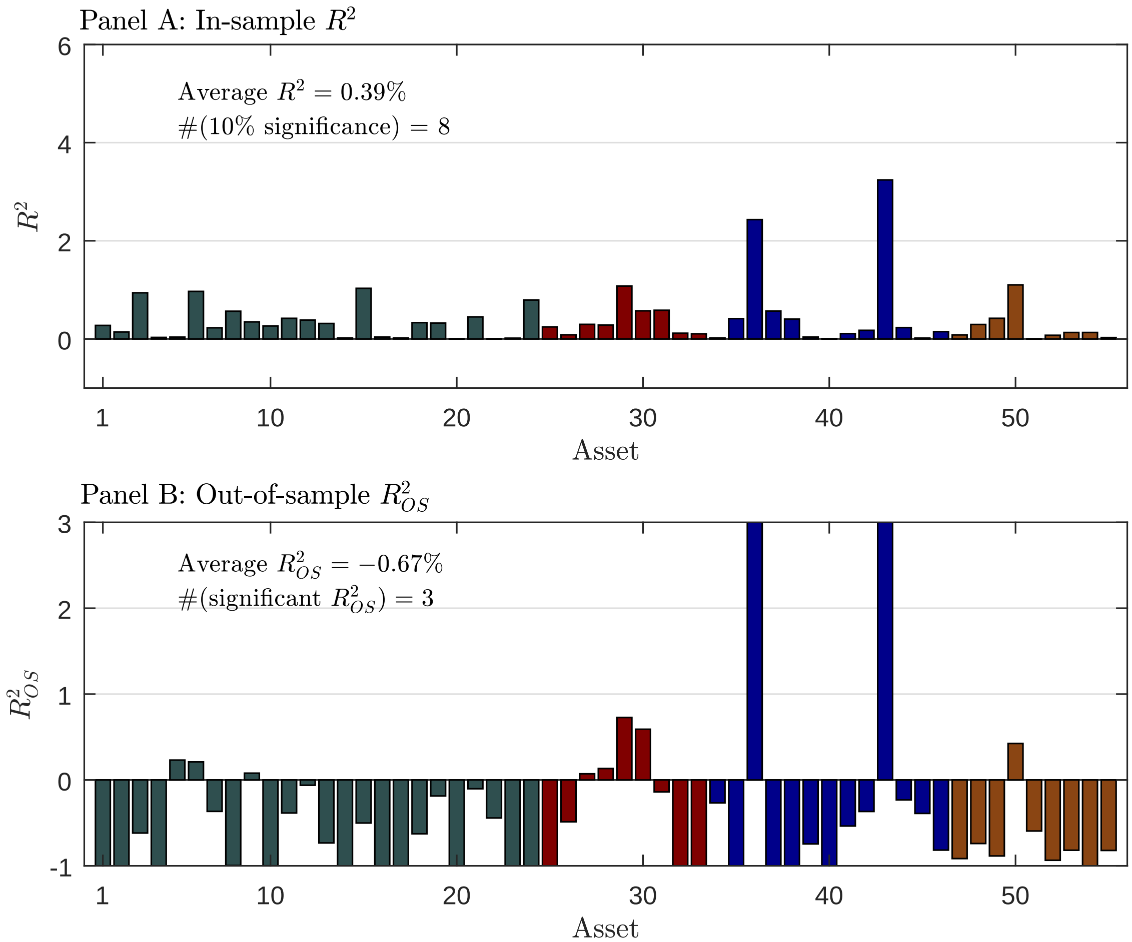


Fig. 1. Time series momentum (TSM) with asset-by-asset regression. This figure plots the in- and out-of-sample R^2s of forecasting a futures contract return with time series regression as

$$r^i_{t+1} = \alpha_i + \beta_i r^i_{t-12,t} + \varepsilon^i_{t+1},$$

where $r^i_{t-12,t}$ is asset i's past 12-month return. The in- and out-of-sample periods are 1985:01–2015:12 and 2000:01–2015:12, respectively.

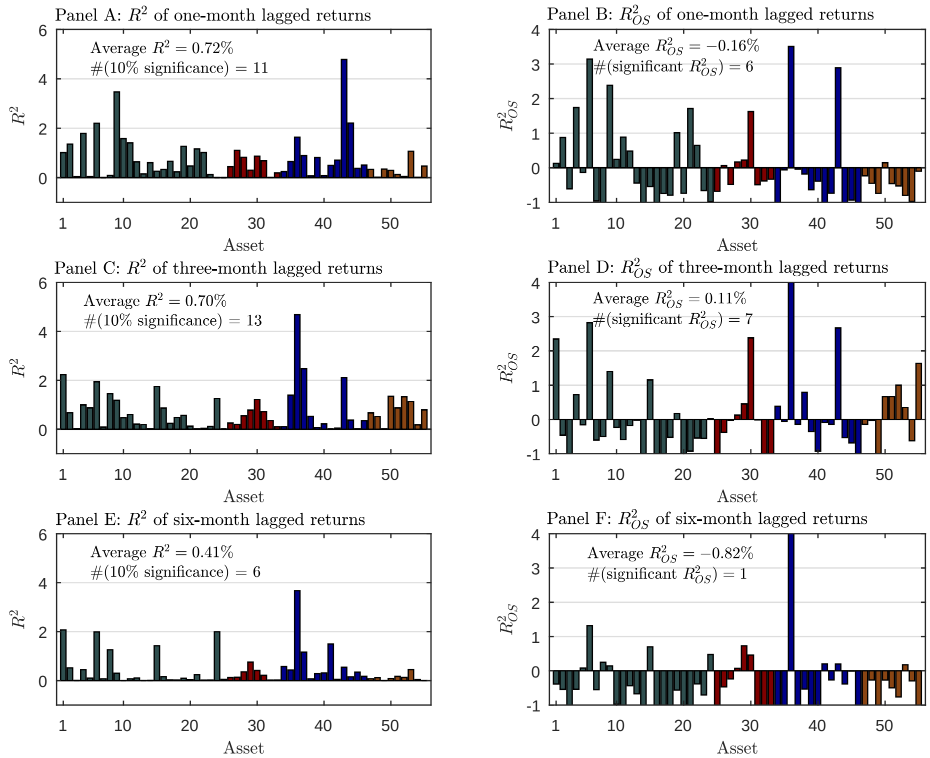


Fig. 2. Time series momentum (TSM) with asset-by-asset regression based on different lags. This figure plots the in- and out-of-sample R^2s of forecasting a futures contract return with time series regression as

$$r^i_{t+1} = \alpha_i + \beta_i r^i_{t-h,t} + \varepsilon^i_{t+1},$$

where $r^i_{t-h,t}$ is asset i's past h-month return ($h = 1$, 3, and 6). The in- and out-of-sample periods are 1985:01–2015:12 and 2000:01–2015:12, respectively.

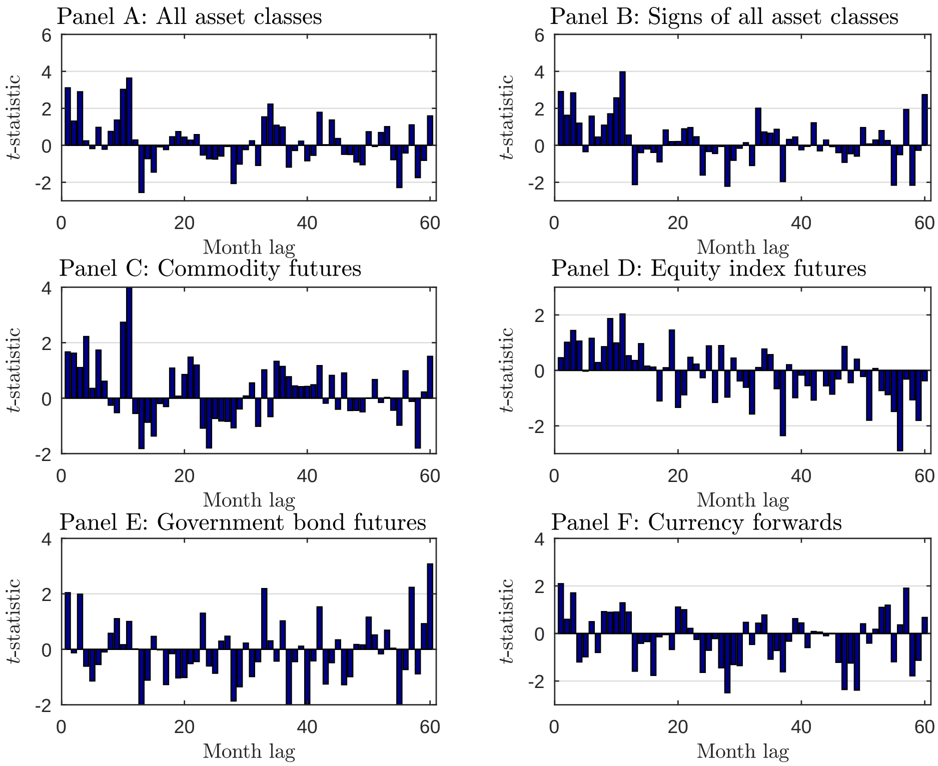


Fig. 3. Time series momentum (TSM) with pooled regression: in-sample performance. This figure plots the t-statistics of the pooled regression slopes that regress month t returns on month $t-h$ returns as

$$r^i_{t+1}/\sigma^i_t = \alpha_h + \beta_h r^i_{t-h+1}/\sigma^i_{t-h} + \varepsilon^i_{t+1},$$

for Panel A and

$$r^i_{t+1}/\sigma^i_t = \alpha_h + \beta_h sign(r^i_{t-h+1}) + \varepsilon^i_{t+1}$$

for Panels B to F, where r^i_{t-h+1} is asset i's return in month $t-h+1$ for $h = 1, 2, \cdots, 60$. The t-statistics are clustered by time (month). The sample period is 1985:01–2015:12.

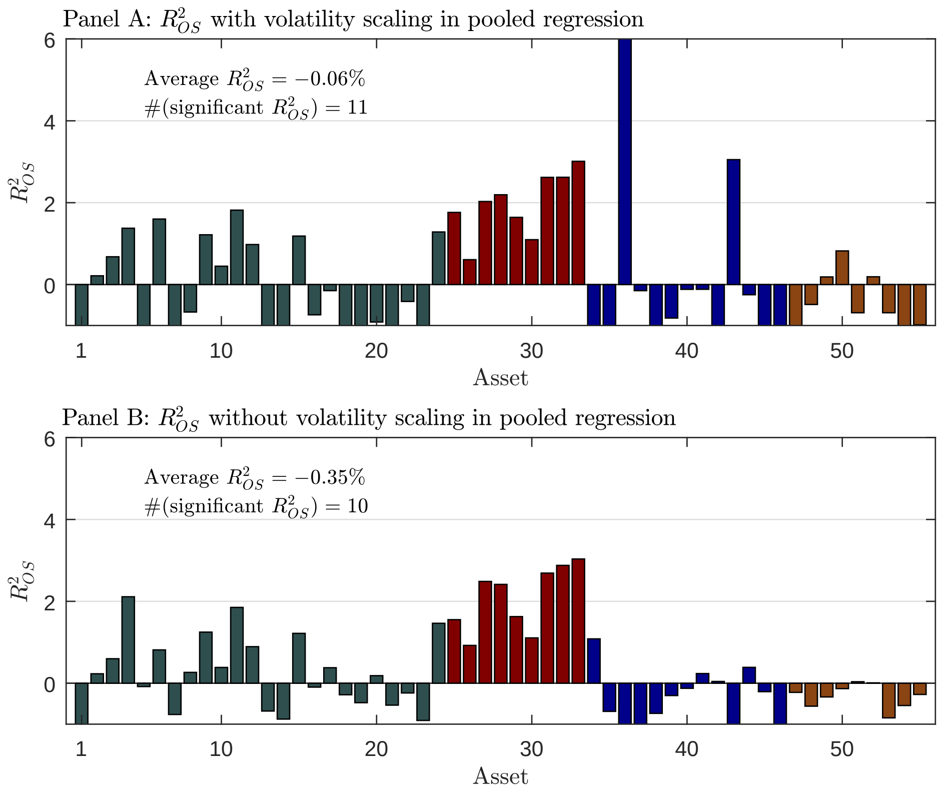


Fig. 4. Time series momentum (TSM) with pooled regression: out-of-sample performance. This figure plots the out-of-sample R^2_{OS} of forecasting a futures contract return with pooled regression

$$r^i_{t+1}/\sigma^i_t = \alpha + \beta r^i_{t-12,t}/\sigma^i_{t-1} + \varepsilon^i_{t+1},$$

for Panel A and

$$r^i_{t+1} = \alpha + \beta r^i_{t-12,t} + \varepsilon^i_{t+1}$$

for Panel B, where $r^i_{t-12,t}$ is asset i's past 12-month return. We calculate asset i's out-of-sample R^2_{OS} by applying the same $\hat{\alpha}_t$ and $\hat{\beta}_t$ to all assets in estimating the expected return as $\mathrm{E}_t(r^i_{t+1}) = \hat{\alpha}_t \sigma^i_t + \hat{\beta}_t \frac{r^i_{t-12,t}}{\sigma^i_{t-1}} \sigma^i_t$ in Panel A and $\mathrm{E}_t(r^i_{t+1}) = \hat{\alpha}_t + \hat{\beta}_t r^i_{t-12,t}$ in Panel B, where $\hat{\alpha}_t$ and $\hat{\beta}_t$ are the pooled regression estimates with data up to month t. The out-of-sample period is 2000:01–2015:12.

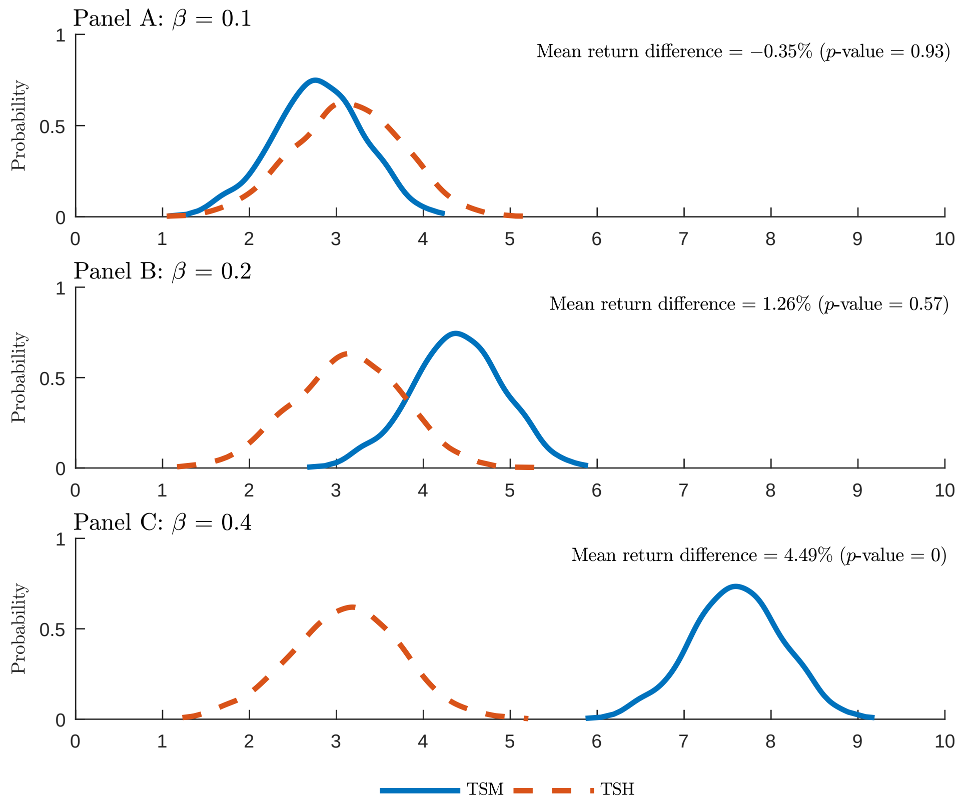


Fig. 5. Annualized mean return difference between time series momentum (TSM) and time series history (TSH). This figure plots the distributions of simulated annualized mean returns of the TSM and TSH strategies, where each asset is assumed to follow

$$r^i_{t+1} = \alpha^i + \beta \frac{r^i_{t-12,t}}{12} + \varepsilon^i_{t+1},$$

where β equals 0.1, 0.2, and 0.4. For each asset, we assume it has the same mean and variance as that in Table 1. Then, given the common slope β, α^i is estimated with asset i's real returns. We simulate a path of $T = 372$ observations, construct the TSM and TSH strategies, and calculate their mean returns. We repeat this procedure one thousand times to test whether these two strategies generate the same mean returns.

Table 1
Summary statistics of 55 assets across four asset classes

This table reports the mean return, volatility (standard deviation), and first-order autocorrelation [$\rho(1)$], where the mean and volatility are annualized and represented in percentage. "Average" refers to the average value within asset class. The sample period is 1985:01–2015:12.

Asset	Start date	Mean	Volatility	$\rho(1)$
Panel A: Commodity futures				
Aluminum	November 1986	−2.09	19.92	0.10
Brent oil	February 1992	7.77	30.62	0.12
Cattle	January 1985	1.64	13.56	0.02
Cocoa	January 1985	−2.54	28.11	−0.13
Coffee	January 1985	−2.06	37.95	−0.02
Copper	January 1985	11.48	26.80	0.15
Corn	January 1985	−4.91	26.65	0.00
Cotton	January 1985	1.36	25.83	0.03
Crude	January 1985	6.37	34.98	0.19
Gas oil	April 1989	8.21	33.23	0.12
Gold	January 1985	1.54	15.65	−0.12
Heating oil	January 1985	6.41	32.60	0.08
Hogs	January 1985	−3.70	24.23	−0.04
Natural gas	April 1990	−11.33	50.79	0.08
Nickel	February 1993	7.06	34.37	0.05
Platinum	January 1992	6.36	20.76	0.06
Silver	January 1985	2.02	27.73	−0.08
Soybeans	January 1985	4.01	23.17	−0.05
Soymeal	January 1985	6.23	28.75	−0.11
Soy oil	October 1990	4.33	26.12	−0.07
Sugar	January 1985	6.24	34.75	0.11
Unleaded gasoline	January 1985	12.31	35.71	0.10
Wheat	January 1985	−4.33	26.68	−0.03
Zinc	February 1991	−0.33	25.08	0.00
Average		2.59	28.50	0.02
Panel B: Equity index futures				
SPI 200	January 1985	7.41	16.11	0.00
DAX	January 1985	9.69	21.70	0.07
IBEX 35	January 1985	9.27	22.66	0.10
CAC 40	January 1985	6.73	19.69	0.09
FTSE/MIB	January 1985	6.29	23.21	0.05
TOPIX	January 1985	3.09	19.70	0.09
AEX	January 1985	6.96	19.16	0.08
FTSE 100	January 1985	6.51	15.87	−0.01
S&P 500	January 1985	9.21	15.20	0.04
Average		7.24	19.26	0.06

Panel C: Government bond futures				
Three-year Australian	July 1988	3.34	4.58	−0.05
Ten-year Australian	January 1985	5.57	6.90	0.08
Two-year European	January 1989	1.47	3.41	0.13
Five-year European	January 1989	1.83	4.26	0.09
Ten-year European	January 1985	4.16	9.66	0.03
Thirty-year European	January 1987	7.56	10.39	0.09
Ten-year Canadian	January 1985	6.44	10.79	−0.03
Ten-year Japanese	December 1986	3.66	13.78	0.07
Ten-year UK	January 1985	4.14	8.28	0.08
Two-year US	January 1989	1.49	1.67	0.22
Five-year US	January 1989	2.84	4.30	0.15
Ten-year US	January 1985	3.64	7.60	0.06
Thirty-year US	January 1985	12.59	16.44	0.07
Average		4.54	7.85	0.08
Panel D: Currency forwards				
AUD/USD	January 1985	1.10	12.03	0.06
EUR/USD	January 1985	2.06	11.02	0.01
CAD/USD	January 1985	0.43	7.44	−0.06
JPY/USD	January 1985	1.72	11.12	0.05
NOK/USD	January 1985	0.51	11.01	0.04
NZD/USD	January 1985	2.13	12.31	−0.02
SEK/USD	January 1985	0.39	11.25	0.10
CHF/USD	January 1985	2.92	11.77	−0.01
GBP/USD	January 1985	−0.28	10.14	0.07
Average		1.22	10.90	0.03

Table 2

In- and out-of-sample performance of time series momentum (TSM) with time series regression

This table reports the slope, t-statistic, in-sample R^2, and out-of-sample R^2_{OS} of $r^i_{t+1} = \alpha_i + \beta_i r^i_{t-12,t} + \varepsilon^i_{t+1}$. "Average" refers to the average value within each asset class. #(10% significance) refers to the number of significant in-sample regression slopes or significant R^2_{OS}s at the 10% level or stronger. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. The in- and out-of-sample periods are 1985:01–2015:12 and 2000:01–2015:12, respectively.

Asset	β_i	t-stat	R^2	R^2_{OS}
Panel A: Commodity futures				
Aluminum	0.30	0.88	0.28	−1.42
Brent oil	0.34	0.69	0.14	−1.29
Cattle	0.38**	2.23	0.94	−0.62
Cocoa	−0.14	−0.28	0.03	−1.51
Coffee	0.21	0.40	0.04	0.23
Copper	0.77*	1.69	0.97	0.21
Corn	−0.37	−0.94	0.23	−0.37
Cotton	0.57	1.23	0.56	−1.00
Crude	0.60	1.34	0.35	0.08
Gas oil	0.49	1.11	0.26	−1.00
Gold	0.29	1.43	0.42	−0.38
Heating oil	0.59	1.39	0.38	−0.06
Hogs	0.39	1.25	0.31	−0.73
Natural gas	−0.21	−0.30	0.02	−4.51
Nickel	1.01*	1.83	1.03	−0.50
Platinum	−0.12	−0.31	0.04	−1.83
Silver	−0.11	−0.25	0.02	−2.31
Soybeans	−0.39	−1.05	0.33	−0.63
Soymeal	−0.47	−1.06	0.32	−0.19
Soy oil	0.04	0.09	0.00	−1.93
Sugar	−0.65	−1.33	0.45	−0.10
Unleaded gasoline	0.05	0.12	0.00	−0.44
Wheat	−0.10	−0.26	0.02	−1.21
Zinc	0.65	1.24	0.79	−2.29
Average			0.33	−0.99

Panel B: Equity index futures				
SPI 200	−0.23	−0.49	0.25	−3.99
DAX	0.18	0.54	0.08	−0.49
IBEX 35	0.36	1.20	0.30	0.07
CAC 40	0.30	0.98	0.28	0.13
FTSE/MIB	0.70*	1.92	1.08	0.73
TOPIX	0.44*	1.69	0.57	0.59
AEX	0.43	1.22	0.58	−0.14
FTSE 100	−0.16	−0.48	0.12	−5.24
S&P 500	0.14	0.45	0.10	−1.78
Average			0.37	−1.12
Panel C: Government bond futures				
Three-year Australian	0.01	0.29	0.02	−0.27
Ten-year Australian	0.13	1.42	0.41	−1.74
Two-year European	0.15*	1.84	2.43	8.02**
Five-year European	0.09	1.11	0.57	−1.03
Ten-year European	0.17	1.25	0.40	−1.03
Thirty-year European	−0.06	−0.31	0.04	−0.74
Ten-year Canadian	−0.02	−0.17	0.01	−1.32
Ten-year Japanese	0.11	0.47	0.11	−0.54
Ten-year UK	0.10	1.06	0.18	−0.37
Two-year US	0.08***	3.57	3.24	4.26***
Five-year US	0.06	1.16	0.23	−0.23
Ten-year US	−0.03	−0.35	0.02	−0.39
Thirty-year US	0.17	0.60	0.15	−0.82
Average			0.60	0.29
Panel D: Currency forwards				
AUD/USD	−0.10	−0.47	0.08	−0.92
EUR/USD	0.17	1.08	0.29	−0.74
CAD/USD	0.14	1.18	0.42	−0.88
JPY/USD	0.33***	2.60	1.10	0.43*
NOK/USD	−0.01	−0.06	0.00	−0.59
NZD/USD	0.09	0.40	0.07	−0.94
SEK/USD	0.12	0.70	0.13	−0.82
CHF/USD	0.12	0.75	0.13	−1.42
GBP/USD	−0.05	−0.29	0.03	−0.82
Average			0.25	−0.75
Average across asset classes			0.39	−0.67
#(10% significance)	8			3

Table 3
p-value from the test that all assets have the same mean or Sharpe ratio

This table reports the *p*-value from the test that all assets have the same mean or Sharpe ratio. We perform four tests, including the analysis of variance (ANOVA) in Fisher (1918), Welch's ANOVA in Welch (1951), Kruskal-Wallis test in Kruskal and Wallis (1952), and bootstrap test. The sample period is 1985:01–2015:12.

	ANOVA	Welch's ANOVA	Kruskal-Wallis	Bootstrap
Mean	0.08	$< 10^{-3}$	$< 10^{-10}$	0
Sharpe ratio	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-15}$	0

Table 4
t-statistic of pooled regression without controlling for fixed effects

This table reports the t-statistic of pooled regression with real data and the simulated t-statistics with wild and pairs bootstrap. For each asset, we bootstrap a path with T observations and run pooled regression without controlling for fixed effects to calculate the t-statistic. We repeat this procedure one thousand times and obtain the distribution of the t-statistic for testing the null hypothesis that there is no time series momentum. The bootstrapped t-statistic is defined as the 97.5% percentile of the simulated t-statistics. The sample period is 1985:01–2015:12.

		Bootstrapped t-statistic			Bootstrapped t-statistic	
h	t-statistic	Wild	Pairs	t-statistic	Wild	Pairs
Panel A: Forecast with return lagged h months						
	$r^i_{t+1}/\sigma^i_t = \alpha_h + \beta_h r^i_{t-h+1}/\sigma^i_{t-h} + \varepsilon^i_{t+1}$			$r^i_{t+1}/\sigma^i_t = \alpha_h + \beta_h sign(r^i_{t-h+1}) + \varepsilon^i_{t+1}$		
1	3.11	9.26	3.63	2.90	8.18	3.41
2	1.31	4.98	1.98	1.62	4.44	2.31
3	2.89	8.61	3.45	2.83	6.84	3.45
4	0.24	2.46	1.06	1.20	2.12	1.99
5	−0.17	1.88	0.60	−0.34	1.83	0.54
6	0.97	4.18	1.71	1.58	3.62	2.28
7	−0.21	1.52	0.65	0.44	1.55	1.29
8	0.75	3.81	1.49	1.09	3.20	1.84
9	1.36	4.76	2.10	1.70	4.20	2.47
10	3.02	8.12	3.60	2.56	6.46	3.30
11	3.63	10.34	4.13	3.97	7.61	4.39
12	0.29	2.52	1.00	0.55	2.35	1.26
Panel B: Forecast with past h-month returns						
	$r^i_{t+1}/\sigma^i_t = \alpha_h + \beta_h r^i_{t-h,t}/\sigma^i_{t-1} + \varepsilon^i_{t+1}$			$r^i_{t+1}/\sigma^i_t = \alpha_h + \beta_h sign(r^i_{t-h,t}) + \varepsilon^i_{t+1}$		
1	3.11	9.26	3.63	2.90	8.18	3.41
2	2.92	9.46	3.46	3.07	8.32	3.61
3	3.74	11.45	4.22	4.15	10.20	4.61
4	3.49	10.71	3.97	4.57	9.49	4.96
5	3.11	9.58	3.63	4.24	8.85	4.72
6	3.29	9.65	3.80	3.88	8.88	4.39
7	3.03	9.30	3.62	3.93	8.31	4.40
8	3.05	9.44	3.62	3.74	8.33	4.24
9	3.38	9.85	3.95	4.44	8.99	4.78
10	3.94	11.38	4.46	5.27	10.22	5.63
11	4.64	13.12	5.08	5.71	11.73	6.05
12	4.34	12.53	4.83	5.14	11.05	5.53

Table 5
t-statistic of pooled regression within each asset class without controlling for fixed effects

This table reports the t-statistic of pooled regression with real data and the simulated t-statistics with wild and pairs bootstrap. For each asset, we bootstrap a path of T observations and run pooled regression without controlling for fixed effects to calculate the t-statistic. We repeat this procedure one thousand times and obtain the distribution of the t-statistic for testing the null hypothesis that there is no time series momentum. The bootstrapped t-statistic is defined as the 97.5% percentile of the simulated t-statistics. The sample period is 1985:01–2015:12.

	$r^i_{t+1}/\sigma^i_t = \alpha_h + \beta_h r^i_{t-h,t}/\sigma^i_{t-1} + \varepsilon^i_{t+1}$			$r^i_{t+1}/\sigma^i_t = \alpha_h + \beta_h sign(r^i_{t-h,t}) + \varepsilon^i_{t+1}$		
		Bootstrapped t-statistic			Bootstrapped t-statistic	
h	t-statistic	Wild	Pairs	t-statistic	Wild	Pairs
Panel A: Commodity futures						
1	1.74	5.09	2.49	1.66	4.86	2.40
3	2.32	6.28	2.97	2.95	5.82	3.52
6	2.98	7.08	3.65	2.89	6.59	3.54
12	3.46	8.20	4.13	4.20	7.43	4.68
Panel B: Equity index futures						
1	1.77	5.56	2.41	0.46	4.85	1.27
3	1.97	6.07	2.68	1.03	5.33	1.79
6	1.92	5.88	2.57	2.29	5.37	2.89
12	2.20	6.49	2.92	3.00	6.05	3.60
Panel C: Government bond futures						
1	2.35	6.58	3.04	2.04	5.78	2.75
3	2.37	6.68	2.99	2.87	5.75	3.41
6	0.60	3.22	1.53	0.73	2.93	1.61
12	1.68	5.39	2.44	1.69	4.77	2.42
Panel D: Currency forwards						
1	1.67	4.97	2.46	2.09	4.32	2.88
3	2.46	6.95	3.09	2.46	5.97	3.12
6	1.40	4.73	2.19	2.23	4.27	2.94
12	1.73	5.49	2.61	1.96	5.08	2.74

Table 6

t-statistic of pooled regression without volatility scaling and without controlling for fixed effects

This table reports the t-statistic of pooled regression with real data and the simulated t-statistics with wild and pairs bootstrap. For each asset, we bootstrap a path with T observations and run pooled regression without volatility scaling and without controlling for fixed effects to calculate the t-statistic. We repeat this procedure one thousand times and obtain the distribution of the t-statistic for testing the null hypothesis that there is no time series momentum. The bootstrapped t-statistic is defined as the 97.5% percentile of the simulated t-statistics. The sample period is 1985:01–2015:12.

		Bootstrapped t-statistic			Bootstrapped t-statistic	
h	t-statistic	Wild	Pairs	t-statistic	Wild	Pairs
Panel A: Forecast with return lagged h months						
	$r^i_{t+1} = \alpha_h + \beta_h r^i_{t-h+1} + \varepsilon^i_{t+1}$			$r^i_{t+1} = \alpha_h + \beta_h sign(r^i_{t-h+1}) + \varepsilon^i_{t+1}$		
1	1.80	5.49	2.51	2.20	6.13	2.85
2	0.52	2.58	1.47	1.65	2.67	2.45
3	1.43	4.57	2.19	1.84	4.58	2.58
4	0.67	3.21	1.58	1.47	3.21	2.26
5	−1.33	−0.10	−0.14	−0.89	−0.08	0.28
6	1.03	3.37	1.92	1.77	3.45	2.48
7	−1.21	−0.47	−0.18	−0.18	−0.50	0.77
8	−0.64	0.60	0.42	0.17	0.54	1.12
9	−0.97	0.23	0.28	0.22	0.19	1.30
10	2.52	6.11	3.21	2.72	5.87	3.42
11	5.04	9.88	5.30	5.17	9.89	5.51
12	−1.04	−0.17	0.08	−0.11	−0.06	0.85
Panel B: Forecast with past h-month returns						
	$r^i_{t+1} = \alpha_h + \beta_h r^i_{t-h,t} + \varepsilon^i_{t+1}$			$r^i_{t+1} = \alpha_h + \beta_h sign(r^i_{t-h,t}) + \varepsilon^i_{t+1}$		
1	1.80	5.49	2.51	2.20	6.13	2.85
2	1.39	4.56	2.21	2.57	5.10	3.18
3	1.71	5.26	2.45	3.06	5.81	3.62
4	1.82	5.30	2.59	3.75	5.94	4.25
5	1.27	4.27	2.09	3.23	4.75	3.77
6	1.55	4.85	2.39	2.71	5.38	3.32
7	1.04	3.89	1.90	2.54	4.26	3.19
8	0.78	3.49	1.58	2.24	3.64	2.94
9	0.62	3.12	1.50	2.57	3.31	3.29
10	1.08	3.75	1.96	3.62	4.23	4.14
11	2.09	5.61	2.84	4.17	6.41	4.72
12	1.68	4.96	2.50	3.72	5.64	4.16

Table 7

t-statistic of pooled regression without controlling for fixed effects over 1985:01–2009:12

This table reports the t-statistic of pooled regression with real data and the bootstrapped t-statistics with wild and pairs bootstrap. For each asset, we bootstrap a path with T observations and run pooled regression without controlling for fixed effects to calculate the t-statistic. We repeat this procedure one thousand times and obtain the distribution of the t-statistic for testing the null hypothesis that there is no time series momentum. The bootstrapped t-statistic is defined as the 97.5% percentile of the simulated t-statistics.

		Bootstrapped t-statistic			Bootstrapped t-statistic	
h	t-statistic	Wild	Pairs	t-statistic	Wild	Pairs
Panel A: Forecast with return lagged h months						
	$r^i_{t+1}/\sigma^i_t = \alpha_h + \beta_h r^i_{t-h+1}/\sigma^i_{t-h} + \varepsilon^i_{t+1}$			$r^i_{t+1}/\sigma^i_t = \alpha_h + \beta_h sign(r^i_{t-h+1}) + \varepsilon^i_{t+1}$		
1	3.71	10.68	4.20	3.75	9.31	4.19
2	0.97	4.07	1.68	1.34	3.65	2.02
3	2.48	7.43	3.11	2.44	6.09	3.01
4	0.22	2.40	1.14	0.65	2.28	1.59
5	−0.15	1.53	0.67	−0.38	1.56	0.66
6	0.52	3.08	1.30	1.35	2.78	2.15
7	0.39	3.07	1.24	0.95	2.74	1.84
8	0.59	3.32	1.37	1.20	2.84	2.06
9	1.68	5.26	2.42	1.96	4.59	2.66
10	2.70	7.37	3.32	2.11	5.83	2.84
11	3.70	10.37	4.23	4.04	7.61	4.54
12	0.37	2.74	1.14	0.54	2.39	1.37
Panel B: Forecast with past h-month returns						
	$r^i_{t+1}/\sigma^i_t = \alpha_h + \beta_h r^i_{t-h,t}/\sigma^i_{t-1} + \varepsilon^i_{t+1}$			$r^i_{t+1}/\sigma^i_t = \alpha_h + \beta_h sign(r^i_{t-h,t}) + \varepsilon^i_{t+1}$		
1	3.71	10.68	4.20	3.75	9.31	4.19
2	3.09	9.53	3.54	3.19	8.39	3.70
3	3.74	11.53	4.27	4.43	9.96	4.94
4	3.45	10.37	3.98	4.78	9.19	5.19
5	3.06	9.27	3.63	4.36	8.39	4.85
6	3.17	9.31	3.73	4.03	8.52	4.46
7	3.05	9.09	3.60	4.19	8.32	4.62
8	3.12	9.06	3.69	4.13	8.11	4.58
9	3.59	10.31	4.13	4.70	9.27	5.13
10	4.00	11.68	4.54	5.46	10.44	5.85
11	4.69	13.16	5.14	5.64	11.77	6.07
12	4.48	12.76	4.96	5.24	11.17	5.62

Table 8
t-statistic of pooled regression controlling for fixed effects

This table reports the t-statistic of pooled regression with real data and the bootstrapped t-statistics with wild and pairs bootstrap. For each asset, we bootstrap a path with T observations and run pooled regression controlling for fixed effects to calculate the t-statistic. We repeat this procedure one thousand times and obtain the distribution of the t-statistic for testing the null hypothesis that there is no time series momentum. The bootstrapped t-statistic is defined as the 97.5% percentile of the simulated t-statistics. The sample period is 1985:01–2015:12.

		Bootstrapped t-statistic			Bootstrapped t-statistic	
h	t-statistic	Wild	Pairs	t-statistic	Wild	Pairs
Panel A: Forecast with return lagged h months						
	$r^i_{t+1}/\sigma^i_t = \alpha^i_h + \beta_h r^i_{t-h+1}/\sigma^i_{t-h} + \varepsilon^i_{t+1}$			$r^i_{t+1}/\sigma^i_t = \alpha^i_h + \beta_h sign(r^i_{t-h+1}) + \varepsilon^i_{t+1}$		
1	2.80	8.51	3.39	2.66	7.60	3.19
2	0.96	4.17	1.66	0.94	3.85	1.67
3	2.53	7.77	3.12	2.17	6.36	2.83
4	−0.19	1.56	0.70	0.36	1.56	1.27
5	−0.56	1.00	0.25	−0.94	1.03	0.02
6	0.58	3.26	1.36	1.07	2.90	1.79
7	−0.62	0.59	0.27	0.20	1.01	1.04
8	0.37	2.90	1.14	0.80	2.64	1.53
9	0.94	3.84	1.73	0.94	3.53	1.79
10	2.57	7.21	3.22	1.87	5.71	2.61
11	3.22	9.40	3.75	3.53	7.03	4.17
12	−0.12	1.63	0.65	0.37	1.70	1.13
Panel B: Forecast with past h-month returns						
	$r^i_{t+1}/\sigma^i_t = \alpha^i_h + \beta_h r^i_{t-h,t}/\sigma^i_{t-1} + \varepsilon^i_{t+1}$			$r^i_{t+1}/\sigma^i_t = \alpha^i_h + \beta_h sign(r^i_{t-h,t}) + \varepsilon^i_{t+1}$		
1	2.80	8.51	3.39	2.66	7.60	3.19
2	2.51	8.43	3.07	2.62	7.41	3.19
3	3.23	10.17	3.74	3.56	9.08	4.17
4	2.89	9.24	3.46	3.60	8.36	4.11
5	2.44	7.89	2.99	3.17	7.45	3.66
6	2.53	7.97	3.12	3.15	7.27	3.70
7	2.22	7.32	2.86	2.97	6.86	3.43
8	2.19	7.35	2.80	2.55	6.67	3.24
9	2.49	7.75	3.09	3.43	7.19	3.92
10	3.00	9.23	3.58	3.94	8.34	4.38
11	3.68	10.80	4.20	4.49	9.71	4.94
12	3.37	10.13	3.93	4.04	9.14	4.53

Table 9

Time series momentum (TSM) versus time series history (TSH) at the asset level

This table reports the mean returns and Sharpe ratios of the time series momentum and time series history strategies, as well as their difference, on the basis of individual assets. TSM refers to the strategy that buys the future contract if its past 12-month return is non-negative and sells it if its past 12-month return is negative, and TSH refers to the strategy that buys the futures contract if its historical sample mean is non-negative and sells it if its historical sample mean is negative. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. The investment period is 1986:01–2015:12.

Asset	TSM return	TSH return	TSM Sharpe ratio	TSH Sharpe ratio	Return difference	*p*-value of return difference	Sharpe ratio difference	*p*-value of Sharpe ratio difference
Aluminum	0.27	−0.47	0.05	−0.08	0.74**	0.04	0.13**	0.04
Brent oil	0.80	0.32	0.09	0.04	0.48	0.44	0.05	0.44
Cattle	0.28	0.08	0.07	0.02	0.20	0.48	0.05	0.48
Cocoa	−0.46	0.13	−0.06	0.02	−0.59	0.32	−0.08	0.31
Coffee	0.18	−0.55	0.02	−0.05	0.73	0.30	0.07	0.30
Copper	0.77	0.94	0.10	0.12	−0.17	0.74	−0.02	0.73
Corn	0.11	0.14	0.01	0.02	−0.04	0.94	−0.01	0.94
Cotton	1.04	−0.07	0.14	−0.01	1.11**	0.04	0.15**	0.04
Crude	1.07	0.21	0.11	0.02	0.86	0.19	0.09	0.19
Gas oil	0.98	0.53	0.10	0.05	0.45	0.49	0.05	0.49
Gold	0.55	0.05	0.12	0.01	0.50*	0.10	0.11*	0.10
Heating oil	1.09	0.30	0.12	0.03	0.79	0.21	0.09	0.21
Hogs	0.29	−0.17	0.04	−0.02	0.46	0.37	0.06	0.37
Natural gas	1.26	0.05	0.09	0.01	1.21	0.25	0.08	0.25
Nickel	0.72	0.43	0.07	0.04	0.29	0.70	0.03	0.70
Platinum	0.30	0.53	0.05	0.09	−0.23	0.61	−0.04	0.61
Silver	0.33	−0.09	0.04	−0.01	0.42	0.46	0.05	0.47
Soybean	−0.15	0.02	−0.02	0.01	−0.17	0.68	−0.03	0.67
Soymeal	0.20	0.51	0.02	0.06	−0.31	0.58	−0.04	0.58
Soy oil	0.42	0.14	0.06	0.02	0.28	0.57	0.04	0.57
Sugar	0.05	0.49	0.01	0.05	−0.44	0.50	−0.04	0.50
Unleaded gasoline	1.00	0.83	0.10	0.08	0.17	0.77	0.02	0.77
Wheat	0.38	0.26	0.05	0.03	0.12	0.81	0.02	0.81
Zinc	0.67	−0.22	0.09	−0.03	0.89*	0.08	0.12*	0.08

SPI 200	0.18	0.55	0.04	0.12	−0.37	0.17	−0.08	0.17
DAX	0.79	0.68	0.13	0.11	0.11	0.80	0.02	0.79
IBEX 35	0.71	0.72	0.11	0.11	−0.01	0.98	0.00	0.98
CAC 40	0.43	0.49	0.08	0.09	−0.06	0.89	−0.01	0.88
FTSE/MIB	0.90	0.36	0.14	0.05	0.54	0.27	0.09	0.27
TOPIX	0.84	0.25	0.15	0.04	0.59	0.18	0.11	0.18
AEX	0.73	0.55	0.13	0.10	0.18	0.66	0.03	0.66
FTSE 100	0.27	0.52	0.06	0.11	−0.25	0.40	−0.05	0.40
S&P 500	0.67	0.73	0.15	0.16	−0.06	0.84	−0.01	0.83
Three-year Australian bond	0.24	0.28	0.24	0.28	−0.04	0.39	−0.04	0.37
Ten-year Australian bond	0.32	0.42	0.15	0.21	−0.10	0.27	−0.06	0.26
Two-year European bond	0.13	0.10	0.13	0.10	0.03	0.57	0.03	0.56
Five-year European bond	0.08	0.13	0.06	0.11	−0.05	0.46	−0.05	0.45
Ten-year European bond	0.05	0.25	0.02	0.09	−0.20	0.18	−0.07	0.18
Thirty-year European bond	0.14	0.56	0.05	0.19	−0.42***	0.01	−0.14***	0.01
Ten-year Canadian bond	0.34	0.52	0.11	0.17	−0.18	0.32	−0.06	0.31
Ten-year Japanese bond	0.21	0.08	0.06	0.02	0.13	0.55	0.04	0.56
Ten-year UK bond	0.34	0.28	0.14	0.12	0.06	0.68	0.02	0.68
Two-year US bond	0.13	0.12	0.28	0.26	0.01	0.65	0.02	0.63
Five-year US bond	0.19	0.23	0.15	0.19	−0.04	0.41	−0.04	0.40
Ten-year US bond	0.19	0.28	0.09	0.13	−0.09	0.40	−0.04	0.40
Thirty-year US bond	0.54	0.89	0.12	0.20	−0.35*	0.07	−0.08*	0.06
AUD/USD	0.06	−0.16	0.02	−0.05	0.22	0.37	0.07	0.37
EUR/USD	0.11	0.11	0.03	0.03	0.00	1.00	0.00	1.00
CAD/USD	0.23	−0.08	0.11	−0.04	0.31**	0.05	0.15**	0.05
JPY/USD	0.46	0.09	0.14	0.03	0.37	0.11	0.11	0.11
NOK/USD	0.06	−0.06	0.02	−0.02	0.12	0.58	0.04	0.58
NZD/USD	0.24	0.02	0.07	0.01	0.22	0.38	0.06	0.38
SEK/USD	0.04	−0.05	0.01	−0.02	0.09	0.70	0.03	0.70
CHF/USD	0.18	0.19	0.05	0.06	−0.01	0.96	−0.01	0.97
GBP/USD	0.00	−0.03	0.00	−0.01	0.03	0.87	0.01	0.87
#(significance)					7		7	

Table 10

Time series momentum (TSM) versus time series history (TSH) at the portfolio level

This table reports the average and risk-adjusted returns of the TSM and TSH strategies, where we restrict portfolio weights on individual assets to be the same for comparison when constructing these two strategies. TSM refers to the strategy that buys futures contracts with non-negative past 12-month return and sells futures contracts with negative past 12-month return, and TSH refers to the strategy that buys futures contracts with non-negative historical sample mean and sells futures contracts with negative historical sample mean. The benchmarks are the Fama-French four-factor model that includes MSCI World Index, SMB (small minus big), HML (high minus low), and UMD (momentum), and the Asness, Moskowitz and Pedersen (2013) three-factor model. Newey-West t-statistics and p-values are reported in parentheses and brackets, respectively. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. The investment period is 1986:01–2015:12.

		Fama-French four-factor model						Asness-Moskowitz-Pedersen three-factor model				
	Mean	Alpha	MSCI World Index	SMB	HML	UMD	R^2	Alpha	MSCI World Index	Value everywhere	Momentum everywhere	R^2
Panel A: Equal weighting, i.e., portfolio weight $= \frac{1}{N}$												
TSM strategy												
Long leg	0.34***	0.12**	0.16***	0.05**	0.09***	0.32***	42.57%	0.09	0.16***	0.14***	0.38***	41.92%
	(4.92)	(2.29)	(7.73)	(2.24)	(2.62)	(7.19)		(1.60)	(7.21)	(2.99)	(7.33)	
Short leg	−0.05	−0.03	0.14***	0.11***	0.03	−0.28***	48.31%	0.02	0.13***	−0.09	−0.32***	45.85%
	(−0.72)	(−0.46)	(5.30)	(4.01)	(0.96)	(−8.34)		(0.23)	(5.08)	(−1.42)	(−6.89)	
Long - short	0.39***	0.15*	0.02	−0.06*	0.06	0.60***	46.03%	0.07	0.03	0.23**	0.70***	47.39%
	(4.73)	(1.94)	(0.61)	(−1.83)	(1.01)	(9.99)		(1.01)	(0.93)	(2.50)	(8.79)	
TSH strategy												
Long leg	0.27***	0.07	0.28***	0.14***	0.10***	0.08*	49.07%	0.10	0.26***	0.01	0.08*	45.30%
	(2.56)	(0.95)	(8.79)	(4.41)	(3.90)	(1.93)		(1.13)	(7.73)	(0.24)	(1.65)	
Short leg	0.02	0.02	0.03***	0.03*	0.02	−0.04**	8.73%	0.01	0.03***	0.03	−0.03	8.04%
	(0.61)	(0.52)	(3.51)	(1.83)	(1.47)	(−2.01)		(0.25)	(3.17)	(1.15)	(−1.04)	
Long - short	0.25***	0.05	0.25***	0.11***	0.08***	0.13***	44.83%	0.09	0.23***	−0.02	0.11*	42.01%
	(2.70)	(0.80)	(8.54)	(3.70)	(2.96)	(2.76)		(1.14)	(7.59)	(−0.29)	(1.94)	
TSM versus TSH												
Mean difference	0.14											
	[0.19]											
Alpha difference		0.10						−0.02				
		[0.26]						[0.84]				

Panel B: Volatility weighting, i.e., portfolio weight $= \frac{1}{N}\frac{40\%}{\sigma_t^i}$												
TSM strategy												
Long leg	1.02***	0.58***	0.28***	−0.04	0.12	0.68***	36.04%	0.48***	0.28***	0.36***	0.84***	37.46%
	(7.66)	(5.31)	(7.34)	(−0.67)	(1.85)	(8.02)		(4.29)	(7.96)	(3.21)	(8.41)	
Short leg	−0.14	−0.06	0.22***	0.19***	0.04	−0.52***	44.46%	0.02	0.20***	−0.17	−0.59***	42.42%
	(−1.07)	(−0.53)	(5.75)	(3.06)	(0.79)	(−8.67)		(0.13)	(5.52)	(−1.48)	(−9.19)	
Long - short	1.16***	0.64***	0.05	−0.22**	0.08	1.19***	40.62%	0.46***	0.08	0.53***	1.43***	41.58%
	(6.31)	(3.68)	(0.80)	(−2.24)	(0.85)	(10.31)		(2.90)	(1.25)	(2.75)	(10.81)	
TSH strategy												
Long leg	0.81***	0.42***	0.48***	0.12**	0.16**	0.25***	40.10%	0.43***	0.46***	0.13	0.29***	38.80%
	(4.72)	(3.06)	(12.17)	(2.29)	(2.48)	(2.97)		(2.71)	(10.96)	(1.03)	(3.05)	
Short leg	0.07	0.10*	0.02	0.03	−0.01	−0.08***	4.27%	0.07	0.02	0.06	−0.04	4.35%
	(1.50)	(1.93)	(1.28)	(1.33)	(−0.11)	(−2.73)		(1.39)	(1.42)	(1.07)	(−1.06)	
Long - short	0.74***	0.32**	0.46***	0.09	0.16**	0.33***	36.23%	0.36**	0.44***	0.06	0.33***	35.01%
	(4.36)	(2.18)	(9.81)	(1.46)	(2.17)	(3.76)		(2.25)	(8.85)	(0.43)	(3.21)	
TSM versus TSH												
Mean difference	0.42*											
	[0.07]											
Alpha difference		0.32*						0.10				
		[0.08]						[0.56]				

Panel C: Past 12-month return weighting, i.e, portfolio weight = $\frac{r^i_{t-12,t}}{\sum_{i=1}^{N} \lvert r^i_{t-12,t} \rvert}$												
TSM strategy												
Long leg	0.50***	0.08	0.27***	0.17***	0.14**	0.66***	32.83%	0.06	0.25***	0.15*	0.73***	30.95%
	(3.54)	(0.72)	(5.85)	(3.62)	(2.06)	(8.65)		(0.52)	(5.09)	(1.65)	(7.71)	
Short leg	−0.07	−0.01	0.19***	0.16***	0.01	−0.43***	39.93%	0.05	0.18***	−0.15	−0.49***	38.05%
	(−0.65)	(−0.08)	(5.06)	(3.61)	(0.18)	(−7.74)		(0.49)	(5.06)	(−1.36)	(−6.97)	
Long - short	0.57***	0.09	0.08	0.01	0.13	1.09***	40.17%	0.01	0.08	0.30**	1.22***	40.55%
	(3.79)	(0.69)	(1.48)	(0.14)	(1.35)	(12.55)		(0.08)	(1.57)	(2.04)	(11.33)	
TSH strategy												
Long leg	0.30*	−0.03	0.43***	0.23***	0.14***	0.19***	48.23%	0.03	0.41***	−0.03	0.17**	44.26%
	(1.78)	(−0.23)	(8.10)	(4.71)	(3.22)	(2.77)		(0.21)	(6.70)	(−0.29)	(1.96)	
Short leg	0.02	0.02	0.02***	0.01	0.01	−0.04*	5.23%	0.01	0.02***	0.03	−0.03	5.18%
	(0.77)	(0.83)	(2.82)	(1.09)	(1.06)	(−1.70)		(0.50)	(2.64)	(1.08)	(−0.86)	
Long - short	0.28*	−0.05	0.41***	0.22***	0.13***	0.23***	46.18%	0.02	0.39***	−0.06	0.20**	42.65%
	(1.71)	(−0.42)	(7.87)	(4.39)	(2.93)	(3.10)		(0.10)	(6.85)	(−0.56)	(1.99)	
TSM versus TSH												
Mean difference	0.29											
	[0.12]											
Alpha difference		0.14						−0.01				
		[0.34]						[0.98]				

Panel D: Zero investment, i.e., long = $\frac{1}{N^{\text{buy}}}$ and short = $\frac{1}{N^{\text{sell}}}$												
TSM strategy												
Long leg	0.60***	0.24***	0.25***	0.08**	0.11**	0.53***	39.74%	0.20**	0.24***	0.17**	0.61***	39.10%
	(5.29)	(2.65)	(6.93)	(2.02)	(2.26)	(7.60)		(2.14)	(6.41)	(2.19)	(7.17)	
Short leg	−0.12	−0.06	0.28***	0.23***	0.05	−0.56***	42.77%	0.02	0.26***	−0.17	−0.64***	40.27%
	(−0.74)	(−0.42)	(5.76)	(3.94)	(0.79)	(−7.40)		(0.10)	(5.33)	(−1.24)	(−6.73)	
Long - short	0.72***	0.30**	−0.03	−0.16**	0.06	1.10***	45.91%	0.18	−0.02	0.34**	1.25***	46.18%
	(4.32)	(1.96)	(−0.65)	(−2.43)	(0.69)	(10.26)		(1.25)	(−0.40)	(1.97)	(8.52)	
TSH strategy												
Long leg	0.35***	0.10	0.35***	0.17***	0.12***	0.10*	49.49%	0.13	0.33***	0.02	0.10	45.78%
	(2.64)	(1.06)	(8.98)	(4.39)	(3.85)	(1.90)		(1.25)	(7.92)	(0.21)	(1.56)	
Short leg	0.08	0.04	0.15***	0.14**	0.09	−0.18*	8.16%	−0.01	0.14***	0.17	−0.09	7.37%
	(0.51)	(0.30)	(3.66)	(2.13)	(1.62)	(−1.86)		(−0.05)	(3.28)	(1.48)	(0.75)	
Long - short	0.27**	0.06	0.20***	0.03	0.03	0.28***	11.89%	0.14	0.19***	−0.16	0.19	12.27%
	(2.10)	(0.40)	(5.30)	(0.45)	(0.49)	(2.75)		(1.02)	(4.83)	(−1.24)	(1.44)	
TSM versus TSH												
Mean difference	0.45**											
	[0.03]											
Alpha difference		0.24						0.04				
		[0.13]						[0.76]				

Table 11

Time series momentum (TSM) and time series history (TSH) forecast comparison

This table reports the results of regressing r^i_{t+1} on the expected return ($\hat{r}^{\text{TSM},i}_{t+1}$) estimated at time t with the TSM pooled regression Eq. (3) and regressing $\hat{r}^{\text{TSM},i}_{t+1}$ on the expected return ($\hat{r}^{\text{TSH},i}_{t+1}$) estimated with the TSH approach (i.e., historical sample mean). ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively.

	$r^i_{t+1} = \alpha + \beta \hat{r}^{\text{TSM},i}_{t+1} + \varepsilon^i_{t+1}$			$\hat{r}^{\text{TSM},i}_{t+1} = d\hat{r}^{\text{TSH},i}_{t+1} + u^i_t$		
Asset class	β	t-statistic	R^2	d	t-statistic	R^2
Panel A: $\hat{r}^{\text{TSM},i}_{t+1}$ is estimated with volatility scaling						
Overall	0.19	0.61	0.04	1.09***	18.56	40.33
Commodity	0.15	0.42	0.02	1.24***	11.62	23.53
Equity	0.07	0.10	0.01	0.84***	14.90	45.06
Bond	0.23	0.60	0.08	0.99***	68.75	92.27
Currency	−0.08	−0.12	0.01	1.01***	14.95	4.45
Panel B: $\hat{r}^{\text{TSM},i}_{t+1}$ is estimated without volatility scaling						
Overall	0.30	0.45	0.03	1.04***	41.89	54.96
Commodity	0.09	0.11	0.00	1.01***	26.53	37.65
Equity	−0.37	−0.32	0.07	0.93***	35.34	77.93
Bond	−0.49	−0.52	0.07	1.00***	72.40	91.38
Currency	0.03	0.03	0.00	1.64***	19.78	14.32

When and where does the market notice the inequality?

Evidence from the mass shooting*

Jiangyuan Li
Singapore Management University

Monday 9th March, 2020

*I thank for the comments from Dashan Huang, Jun Tu, Weikai Li, William Xiang Xia, Yilin He, Li Guo, Yubo Tao, David Xiaoyu Xu, Zentao Zou and discussion at Point72 Cubist Systematic Strategies.

Send correspondence to Jiangyuan LI, e-mail: jiangyuanli.2015@pbs.smu.edu.sg; phone: +65 83441406

When does market notice the inequality?
Evidence from mass shooting

Abstract

I investigate the cross-sectional inequality-return relationship in the equity market from a geographical angle. I document that the negative inequality-return relationship still holds in cross-sectional level. Additionally, the negative inequality-return relationship is largely driven by the firms that locate on mass shooting treated regions and mass shooting periods. I attribute the empirical fact to attention channel that after observing long-term cooccurrence between inequality increase and mass shootings, public attention to inequality issues would be spurred when treated by a mass shooting.

Keywords : Inequality, Mass shooting, Attention, Equity market

JEL Classification: D31, G12, Z13

We'll see fewer mass shootings by taking care of income inequality.

- Mark Cuban, 4th June, 2019

1 Introduction

The paper investigates the cross-sectional relationship between economic inequality and equity return. The question has been discussed as early as Fisher (1910). He argues that the "enterpriser-borrower" (the rich) becomes richer, the demand for risky assets increases as well. Consequently, risky assets would be more volatile.

With the abundant data source nowadays, one consensus the scholars have just reached is that inequality-return relationship should be negative. Campbell, Delikouras, Jiang, and Korniotis (2016) argue that the top 1% earners' income growth is one priced risk factor and would compensate part of the value premium. Toda and Walsh (2019) argue that when rich get richer, their increasing demand for the risky asset would lead overpricing, which means a negative subsequent return. Both two works mentioned above focus on the aggregate market. However, still water runs deep. At the cross-sectional level, we still know very little about when and where the inequality information is actually priced into the equity market.

As economic inequality is largely a geographical measure (Chetty, Hendren, Kline, and Saez 2014a), one natural question is if the inequality-return relationship is also geographically conditional. With this question in mind, I intend to investigate the geographical inequality-return relationship. However, from aggregate to cross-sectional, there are countless different criteria to set the cross-sectional group. One can not arbitrarily choose one criterion to group observation cross-sectionally with statistical evidence but without valid economic motivation.

My motivation for cross-sectional group emerges from the recent interview of billionaire Mark Cuban by Yahoo Finance on 4th June 2019, in which he connects two timely topics inequality and mass shooting. He argues that one remedy to control for a mass shooting is by mitigating the economic inequality gap. As Piketty and Saez (2003) document, economic inequality in the U.S. experience a dramatic increase in the post-war era. At the same time, the mass shooting in the U.S. also experiences a severe increase. As the Los Angeles Post describes that the mass shootings "aren't just increasing, theyre getting deadlier". One recent sociology study Kwon and Cabrera (2019) point out that inequality and mass shooting is statistically correlated cross countries in the U.S.

The empirical fact reminds the financial economists if one could observe a conditional inequality-return relationship based on the information of mass shooting. Next, if there exists a conditional inequality-return relationship, what is the channel.

To answer these questions, I need to capture geographical inequality. I mainly employ the recent contribution in the vein of inequality study by Toda and Walsh (2019), the capital gain ratio (KGR) as the inequality measure. There are several main reasons to employ KGR as the inequality measure. First, in the economic sense, KGR captures the difference between the top 1% earner's income share with or without realized capital gain scaled by the bottom 99% income share. Because the top earners are also more likely to be the capital holder, the KGR should both capture the capital wealth and income inequality. Second, in an econometric sense, most inequality measures like the Gini index or top 1% earner's income share are persistent, which could lead to serious spurious regression issues (Granger 1981). As argued by Toda and Walsh (2019), top income share inclusive or exclusive of capital gain has a very similar trend. Differentiating these two series would cancel out the common trend and generate a stationary series. Third, the data from Piketty and Saez (2003) and World Inequality Database would allow for the construction of state-level inequality measures, which makes the geographical cross-sectional analysis possible.

Next, I need to pin down the time and location of mass shootings. I obtain the mass shooting record mainly from Mass Shootings in America project from Stanford Geospatial Center and Stanford Libraries and USA Today mass shooting list. However, not all of the records could be regarded as mass shootings. The reason is that the traditional concept of mass shooting mainly means for the shootings that are unpredictable and without any specific targets. Usually, the consequence of the shootings should be "mass" enough as well. Thus, in the baseline analysis, I mainly use the mass shooting records with at least 3 casualties and is not caused by an economic or romantic reason. For each mass shooting, the database also provides very specific location information for the county and Metropolitan Statistical Area (MSA). For firm-level location information, I follow the argument in Chaney, Sraer, and Thesmar (2012) that the firm's headquarter location represents great friction of the overall firm's assets location, thus headquarter location is a proper proxy for firm's location.

I first document an empirical finding that at the geographical cross-sectional level, the negative inequality-return relationship remains, which supports the aggregate market analysis in Campbell, Delikouras, Jiang, and Korniotis (2016) and Toda and Walsh (2019). The mass shooting also negatively contributes to the contemporaneous return. The novelty of this study emerges from the conditional analysis. By the panel analysis method from Hirshleifer and Shumway (2003) and Edmans, Garcia, and Norli (2007), I find the inequality and mass shooting interactively explain the negative subsequent return. After controlling for the interactive term, the inequality measure and mass shooting variable do not present significance. The direct interpretation is that inequality innovation is largely priced into the stock market while there is a mass shooting.

The next in-depth subsample analysis illustrates even more puzzling results. If I separate the sample into mass shooting treated region and non-mass shooting treated region. The negative inequality-return relationship only exists in mass shooting treated regions. The result indicates that potentially, the mass shooting treated region could be different from the mass shooting untreated region. However, it could be either the following cases. Mass shooting per se changes the economic channel between inequality and subsequent return, or mass shooting is just a proxy for some other channel.

I examine the question in two ways. First, to test if the mass shooting is just a proxy for some unobservable channel, I compare two region subsample. One is the region that has ever been treated by a mass shooting. The other one is the region just adjacent to the treated region. The reason is that these two subsamples should share as much same geographical and economic conditions as each other. The only difference is the mass shooting treatment condition. Thus, it is highly likely that the different inequality-return relationship should be due to mass shooting per se.

Next, I follow the hint from Mark Cuban's interview. If he intuitively connects the mass shooting with inequality, is that possible it is a general pattern, that the public would also pay more attention to inequality when there is one mass shooting? When attention concentrate on these regions, the firm that locates in these regions would also be more sensitive in the inequality return relationship.

As early as Pavlov and Gantt (1928), scholars have noticed the widespread reflex reaction in many aspects of animal behaviors. Recent psychological studies like Lang, Bradley, and Cuthbert (1990) and Lang et al. (1997) also document that human beings' attention also exhibits the reflex reaction pattern. If one frequently observes two concrete or abstract subjects appear at the same time. Her attention to one subject would be spurred when the other subject appears.

To test this attention channel, I employ the google trends data at the state level to capture public attention. By a standard difference in difference method, I document that the search volume of the keyword "income inequality" or "wealth inequality" significantly increases if there is a mass shooting happens in this state.

This study contributes to the on-going discussion about economic inequality and its connection to the asset market. Piketty and Saez (2003) document the upward sloping inequality trend in U.S. Toda and Walsh (2019) and Campbell, Delikouras, Jiang, and Korniotis (2016) extend the inequality to the equity market at the aggregate level. I investigate the cross-sectional inequality-return relationship and document the inequality-return relationship mainly holds in mass shooting treated regions.

The rest of the paper proceeds as follows. Section II describes the inequality, mass shooting, and geographical data. Section III describes the background of mass shooting and inequality, and briefly discuss their statistical connection. Section IV tests the inequality-return relationship that conditions on mass shooting treatment. Section V presents the robustness of the results. Section VI and VII discuss the channel, and Section VIII concludes.

2 Data

In the paper, the data source is fivefold, inequality, mass shooting, geographical definition, public attention, and equity return. I describe the data properties and measure constructions in this section.

2.1 Inequality

The economic implication of inequality is one major research subject. In their well-cited work, Piketty and Saez (2003) employ the tax return to inference the top income earners' share in the total population. The following works in the vain of inequality mainly keep a similar method or similar concept in general to measure the inequality. David, Katz, and Kearney (2006) study the polarization of the US labor market. Chetty, Hendren, Kline, Saez, and Turner (2014b) study the mobility of the United States.

However, the data of inequality like top earner's income share are generally very persistent, which will induce the spurious regression problem Granger (1981). To employ the inequality information, one needs to translate the persistent independent variable into a stationary variable. In one recently progress, Toda and Walsh (2019) resolve this problem by capturing the evolution of the capital gain ratio (KGR, henceforth). Before explaining the intuition, I define the measure and show how to achieve the measure,

$$\text{KGR}(1) = \frac{\text{top}(1) - \text{top}(1)^{excg}}{1 - \text{top}(1)}, \tag{1}$$

where the KGR is the stationary inequality measure capital gain ratio, $\text{top}(1)$ is the top 1% earners' income share in overall population and $\text{top}(1)^{excg}$ is the same concept but excluding capital gain. I repeat the brief algebra just below for the completeness of the paper. One should turn to Toda and Walsh (2019) for a detailed technical reference.

Assume two types of agents exist in the economy, that is, the top 1% earner and the rest (bottom 99%). I denote the type of them by $i = A, B$. The Y_i^k, and Y_i^l represent the total capital and labor income, respectively. Accordingly, the top 1% type agent's share could be represented by

$$\text{top}(1) = \frac{Y_A}{Y} = \frac{Y_A^k + Y_A^l}{Y_A^k + Y_A^l + Y_B^k + Y_B^l}, \tag{2}$$

Next, I define ρ_i as the proportion of type i agent's realized capital gains in her total capital income. The top income share excluding realized capital gains is

$$\text{top}(1)^{excg} = \frac{(1-\rho_A)Y_A^k + Y_A^l}{(1-\rho_A)Y_A^k + Y_A^l + (1-\rho_B)Y_B^k + Y_B^l} = f(\rho_A, \rho_B). \tag{3}$$

Apply the Taylor's expansion, the difference between $\text{top}(1)$ and $\text{top}(1)^{excg}$ could be approximated as

$$\begin{aligned} \text{top}(1) - \text{top}(1)^{excg} &= f(0,0) - f(\rho_A, \rho_B) \\ &\approx -\rho_A \frac{\partial f}{\partial \rho_A}(0,0) - \rho_B \frac{\partial f}{\partial \rho_B}(0,0) \\ &= \rho_A \frac{Y_A^k Y_B}{Y^2} - \rho_B \frac{Y_B^k Y_A}{Y^2} \end{aligned} \tag{4}$$

For brevity, I denote the capital income share in aggregate income, so $Y^k = \alpha Y$, and write the Eq. (4),

$$\begin{aligned} \text{top}(1) - \text{top}(1)^{excg} &\approx \alpha \left(\rho_A \frac{Y_A^k}{Y} (1 - \text{top}(1)) - \rho_B \left(1 - \frac{Y_A^k}{Y^k} \right) \text{top}(1) \right) \\ &\approx \alpha \rho_A \frac{Y_A^k}{Y} (1 - \text{top}(1)). \end{aligned} \tag{5}$$

The last step holds because the second term is ignored. The reason to ignore the second term are 1) top(1) usually resides between 0.1 to 0.2 since the 1960s, 2) the top 1% is more likely to be the capital owners, thus the scale of Y_A^k/Y^k is at least that of $1 - Y_A^k/Y^k$. After some simple rearrangements, I obtain

$$\text{KGR}(1) = \frac{\text{top}(1) - \text{top}(1)^{excg}}{1 - \text{top}(1)} \approx \alpha_A \rho_A \frac{Y_A^k}{Y^k}. \tag{6}$$

I first address the issue of persistence variable. In Toda and Walsh (2019), they motivate to make the difference between top(1) and $\text{top}(1)^{excg}$ because they share "common trends" from plot. Thus, the method to address the persistence issue is in line with the class concept of differentiating two persistent variables. I next explain the economic intuition. The difference between the two measures is mainly driven by the flow of realized capital gain. Given the fact that wealth is more likely to be the capital holder Piketty and Saez (2003), KGR captures the wealth and income inequality from the capital. I present the summary statistic and the stationary test result of each state's KGR in Table 1. In summary, all of the states' KGR measure is stationary.

[Insert Table 1 around here]

2.2 Mass shooting

The second type of data is about the mass shooting. In the United States, there are around 30000 gun deaths each year, but only less than 100 of them pass away in a mass shooting (Luca, Malhotra, and Poliquin, 2019). The reason is that not all of the gun violence could be defined as a mass shooting. To single out the mass shootings from the general gun violence, I combine several different sources of mass shooting database from academia or media since 1966. Specifically, I use Mass Shootings in America project from the Stanford Geospatial Center and Stanford Libraries and USA Today mass shooting list. After compiling different database information together, I refine the mass shooting list according to the following definition by manual check. In this study, I follow Luca, Malhotra, and Poliquin (2019) and define a mass shooting as an incident with 3 or more casualties, other than the killer, are illegally influenced. Additionally, I also remove cases that shooters conduct the crime due to an economic or romantic relationship with the victim. The overarching criteria are to ensure the definition goes in line with the first impression of the mass shooting that mass shooting is a kind of plausibly random and salient public event.

2.3 Geographical definition

For any event, its influence is not unlimited. Mass shootings should follow the same pattern. Coval and Moskowitz (2001) shows that the fund manager could obtain information advantage due to geographical proximity. Applying a similar intuition, I assume that a mass shooting should influence more to local economic entities than that of other places. Accordingly, the third type of data source is location data.

As discussed in Chaney, Sraer, and Thesmar (2012), they use he hand-collected 10K file data to prove headquarters and facilities usually cluster in the same Metropolitan Statistical Area (MSA), and usually represents a great fraction of the overall firm's real estate asset. Thus, I follow the same assumption to use the headquarter location as a proxy for the firm location. The measure of the firm headquarter location is the headquarter's county code from Compustat [1]. However, the economic connection is not simply confined by county boundaries. Applying the same argument from Chaney, Sraer, and Thesmar (2012), I define the companies with the same MSA as one regional cluster. MSA is the geographic definition constructed by the Office of Management and Budget that contains over 50000 population. In each MSA, there are one or more counties that are highly integrated economically and culturally. Recently, the Census Bureau uses the term "Core Based Statistical Area" (CBSA) to refer to both metro and micro areas. Usually, MSA and metro CBSA terms are used interchangeably in the Census Bureau website, but the geographical definition is slightly different between MSA and CBSA. In the following content, I also use the county and CBSA as a robustness check.

[1] I follow the Engelberg, Ozoguz, and Wang (2018) to account for the headquarter location change.

2.4 Attention

The fourth type of data is public attention. I use the google trend search volume as the measure of public attention. There are three reasons to use google trends search data. First, google trend search volume data could be attained at a relative high weekly frequency. One needs weekly frequency data because attention is a kind of short-term concept. As documented in Da, Engelberg, and Gao (2011), google search volume attention measures predict the subsequent stock return in the two-week horizon. In the analysis, I apply the annual and weekly frequency to check for the robustness. Second, google trend search volume data has a very flexible topic range. The topic is not limited to economic topics, one is able to capture the public attention to some general economic issue. In my case, I use the search item ”income inequality” to capture the public attention to the inequality issue. Third, the google search volume is an active measure. If someone is searching for some keywords in Google, he or she must be paying attention to it. Compared with some traditional attention measures like media news coverage Liu, Sherman, and Zhang (2014), the measure is passive. If a piece of information is broadcast in media people do not necessarily pay attention to it. Based on these reasons, I use the google trends search volume data as the attention to inequality measure.

2.5 Equity premium

The fifth type of data is the equity premium. I use the CRSP monthly return data as the equity return data source. I compound the monthly return from January to December as the annual return of each firm. I remove the observations of lowest two deciles by NYSE breakpoint and price below five dollars to remove the distortion from small caps. Throughout the paper, I define the industry as the Fama-French 48 industry classification by SIC code. I also remove the firms with FF48 Guns industry classification as code 26. The reason there are anecdotal debate if mass shooting is a piece of good or bad news for gun companies. It could be good news, because mass shooting may boost the public defense demand to buy more guns. It could also be bad news because mass shootings could drive up the legislation process to constrain gun accessibility Luca, Malhotra, and Poliquin (2019). The risk-free rate is obtained from Kenneth French’s personal website. In the next section, I briefly explain the background and motivate the exercise to examine the inequality information sensitivity around mass shootings.

3 Background: mass shooting and inequality

As suggested by the anecdotal evidence, Mark Cuban argues that mass shooting is associated with the income inequality, and one way to mitigate the mass shooting problem in the U.S. is to mitigate the deteriorating condition of income inequality. In the academy, as early as Merton and Merton (1968) argue that community with severer income inequality would suffer more serious anger, frustration, resentment, and hostility atmosphere. Countless studies connect income inequality with crime (Blau and Blau, 1982), and homicide (Pridemore, 2002). Specifically, Kwon and Cabrera (2019) document that income inequality is associated with mass shootings across counties.

In this section, I briefly redo the summary statistical analysis about inequality and mass shooting for the logical completeness of the paper [2]. First, I compare the trends between inequality and mass shooting at the aggregate level for the United State. I capture the inequality by the share of income from the top 1% earners in that of the overall population. I define the mass shooting according to Luca, Malhotra, and Poliquin (2019) the definition of gun violence with 3 casualties. The shooting due to a romantic, economic, or family relationship will be excluded from the sample. Generally, a mass shooting is serious gun violence without a specific target.

I depict the income inequality trend in Figure 1 by the solid line. The income inequality slightly decreases in the first ten years in our sample from 1966 to 1976. After 1976, the trend presents a long-term increase. As documented by Piketty and Saez (2003) and Piketty, Saez, and Zucman (2017), income inequality, wealth inequality and a general sense of unbalanced economic distribution in the United States experience a long-term deterioration. At the same time, the trend of mass shootings in the United State also experiences a long-term increase. From the 1960s to the 1970s, there are usually less than 2 mass shootings with at least 3 casualties each year. In some years, there is no such serious mass shooting. However, after the 1980s, a serious mass shooting happens nearly on an annual or even monthly basis. Thus, one could get one statistical conclusion that income inequality and mass shootings share a similar upward trend from the 1960s to the 2010s at the aggregate U.S. level.

[Insert Figure 1 around here]

[2] I do not take a stand on whether the income inequality and mass shooting are causally correlated or just statistical correlated.

Next, I investigate the relationship between income inequality and mass shooting at the cross-state level. In the upper panel of Figure 2, I depict the cross-state income inequality and mass shooting information in a geographical manner. In the state map of the U.S., the color shade represents the income inequality, which is captured by the time-series mean of top 1% earners' income share in each state from 1966-2014. The darker the shade, the higher the income inequality. The area of solid circles in the center of each state stands for the number of the mass shooting in each state in the same period. The bigger the circle, the more mass shooting. In the lower panel of Figure 2, I present the same information in the bar plot manner. The clear bar and solid bar represent the income inequality and number of the mass shooting, respectively. Among all the states, California, Florida, and Texas are the three states that have the worst economic inequality condition. Not surprisingly, these states also have relatively more mass shootings than most of other states. Several exceptions are New York and Delaware. Both states have nearly the highest income inequality, but their mass shootings do not rank the highest.

[Insert Figure 2 around here]

The detailed information of State level, top 1% and 10% earner's income share and mass shooting are listed in Table 1. Because the top 1% of earner's income share is a persistent variable, before any serious empirical test, I need to generate the stationary variables. I also check if the stationary variable KGR still keeps a similar pattern. As expected, a similar pattern emerges from Figure 3, the California, Texas, and Florida also have relatively high KGR from 1966-2014.

[Insert Figure 3 around here]

I conclude the graphical patterns by correlation analysis in Table 2. The income inequality is captured by three measures, top 1%, and 10% earners' income share. At the time-series level, the income inequality measure is around 60% co-move with the total number of the mass shooting in that year, and 50% co-move with the total casualties. I next compare the median of income inequality measures across the state and the overall mass shooting and casualties in that state. The pattern still keeps that income inequality and mass shooting co-move over 30% across the state. After comparing the first order inequality information, I next compare the stationary inequality measure KGR's correlation. Also in Table 2, the last two rows of each panel present the correlations between KGR and the number of the mass shooting and the number of casualties. The pattern keeps that inequality and mass shootings are highly positively correlated with each other. At the time-series level and cross-state level, the correlation between KGR and the number of the mass shooting is above 30% and 40 %, respectively.

[Insert Table 2 around here]

However, the mass shooting is a very complex function of many different inputs. The states' area, population density, gun control policy and cultural atmosphere could influence the mass shootings as well. The key information I would like to delivery in this section is that inequality and mass shootings do co-move greatly in the United States. Anecdotal evidence and academic evidence have established a positive correlation between inequality and mass shooting. My next question is whether the equity market reacts to inequity information when a mass shooting happens.

4 The impact of mass shooting on inequality-return sensitivity

In this section, before establishing any channel, I investigate one asset pricing fact how the market reacts to the inequality when mass shootings. This exercise is simply motivated by the statistical fact I present in the previous section that inequality and mass shootings are highly correlated with each other. If public attention to inequality does spike with a mass shooting, one intuitive asset pricing phenomenon should emerge that inequality's predictability for stock price concentrate on the period of mass shootings. I present the result in the sequence of inequality, mass shooting, and interaction to gradually introduce from first-order effects to the interaction effect. In the next section, I present severally robustness check to make sure the result is not driven by mechanical sample reasons.

I first test the cross-regional inequality-return relationship. As discussed in Toda and Walsh (2019), the wealthy class is generally more optimistic and less risk-averse. If more wealth shifts into the wealthy class, prices of risky asset increases and subsequent market return decrease. Accordingly, KGR is a negative predictor of the equity premium. The empirical setting of Toda and Walsh (2019) is mainly on U.S. annual time-series, and international annual panel. In other word, aggregate annual frequency. Empirically, whether an aggregate predictor' statistical property remains at the cross-sectional level is not guaranteed, like Hirshleifer, Hou, and Teoh (2009)[3]. Thus, I first test if the negative inequality-return relation still exists at the cross-sectional level by the following model,

$$r_{i,j,t} = \alpha + \beta_1 \text{KGR}_{j,t-1} + \mathbf{FE}(t) + \varepsilon_t, \tag{7}$$

where the $r_{i,j,t}$ is the annual return of firm i, the subscript j stands for the firm headquarter's location region, and KGR_{t-1} is the previous year's capital gain ratio in region j. Additionally, I also incorporate two different combinations of fixed effects **FE**(t). One combination is the firm and year fixed effects, and the other combination is firm, industry and region and year fixed effects. The standard errors are clustered at firm, industry and region level. As discussed in Toda and Walsh (2019), I employ the lagged KGR to resolve the simultaneity concern, that stock return could also influence the top shares contemporaneously, but cannot influence the lagged KGR. As documented in Toda and Walsh (2019), I expect β_1 in Eq. (7) is negative. I present the regression results in columns (1) and (4) of Table 3. In line with the time-series evidence of Toda and Walsh (2019), at the cross-sectional level, the one-percent-point increase in KGR predicts on average -1.07% the subsequent return, which is marginally significant at 10% level. Compared with the time-series annual level results in Toda and Walsh (2019), the one-percent-point increase in KGR predicts -2.69% subsequent return. The main difference may come from different sample period (1913-2015 in Toda and Walsh 2019), or the cross-sectional deviation. However, one could conclude that the cross-sectional and aggregate time-series analysis for inequality-return relationship is largely consistent for both scale and significance.

[Insert Table 3 around here]

After checking the inequality-return relationship at the cross-sectional level, I next turn to the empirical question when the inequality information is incorporated into the stock market. I implement a difference in difference strategy that compares the inequality-return sensitivity before and after the mass shooting, in regions where the mass shooting happens compared with rest regions. My dependent variable is the firm-level equity premium. My independent variable to capture inequality is KGR from Toda and Walsh (2019), which essentially is the difference method between top income share with or without capital gain

[3] In Hirshleifer, Hou, and Teoh (2009), they document that accrual is a cross-sectional negative predictor but time-series positive predictor.

for stationary concerns. The other main independent variable is the mass shooting as a plausible random shock. Based on the main argument 1) the mass-shooting and inequality are statistically connected; 2) after decades of connection, public attention to inequality would spike when there is a mass shooting, I formally hypothesize:

Hypothesis 1. *The inequality information is even incorporated into stock market regardless of mass shootings.*

To test the Hypothesis1, I first estimate the similar regression as Hirshleifer and Shumway (2003) and Edmans, Garcia, and Norli (2007),

$$r_{i,j,t} = \alpha + \beta_1 \text{KGR}_{j,t-1} + \beta_2 I^{\text{Shooting}}_{j,t} + \beta_3 I^{\text{Shooting}}_{j,t} \text{KGR}_{j,t-1} + \mathbf{FE}(t) + \varepsilon_t, \tag{8}$$

$$r_{i,j,t} = \alpha + \beta_1 \text{KGR}_{j,t-1} + \beta_2 \text{Num}^{\text{Shooting}}_{j,t} + \beta_3 \text{Num}^{\text{Shooting}}_{j,t} \text{KGR}_{j,t-1} + \mathbf{FE}(t) + \varepsilon_t, \tag{9}$$

where $I^{\text{Shooting}}_{j,t}$ is the indicator that equals one if there is at least one mass shooting in location j, and $\text{Num}^{\text{Shooting}}_{j,t}$ is the logarithm total number of mass shooting in this MSA as $\ln(1 + \text{number of mass shooitng})$. I also keep the same fixed effect settings as in Eq. (7). Next, I could translate the Hypothesis 1 into the test the direction and statistical significance β_3. Effectively, I investigate if the inequality-return relationship would be dramatically different between the mass shooting and the non-mass shooting returns. I present the regressions in Table 3. In column (2) Panel A, I first test if the mass shooting influence the contemporaneous stock market return and the coefficient β_2 is -3.59 with marginal 10% significance. As mentioned, I have removed all the guns or defense-related firms with the FF48 industry code of 26. Thus, for guns unrelated firms, the coefficient implies that treated with at least one unexpected mass shooting will lead to -3.59 percent return relative to non-treated firm and year.

The main result is presented in column (3). After controlling for the interaction term of mass shooting and KGR, both mass shooting indicator and KGR are not significant anymore, but the signs do not change. The coefficient of interaction implies that conditioning on mass shooting treatment, one-percent-point increase in KGR predicts -5.07% equity premium. Similar results emerge if employ the logarithm local mass shooting number. In column (5) to (8), I control for firm, industry, region and year fixed effects, the result barely change. Thus, the industry and region heterogeneity is not a major concern for the robustness of the result. Another concern is that the sample selection of mass shootings may introduce unobservable bias.

In summary, I reject the Hypothesis1 that the inequality information is even incorporated into stock market regardless of mass shootings. The empirical evidence indicates that the negative inequality-return relationship is much stronger when being treated by mass shooting. I next turn to a series robust test to ensure the results still survive.

5 Robustness

In this section, I keep testing the main results of inequality-return sensitivity treated by mass shooting under a series of robustness settings.

One may concern that the mass shooting is filtered by some criteria, it is still possible that the result is largely driven by a subsample of the mass shooting. In Panel B of Table 3, I present regression results with the same setting but consider all samples of the mass shooting. The results show that all the empirics keep a similar pattern. The minor difference between Panel A and B indicate the empirical result is not driven by mass shooting's filter selection.

The next concern is that the results could simply be driven by a very specific region definition. First, the MSA definition is one of the most pervasive regional definitions (Tuzel and Zhang, 2017). If the result in the MSA region definition presents results, it should be fairly representative of local economic clusters. However, to mitigate potential regional definition concern. I apply the analysis to two different regional definitions. The first one is the administrative county, and the second one is the CBSA definition. I do not use the state as the regional cluster due to several reasons. First, compared with the county the state is a too large definition to capture economic and cultural integration. Second, in the same year across the whole state, there could be many different legislative, administrative and natural treatment affect the whole state. However, at the relative small regional definition like the county, CBSA, and MSA, the concern of uncontrolled treatment could be largely mitigated.

[Insert Table 4 around here]

The results are presented in Table 4. In column (1) and (2), the results for the county region cluster keep the same pattern as the MSA cluster. The main difference is that the interaction term is only marginally significant. The potential reason is that the influence of mass shootings does not confine in the boundary of a county. To well capture the regional influence of mass shooting, one needs to use a geographical definition at the urban cluster level. I next turn to the CBSA urban cluster definition. In column (3) and (4), the results for the CBSA region cluster also keep the pattern and significance pretty well, which indicates that the result is not driven by the urban area selection.

In summary, neither sample selection of mass shooting or definition of the urban cluster should mechanically drive mass shooting conditional inequality-return relationship. Next, I test one potential channel that mass shootings and negative equity premiums are both the results inequality increase.

6 The Mechanisms: simultaneity of return and mass shooting?

One potential channel that could lead the conditional inequality-return relationship is that increasing inequality leads to both mass shootings and negative stock market returns. To simplify the description, I name this channel as simultaneity channel.

If the simultaneity channel does hold, one direct implication is that if two regions in the U.S. share the same conditions, the increase in inequality should generate commensurable inequality-return sensitivity with or without happening a mass shooting. To test this implication, the ideal case is that one could find two sets of locations with the same condition, the only difference is the mass shooting happens or not. However, there is no such data in the real lift at all. To approach the ideal case as close as possible, I employ the subsample analysis strategy. I first single out these firms located the counties where there was a mass shooting, and group these observations as the treated sample. I employ the county adjacent table from Census Bureau to identify the counties that just alone side the counties in the treated sample. The overarching objective to create two subsamples with as much geographical similarity as possible. Next, I formalize the Hypothesis 2,

Hypothesis 2. *The inequality-return relationships are indifferent between mass shooting treated sample and adjacent sample with similar economic, cultural and geographical conditions.*

To test the hypothesis, I run the following regression in the two subsamples separately,

$$r_{i,j,t}^{\text{Treated}} = \alpha + \beta_1 \text{KGR}_{j,t-1}^{\text{Treated}} + \mathbf{FE}(t) + \varepsilon_t, \tag{10}$$

$$r_{i,j,t}^{\text{Adjacent}} = \alpha + \beta_1 \text{KGR}_{j,t-1}^{\text{Adjacent}} + \mathbf{FE}(t) + \varepsilon_t, \tag{11}$$

where the empirical design is largely similar to Eq. (7). The only difference is that Eq. (10) only includes the observations in the county where mass shootings ever happen in the sample and the Eq. (11) only includes the observations just adjacent to these counties. To figure out how the inequality-return relationship in these two samples could statistically differentiate from each other, I also design a pooled-regress with both subsample group together, and label them by indicator as,

$$r_{i,j,t} = \alpha + \beta_1 I^{\text{Treated}} \text{KGR}_{j,t-1}^{\text{Treated}} + \beta_2 I^{\text{Adjacent}} \text{KGR}_{j,t-1}^{\text{Adjacent}} + \mathbf{FE}(t) + \varepsilon_t, \tag{12}$$

where I^{Treated} and I^{Adjacent} mean which subsample the observations belong to. By employing the pooled regression, I can design a Wald test to examine the difference of inequality-return relationship in two subsamples. One could translate the Hypothesis 2 as to test if $\beta_1 = \beta_2$ in Eq. (12).

[Insert Table 5 around here]

I present the regression results in Table 5. Column (1) represents Eq. (10) and indicates that the inequality-return relationship presents a strong negative pattern. In this subsample that has been treated by a mass shooting, a one percent increase in KGR leads to -3.83 % equity premium in the subsequent year with 5% significance. However, in the adjacent subsample where the inequality and other economic condition should be very close to treated subsample, the inequality-return relation does not present any significance at all, which corresponds to column (2) where the coefficient is -0.31 with one decimal scale lower that of the treated subsample. The formal test of the hypothesis 2 is in column (3). A similar pattern appears that only the coefficient of mass shooting treated indicator shows significance. In the lower panel, I present the result of the Wald test of the null hypothesis $\beta_1 = \beta_2$. The F-value is 7.32, which indicates that $\beta_1 = \beta_2$ in Eq. (12) has been highly rejected with 1% significance.

One concern of this subsample analysis is that, the sample selection is totally an ex-post criteria and the implicit assumption is that the influence of mass shooting exists forever. To mitigate the look-ahead bias, I design a simple regime switching model to take all observation into consideration,

$$r_{i,j,t} = \alpha + \beta_1 I^{\text{Treated}}_{j,t}\text{KGR}_{j,t-1} + \beta_2 I^{\text{Untreated}}_{j,t}\text{KGR}_{j,t-1} + \mathbf{FE}(t) + \varepsilon_t, \tag{13}$$

where $I^{\text{Treated}}_{j,t}$ indicates the observation is treated by mass shooting in year t with value of unity, otherwise 0. $I^{\text{Untreated}}_{j,t}$ is a compensate indicator with opposite value of 0 and 1 with $I^{\text{Treated}}_{j,t}$. The implicit assumption is that the influence of mass shootings just exists for one year. Also, to mitigate the concern that the definition of the region could influence the results, I employ county, MSA and CBSA definition for robustness, and present the result in column (1) - (3) of Table 6, respectively. As expected, even with the assumption that the effect lasts for just one year, the results still keep a similar pattern that the negative inequality-return relationship mainly concentrates on the mass shooting treated year and region.

[Insert Table 6 around here]

Next, I also test the subsequent reaction of the stock market return. If the current stock price does incorporate the contemporaneous inequality information, one could also expect the reserve subsequent reaction. The intuition is that as $r_t = (P_t + D_t)/P_{t-1}$, the decrease in P_t will lead decrease in r_t for contemporaneous relation. However, a decrease in P_{t-1} will lead to an increase in r_t as subsequent relation. Specifically, I run the following regression to test the subsequent reaction,

$$r_{i,j,t} = \alpha + \beta_1\text{KGR}_{j,t-2} + \beta_2 I^{\text{Shooting}}_{j,t-1} + \beta_3 I^{\text{Shooting}}_{j,t-1}\text{KGR}_{j,t-2} + \mathbf{FE}(t) + \varepsilon_t, \tag{14}$$

$$r_{i,j,t} = \alpha + \beta_1\text{KGR}_{j,t-2} + \beta_2\text{Num}^{\text{Shooting}}_{j,t-1} + \beta_3\text{Num}^{\text{Shooting}}_{j,t-1}\text{KGR}_{j,t-2} + \mathbf{FE}(t) + \varepsilon_t, \tag{15}$$

where the main difference is that there is a two-year distance between KGR and subsequent return. My coefficient of interest is β_3, and I expect β_3 is significantly positive. I present the result in Table 7. In column (1) and (2), regression subsequent return with a two-year lag of KGR and one-year lag of mass shooting indicator, there is no conclusive result. Thus, one can summarize that with longer time distance, KGR and mass shootings are not informative to subsequent return. However, in column (3)-(4), the interaction term present positive significance, which is in line with my expectations. The results also indicate that ceteris paribus, mass shooting and inequality interactively be priced into the stock market previously.

[Insert Table 7 around here]

In summary, the empirical results highly reject Hypothesis 2 that the inequality-return relationships are indifferent between mass shooting treated samples and adjacent samples with similar economic, cultural and geographical conditions. The negative inequality-return relationship is largely driven by the mass shooting treated subsamples. The results are consistent with the subsample analysis and regime-switching method. I next turn to the attention channel analysis.

7 Attention to inequality and mass shooting

As early as Pavlov and Gantt (1928), scholars have documented a widespread conditioned reflex phenomenon in advanced animals. Human beings are not exceptions in this aspect. The modern studies also confirm that human's attention also exhibits the conditioned reflex pattern (Lang, Bradley, and Cuthbert 1990, Lang, Bradley, Cuthbert, et al. 1997). If human beings observe two objects A and B happened at the same time, he or she's attention to B would be spurred when observing A. However, the underlying relation between A and B could be independent by experiment setting.

In this section, I proceed with the analysis of mass shootings and inequality. Given inequality and mass shooting are significantly positively correlated with each other in the United States. The main point in this section is to answer the question if a mass shooting would spike the public attention to inequality so that the information of inequality would be priced into the equity market mainly along with mass shootings. The first challenge to answer this question is the measure. Attention per se is not directly observable. One needs to rely on a measure to capture public attention to a certain issue. To overcome this challenge, I turn to google trends search volume as the attention measure from Da, Engelberg, and Gao (2011) and Drake, Roulstone, and Thornock (2012).

I obtain the search volume index data from the google trends API for each state in the United States. The keyword I employ as the baseline is "income inequality". Google provides an accurate search volume index for weekly frequency. Because the mass shooting is a salient event and plausibly unpredictable, I employ the difference in difference method around the timing of mass shooting to test its impact on public attention to inequality by the following model at the annual frequency,

$$\mathrm{SVI}^{\mathrm{Inequality}}_{j,t} = \alpha + \beta_1 I^{\mathrm{Shooting}}_{t,j} + \gamma X_{j,t-1} + \mathrm{FE}(t) + \varepsilon_t, \tag{16}$$

$$\mathrm{SVI}^{\mathrm{Inequality}}_{j,t} = \alpha + \beta_1 \mathrm{Num}^{\mathrm{Shooting}}_{t,j} + \gamma X_{j,t-1} + \mathrm{FE}(t) + \varepsilon_t, \tag{17}$$

where $\mathrm{SVI}^{\mathrm{Inequality}}_{j,t}$ is the google search volume index for inequality issue for region j in time t. I follow Da, Engelberg, and Gao (2011) to employ the logarithm of raw volume index as the SVI. $I^{\mathrm{Shooting}}_{t,j}$ is the treatment indicator of mass shooting, and $\mathrm{Num}^{\mathrm{Shooting}}_{t,j}$ is the logarithm of mass shooting number in region j. $X_{j,t}$ represents a bunch of control variables that could also potentially influence the search volume of the inequality issue. $X_{j,t}$ include the inequality variable KGR, unemployment rate, GDP growth rate, population density, and households density. As the inequality measure KGR, it could increase the public attention to the inequality issue. The unemployment rate and GDP growth rate are to capture the local economic condition. A high unemployment rate or low GDP growth rate could increase public attention to the inequality issue. The population density and household density are to capture the size of the potential searcher that will use google to search the inequality issue. I control for time and region fixed effects to account for the unobservable time and geographical factors that influence search volume.

[Insert Table 8 around here]

The results are presented in Table 8. Columns (1) and (3) indicate that, without any other control variables, the google search volume about income inequality significantly co-move with the mass shooting. Column (2) and (4) present the full variable tests. After controlling for other variables, the mass shootings still significantly introduce the increase in google search volume about income inequality. If the state and in a given year treated by a mass shooting, the search volume would be 21% higher than untreated state-years. The scale is not only significant at 1% level, but also economically great.

Because in many settings like Da, Engelberg, and Gao (2011), the attention is a short term economic variable, for the concreteness of the result, I also apply the same setting at the weekly frequency. For there variable that is at annual frequency, I simply use the last year's value as the imputation. The weekly result is very similar to the annual results. In column (2), mass shooting treatment indicates 49% google search volume increase about inequality than other untreated state-week. Figure 4 presents the movement of google search volume about inequality issue around the mass shootings. In panel A, I present the time-series dynamics from minus 3 week to plus 3 week. The pattern is obvious that when mass shooting happens in a given region, the attention to mass shooting dramatically increases compared with previous weeks. In cross-sectional level, Panel B shows that in the same week, the attention to inequality is much higher in the mass shooting treated regions than control regions.

[Insert Figure 4 around here]

In summary, given the very long-term co-occurrence between mass shootings and inequality increases, when a mass shooting happens, the public attention to inequality issue does increase greatly. The evidence supports the channel that mass shooting facilitates as a reminder to the equity market to pay attention to inequality.

8 Conclusion

Even just account for a very small fraction for gun violence, mass shootings do play an important role in U.S. society and draw great attention from the public, media, and policymakers. This study documents the asset pricing fact that the inequality-return relationship is largely driven by the mass shooting treated stock sample from a geographical standpoint of view. The potential channel should be that public attention would dramatically concentrate on the inequality conditions to the place when mass shootings happen.

References

Blau, J. R., Blau, P. M., 1982. The cost of inequality: Metropolitan structure and violent crime. American sociological review pp. 114–129.

Campbell, S. D., Delikouras, S., Jiang, D., Korniotis, G. M., 2016. The human capital that matters: Expected returns and high-income households. The Review of Financial Studies 29, 2523–2563.

Chaney, T., Sraer, D., Thesmar, D., 2012. The collateral channel: How real estate shocks affect corporate investment. American Economic Review 102, 2381–2409.

Chetty, R., Hendren, N., Kline, P., Saez, E., 2014a. Where is the land of opportunity? the geography of intergenerational mobility in the united states. The Quarterly Journal of Economics 129, 1553–1623.

Chetty, R., Hendren, N., Kline, P., Saez, E., Turner, N., 2014b. Is the united states still a land of opportunity? recent trends in intergenerational mobility. American Economic Review 104, 141–47.

Coval, J. D., Moskowitz, T. J., 2001. The geography of investment: Informed trading and asset prices. Journal of political Economy 109, 811–841.

Da, Z., Engelberg, J., Gao, P., 2011. In search of attention. The Journal of Finance 66, 1461–1499.

David, H., Katz, L. F., Kearney, M. S., 2006. The polarization of the us labor market. American economic review 96, 189–194.

Drake, M. S., Roulstone, D. T., Thornock, J. R., 2012. Investor information demand: Evidence from google searches around earnings announcements. Journal of Accounting research 50, 1001–1040.

Edmans, A., Garcia, D., Norli, Ø., 2007. Sports sentiment and stock returns. The Journal of Finance 62, 1967–1998.

Engelberg, J., Ozoguz, A., Wang, S., 2018. Know thy neighbor: Industry clusters, information spillovers, and market efficiency. Journal of Financial and Quantitative Analysis 53, 19371961.

Fisher, I., 1910. Introduction to Economic Science. Macmillan.

Granger, C. W., 1981. Some properties of time series data and their use in econometric model specification. Journal of econometrics 16, 121–130.

Hirshleifer, D., Hou, K., Teoh, S. H., 2009. Accruals, cash flows, and aggregate stock returns. Journal of Financial Economics 91, 389–406.

Hirshleifer, D., Shumway, T., 2003. Good day sunshine: Stock returns and the weather. The Journal of Finance 58, 1009–1032.

Kwon, R., Cabrera, J. F., 2019. Income inequality and mass shootings in the united states. BMC public health 19, 1147.

Lang, P. J., Bradley, M. M., Cuthbert, B. N., 1990. Emotion, attention, and the startle reflex. Psychological review 97, 377.

Lang, P. J., Bradley, M. M., Cuthbert, B. N., et al., 1997. Motivated attention: Affect, activation, and action. Attention and orienting: Sensory and motivational processes 97, 135.

Liu, L. X., Sherman, A. E., Zhang, Y., 2014. The long-run role of the media: Evidence from initial public offerings. Management Science 60, 1945–1964.

Luca, M., Malhotra, D., Poliquin, C., 2019. The impact of mass shootings on gun policy. Wokring Paper .

Merton, R. K., Merton, R. C., 1968. Social theory and social structure. Simon and Schuster.

Pavlov, I. P., Gantt, W., 1928. Lectures on conditioned reflexes: Twenty-five years of objective study of the higher nervous activity (behaviour) of animals. .

Piketty, T., Saez, E., 2003. Income inequality in the united states, 1913–1998. The Quarterly journal of economics 118, 1–41.

Piketty, T., Saez, E., Zucman, G., 2017. Distributional national accounts: methods and estimates for the united states. The Quarterly Journal of Economics 133, 553–609.

Pridemore, W. A., 2002. What we know about social structure and homicide: A review of the theoretical and empirical literature. Violence and victims 17, 127.

Toda, A. A., Walsh, K. J., 2019. The Equity Premium and the One Percent. The Review of Financial Studies Hhz121.

Tuzel, S., Zhang, M. B., 2017. Local risk, local factors, and asset prices. The Journal of Finance 72, 325–370.

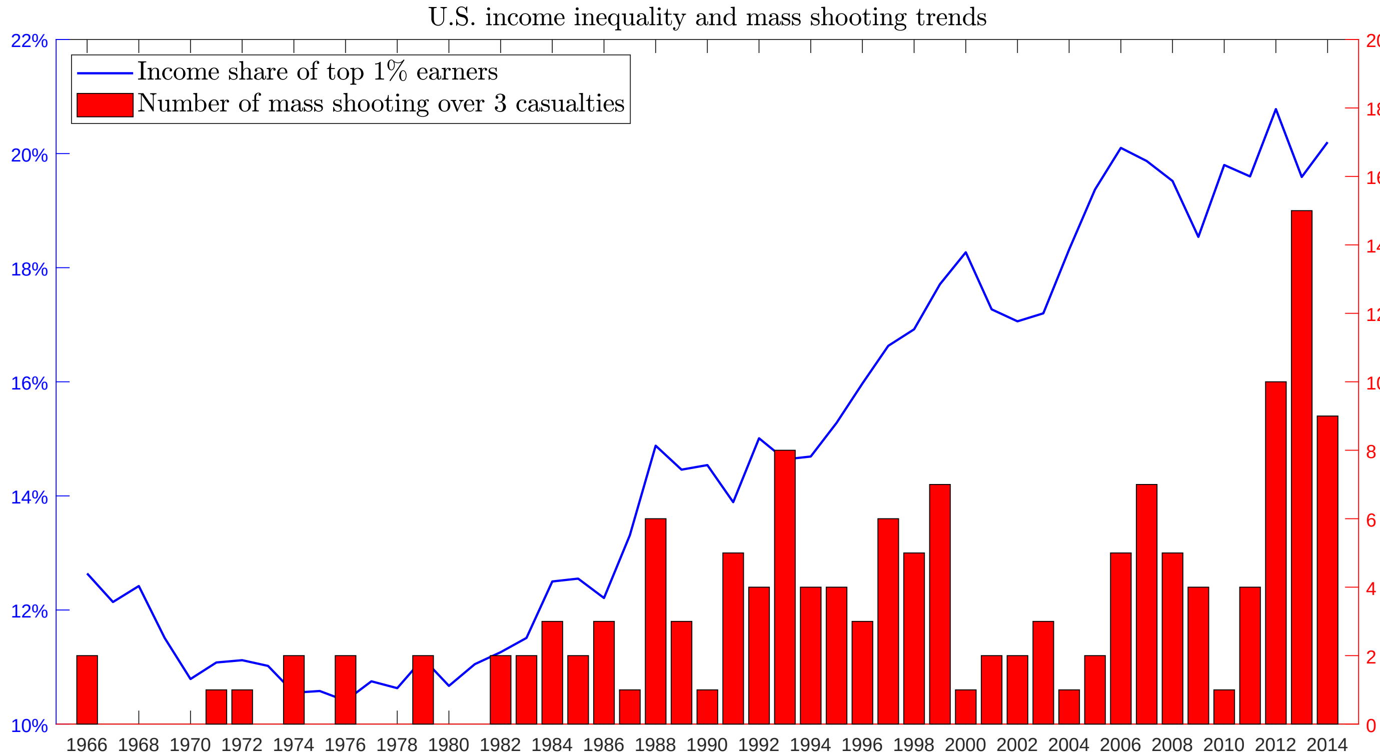


Figure 1 This figure plots trends of economic inequality and counts of the mass shooting in the United States from 1966-2014. The solid curve presents the inequality, and the solid bar presents mass shooting count on an annual basis. The inequality is defined by the income share of top 1% earners in that of the overall population. The mass shooting is defined as gun violence with at least 3 casualties. The shootings due to a romantic, economic, or family relationship are removed from the sample.

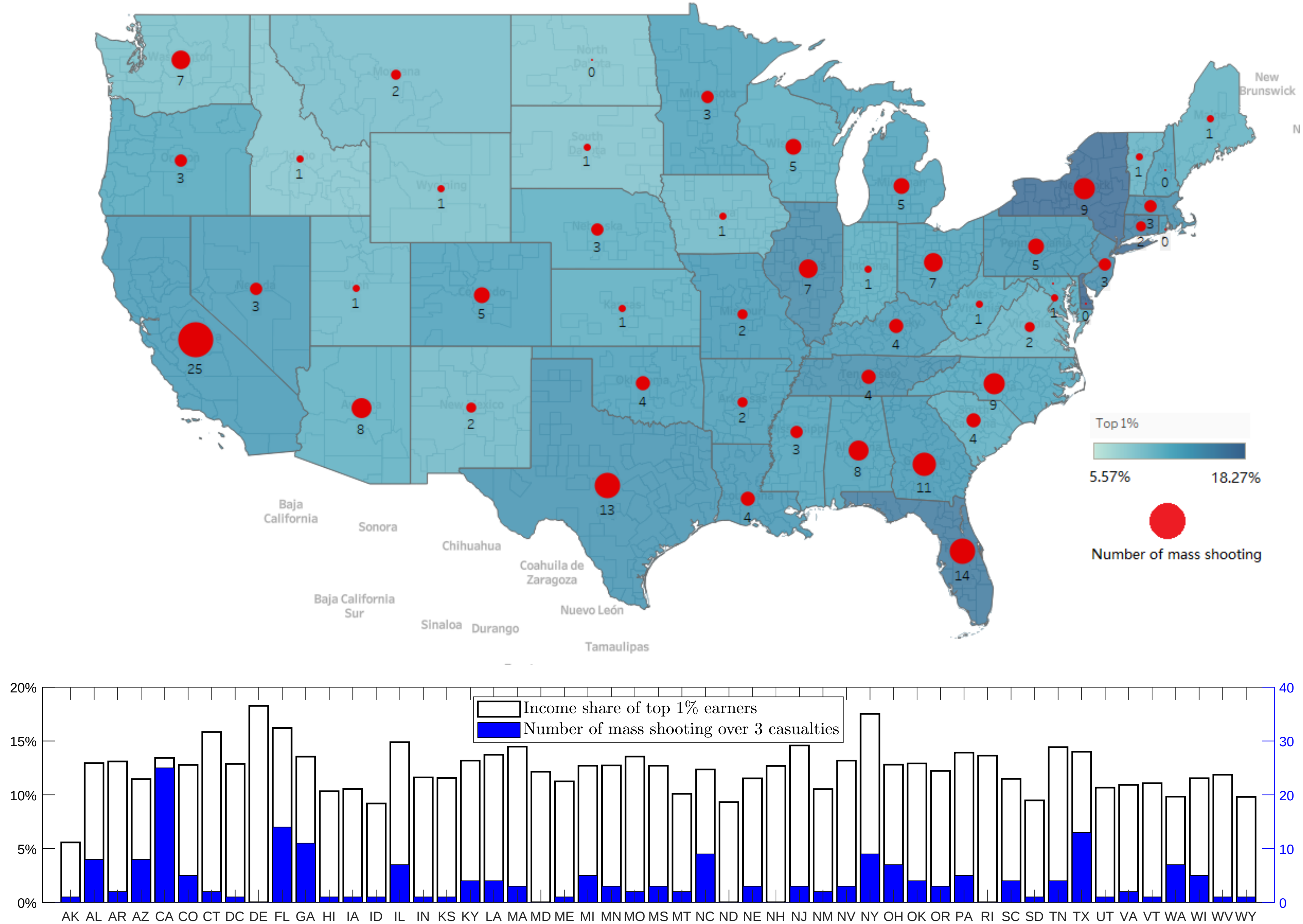


Figure 2 This figure plots the state level of income inequality and the number of mass shootings from 1966-2014. The inequality is defined by the income share of the top 1% earners in the overall population. In the upper panel, the median inequality measures of each state are presented by the color shadow. The heavier the color higher the top 1% earners' share. The area of the solid circle in each state presents the number of mass shooting happened in that state from 1966-2014. In the lower panel, the same information is presented by a bar plot manner by each state.

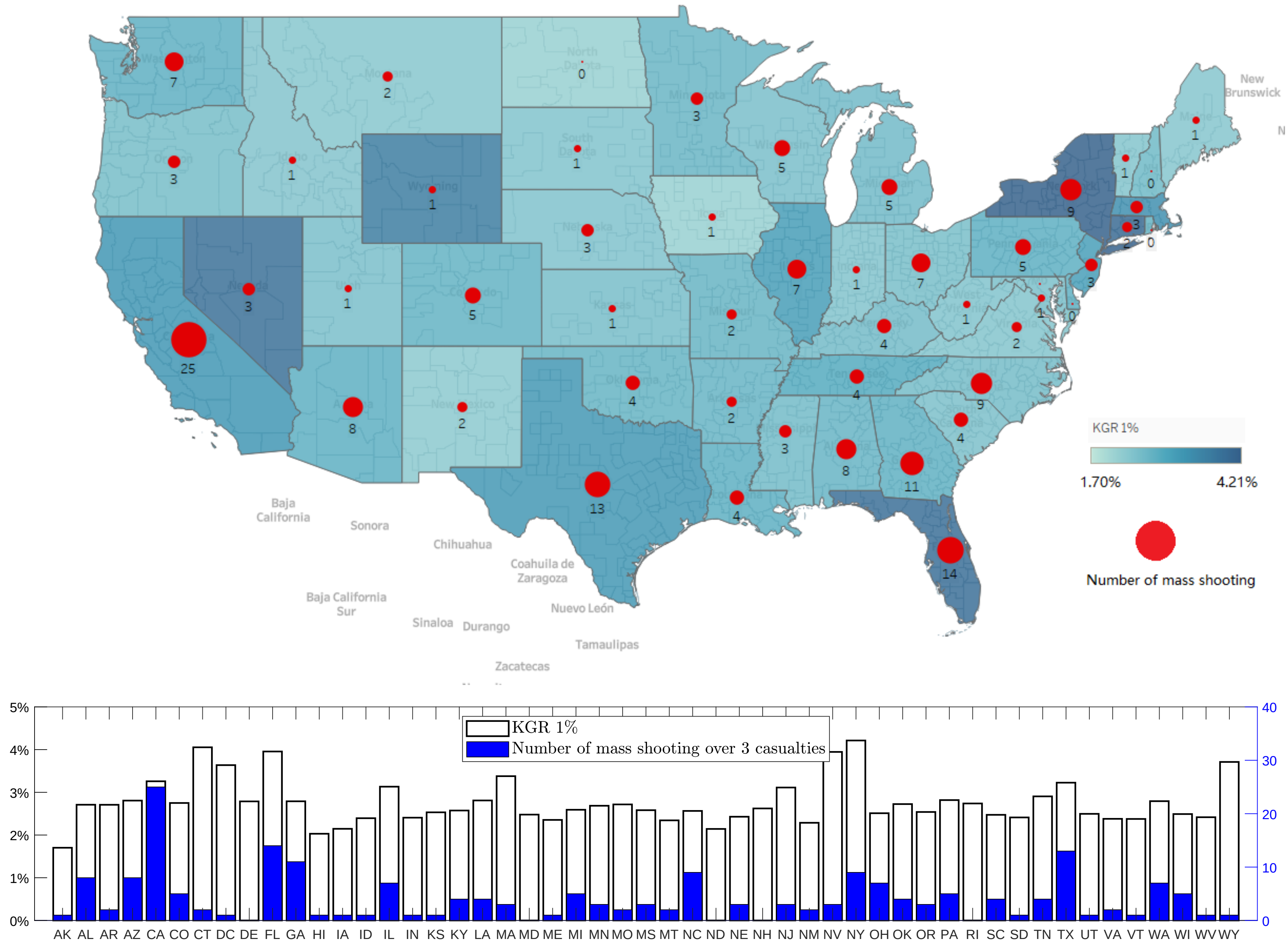
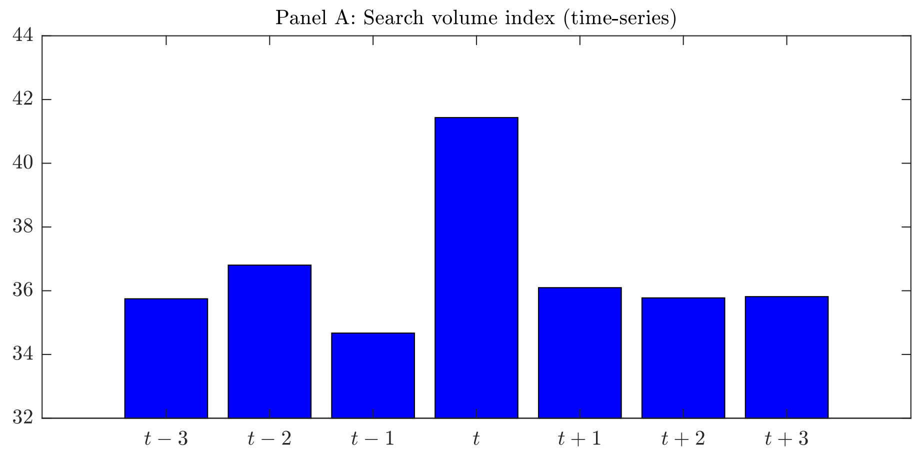


Figure 3 This figure plots the state level of income inequality and the number of mass shootings from 1966-2014. The inequality is defined by KGR computed by the 1% cutoff in each state's population. In the upper panel, the median inequality measures of each state are presented by the color shadow. The heavier the color higher the top 1% earners' share. The area of the solid circle in each state presents the number of mass shooting happened in that state from 1966-2014. In the lower panel, the same information is presented by a bar plot manner by each state.



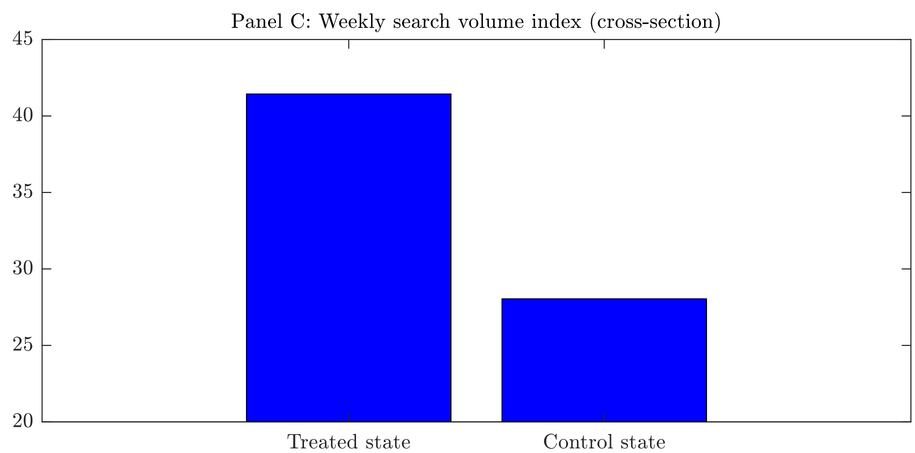


Figure 4 This figure plots the public attention about issue "income inequality" around an unexpected mass shooting. I measure the public attention by google trends search volume index. In a given week, if a mass shooting happens in one state, I denote this week as week t and this state as treated state. Accordingly, +1 and -1 stand for the week before and after and rest of states as control state. First row presents the time-series dynamics of the google search volume from week $t-3$ to $t+3$. Second, row present the search volume in treated states and control states. In the left column, I capture the attention by search volume index. In the right column, I capture the attention by abnormal search volume in google trends as in Da, Engelberg, and Gao (2011).

Table 1 Summary statistics

The table presents top 1% and 10% earner's income share and mass shooting summary statistics in the state level.

Panle A	Top 1%	Top 10%	Num of mass shooting	Num of casualty
AK	0.07	0.26	1	4
AL	0.13	0.38	15	79
AR	0.13	0.38	2	18
AZ	0.12	0.37	11	54
CA	0.14	0.39	38	369
CO	0.13	0.38	6	69
CT	0.17	0.42	2	33
DC	0.15	0.39	1	14
DE	0.20	0.44	1	4
FL	0.18	0.45	31	338
GA	0.14	0.39	16	92
HI	0.10	0.30	1	7
IA	0.11	0.35	1	6
ID	0.10	0.34	1	4
IL	0.15	0.40	11	77
IN	0.12	0.36	2	9
KS	0.12	0.36	4	17
KY	0.13	0.38	5	42
LA	0.14	0.40	7	51
MA	0.15	0.40	5	34
MD	0.13	0.39	6	31
ME	0.12	0.36	1	3
MI	0.13	0.38	10	57
MN	0.13	0.37	4	21
MO	0.14	0.39	4	21
MS	0.13	0.38	3	28
MT	0.11	0.36	2	8
NC	0.13	0.36	10	47
ND	0.10	0.33	0	0
NE	0.12	0.36	3	25
NH	0.12	0.36	0	0
NJ	0.14	0.40	4	17
NM	0.11	0.36	5	22
NV	0.15	0.38	5	563
NY	0.19	0.45	12	103
OH	0.13	0.37	9	41
OK	0.14	0.39	5	37
OR	0.12	0.37	3	55
PA	0.14	0.39	8	48
RI	0.14	0.38	0	0
SC	0.12	0.35	5	32
SD	0.11	0.34	1	5
TN	0.14	0.40	11	54
TX	0.15	0.40	23	273
UT	0.11	0.35	1	9
VA	0.11	0.35	8	34
VT	0.12	0.36	1	4
WA	0.12	0.36	10	75
WI	0.12	0.35	7	45
WV	0.12	0.38	1	4
WY	0.14	0.36	1	4

Table 1 (contiuned)

The table presents summary statistics of capital gain ratio (KGR) and associated augmented Dickey-Fuller test in the state level.

Panel B	KGR mean (%)	KGR std (%)	KGR skew	KGR kurt	Stationary	ADF test
AK	1.70	0.90	0.93	0.18	Yes***	-3.46
AL	2.71	1.39	1.39	2.29	Yes***	-3.61
AR	2.71	1.28	1.20	2.03	Yes***	-3.57
AZ	2.81	1.59	1.13	0.68	Yes***	-3.68
CA	3.26	1.98	1.15	0.68	Yes***	-3.63
CO	2.75	1.55	1.03	0.22	Yes***	-3.57
CT	4.05	2.53	1.02	0.21	Yes***	-3.77
DC	3.64	2.28	1.71	4.14	Yes***	-3.30
DE	2.79	1.45	0.99	0.44	Yes***	-3.92
FL	3.96	2.42	1.25	1.28	Yes***	-3.97
GA	2.79	1.48	1.14	1.23	Yes***	-3.75
HI	2.03	1.04	1.26	1.43	Yes***	-3.48
IA	2.15	1.00	1.13	1.42	Yes***	-3.64
ID	2.39	1.33	1.19	0.68	Yes***	-3.71
IL	3.13	1.72	1.01	0.36	Yes***	-3.82
IN	2.41	1.18	1.30	2.18	Yes***	-3.68
KS	2.53	1.21	1.02	0.75	Yes***	-3.68
KY	2.57	1.26	1.63	4.22	Yes***	-3.56
LA	2.81	1.37	1.25	2.06	Yes***	-3.62
MA	3.38	2.10	1.18	0.85	Yes***	-3.59
MD	2.48	1.35	1.12	0.95	Yes***	-3.58
ME	2.35	1.26	1.53	3.22	Yes***	-3.56
MI	2.59	1.27	0.99	0.76	Yes***	-3.81
MN	2.69	1.39	1.12	1.07	Yes***	-3.71
MO	2.72	1.36	1.20	1.74	Yes***	-3.68
MS	2.58	1.26	1.45	3.32	Yes***	-3.47
MT	2.34	1.21	0.96	0.18	Yes***	-3.72
NC	2.56	1.31	1.26	2.00	Yes***	-3.63
ND	2.14	1.10	1.05	0.61	Yes***	-3.48
NE	2.43	1.25	1.12	0.78	Yes***	-3.73
NH	2.62	1.49	1.42	2.03	Yes***	-3.56
NJ	3.11	1.77	1.24	1.54	Yes***	-3.55
NM	2.29	1.16	1.22	1.53	Yes***	-3.61
NV	3.95	2.43	1.06	0.35	Yes***	-3.97
NY	4.21	2.56	1.00	0.31	Yes***	-3.87
OH	2.51	1.21	1.16	1.73	Yes***	-3.63
OK	2.72	1.35	1.07	0.93	Yes***	-3.47
OR	2.54	1.29	1.08	0.73	Yes***	-3.73
PA	2.82	1.47	1.23	1.79	Yes***	-3.65
RI	2.74	1.44	1.48	3.07	Yes***	-3.50
SC	2.47	1.35	1.20	1.25	Yes***	-3.78
SD	2.41	1.30	0.96	-0.03	Yes***	-4.11
TN	2.91	1.49	1.35	2.40	Yes***	-3.72
TX	3.23	1.66	0.95	0.28	Yes***	-3.72
UT	2.50	1.38	1.07	0.30	Yes***	-3.87
VA	2.38	1.28	1.09	0.81	Yes***	-3.61
VT	2.38	1.31	1.38	2.19	Yes***	-3.75
WA	2.79	1.72	1.22	0.63	Yes***	-3.56
WI	2.49	1.25	1.02	0.76	Yes***	-3.68
WV	2.42	1.11	1.79	5.43	Yes***	-3.37
WY	3.71	3.16	2.22	6.19	Yes***	-4.03

Table 2 Correlation between income inequality and mass shooting

The table presents the Pearson correlation coefficients between inequality measures and mass shooting. In panel A of the time-series analysis, the correlation coefficients are calculated for the aggregate U.S. on an annual basis. In panel B, the correlation coefficients are calculated for each state. The income inequality is the median of the inequality measures from 1966-2014, and the number and casualties of a mass shooting are the sums for the same periods. The income inequality is captured by the top 1%, 5%, 10% earner's income share from 1966-2014. The mass shooting is defined by shootings with at least 3 casualties.

Panel A: Time-series	Number of mass shooting (%)	Casualties of mass shooting (%)
Top 1%	62.69***	52.13***
Top 10%	64.85***	53.46***
KGR 1%	32.75**	27.76**
KGR 10%	31.56**	28.13**
Panel B: Cross-state	**Number of mass shooting (%)**	**Casualties of mass shooting (%)**
Top 1%	36.87***	30.13**
Top 10%	36.03***	33.35**
KGR 1%	40.23***	55.06***
KGR 10%	47.24***	48.57***

Table 3 Inequality-return sensitivity and mass shooting

The table tests the regional stock market reaction to inequality information when a mass shooting happens using the following model,

$$r_{i,j,t} = \alpha + \beta_1 \text{KGR}_{j,t-1} + \beta_2 I^{\text{Shooting}}_{j,t} + \beta_3 I^{\text{Shooting}}_{j,t} \text{KGR}_{j,t-1} + \mathbf{FE}(t) + \varepsilon_t,$$

$$r_{i,j,t} = \alpha + \beta_1 \text{KGR}_{j,t-1} + \beta_2 \text{Num}^{\text{Shooting}}_{j,t} + \beta_3 \text{Num}^{\text{Shooting}}_{j,t} \text{KGR}_{j,t-1} + \mathbf{FE}(t) + \varepsilon_t,$$

where $r_{i,j,t}$ is the firm i's annul equity premium, the KGR is the previous year's inequality measure as capital gain ratio in region j, and indicator I^{Shooting} equals one if there is at least one mass shooting happens in region j in period t. $\text{Num}^{\text{Shooting}}$ is the logarithm of the mass shooting number. The firm's region is defined as the headquarters ' MSA location. In column (1) and (2), I present the regression result only with KGR and I^{Shooting} as the baseline case. In column (3) and (4), I employ the treated dummy and log(1+number of shooting), respectively, to investigate the interaction effect. In panel A, only mass shootings with at least 3 casualties are included in the test. In panel B, all mass shootings are included in the tests. I follow Hirshleifer and Shumway (2003) and Edmans, Garcia, and Norli (2007) to control for firm, industry, region and time fixed effects. Standard errors are clustered by firm, industry, and region. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Mass shooting with at least three casualties

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
KGR	-1.07*		-0.89	-0.88	-1.09*		-0.91	-0.90
	1.68		-1.60	-1.58	-1.68		-1.61	-1.58
I^{Shooting}		-3.59*	-1.77			-3.64*	-1.85	
		-1.77	-1.14			-1.77	-1.17	
$I^{\text{Shooting}} \times$ KGR			-5.07***				-5.00**	
			-2.60				-2.55	
$\text{Num}^{\text{Shooting}}$				-0.01				-1.01
				-0.56				-0.59
$\text{Num}^{\text{Shooting}} \times$ KGR				-7.53***				-7.43**
				-2.61				-2.50
Firm, time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ind, region FE	No	No	No	No	Yes	Yes	Yes	Yes
Obs (Firm-Year)	49707	49707	49707	49707	49707	49707	49707	49707

Panel B: Full sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
KGR	-1.07*		-0.89	-0.87	-1.09*		-0.91	-0.90
	1.68		-1.60	-1.57	-1.68		-1.61	-1.58
I^{Shooting}		-3.61*	-1.83			-3.67*	-1.90	
		-1.80	-1.19			-1.80	-1.22	
$I^{\text{Shooting}} \times$ KGR			-5.03***				-4.96**	
			-2.58				-2.53	
$\text{Num}^{\text{Shooting}}$				-0.79				-0.85
				-0.47				-0.50
$\text{Num}^{\text{Shooting}} \times$ KGR				-7.43**				-7.33**
				-2.50				-2.46
Firm, time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ind, region FE	No	No	No	No	Yes	Yes	Yes	Yes
Obs (Firm-Year)	49707	49707	49707	49707	49707	49707	49707	49707

Table 4 Reaction to inequality and mass shooting in county and CBSA region definitions

The table tests the regional stock market reaction to inequality information when a mass shooting happens using regional boundary as country and CBSA. The regression models are

$$r_{i,j,t} = \alpha + \beta_1 \text{KGR}_{j,t-1} + \beta_2 I^{\text{Shooting}}_{j,t} + \beta_3 I^{\text{Shooting}}_{j,t} \text{KGR}_{j,t-1} + \mathbf{FE}(t) + \varepsilon_t,$$

$$r_{i,j,t} = \alpha + \beta_1 \text{KGR}_{j,t-1} + \beta_2 \text{Num}^{\text{Shooting}}_{j,t} + \beta_3 \text{Num}^{\text{Shooting}}_{j,t} \text{KGR}_{j,t-1} + \mathbf{FE}(t) + \varepsilon_t,$$

where $r_{i,j,t}$ is the firm i's annul equity premium, the KGR is the previous year's inequality measure as capital ratio in region j, and indicator I^{Shooting} equals one if there is at least one mass shooting happens in region j. $\text{Num}^{\text{Shooting}}$ is the logarithm of the mass shooting number. In column (1) and (2), I present the regression results with county region boundary. In column (3) and (4), I present the regression results with CBSA region boundary. I follow Hirshleifer and Shumway (2003) and Edmans, Garcia, and Norli (2007) to control for firm, industry, region and time fixed effects. Standard errors are clustered by firm, industry and region. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	County		CBSA	
	(1)	(2)	(3)	(4)
KGR	-0.92	-0.92	-0.91*	-0.90
	-1.64	-1.63	-1.70	-1.64
I^{Shooting}	-1.44		-1.37	
	-1.05		-1.25	
$I^{\text{Shooting}} \times$ KGR	-1.44*		-1.60^{**}	
	-1.90		-2.10	
$\text{Num}^{\text{Shooting}}$		-0.96		-0.77
		-0.74		-0.67
$\text{Num}^{\text{Shooting}} \times$ KGR		-1.99*		-2.37**
		-1.71		-2.49
Firm, ind, region, time FE	Yes	Yes	Yes	Yes
Obs (Firm-Year)	49707	49707	49707	49707

Table 5 Inequality-return relationship in mass shooting treated counties and adjacent counties

The table presents a subsumple analysis on inequality-return relationship in mass shooting treated counties and adjacent counties. Specifically, I run the regression of annual KGR and subsequent return in two subsamples,

$$r^{\text{treated}}_{i,j,t} = \alpha + \beta_1 \text{KGR}^{\text{Treated}}_{j,t-1} + \mathbf{FE}(t) + \varepsilon_t,$$

$$r^{\text{adjacent}}_{i,j,t} = \alpha + \beta_1 \text{KGR}^{\text{Adjacent}}_{j,t-1} + \mathbf{FE}(t) + \varepsilon_t,$$

$$r_{i,j,t} = \alpha + \beta_1 I^{\text{Treated}} \text{KGR}^{\text{Treated}}_{j,t-1} + \beta_2 I^{\text{Adjacent}} \text{KGR}^{\text{Adjacent}}_{j,t-1} + \mathbf{FE}(t) + \varepsilon_t,$$

where $r_{i,j,t}$ is the firm i's annual equity premium, the KGR is the previous year's inequality measure as the capital gain ratio in the region j, the superscript indicates the observations are from mass shooting treated sample or adjacent samples. I follow Hirshleifer and Shumway (2003) and Edmans, Garcia, and Norli (2007) to control for firm, industry, region and time fixed effects. Standard errors are clustered by firm, industry, and region. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	Treated sample	Adjacent sample	Pooled sample
	(1)	(2)	(3)
$\text{KGR}^{\text{Treated}}$	-3.83**		
	-2.47		
$\text{KGR}^{\text{Adjacent}}$		-0.38	
		-0.52	
$I^{\text{Treated}}\text{KGR}^{\text{Treated}}$			-7.76**
			-2.53
$I^{\text{Adjacent}}\text{KGR}^{\text{Adjacent}}$			-1.03
			-1.21
Firm, ind, region, time FE	Yes	Yes	Yes
Obs (Firm-Year)	7161	14063	20,957
$H_0: \beta_1 = \beta_2$	-	-	Reject***
F-value	-	-	7.32
P-value	-	-	0.01

Table 6 Regime switching

The table tests the regional stock market reaction to inequality information when a mass shooting happens using the following model

$$r_{i,j,t} = \alpha + \beta_1 I_{j,t}^{\text{Untreated}} \text{KGR}_{j,t-1} + \beta_2 I_{j,t}^{\text{Shooting}} \text{KGR}_{j,t-1} + \mathbf{FE}(t) + \varepsilon_t,$$

where $r_{i,j,t}$ is the firm i's annul equity premium, the KGR is the previous year's inequality measure as capital gain ratio in region j, and indicator I^{Shooting} equals one if there is at least one mass shooting happens in region j. $\text{Num}^{\text{Shooting}}$ is the logarithm of the mass shooting number. The region boundary is defined as MSA boundary, county boundary and CBSA boundary. I follow Hirshleifer and Shumway (2003) and Edmans, Garcia, and Norli (2007) to control for firm, industry, region and time fixed effects. Standard errors are clustered by firm, industry and region. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Regression			
	MSA	County	CBSA
$I^{\text{Untreated}} \times$ KGR	-0.87	-0.90*	-0.88^{*}
	-1.61	-1.81	-1.83
$I^{\text{Shooting}} \times$ KGR	-6.70***	-2.99**	-3.01**
	-3.18	-2.25	-3.42
Firm, ind, region, time FE	Yes	Yes	Yes
Obs (Firm-Year)	49707	49707	49707
Panel B: Wald test			
$H_0 : \beta_1 = \beta_2$	Reject***	Reject**	Reject***
F-value	8.16	5.36	6.97
P-value	0.00	0.02	0.01

Table 7 Subsequent reaction to inequality and mass shooting

The table tracks the regional stock market reaction to inequality information after previous year's mass shooting using the following model

$$r_{i,j,t} = \alpha + \beta_1 \text{KGR}_{j,t-2} + \beta_2 I^{\text{Shooting}}_{j,t-1} + \beta_3 I^{\text{Shooting}}_{j,t-1} \text{KGR}_{j,t-2} + \mathbf{FE}(t) + \varepsilon_t,$$

$$r_{i,j,t} = \alpha + \beta_1 \text{KGR}_{j,t-2} + \beta_2 \text{Num}^{\text{Shooting}}_{j,t-1} + \beta_3 \text{Num}^{\text{Shooting}}_{j,t-1} \text{KGR}_{j,t-2} + \mathbf{FE}(t) + \varepsilon_t,$$

where $r_{i,j,t}$ is the firm i's annul equity premium, the KGR is the previous year's inequality measure as capital gain ratio in region j, and indicator I^{Shooting} equals one if there is at least one mass shooting happens in region j. $\text{Num}^{\text{Shooting}}$ is the logarithm of mass shooting number. The firm's region is defined as headquarter's MSA location. In column (1) and (2), I present the regression result only with KGR and I^{Shooting} as the baseline case. In column (3) and (4), I employ the treated dummy and log(1+number of shooting), respectively, to investigate the interaction effect. I follow Hirshleifer and Shumway (2003) and Edmans, Garcia, and Norli (2007) to control for firm, industry, region and time fixed effects. Standard errors are clustered by firm, industry and region. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)
KGR	-0.01		-0.91	-0.91
	-1.10		-1.17	-0.76
I^{Shooting}		0.00	-0.06	
		0.25	-0.57	
$I^{\text{Shooting}} \times$ KGR			1.99**	
			1.99	
$\text{Num}^{\text{Shooting}}$				-0.70
				-0.76
$\text{Num}^{\text{Shooting}} \times$ KGR				2.92**
				2.04
Firm, time FE	Yes	Yes	Yes	Yes
Ind, region FE	Yes	Yes	Yes	Yes
Obs (Firm-Year)	49707	49707	49707	49707

Table 8 Attention channel test

The table presents the difference in difference results of testing the relationship between public attention to inequality issue and mass shooting. I employ the similar DID method as Luca, Malhotra, and Poliquin (2019),

$$\mathrm{SVI}_{j,t}^{\mathrm{Inequality}} = \alpha + \beta_1 I_{t,j}^{\mathrm{Shooting}} + \gamma' X_{j,t} + \mathrm{FE}(t) + \varepsilon_t,$$

$$\mathrm{SVI}_{j,t}^{\mathrm{Inequality}} = \alpha + \beta_1 \mathrm{Num}_{t,j}^{\mathrm{Shooting}} + \gamma' X_{j,t} + \mathrm{FE}(t) + \varepsilon_t,$$

the $\mathrm{SVI}_{j,t}^{\mathrm{Inequality}}$ is the logarithm of the google search volume index about the income inequality issue in region j. $I_{t,j}^{\mathrm{Shooting}}$ and $\mathrm{Num}_{t,j}^{\mathrm{Shooting}}$ are mass shooting treatment indicator and logarithm of total shooting numbers, respectively. The $X_{j,t}$ represent a bunch of control variables that could potentially influence the google search volume about inequality. They are capital gain ratio (KGR), unemployment rate, GDP growth rate, population density and households density. $\mathrm{FE}(t)$ stands for the region and time fixed effects. Standard errors are clustered by , industry and region. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A: Annual	(1)	(2)	(3)	(4)
I^{Shooting}	0.56***	0.21***		
	7.53	2.92		
$\mathrm{Num}^{\mathrm{Shooting}}$			0.61***	0.27***
			11.99	4.15
KGR		0.29		0.49
		0.10		0.16
Unemployment rate		0.07***		0.07***
		3.03		3.33
GDP growth rate		0.05***		0.04**
		2.69		2.63
Population density		-2.64		-3.54
		-0.58		-0.78
Household unit density		15.32***		15.70***
		3.09		3.28
Region, time FE	Yes	Yes	Yes	Yes
Obs (Firm-Year)	541	541	541	541

Table 8 (contiuned)

The table presents

Panel A: Weekly	(1)	(2)	(3)	(4)
I^{Shooting}	0.52**	0.49**		
	2.62	2.45		
$\text{Num}^{\text{Shooting}}$			0.75**	0.71***
			2.62	2.45
KGR		3.22		3.22
		1.50		1.50
Unemployment rate		0.08		0.08
		1.31		1.31
GDP growth rate		-0.03		-0.03
		-1.38		-1.38
Population density		6.37		6.37
		1.24		1.24
Household unit density		-8.36		-8.36
		-1.34		-1.34
Region, time FE	Yes	Yes	Yes	Yes
Obs (Firm-Year)	4628	4628	4628	4628