Weak Identification of Long Memory with Implications for Inference *

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

†Singapore Management University

††Yale University, University of Auckland, University of Southampton,

Singapore Management University

†††Macquarie University

June 18, 2022

Abstract

This paper explores weak identification issues arising in commonly used models of economic and financial time series. Two highly popular configurations are shown to be asymptotically observationally equivalent: one with long memory and weak autoregressive dynamics, the other with antipersistent shocks and a near-unit autoregressive root. We develop a data-driven semiparametric and identification-robust approach to inference that reveals such ambiguities and documents the prevalence of weak identification in many realized volatility and trading volume series. The identification-robust empirical evidence generally favors long memory dynamics in volatility and volume, a conclusion that is corroborated using social-media news flow data.

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

JEL classification: C12, C13, C58

*Thanks go to Xu Cheng, Frank Diebold, and Zhongjun Qu for helpful discussions. 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.

1 Introduction

For several decades models of cointegrated systems have enjoyed a vast range of empirical applications in economics and finance.[1] A leading condition that underlies much of the empirical research on such systems, particularly when vector autoregressive approaches are employed, is that the individual time series are integrated with most attention focused on simple $I(1)$ cointegrated systems. The $I(1)$ condition has proved extremely convenient in the development of an asymptotic theory of estimation and testing and has given rise to a wealth of empirical findings. Nonetheless, it is well understood that the $I(1)$ assumption is usually just an approximate representation of more general nonstationary processes within a wider class such as the $I(d)$ processes where the order of integration d may take on fractional values.

There is ample evidence in the literature for fractional processes taking positive values of the memory parameter d, yielding what is known as long memory. Nonstationary economic time series have spectral shapes in which the low frequency part of the spectrum is strongly dominant. This characteristic matches well with one of the fundamental empirical properties of an $I(d)$ process with $d > 0$, as noted in much early economic research (Adelman, 1965; Granger, 1966; Diebold and Rudebusch, 1989; Ding et al., 1993; Baillie et al., 1996). Correspondingly, considerable research has been devoted to explain this phenomenon in terms of more primitive generating mechanisms that have empirical justification. Simple unit root cases where $d = 1$ gain support from the theory of efficient markets. For more general $I(d)$ cases with $d \neq 1$, it is 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 and Mavroeidis, 2017), 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, the class of ARFIMA(p, d, q) models has

[1] Amongst a wide and diverse literature we mention the following short list of papers: Friedman and Kuttner (1992); Beaudry and Portier (2006); Ireland (2009); Bauer and Rudebusch (2020).

been widely used in empirical work to model economic time series that manifest both short and long memory as well as possible nonstationarity. The impulse response function implied by this general model with fractional integration is substantially different from that of an ARMA(p, q) model, allowing for long decays in responses. This feature of impulse responses is present in much financial data, making the class attractive in financial applications. 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.

In the large literature on volatility two different forms of estimated ARFIMA$(1, d, 0)$ models have been discovered in empirical research. First, log volatility is often found to be well modelled by an ARFIMA$(1, d, 0)$ model with $d > 0$. Among the many references and empirical contexts we mention the following studies: Baillie et al. (1996) for GARCH models; Comte and Renault (1996) and Breidt et al. (1998) for stochastic volatility models; Andersen and Bollerslev (1997), Andersen et al. (2001, 2003), Andersen, Bollerslev, Diebold, and Ebens (2001), Bandi and Perron (2006); Baillie et al. (2019) with realized volatility (RV); and Bandi and Perron (2006) with implied volatility. Popular estimation methods for d include semiparametric methods, such as local Whittle estimation (Künsch, 1987; Robinson, 1995a) and log periodogram regression (Geweke and Porter-Hudak, 1983; Robinson, 1995b). These two methods rely on the asymptotic behavior of periodogram ordinates at frequencies near zero and are valid in the stationary region with $-0.5 < d < 0.5$. A method that is asymptotically valid over wider regions and provides uniformly valid confidence intervals covering stationary and nonstationary cases ($d \geq 0.5$) is the exact local Whittle procedure (Shimotsu and Phillips, 2005). Estimated values of d with log volatility data usually turn out to be close to 0.5, the boundary point of nonstationarity for d, and these estimates are typically statistically significantly greater than zero and less than unity. After semiparametric estimation of d and data filtering to remove the fractional component, first order autoregressive estimation can be conducted on the filtered series, which often leads to autoregressive coefficient estimates that are close to zero (Andersen et al., 2003; Shi and Yu, 2022).

Second, ample empirical evidence now points towards the same ARFIMA$(1, d, 0)$ model but with a negative value for d (so-called antipersistence) producing what is called 'rough-

volatility' with an autoregressive parameter taken to be unity or near unity.[2] Amongst many studies that support such a model we mention the following: Gatheral et al. (2018); Bayer et al. (2016); Wang et al. (2021); Bolko et al. (2022); Liu et al. (2020); Fukasawa et al. (2021). 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. The 2021 Risk Awards were presented for introducing rough-volatility models – see the Risk website for the citation;[3] and the Rough Volatility website[4] has a collection of some 200 papers in this rapidly growing literature. In spite of this considerable interest only a single study to our knowledge has advanced an economic microstructure foundation to explain the rough-volatility phenomenon (El Euch et al., 2018).

These two empirical findings seem contradictory. Yet they reveal that this simple model has dual capabilities of matching the data. Only two parameters in the model, the autoregressive parameter α and the memory parameter d, control dependency. The empirical evidence above indicates that nonstationarity/near-nonstationarity in the data may be captured either by a long-memory parameter $d \approx 0.5$ (the nonstationarity border for d) or an autoregressive parameter α near unity. The first group of studies point to the values $(\alpha \approx 0, d \approx 0.5)$ where autoregressive effects are small (and often negative) but there is strong long-run dependence in the data. The second group point to values of α near unity capturing nonstationary/near-nonstationarity dependence coupled with $d < 0$ which captures volatility roughness or possible slight overdifferencing in the raw data.

Inspired by what is now an extensive literature on weak identification (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), 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

[2]The rough-volatility literature, pioneered by Gatheral et al. (2018), use a class of continuous-time models driven by the fractional Brownian motion with the 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).

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

[4]https://sites.google.com/site/roughvol/home/risks-1.

where the autoregressive coefficient lies either near unity or near zero.[5] More specifically, we show that the 'distance' between a near-unit-root ARFIMA$(1, d, 0)$ model and a near-zero-root ARFIMA$(1, d+1, 0)$ model goes to zero when nearness parameters shrink.[6] The two seemingly distinct parameterizations that have been studied extensively in the literature therefore generate observationally nearly equivalent dynamics. The consequences of such observational 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. Further, 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). 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 (2022).

To address this issue we propose using an identification-robust confidence set obtained by inverting tests for zero serial correlation in the model-implied residual series. The implied inferences are 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 sample of realized volatility and trading volume data 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. This highlights the empirical difficulty in determining whether a time series is driven by long-memory disturbances with

[5] 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.

[6] The result is unsurprising upon noting that ARFIMA model identification failure occurs when the nonstationary long-memory operator $(1-\alpha_1 L)^{-1}(1-L)^{-d_1}$, with $\alpha_1 = 0$ and $d_1 = 0.5$ takes the equivalent ARFIMA form $(1-\alpha_2 L)^{-1}(1-L)^{-d_2}$ with $\alpha_2 = 1$ and $d_2 = -0.5$. See (2.5) in Section 2.2.

$d > 0$, a property that is highly relevant for forecasting, pricing applications, or improved understanding of a network structure in an economic system (Schennach, 2018). 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 price volatility and trading volume are driven by underlying information flows. Applying our robust inference method on a Twitter-based economic uncertainty measure (Baker et al., 2021), we find that inferences concerning this news arrival process do not appear to be affected by weak identification, and the resulting confidence sets of the model parameters support the long-memory specification. An implication of our empirical findings is that the rough-volatility narrative advocated by Gatheral et al. (2018), among others, may need further evaluation to account for potential weak identification issues and possible analysis with an extended model that allows for joint dynamics with trading volume and news arrivals.

The paper is organized as follows. Section 2 reviews model specifications, details nearness measures, and shows how weak identification manifests by establishing conditions under which the spectral densities of the processes are asymptotically equivalent. Section 3 explains how to invert tests for zero serial correlation to construct Anderson–Rubin confidence sets for both parameters. Section 4 presents Monte Carlo results that explore the relevance of weak identification by examining these identification-robust confidence sets. Section 5 reports extensive empirical studies using RV and trading volume series for many assets and Twitter-based economic uncertainty indices. Section 6 concludes. The appendix collects proofs. The online supplement contains details about the data, various empirical robustness checks, and additional empirical applications to economic and climate data.

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 measure, trading volume, or news-based uncertainty. 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 power 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 and 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).

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). Evidently from (2.3), 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).[7] 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, among that of many other economic time series (Diebold and Rudebusch, 1989), has been deemed a stylized fact. If this is debatable for volatility modeling, similar conflicting outcomes can arise with other economic time series or data from a wider field of disciplines such as hydrology, meteorology, and geophysics where long memory has been documented (Graves et al., 2017).

[7] 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. (2021).

Some examples from these fields are given in the Online Supplement.

Besides its long-run implications, the distinction between long memory and rough-volatility 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. As a case in point, consider the nonparametric estimation of spot volatility (say, over an event window before or after a macro news announcement). When the volatility is 'rough', a small bandwidth in nonparametric estimation is needed to tame the large nonparametric estimation bias, which leads to a slower optimal rate of convergence. In the boundary case with d approaching -0.5 and fBM becoming 'barely' continuous, the rate of convergence can be 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 empirical and theoretical considerations, we aim to shed new light on long memory versus rough-volatility empirical issues that should be helpful in analyzing other economic time series where long memory seems evident. 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 9 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. Recognizing this connection, it is natural to adapt Le Cam's notion of distance between statistical experiments (see, e.g., Section 2.2 in Le Cam and Yang (2000)) to the present setting as follows. We define the *deficiency* of $\widetilde{\mathcal{M}}_T$ with respect to $\mathcal{M}_T$ as

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

where the lower bound $\underline{\lambda}_T > 0$ may possibly shrink to zero.[8] The idea under this definition

[8]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 $T \to \infty$.

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 can then gauge the (pseudo) distance between the two local models using

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

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 $\widetilde{\mathcal{M}}_T$.

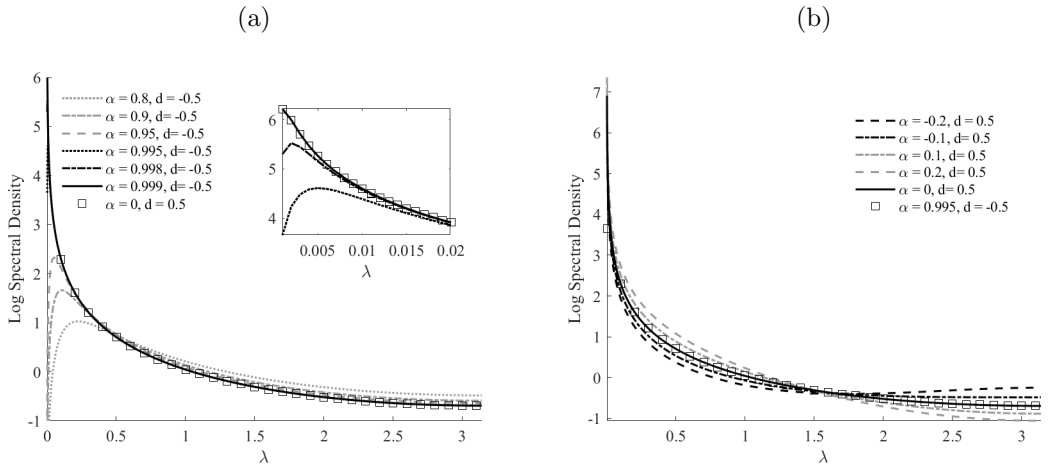
Theorem 1 *Let R_T and $\widetilde{R}_T$ be defined as (2.6) for some positive sequences $\gamma_T = o(1)$, $\tilde{\gamma}_T = o(1)$, and $\eta_T = O(1)$. Then, $\Delta(\mathcal{M}_T, \widetilde{\mathcal{M}}_T) = O((\underline{\lambda}_T^{-2}\gamma_T) \vee \tilde{\gamma}_T)$.*

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).[9] 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, i.e. 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 1 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

[9] 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 (2022).

Figure 1: Log Spectral Densities of Two Local Models: An Illustration



spectral densities for the remaining parameter settings are in $\widetilde{\mathcal{M}}_T$ with α between -0.2 and 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. Evidently, 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 $d = 0.5$ evidently 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., 2021; Wang et al., 2021).[10] This corresponds to the

[10]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.

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 (2022) 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 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. Conven-

tional 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 $\theta_0 = \theta$. 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. 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' comprise 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 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^2_p. 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.[11]

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,

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

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

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

[11] The AP test is also easy to implement as it does not require re-sampling the data as with bootstrap methods. Computational efficiency is highly relevant for our application since we need to invert the test across a large number of candidate parameter values.

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. 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^2_1 distribution. □

3 Simulations

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 4. 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.[12] Specifically, we consider $d = -0.4$ or 0.4 under which the process u_t 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.[13]

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 $\alpha \in [0.99, 1]$.

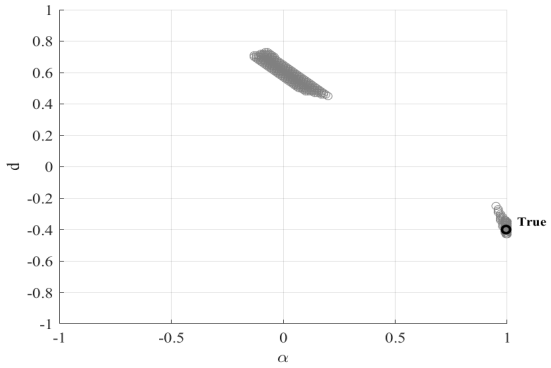
It is instructive to illustrate the workings of the proposed confidence set for a single random draw in the Monte Carlo experiment. Figure 2 plots (one-shot) realizations of the estimated confidence sets under four Monte Carlo configurations. Specifically, we consider two sample sizes, $T = 2,000$ or $T = 5,000$, 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

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

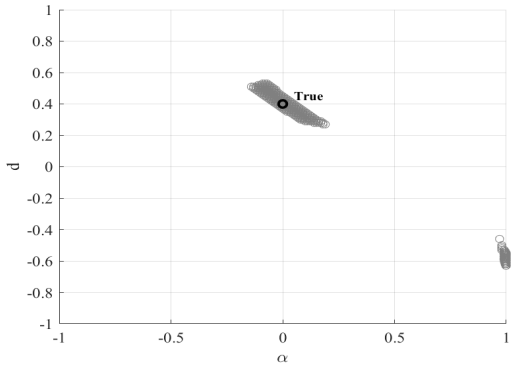
[13] This value matches the estimate reported in Shi and Yu (2022) for the S&P 500 ETF (SPY) from January 2010 to May 2021 based on the Whittle method.

Figure 2: Identification-Robust Confidence Sets: One-Shot Realizations

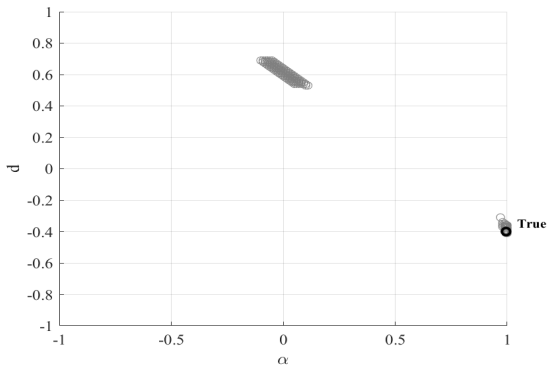
(a) $\alpha = 0.995$, $d = -0.4$, and $T = 2000$



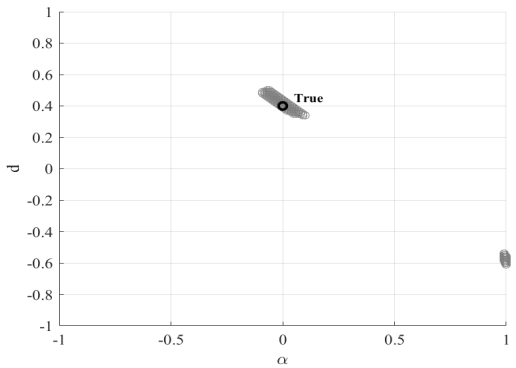
(b) $\alpha = 0$, $d = 0.4$, and $T = 2000$



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



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



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 2 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. 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.

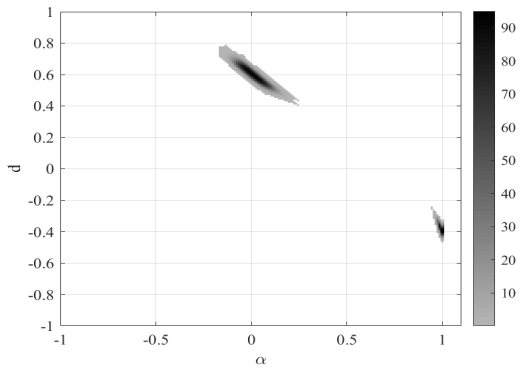
The pattern depicted in these illustrations is representative. This may be shown by overlaying the plots in Figure 2 across all Monte Carlo trials. More precisely, for each candidate parameter value on the (α, d) plane, we compute the frequency that it falls in the confidence set; we then plot these coverage rates as a heatmap. 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 3 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 (2022) 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 3. 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 non-standard 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).

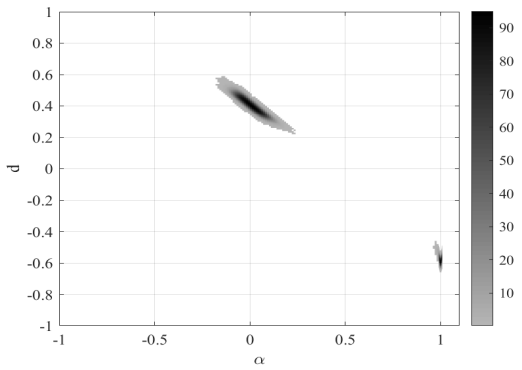
To reveal the incidence of bifurcation Figure 4 plots its frequency of occurrence as a function of the true value of the autoregressive coefficient $\alpha \in \mathcal{A}$ while fixing $d = -0.4$ (left) or $d = 0.4$ (right). For brevity the $T = 5,000$ case is reported. In the left panel where $d = -0.4$ the confidence set almost always contains two disjoint regions when $\alpha = 0.995$, just as in panel (a) of Figure 3. When $\alpha = 0.9$, the bifurcation frequency drops to approximately

Figure 3: Identification-Robust Confidence Sets: Coverage Rates

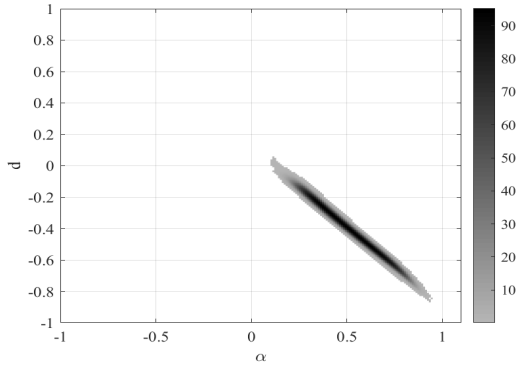
(a) $\alpha = 0.995, d = -0.4$ (Weak ID)



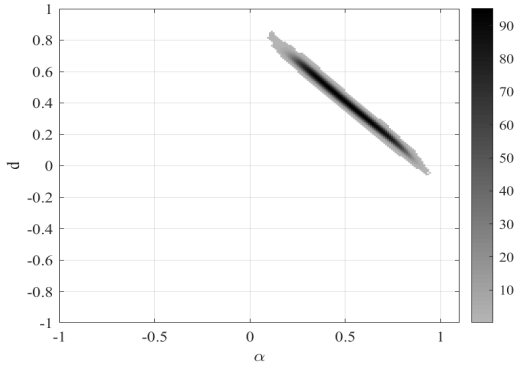
(b) $\alpha = 0, d = 0.4$ (Weak ID)



(c) $\alpha = 0.5, d = -0.4$ (Strong ID)



(d) $\alpha = 0.5, d = 0.4$ (Strong ID)



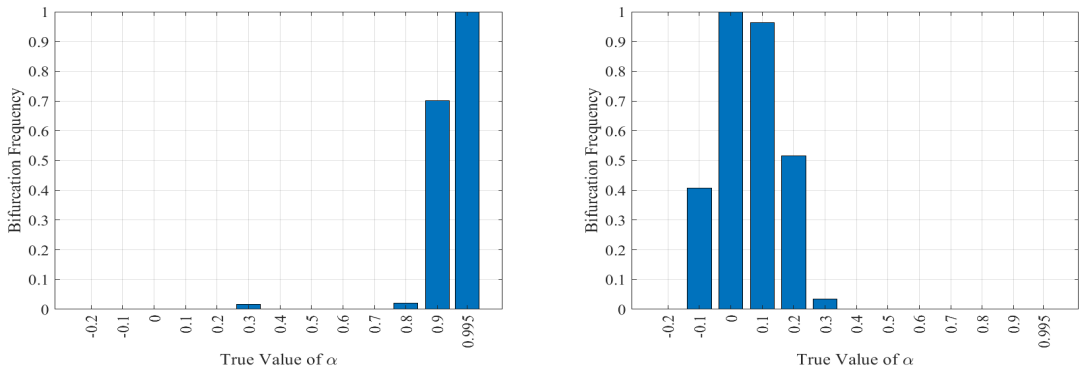
Note: See the online version for this figure in color.

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 4 shows that when $d > 0$, weak identification is more severe when α is close to zero, complementing intuition and Theorem 1.

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 empirical findings, to which we now turn.

Figure 4: Bifurcation Frequencies of the Identification-Robust Confidence Set

(a) $d = -0.4$ (b) $d = 0.4$



4 Empirical Applications

The proposed identification-robust inference approach is applied to several economic time series. Section 4.1 presents results based on realized volatility (RV) measures for a broad range of U.S. ETFs and stocks, and international stock market indices. Sections 4.2 and 4.3 report additional empirical findings for trading volume and social-media news flow data. Further empirical results for a wider range of time series are given in the Online Supplement.

4.1 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.[14] 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[15] – 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 online supplement. 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 online supplement for a summary.

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

[15] 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.

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.[16] This extension is not pursued here to retain the weak identification focus of the paper.

As in the Monte Carlo simulations, 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 2) 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.

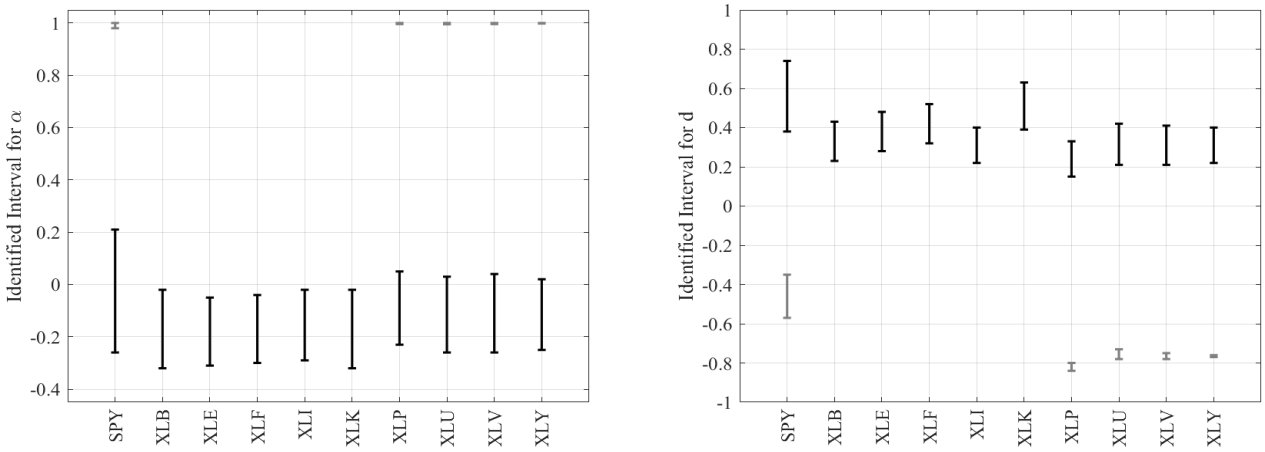
Figure 5 plots the one-dimensional confidence sets for the SPY and the nine industry ETFs, with panel (a) and panel (b) showing the results for α and d, respectively. Since the confidence set for (α, d) often contains two disjoint regions, we use two gray scales (dark and light) to signify them, so that 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. $d < 0$).

From Figure 5 five of the ten ETFs (including SPY, XLP, XLU, XLV, and XLY) have

[16] 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. (2022) 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.

Figure 5: Confidence Sets for Selected ETFs (1996–2021)

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



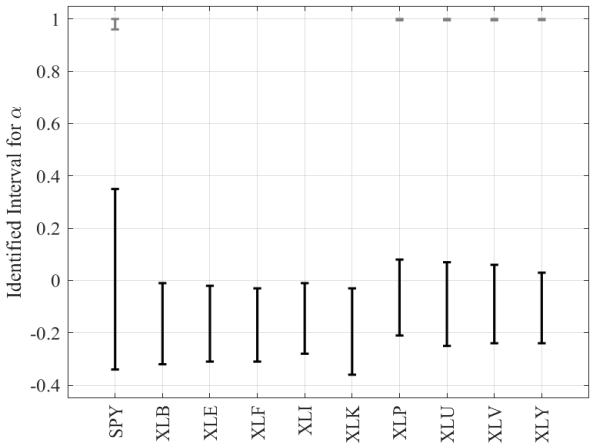
confidence sets with disjoint regions. In dark-colored regions the autoregressive coefficient α is near zero and the positive fractional parameter indicates long memory, whereas in light-colored regions α is near unity and d takes large negative values. These patterns are consistent with theory, earlier 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.

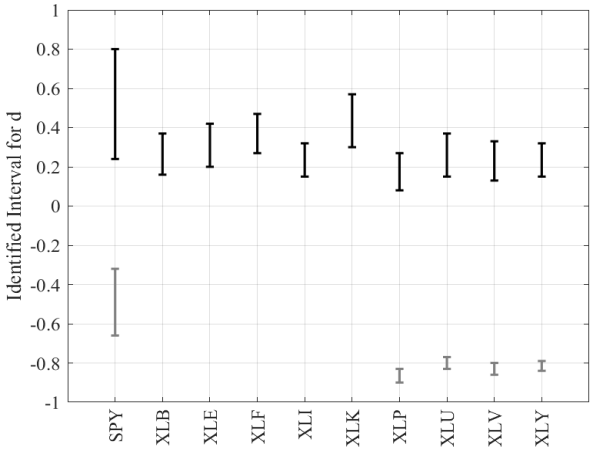
Figure 5 also reveals that the confidence sets for some assets (including XLB, XLE, XLF, XLI, and XLK) consist of only a single region, becoming confidence intervals. These confidence intervals for the fractional parameter d all hover around 0.4, which is close to

Figure 6: Confidence Sets for Selected ETFs: Subsample Analysis

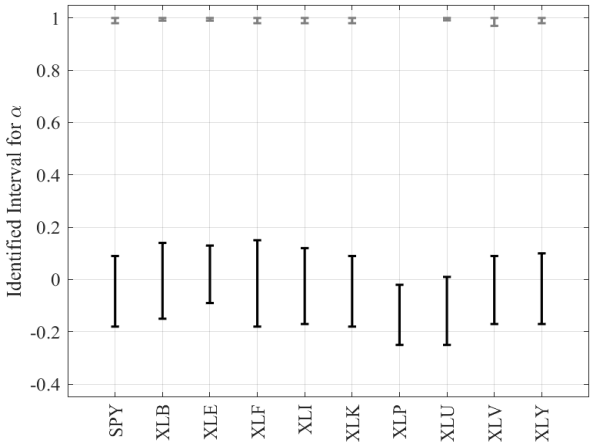
(a) 95% Confidence Set of α (1996–2009)



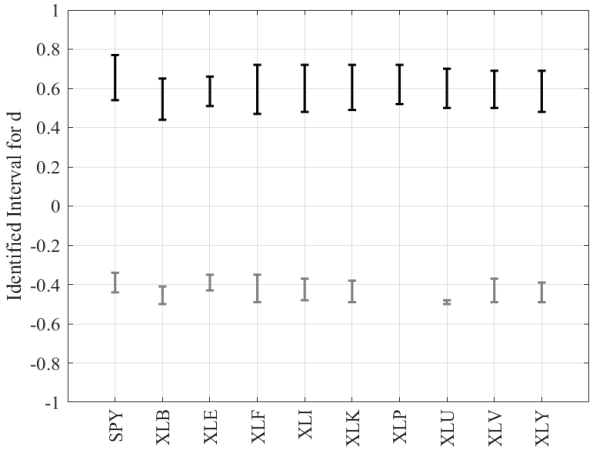
(b) 95% Confidence Set of d (1996–2009)



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



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



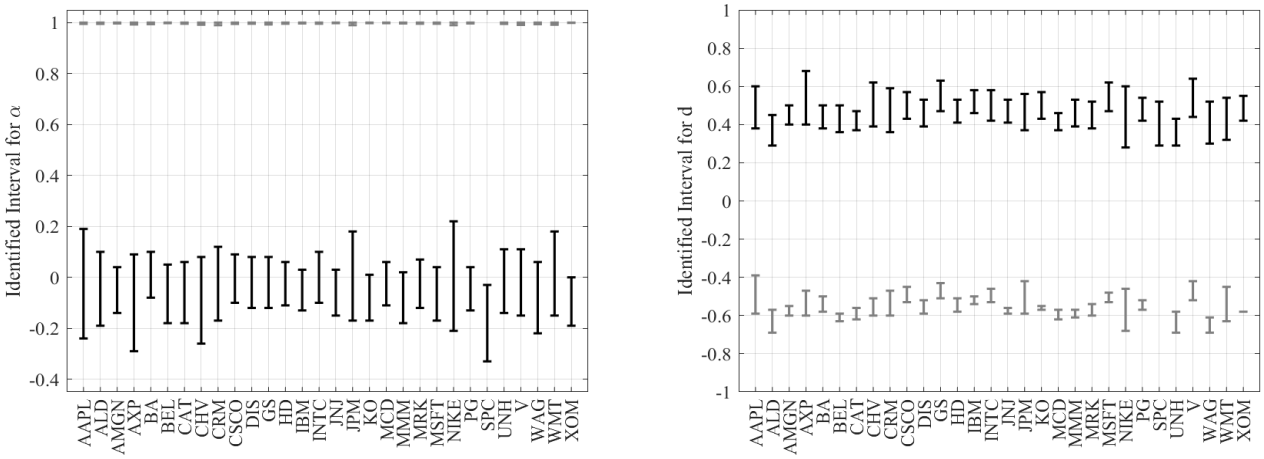
the estimate reported in Andersen et al. (2003) and other work, thereby favoring the long memory narrative advocated in the early RV literature.

How robust are these findings to different 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 6 reports the estimated confidence sets. The results for 1996–2009 (in the top row) are qualitatively similar to those of the full-sample shown in Figure 5. But the bottom row of Figure 6 shows that 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 7 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 7: Confidence Sets for Dow Jones Industrial Average Stocks

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



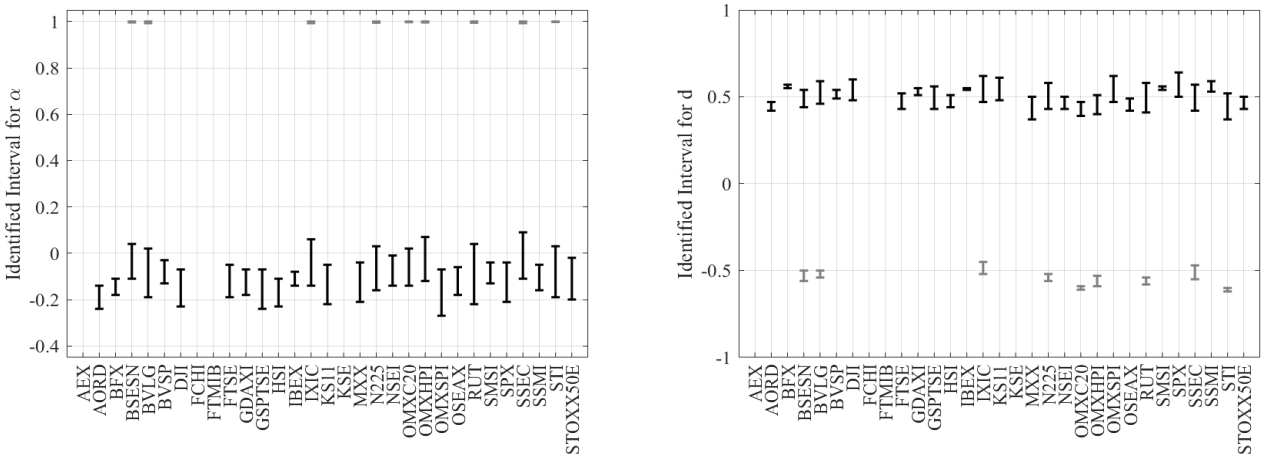
Additional empirical evidence is obtained with data from a broad range of international markets. The analysis relies on the daily RV series for all 31 stock market indices that are available online from the Oxford–Man Realized Library.[17] The same procedure is conducted for these market-level volatility measures and Figure 8 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.[18] These findings show that weak identification issues occur over a broad set of markets but that the overall evidence tends in

[17] Robustness checks based on alternative RV measures are provided in the online supplement.

[18] 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 8: Confidence Sets for International Stock Market Indices

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



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 memory can arise from aggregation (Robinson, 1978; Granger, 1980). The online supplement 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. But this conclusion does not rule out the possibility that a rough-volatility model may outperform long memory in volatility forecasting or option pricing applications, which are certainly of interest but lie beyond the scope of the present paper.

4.2 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) shows that stochastic liquidity can drive this relationship in a Kyle (1985)-type noisy rational expectations model.

We study the same 40 assets in the U.S. equity market as those in Section 4.1. 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 online supplement 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 in Figure 9 and the 30 Dow Jones stocks in Figure 10. 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 $\alpha \approx 0$).

Figure 9: Confidence Sets for Trading Volume of Selected ETFs

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

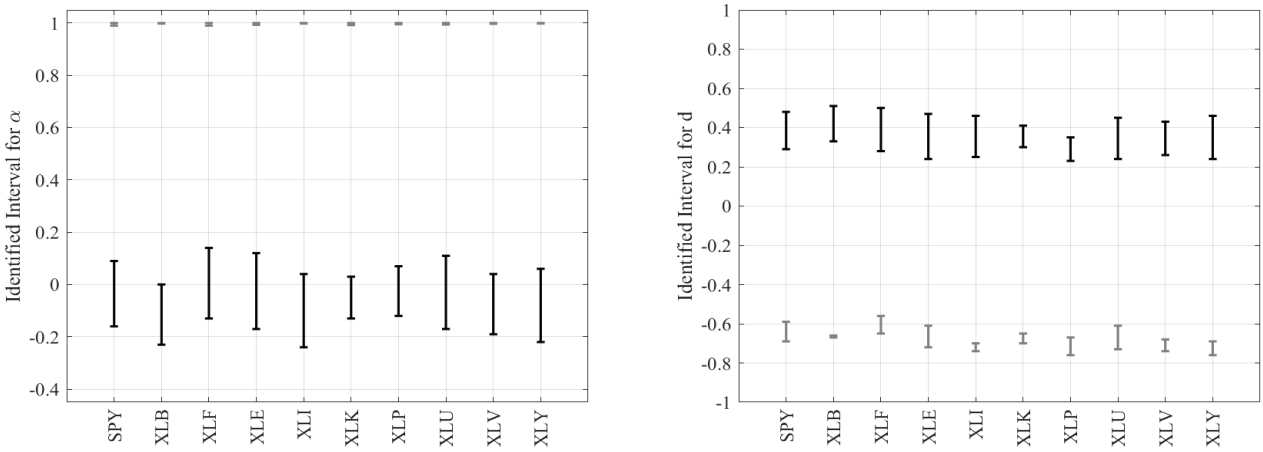
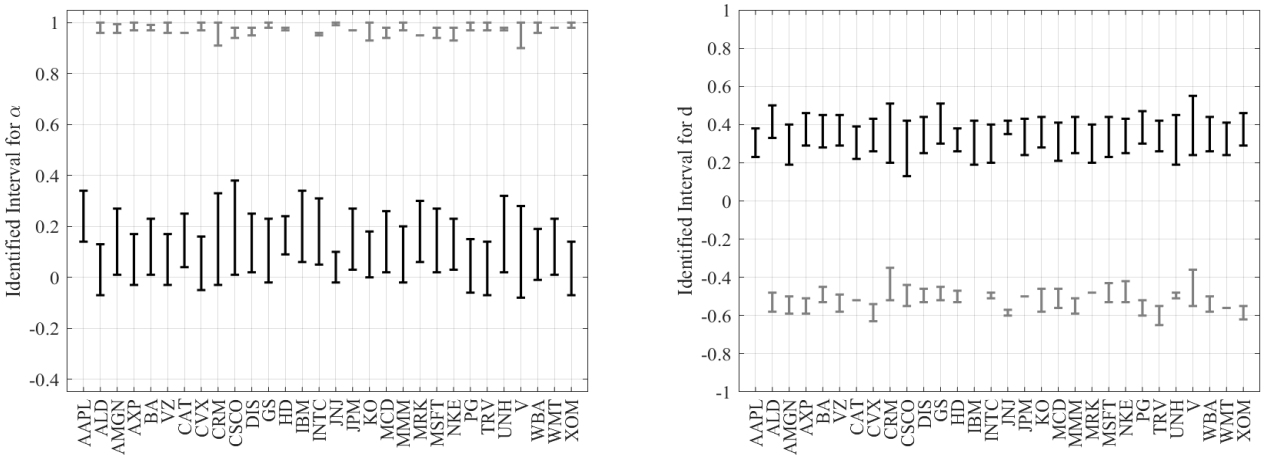


Figure 10: Confidence Sets for Trading Volume of Dow Jones Industrial Average Stocks

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



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 9 and 10 are qualitatively similar to those in Figures 5 and 7, 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[19].

4.3 Twitter-based economic uncertainty

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)[20]. 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 possibilites involve using prior information from relevant theory or additional data that provide new information to assist in resolving the indeterminacy empirically.

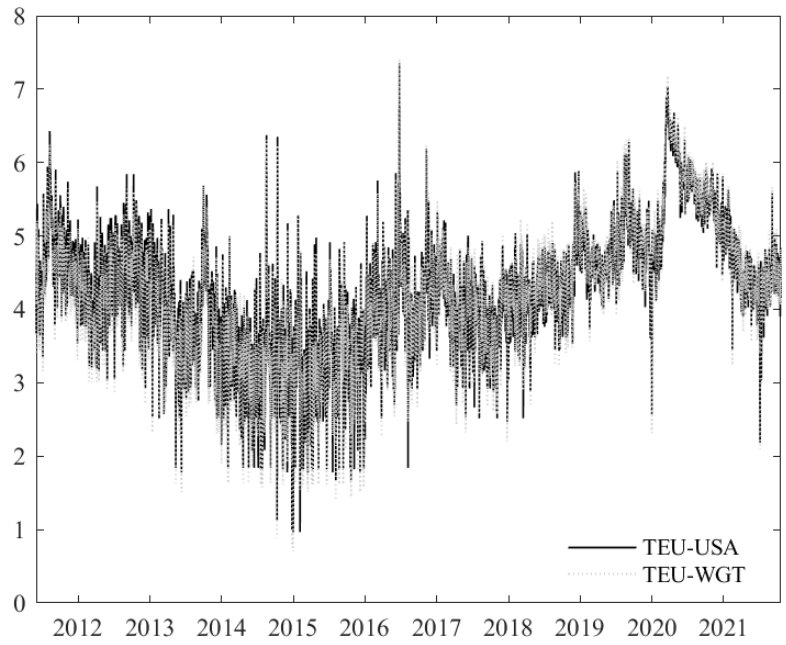
The mixture-of-distributions hypothesis suggests a strong volume-volatility relationship and further postulates that both volume and volatility are driven by underlying information flows. The precise manner in which these economic quantities are linked depends partly on trading behavior and is generally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

This consideration provides motivation to examine the dynamics of news arrivals. In view of the enormous attention now paid to social media, we focus on the Twitter-based Economic

[19]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.

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

Figure 11: Twitter Economic Uncertainty Indices



Uncertainty (TEU) index, which is publicly available from the Economic Policy Uncertainty website.[21] 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.[22] 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 11 plots these two TEU indices in logarithmic units. Unsurprisingly, they are highly correlated (correlation coefficient $= 0.996$) and highly persistent.

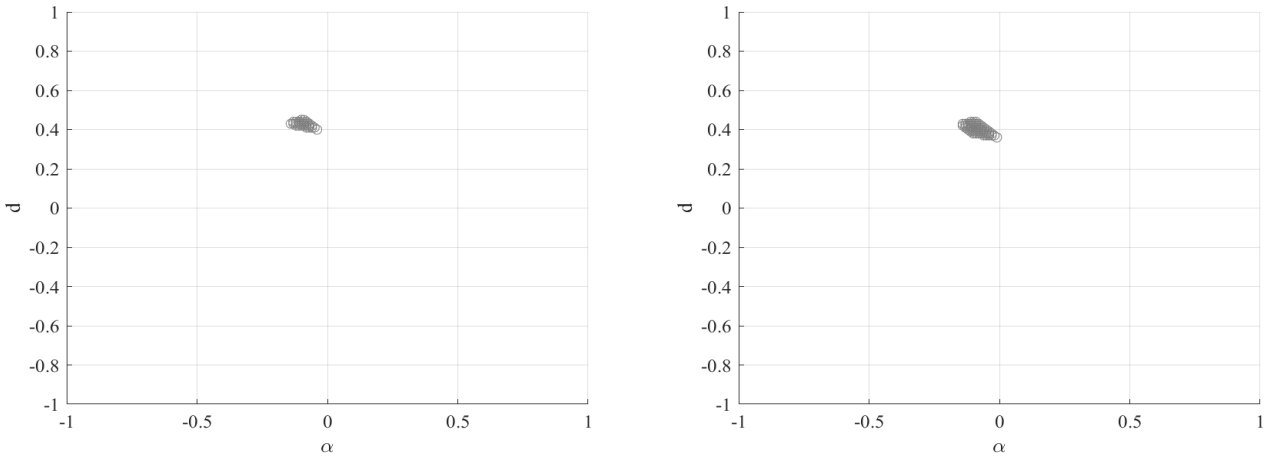
Following the same procedure as before, we compute identification-robust confidence sets of (α, d) for the two (log-transformed and then demeaned) TEU indices. The confidence sets are plotted in Figure 12. Interestingly, the confidence sets for these TEU indices contain only one region, with memory parameter around 0.4 and autoregressive coefficient taking

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

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

Figure 12: Confidence Sets for Twitter-Based Economic Uncertainty Indices

(a) Estimated interval of α (b) Estimated interval of d



negative values but close to zero.[23] Weak identification does not appear to be a 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 TEU indices center around $d = 0.4$, 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.

This empirical exercise is based on specific measures of news flows, and the extent to which the empirical outcomes may apply over a broader set of measures remains to be investigated in future research. A further 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

[23]Specifically, the confidence intervals for d are $[0.4, 0.45]$ for TEU-USA and $[0.36, 0.44]$ for TEU-WGT, with the corresponding intervals for α being $[-0.14, -0.04]$ and $[-0.14, -0.01]$.

it highlights the potential usefulness of studying the joint dynamics of volatility, volume, and news arrivals, in the hope of revealing a suitable model for the 'forest' and not just the 'trees'. A comprehensive analysis, possibly based on more complete measures of economic news flows, is warranted for future research.

5 Conclusion

Economic time series often manifest long memory characteristics. Asset price volatility has received arguably the most attention, partly due to advances in high-frequency-based realized volatility estimation where fractional processes have proved particularly useful. 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 (gauged by a deficiency measure adapted from Le Cam's theory on limits of experiments) converges to zero when the autoregressive parameter is localized to unity or zero.

To address potential weak identification in practical work 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. The finding is further corroborated by an analysis of news

arrivals, measured by Twitter-based economic uncertainty indices, for which identification-robust inference lends support to long memory in the ARFIMA$(1, d, 0)$ model.

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 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.

References

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

Adelman, I. (1965). Long cycles: Fact or artifact? *American Economic Review 55*(3), 444–463.

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 (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 74*(1), 3–30.

Baillie, R. T., F. Calonaci, D. Cho, and S. Rho (2019). Long memory, realized volatility and heterogeneous autoregressive models. *Journal of Time Series Analysis 40*(4), 609–628.

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

Bandi, F. M. and B. Perron (2006). Long memory and the relation between implied and realized volatility. *Journal of Financial Econometrics 4*(4), 636–670.

Bauer, M. D. and G. D. Rudebusch (2020). Interest rates under falling stars. *American Economic Review 110*(5), 1316–54.

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

Beaudry, P. and F. Portier (2006). Stock prices, news, and economic fluctuations. *American Economic Review 96*(4), 1293–1307.

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 (2022). A GMM approach to estimate the roughness of stochastic volatility. *Journal of Econometrics, forthcoming*.

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

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.

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.

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.

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.

Diebold, F. X. and G. D. Rudebusch (1989). Long memory and persistence in aggregate output. *Journal of Monetary Economics 24*(2), 189–209.

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 (2018). The microstructural foundations of leverage effect and rough volatility. *Finance Stochastics 22*, 241–280.

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

Friedman, B. M. and K. N. Kuttner (1992). Money, income, prices, and interest rates. *American Economic Review 82*(3), 472–492.

Fukasawa, M., T. Takabatake, and R. Westphal (2021). Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics. *Mathematical Finance, forthcoming*.

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

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. (1966). The typical spectral shape of an economic variable. *Econometrica 34*(1), 150–161.

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.

Graves, T., R. Gramacy, N. Watkins, and C. Franzke (2017). A brief history of long memory: Hurst, Mandelbrot and the road to ARFIMA, 1951–1980. *Entropy 19*(9), 437.

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).

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), 165–176.

Ireland, P. N. (2009). On the welfare cost of inflation and the recent behavior of money demand. *American Economic Review 99*(3), 1040–1052.

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.

Le Cam, L. and G. L. Yang (2000). *Asymptotics in Statistics: Some Basic Concepts*. Springer Science & Business Media.

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 nonneutrality: 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. (2022). Estimation and inference with near unit roots. *Econometric Theory, forthcoming*.

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.

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 (2022). Volatility puzzle: Long memory or anti-persistency. *Management Science, forthcoming*.

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 (2021). Modeling and forecasting realized volatility with the fractional Ornstein-Uhlenbeck process. *Journal of Econometrics, forthcoming*.

A Appendix: 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,

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

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

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

Hence,

$$\begin{aligned}
&\sup_{\lambda} \left| \log f_{(\alpha_T, d_T)}(\lambda) - \log f_{(\tilde{\alpha}_T, d_T+1)}(\lambda) \right| \\
&\quad = \sup_{\lambda} \left| \log\left(\frac{2 - 2\cos(\lambda)}{1 - 2\alpha_T \cos(\lambda) + \alpha_T^2}\right) + \log\left(1 - 2\tilde{\alpha}_T \cos(\lambda) + \tilde{\alpha}_T^2\right) \right|
\end{aligned}$$

$$\leq \sup_{\lambda} \left| \log \left(\frac{2 - 2\cos(\lambda)}{1 - 2\alpha_T \cos(\lambda) + \alpha_T^2} \right) \right| + \sup_{\lambda} \left| \log \left(1 - 2\tilde{\alpha}_T \cos(\lambda) + \tilde{\alpha}_T^2 \right) \right|. \quad \text{(A.2)}$$

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

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

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

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

Combining (A.1)–(A.4) yields

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

Since the upper bound in the above display holds for a universal constant K not depending on θ, it also holds for the supremum over $\theta \in R_T$, which further 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, we can also show that $\delta(\mathcal{M}_T, \widetilde{\mathcal{M}}_T) = O\left((\underline{\lambda}_T^{-2} \gamma_T) \vee \tilde{\gamma}_T \right)$. The assertion of the theorem then readily follows. *Q.E.D.*