Weak Identification of Long Memory with Implications for Volatility Modeling *

Jia Li[†], Peter C. B. Phillips[††], Shuping Shi[†††], Jun Yu[†]

[†]Singapore Management University

[††]Yale University, University of Auckland, Singapore Management University

[†††]Macquarie University

October 4, 2023

Abstract

This paper explores the implications of weak identification in common 'long memory' and recent alternative 'rough' approaches to modeling volatility dynamics of financial assets. Our analysis unveils asymptotic near observational equivalence between a long memory model with weak autoregressive dynamics and a rough model with anti-persistent shocks and a near-unit autoregressive root. A data-driven semiparametric and identification-robust approach to inference is developed, revealing the effect of these model ambiguities and documenting the prevalence of weak identification in many realized volatility and trading volume series. The forecasting performance and economic value of these two models are examined across a wide range of tradable financial assets.

Keywords: Realized volatility; Weak identification; Disjoint confidence sets; Trading volume; Long memory.

JEL classification: C12, C13, C58

*Thanks go to the Editor and two referees for searching comments and many useful suggestions that assisted the revision of this paper. The authors also thank Yacine Aït-Sahalia, Torben Andersen, Isaiah Andrews, Federico M. Bandi, Tim Bollerslev, Xu Cheng, Frank Diebold, Jim Gatheral, Peter Hansen, Morten Nielsen, Mathieu Rosenbaum, and Zhongjun Qu for helpful discussions and comments. Li acknowledges support from Singapore Ministry of Education (Grant No. 22-SOE-SMU-016). Shi acknowledges support from the Australian Research Council (Grant No. DE190100840). Yu acknowledges research/project support from the Ministry of Education, Singapore, under its Academic Research Fund (AcRF) Tier 2 (Award Number MOE-T2EP402A20-0002). Phillips acknowledges research support from the NSF (Grant No. SES 18-50860) and a Kelly Fellowship at the University of Auckland. Jia Li, School of Economics, Singapore Management University. Email: jiali@smu.edu.sg. Peter C.B. Phillips, Yale University, University of Auckland, & Singapore Management University; Email: peter.phillips@yale.edu. Shuping Shi, Department of Economics, Macquarie University; Email: shuping.shi@mq.edu.au. Jun Yu, School of Economics and Lee Kong Chian School of Business, Singapore Management University. Email: yujun@smu.edu.sg.

In the past five years ... a new family of models, known as rough volatility models, has sprung up ... Arguably, this is the breakthrough that volatility quants have been waiting for ... High-frequency market-makers such as Jump Trading have adopted the models. Hedge funds hint at using them in arbitrage trading strategies. Banks are scrutinising the approach. — *Risk.net 2/1/2021* [1]

1 Introduction

Since the seminal studies of Engle (1982) and Bollerslev (1986), the volatility literature seems to have reached a consensus concerning the presence of strong persistence in the volatility of financial assets. Nonetheless, the question of how best to model the persistence in volatility continues to be debated. Several approaches have been employed in what is now an extensive literature. Much of the research points to the presence of long range dependence or long memory in volatility data as a means of capturing volatility persistence.[2] Most recently, some evidence and industry practice has emphasized the importance of capturing 'rough volatility' in modeling, as the quotation that heads this article suggests.

Considerable research has been devoted to explain the long memory phenomenon in terms of more primitive generating mechanisms that have empirical justification. It is now known that mechanisms such as cross section aggregation (Robinson, 1978; Granger, 1980; Abadir and Talmain, 2002), structural breaks (Klemeš, 1974; Perron and Qu, 2010), trends (Bhattacharya et al., 1983), regime switching (Potter, 1976; Diebold and Inoue, 2001), learning (Alfarano and Lux, 2007; Chevillon and Mavroeidis, 2017), nonlinearity (Chen et al., 2010), marginalization (Chevillon et al., 2018), and networking (Schennach, 2018) can all generate long memory.

By combining short-run autoregressive (order p) and moving average (order q) components parametrically with fractional integration ($\mathrm{I}(d)$) to capture long range dependence, the class of ARFIMA(p, d, q) models has been widely used in empirical work to model economic time series that manifest both short and long memory as well as possible nonstationarity.

[1] Risk Awards 2021: Rough volatility models could make the options market more efficient.

[2] See Ding, Granger, and Engle (1993); Baillie, Bollerslev, and Mikkelsen (1996a,b); Comte and Renault (1996); Breidt, Crato, and De Lima (1998); Harvey (2007); Andersen and Bollerslev (1997); Andersen, Bollerslev, Diebold, and Labys (2001, 2003); Andersen, Bollerslev, Diebold, and Ebens (2001), among others.

The impulse response function implied by this general model with fractional integration is substantially different from that of an ARMA(p, q) model, particularly allowing for long slow decays in responses when the fractional parameter d is positive. This feature of impulse responses in volatility has important implications for financial decision makings, such as volatility forecasting, portfolio choice, and option pricing. Within the class of ARFIMA(p, d, q) models, the ARFIMA($1, d, 0$) model has been found to be especially useful and a leading example that motivates the present paper is the stochastic volatility of financial assets. We refer to the autoregressive coefficient of the ARFIMA($1, d, 0$) model as α in the subsequent discussions.

Many statistical procedures are now available for the estimation of the memory parameter d in both the ARFIMA parametric class and various semiparametric classes that do not prescribe autoregressive or moving average specifications.[3] Estimated values of d from log daily realized volatility (RV) data usually turn out to be positive and close to the non-stationary boundary 0.5, which implies long memory dynamics, and the first order autoregressive coefficient estimates are found to be near zero (Andersen and Bollerslev, 1997; Andersen, Bollerslev, Diebold, and Ebens, 2001; Andersen, Bollerslev, Diebold, and Labys, 2001, 2003; Shi and Yu, 2023).

The idea that underlies the long memory specification is well understood by both financial analysts and informed market participants: news stories and their financial market impact tend to be 'remembered' by the market for a long time, longer than in typical GARCH-type models. In order to determine the current volatility level and investment opportunities, one needs to examine shocks that extend from the distant past up to the current moment, because even 'old' news carries information about the market mechanism and can still have its own distinct impact on current volatility. This long memory idea is intuitively appealing, as investors routinely review historical events before trading or consider long term behavior such as cyclically adjusted price-earnings ratios, which drive volatility and volume; and market commentators and financial analysts frequently make similar connections.

Interestingly and as attested in the lead quotation, a new body of empirical evidence

[3] See, for example, Künsch (1987); Robinson (1995a); Geweke and Porter-Hudak (1983); Robinson (1995b); Shimotsu and Phillips (2005).

points towards the same ARFIMA$(1, d, 0)$ model structure but with a negative value for d (so-called antipersistence) in log RV, producing what is called 'rough-volatility' with an autoregressive parameter taken to be unity or near unity.[4] Amongst many studies that support such a model we mention the following: Gatheral et al. (2018); Bayer et al. (2016); Wang et al. (2023b); Bolko et al. (2023); Liu et al. (2020); Fukasawa et al. (2022). Although still in its early stages of development, some attempts have been made within this body of literature to comprehend its underlying mechanisms. For example, El Euch et al. (2018a) show how, in a highly endogenous market, roughness in volatility is generated by large sets of split orders. In a more general framework, Jusselin and Rosenbaum (2020) show that the no-arbitrage condition implies rough volatility.

Rough volatility modeling has received considerable attention in the financial industry and financial engineering as well as in academic research in quantitative finance, mathematical finance, and financial econometrics.[5] Notably, the 2021 Risk Awards were presented for introducing rough-volatility models – see the Risk website for the citation;[6] Based on evidence of roughness in volatility, many studies have conducted financial applications using rough volatility models. See, for example, Bayer et al. (2016); Garnier and Sølna (2017, 2018), Fouque and Hu (2018) in portfolio choice, and El Euch et al. (2018b) in dynamic hedging. The profound implications of rough volatility models for asset pricing was detailed in an article by Jean-Philippe Bouchaud, Chairman of Capital Fund Management. According to Bouchaud (2020), one of recent rough specifications, the quadratic rough Heston model of Gatheral et al. (2020), effectively resolves a long-standing puzzle by jointly calibrating the volatility smile of the S&P 500 and VIX options – a challenging task for quantitative analysts for many years. Many hedge funds and non-bank market-makers claim to have adopted rough volatility models (Risk Staff, 2021).

Under the rough ARFIMA$(1, d, 0)$ model, volatility determination operates through a mechanism distinct from that of the long memory model: news shocks from the past can

[4]The rough-volatility literature, pioneered by Gatheral et al. (2018), uses a class of continuous-time models driven by fractional Brownian motion with Hurst parameter H, whose discrete-time representations are asymptotically equivalent to the ARFIMA$(1, d, 0)$ model with $H = d + 1/2$; see Tanaka (2013); Wang and Yu (2023).

[5]The Rough Volatility website `https://sites.google.com/site/roughvol/home/risks-1` provides a collection of over 200 papers in this rapidly growing literature.

[6]`https://www.risk.net/awards/7736196/quants-of-the-year-jim-gatheral-and-mathieu-rosenbaum.`

be effectively summarized by the immediately preceding volatility; but volatility is extended outwards with a slight depreciation since the autoregressive coefficient is near but below unity. Importantly, in addition, the day-to-day shocks, i.e., volatility innovation, are strongly negatively autocorrelated, so that an upward movement of volatility tends to be followed by a downward drop. The resulting 'zigzag' pattern gives rise to the so-called roughness, and the spikes in the volatility path tend to reflect the granularity of the information arrival process.

These two groups of empirical findings on volatility dynamics seem contradictory. Yet they reveal that this simple parametric time series model has dual capabilities of matching the RV data. Only two parameters in the model, the autoregressive parameter α and the memory parameter d, control dependency. While both groups of empirical literature concur that past shocks have a long-lasting influence on volatility, they diverge in how this effect is captured in the models. In the long memory model, this impact is captured by a positive fractional parameter d (accompanied by small and often negative values for the autoregressive coefficient α). On the other hand, the rough volatility model encapsulates this influence with α close to unity (coupled with a rough fractional parameter, i.e., $d < 0$). Importantly, the different parameter configurations may have distinct implications for long-term volatility forecasting, thereby influencing portfolio allocation, as demonstrated later in the paper. Such distinction also holds significance for asset pricing practice as noted by Bouchaud (2020) among others.

Inspired by what is now an extensive literature on weak identification,[7] we recognize that the contradictory findings may be symptomatic of weak identification in the fitted ARFIMA$(1, d, 0)$ model for this type of economic data. To clarify these findings in the present case we show that for two well isolated local parameterizations the model-implied spectral densities are nearly indistinguishable. The two parameterizations correspond, as described above, to cases where the autoregressive coefficient lies either near unity or near zero.[8] More specifically, we show that the 'distance' between a near-unit-root ARFIMA$(1, d, 0)$ model and

[7] See, Phillips (1989); Staiger and Stock (1997); Stock and Wright (2000); Stock and Yogo (2005); Andrews and Cheng (2012); Andrews et al. (2019); Andrews and Mikusheva (2022); Cheng et al. (2022), among others.

[8] Precise definitions of 'near unity' and 'near zero' involve sample size dependencies as commonly used in the time series literature and these are discussed explicitly in Section 2 of the paper.

a near-zero-root ARFIMA$(1, d+1, 0)$ model goes to zero when nearness parameters shrink. The two seemingly distinct parameterizations that have been studied extensively in the literature therefore generate observationally *nearly* equivalent dynamics.[9] The consequences of such observationally near equivalence have been well studied in other contexts and it is known, for instance, that methods of inference based on conventional (identified) asymptotic theory lead to major finite sample distortions under identification failure (Phillips, 1989; Dufour, 1997) that can include unbounded confidence intervals.[10] Against the background of these findings some related phenomena involving inferential distortions are to be expected in the present context and have been documented in recent work by Shi and Yu (2023).

To address this issue we propose using an identification-robust confidence set of the model parameters obtained by inverting tests for zero serial correlation in the model-implied residuals. The inference procedure is semiparametric, data-driven, and do not rely on Gaussian errors. Consonant with theory, simulations show that the robust confidence sets generally 'bifurcate' in the sense that they include two distinctly isolated regions in which either (i) the autoregressive parameter is close to unity and the memory parameter is negative or (ii) the autoregressive parameter is close to zero and the memory parameter is positive.

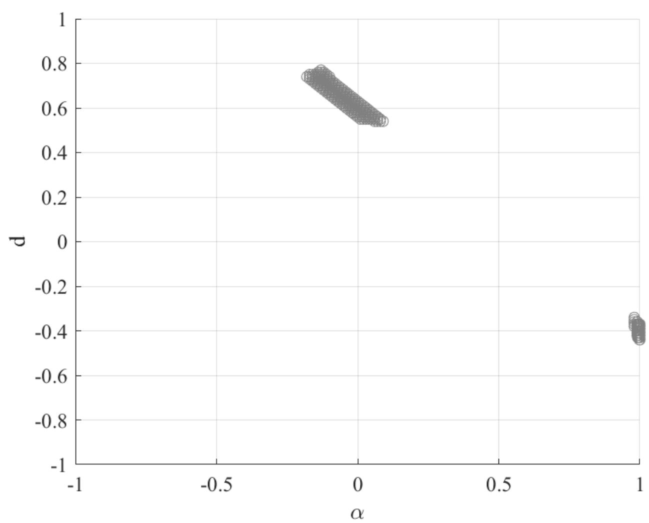
Real data studies are undertaken to explore how prevalent this empirical phenomenon is in practical work with financial data. We report results for a large number of realized volatility and trading volume time series for a broad variety of U.S. equities and international stock market indices. Our findings indicate that identification-robust confidence sets often do bifurcate, exhibiting precisely the same pattern observed in simulations. Figure 1 provides an illustration. This bifurcation highlights the empirical difficulty of determining whether or not a time series is driven by long-memory disturbances with $d > 0$, a property that is highly relevant for pricing applications, one that can improve understanding of network structures within an economic system (Schennach, 2018), and one that has implications for forecasting. To shed more light on this question we draw on insights from the mixture-of-distribution hypothesis (Clark, 1973; Tauchen and Pitts, 1983; Andersen, 1996), which postulates that

[9] This does not imply that the two models are identical; in fact, as we discuss later in the paper, their spectral densities differ at near-zero frequencies and there are differences in their autocorrelation functions.

[10] Duffy and Kasparis (2021) have discovered asymptotic affinities between time series with long memory parameter at the nonstationary boundary $d = 0.5$ and the class of mildly integrated processes with roots near unity studied in Phillips and Magdalinos (2007).

price volatility and trading volume are driven by underlying information flows. Applying our robust inference methodology to the Refinitiv buzz indices for U.S. stocks, which are textual-based measures of information flows derived from news and social media outlets, we find that inferences concerning news flow do not appear to be affected by weak identification, and the resulting confidence sets of the model parameters support the long-memory specification. Our empirical findings suggest that the recently documented statistical evidence for volatility roughness in continuous-time models may need re-examination and an empirical threat of weak identification cannot be ignored.

Figure 1: Confidence Set for the S&P 500 Index Exchange Traded Fund (SPY)



Identification-robust inference was applied to the (demeaned) log realized volatility of SPY from 1996 to 2021. The x-axis displays the autoregressive coefficient α and the y-axis shows the fractional parameter d. The procedure identifies two isolated regions for (α, d) at the 95% confidence level: (1) α is close to zero and $d > 0$ (long memory); (2) α is near unity and $d < 0$ (rough).

Although weak identification suggests that the two models—long memory and rough—are difficult to distinguish statistically based on observed autocovariance structures, it does not necessarily lead to equivalent performance in specific practical applications, because a practitioner's loss or reward is often not aligned with the econometrician's statistical measure of model fitting. In fact, the weak identification phenomenon highlights the lack of strong convincing statistical evidence for either model and instead the potential relevance of elements of both models. So the relative rankings of the models are more likely to be context specific. The cited recent literature on rough volatility models tells the story from one side, demonstrating the rough model merits in pricing options. To shed light from an alternative perspective, we conduct an empirical analysis in a forecast comparison context. We examine the forecasting performance of the two models at various horizons for a large set of real data on realized volatility (as well as simulated data) and find that the long memory model is unequivocally superior to the rough model in forecasting exercises. The main takeaway of our analysis is this: to judge the relative merit of the two models, it is essential to articulate the applied scenario and be especially careful when extrapolating lessons learned from one context to another. Pure statistical evidence gained from standard econometric methods is unlikely to be sufficiently informative because of the fundamental difficulty in resolving the weak identification and associated model ambiguity that is emphasized in the present research.

The paper is organized as follows. Section 2 reviews model specifications, details nearness measures between models, and shows how weak identification manifests. This section also explains how to invert tests for zero residual serial correlation, providing Anderson–Rubin type confidence sets for model parameters; and simulation results were conducted and are reported online that explore the relevance of weak identification by examining the identification-robust confidence sets. Section 3 reports extensive empirical studies using RV and trading volume series for many assets and news and social media based measures of information flows. Section 4 compares the forecasting accuracy and economic value of the two competing models (long memory and rough), benchmarked against other popular volatility models. Section 5 concludes. An internet appendix presents proofs, comprehensive Monte Carlo results concerning weak identification and forecasting performance, details about the data and forecasting methodology, additional empirical findings, and robustness checks.

2 The Econometric Method

This section describes the identification-robust inference methods used in our simulation and empirical analysis. Section 2.1 provides a brief background on ARFIMA processes and Section 2.2 explains the weak-identification issue under study. The procedure for the construction of identification-robust confidence sets is given in Section 2.3.

2.1 Fractionally integrated processes

We start with introducing the econometric model. Let L denote the lag operator. The observed time series y_t is modeled as an ARFIMA$(1, d, 0)$ process:

$$(1-\alpha L)\, y_t = u_t, \qquad u_t = \sigma\,(1-L)^{-d}\,\varepsilon_t, \tag{2.1}$$

where α is the autoregressive coefficient, u_t is a fractionally integrated process with memory parameter d, $\sigma > 0$ is a scale parameter and ε_t is a stationary martingale difference sequence (MDS) with unit variance. In the stationary case where $d \in (-0.5, 0.5)$ the fractional operator in (2.1) can be defined by binomial series expansion as

$$(1-L)^{-d} = \sum_{j=0}^{\infty} \binom{-d}{j} (-L)^j = \sum_{j=0}^{\infty} \frac{(d)_j}{j!} L^j \tag{2.2}$$

giving $u_t = \sigma\,(1-L)^{-d}\,\varepsilon_t = \sigma \sum_{j=0}^{\infty} \frac{(d)_j}{j!}\varepsilon_{t-j}$. In (2.2), $(d)_j = d(d+1)...(d+j-1) = \frac{\Gamma(d+j)}{\Gamma(d)}$ is a forward factorial and $\Gamma(\cdot)$ is the gamma function. In nonstationary cases where $d \geq 0.5$ initial conditions are set to a fixed origin such as $t = 0$ and the series is truncated giving $u_t = \sigma\,(1-L)^{-d}\,\varepsilon_t\, \mathbf{1}\{t \geq 1\} = \sigma \sum_{j=0}^{t-1} \frac{(d)_j}{j!}\varepsilon_{t-j}$ (Phillips, 1999; Shimotsu and Phillips, 2005). When $d = 0$ the series reduces to the identity and $u_t = \sigma\varepsilon_t$. We denote the parameter of interest by $\theta = (\alpha, d)$ and the variance σ^2 is treated as a nuisance parameter.

In our empirical work the observed series y_t may be (after demeaning) a volatility proxy, trading volume, or textual measures of news flow. While these series are highly persistent, they evidently do not wander without bounds as random walks. We therefore focus on the empirically relevant scenario by restricting the autoregressive coefficient to $|\alpha| < 1$. When $0 < |d| < 0.5$, the innovation u_t is stationary (Granger and Joyeux, 1980; Hosking, 1981) with autocorrelation function (acf) at lag k

$$\rho_u(k) \;=\; \frac{(-d)!\,(k+d-1)!}{(d-1)!\,(k-d)!} \sim_a \frac{(-d)!}{(d-1)!}\frac{1}{k^{1-2d}} \text{ as } k \to \infty, \tag{2.3}$$

which decays at a polynomial rate (compared with the exponential rate of a stationary ARMA model) as $k \to \infty$. The spectral density of y_t is

$$f_\theta(\lambda) = \frac{\sigma^2}{2\pi} \frac{[2 - 2\cos(\lambda)]^{-d}}{1 - 2\alpha\cos(\lambda) + \alpha^2} \text{ for } -\pi \leq \lambda \leq \pi, \tag{2.4}$$

which encodes the dynamics of the observed series y_t.[11]

The sign of d determines whether the fractional process u_t has long or short memory. McLeod and Hipel (1978) define a stationary process as having a long (resp. short) memory if its acf is not summable (resp. summable). Hence, u_t has long memory when $d > 0$ and short memory when $d \leq 0$. The memory parameter d in u_t relates to the Hurst parameter H in the increment of fractional Brownian motion (fBM) through the relationship $d = H - 1/2$; see Giraitis et al. (2012, chap. 3).[12] In continuous time models driven by fBM increments, the long memory (resp. short memory) setting corresponds to $H > 1/2$ (resp. $H \leq 1/2$). The Hurst index H controls the smoothness of the sample path of fBM and the process has 'rough' paths when $H \in (0, \frac{1}{2})$.

The empirical literature on volatility modeling has yielded apparently conflicting results on the memory parameter d (which we identify with its continuous-time analogue H). For example, Comte and Renault (1998) found that $d \approx 0.25$ in a continuous-time fBM-driven model, and Andersen et al. (2003) estimated $d \approx 0.4$ in a discrete-time ARFIMA$(1, d, 0)$ model. In sharp contrast, the recent literature on 'rough volatility' starting with Gatheral et al. (2018) provides empirical evidence for $d < 0$, where the typical estimated value of d is close to -0.5. The corresponding long memory and rough sample path models can have different implications for volatility forecasting and option pricing. The conflicting empirical findings are surprising because the long memory property of volatility has been deemed a stylized fact. This in turn has stimulated an active area of research in the recent financial econometrics literature.

Besides its long-run implications, the distinction between long memory and rough-volatility

[11] The spectral density is divergent with a fractional pole at the zero frequency when $d > 0$. When $d \geq 0.5$ and y_t is nonstationary, $f_\theta(\lambda)$ is no longer integrable but is still defined for $\lambda \neq 0$ (Solo, 1992; Velasco and Robinson, 2000).

[12] For the same reason, the d and α parameters in an ARFIMA$(1, d, 0)$ model correspond to two parameters in the fractional Ornstein–Uhlenbeck process; see Tanaka (2013); Wang et al. (2023a).

models is also extremely important for the large literature on high-frequency-based nonparametric volatility estimation, as most of the existing work in that literature requires (in a stochastic sense) sufficient smoothness in the volatility path that is incompatible with the rough-volatility model. For instance, nonparametric estimation of spot volatility (e.g., over an event window before or after a critical news announcement) requires a small bandwidth to reduce bias when volatility is 'rough', which produces a slow optimal rate of convergence. In the boundary case with d approaching -0.5, the underlying fBM process is barely continuous and the convergence rate becomes arbitrarily close to zero even with optimal tuning. This in turn would severely limit the use of the high-frequency identification strategy (Nakamura and Steinsson, 2018a,b) based on heteroskedasticity (Rigobon, 2003), or high-frequency regression-discontinuity designs (Bollerslev et al., 2018).

Motivated by these considerations, we aim to shed new light on long memory versus rough-volatility empirical issues. While the aforementioned empirical studies have focused on alternative ways of estimating the fractional parameter d, we ask a more fundamental question: whether this parameter is strongly or weakly identified along with the companion autoregressive coefficient α. If these parameters are weakly identified, standard econometric inference may be severely distorted, and identification-robust inference is required to reveal the underlying ambiguities in inference.

2.2 The weak identification problem

Why are the fractional parameter d and the autoregressive parameter α weakly identified? To guide intuition, note that the ARFIMA$(1, d, 0)$ model with $\alpha = 1$ is observationally equivalent to the ARFIMA$(1, d+1, 0)$ model with $\alpha = 0$. That is,

$$(1-L)y_t = \sigma\,(1-L)^{-d}\,\varepsilon_t \iff y_t = \sigma\,(1-L)^{-(d+1)}\,\varepsilon_t. \tag{2.5}$$

Thus, for any $d \in \mathbb{R}$ there is identification failure between the two configurations $(\alpha, d) = (1, d)$ and $(\tilde{\alpha}, \tilde{d}) = (0, d+1)$. This failure is clearly specific to $\alpha = 1$, as the $1 - \alpha L$ operator can only be absorbed into the differencing filter $(1-L)^{-d}$ when $\alpha = 1$. Failure manifests here in a separable manner as these two isolated points on the parameter space become observationally equivalent.

This simple identification failure in the ARFIMA model may appear irrelevant if the unit autoregressive root $\alpha = 1$ is ruled out *a priori*. But such a restriction does not prevent *weak identification* when α is near unity and there is near observational equivalence in the two structures. For whenever α is close to unity (and $\tilde{\alpha}$ close to zero), a breakdown of identification between (α, d) and $(\tilde{\alpha}, \tilde{d}) = (0, d+1)$ holds approximately. In particular, a 'rough' parametric configuration with $d \approx -0.5$ is observationally nearly equivalent to a 'long memory' configuration with $d \approx 0.5$, provided that the autoregressive coefficients α and $\tilde{\alpha}$ are adjusted accordingly.

This weak identification issue is qualitatively distinct from the 'common root' identification failure in ARMA models. In that setting common AR and MA roots are well known to lead to identification failure (Ansley and Newbold, 1980) and the related weak identification issue has been studied in detail for stationary ARMA models by Andrews and Cheng (2012). Identification failure in ARMA models can arise for any corresponding AR/MA parameter values in the parameter space. In contrast, weak identification in the present ARFIMA setting is specific to the joint 'near-unity and near-zero' scenario for the AR coefficient. Further, when weak identification occurs it manifests as a discrete 'phase transition' between one set of parameters $(\alpha, d) = (1, d)$ and the other $(\tilde{\alpha}, \tilde{d}) = (0, d+1)$. Complications related to unit-root asymptotics also prevent any application of the common root ARMA weak identification analysis in the present setting.

The weak identification issue considered here is related to but also distinct from the well known long memory estimation bias phenomenon in which both Gaussian maximum likelihood and semiparametric Whittle estimates of long memory exhibit large finite sample bias in the presence of a substantial autoregressive component. This bias problem was shown in early simulations in Agiakloglou et al. (1993) and bias correction methods were considered in subsequent research, e.g., Andrews and Guggenberger (2003) and Poskitt et al. (2017).

To fix ideas we now formalize the intuition on weak identification by quantifying the 'distance' between two isolated local ARFIMA models. Let $d^* \in (-1, 0)$ be a fixed constant. We consider two models indexed by the following local parameter regions: for some positive sequences $\gamma_T = o(1)$, $\tilde{\gamma}_T = o(1)$, and $\eta_T = O(1)$ as $T \to \infty$, define the regions

$$\begin{cases} R_T = \{(\alpha_T, d_T) : |\alpha_T - 1| < \gamma_T, |d_T - d^*| < \eta_T\}, \\ \widetilde{R}_T = \{(\tilde{\alpha}_T, \tilde{d}_T) : |\tilde{\alpha}_T| < \tilde{\gamma}_T, |\tilde{d}_T - d^* - 1| < \eta_T\}. \end{cases} \tag{2.6}$$

Note that R_T and $\widetilde{R}_T$ are respectively near the identification-failure points $(1, d^*)$ and $(0, d^* + 1)$ in the autoregressive dimension (i.e. α), shrinking at rates γ_T and $\tilde{\gamma}_T$; we only require $\gamma_T \to 0$ and $\tilde{\gamma}_T \to 0$ without setting any specific rates on these sequences. Whether the width η_T of these regions along the fractional dimension (i.e., d) shrinks to zero is not essential for our analysis, so it is kept unspecified. It is convenient, but not essential, to think of these sequences as depending on the sample size $T \to \infty$. In such cases, the formulation subsumes a large class of near unity and near zero local parameterizations that have been used in the econometric literature, as described in footnote 14 below.

Since the dynamics implied by each parameter vector $\theta = (\alpha, d)$ is summarized by the spectral density $f_\theta(\cdot)$, the two local models may be represented as the corresponding collections of spectral densities, $\mathcal{M}_T = \{f_\theta(\cdot) : \theta \in R_T\}$ and $\widetilde{\mathcal{M}}_T = \{f_\theta(\cdot) : \theta \in \widetilde{R}_T\}$. This definition mirrors the usual definition of a statistical experiment as a collection of probability laws, but our focus is more specifically on the dynamics, with other model ingredients treated as nuisance. To quantify the distance between the two local models, we define the deficiency of $\widetilde{\mathcal{M}}_T$ with respect to $\mathcal{M}_T$ as

$$\delta\big(\widetilde{\mathcal{M}}_T, \mathcal{M}_T\big) \equiv \sup_{\theta \in R_T} \inf_{\tilde{\theta} \in \widetilde{R}_T} \sup_{\underline{\lambda}_T \leq |\lambda| \leq \pi} \left|\log f_\theta(\lambda) - \log f_{\tilde{\theta}}(\lambda)\right|,$$

where the lower bound $\underline{\lambda}_T > 0$ may possibly shrink to zero.[13] The idea under this definition is, for any $\theta \in R_T$ under the model $\mathcal{M}_T$, one can find $\tilde{\theta} \in \widetilde{R}_T$ under the other model $\widetilde{\mathcal{M}}_T$, such that the uniform distance between their spectral densities (in log form) is bounded by $\delta(\widetilde{\mathcal{M}}_T, \mathcal{M}_T)$. The deficiency measure thus quantifies the extent to which the dynamics generated by $\mathcal{M}_T$ cannot be captured by $\widetilde{\mathcal{M}}_T$. Symmetrizing the roles of $\mathcal{M}_T$ and $\widetilde{\mathcal{M}}_T$, we

[13] The $\underline{\lambda}_T$ lower bound is needed to properly define the uniform distance between spectral densities for ARFIMA models because these densities have fractional poles at frequency zero. When $d_T \neq \tilde{d}_T$ the fractional asymptotes differ and the lower bound $\lambda_T \to 0$ controls the rate at which comparisons are made in the relative differences between the spectral densities as $T \to \infty$.

can then gauge the distance between the two local models using

$$\Delta\big(\mathcal{M}_T, \widetilde{\mathcal{M}}_T\big) = \max\Big\{\delta\big(\widetilde{\mathcal{M}}_T, \mathcal{M}_T\big), \delta\big(\mathcal{M}_T, \widetilde{\mathcal{M}}_T\big)\Big\},$$

which is, in fact, the Hausdorff distance between the $\mathcal{M}_T$ and $\widetilde{\mathcal{M}}_T$ sets of functions induced by the local uniform metric on the space of spectral density functions. When the distance between the two models is zero, they generate exactly the same spectral densities. Theorem 1 shows that this equivalence nearly holds for $\mathcal{M}_T$ and $\widetilde{\mathcal{M}}_T$.

Theorem 1 *Let R_T and $\widetilde{R}_T$ be defined as (2.6) for some positive sequences $\gamma_T = o(1)$, $\tilde{\gamma}_T = o(1)$, and $\eta_T = O(1)$. Then, $\Delta(\mathcal{M}_T, \widetilde{\mathcal{M}}_T) = O((\underline{\lambda}_T^{-2}\gamma_T) \vee \tilde{\gamma}_T)$, where $\vee$ is the supremum operator.*

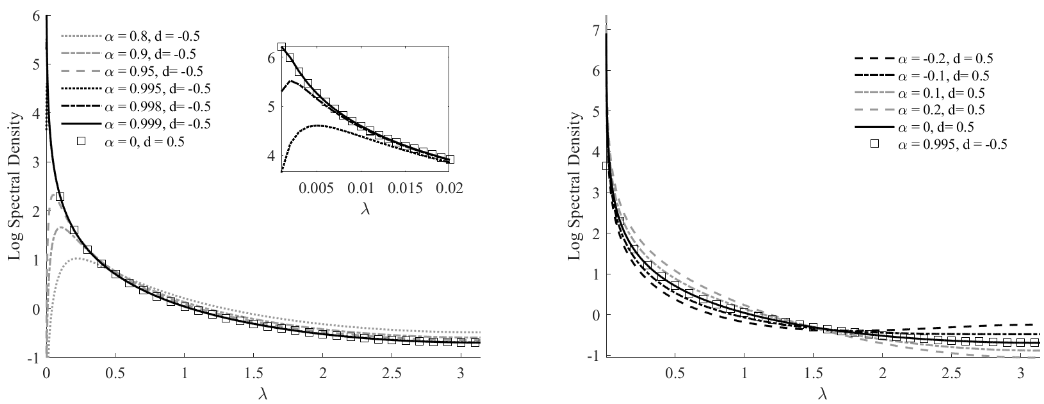
This result formally clarifies the weak identification issue in the ARFIMA context. It shows that the $\Delta(\mathcal{M}_T, \widetilde{\mathcal{M}}_T)$ distance between the two local models, parameterized by R_T and $\widetilde{R}_T$, asymptotically shrinks to zero when the former is near-unity and the latter is near-zero (in the autoregressive dimension).[14] This leads to a rather severe form of weak identification because the two sets of parameters R_T and $\widetilde{R}_T$ are *not* close to each other, as they are centered around the two isolated points $(1, d^*)$ and $(0, d^* + 1)$ in the (α, d) plane. As γ_T and $\tilde{\gamma}_T$ approach zero, these two regions become further apart in the parameter space but, as shown in Theorem 1, the difference between their dynamic implications also vanishes, provided the lower frequency bound $\underline{\lambda}_T$ does not tend to zero too fast, that is, faster than $\sqrt{\gamma_T}$. As such, weak identification arises in a 'bimodal' form, with two distinct sets of parameters being observationally nearly equivalent.

To illustrate this point, Figure 2 plots the log spectral densities of the ARFIMA model under these two local models. In panel (a), the configuration with $\alpha = 0$ and $d = 0.5$ belongs to $\widetilde{\mathcal{M}}_T$, while the others fall in $\mathcal{M}_T$ with α ranging from 0.8 to 0.999 and d fixed at -0.5. Similarly, in panel (b), the configuration $\alpha = 0.995$ and $d = -0.5$ falls in $\mathcal{M}_T$, while the spectral densities for the remaining parameter settings are in $\widetilde{\mathcal{M}}_T$ with α between -0.2 and

[14] The near-unity restriction covers a wide spectrum of near unit root behavior, including the local-to-unity specification of Phillips (1987) and Chan and Wei (1987), the mildly integrated specification of Phillips and Magdalinos (2007), and the general near-unity specification of Phillips (2023).

Figure 2: Log Spectral Densities of Two Local Models: An Illustration

(a) (b)



The spectral density of the ARFIMA$(1, d, 0)$ model is given in (2.4). Panel (a) shows that when $(\alpha, d) \in \mathcal{R}_T$, as α moves closer to unity, the log spectral density of ARFIMA$(1, d, 0)$ approaches that of $(0, 1+d) \in \widetilde{\mathcal{R}}_T$. Panel (b) illustrates the convergence in the opposite direction: when $(\alpha, d) \in \widetilde{\mathcal{R}}_T$, the log spectral density of ARFIMA$(1, d, 0)$ converges to that of $(\alpha = 0.995, d-1) \in \mathcal{R}_T$ as $\alpha \to 0$.

0.2. Evidently, as $\alpha \to 1$ (resp. $\alpha \to 0$), the log spectral density generated from $\mathcal{M}_T$ (resp. $\widetilde{\mathcal{M}}_T$) approaches and eventually becomes virtually indistinguishable from that associated with $\widetilde{\mathcal{M}}_T$ (resp. $\mathcal{M}_T$), revealing the weak identification between them, subject to the lower frequency bound $\underline{\lambda}_T$ not passing to zero so fast that the different order of the fractional poles dominates the discrepancy in the spectral densities. To further appreciate the impact of $\underline{\lambda}_T$ on the discrepancy measure, we show in the inset in panel (a) an enlarged graphic of the log spectral densities focused at frequencies closer to zero, ranging between 0.001 and 0.02. For a given $\theta \in \tilde{R}_T$ (panel (a)), the uniform distance between the two spectral densities is affected by the lower bound of λ but diminishes rapidly provided $\underline{\lambda}_T$ does not pass to zero too fast. When θ is in R_T, the densities are very close (panel (b)). In this case, the impact of small negative autoregressive coefficients $\alpha < 0$ under long memory with $d = 0.5$ also raises spectral power at high frequency.

This weak-identification perspective on ARFIMA specifications provides a plausible explanation for the conflicting empirical findings in the literature regarding long memory or roughness in volatility dynamics. In the empirical rough volatility literature when α_T is assumed to be unity or local to unity the estimated value of d is negative and often close to

-0.5 (Gatheral et al., 2018; Fukasawa et al., 2022; Wang et al., 2023b; Bolko et al., 2023).[15] This corresponds to the R_T region in our analysis. As discussed above, such a parametric configuration is essentially indistinguishable from its counterpart in $\widetilde{R}_T$ with d around 0.5 and α_T near zero. The latter parameter values are actually in line with the estimates reported in the long memory RV literature reviewed in the Introduction.

If weak identification is indeed in force, conventional asymptotic inference based on strong identification may be unreliable. This explains why Shi and Yu (2023) find two disjoint intervals in the highest density set for the time-domain maximum likelihood estimators and frequency-domain maximum likelihood estimators. Similar effects have been extensively studied in the literature on weak instrumental variables (Staiger and Stock (1997), Moreira (2003)) and more generally in the setting of weak GMM (Stock and Wright (2000), Andrews and Mikusheva (2022)). Andrews and Cheng (2012) analyze the weak identification problem in a broad range of problems, including weakly identified stationary ARMA models (Ansley and Newbold (1980)). A key lesson from the weak identification literature is this: if the strength of identification is in doubt, it is better to apply inferential methods that are robust to identification failure. This idea motivates the approach we now propose.

2.3 Identification-robust confidence sets

Theory suggests that parameters α and d are jointly weakly identified when α is near-unity or near-zero. To prevent weak identification from distorting statistical inference, we now construct identification-robust confidence sets for $\theta = (\alpha, d)$. Using a standard approach from the weak identification literature we construct Anderson–Rubin confidence sets by inverting tests for null hypotheses of the form $H_0 : \theta_0 = \theta$, where θ_0 denotes the true parameter value and θ denotes a generic candidate parameter that runs over the parameter space Θ. Specifically, equipped with a test that has asymptotic size β and following Anderson and Rubin (1949), the associated $1 - \beta$ level confidence set is constructed as the collection of all

[15] The discrete-time representation of fBM implies that α_T is unity while the discrete-time representation of fOU under an infill scheme implies that α_T is local to unity.

non-rejected parameter values, viz.,

$$\mathrm{CS}_{1-\beta} = \{\theta \in \Theta : \text{The null hypothesis } H_0 : \theta_0 = \theta \text{ is not rejected at level } \beta\}. \tag{2.7}$$

The remaining task is to construct a test that is robust to weak identification. Conventional tests derived from (quasi) maximum likelihood or GMM estimators are not suitable for this task, because the classical inferential theory relies heavily on strong identification. We instead consider a test that targets moment conditions implied by the null hypothesis $\theta_0 = \theta$.[16] Under the null the θ-implied disturbance term $\varepsilon_t(\theta) \equiv (1-L)^d (y_t - \alpha y_{t-1})$ coincides with the true ε_t error term and forms an MDS. This in turn implies

$$H_0^{WN} : \gamma_j(\theta) = 0 \text{ for all } j \geq 1, \tag{2.8}$$

where $\gamma_j(\theta)$ denotes the autocovariance of $\varepsilon_t(\theta)$ of order j.[17] We may test the original null hypothesis $H_0 : \theta_0 = \theta$ by testing the moment conditions in (2.8), viz., $\varepsilon_t(\theta)$ forms a white noise sequence for the candidate parameter value θ.

It is worth clarifying that the 'white-noise' null hypothesis H_0^{WN} does not fully exhaust the model restrictions implied by the maintained stationary MDS assumption on ε_t. By testing the weaker null hypothesis H_0^{WN}, we intentionally direct test power towards the detection of non-zero serial correlations rather than general forms of nonlinear serial dependence (which are not intended to be captured by the ARFIMA model). Evidently, this technical gap would not have appeared if we had assumed from the outset that the ε_t 'only' comprised white noise. We adopt the stationary MDS structure for technical convenience, which is common in the literature, because it simplifies the computation of the test statistic.

Our proposal for constructing identification-robust confidence sets is simply to invert tests for zero serial correlation in the implied disturbance. Testing for serial correlation is a well studied topic in time series analysis. We can therefore address weak identification in

[16] This idea is similar in spirit to the Anderson-Rubin test proposed by Chevillon et al. (2010) dealing with inferences in a structural model with strongly persistent data.

[17] When dealing with an ARFIMA specification that includes q moving average terms, one may initiate the test starting from and including $\gamma_{q+1}(\theta)$, effectively treating the moving average parameters as an unknown nuisance.

the present context by drawing from the broad literature on serial correlation tests.

In principle, any reasonable test for serial correlation may be used for robust inference here. Arguably the most popular test in practical work is the portmanteau test proposed by Box and Pierce (1970) and Ljung and Box (1978), for which the test statistic is formed as a (weighted) sum of the first $p \geq 1$ squared sample autocorrelation coefficients. Under the baseline setting with i.i.d. errors, the asymptotic distribution of this test statistic under the null hypothesis (i.e., no serial correlation) is χ_p^2. This classical test has also been adapted to accommodate non-i.i.d. errors; see, for example, Diebold (1986), Guo and Phillips (2001), Lobato et al. (2002), Escanciano and Lobato (2009), Dalla et al. (2022), and the many references therein. The portmanteau test is designed to detect violations of the null hypothesis up to the pth lag. It is also possible to allow the lag length p to slowly diverge to infinity as $T \to \infty$ so that the test is consistent against alternatives with unknown forms; see Hong (1996). In finite samples, however, it is evident that choosing p too large dilutes power if violations against the null can already be detected from the first few lags (which is not uncommon in practice). The power of this type of test crucially depends on the choice of p.

Motivated by these considerations, we adopt the Adaptive Portmanteau (AP) test proposed by Escanciano and Lobato (2009). The key advantage of their approach is to choose p in a data-driven fashion, which makes the test 'adaptive' with respect to the unknown complexity and nonparametric nature of the alternative. The test also readily accommodates an MDS structure of the error without requiring ε_t to be i.i.d. The simulation evidence provided in Escanciano and Lobato (2009) shows that the AP test is generally more powerful than commonly used competitors.[18]

We implement the AP test for a given candidate parameter θ as follows. Let $\hat{\gamma}_j(\theta)$ denote the jth sample autocovariance of $\varepsilon_t(\theta)$, that is,

$$\hat{\gamma}_j(\theta) \equiv \frac{1}{T-j} \sum_{t=j+1}^{T} \big(\varepsilon_t(\theta) - \bar{\varepsilon}(\theta)\big)\big(\varepsilon_{t-j}(\theta) - \bar{\varepsilon}(\theta)\big),$$

[18]The AP test is easy to implement as it does not require re-sampling unlike bootstrap methods. Computational efficiency is important as the test must be inverted over many candidate parameter values.

where $\bar{\varepsilon}(\theta)$ is the sample average of $\varepsilon_t(\theta)$. The asymptotic variance of $\hat{\gamma}_j(\theta)$ is estimated by

$$\hat{\tau}_j(\theta) \equiv \frac{1}{T-j} \sum_{t=j+1}^{T} \big(\varepsilon_t(\theta) - \bar{\varepsilon}(\theta)\big)^2 \big(\varepsilon_{t-j}(\theta) - \bar{\varepsilon}(\theta)\big)^2.$$

For a generic lag order $p \geq 1$, the portmanteau test statistic is defined as the sum of squared t-statistics in the following form

$$Q_p(\theta) \equiv T \sum_{j=1}^{p} \frac{\hat{\gamma}_j(\theta)^2}{\hat{\tau}_j(\theta)}. \tag{2.9}$$

The data-driven choice of p underlying the AP test relies on a combination of the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Specifically, let $\bar{p} \geq 1$ be a user-specified upper bound for p. Define the hybrid penalty function $\pi(p, T)$ as

$$\pi(p, T) \equiv \begin{cases} p \log T & \text{if } \max_{1 \leq j \leq \bar{p}} \sqrt{T} |\hat{\gamma}_j(\theta)| / \sqrt{\hat{\tau}_j(\theta)} \leq \sqrt{2.4 \log T}, \\ 2p & \text{otherwise.} \end{cases} \tag{2.10}$$

The lag order actually used in the AP test, denoted $p^*(\theta)$, is determined as (the smallest element of) the argmax of $Q_p(\theta) - \pi(p, T)$, with T and θ taken as given.

With this notation, the AP test statistic is defined by

$$Q^*(\theta) \equiv T \sum_{j=1}^{p^*(\theta)} \frac{\hat{\gamma}_j(\theta)^2}{\hat{\tau}_j(\theta)}. \tag{2.11}$$

Escanciano and Lobato (2009) show that the asymptotic distribution of this test statistic under the null hypothesis is χ^2_1, since the optimal lag order is one under the null hypothesis. Hence, we reject the null at the significance level β when $Q^*(\theta)$ exceeds the $1-\beta$ quantile of χ^2_1, denoted $\chi^2_{1,1-\beta}$. Recalling (2.7), the $1-\beta$ level identification-robust confidence set that we propose can thus be written explicitly as follows

$$\mathrm{CS}_{1-\beta} = \left\{\theta \in \Theta : Q^*(\theta) \leq \chi^2_{1,1-\beta}\right\}. \tag{2.12}$$

For ease of application, we summarize the proposed procedure in the following algorithm.

Algorithm 1 (Construction of Identification-Robust Confidence Sets).

Step 1: For a given candidate parameter vector $\theta = (\alpha, d) \in \Theta$, obtain the implied residual sequence $\varepsilon_t(\theta) = (1-L)^d (y_t - \alpha y_{t-1})$.

Step 2: Given a user-specified upper bound $\bar{p}$, compute $Q_p(\theta)$ according to (2.9) for all $p \in \{1, \ldots, \bar{p}\}$. Set $p^*(\theta)$ as the smallest p that maximizes $Q_p(\theta) - \pi(p, T)$, with $\pi(p, T)$ defined by (2.10).

Step 3: Compute the AP test statistic $Q^*(\theta)$ as in (2.11).

Step 4: Repeat Steps 1–3 for all θ on a (fine) discretization of the parameter space Θ. Form the $1-\beta$ level confidence set as (2.12), which collects all θ's such that $Q^*(\theta)$ is below the $1-\beta$ quantile of the χ_1^2 distribution. $\square$

We also investigated the weak identification issue via Monte Carlo with data simulated from the ARFIMA$(1, d, 0)$ model (2.1) setting $d = -0.4$ or 0.4 and α taking a wide range of values. The identification-robust inference procedure was implemented as described in Algorithm 1. Details of the simulation design, numerical implementation of the inference procedure, and findings are reported in Appendix S2. The results show that ARFIMA$(1, d, 0)$ has a severe weak identification issue when α is near unity with negative d or when α is near zero with positive d. In such cases, the identification-robust confidence sets typically exhibit bifurcation, the same pattern predicted by theory, and this outcome persists even when the sample size rises from $2,000$ to $5,000$. The probability of bifurcation is higher as α moves closer to zero (unity) when d is positive (negative).

3 Empirical Applications

The proposed identification-robust inference approach is applied to volatility measures and related quantities. Section 3.1 presents results based on realized volatility (RV) measures for a broad range of U.S. ETFs and stocks, and international stock market indices. Sections 3.2 and 3.3 report additional empirical findings for trading volume and news flow data.

3.1 Realized volatility measures

Daily RV measures from two publicly available databases are employed: the Realized Library of the Oxford–Man Institute of Quantitative Finance and the Risk Lab constructed by Dacheng Xiu.[19] For analysis of U.S. equity market data we use daily RV time series of the S&P 500 market ETF, nine industry ETFs, and the Dow Jones Industrial Average 30 stocks from the Risk Lab[20] – see Da and Xiu (2021) for the construction of these measures. The list of assets and summary statistics are reported in Table S1 of the internet appendix. We also conduct empirical analyses of international stock market indexes, for which the RV measures are obtained from the Realized Library and constructed as the sum of squared 5-minute intraday returns – see Table S2 in the internet appendix for a summary.

Following Andersen et al. (2003), we model each demeaned log RV series using the ARFIMA model in (2.1). For each series we compute the 95%-level robust confidence set for (α, d) by inverting the AP test at the 5% significance level, as in Algorithm 1. The implied inferences are constructed in a semiparametric data-driven manner with respect to potential serial correlation, employ asymptotic theory under the null, and do not rely on Gaussian errors. To simplify interpretation the RV measures are treated as stand-alone time series and attention is confined to their individual properties. In principle it is possible to translate empirical evidence obtained from the RV measures into statements regarding certain latent volatility-related functionals (e.g., integrated variance or quadratic variation) by invoking the so-called asymptotic negligibility argument as in Corradi and Distaso (2006) (see also Li and Patton (2018) for similar results designed more specifically for hypothesis testing). Extensions to obtain such further interpretation requires additional assumptions and asymptotic approximations with no changes in the robust approach to inference.[21] This extension is not pursued here to retain the weak identification focus of the paper.

[19] See https://realized.oxford-man.ox.ac.uk/ and https://dachxiu.chicagobooth.edu/#risklab.

[20] Since Dow Inc. (NYSE: DOW) is listed on NYSE only since 2019, its sample size is substantially shorter than all the other stocks. For this reason we replace it with Exxon Mobil Co. (NYSE: XOM), which belonged to the Dow Jones index until August 31, 2020.

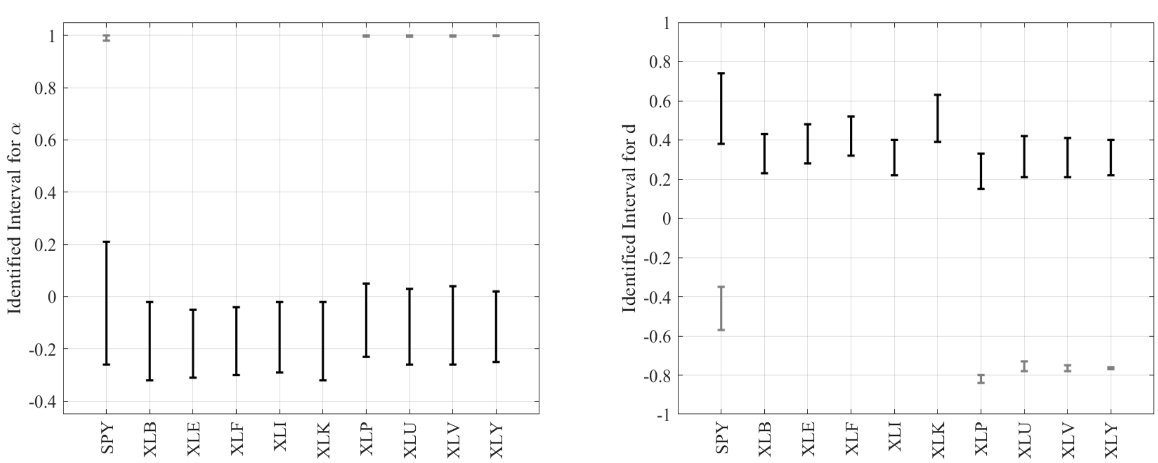
[21] Asymptotic negligibility arguments use sufficient conditions to ensure the errors between a realized measure and its 'population' continuous time counterpart can be ignored provided the high-frequency measures converge sufficiently fast. In recent work Bolko et al. (2023) explicitly address the measurement error problem by introducing assumptions on the proxy error which require primitive conditions on the continuous model and may be misspecified in general.

We carry out the test inversion via a grid search for $(\alpha, d) \in [-1, 1] \times [-1, 1]$. Given the large number of assets under consideration, presenting and comparing the two-dimensional confidence sets for all data series (say, in the form of Figure 1) is challenging in limited space. To achieve a concise presentation, one-dimensional confidence sets are reported for the autoregressive coefficient α and the fractional parameter d obtained by projecting the two-dimensional confidence sets onto each dimension.

Figure 3 plots the one-dimensional confidence sets for the SPY and the nine industry ETFs, with panel (a) and panel (b) showing the results for α and d, respectively. Since the confidence set for (α, d) often contains two disjoint regions, we use two gray scales (dark and light) to signify them, so that the same-colored one-dimensional confidence sets of α and d are projected from the same parent two-dimensional confidence set. By convention, the dark-colored (resp. light-colored) confidence sets are associated with $d > 0$ (resp. $d < 0$).

Figure 3: Confidence Sets for Selected ETFs

(a) 95% Confidence Set of α (b) 95% Confidence Set of d



95% identification-robust confidence sets were computed for the demeaned log RV of the ten index ETFs from 1996 to 2021 (downloaded from the Risk Lab) and the projected confidence intervals of α and d are plotted in Panels (a) and (b). Labels on the x-axis are the tickers of the ETFs.

From Figure 3 five of the ten ETFs (including SPY, XLP, XLU, XLV, and XLY) have confidence sets with disjoint regions. In dark-colored regions the autoregressive coefficient α is near zero and the positive fractional parameter indicates long memory, whereas in light-colored regions α is near unity and d takes large negative values. These patterns are consistent with theory, earlier intuition, and mirror the simulation findings, revealing evidence of weak identification in these cases.

These findings go some way to reconcile conflicting empirical evidence on long memory and rough-volatility in the existing literature. By accommodating the possibility of weak identification and adopting identification-robust inference, the present approach offers a rationale for different modeling schemes to 'co-exist', with both showing statistical support in the data. Lessons from the wider literature on weak identification suggest caution in the use of conventional methods that presume strong identification in the present setting of ARFIMA inference. Prior restriction of attention to one region in the parameter space (e.g., by imposing $\alpha = 1$ or by conducting optimization within a local neighborhood) removes the opportunity to introduce evidence in partial support of an alternative parameter region of d, thereby influencing forecasting and decision making.

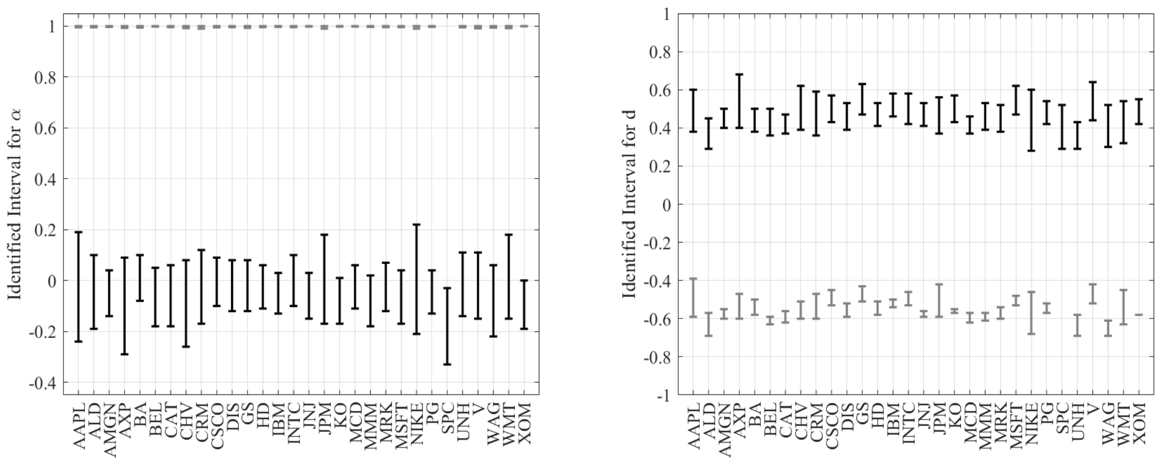
Figure 3 also reveals that the confidence sets for some assets (including XLB, XLE, XLF, XLI, and XLK) consist of only a single region, becoming confidence intervals. These confidence intervals for the fractional parameter d all hover around 0.4, which is close to the estimate reported in Andersen et al. (2003) and other work, thereby favoring the long memory narrative advocated in the early RV literature.

Are these findings robust to sample periods? To investigate, the full sample is divided into two subsamples of similar size, spanning 1996–2009 and 2010–2021. Each subsample is analyzed and Figure S9 in the internet appendix reports the estimated confidence sets. While the results for 1996–2009 are qualitatively unchanged to those of the full-sample shown in Figure 3, bifurcation into disjoint confidence regions is more prevalent for the 2010–2021 subsample, for which the confidence sets of all but one ETF contain regions of long memory and antipersistence.

So far, the evidence from the ten ETFs clearly demonstrates the empirical relevance of weak identification and the difficulty in robustly discriminating between long memory and rough dynamics. This phenomenon is not specific to ETFs. Similar analyses were conducted for each of the 30 constituent stocks of the Dow Jones Industrial Average and Figure 4 plots the resulting one-dimensional confidence sets for α and d. Almost all these sets bifurcate, suggesting that weak identification issues are even more prevalent for individual stocks than market indices. The estimated confidence sets are again fairly stable across assets.

Figure 4: Confidence Sets for Dow Jones Industrial Average Stocks

(a) 95% Confidence Set of α (b) 95% Confidence Set of d



The figures show 95% identification-robust confidence sets for the (demeaned) log RV of the 30 Dow Jones industrial average stocks from 1996 to 2021. The left (right) panel provides projections of the confidence sets on the α-axis (d-axis) for each asset. Labels on the x-axis are the tickers of the 30 stocks.

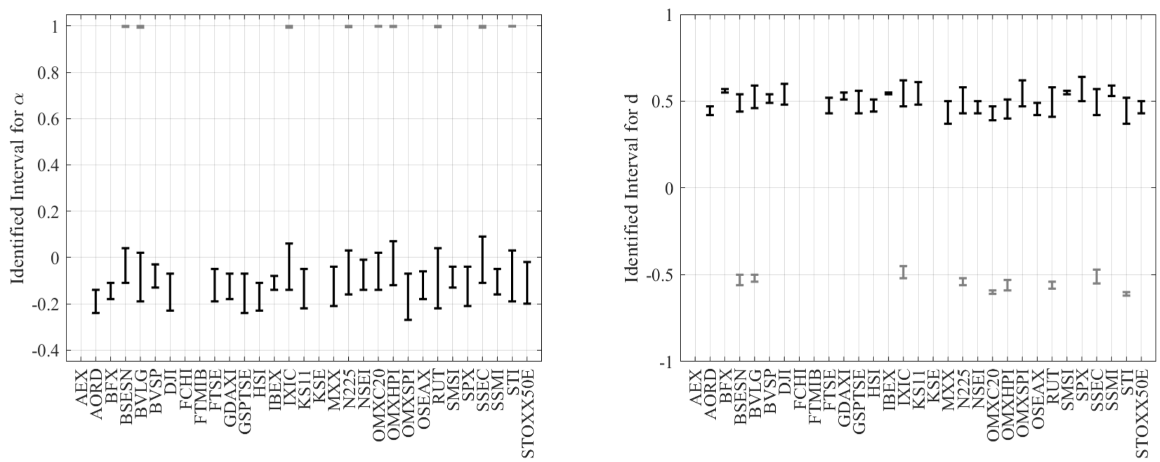
Additional empirical evidence is obtained with data from a broad range of international markets. The analysis relies on the daily RV series for all 31 stock market indices that are available online from the Oxford–Man Realized Library.[22] The same procedure is conducted for these market-level volatility measures and Figure 5 plots the projection-based one-dimensional confidence sets. The confidence sets for nine of the 31 indices exhibit bifurcation, eighteen of them are single-region, and the confidence sets for the remaining four indices (i.e., AEX, FCHI, FTMIB, and KSE) are empty.[23] These findings show that weak identification issues occur over a broad set of markets but that the overall evidence tends in favor of the long memory configuration, which is always present in the confidence set irrespective of whether there is one region or two regions. The volatility of stock market indices are weighted sums of individual stock variances and covariances. The fact that their RV measures exhibit stronger support for long memory is consistent with the property that long

[22]Robustness checks based on alternative RV measures are provided in the internet appendix.

[23]An empty confidence set may be interpreted as a specification test leading to a rejection of the hypothesis that the ARFIMA$(1, d, 0)$ model is correctly specified for a given data series. However, in view of the number of time series analyzed in the empirical analysis, these rejections seem tolerable with respect to possible false rejections (type I errors).

Figure 5: Confidence Sets for International Stock Market Indices

(a) 95% Confidence Set of α (b) 95% Confidence Set of d



95% identification-robust confidence sets were computed for the (demeaned) log RV of the 31 international stock market indices in the Realized Library of the Oxford–Man Institute of Quantitative Finance and the confidence sets were projected onto the α- and d-axes. The left (right) panel displays the projected confidence intervals on the α-axis (d-axis) for each asset. Labels on the x-axis are the tickers of the 31 international stock market indices and their names can be found in Table S2 of the internet appendix.

memory can arise from aggregation (Robinson, 1978; Granger, 1980). The internet appendix provides additional subsample results which support the same conclusion.

These empirical results for RV measures are summarized as follows. First, weak identification is prevalent in volatility dynamics analyzed via ARFIMA modeling. Robust inference manifests the issue in bifurcated confidence sets, suggesting caution in any statements about the generating mechanism relating to long memory versus roughness when the methodology relies on a presumption of strong identification. Second, for some assets the robust confidence sets reveal only a single region, which is always associated with a long memory configuration ($d > 0$). Long memory therefore appears to be more compatible with the in-sample RV dynamics for these assets.

3.2 Trading volume

The methodology is next applied to trading volume data, which are of independent economic interest and somewhat easier to interpret than RV measures because volume series are directly observable (in contrast to RV measures which are often regarded as proxies for latent volatility functionals). Given the well-known relationship between volume and volatility, these processes are expected to share similar qualitative dynamic properties and may therefore assist in interpreting results for volatility dynamics. A leading theoretical explanation of the volume-volatility relationship is the mixture-of-distributions hypothesis (Clark, 1973; Tauchen and Pitts, 1983; Andersen, 1996) which postulates that both volume and volatility are driven by common underlying information flows. The volume-volatility relationship is supported more formally in an equilibrium model. For instance, Collin-Dufresne and Fos (2016) show that stochastic liquidity can drive this relationship in a Kyle (1985)-type noisy rational expectation model.

We study the same 40 assets in the U.S. equity market as those in Section 3.1. For ease of replication, we use publicly available trading volume data obtained from *Yahoo Finance*. The sample period is February 1, 1993 to June 4, 2021. In parallel to the RV analysis the volume series are measured in logarithms. Because trading volume in the U.S. equity market exhibits a salient trend during earlier samples, we detrend the log volume series following standard practice, adopting a procedure similar to Andersen (1996) in which an additive (in logs) trend component is removed using a two-sided moving average spanning 512 trading days. Table S3 in the internet appendix reports summary statistics for the de-trended log trading volume. With only a few exceptions, the sample sizes of these volume series are greater than 5,000.

Using the same approach as before robust confidence sets are computed for each of the 40 (de-trended) log volume series. The projected one-dimensional confidence sets of α and d are plotted for the ten ETFs and the 30 Dow Jones stocks in Figure S10 in the internet appendix. As before, there is overwhelming evidence of weak identification in the trading volume data. Almost all the confidence sets have two disjoint regions, one suggesting rough near unit root dynamics (i.e., $d < 0$ and $\alpha \approx 1$) and the other implying long memory with weak short-run dynamics (i.e., $d > 0$ and $\alpha \approx 0$).

Overall, the findings point to the relevance of weak identification in ARFIMA modeling of trading volume, which is sometimes used as a measure of liquidity or investor sentiment, and thereby lends support to the results for volatility in view of the volume-volatility relationship.

The patterns exhibited in Figures S10 are qualitatively similar to those in Figures 3 and 4, but the 'statistically acceptable' values of d (i.e., those in the confidence sets) for trading volume are generally lower than those of the RV measures. Restricting attention to the long memory scheme, this outcome suggests that the volatility process may have longer memory than the volume process, a matter that deserves further investigation and may be associated with prior detrending of the volume series.[24]

3.3 News and Social Media Information Flow

The above findings highlight a central message of the paper concerning the common difficulty in economic data of empirically determining the dynamics in an ARFIMA generating mechanism when disjoint confidence sets suggest plausible mechanisms of either long memory or near unit root roughness as compatible with observed data. This duality in interpreting empirical outcomes is an advantage of identification-robust confidence regions but it does not resolve a definitive generating mechanism for use in practice.

Weak identification issues of this type are not necessarily purely finite sample problems because indeterminacy may persist asympotically as central limit theory can apply internally under either unidentified specifications or local 'drifting to unidentified' specifications, each of which reproduces finite sample characteristics of uncertainty asymptotically (Phillips, 1989; Staiger and Stock, 1997).[25] The sample sizes used in the current paper are already large by normal standards and using a few more years of data is unlikely to lead to any meaningful change in the results, as simulation findings affirm. Other possibilities involve using prior information from relevant theory or additional data that provide new information to assist in resolving the indeterminacy empirically.

The mixture-of-distributions hypothesis suggests a strong volume-volatility relationship and further postulates that both volume and volatility are driven by underlying information flows. The precise manner in which these economic quantities are linked depends partly on trading behavior and is generally unknown, but that they are linked raises little doubt. It is therefore reasonable to conjecture the news arrival process may well share dynamics similar

[24]For instance, Phillips and Jin (2021) show that detrending a stochastic trend by the HP filter materially influences the nature of the fitted trend and the residual series.

[25]The internal nature of the asymptotics is explained and established in Phillips (1989).

to volatility and volume.

This consideration provides motivation to examine the dynamics of news arrivals. We gauge the flow of information using Refinitiv MarketPsych buzz indices. The buzz indices are computed by aggregating references to U.S. stocks, sourced from both news and social media platforms. The news-based buzz index incorporates data from over $4,000$ mainstream news outlets, while the social media index is formulated using content from blogs, forums, and tweets from a selected pool of around $2,000$ social media sources. Although the data is available from January 1, 1998, a series of significant modifications has been introduced over time. In particular, internet news gathered by LexisNexis was added to the news source, in addition to the existing Reuters news starting from March 2005. LexisNexis social media content was integrated into the social media database starting from late 2008, while tweets were included from late 2009. To avoid the impact of these changes, our analysis commences from March 1, 2005 for the news index, and from January 1, 2010 for the social media index.

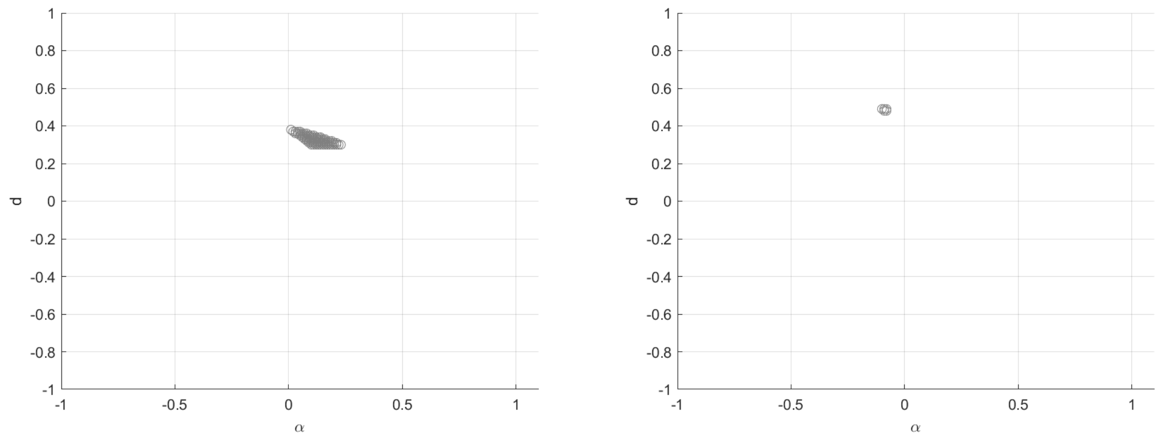
For a visual representation of the dynamics exhibited by the buzz indexes, refer to the left column of Figure S11 in the internet appendix. Both data series display an obvious upward trend and weekly seasonal patterns as shown by their autocorrelation functions (ACF) in the right column. Similar to the approach taken with trading volumes, we eliminate the trend component by applying a two-sided moving average across 512 trading days. Additionally, we address the weekly seasonal pattern by employing the moving average technique suggested by Cleveland et al. (1990). The resulting transformed data series, which is free from trend, seasonality, and has zero mean, is showcased in Figure S12. We observe a slowly decaying pattern in the ACF of both data series.

We apply the identification robust procedure to the two transformed data series and illustrate the resulting confidence sets in Figure 6. Interestingly, the confidence sets for both indexes contain only one region, featuring a memory parameter around 0.4 and an autoregressive coefficient near zero. Specifically, the confidence intervals for d are $[0.30, 0.38]$ for the news index and $[0.48, 0.49]$ for the social media index, with the corresponding intervals for α being $[0.01, 0.23]$ and $[-0.10, -0.08]$.[26] Weak identification does not appear to be a

[26]Similar corroborative findings based on an alternative measure of news flow, the Twitter economic uncertainty index (Baker et al., 2021), are presented in the internet appendix.

Figure 6: Confidence Sets for News Flow Indices

(a) Buzz Index: News (b) Buzz Index: Social Media



The left (right) panel displays the 95% identification-robust confidence set for the transformed buzz news (social media) index. The x-axis and y-axis correspond to α and d, respectively.

major issue here, and the statistical evidence supports long memory and weak short-run autoregressive (negatively correlated) dynamics. Notably, the confidence sets for the memory parameter d in both buzz indices center around $d = 0.4$,[27] which, incidentally, is the estimate reported by Andersen et al. (2003) for RV measures, despite the fact that these estimates are obtained using quite different datasets, over different periods, and by different econometric methods.

One caveat is that the finding that the news arrival process exhibits long memory does not in itself eliminate the ambiguity revealed in the bifurcated confidence sets for volatility and trading volume; and it does not rule out the possibility that volatility may be generated by a rough model involving antipersistence and near unit root dynamics. But this evidence does pose a conceptual challenge for the rough-volatility narrative from an economic standpoint. If volatility dynamics are driven by news arrivals, then the question remains what economic mechanism might explain their distinct empirical behavior along the long memory/roughness spectrum – that is, why is volatility rough, but news arrivals have long memory? Nonetheless, the small-scale analysis conducted here is suggestive and it highlights the potential usefulness of studying the joint dynamics of volatility, volume, and news arrivals, in the hope of revealing

[27] Interestingly, the fractional parameter obtained from the social media buzz index is very close to the nonstationary boundary 0.5, even after de-trending.

a suitable model for the 'forest' and not just the 'trees'.

4 Volatility Forecasting and Economic Value

Previous sections have so far shown that the theoretical and empirical difficulties in distinguishing the two model specifications in finite samples. The ambiguity manifests even in reasonably large samples with $5,000$ observations. Weak identification does not, however, necessarily imply that the models perform equally well in out-of-sample forecasting because forecasting behavior depends on elements such as the internal dynamics of the model, the forecast horizon and the particular loss function employed. Indeed, the spectral densities of these two models at near-zero frequencies are, in fact, distinct. Specifically, as the frequency approaches zero, the spectral density of the rough model tends towards zero, while the spectral density of the long memory model diverges to infinity (see Figure 2). Furthermore, from (2.3) and despite their similarities, the ACF of the rough model can potentially take small negative values at long lags, while the ACF of the long memory model will be positive at all lags. These disparities are anticipated to result in different forecasting performances at long horizons. This section therefore compares performance of the two model specifications in forecasting exercises at various horizons with large sets of empirical data. The dataset consists of 40 series spanning from 1996 to 2021, including the ten index ETFs and the 30 Dow Jones industrial average stocks examined in Section 3.1.[28]

We employ the ARFIMA$(1, d, 0)$ model for the (demeaned) log RV and consider two different estimators of $\theta = (\alpha, d)$. The two estimators, denoted $\hat{\theta}_L$ and $\hat{\theta}_R$, provide the smallest $Q^*(\theta)$ statistic across, respectively, the long memory and rough regions, i.e.,

$$\hat{\theta}_L = \arg\min_{\theta\in\Theta^L} Q^*(\theta) \text{ and } \hat{\theta}_R = \arg\min_{\theta\in\Theta^R} Q^*(\theta),$$

where $\Theta^L = [-1, 1] \times [0, 1]$, $\Theta^R = [-1, 1] \times [-1, 0]$, and $Q^*(\theta)$ is the adaptive Portmanteau statistic. This approach to estimation is in the spirit of Chevillon et al. (2010) and can be interpreted as yielding 'the parameter value that is least likely to be rejected'.

[28]We do not consider the 31 international stock indexes as they are not directly tradable.

In this approach the parameter space of d is allowed to span the full domain $[-1, 1]$. The same domain cannot be covered in standard time domain or frequency domain maximum likelihood as construction of the likelihood in these methods requires the fractional parameter d to lie within the stationary and invertible range, viz., $d \in (-0.5, 0.5)$. Moreover, popular semi-parametric methods (Künsch, 1987; Robinson, 1995a; Geweke and Porter-Hudak, 1983; Robinson, 1995b) are known to suffer bias that can be severe when the autoregressive parameter deviates far from zero. So those methods are not employed here.[29]

4.1 Out-of-sample Forecasting

We conduct a rolling forecasting exercise for RV, utilizing a window size of five years, corresponding to $T_0 = 251 \times 5 = 1,255$ data points. Three forecasting horizons (h) were considered: one day ($h = 1$), one week ($h = 5$), one month ($h = 21$). The loss functions employed are squared forecast error (SFE), $L^S_{t,h} = \left(\widehat{RV}_{t+h} - RV_{t+h}\right)^2$, and QLIKE, $L^Q_{t,h} = \log(\widehat{RV}_{t+h}) + RV_{t+h}/\widehat{RV}_{t+h}$, where $t = T_0, \cdots, T - h$. In order to derive the RV forecast from the ARFIMA$(1, d, 0)$ model for log RV, we assume that the error term of the model is Gaussian[30] and utilize the following relationship:

$$\widehat{RV}_{t+h} = \exp\left(\widehat{\log RV_{t+h}} + \frac{1}{2}\hat{\sigma}^2\right),$$

where $\hat{\sigma}^2$ is the sample variance of estimated residuals. Details of the forecasting method for obtaining $\widehat{\log RV_{t+h}}$ from the ARFIMA model are provided in Appendix S5.

Besides the two ARFIMA models, we also include the heterogeneous autoregressive model proposed by Corsi (2009) for RV (referred to as HAR-RV hereafter) in the comparison. The HAR-RV model is a popular benchmark for forecasting RV with the following specification:

$$RV_{t+h} = \alpha + \beta_d RV_t + \beta_w RV_t^{(w)} + \beta_m RV_t^{(m)} + \eta_t, \tag{4.1}$$

where $RV_t^{(w)} = \frac{1}{4}\sum_{k=1}^{4} RV_{t-k}$ and $RV_t^{(m)} = \frac{1}{17}\sum_{k=5}^{21} RV_{t-k}$ represent, respectively, the

[29]See Shi and Yu (2023) for more detailed discussions and comparisons of various estimation methods for the ARFIMA$(1, d, 0)$ model.

[30]In general, forecasting models are viewed as misspecified. The Gaussian assumption gives us a reasonable and standard way to carry out the transformation, as done in Andersen et al. (2003).

weekly and monthly averaged RV and ε_t is a disturbance term. The HAR-RV specification can be viewed as a constrained version of an AR(21) model and its parameters can be estimated using the ordinary least square (OLS) method.

To assess whether the three competing models generate statistically distinct forecasts, we employ the model confidence set (MCS) approach of Hansen et al. (2011), which provides formal inference for the best performing models.[31] Under this evaluation procedure, the 'p-value' for each individual model is used to provide a ranking of the models; the higher the p-value is, the more likely the corresponding model is included in the MCS at any given confidence level. In lieu of this mechanic, we shall report the models' p-values below without pre-fixing any specific confidence level. The MCS procedure is applied to each forecasting horizon h separately, involving $T^* = T - h - T_0 + 1$ out-of-sample periods.

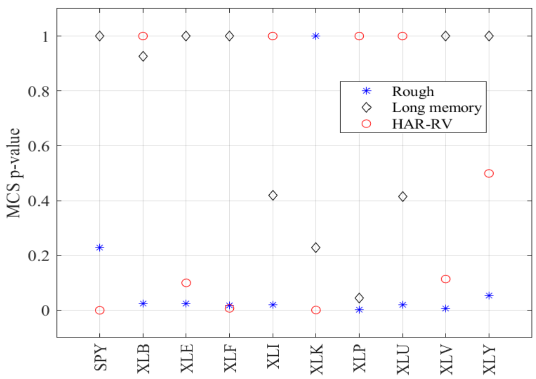
Figures 7 and 8 plot the MCS p-values based on the squared forecast errors (left panel) and QLIKE (right panel) for selected ETFs and individual stocks. Forecasting results for the ARFIMA model with $\hat{\theta}_R$ and $\hat{\theta}_L$ are labeled as 'Rough' and 'Long memory', respectively.[32] For the majority of studied assets, the long memory specification outperforms the rough specification for both loss measures (i.e., SFE and QLIKE) and across all three forecasting horizons (i.e., one-day, one-week, and one month), as the plotted p-values suggest that the former is generally much more likely to be included in the MCS than the latter. In addition, we observe that the long memory specification outperforms the benchmark HAR-RV model in a similar fashion. The superiority of the long memory specification is especially pronounced for Dow Jones industrial average stocks: most of the p-values are indistinguishable from 1, suggesting that the long memory specification almost always falls in the MCS. To further appreciate the economic value of the long memory specification's superior forecasting performance, we turn next to a portfolio construction problem that utilizes volatility forecasts as inputs.

[31] Model comparison is based on forecast loss differentials, which are assumed in Hansen et al. (2011) to be stationary and α-mixing with sufficient moments. These conditions are not necessarily satisfied in the present framework when there is nonstationary and long memory behavior in the data. Extending the MCS approach to accommodate these complications is beyond the scope of the present paper.

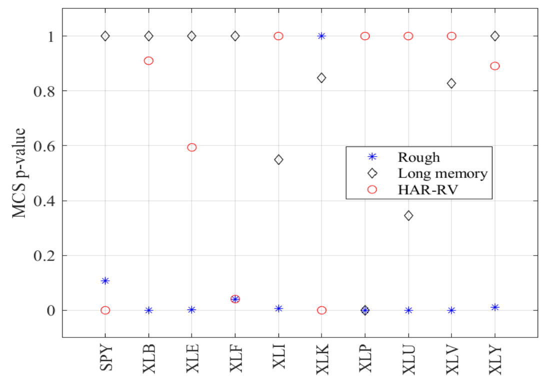
[32] Detailed summary statistics for the mean squared forecast errors and QLIKE are reported in Section S6 of the internet appendix.

Figure 7: MCS p-values: Selected ETFs

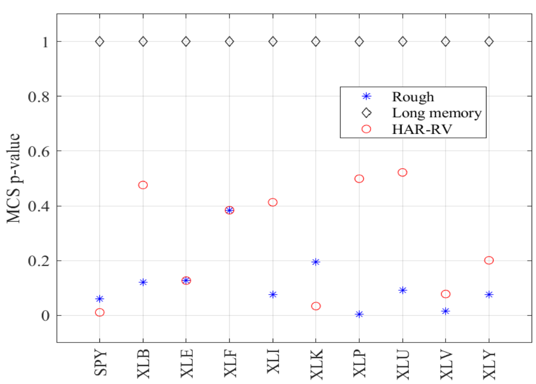
(a) One-day-ahead: SFE



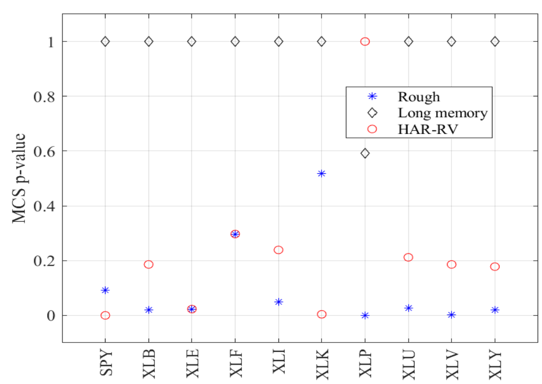
(b) One-day-ahead: QLIKE



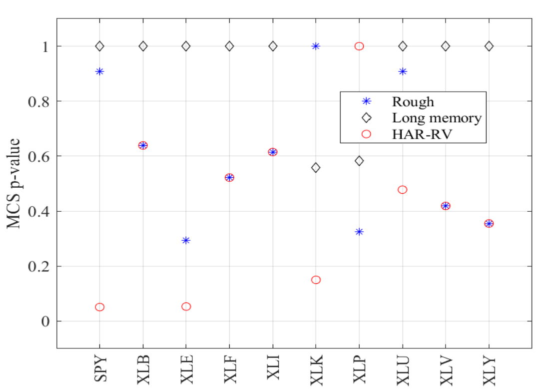
(c) One-week-ahead: SFE



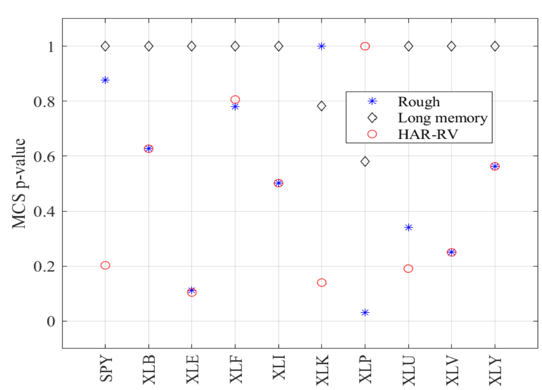
(d) One-week-ahead: QLIKE



(e) One-month-ahead: SFE



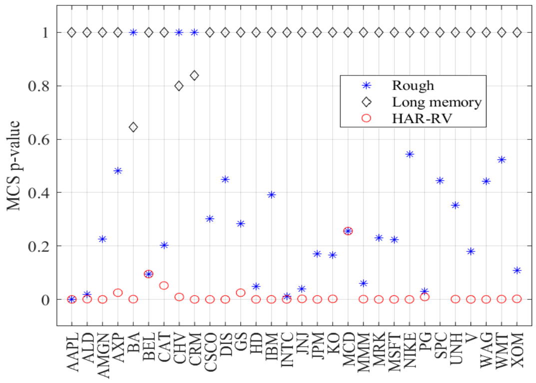
(f) One-month-ahead: QLIKE



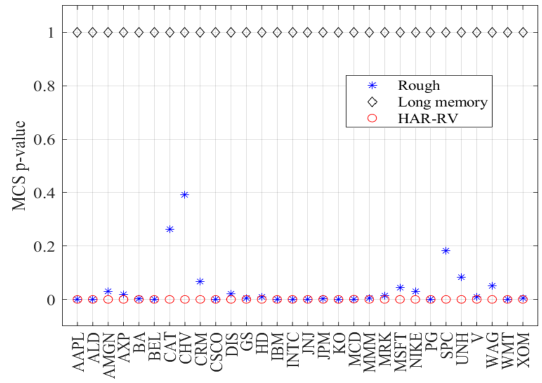
Five-year rolling window forecasts were conducted for the RVs of the ten index ETFs for the period 1996–2021, based on the ARFIMA model with $\hat{\theta}_R$ (Rough) and $\hat{\theta}_L$ (Long memory) and the HAR-RV model (HAR). The MCS p-values for the one-day, one-week, and one-month-ahead forecasts are presented in the first, second, and third rows, respectively. Results on the left (right) column are based on SFE (QLIKE). The x-axis displays asset tickers.

Figure 8: MCS p-values: Dow Jones Industrial Average Stocks

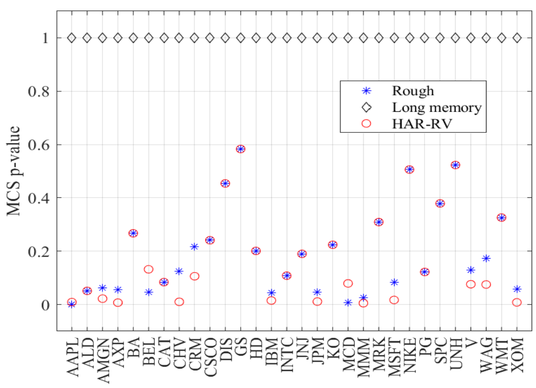
(a) One-day-ahead: SFE



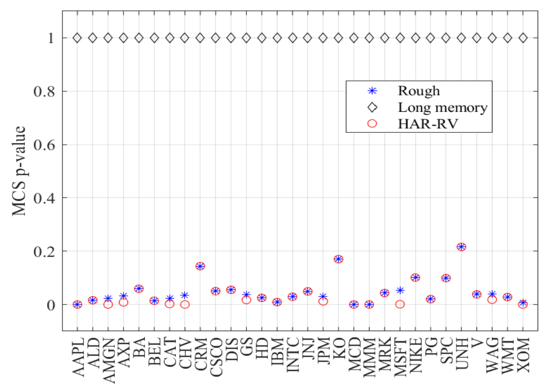
(b) One-day-ahead: QLIKE



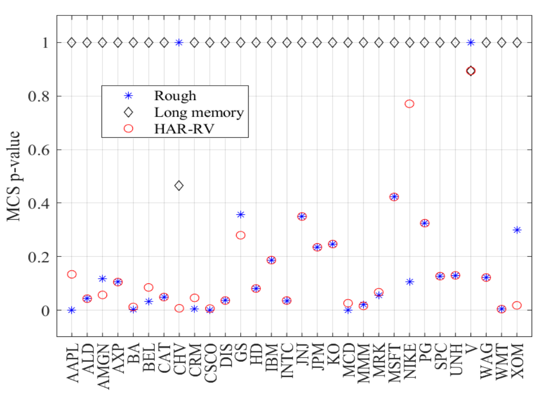
(c) One-week-ahead: SFE



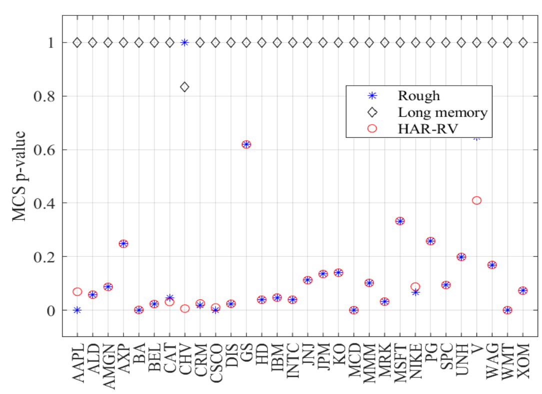
(d) One-week-ahead: QLIKE



(e) One-month-ahead: SFE



(f) One-month-ahead: QLIKE



Five-year rolling window forecasts were conducted for the RVs of the 30 Dow Jones industrial average stocks for the period 1996–2021, based on the ARFIMA model with $\hat{\theta}_R$ (Rough) and $\hat{\theta}_L$ (Long memory) and the HAR-RV model (HAR). The MCS p-values for the one-day, one-week, and one-month-ahead forecasts are presented in the first, second, and third rows, respectively. Results on the left (right) are based on SFE (QLIKE). The x-axis displays asset tickers.

4.2 The Economic Values of Forecasting Models

To help reveal the gains of superior volatility forecasts in economic terms, we consider an investor with mean-variance preference, allocating a fraction of wealth ω_t to a risky asset with return r_{t+1} and the rest to a risk-free asset with return r^f.[33] His future wealth is $W_{t+1} = W_t\left(1 + r^f + \omega_t r^e_{t+1}\right)$, where $r^e_{t+1} = r_{t+1} - r^f$ denotes the excess return. The conditional expected utility is

$$U(\omega_t) = \mathbb{E}_t(W_{t+1}) - \frac{\gamma}{2}\mathbb{V}_t(W_{t+1}) = W_t\left[1 + r^f + \omega_t\,\mathbb{E}_t\left(r^e_{t+1}\right) - \frac{\gamma}{2}\omega_t^2\,\mathbb{V}_t\left(r^e_{t+1}\right)\right], \tag{4.2}$$

where γ is the absolute risk aversion of the investor.

We adopt a risk parity strategy as in Bollerslev et al. (2018), where investors maintain a constant conditional Sharpe ratio, i.e., $SR = \mathbb{E}_t\left(r^e_{t+1}\right)/\sqrt{\mathbb{V}_t\left(r^e_{t+1}\right)}$. The optimal portfolio that maximizes the mean-variance utility is obtained by investing a fraction of wealth

$$\omega_t^* = \frac{\mathbb{E}_t\left(r^e_{t+1}\right)}{\gamma\,\mathbb{V}_t\left(r^e_{t+1}\right)} = \frac{SR/\gamma}{\sqrt{\mathbb{E}_t(RV_{t+1})}} \tag{4.3}$$

in the risky asset, where we use the relationship of $\mathbb{E}_t(RV_{t+1}) = \mathbb{V}_t\left(r^e_{t+1}\right)$. This trading strategy mimics in a simple way the volatility targeting strategy employed by practitioners and the investor is often referred to as 'risk parity investor'. When the predicted volatility $\sqrt{\mathbb{E}_t(RV_{t+1})}$ is below the target SR/γ, the investor applies leverage $\omega_t^* > 1$; when the predicted volatility $\sqrt{\mathbb{E}_t(RV_{t+1})}$ is above the target, the agent only invests a part of his wealth in the risky asset $\omega_t^* < 1$.

Assume that we use model m (e.g., long memory, rough, or HAR-RV) to generate the one-period-ahead forecast for RV, denoted by $\mathbb{E}_t^m(RV_{t+1})$. The model-implied optimal portfolio weight becomes

$$\omega_t^{m*} = \frac{SR/\gamma}{\sqrt{\mathbb{E}_t^m(RV_{t+1})}}. \tag{4.4}$$

[33]The single-asset setting is used as an illustration because it is aligned with the current paper's focus on univariate volatility modeling. Our discussion on the long memory and rough volatility models may have further implications in a multivariate setting with covariance matrices. The latter problem involves additional complications concerning factor structure and/or high dimensionality issues, and so, is left for future research.

By combining equations (4.2) and (4.4), the utility per unit of wealth at period t, denoted by UoW_t, is:

$$UoW_t^m = 1 + r^f + \frac{SR^2}{\gamma}\left[\frac{\sqrt{RV_{t+1}}}{\sqrt{\mathbb{E}_t^m\left(RV_{t+1}\right)}} - \frac{1}{2}\frac{RV_{t+1}}{\mathbb{E}_t^m\left(RV_{t+1}\right)}\right]. \tag{4.5}$$

The realized utility is then computed as:

$$UoW^m = \frac{1}{T-h-T_0+1}\sum_{t=T_0}^{T-h} UoW_t^m. \tag{4.6}$$

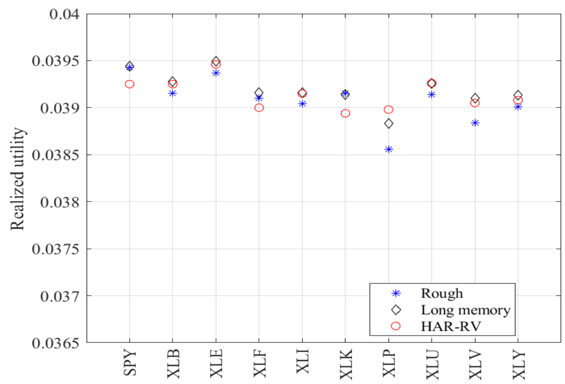
To carry out the comparison, we exclude the constant term $1 + r_f$ in equation (4.5) without loss of generality, and assume an annualized Sharpe ratio of $SR = 0.4$ and a coefficient of risk aversion of $\gamma = 2$, following Bollerslev et al. (2018). Consequently, assuming perfect forecast, i.e., $\mathbb{E}_t^m\left(RV_{t+1}\right) = RV_{t+1}$, the realized utility UoW^m amounts to 4%. The better the forecasting performance of the model for RV, the closer the realized utility is to 4%.

Figure 9 presents the realized utilities corresponding to the forecasting results discussed in the previous section. The left panel represents the selected index ETFs, while the right panel displays the Dow Jones Industrial Average stocks. These utilities are computed for risk parity investors using equation (4.6) with the three forecasting models (rough, long memory, and HAR-RV) applied in a five-year rolling window fashion over the period 1996–2021. The first, second, and third rows of the figure illustrate the realized one-day, five-day, and 21-day utilities of the investor, respectively. As we can see, all realized utilities are below the optimal level of 4% and decrease as the forecasting horizon extends from one day to one month. A forecasting model is considered better when its realized utility is closer to the optimal level. According to Bollerslev et al. (2018), improvements of the order of multiple basis points are noticeable, while tens of basis points are considered substantial, and less than one basis point is relatively insignificant.

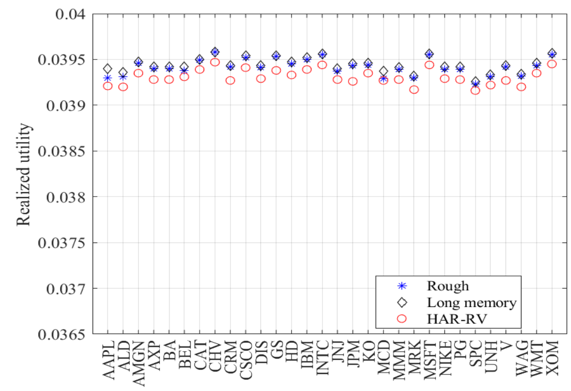
Consistent with the statistical evidence from the forecast comparison in Section 4.1, all three models demonstrate strong performance in predicting volatility, with the long memory model consistently achieving the highest realized utilities with only a few exceptions. The advantage of the long memory model over the rough and HAR-RV models is more pronounced

Figure 9: Realized Utility of Risk Parity Investors

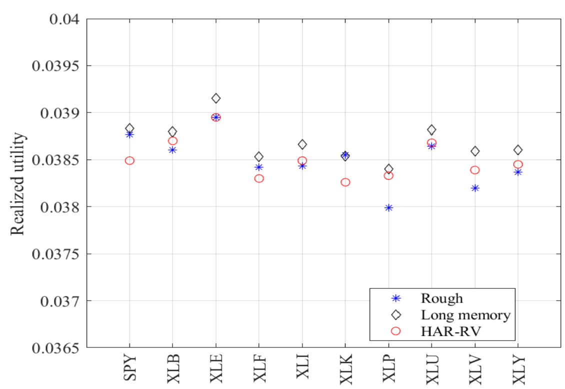
(a) Selected ETFs: one-day-ahead



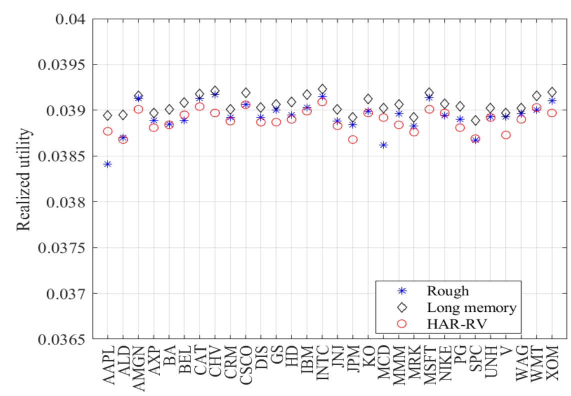
(b) Dow Jones industrial average: One-day-ahead



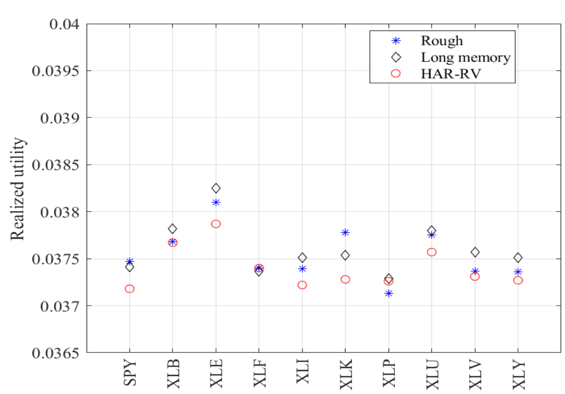
(c) Selected ETFs: one-week-ahead



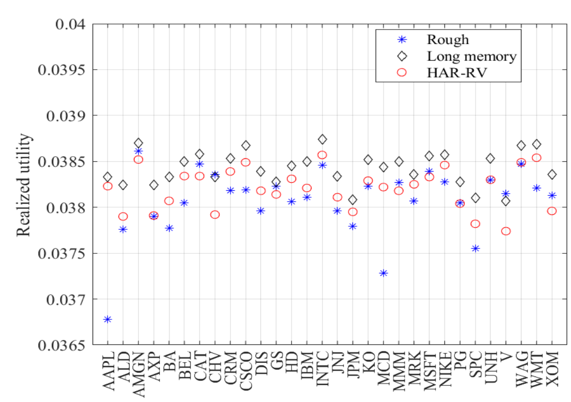
(d) Dow Jones industrial average: One-week-ahead



(e) Selected ETFs: one-month-ahead



(f) Dow Jones industrial average: One-month-ahead



The figure presents the realized utilities based on one-day-ahead, one-week-ahead, and one-month-ahead forecasts in the first, second, and third rows, respectively. These utilities are computed using (4.6) for risk parity investors with five-year rolling window forecasts from the rough, long memory, and HAR-RV models for the period 1996–2021. The left (right) panel represents the selected index ETFs (Dow Jones Industrial Average stocks). The x-axis shows the tickers of the assets.

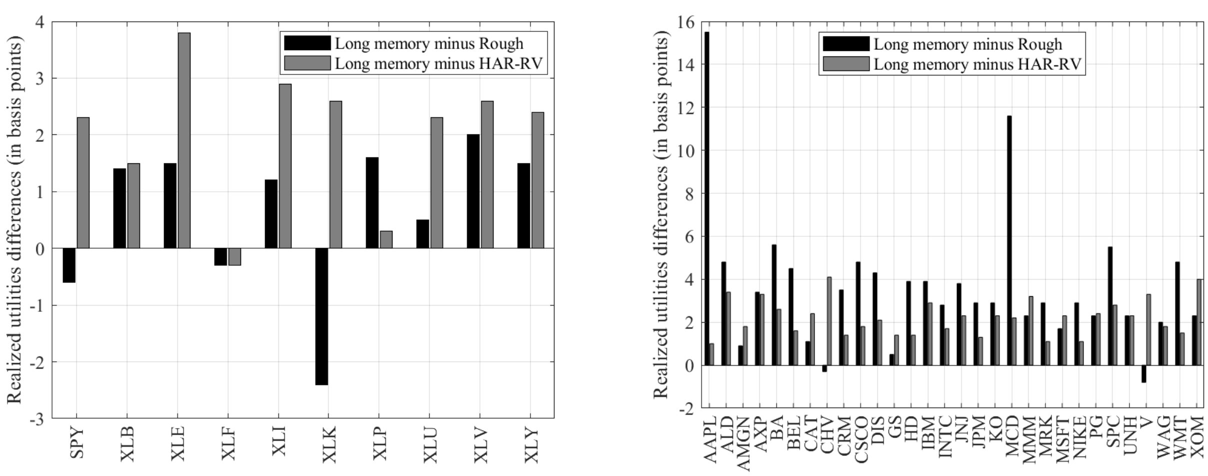
when applied to the Dow Jones 30 stocks as compared to the index ETFs. Furthermore, for the Dow Jones 30 stocks, the superiority of the long memory model becomes increasingly evident for longer forecasting horizons.

To visualize the differences in performance more directly, we plot the realized 21-day utility differentials between the long memory model and the alternative models (in basis points) for both the selected ETFs and Dow Jones 30 stocks in Figure 10. For the ETFs, the long memory model shows an average improvement of 1.9 basis points over HAR-RV, while the difference between the long memory and rough models is less pronounced, with an average value of 0.6 basis points. However, for the Dow Jones 30 stocks, the average realized 21-day utility differentials between the long memory model and the rough and HAR-RV models were 3.6 and 2.2 basis points, respectively. These results underscore the advantageous nature of the long memory model in accurately predicting volatility, particularly for the Dow Jones 30 stocks and when considering longer forecasting horizons.[34]

Figure 10: The 21-day Realized Utility Differentials

(a) Selected ETFs

(b) Dow Jones Industrial Average



The figure presents the 21-day realized utilities differentials between the long memory and alternative models. The realized utilities are computed using (4.6) for risk parity investors with five-year rolling window 21-day-ahead forecasts from the three models for the period 1996–2021. The x-axis shows the tickers of the assets.

[34]The three models for high-frequency RV measures (rough, long memory, and HAR-RV) exhibit substantial improvement over the low-frequency GARCH model based on daily returns. On average, the realized 21-day utilities of the long memory model surpass that of the GARCH model by 13.2 basis points for the selected ETFs and by 20.2 basis points for the Dow Jones 30 stocks. Refer to the online appendix for details regarding the GARCH results.

4.3 Discussion

The forecasting exercise discussed above provides a concrete empirical setting in which the long memory specification outperforms the rough specification. As alluded to in the introduction, this finding is not at odds with the rough model's superior performance for option pricing (Gatheral et al., 2020). Instead, our new empirical evidence complements those prior findings with a distinct perspective, implying that neither the long memory nor the rough specification is likely to 'rule it all' in practice. Such ambiguity reflects the fundamental difficulty in decision making when alternative models are weakly identified, as recently emphasized by Andrews and Mikusheva (2022). Indeed, optimal decisions including choosing the "correct" model could be sensitive to the decision maker's loss function and/or prior belief. Needless to say, the empirical evidence also clearly demonstrates that models being weakly identified should not be misunderstood as their being completely the same; the models are statistically difficult to disentangle given the available information, but may nevertheless lead to decisions with nontrivially distinct economic consequences.

To shed more light on how the competing models generate distinct volatility forecasts, we take the estimated models for the material sector ETF, XLB, as an example. We estimate the ARFIMA$(1, d, 0)$ model for the log RV of XLB using data from 2010 to 2021. The estimated model coefficients for this period are $\hat{\theta}_R = (\hat{d}_R, \hat{\alpha}_R) = (-0.45, 0.992)$ and $\hat{\theta}_L = (\hat{d}_L, \hat{\alpha}_L) = (0.47, 0.07)$. The h-step-ahead predictors of ARFIMA are calculated using the formula:

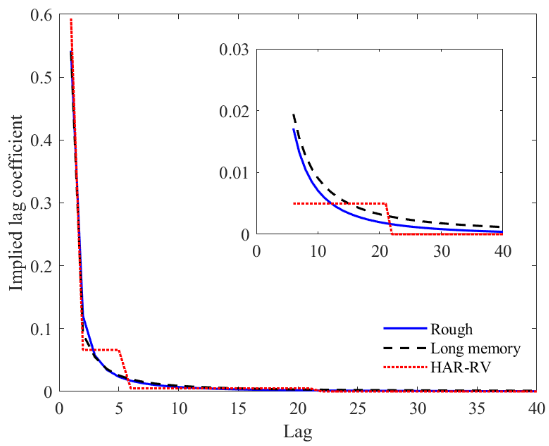
$$\hat{y}_{T+h} = \phi_{T,1}^{(h)}(\theta)\, y_T + \phi_{T,2}^{(h)}(\theta)\, y_{T-1} + \cdots + \phi_{T,T}^{(h)}(\theta)\, y_1 \tag{4.7}$$

where $\phi_T^{(h)}(\theta) = (\phi_{T,1}^{(h)}(\theta), \ldots, \phi_{T,T}^{(h)}(\theta))'$ denotes the model implied forecasting coefficients detailed in Appendix S5.

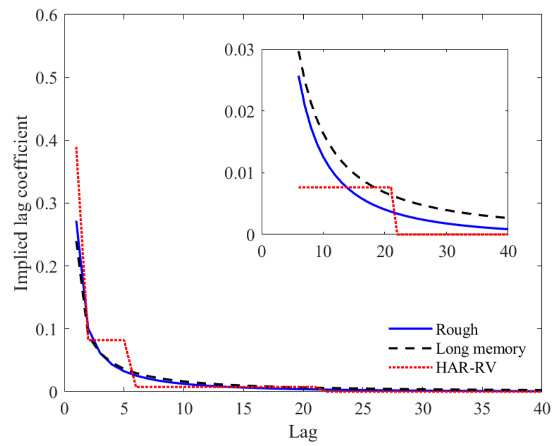
In Figure 11, we plot the model implied forecasting coefficients of the three forecasting models at various horizons. As expected, all three models assign higher weights to more recent observations than to more distant observations. For the rough volatility model, there is no obvious increase in the weights of distant observations as forecasting horizon increases, whereas the long memory and HAR-RV models allocate higher weight to distant observations as forecasting horizon increases. However, the HAR-RV specification lacks flexibility, as the weights are determined by a step function truncated at lag 21. These differences help explain why the long memory model outperforms the other two models in long-term forecasts.

Figure 11: Implied lag coefficients for different forecating models: XLB

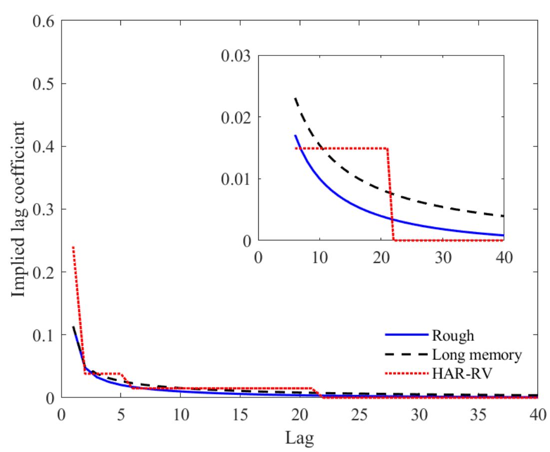
(a) One-day-ahead



(b) One-week-ahead



(c) One-month-ahead



Note: The ARFIMA and HAR-RV models are applied to the log RV and RV, respectively, estimated using data from 2010 to 2021. The implied lag coefficients for the rough and long memory models ($\phi_T^{(h)}(\hat{\theta}_R)$ and $\phi_T^{(h)}(\hat{\theta}_L)$) are computed from equation (S5.2) in Appendix S5. The subplots display the lag coefficients from the sixth lag onward.

5 Conclusion

In early applied literature the ARFIMA$(1, d, 0)$ model was found to be an adequate model for log realized volatility with a fitted autoregressive parameter (α) near zero and estimated memory parameter (d) close to 0.5, signifying long memory in volatility. Recent literature using the same ARFIMA model has found autoregressive parameters near unity and memory parameters close to -0.5, providing empirical evidence for 'rough volatility', reflecting a primary characteristic of antipersistent time series in contrast to long memory. This paper explains these co-existing, yet conflicting, empirical outcomes as a symptom of intrinsic weak identification within the ARFIMA model itself. Our theory suggests that, while the two parameter configurations appear very different, the distance between the corresponding models converges to zero when the autoregressive parameter is localized to unity or zero. The resulting weak identification produces model ambiguities that inevitably affect empirical work.

To address this potential ambiguity in practical applications our approach proposes the use of Anderson–Rubin identification-robust confidence sets for the model parameters by inverting tests for zero serial correlation in the implied disturbances. Extensive applications of this approach conducted on a broad range of realized volatility and trading volume series, document the prevalence of weak identification in ARFIMA inference. Robust confidence sets are often found to bifurcate, containing two disjoint regions that signal a severe form of weak identification and reveal an indeterminacy between the two parameter configurations reported in the literature. The overall empirical evidence we have examined leans in favor of the long memory configuration in that the corresponding parameter region is always part of the estimated confidence set. These findings are further corroborated by the empirical analysis of news arrivals, measured by news and social media buzz indices for US stock index, for which identification-robust inference lends support to long memory in the ARFIMA$(1, d, 0)$ model. Forecasting analysis also points to generally superior performance of the long memory model at both short and long horizons.[35]

The weakness in ARFIMA modeling revealed in our analysis is a cautionary message to empirical investigators using this model. A deeper implication is that the observed data are often not rich enough to discriminate between disjoint memory structures in the ARFIMA model framework. Practical econometric work can face this reality by reporting confidence regions that reflect any ambiguity, as demonstrated here; and if this robustness is insufficient

[35]The rough volatility model inherently predicts the high-frequency behavior of volatility strongly. There is a potential to assess its fit to higher frequency volatilities using spot volatility proxies, although spot volatility proxies from ultra high-frequency data may be imprecise.

for a task at hand, such as prediction, then the framework must be extended to accommodate data that might assist in resolving the ambiguity. Possible extensions include the use of varying coefficient regression so that memory and autoregressive parameters vary according to news flow covariates that import information about memory in the data to assist in resolving ambiguities; another is to incorporate such covariates in a multivariate system to jointly model news flows with the relevant economic variables. Progress on such extensions is left for future research.

References

Abadir, K. and G. Talmain (2002). Aggregation, Persistence and Volatility in a Macro Model. *Review of Economic Studies 69*(4), 749–779.

Agiakloglou, C., P. Newbold, and M. Wohar (1993). Bias in an estimator of the fractional difference parameter. *Journal of Time Series Analysis 14*(3), 235–246.

Alfarano, S. and T. Lux (2007). A noise trader model as a generator of apparent financial power laws and long memory. *Macroeconomic Dynamics 11*(S1), 80–101.

Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *The Journal of Finance 51*(1), 169–204.

Andersen, T. G. and T. Bollerslev (1997). Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *The Journal of Finance 52*(3), 975–1005.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics 61*(1), 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association 96*(453), 42–55.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*(2), 579–625.

Anderson, T. W. and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics 20*(1), 46–63.

Andrews, D. W. K. and X. Cheng (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica 80*(5), 2153–2211.

Andrews, D. W. K. and P. Guggenberger (2003). A bias–reduced log–periodogram regression estimator for the long–memory parameter. *Econometrica 71*(2), 675–712.

Andrews, I. and A. Mikusheva (2022). Optimal decision rules for weak GMM. *Econometrica 90*(2), 715–748.

Andrews, I., J. Stock, and L. Sun (2019). Weak instruments in IV regression: Theory and practice. *Annual Review of Economics 11*, 727–753.

Ansley, C. F. and P. Newbold (1980). Finite sample properties of estimators for autoregressive moving average models. *Journal of Econometrics 13*(2), 159–183.

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1996a). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 74*(1), 3–30.

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1996b). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 74*(1), 3–30.

Baker, S. R., N. Bloom, S. Davis, and T. Renault (2021). Twitter-derived measures of economic uncertainty.

Bayer, C., P. Friz, and J. Gatheral (2016). Pricing under rough volatility. *Quantitative Finance 16*(6), 887–904.

Bhattacharya, R. N., V. K. Gupta, and E. Waymire (1983). The Hurst effect under trends. *Journal of Applied Probability 20*(3), 649–662.

Bolko, A. E., K. Christensen, M. S. Pakkanen, and B. Veliyev (2023). A GMM approach to estimate the roughness of stochastic volatility. *Journal of Econometrics 235*, 745–778.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Bollerslev, T., B. Hood, J. Huss, and L. H. Pedersen (2018). Risk everywhere: Modeling and managing volatility. *The Review of Financial Studies 31*(7), 2729–2773.

Bollerslev, T., J. Li, and Y. Xue (2018). Volume, volatility, and public news announcements. *Review of Economic Studies 85*(4), 2005–2041.

Bouchaud, J.-P. (2020, December). A step closer to the perfect volatility model. *Risk.net*.

Box, G. E. and D. A. Pierce (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association 65*(332), 1509–1526.

Breidt, F. J., N. Crato, and P. De Lima (1998). The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics 83*(1-2), 325–348.

Chan, N. H. and C. Z. Wei (1987). Asymptotic Inference for Nearly Nonstationary AR(1) Processes. *The Annals of Statistics 15*(3), 1050 – 1063.

Chen, X., L. P. Hansen, and M. Carrasco (2010). Nonlinearity and temporal dependence. *Journal of Econometrics 155*(2), 155–169.

Cheng, X., W. W. Dou, and Z. Liao (2022). Macro-finance decoupling: Robust evaluations of macro asset pricing models. *Econometrica 90*(2), 685–713.

Chevillon, G., A. Hecq, and S. Laurent (2018). Generating univariate fractional integration within a large var (1). *Journal of Econometrics 204*(1), 54–65.

Chevillon, G., M. Massmann, and S. Mavroeidis (2010). Inference in models with adaptive learning. *Journal of Monetary Economics 57*(3), 341–351.

Chevillon, G. and S. Mavroeidis (2017). Learning can generate long memory. *Journal of Econometrics 198*(1), 1–9.

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica 41*(1), 135–155.

Cleveland, R. B., W. S. Cleveland, J. E. McRae, and I. Terpenning (1990). Stl: A seasonal-trend decomposition. *J. Off. Stat 6*(1), 3–73.

Collin-Dufresne, P. and V. Fos (2016). Insider trading, stochastic liquidity, and equilibrium prices. *Econometrica 84*(4), 1441–1475.

Comte, F. and E. Renault (1996). Long memory continuous time models. *Journal of Econometrics 73*(1), 101–149.

Comte, F. and E. Renault (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance 8*(4), 291–323.

Corradi, V. and W. Distaso (2006). Semi-parametric comparison of stochastic volatility models using realized measures. *Review of Economic Studies 73*(3), 635–667.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7*(2), 174–196.

Da, R. and D. Xiu (2021). When moving-average models meet high-frequency data: Uniform inference on volatility. *Econometrica 89*(6), 2787–2825.

Dalla, V., L. Giraitis, and P. C. B. Phillips (2022). Robust tests for white noise and cross-correlation. *Econometric Theory, forthcoming*.

Diebold, F. X. (1986). Testing for serial correlation in the presence of ARCH. In *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, Volume 323, pp. 328. American Statistical Association Washington DC.

Diebold, F. X. and A. Inoue (2001). Long memory and regime switching. *Journal of Econometrics 105*(1), 131–159.

Ding, Z., C. W. Granger, and R. F. Engle (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance 1*(1), 83–106.

Duffy, J. A. and I. Kasparis (2021). Estimation and inference in the presence of fractional d= 1/2 and weakly nonstationary processes. *The Annals of Statistics 49*(2), 1195–1217.

Dufour, J.-M. (1997). Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica 65*(6), 1365–1387.

El Euch, O., M. Fukasawa, and M. Rosenbaum (2018a). The microstructual foundations of leverage effect and rough volatility. *Finance and Stochastics 22*(2), 241–280.

El Euch, O., M. Fukasawa, and M. Rosenbaum (2018b). The microstructural foundations of leverage effect and rough volatility. *Finance Stochastics 22*, 241–280.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica 50*(4), 987–1007.

Escanciano, J. C. and I. N. Lobato (2009). An automatic portmanteau test for serial correlation. *Journal of Econometrics 151*(2), 140–149.

Fouque, J. P. and R. Hu (2018). Optimal portfolio under fast mean–reverting fractional stochastic environment. *SIAM Journal of Financial Mathematics 9*, 564–601.

Fukasawa, M., T. Takabatake, and R. Westphal (2022). Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics. *Mathematical Finance 32*, 1086–1132.

Garnier, J. and K. Sølna (2017). Correction to black-scholes formula due to fractional stochastic volatility. *SIAM Journal Financial Mathematics 8*, 560–588.

Garnier, J. and K. Sølna (2018). Option pricing under fast-varying and rough stochastic volatility. *Ann Finance 14*, 489–516.

Gatheral, J., T. Jaisson, and M. Rosenbaum (2018). Volatility is rough. *Quantitative Finance 18*(6), 933–949.

Gatheral, J., P. Jusselin, and M. Rosenbaum (2020). The quadratic rough heston model and the joint S&P 500/VIX smile calibration problem. *arXiv preprint arXiv:2001.01789*.

Geweke, J. and S. Porter-Hudak (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis 4*(4), 221–238.

Giraitis, L., H. L. Koul, and D. Surgailis (2012). *Large Sample Inference for Long Memory Processes*. Imperial College Press.

Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics 14*(2), 227–238.

Granger, C. W. J. and R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis 1*(1), 15–29.

Guo, B. and P. C. B. Phillips (2001). Testing for autocorrelation and unit roots in the presence of conditional heteroskedasticity of unknown form. *UC Santa Cruz Economics Working Paper* (540).

Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica 79*(2), 453–497.

Harvey, A. C. (2007). Long memory in stochastic volatility. In *Forecasting Volatility in the Financial Markets*, pp. 351–363. Elsevier.

Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica 64*(4), 837–864.

Hosking, J. R. (1981). Fractional differencing. *Biometrika 68*(1), 165–76.

Jusselin, P. and M. Rosenbaum (2020). No-arbitrage implies power-law market impact and rough volatility. *Mathematical Finance 30*(4), 1309–1336.

Klemeš, V. (1974). The Hurst phenomenon: A puzzle? *Water Resources Research 10*(4), 675–688.

Künsch, H. (1987). Statistical aspects of self-similar processes. In *Proceedings of the First Congress of the Bernoulli Society, 1987.*

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica 53*(6), 1315–1335.

Li, J. and A. J. Patton (2018). Asymptotic inference about predictive accuracy using high frequency data. *Journal of Econometrics 203*(2), 223–240.

Liu, X., S. Shi, and J. Yu (2020). Persistent and rough volatility. *Available at SSRN 3724733*.

Ljung, G. M. and G. E. Box (1978). On a measure of lack of fit in time series models. *Biometrika 65*(2), 297–303.

Lobato, I. N., J. C. Nankervis, and N. Savin (2002). Testing for zero autocorrelation in the presence of statistical dependence. *Econometric Theory 18*(3), 730–743.

McLeod, A. I. and K. W. Hipel (1978). Preservation of the rescaled adjusted range: 1. a reassessment of the Hurst Phenomenon. *Water Resources Research 14*(3), 491–508.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica 71*(4), 1027–1048.

Nakamura, E. and J. Steinsson (2018a). High-frequency identification of monetary non-neutrality: The information effect. *Quarterly Journal of Economics 133*(3), 1283–1330.

Nakamura, E. and J. Steinsson (2018b). Identification in macroeconomics. *Journal of Economic Perspectives 32*(3), 59–86.

Perron, P. and Z. Qu (2010). Long-memory and level shifts in the volatility of stock market return indices. *Journal of Business & Economic Statistics 28*(2), 275–290.

Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika 74*(3), 535–547.

Phillips, P. C. B. (1989). Partially identified econometric models. *Econometric Theory 5*(2), 181–240.

Phillips, P. C. B. (1999). Discrete fourier transforms of fractional processes. *Cowles Foundation Discussion Paper No.1243 Yale University*.

Phillips, P. C. B. (2023). Estimation and inference with near unit roots. *Econometric Theory 39*, 221–263.

Phillips, P. C. B. and S. Jin (2021). Business cycles, trend elimination, and the HP filter. *International Economic Review 62*(2), 469–520.

Phillips, P. C. B. and T. Magdalinos (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics 136*(1), 115–130.

Poskitt, D. S., G. M. Martin, and S. D. Grose (2017). Bias correction of semiparametric long memory parameter estimators via the prefiltered sieve bootstrap. *Econometric Theory 33*(3), 578–609.

Potter, K. W. (1976). Evidence for nonstationarity as a physical explanation of the Hurst Phenomenon. *Water Resources Research 12*(5), 1047–1052.

Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics 85*(4), 777–792.

Risk Staff (2021, February). Quants of the year – jim gatheral and mathieu rosenbaum. *Risk.net*.

Robinson, P. M. (1978). Statistical inference for a random coefficient autoregressive model. *Scandinavian Journal of Statistics 5*, 163–168.

Robinson, P. M. (1995a). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics 23*(5), 1630–1661.

Robinson, P. M. (1995b). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics 23*(4), 1048–1072.

Schennach, S. M. (2018). Long memory via networking. *Econometrica 86*(6), 2221–2248.

Shi, S. and J. Yu (2023). Volatility puzzle: Long memory or anti-persistency. *Management Science 7*(69), 3861–3883.

Shimotsu, K. and P. C. B. Phillips (2005). Exact local whittle estimation of fractional integration. *The Annals of Statistics 33*(4), 1890–1933.

Solo, V. (1992). Intrinsic random functions and the paradox of 1/f noise. *SIAM Journal on Applied Mathematics 52*(1), 270–291.

Staiger, D. and J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Stock, J. H. and J. H. Wright (2000). GMM with weak identification. *Econometrica 68*(5), 1055–1096.

Stock, J. H. and M. Yogo (2005). *Testing for Weak Instruments in Linear IV Regression*, pp. 80–108. New York: Cambridge University Press.

Tanaka, K. (2013). Distributions of the maximum likelihood and minimum contrast estimators associated with the fractional ornstein–uhlenbeck process. *Statistical Inference for Stochastic Processes 16*(3), 173–192.

Tauchen, G. E. and M. Pitts (1983). The price variability-volume relationship on speculative markets. *Econometrica 51*(2), 485–505.

Velasco, C. and P. M. Robinson (2000). Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association 95*(452), 1229–1243.

Wang, X., W. Xiao, and J. Yu (2023a). Asymptotic properties of least squares estimator in local to unity processes with fractional Gaussian noises. *Advances in Econometrics 45A*, 73–95.

Wang, X., W. Xiao, and J. Yu (2023b). Modeling and forecasting realized volatility with the fractional Ornstein-Uhlenbeck process. *Journal of Econometrics 232*(2), 389–415.

Wang, X. and J. Yu (2023). Latent local-to-unity models. *Econometric Reviews 42*(7), 586–611.

Internet Appendix to 'Weak Identification of Long Memory with Implications for Volatility Modelling' *

Jia Li†, Peter C. B. Phillips††, Shuping Shi†††, Jun Yu†

†Singapore Management University

††Yale University, University of Auckland, Singapore Management University

†††Macquarie University

October 4, 2023

Abstract

This internet appendix contains three sections. Section S3 provides details of the datasets in our empirical analysis. Section S4 presents various robustness checks for the empirical findings in the main paper. Section S6 provides additional forecasting results.

S1 Proof of Theorem 1

Proof. Throughout this proof K denotes a generic finite positive constant that may change from line to line but does not depend on T or parameter values in R_T or $\widetilde{R}_T$. Let $\theta = (\alpha_T, d_T) \in R_T$. Consider a generic sequence $\tilde{\alpha}_T$ such that $|\tilde{\alpha}_T| < \tilde{\gamma}_T$ and set $\theta' = (\tilde{\alpha}_T, d_T + 1)$. It is easy to see that $\theta' \in \widetilde{R}_T$. Hence,

$$\inf_{\tilde{\theta} \in \widetilde{R}_T} \sup_{\lambda} \left|\log f_{\theta}(\lambda) - \log f_{\tilde{\theta}}(\lambda)\right| \leq \sup_{\lambda} \left|\log f_{\theta}(\lambda) - \log f_{\theta'}(\lambda)\right|, \tag{S1.1}$$

where we have written $\sup_{\lambda}$ in place of $\sup_{\underline{\lambda}_T \leq |\lambda| \leq \pi}$ for brevity. By definition (recall (2.4)),

*Thanks go to the Editor and two referees for searching comments and many useful suggestions that assisted the revision of this paper. The authors also thank Yacine Aït-Sahalia, Torben Andersen, Isaiah Andrews, Federico M. Bandi, Tim Bollerslev, Xu Cheng, Frank Diebold, Jim Gatheral, Peter Hansen, Morten Nielsen, Mathieu Rosenbaum, and Zhongjun Qu for helpful discussions and comments. Shi acknowledges support from the Australian Research Council (Grant No. DE190100840). Yu acknowledges research/project support from the Ministry of Education, Singapore, under its Academic Research Fund (AcRF) Tier 2 (Award Number MOE-T2EP402A20-0002). Phillips acknowledges research support from the NSF (Grant No. SES 18-50860) and a Kelly Fellowship at the University of Auckland. Jia Li, School of Economics, Singapore Management University. Email: jiali@smu.edu.sg. Peter C. B. Phillips, Yale University, University of Auckland, University of Southampton & Singapore Management University; Email: peter.phillips@yale.edu. Shuping Shi, Department of Economics, Macquarie University; Email: shuping.shi@mq.edu.au. Jun Yu, School of Economics and Lee Kong Chian School of Business, Singapore Management University. Email: yujun@smu.edu.sg.

$$\log f_{(\alpha_T,d_T)}(\lambda) = \log\left(\frac{\sigma^2}{2\pi}\right) - d_T\log\left(2-2\cos(\lambda)\right) - \log\left(1-2\alpha_T\cos(\lambda)+\alpha_T^2\right),$$
$$\log f_{(\tilde{\alpha}_T,d_T+1)}(\lambda) = \log\left(\frac{\sigma^2}{2\pi}\right) - (d_T+1)\log\left(2-2\cos(\lambda)\right) - \log\left(1-2\tilde{\alpha}_T\cos(\lambda)+\tilde{\alpha}_T^2\right),$$

so that $\sup_\lambda \left|\log f_{(\alpha_T,d_T)}(\lambda) - \log f_{(\tilde{\alpha}_T,d_T+1)}(\lambda)\right|$ is

$$\begin{aligned}
&\sup_\lambda \left|\log\left(\frac{2-2\cos(\lambda)}{1-2\alpha_T\cos(\lambda)+\alpha_T^2}\right) + \log\left(1-2\tilde{\alpha}_T\cos(\lambda)+\tilde{\alpha}_T^2\right)\right| \\
&\leq \sup_\lambda \left|\log\left(\frac{2-2\cos(\lambda)}{1-2\alpha_T\cos(\lambda)+\alpha_T^2}\right)\right| + \sup_\lambda \left|\log\left(1-2\tilde{\alpha}_T\cos(\lambda)+\tilde{\alpha}_T^2\right)\right|. \qquad \text{(S1.2)}
\end{aligned}$$

Since $\alpha_T \to 1$, we may assume that $\alpha_T \in (1/2, 2)$ without loss of generality. Hence, uniformly for λ satisfying $\underline{\lambda}_T \leq |\lambda| \leq \pi$,

$$\frac{2-2\cos(\lambda)}{1-2\alpha_T\cos(\lambda)+\alpha_T^2} \geq \frac{2-2\cos(\underline{\lambda}_T)}{9} > 0.$$

By the mean-value theorem, we further have

$$\begin{aligned}
&\sup_\lambda \left|\log\left(\frac{2-2\cos(\lambda)}{1-2\alpha_T\cos(\lambda)+\alpha_T^2}\right)\right| = \sup_\lambda \left|\log\left(\frac{2-2\cos(\lambda)}{1-2\alpha_T\cos(\lambda)+\alpha_T^2}\right) - \log(1)\right| \\
&\leq \frac{9}{2-2\cos(\underline{\lambda}_T)} \sup_\lambda \left|1-2(1-\alpha_T)\cos(\lambda)-\alpha_T^2\right| \leq K\left(\frac{|1-\alpha_T|+|1-\alpha_T|^2}{\underline{\lambda}_T^2}\right). \qquad \text{(S1.3)}
\end{aligned}$$

Similarly, since $\tilde{\alpha}_T \to 0$, $1-2\tilde{\alpha}_T\cos(\lambda)+\tilde{\alpha}_T^2 \to 1$ uniformly for all λ, and so, is uniformly bounded away from zero. Applying the mean-value theorem again yields

$$\sup_\lambda \left|\log\left(1-2\tilde{\alpha}_T\cos(\lambda)+\tilde{\alpha}_T^2\right)\right| \leq K\left(|\tilde{\alpha}_T|+\tilde{\alpha}_T^2\right). \qquad \text{(S1.4)}$$

Combining (S1.1)–(S1.4) yields

$$\inf_{\tilde{\theta}\in\widetilde{R}_T} \sup_\lambda \left|\log f_\theta(\lambda) - \log f_{\tilde{\theta}}(\lambda)\right| \leq K\left(\underline{\lambda}_T^{-2}|1-\alpha_T| + |\tilde{\alpha}_T|\right) = O\left((\underline{\lambda}_T^{-2}\gamma_T) \vee \tilde{\gamma}_T\right).$$

The upper bound in the above display holds for a constant K independent of θ and so for the sup over $\theta \in R_T$, which implies that $\delta(\widetilde{\mathcal{M}}_T, \mathcal{M}_T) = O\left((\underline{\lambda}_T^{-2}\gamma_T) \vee \tilde{\gamma}_T\right)$. By symmetry, $\delta(\mathcal{M}_T, \widetilde{\mathcal{M}}_T) = O\left((\underline{\lambda}_T^{-2}\gamma_T) \vee \tilde{\gamma}_T\right)$ and the theorem follows. ■

S2 Simulations: Identification-robust confidence sets

We apply the proposed identification-robust inference method in a Monte Carlo setting that is designed to illustrate the intuition discussed in Section 2.2 and, at the same time, match some key patterns seen in empirical work, including our own study in Section 3. We generate the observed y_t series from the ARFIMA$(1, d, 0)$ model (2.1) for different (α, d) parameter values, with the ε_t error terms simulated as i.i.d. standard normal variables.[1] Specifically, we consider $d = -0.4$ or 0.4 under which the process u_t

[1] Since the inference procedure is scale-invariant, the scale parameter σ is set to unity without loss of generality.

exhibits roughness or long-memory, respectively. Whether the model is weakly or strongly identified depends critically on the value of the autoregressive coefficient α. Accordingly, we consider a broad range of configurations for this parameter by varying its value over the set $\mathcal{A} = \{-0.2, -0.1, 0, \ldots, 0.9\} \cup \{0.995\}$, with the point $\alpha = 0.995$ being representative of the near-unity region.[2]

Under each configuration, we compute the 95%-level robust confidence set for (α, d) as described in Algorithm 1. Test inversion is carried out via a grid search over the set $[-1, 1] \times [-1, 1]$. This candidate set is sufficiently wide so that empirical estimates seen in the prior literature are not ruled out *a priori*. To keep the computation manageable, we discretize each dimension of the parameter space with mesh size 0.01. Since the near-unity region for the autoregressive parameter α is of special importance, we refine the mesh size for α down to 0.001 when $\alpha \in [0.99, 1]$.

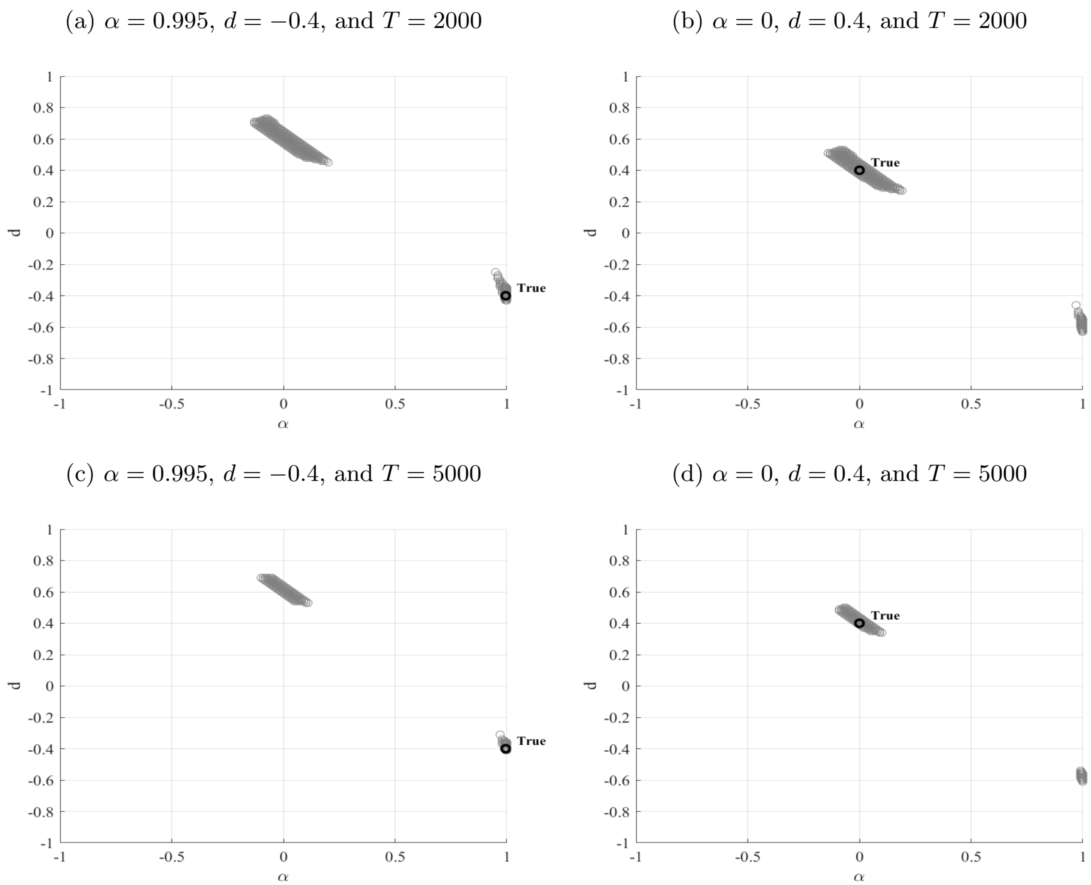
It is instructive to illustrate the workings of the proposed confidence set for a single random draw in the Monte Carlo experiment. Figure S1 plots the estimated confidence sets constructed for a single sample path in each of four Monte Carlo configurations. Specifically, we consider two sample sizes, $T = 2,000$ or $T = 5,000$, that are in line with the real datasets used in our empirical study. For each sample size, we consider two parameter configurations $(\alpha, d) = (0.995, -0.4)$ or $(0, 0.4)$, which are representative of the empirical estimates in prior studies that support rough or long-memory volatility dynamics, respectively (see, e.g., Gatheral et al. (2018) and Andersen et al. (2003)). Also note that these configurations directly mirror the two parameterizations analyzed in Theorem 1.

Inspection of the plotted confidence sets for all four settings in Figure S1 reveals that they share a common 'bifurcation' pattern in which there are two disjoint regions irrespective of which region actually contains the true parameters. One region features near-unity α and $d < 0$ (signifying roughness), while the other features near-zero α and $d > 0$ (signifying long-memory). Viewed through the lens of robust confidence set inference, neither of these two possibilities can be ruled out at the given confidence level, despite the fact that the parameter values seem highly disjoint and individually very different between the two regions. Although this indeterminacy may be disconcerting and unsatisfying from a practical viewpoint, it reflects the intrinsic difficulty arising from the near indistinguishability of the two parameter schemes, given the available data and the focus on autocovariances. Increasing the sample size from $2,000$ to $5,000$ sharpens the individual regions in the confidence sets as may be expected but it does not eliminate the bifurcation phenomenon.

The pattern depicted in these illustrations is representative. This may be shown by overlaying the plots in Figure S1 across all 1000 Monte Carlo trials. More precisely, for each candidate parameter value on the (α, d) plane, we compute the frequency that it falls in the confidence set; we then plot these coverage rates as a heatmap, where darker colors represent higher frequencies. For brevity, we focus on the case $T = 5,000$, which is roughly the average sample size for datasets used in our empirical work. The top row of Figure S2 plots the coverage rate heatmaps for $(\alpha, d) = (0.995, -0.4)$ and $(\alpha, d) = (0, 0.4)$, as in the illustrative examples. The bifurcation pattern is again self-evident, suggesting that the identification-robust confidence sets generally contain those two disjoint regions. While our approach does not estimate any parameter, our findings reinforce what Shi and Yu (2023)

[2]This value matches the estimate reported in Shi and Yu (2023) for the S&P 500 ETF (SPY) from January 2010 to May 2021 based on the Whittle method.

Figure S1: Identification-Robust Confidence Sets: A One-Path Illustration



(a) $\alpha = 0.995$, $d = -0.4$, and $T = 2000$

(b) $\alpha = 0$, $d = 0.4$, and $T = 2000$

(c) $\alpha = 0.995$, $d = -0.4$, and $T = 5000$

(d) $\alpha = 0$, $d = 0.4$, and $T = 5000$

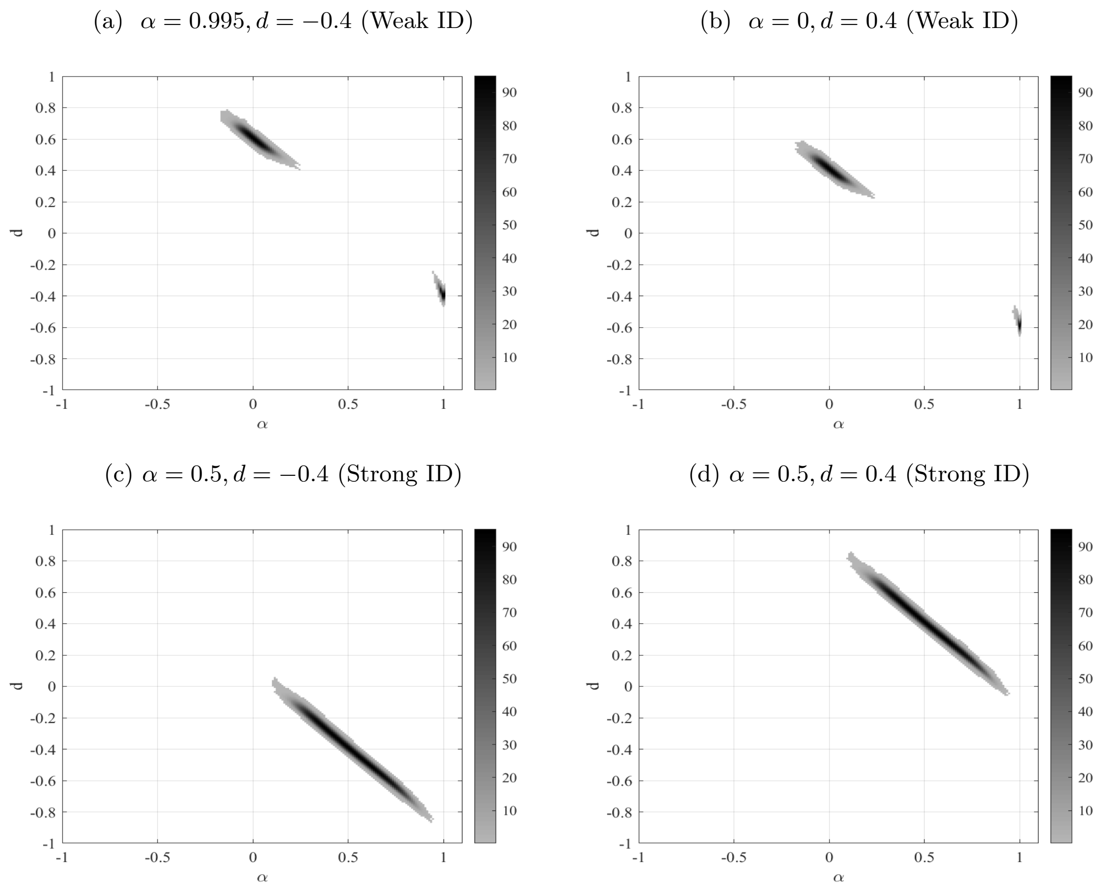
Data are simulated from the ARFIMA$(1, d, 0)$ model (2.1) with different parameter settings. The figure displays the 95% identification-robust confidence set of the four data series, with α on the x-axis and d on the y-axis. The black circle indicates the true model parameters.

found when maximum likelihood methods are used.

For comparison, we plot heatmaps for the true parameter values $(\alpha, d) = (0.5, -0.4)$ or $(0.5, 0.4)$ in the bottom row of Figure S2. From the analysis in Section 2.2, weak identification is mainly relevant when α is near unity or near zero. Therefore, the two configurations with $\alpha = 0.5$ are expected to deliver strong identification and this behavior is evident in the plotted heatmaps. Indeed, an 'average' confidence set for (α, d) has the familiar (single-region) elliptic shape and is centered at the true parameter value, precisely what is expected in classical likelihood or moment-based inference. On the other hand, confidence sets under strong identification are not necessarily small. Instead, weak identification is revealed through nonstandard shapes, such as the bifurcation pattern seen here in the robust confidence sets, rather than by the size of the confidence set. Readers are referred to the literature for more discussion of these differences (Staiger and Stock, 1997; Stock and Wright, 2000; Stock and Yogo, 2005; Andrews and Cheng, 2012; Andrews et al., 2019).

To reveal the incidence of bifurcation Figure S3 plots its frequency of occurrence as a function of the true value of the autoregressive coefficient $\alpha \in \mathcal{A}$ while fixing $d = -0.4$ (left) or $d = 0.4$ (right).

Figure S2: Identification-Robust Confidence Sets: Coverage Rates

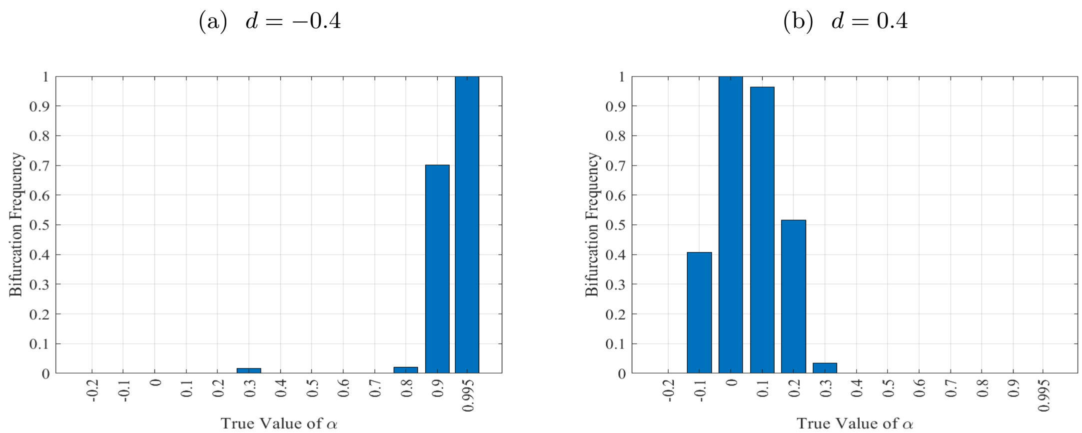


For each parameter setting, 1000 data series were simulated from the ARFIMA(1,d,0) model (2.1) and the 95% identification-robust confidence set computed for each series. The graphs show the frequency of a given (α, d) falling in the 95% identification-robust confidence set. Darker shading colors imply more frequent presence of the value in the confidence set. ID is short for identification.

For brevity the $T = 5,000$ case is reported. In the left panel where $d = -0.4$ the confidence set almost always contains two disjoint regions when $\alpha = 0.995$, just as in panel (a) of Figure S2. When $\alpha = 0.9$, the bifurcation frequency drops to approximately 70%, suggesting that weak identification is still largely in play. As the true value of α moves further away from the near-unity region, the bifurcation frequency drops essentially to zero. The overall pattern is consistent with the intuition that when $d < 0$ the parameters tend to be weakly identified when α is near unity. Mirroring this finding, the right panel of Figure S3 shows that when $d > 0$, weak identification is more severe when α is close to zero, complementing intuition and Theorem 1.

In cases of weak identification, the long-memory region is generally larger than the rough region, a result that is expected for the following reasons. First, from Figure S3, in the case of rough volatility (panel (a)) the weak identification issue disappears almost completely when $\gamma_T = 0.2$ (i.e., $\alpha = 0.8$), whereas in the case of long-memory (panel (b)) the bifurcation frequency remains more than 50% when $\tilde{\gamma}_T = 0.2$ (i.e. $\alpha = 0.2$). The asymmetric impact of γ_T and $\tilde{\gamma}_T$ on weak identification is also manifest

Figure S3: Bifurcation Frequencies of the Identification-Robust Confidence Set



The autoregressive coefficient α takes a wide range of values (displayed on the horizontal axis). Under each parameter setting, we simulate 1000 data series from ARFIMA(1,d,0) and compute their 95% identification-robust confidence sets. The bifurcation frequency is the percentage of replications with two isolated areas in the confidence set. This figure plots the bifurcation frequency against the value of α when $d = -0.4$ (left panel) and $d = 0.4$ (right panel).

in Theorem 1, where γ_T is shown to have a more significant impact on the order magnitude of the distance $O((\underline{\lambda}_T^{-2}\gamma_T) \vee \tilde{\gamma}_T))$ than $\tilde{\gamma}_T$. It is therefore unsurprising that the long-memory region is larger than the rough volatility region. Second, we impose the restriction that the autoregressive coefficient $|\alpha| < 1$ and so the rough volatility region is truncated at $\alpha = 1$. Third, the power of the adaptive Portmanteau test may well differ in different regions of the parameter space. Alternative omnibus tests for serial correlation might be considered but are left to future work.

These Monte Carlo findings corroborate theory and intuition. In ARFIMA model simulations the (α, d) parameters show strong evidence of joint weak identification when α is near unity or near zero. In such cases, identification-robust confidence sets typically contain two distinct regions that exhibit the bifurcation pattern predicted by theory and guide the interpretation of our empirical findings presented in Section 3.

S3 Data Description

This section details the data series used in our empirical analysis. Tables S1 and S2 describe the two datasets for realized volatility (RV) measures employed in the empirical analysis of Section 3.1 of the main text. This includes ten ETFs and 30 Dow Jones Industrial Average stocks obtained from Dacheng Xiu's Risk Lab (see Table S1) and 31 internal stock market indices obtained from the Oxford–Man Realized Library (see Table S2). Table S3 provides summary statistics for the trading volume series employed in Section 3.2 of the main text.

Table S1: Summary Statistics for Log Realized Variances of U.S. Equities

Ticker	Start date	Nob	Mean	Std.	Skew.	Kurto.
S&P 500 market ETF (SPY)	03-Jan-1996	6164	-2.19	0.55	-0.20	8.62
Industry ETF: Material (XLB)	05-Jan-1999	5161	-1.98	0.65	-1.41	12.90
Industry ETF: Energy (XLE)	23-Dec-1998	5392	-1.81	0.62	-1.77	14.53
Industry ETF: Financial (XLF)	23-Dec-1998	5403	-1.95	0.65	-0.37	8.05
Industry ETF: Industrial (XLI)	05-Jan-1999	4943	-2.17	0.69	-1.64	14.19
Industry ETF: Technology (XLK)	23-Dec-1998	5403	-1.93	0.60	-0.20	8.33
Industry ETF: Consumer staples (XLP)	23-Dec-1998	5380	-2.40	0.70	-2.01	16.23
Industry ETF: Utilities (XLU)	24-Dec-1998	5108	-2.10	0.64	-1.67	14.36
Industry ETF: Health care (XLV)	05-Jan-1999	5013	-2.25	0.64	-2.58	28.33
Industry ETF: consumer discretionary (XLY)	07-Jan-1999	4948	-2.17	0.71	-1.70	15.31
Dow Jones 30						
Apple Inc (AAPL)	03-Jan-1996	6174	-1.35	0.55	-0.34	9.40
Honeywell International Inc (ALD or HON)	03-Jan-1996	6152	-1.60	0.56	-0.51	9.32
Amgen Inc (AMGN)	03-Jan-1996	6172	-1.47	0.42	0.54	4.15
American Express Co (AEXP or AXP)	03-Jan-1996	6173	-1.56	0.58	0.04	8.99
Boeing Co (BA)	03-Jan-1996	6174	-1.53	0.46	0.49	3.90
Verizon Communications Inc (BEL or VZ)	03-Jan-1996	6145	-1.72	0.47	0.34	5.39
Caterpillar Inc (CAT)	03-Jan-1996	6173	-1.48	0.45	-0.35	12.10
Chevron Corp (CHV or CVX)	03-Jan-1996	6174	-1.68	0.43	0.12	8.35
Salesforce.Com Inc (CRM)	24-Jun-2004	4110	-1.27	0.45	0.37	3.71
Cisco Systems Inc (CSCO)	03-Jan-1996	6173	-1.45	0.47	0.47	3.08
Walt Disney Co (DIS)	03-Jan-1996	6172	-1.60	0.47	0.53	3.29
Goldman Sachs Group Inc (GS)	05-May-1999	5359	-1.48	0.47	1.01	4.46
Home Depot Inc (HD)	03-Jan-1996	6174	-1.56	0.47	0.58	3.29
International Business Machines Corp (IBM)	03-Jan-1996	6174	-1.72	0.47	0.64	3.28
Intel Corps (INTC)	03-Jan-1996	6172	-1.43	0.43	0.58	3.31
Johnson & Johnson (JNJ)	03-Jan-1996	6173	-1.88	0.45	0.60	3.47
JPMorgan Chase & Co (JPM)	03-Jan-1996	6153	-1.50	0.55	0.23	7.69
Coca-Cola Co (KO)	03-Jan-1996	6174	-1.84	0.46	0.60	3.40
McDonald's Corp (MCD)	03-Jan-1996	6173	-1.75	0.47	0.47	3.50
3M Co (MMM)	03-Jan-1996	6173	-1.75	0.47	0.21	5.27
Merck & Co Inc (MRK)	03-Jan-1996	6172	-1.65	0.42	0.69	4.08
Microsoft Corp (MSFT)	03-Jan-1996	6172	-1.60	0.42	0.52	3.31
Nike Inc (NIKE)	03-Jan-1996	6174	-1.55	0.48	-0.13	11.20
Procter & Gamble Co (PG)	03-Jan-1996	6172	-1.83	0.46	0.76	3.96
Travelers Companies Inc (SPC or TRV)	03-Jan-1996	6153	-1.73	0.65	-1.05	11.17
UnitedHealth Group Inc (UNH)	03-Jan-1996	6156	-1.48	0.48	-0.09	9.40
Visa Inc (V)	20-Mar-2008	3183	-1.68	0.48	1.00	4.26
Walgreens Boots Alliance Inc (WAG or WBA)	03-Jan-1996	6141	-1.53	0.43	-0.29	9.54
Walmart Inc (WMT)	03-Jan-1996	6139	-1.73	0.49	0.22	5.99
Exxon Mobil Co (XOM)	03-Jan-1996	6345	-1.70	0.45	0.55	3.79

Notes: The realized volatility data are downloaded from the Risk Lab constructed by Dacheng Xiu. The end date is May 25, 2021. The top panel is for the ten index ETFs and the bottom panel is for the 30 Dow Jones industrial average stocks.

Table S2: Summary Statistics for Log Realized Variances of Stock Market Indices

Name (Ticker)	Start date	Nob	Mean	Std.	Skew.	Kurto.
AEX index (AEX)	03-Jan-2000	5459	-9.68	1.02	0.53	3.40
All Ordinaries (AORD)	04-Jan-2000	5409	-10.47	0.96	0.71	4.18
Bell 20 Index (BFX)	03-Jan-2000	5457	-9.81	0.93	0.53	3.41
S&P BSE Sensex (BSESN)	03-Jan-2000	5308	-9.44	0.98	0.60	3.69
PSI All-Share Index (BVLG)	15-Oct-2012	2193	-10.23	0.76	0.66	4.03
BVSP BOVESPA Index (BVSP)	03-Jan-2000	5268	-9.21	0.83	0.62	4.52
Dow Jones Industrial Average (DJI)	03-Jan-2000	5366	-9.86	1.12	0.43	3.52
CAC 40 (FCHI)	03-Jan-2000	5461	-9.48	0.99	0.39	3.28
FTSE MIB (FTMIB)	01-Jun-2009	3042	-9.50	0.88	0.36	3.59
FTSE 100 (FTSE)	04-Jan-2000	5400	-9.65	1.01	0.58	3.66
DAX (GDAXI)	03-Jan-2000	5427	-9.36	1.04	0.42	3.22
S&P/TSX Composite index (GSPTSE)	02-May-2002	4772	-10.30	1.10	0.90	4.47
HANG SENG Index (HSI)	03-Jan-2000	5245	-9.69	0.84	0.68	3.95
IBEX 35 Index (IBEX)	03-Jan-2000	5426	-9.33	0.95	0.17	3.04
Nasdaq 100 (IXIC)	03-Jan-2000	5370	-9.68	1.10	0.44	3.09
Korea Composite Stock Price Index (KS11)	04-Jan-2000	5269	-9.64	1.03	0.47	3.05
Karachi SE 100 Index (KSE)	03-Jan-2000	5216	-9.78	1.04	0.18	4.79
IPC Mexico (MXX)	03-Jan-2000	5370	-9.94	0.92	0.67	3.88
Nikkei 225 (N225)	02-Feb-2000	5205	-9.70	0.94	0.30	3.43
NIFTY 50 (NSEI)	03-Jan-2000	5300	-9.66	1.05	0.42	3.90
OMX Copenhagen 20 Index (OMXC20)	03-Oct-2005	3888	-9.59	0.89	1.10	5.15
OMX Helsinki All Share Index (OMXHPI)	03-Oct-2005	3929	-9.82	1.01	0.91	4.15
OMX Stockholm All Share Index (OMXSPI)	03-Oct-2005	3929	-9.95	1.03	0.81	3.97
Oslo Exchange All-share Index (OSEAX)	03-Sep-2001	4917	-9.56	0.97	0.78	4.11
Russel 2000 (RUT)	03-Jan-2000	5367	-10.07	1.00	0.33	7.28
Madrid General Index (SMSI)	04-Jul-2005	4055	-9.48	0.96	0.27	3.45
S&P 500 Index (SPX)	03-Jan-2000	5369	-9.89	1.15	0.38	3.39
Shanghai Composite Index (SSEC)	04-Jan-2000	5171	-9.38	1.06	0.40	3.06
Swiss Stock Market Index (SSMI)	04-Jan-2000	5363	-9.94	0.90	1.07	4.61
Straits Times Index (STI)	03-Jan-2000	3425	-10.05	0.68	0.56	3.91
EURO STOXX 50 (STOXX50E)	03-Jan-2000	5458	-9.38	1.05	0.01	5.70

Notes: The realized volatility data are downloaded from the Realized Library of the Oxford–Man Institute of Quantitative Finance. The end date is Oct 12, 2021.

Table S3: Summary Statistics: De-trended Log Trading Volume

Ticker	Start date	Nob	Mean	Std.	Skew.	Kurto.
S&P 500 market ETF (SPY)	01-Feb-1993	7137	-0.16	0.51	-0.66	6.86
Industry ETF: Material (XLB)	23-Dec-1998	5647	-0.30	0.81	-0.99	6.18
Industry ETF: Energy (XLE)	23-Dec-1998	5647	-0.25	0.66	-0.58	4.35
Industry ETF: Financial (XLF)	23-Dec-1998	5647	-0.24	0.66	-0.66	4.85
Industry ETF: Industrial (XLI)	23-Dec-1998	5646	-0.35	0.87	-1.04	5.81
Industry ETF: Technology (XLK)	23-Dec-1998	5647	-0.17	0.55	0.09	3.75
Industry ETF: Consumer staples (XLP)	23-Dec-1998	5647	-0.27	0.68	-0.12	4.47
Industry ETF: Utilities (XLU)	23-Dec-1998	5647	-0.28	0.73	-0.79	5.79
Industry ETF: Health care (XLV)	23-Dec-1998	5647	-0.32	0.76	-0.44	5.33
Industry ETF: consumer discretionary (XLY)	23-Dec-1998	5647	-0.38	0.89	-1.09	6.13
Dow-Jones 30						
Apple Inc (AAPL)	01-Feb-1993	7137	-0.13	0.47	0.39	4.40
Honeywell International Inc (ALD or HON)	01-Feb-1993	7137	-0.10	0.42	0.30	5.04
Amgen Inc (AMGN)	01-Feb-1993	7137	-0.11	0.43	0.54	6.58
American Express Co (AEXP or AXP)	01-Feb-1993	7137	-0.10	0.42	0.31	4.26
Boeing Co (BA)	01-Feb-1993	7137	-0.12	0.46	0.12	4.81
Verizon Communications Inc (BEL or VZ)	01-Feb-1993	7137	-0.09	0.39	0.54	6.02
Caterpillar Inc (CAT)	01-Feb-1993	7137	-0.10	0.42	0.29	4.20
Chevron Corp (CHV or CVX)	01-Feb-1993	7137	-0.07	0.34	0.27	4.67
Salesforce.Com Inc (CRM)	23-Jun-2004	4267	-0.15	0.51	0.48	4.94
Cisco Systems Inc (CSCO)	01-Feb-1993	7137	-0.09	0.40	0.11	6.99
Walt Disney Co (DIS)	01-Feb-1993	7137	-0.10	0.41	0.71	5.17
Goldman Sachs Group Inc (GS)	04-May-1999	5558	-0.12	0.44	0.32	4.92
Home Depot Inc (HD)	01-Feb-1993	7137	-0.09	0.40	0.48	4.50
International Business Machines Corp (IBM)	01-Feb-1993	7137	-0.09	0.39	0.58	4.67
Intel Corps (INTC)	01-Feb-1993	7137	-0.08	0.39	0.08	5.69
Johnson & Johnson (JNJ)	01-Feb-1993	7137	-0.07	0.36	0.41	5.03
JPMorgan Chase & Co (JPM)	01-Feb-1993	7137	-0.10	0.40	0.22	3.99
Coca-Cola Co (KO)	01-Feb-1993	7137	-0.07	0.36	0.41	4.42
McDonald's Corp (MCD)	01-Feb-1993	7137	-0.09	0.39	0.52	4.44
3M Co (MMM)	01-Feb-1993	7137	-0.09	0.39	0.56	4.74
Merck & Co Inc (MRK)	01-Feb-1993	7137	-0.09	0.39	0.65	4.99
Microsoft Corp (MSFT)	01-Feb-1993	7137	-0.08	0.39	0.35	4.68
Nike Inc (NIKE)	01-Feb-1993	7137	-0.12	0.46	0.48	4.76
Procter & Gamble Co (PG)	01-Feb-1993	7137	-0.09	0.39	0.72	6.51
Travelers Companies Inc (TRV)	01-Feb-1993	7137	-0.12	0.46	0.03	4.48
UnitedHealth Group Inc (UNH)	01-Feb-1993	7137	-0.13	0.47	0.41	5.03
Visa Inc (V)	19-Mar-2008	3326	-0.11	0.42	0.72	5.66
Walgreens Boots Alliance Inc (WAG or WBA)	01-Feb-1993	7137	-0.10	0.41	0.52	4.54
Walmart Inc (WMT)	01-Feb-1993	7137	-0.08	0.38	0.58	4.66
Exxon Mobil Co (XOM)	01-Feb-1993	7137	-0.06	0.33	0.31	5.12

Notes: The trading volume data are downloaded from *Yahoo Finance*. The end date is Oct 12, 2021.

S4 Empirical Weak Identification Robustness Checks

S4.1 Subsample analysis for international stock market indices

We perform a subsample analysis for the realized volatility (RV) measures of the 31 international stock market indices obtained from the Oxford–Man Realized Library. This complements the full-sample analysis shown in Figure 5 of the main text. The full sample is divided into two subsamples using the first day of 2010 as the cutoff. The confidence sets for α and d are plotted in the top (resp. bottom) row of Figure S4 for the 2000–2009 (resp. 2010–2021) subsample. Due to data availability, the first subsample for the FTSE MIB index (FTMIB) is very short, containing only seven months of data, and results in highly inaccurate estimates. In addition, the PSI all-share index (BVLG) is unavailable for the first subsample as the data only starts from 2012.

We find similar patterns as those in the full-sample analysis shown in Figure 5 of the main text. Indeed, we again observe that the confidence sets of some indices contain two disjoint regions, but many of them have only a single region associated with long memory. As expected, the bifurcation, or weak identification, phenomenon is more severe for the shorter subsamples. The number of bifurcated confidence sets increases from 9 to 18 (resp. 15) for the first (resp. second) subsample.

S4.2 Analysis using alternative realized volatility measures

Our main empirical analysis (see Section 3.1 of the main text) of the international stock market indices is based on the classical 5-minute RV from the Oxford–Man Realized Library. In this section, we further check whether the weak identification issue is specific for this particular measure by considering several alternative RV measures that are also available in the Realized Library. For brevity, we focus on the S&P 500 index (SPX). The list of RV measures and their summary statistics are reported in Table S4.

Table S4: Summary Statistics: log volatility estimators of SPX

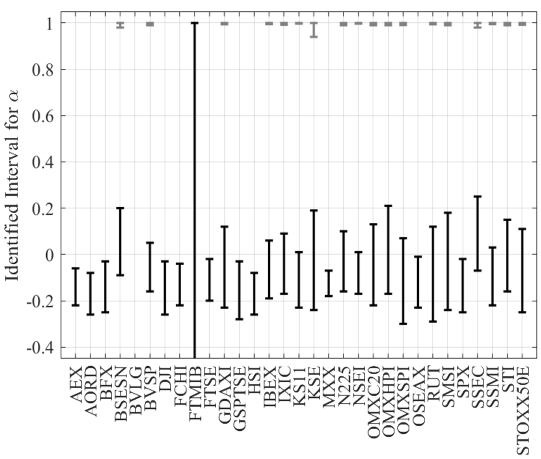
Name (Ticker)	Nob	Mean	Std.	Skew.	Kurto.
Median realized variance (5-min,*medrv*)	5369	-10.81	1.22	0.37	3.34
Realized kernel variance (Two-Scale/Bartlett, *rk_twoscale*)	5369	-10.00	1.10	0.43	3.46
Bipower variation (5-min sub-sampled, *bv_ss*)	5369	-10.09	1.14	0.44	3.46
Realized variance (5-min, *rv*5)	5369	-9.89	1.15	0.38	3.39
Realized kernel variance (Tukey-Hanning(2), *rk_th*2)	5369	-9.99	1.11	0.43	3.42
Bipower variation (5-min, *bv*)	5369	-10.09	1.14	0.44	3.46
Realized semi-variance (5-min, *rsv*)	5369	-10.69	1.24	0.33	3.26
Realized variance (10-min sub-sampled, *rv*10*_ss*)	5369	-9.90	1.17	0.35	3.35
Realized kernel variance (non-flat Parzen, *rk_parzen*)	5369	-10.01	1.22	0.29	3.25
Realized semi-variance (5-min sub-sampled, *rsv_ss*)	5369	-10.69	1.24	0.33	3.26
Realized variance (5-min, sub-sampled, *rv*5*_ss*)	5369	-9.89	1.15	0.38	3.39
Realized variance (10-min, rv10)	5369	-9.90	1.17	0.35	3.35

Notes: This table lists alternative estimators for integrated volatility of SPX available in the Realized Library. The start date is January 3, 2000 and the end date is October 12, 2021.

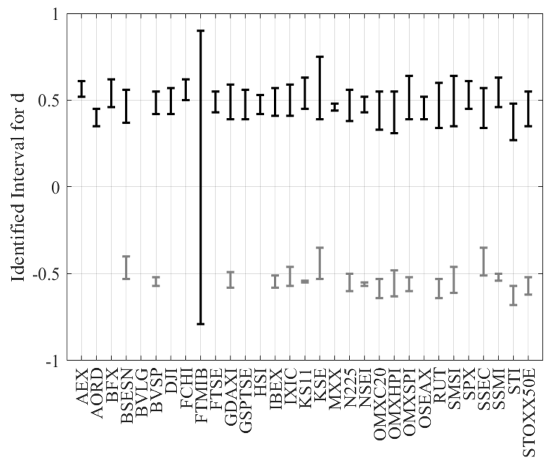
Figure S5 reports the projected one-dimensional confidence sets for the autoregressive parameter α and the memory parameter d, computed separately for each RV measure. These results are in line with those reported in the main text. Some confidence sets bifurcate with two disjoint regions, suggesting

Figure S4: Confidence Sets for International Stock Market Indices: Subsample Analysis

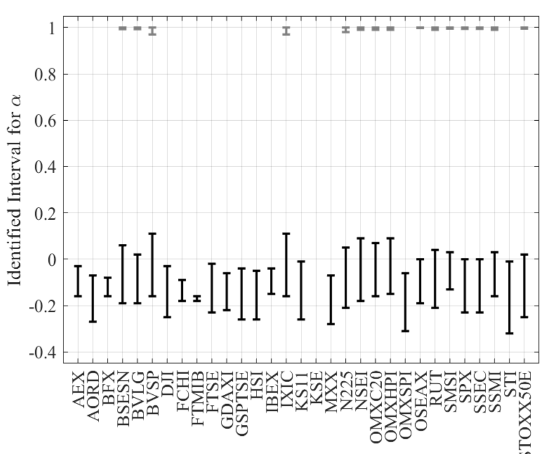
(a) 95% Confidence Set of α (2000–2009)



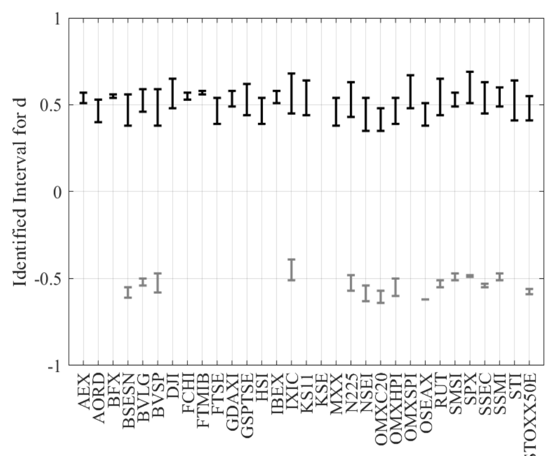
(b) 95% Confidence Set of d (2000–2009)



(c) 95% Confidence Set of α (2010–2021)



(d) 95% Confidence Set of d (2010–2021)

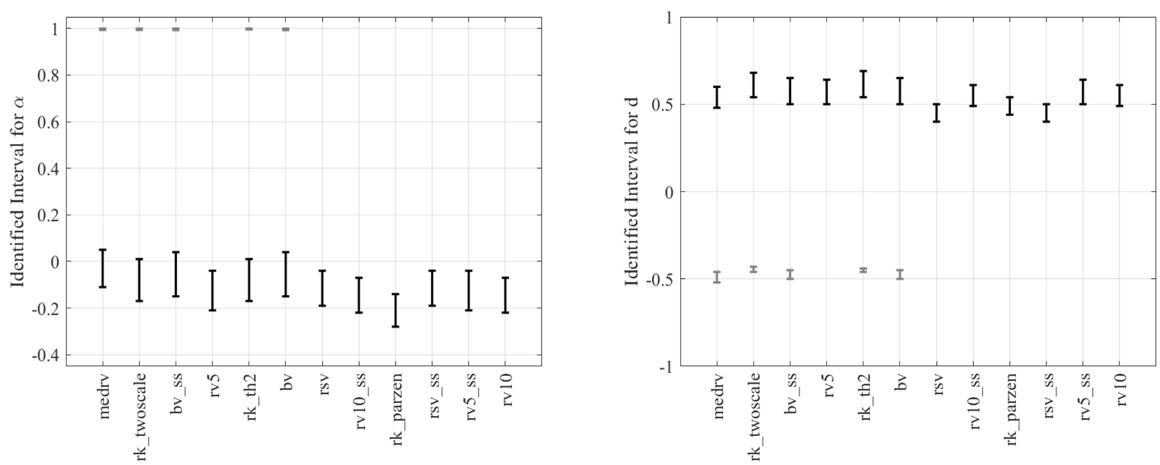


This figure presents the 95% identification-robust confidence sets for the 31 (demeaned) log RV of the international stock market indices for the period 2000-2009 in the first row and 2010-2021 in the second row. We project the confidence sets onto the α-axis on the left panels and the d-axis on the right panels. Labels on the x-axis are the tickers of the indices.

the presence of weak identification. When the confidence set contains only one region, the evidence points to long memory. It is also interesting to note that the long memory regions of the confidence sets are relatively stable across different RV measures.

Figure S5: Confidence Sets for Alternative Realized Volatility Measures of SPX

(a) 95% Confidence Set of α (b) 95% Confidence Set of d



This figure presents the 95% identification robust confidence sets for the (demeaned) log volatility estimators for SPX. The left (right) panel shows the projection of the confidence sets on the α-axis (d-axis) for each asset. Labels on the x-axis are the labels of different volatility estimators with their names in Table S4.

S4.3 Twitter Economic Uncertainty Index

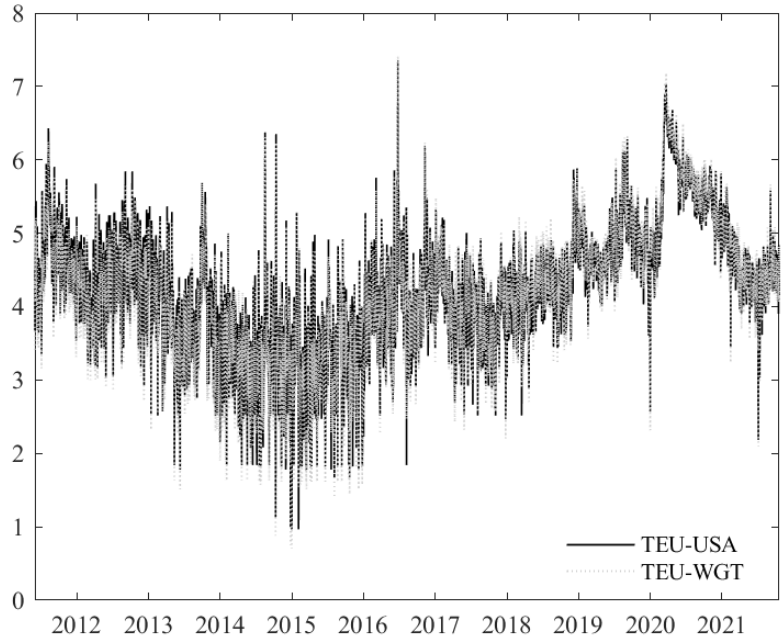
As an alternative measure of information flow, we examine the Twitter-based Economic Uncertainty (TEU) index, which is publicly available from the Economic Policy Uncertainty website.[3] The TEU index is constructed based on Twitter messages containing keywords related to uncertainty and economy – see Baker et al. (2021) for details. Since our empirical analysis mainly concerns the U.S. equity market, we use the TEU-USA index, which is built on tweets originating from the U.S., and its attention-weighted variant TEU-WGT.[4] The TEU indices are available at the daily frequency and our sample spans the period June 1, 2011 to July 24, 2021, giving a sample size of $T = 3,789$ observations. Figure S6 plots these two TEU indices in logarithmic units. Unsurprisingly, they are highly correlated (correlation coefficient $= 0.996$) and highly persistent.

Following the same procedure as before, we compute identification-robust confidence sets of (α, d) for the two (log-transformed and then demeaned) TEU indices. The confidence sets are plotted in Figure S7. Interestingly, the confidence sets for these TEU indices contain only one region, with memory parameter around 0.4 and autoregressive coefficient taking negative values but close to zero. Specifically, the confidence intervals for d are $[0.4, 0.45]$ for TEU-USA and $[0.36, 0.44]$ for TEU-WGT,

[3] https://www.policyuncertainty.com/twitter_uncert.html

[4] The TEU-WGT index adjusts the weight of each tweet according to the number of its re-tweets.

Figure S6: Twitter Economic Uncertainty Indices



The daily Twitter economic uncertainty indices run from June 1, 2011 to July 24, 2021. TEU-USA measures the uncertainty using tweets originating from the U.S. and TEU-WGT is its attention-weighted variant. The figure plots the two data series in logarithmic units.

with the corresponding intervals for α being $[-0.14, -0.04]$ and $[-0.14, -0.01]$. The results align with those obtained from the news and social media buzz indices provided by Refinitive.

S5 Forecasting Method and Simulations

The forecasting method is given in Section S5.1 and the simulation setting and forecasting results are in Section S5.2.

S5.1 Forecasting method

The h-step-ahead predictor of the model can be written as

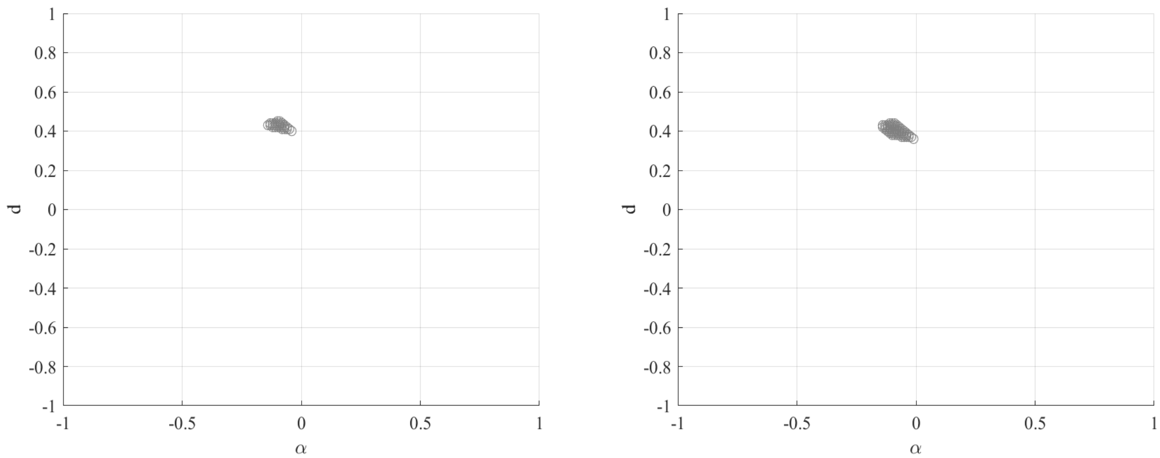
$$\hat{z}_{t+h} = \phi_{t,1}^{(h)}(\theta)\, z_t + \phi_{t,2}^{(h)}(\theta)\, z_{t-1} + \cdots + \phi_{t,t}^{(h)}(\theta)\, z_1 \text{ with } t > 1. \tag{S5.1}$$

Let $\phi_t^{(h)}(\theta) = (\phi_{t,1}^{(h)}(\theta), \ldots, \phi_{t,t}^{(h)}(\theta))'$. Optimal prediction (in the sense of minimum mean squared forecast error) is achieved when

$$\phi_t^{(h)}(\theta) = \mathbf{\Gamma}_t(\theta)^{-1}\, \gamma_{t,h}(\theta), \tag{S5.2}$$

Figure S7: Confidence Sets for Twitter-Based Economic Uncertainty Indices

(a) TEU-USA (b) TEU-WGT



The left panel displays the 95% identification-robust confidence set for the Twitter economic uncertainty index (TEU-US) and the right panel for its attention weighted variant (TEU-WGT). The x-axis and y-axis correspond to α and d, respectively.

where $\gamma_{t,h}(\theta) = (\gamma_{z,h}(\theta), \ldots, \gamma_{z,h+t-1}(\theta))'$ and $\mathbf{\Gamma}_t(\theta) \equiv [\gamma_{z,i-j}(\theta)]_{i,j=1}^{t}$ with $\gamma_{z,k}(\theta) \equiv Cov(z_t, z_{t-k})$. Simple calculation reveals the theoretical mean squared prediction errors, which have the form

$$\mathbb{E}(z_{t+h} - \hat{z}_{t+h})^2 = \gamma_{z,0}(\theta) - \gamma_{t,h}(\theta)' \mathbf{\Gamma}_t(\theta)^{-1} \gamma_{t,h}(\theta).$$

The forecasting framework requires stationarity of the model. If the model parameter $d \in (-0.5, 0.5)$ (i.e., stationary), we have $z_{t+h} = y_{t+h}$. When $d \in [0.5, 1]$, the process y_t is nonstationary and can be rewritten as

$$(1 - \alpha L)\Delta y_t = \sigma (1 - L)^{-(d-1)} \varepsilon_t,$$

so that stationarity is achieved by taking first differences. Instead of forecasting y_{t+h} directly, we forecast the first differenced series Δy_{t+h} (i.e., $z_{t+h} = \Delta y_{t+h}$) and compute $\hat{y}_{t+h}$ as $\hat{y}_{t+h} = y_t + \sum_{j=1}^{h} \hat{z}_{t+j}$.

Computation of the $\gamma_{z,k}(\theta)$ is based on the following property for stationary processes (**?**):

$$\gamma_{z,h}(\theta) = \sum_{s=-\infty}^{\infty} \tilde{\gamma}_s(\alpha)\, \gamma_{u,h-s}\left(d^{\dagger}\right), \tag{S5.3}$$

where $\tilde{\gamma}_s(\alpha)$ is the autocovariance of the pure AR component, $d^{\dagger}$ is the fractional parameter associated with z_t, and $\gamma_{u,k}\left(d^{\dagger}\right)$ is the autocovariance of the fractional integrated noise of z_t and takes the form of

$$\gamma_{u,k}\left(d^{\dagger}\right) = \frac{(-1)^k \Gamma\left(1 - 2d^{\dagger}\right)}{\Gamma\left(k - d^{\dagger} + 1\right)\Gamma\left(1 - k - d^{\dagger}\right)}. \tag{S5.4}$$

The summand in (S5.3) is truncated at a finite value K, following the rule recommended by Shi and Yu (2023). That is, $K = 200$ for $\alpha \leq 0.9$, $K = 300$ for $0.9 < \alpha \leq 0.95$, $K = 1700$ for $0.95 < \alpha \leq 0.99$, $K = 3000$ for $0.99 < \alpha \leq 0.995$, and $K = 7000$ for $0.995 < \alpha < 1$. These settings for K strike a balance

Table S5: Model Forecasting: Rolling window implementation with $T_0 = 251 \times 5$ and $T = 251 \times 10$. The number of out-of-sample forecasts is 1,256. The simulation is repeated 200 times. MSFE and MAFE are computed over all out-of-sample forecasts and replications. MCS10 (MCS25) shows proportions of replications that the model MSC p-value is above the 10% (25%) threshold (i.e., in the best model set). The best models are highlighted in bold.

	$\alpha = 0.995, d = -0.4$						$\alpha = 0, d = 0.4$					
	SFE			AFE			SFE			AFE		
	MSFE	MCS10	MCS25	MAFE	MCS10	MCS25	MSFE	MCS10	MCS25	MAFE	MCS10	MCS25
$h = 1$												
Rough $\hat{\theta}_R$	**1.01**	**0.89**	**0.79**	**0.80**	**0.92**	**0.84**	1.06	0.30	0.23	0.82	0.35	0.29
Long memory $\hat{\theta}_L$	1.01	0.85	0.77	0.80	0.87	0.74	**1.01**	**0.96**	**0.93**	**0.80**	**0.96**	**0.92**
$h = 5$												
Rough $\hat{\theta}_R$	1.96	0.78	0.62	1.12	0.80	0.66	1.61	0.20	0.17	1.01	0.23	0.17
Long memory $\hat{\theta}_L$	**1.92**	**0.92**	**0.80**	**1.11**	**0.92**	**0.83**	**1.34**	**0.96**	**0.95**	**0.92**	**0.97**	**0.96**
$h = 21$												
Rough $\hat{\theta}_R$	3.00	0.79	0.65	1.38	0.78	0.67	2.22	0.21	0.17	1.18	0.23	0.17
Long memory $\hat{\theta}_L$	**2.91**	**0.85**	**0.78**	**1.36**	**0.85**	**0.77**	**1.53**	**0.96**	**0.95**	**0.99**	**0.97**	**0.95**

between computational cost and estimation accuracy.

In practical work the true model parameters contained in θ are unknown. So forecasting analysis is based on the three estimated model parameters, $\hat{\theta}$, $\hat{\theta}_R$, and $\hat{\theta}_L$.

S5.2 Forecasting Simulation results

The data generating process (DGP) used is the ARFIMA$(1, d, 0)$ model with two parameter settings that reflect the different models: $(\alpha, d) = (0.995, -0.4)$ and $(\alpha, d) = (0, 0.4)$. For each parameter setting, 200 sample paths were generated from the DGPs, each containing 10 years of daily data (i.e., $T = 251 \times 10$). The parameters of the models were estimated from a rolling five-year window, comprising $T_0 = 251 \times 5 = 1,255$ observations. We consider three forecasting horizons: the one-day $h = 1$, one-week $h = 5$, and one-month $h = 21$. The loss functions employed are squared forecast error (SFE), $L^S_{t,h} = (\hat{y}_{t+h} - y_{t+h})^2$ and absolute forecast error (AFE), $L^A_{t,h} = |\hat{y}_{t+h} - y_{t+h}|$. As the ARFIMA model is designed for log RV and the simulated data can have negative values, the QLIKE criteria cannot be applied in this context.

Table S5 reports out-of-sample forecasting results based on $\hat{\theta}$, $\hat{\theta}_R$, and $\hat{\theta}_L$. It provides the mean SFE (MSFE) and mean AFE (MAFE), computed from all rolling window forecasts and across all simulations, together with the percentages of replications that the model is identified to be in the best model set by the MCS criterion with thresholds 10% (MCS10) and 25% (MCS25), corresponding to 90% and 75% confidence levels, respectively.

The findings are unequivocal: in terms of forecasting performance, the long memory specification is superior to the rough model specification irrespective of the true data generating process. As evident in the table, when the true DGP is a long memory process, the long memory model provides superior forecasting results at all three horizons by a large margin according to both squared and absolute losses, making it a clear winner. When data are generated from a rough process, the long memory model outperforms the rough specification significantly for the week-ahead and month-ahead forecasts, and performs only slightly worse than the rough model for the one-day-ahead forecast with identical MSFE and MAFE but narrowly lower MCS inclusion frequencies.

S6 Forecasting Empirical Results

Tables S6-S11 display the mean squared forecasting errors (MSFEs), QLIKE and realized utilities of RV of the 10 ETFs and the 30 Dow Jones industrial average stocks at three different forecasting horizons. Boldface text is used to signify the best performing model in each case.

Table S6: One-day-ahead Forecasting Performance: Selected ETFs

	MSFE			QLIKE			Realized Utilities			
Ticker	Rough	LM	HAR-RV	Rough	LM	HAR-RV	Rough	LM	HAR-RV	GARCH
SPY	0.001	**0.001**	0.002	-1.217	**-1.217**	-1.208	0.0394	**0.0394**	0.0392	0.0376
XLB	0.004	0.003	**0.003**	-0.899	**-0.906**	-0.906	0.0391	**0.0393**	0.0392	0.0378
XLE	0.005	**0.004**	0.004	-0.692	**-0.698**	-0.698	0.0394	**0.0395**	0.0395	0.0371
XLF	0.006	**0.006**	0.006	-0.928	**-0.932**	-0.926	0.0391	**0.0392**	0.0390	0.0366
XLI	0.004	0.003	**0.003**	-1.044	-1.052	**-1.053**	0.0390	**0.0392**	0.0391	0.0378
XLK	**0.004**	0.004	0.005	**-1.056**	-1.056	-1.047	**0.0391**	0.0391	0.0389	0.0377
XLP	0.004	0.003	**0.002**	-1.303	-1.320	**-1.331**	0.0386	0.0388	**0.0390**	0.0351
XLU	0.004	0.003	**0.003**	-0.988	-0.995	**-0.996**	0.0391	0.0393	**0.0393**	0.0371
XLV	0.004	**0.003**	0.003	-1.163	-1.177	**-1.178**	0.0388	**0.0391**	0.0391	0.0378
XLY	0.004	**0.003**	0.003	-1.063	**-1.071**	-1.071	0.0390	**0.0391**	0.0391	0.0379

This table reports the MSFEs, QLIKEs, and realized utilities of the one-day-ahead rolling window forecast for the ten index ETFs using the ARFIMA model with $\hat{\theta}_R$ ('Rough') and $\hat{\theta}_L$ ('Long memory'), the HAR-RV model, and the GARCH model. The rolling window size is five years. Boldface text signifies the best performing model.

Table S7: One-week-ahead Forecasting Performance: Selected ETFs

	MSFE			QLIKE			Realized Utilities			
Ticker	Rough	LM	HAR-RV	Rough	LM	HA-RV	Rough	LM	HAR-RV	GARCH
SPY	0.003	**0.003**	0.004	-1.187	**-1.190**	-1.174	0.0388	**0.0388**	0.0385	0.0374
XLB	0.007	**0.006**	0.006	-0.873	**-0.884**	-0.880	0.0386	**0.0388**	0.0387	0.0375
XLE	0.008	**0.006**	0.008	-0.672	**-0.682**	-0.674	0.0389	**0.0392**	0.0390	0.0369
XLF	0.010	**0.009**	0.010	-0.896	**-0.904**	-0.894	0.0384	**0.0385**	0.0383	0.0362
XLI	0.006	**0.005**	0.005	-1.015	**-1.027**	-1.022	0.0384	**0.0387**	0.0385	0.0374
XLK	0.006	**0.006**	0.007	-1.027	**-1.029**	-1.016	**0.0386**	0.0385	0.0383	0.0374
XLP	0.005	**0.004**	0.004	-1.277	-1.299	**-1.302**	0.0380	**0.0384**	0.0383	0.0353
XLU	0.006	**0.005**	0.006	-0.965	**-0.975**	-0.970	0.0386	**0.0388**	0.0387	0.0370
XLV	0.006	**0.004**	0.005	-1.133	**-1.154**	-1.148	0.0382	**0.0386**	0.0384	0.0377
XLY	0.006	**0.005**	0.005	-1.032	**-1.045**	-1.040	0.0384	**0.0386**	0.0384	0.0376

This table reports the MSFEs, QLIKEs, and realized utilities of the one-week-ahead rolling window forecast for the ten index ETFs using the ARFIMA model with $\hat{\theta}_R$ ('Rough') and $\hat{\theta}_L$ ('Long memory'), the HAR-RV model, and the GARCH model. The rolling window size is five years. Boldface text signifies the best performing model.

Table S8: One-month-ahead Forecasting Performance: Selected ETFs

	MSFE			QLIKE			Realized Utilities			
Ticker	Rough	LM	HAR-RV	Rough	LM	HAR-RV	Rough	LM	HAR-RV	GARCH
SPY	0.006	**0.006**	0.007	-1.132	**-1.133**	-1.120	**0.0375**	0.0374	0.0372	0.0365
XLB	0.011	**0.010**	0.011	-0.834	**-0.844**	-0.837	0.0377	**0.0378**	0.0377	0.0366
XLE	0.013	**0.012**	0.014	-0.634	**-0.642**	-0.625	0.0381	**0.0383**	0.0379	0.0362
XLF	0.015	**0.014**	0.015	-0.848	**-0.854**	-0.852	**0.0374**	0.0374	0.0374	0.0354
XLI	0.009	**0.008**	0.009	-0.968	**-0.976**	-0.964	0.0374	**0.0375**	0.0372	0.0363
XLK	**0.009**	0.009	0.010	**-0.989**	-0.987	-0.975	**0.0378**	0.0375	0.0373	0.0369
XLP	0.007	0.006	**0.006**	-1.241	-1.254	**-1.257**	0.0371	**0.0373**	0.0373	0.0348
XLU	0.009	**0.009**	0.010	-0.927	**-0.932**	-0.922	0.0377	**0.0378**	0.0376	0.0364
XLV	0.007	**0.007**	0.007	-1.094	**-1.107**	-1.097	0.0374	**0.0376**	0.0373	0.0371
XLY	0.009	**0.008**	0.009	-0.984	**-0.995**	-0.986	0.0374	**0.0375**	0.0373	0.0367

This table reports the MSFEs, QLIKEs, and realized utilities of the one-month-ahead rolling window forecast for the ten index ETFs using the ARFIMA model with $\hat{\theta}_R$ ('Rough') and $\hat{\theta}_L$ ('Long memory'), the HAR-RV model, and the GARCH model. The rolling window size is five years. Boldface text signifies the best performing model.

Table S12 displays the MSFE, QLIKE, and realized utilities for the one-quarter-ahead forecast using the rough, long memory, and HAR models for the 30 Dow Jones industrial average stocks. We also include the GARCH model in the horse race using the realized utility measure. The best performing model is highlighted in boldface. All measures indicate the superior performance of the long memory model over the rough and HAR models.

S7 Forecasting Robustness Check: Alternative Volatility Measures

As a robustness check, we perform a five-year rolling window forecasting analysis using alternative RV measures for the S&P 500 index (SPX), which are listed in Table S4. The results of the one-day-ahead forecast based on the squared and QLIKE losses are presented in Figure S8, displaying the MCS p-values. It is evident that the conclusions remain consistent when using different measures of realized volatility.

Table S9: One-day-ahead Forecasting Performance: Dow Jones Industrial Average Stocks

	MSFE			QLIKE			Realized Utilities			
Ticker	Rough	LM	HAR-RV	Rough	LM	HAR-RV	Rough	LM	HAR-RV	GARCH
AAPL	0.007	**0.007**	0.008	-0.419	**-0.424**	-0.415	0.0393	**0.0394**	0.0392	0.0340
ALD	0.006	**0.006**	0.007	-0.629	**-0.631**	-0.623	0.0393	**0.0394**	0.0392	0.0370
AMGN	0.004	**0.004**	0.005	-0.511	**-0.511**	-0.506	0.0395	**0.0395**	0.0394	0.0375
AXP	0.006	**0.005**	0.006	-0.599	**-0.600**	-0.593	0.0394	**0.0394**	0.0393	0.0364
BA	**0.006**	0.006	0.006	-0.530	**-0.532**	-0.525	0.0394	**0.0394**	0.0393	0.0363
BEL	0.004	**0.004**	0.004	-0.736	**-0.738**	-0.733	0.0394	**0.0394**	0.0393	0.0382
CAT	0.005	**0.005**	0.006	-0.492	**-0.493**	-0.487	0.0395	**0.0395**	0.0394	0.0364
CHV	**0.003**	0.003	0.004	-0.686	**-0.686**	-0.680	0.0396	**0.0396**	0.0395	0.0382
CRM	**0.005**	0.005	0.007	-0.382	**-0.382**	-0.374	0.0394	**0.0394**	0.0393	0.0339
CSCO	0.004	**0.004**	0.005	-0.521	**-0.523**	-0.516	0.0395	**0.0395**	0.0394	0.0357
DIS	0.005	**0.005**	0.005	-0.619	**-0.620**	-0.613	0.0394	**0.0394**	0.0393	0.0380
GS	0.007	**0.007**	0.010	-0.523	**-0.524**	-0.516	0.0395	**0.0395**	0.0394	0.0368
HD	0.005	**0.005**	0.005	-0.604	**-0.605**	-0.599	0.0395	**0.0395**	0.0393	0.0381
IBM	0.003	**0.003**	0.003	-0.801	**-0.802**	-0.796	0.0395	**0.0395**	0.0394	0.0378
INTC	0.004	**0.004**	0.005	-0.456	**-0.457**	-0.450	0.0395	**0.0396**	0.0394	0.0373
JNJ	0.003	**0.003**	0.003	-0.938	**-0.939**	-0.934	0.0394	**0.0394**	0.0393	0.0372
JPM	0.008	**0.008**	0.010	-0.530	**-0.531**	-0.522	0.0394	**0.0394**	0.0393	0.0371
KO	0.003	**0.002**	0.003	-0.909	**-0.911**	-0.905	0.0394	**0.0395**	0.0393	0.0384
MCD	0.004	**0.004**	0.004	-0.789	**-0.793**	-0.788	0.0393	**0.0394**	0.0393	0.0385
MMM	0.003	**0.003**	0.004	-0.794	**-0.795**	-0.788	0.0394	**0.0394**	0.0393	0.0379
MRK	0.005	**0.005**	0.006	-0.665	**-0.666**	-0.659	0.0393	**0.0393**	0.0392	0.0384
MSFT	0.003	**0.003**	0.004	-0.633	**-0.634**	-0.628	0.0396	**0.0396**	0.0394	0.0373
NIKE	0.004	**0.004**	0.005	-0.608	**-0.609**	-0.603	0.0394	**0.0394**	0.0393	0.0374
PG	0.004	**0.004**	0.005	-0.921	**-0.922**	-0.916	0.0394	**0.0394**	0.0393	0.0381
SPC	0.007	**0.007**	0.007	-0.706	**-0.707**	-0.703	0.0392	**0.0393**	0.0392	0.0382
UNH	0.006	**0.006**	0.007	-0.516	**-0.517**	-0.511	0.0393	**0.0393**	0.0392	0.0378
V	0.003	**0.003**	0.004	-0.830	**-0.831**	-0.824	0.0394	**0.0394**	0.0393	0.0381
WAG	0.005	**0.005**	0.006	-0.534	**-0.535**	-0.528	0.0393	**0.0393**	0.0392	0.0369
WMT	0.003	**0.003**	0.003	-0.809	**-0.811**	-0.805	0.0394	**0.0395**	0.0394	0.0384
XOM	0.003	**0.003**	0.004	-0.727	**-0.728**	-0.722	0.0395	**0.0396**	0.0395	0.0369

This table reports the MSFEs, QLIKEs, and realized utilities of the one-day-ahead rolling window forecast for the 30 Dow Jones stocks using the ARFIMA model with $\hat{\theta}_R$ ('Rough') and $\hat{\theta}_L$ ('Long memory'), the HAR-RV model, and the GARCH model. The rolling window size is five years. Boldface signifies the best performing model.

Table S10: One-week-ahead Forecasting Performance: Dow Jones Industrial Average Stocks

	MSFE			QLIKE			Realized Utilities			
Ticker	Rough	LM	HAR-RV	Rough	LM	HAR-RV	Rough	LM	HAR-RV	GARCH
AAPL	0.015	**0.010**	0.012	-0.375	**-0.403**	-0.395	0.0384	**0.0389**	0.0388	0.0328
ALD	0.011	**0.009**	0.011	-0.601	**-0.613**	-0.600	0.0387	**0.0389**	0.0387	0.0359
AMGN	0.007	**0.007**	0.008	-0.496	**-0.499**	-0.491	0.0391	**0.0392**	0.0390	0.0372
AXP	0.011	**0.010**	0.012	-0.576	**-0.581**	-0.573	0.0389	**0.0390**	0.0388	0.0361
BA	0.011	**0.010**	0.012	-0.504	**-0.514**	-0.506	0.0388	**0.0390**	0.0388	0.0359
BEL	0.007	**0.006**	0.006	-0.714	**-0.724**	-0.718	0.0389	**0.0391**	0.0390	0.0380
CAT	0.009	**0.008**	0.009	-0.476	**-0.479**	-0.472	0.0391	**0.0392**	0.0390	0.0363
CHV	0.006	**0.006**	0.008	-0.667	**-0.669**	-0.658	0.0392	**0.0392**	0.0390	0.0379
CRM	0.010	**0.009**	0.010	-0.357	**-0.363**	-0.356	0.0389	**0.0390**	0.0389	0.0334
CSCO	0.008	**0.007**	0.008	-0.501	**-0.508**	-0.502	0.0391	**0.0392**	0.0391	0.0354
DIS	0.009	**0.008**	0.009	-0.597	**-0.603**	-0.595	0.0389	**0.0390**	0.0389	0.0377
GS	0.015	**0.014**	0.016	-0.497	**-0.501**	-0.493	0.0390	**0.0391**	0.0389	0.0364
HD	0.009	**0.008**	0.009	-0.582	**-0.589**	-0.580	0.0390	**0.0391**	0.0389	0.0378
IBM	0.005	**0.004**	0.006	-0.779	**-0.787**	-0.778	0.0390	**0.0392**	0.0390	0.0377
INTC	0.008	**0.007**	0.009	-0.438	**-0.442**	-0.436	0.0392	**0.0392**	0.0391	0.0371
JNJ	0.005	**0.004**	0.005	-0.916	**-0.923**	-0.915	0.0389	**0.0390**	0.0388	0.0376
JPM	0.015	**0.014**	0.017	-0.502	**-0.508**	-0.496	0.0388	**0.0389**	0.0387	0.0367
KO	0.004	**0.004**	0.005	-0.890	**-0.897**	-0.890	0.0390	**0.0391**	0.0390	0.0382
MCD	0.008	**0.006**	0.006	-0.757	**-0.779**	-0.774	0.0386	**0.0390**	0.0389	0.0383
MMM	0.006	**0.005**	0.006	-0.774	**-0.780**	-0.770	0.0390	**0.0391**	0.0388	0.0376
MRK	0.009	**0.008**	0.009	-0.644	**-0.649**	-0.641	0.0388	**0.0389**	0.0388	0.0379
MSFT	0.006	**0.005**	0.006	-0.615	**-0.618**	-0.609	0.0391	**0.0392**	0.0390	0.0370
NIKE	0.007	**0.007**	0.007	-0.587	**-0.594**	-0.589	0.0389	**0.0391**	0.0390	0.0370
PG	0.006	**0.006**	0.007	-0.900	**-0.907**	-0.897	0.0389	**0.0390**	0.0388	0.0384
SPC	0.013	**0.011**	0.012	-0.681	**-0.690**	-0.681	0.0387	**0.0389**	0.0387	0.0382
UNH	0.010	**0.009**	0.010	-0.498	**-0.503**	-0.498	0.0389	**0.0390**	0.0389	0.0376
V	0.006	**0.005**	0.007	-0.808	**-0.812**	-0.800	0.0389	**0.0390**	0.0387	0.0379
WAG	0.008	**0.007**	0.008	-0.517	**-0.521**	-0.515	0.0390	**0.0390**	0.0389	0.0368
WMT	0.005	**0.004**	0.005	-0.789	**-0.798**	-0.792	0.0390	**0.0392**	0.0390	0.0381
XOM	0.006	**0.006**	0.007	-0.706	**-0.711**	-0.700	0.0391	**0.0392**	0.0390	0.0373

This table reports the MSFEs, QLIKEs, and realized utilities of the one-week-ahead rolling window forecast for the 30 Dow Jones stocks using the ARFIMA model with $\hat{\theta}_R$ ('Rough') and $\hat{\theta}_L$ ('Long memory'), the HAR-RV model, and the GARCH model. The rolling window size is five years. Boldface signifies the best performing model.

Table S11: One-month-ahead Forecasting Performance: Dow Jones Industrial Average Stocks

	MSFE			QLIKE			Realized Utilities			
Ticker	Rough	LM	HAR-RV	Rough	LM	HAR-RV	Rough	LM	HAR-RV	GARCH
AAPL	0.031	**0.015**	0.016	-0.296	**-0.383**	-0.376	0.0368	**0.0383**	0.0382	0.0318
ALD	0.016	**0.013**	0.015	-0.559	**-0.586**	-0.569	0.0378	**0.0382**	0.0379	0.0341
AMGN	0.010	**0.010**	0.011	-0.476	**-0.483**	-0.474	0.0386	**0.0387**	0.0385	0.0365
AXP	0.019	**0.016**	0.021	-0.534	**-0.553**	-0.536	0.0379	**0.0382**	0.0379	0.0352
BA	0.021	**0.016**	0.018	-0.454	**-0.487**	-0.475	0.0378	**0.0383**	0.0381	0.0350
BEL	0.011	**0.009**	0.010	-0.678	**-0.701**	-0.694	0.0381	**0.0385**	0.0383	0.0377
CAT	0.014	**0.013**	0.015	-0.447	**-0.455**	-0.442	0.0385	**0.0386**	0.0383	0.0359
CHV	**0.011**	0.012	0.014	**-0.632**	-0.631	-0.613	**0.0384**	0.0383	0.0379	0.0372
CRM	0.018	**0.013**	0.014	-0.319	**-0.343**	-0.336	0.0382	**0.0385**	0.0384	0.0327
CSCO	0.015	**0.010**	0.012	-0.464	**-0.491**	-0.483	0.0382	**0.0387**	0.0385	0.0347
DIS	0.014	**0.012**	0.013	-0.557	**-0.579**	-0.569	0.0380	**0.0384**	0.0382	0.0369
GS	0.024	**0.022**	0.025	-0.459	**-0.465**	-0.458	0.0382	**0.0383**	0.0381	0.0354
HD	0.014	**0.012**	0.013	-0.543	**-0.565**	-0.556	0.0381	**0.0385**	0.0383	0.0370
IBM	0.010	**0.008**	0.010	-0.738	**-0.762**	-0.748	0.0381	**0.0385**	0.0382	0.0373
INTC	0.014	**0.011**	0.012	-0.409	**-0.426**	-0.418	0.0385	**0.0387**	0.0386	0.0368
JNJ	0.007	**0.007**	0.007	-0.876	**-0.896**	-0.884	0.0380	**0.0383**	0.0381	0.0374
JPM	0.025	**0.021**	0.023	-0.455	**-0.473**	-0.464	0.0378	**0.0381**	0.0380	0.0354
KO	0.007	**0.006**	0.007	-0.859	**-0.876**	-0.864	0.0382	**0.0385**	0.0383	0.0378
MCD	0.014	**0.008**	0.009	-0.694	**-0.758**	-0.747	0.0373	**0.0384**	0.0382	0.0377
MMM	0.009	**0.008**	0.009	-0.746	**-0.759**	-0.743	0.0383	**0.0385**	0.0382	0.0372
MRK	0.012	**0.011**	0.012	-0.610	**-0.626**	-0.620	0.0381	**0.0384**	0.0382	0.0376
MSFT	0.010	**0.009**	0.010	-0.583	**-0.592**	-0.581	0.0384	**0.0386**	0.0383	0.0367
NIKE	0.012	**0.010**	0.010	-0.558	**-0.575**	-0.570	0.0383	**0.0386**	0.0385	0.0365
PG	0.009	**0.009**	0.010	-0.864	**-0.880**	-0.868	0.0380	**0.0383**	0.0380	0.0378
SPC	0.020	**0.017**	0.019	-0.633	**-0.659**	-0.644	0.0375	**0.0381**	0.0378	0.0375
UNH	0.015	**0.014**	0.016	-0.470	**-0.484**	-0.472	0.0383	**0.0385**	0.0383	0.0369
V	**0.010**	0.010	0.010	-0.770	**-0.774**	-0.760	**0.0381**	0.0381	0.0377	0.0376
WAG	0.011	**0.010**	0.011	-0.497	**-0.507**	-0.498	0.0385	**0.0387**	0.0385	0.0368
WMT	0.008	**0.006**	0.007	-0.755	**-0.782**	-0.774	0.0382	**0.0387**	0.0385	0.0380
XOM	0.011	**0.010**	0.013	-0.664	**-0.676**	-0.658	0.0381	**0.0384**	0.0380	0.0373

This table reports the MSFEs, QLIKEs, and realized utilities of the one-month-ahead rolling window forecast for the 30 Dow Jones stocks using the ARFIMA model with $\hat{\theta}_R$ ('Rough') and $\hat{\theta}_L$ ('Long memory'), the HAR model, and the GARCH model. The rolling window size is five years. Boldface signifies the best performing model.

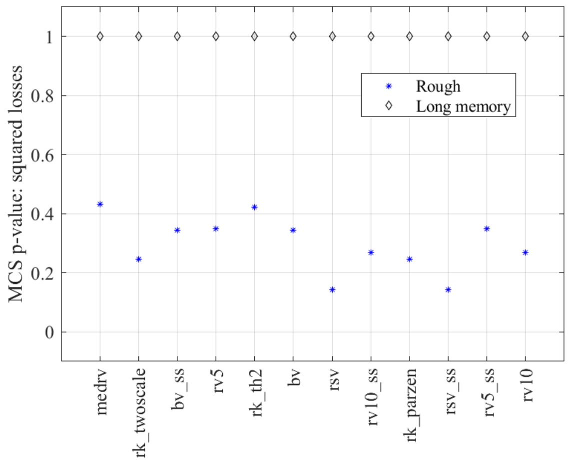
Table S12: One-quarter-ahead Forecasting Performance: Dow Jones Industrial Average Stocks

	MSFE			QLIKE			Realized Utilities			
Ticker	Rough	LM	HAR-RV	Rough	LM	HAR-RV	Rough	LM	HAR-RV	GARCH
AAPL	0.040	0.018	**0.018**	-0.258	**-0.379**	-0.379	0.0358	**0.0380**	**0.0380**	0.0312
ALD	0.019	**0.015**	0.017	-0.531	**-0.576**	-0.565	0.0370	**0.0378**	0.0376	0.0327
AMGN	0.012	**0.011**	0.012	-0.472	**-0.481**	-0.474	0.0384	**0.0385**	0.0384	0.0356
AXP	0.027	**0.022**	0.031	-0.498	**-0.531**	-0.507	0.0370	**0.0376**	0.0372	0.0343
BA	0.024	**0.020**	0.022	-0.432	**-0.468**	-0.445	0.0372	**0.0378**	0.0373	0.0339
BEL	0.015	**0.012**	0.013	-0.644	**-0.684**	-0.680	0.0372	0.0379	**0.0380**	0.0373
CAT	0.017	**0.017**	0.018	-0.429	**-0.436**	-0.422	0.0380	**0.0381**	0.0378	0.0355
CHV	**0.014**	0.017	0.016	**-0.604**	-0.590	-0.591	**0.0378**	0.0374	0.0375	0.0366
CRM	0.022	0.014	**0.014**	-0.295	-0.328	**-0.331**	0.0378	0.0383	**0.0384**	0.0320
CSCO	0.019	**0.013**	0.014	-0.440	**-0.490**	-0.480	0.0374	**0.0384**	0.0382	0.0346
DIS	0.017	**0.014**	0.016	-0.531	**-0.568**	-0.552	0.0373	**0.0379**	0.0377	0.0363
GS	**0.031**	0.032	0.036	-0.415	**-0.422**	-0.414	0.0374	**0.0375**	0.0374	0.0344
HD	0.019	**0.014**	0.017	-0.512	**-0.559**	-0.544	0.0372	**0.0381**	0.0379	0.0362
IBM	0.012	**0.010**	0.012	-0.717	**-0.751**	-0.729	0.0375	**0.0381**	0.0377	0.0370
INTC	0.018	**0.014**	0.016	-0.386	**-0.418**	-0.411	0.0377	**0.0383**	0.0382	0.0364
JNJ	0.009	**0.008**	0.008	-0.856	**-0.883**	-0.876	0.0374	**0.0379**	0.0378	0.0374
JPM	0.032	**0.027**	0.037	-0.424	**-0.452**	-0.433	0.0370	**0.0375**	0.0372	0.0336
KO	0.009	**0.008**	0.009	-0.834	**-0.860**	-0.850	0.0376	**0.0380**	0.0379	0.0376
MCD	0.018	**0.010**	0.010	-0.652	**-0.752**	-0.743	0.0363	**0.0381**	0.0380	0.0375
MMM	0.010	**0.009**	0.011	-0.732	**-0.752**	-0.730	0.0378	**0.0381**	0.0377	0.0363
MRK	0.015	**0.013**	0.014	-0.587	**-0.610**	-0.609	0.0375	0.0379	**0.0380**	0.0373
MSFT	0.012	**0.011**	0.012	-0.570	**-0.578**	-0.573	0.0380	**0.0380**	0.0380	0.0364
NIKE	0.014	**0.013**	0.014	-0.546	**-0.561**	-0.556	0.0379	**0.0381**	0.0380	0.0360
PG	0.011	**0.010**	0.011	-0.849	**-0.868**	-0.859	0.0377	**0.0379**	0.0377	0.0376
SPC	0.025	**0.022**	0.025	-0.593	**-0.630**	-0.610	0.0366	**0.0374**	0.0371	0.0365
UNH	0.020	**0.017**	0.019	-0.444	**-0.466**	-0.456	0.0377	**0.0381**	0.0379	0.0360
V	0.012	0.013	**0.010**	-0.747	-0.745	**-0.763**	0.0378	0.0374	**0.0378**	0.0377
WAG	0.012	**0.011**	0.013	-0.487	**-0.498**	-0.486	0.0382	**0.0384**	0.0382	0.0369
WMT	0.010	**0.007**	0.009	-0.732	**-0.777**	-0.762	0.0376	**0.0384**	0.0381	0.0381
XOM	**0.014**	0.014	0.014	-0.631	**-0.646**	-0.636	0.0374	**0.0376**	0.0375	0.0368

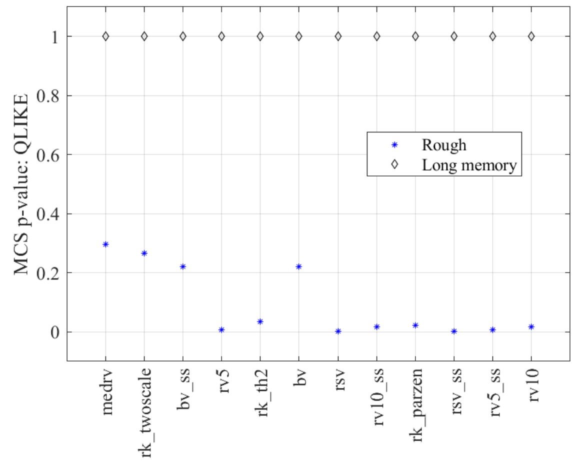
This table reports the MSFEs, QLIKEs, and realized utilities of the one-quarter-ahead rolling window forecast for the 30 Dow Jones stocks using the ARFIMA model with $\hat{\theta}_R$ ('Rough') and $\hat{\theta}_L$ ('Long memory'), the HAR-RV model, and the GARCH model. The rolling window size is five years. Boldface signifies the best performing model.

Figure S8: Forecasting with Alternative Volatility Measures for the S&P 500 Index: One-day-ahead

(a) MCS p-value: SFE



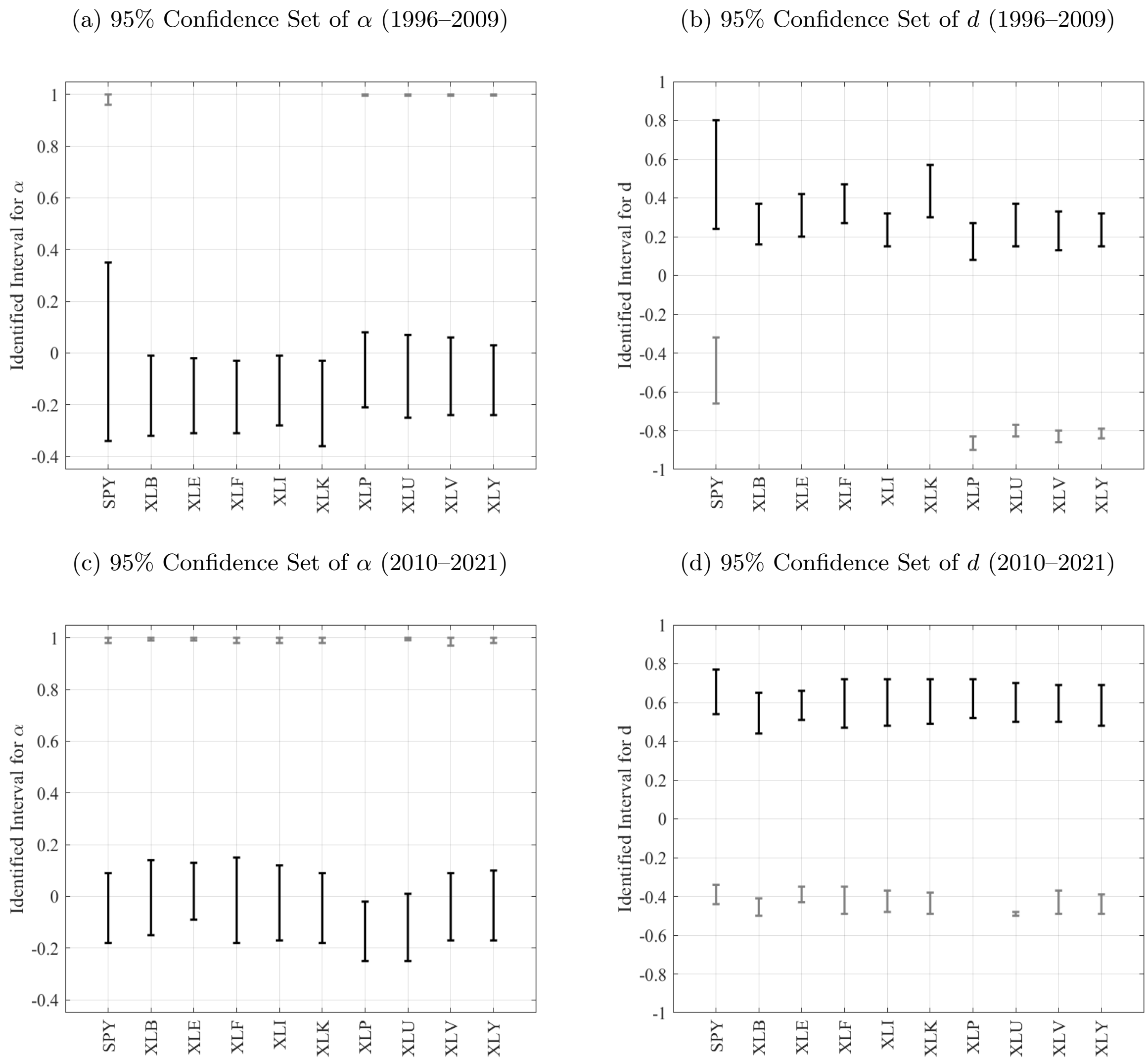
(b) MCS p-value: QLIKE



Five-year rolling window forecasts were conducted for the 12 alternative measures of RV for the period 2000-2021. We considered two forecasting models: the ARFIMA model with $\hat{\theta}_R$ ('Rough') and $\hat{\theta}_L$ ('Long memory'). The forecasting horizon is $h = 1$. The x-axis displays the abbreviated names of the estimators.

S8 Additional Figures

Figure S9: Confidence Sets for Selected ETFs: Subsample Analysis



(a) 95% Confidence Set of α (1996–2009)

(b) 95% Confidence Set of d (1996–2009)

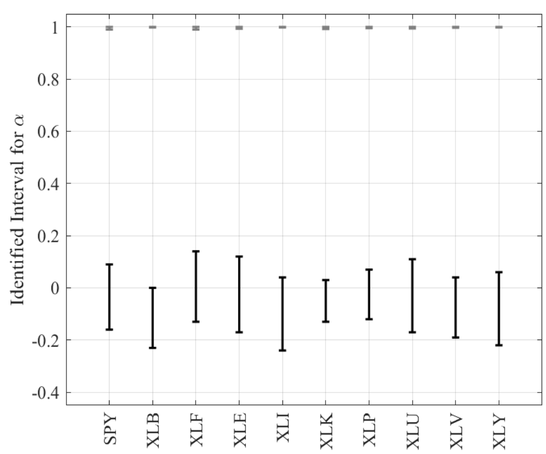
(c) 95% Confidence Set of α (2010–2021)

(d) 95% Confidence Set of d (2010–2021)

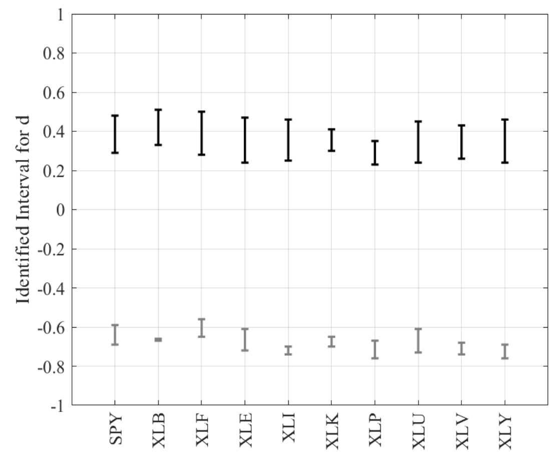
The figures present 95% identification-robust confidence sets for the (demeaned) log RV of the ten index ETFs for the period 1996–2009 in the first row and for 2010–2021 in the second row. The confidence sets are projected onto the α-axis in the left panels and the d-axis in the right panels. Labels on the x-axis are the tickers of the ETFs.

Figure S10: Confidence Sets for Trading Volume of Selected ETFs and Dow Jones Industrial Average (DJIA) Stocks

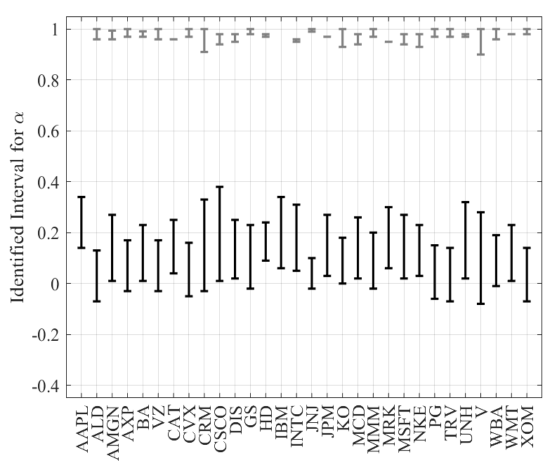
(a) Selected ETFs: 95% Confidence Set of α



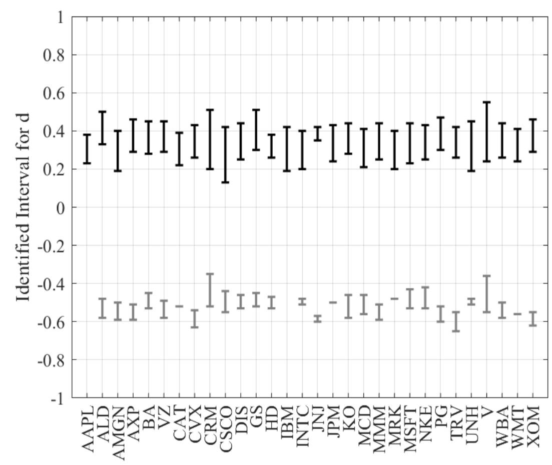
(b) Selected ETFs: 95% Confidence Set of d



(c) DJIA stocks: 95% Confidence Set of α



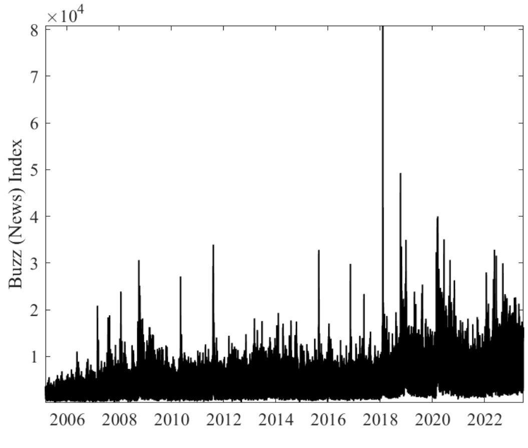
(d) DJIA stocks: 95% Confidence Set of d



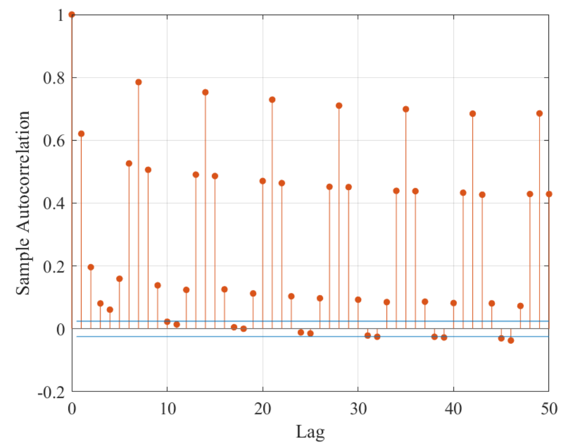
95% identification-robust confidence sets were computed for the (demeaned and de-trended) log trading volume of the ten index ETFs and the 30 Dow Jones industrial average stocks. The data are from 1993 to 2021 and downloaded from *Yahoo Finance*. The left (right) panel shows projections of the confidence sets on the α-axis (d-axis) for each asset. Labels on the x-axis are the tickers of the ETFs and stocks.

Figure S11: Refinitiv Buzz Indices

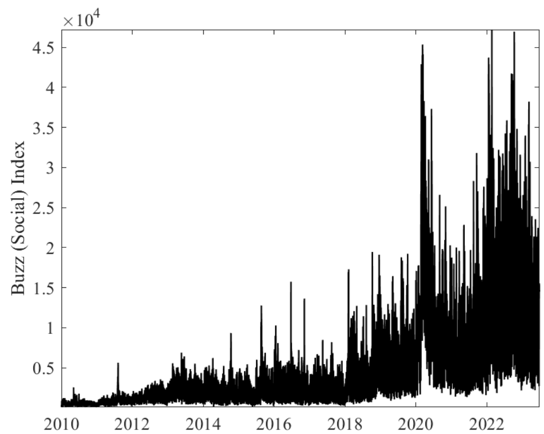
(a) Buzz Index: News



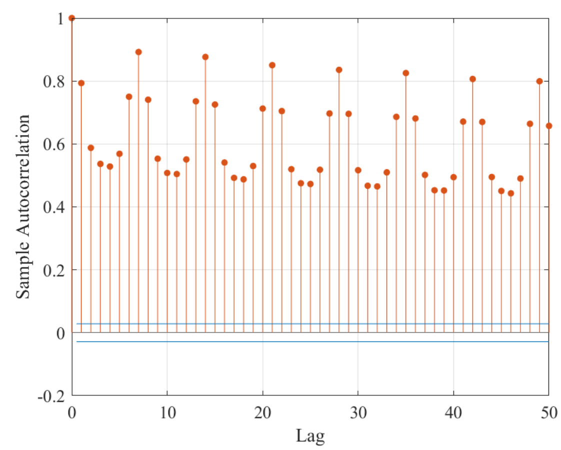
(b) ACF of the Buzz News Index



(c) Buzz Index: Social Media



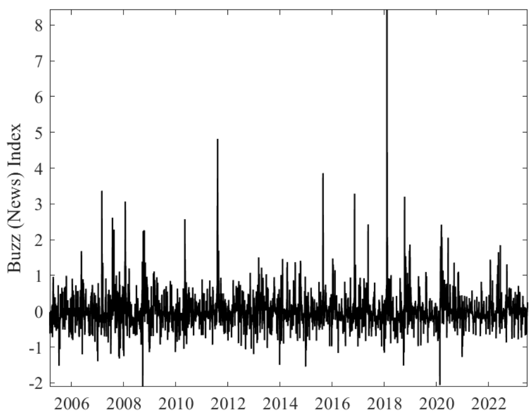
(d) ACF of the Buzz Social Media Index



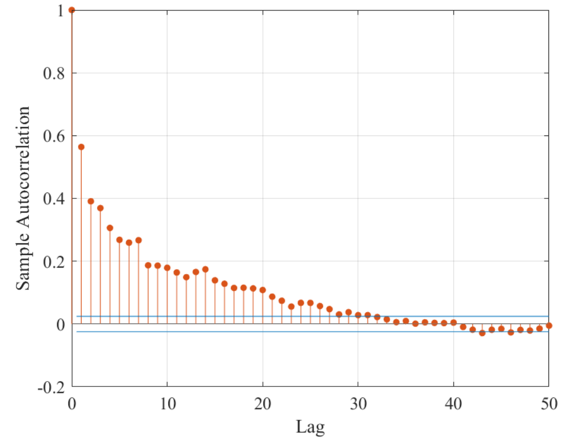
The daily buzz indices are obtained from Refinitiv MarketPsych country market database. The buzz news (resp. social media) index is derived by summing references to U.S. stocks sourced from news (resp. social media) outlets. The buzz news index starts from March 1, 2005, while the buzz social media index is from January 1, 2010. The end date of both data series is July 1, 2023. The two blue (horizontal) lines in the right panel represent the 95% confidence bounds for zero correlation.

Figure S12: Transformed Refinitiv Buzz Indices

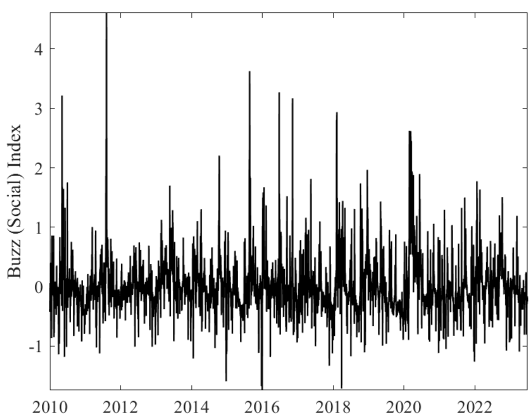
(a) Transformed Buzz News Index



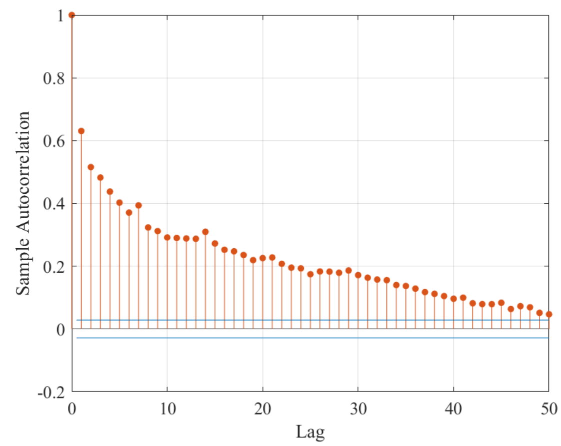
(b) ACF of Transformed Buzz News Index



(c) Transformed Buzz Social Media Index



(d) ACF of Transformed Buzz Social Media Index



This figure presents the transformed daily buzz indicess after detrending, seasonality removal, and demeaning. Detrending is accomplished through a two-sided moving average spanning 512 trading days, while to remove seasonality, we adopt the moving average technique introduced by Cleveland et al. (1990). The two blue (horizontal) lines in the right panel represent the 95% confidence bounds for zero correlation.