The Review of Financial Studies

Weak Identification of Long Memory with Implications for Volatility Modeling

Jia Li
Singapore Management University, Singapore

Peter C. B. Phillips
Yale University, United States
University of Auckland, New Zealand
Singapore Management University, Singapore

Shuping Shi
Macquarie University, Australia

Jun Yu
University of Macau, China

This paper explores implications of weak identification in common 'long memory' and recent 'rough' approaches to modeling volatility dynamics of financial assets. We unveil an asymptotic near-observational equivalence between a long memory model with weak autoregressive dynamics and a rough model with a near-unit autoregressive root. Standard methods struggle to distinguish them, and conventional asymptotics are invalid. We propose an identification-robust approach to construct confidence sets that reveal the uncertainty and aid inference. Empirical studies based on realized volatility and trading volume often fail to statistically reject either model, thereby providing evidence of their potential coexistence. (*JEL* C12, C13, C58)

Received: March 30, 2023; Editorial decision: October 18, 2024
Editor: Juhani Linnainmaa
Authors have furnished an Internet Appendix, which is available on the Oxford University Press Web site next to the link to the final published paper online.

Thanks go to the Editor and two referees for searching comments and many useful suggestions that assisted the revision of this paper. The authors also thank Yacine Aït-Sahalia, Torben Andersen, Isaiah Andrews, Federico M. Bandi, Tim Bollerslev, Xu Cheng, Frank Diebold, Jim Gatheral, Peter Hansen, Morten Nielsen, Mathieu Rosenbaum, and Zhongjun Qu for helpful discussions and comments. Li acknowledges support from Singapore Ministry of Education [grant no. 22-SOE-SMU-016]. Shi acknowledges support from the Australian Research Council [grant no. DE190100840]. Phillips acknowledges research support from a Kelly Fellowship at the University of Auckland. Yu acknowledges support from the Lee Foundation and University of Macau Development Foundation. Supplementary data can be found on *The Review of Financial Studies* web site. Send correspondence to Jun Yu, junyu@um.edu.mo.

The Review of Financial Studies 00 (2025) 1–32

https://doi.org/10.1093/rfs/hhaf022 Advance Access publication April 3, 2025

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

Since the seminal studies of Engle (1982) and Bollerslev (1986), the volatility literature seems to have reached a consensus concerning the presence of strong persistence in the volatility of financial assets. Nonetheless, the question of how best to model the persistence in volatility continues to be debated. Two competing paradigms – long memory and rough volatility – have emerged as leading contenders, each offering a distinct approach to modeling volatility persistence. While much of the research points to the presence of long-range dependence or long memory in volatility data as a means of capturing volatility persistence,[1] more recent work and industry practice increasingly point to rough volatility as a crucial component, as the quotation that heads this article suggests. Despite this focus, neither researchers nor practitioners agree on which model is superior, and this unresolved tension continues to shape both research and industry practice.

Adding to this complexity, Shi and Yu (2023) reveal the poor finite sample performance of existing estimation methods for volatility models. Inspired by findings in Shi and Yu (2023) and the existing literature on weak identification,[2] this paper demonstrates that weak identification is the root cause of this poor performance. This issue not only undermines estimation accuracy and invalidates conventional statistical inference but also contributes to the ongoing divergence between the long memory and rough volatility paradigms. Our analysis therefore helps to explain the uncertainty that researchers have been finding in practical work concerning evidence for rough and long memory specifications

Considerable research has been devoted to explain the long memory phenomenon in terms of more primitive generating mechanisms that have empirical justification. It is now known that mechanisms such as cross-section aggregation, structural breaks, trends, regime switching, learning, nonlinearity, marginalization, and networking can all generate long memory.[3] By combining short-run autoregressive (order p) and moving average (order q) components parametrically with fractional integration (I(d)) to capture long-range dependence, the class of ARFIMA(p,d,q) models has been widely used in empirical work to model economic time series that manifest both short and long memory

[1] See Andersen et al. (2001b), Andersen et al. (2001a, 2003), Harvey (2007), and references therein.

[2] See Phillips (1989); Staiger and Stock (1997); Stock and Wright (2000); Stock and Yogo (2005); Andrews and Cheng (2012); Andrews and Mikusheva (2022); Cheng, Dou, and Liao (2022), among others.

[3] See, for example, Chevillon, Massmann, and Mavroeidis (2010), Schennach (2018), and references therein.

as well as possible nonstationarity. The impulse response function implied by this general model with fractional integration (FI) is substantially different from that of an ARMA(p,q) model, particularly allowing for long slow decays in responses when the fractional parameter d is positive. This feature of impulse responses in volatility has important implications for volatility forecasting and hence financial decisions, such as portfolio choice, and option pricing. For example, in option pricing, small errors in volatility estimates can lead to significant mispricing, affecting trading profitability. In portfolio management, precise volatility forecasts help optimize asset allocation and manage risk. Within the class of ARFIMA(p,d,q) models, the ARFIMA$(1,d,0)$ model has been found to be especially useful and a leading example that motivates the present paper is the stochastic volatility of financial assets. We refer to the autoregressive (AR) coefficient of the ARFIMA$(1,d,0)$ model as α in the subsequent discussions.

Many statistical procedures are now available for the estimation of the memory parameter d in both the ARFIMA parametric class and various semi-parametric classes that do not prescribe AR or moving average specifications.[4] Estimated values of d from log daily realized volatility (RV) data usually turn out to be positive and close to the nonstationary boundary 0.5, which implies long memory dynamics, and estimates of α are found to be near zero.[5]

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

Interestingly and as attested in the lead quotation, a new body of evidence points towards the same ARFIMA$(1,d,0)$ model structure but with a negative value for d (so-called 'antipersistence'), producing what is called 'rough-volatility' with an AR parameter taken to be unity or near-unity.[6] Although still in the early stages of development, some attempts have been made to

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

[5] See Andersen and Bollerslev (1997), Andersen et al. (2001a), Andersen et al.(2001b, 2003), Shi and Yu (2023), among others.

[6] Among many studies that support such a model we mention the following: Bayer, Friz, and Gatheral (2016); Gatheral, Jaisson, and Rosenbaum (2018); Fukasawa, Takabatake, and Westphal (2022); Bolko et al. (2023); Wang, Xiao, and Yu (2023b); Shi, Liu, and Yu (2025).

comprehend the rough volatility model's underlying mechanisms (see, e.g., El Euch, Fukasawa, and Rosenbaum 2018; Jusselin and Rosenbaum 2020).

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 (Risk Staff 2021). Notably, the 2021 Risk Awards were presented for introducing rough-volatility models.[7] Based on evidence of roughness in volatility, many studies have conducted financial applications using rough volatility models.[8] An article by Jean-Philippe Bouchaud, Chairman of Capital Fund Management, details the profound implications of rough volatility models for asset pricing. According to Bouchaud (2020), one of recent rough specifications, the quadratic rough Heston model of Gatheral, Jusselin, and Rosenbaum (2020), effectively resolves a long-standing puzzle by jointly calibrating the volatility smile of the S&P 500 and VIX options – a challenging task for quantitative analysts for many years. Many hedge funds and nonbank market-makers claim to have adopted rough volatility models (Risk Staff 2021).

Under the rough ARFIMA$(1,d,0)$ model, volatility determination operates through a mechanism distinct from that of the long memory model: news shocks from the past can be effectively summarized by the immediately preceding volatility, but volatility is extended outward with a slight depreciation since the AR coefficient is near but below unity. Importantly, in addition, the day-to-day shocks, that is, volatility innovations, are strongly negatively autocorrelated, so that an upward movement of volatility tends to be followed by a downward drop. The resultant 'zigzag' pattern contributes to the roughness characteristic and spikes in the volatility path tend to reflect the granularity of the information arrival process.

It is surprising that these two strands of seemingly contradictory literature coexist. In a recent study, Shi and Yu (2023) examine finite sample performance of alternative estimation methods for the ARFIMA$(1,d,0)$ model. They find that when data are generated from a rough ARFIMA$(1,d,0)$ (i.e., $d<0$) with a close-to-unity autoregressive coefficient α, semiparametric methods such as the local Whittle method produce substantially biased parameter estimates and result in spurious findings of long memory. Moreover, when the DGP is ARFIMA$(1,d,0)$ with α close to unity and negative d or with α close to zero and positive d, the likelihood function exhibits bimodality.

The present paper shows that the poor finite sample performance of present estimation methods is symptomatic of weak identification, which renders standard statistical inference invalid. In fact, the implied spectral

[7] See the Risk website for the citation: https://www.risk.net/awards/7736196/quants-of-the-year-jim-gatheral-and-mathieu-rosenbaum.

[8] See, for example, Bayer, Friz, and Gatheral (2016); Garnier and Sølna (2017, 2018); El Euch, Fukasawa, and Rosenbaum (2018) in portfolio choice; and El Euch, Fukasawa, and Rosenbaum (2018) in dynamic hedging.

densities of the models are nearly indistinguishable at two well-isolated local parameterizations, corresponding to cases in which α lies either near-unity or near-zero. More specifically, we show that the 'distance' between a near-unit-root ARFIMA$(1,d,0)$ model with a negative $d \in (-0.5,0)$ (i.e., a rough model) and a near-zero-root ARFIMA$(1,d+1,0)$ model (i.e., a long memory model) goes to zero when nearness parameters shrink.[9] This explains why standard statistical methods of inference lead to major finite sample distortions under identification failure (Phillips 1989; Dufour 1997). Against the background of these findings some related phenomena involving inferential distortions are to be expected in the present context, helping to elucidate the findings of Shi and Yu (2023) in finite samples. But unlike Shi and Yu (2023) our empirical analysis relies on a new identification robust procedure for statistical inference and our empirical exercises encompass a far broader range of financial data.[10]

To reveal the prevalence of the weak identification issue and address this issue we propose an identification-robust confidence set of the model parameters. The confidence set remains valid even when model suffers from weak identification. It is obtained by inverting tests for zero serial correlation in the model-implied residuals, leveraging the well-established literature on serial correlation tests. The inference procedure is semiparametric, data-driven, and does not rely on Gaussianity. 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 (a) α is close to unity and d is negative or (b) α is close to zero and d is positive.

The identification-robust procedure serves as an effective tool that enables us to explore how prevalent this empirical phenomenon is in practical work with financial data. We report results for 111 data series encompassing realized volatility and trading volume time series for a broad variety of U.S. equities and international stock market indices. Our findings indicate that identification-robust confidence sets often do bifurcate, exhibiting precisely the same pattern observed in simulations. Figure 1 provides an illustration. We apply the identification-robust inference procedure for the ARFIMA$(1,d,0)$ model to the demeaned log realized volatility of the S&P 500 Index exchange-traded fund (ETF) from 1996 to 2021. The x-axis represents values of α, while the y-axis shows values of d. The procedure reveals two clearly defined regions for (α,d) at the 95% confidence level: (1) α is close to zero and $d>0$, consistent with long memory dynamics, and (2) α approaches unity and $d<0$, characteristic of rough volatility behavior. This bifurcation highlights the empirical difficulty of determining whether or not a time series is driven by long-memory disturbances with $d>0$, a property that is highly relevant

[9] 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.

[10] Shi and Yu (2023) applied the ARFIMA$(1,d,0)$ model to 10 U.S. index ETFs employing various estimation techniques. With no valid inferential approach their focus was point estimates.

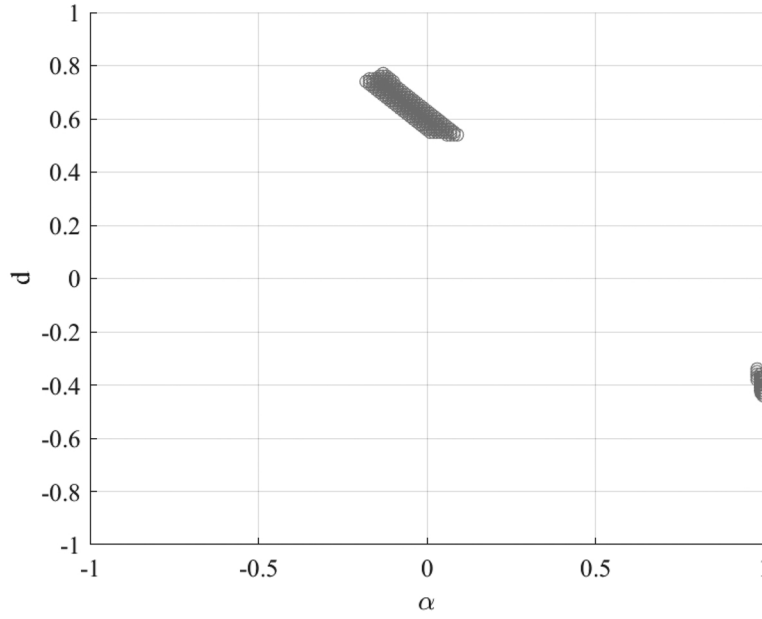

Figure 1
Confidence set for the S&P 500 index ETF: SPY
Identification-robust inference was applied to the (demeaned) log realized volatility of SPY from 1996 to 2021. The x-axis displays α and the y-axis shows d. The procedure identifies two isolated regions for (α, d) at the 95% confidence level: (1) α is close to zero and $d > 0$ (long memory), and (2) α is near-unity and $d < 0$ (rough).

for pricing applications and one that has implications for forecasting. The empirical prevalence of weak identification in the intricate landscape of financial volatility and trading volume helps to explain the present coexistence of two conflicting model approaches and empirical findings in the volatility literature.

The main takeaways of our analysis are as follows: (a) The ARFIMA$(1, d, 0)$ model has a weak identification issue that invalidates standard methods and explains the performance of various estimation methods documented in Shi and Yu (2023). (b) The problem of weak identification is widespread in the landscape of financial return volatility and trading volume. This finding explains the conflicting empirical findings in the volatility literature and reveals an important cause of some heated debates regarding long memory versus roughness in the recent literature.[11]

[11] While weak identification suggests it is difficult to distinguish the two models from a statistical perspective, it does not necessarily imply that the two models produce equal performance in applied scenarios, such as option pricing and forecasting. The weak identification arises from the first-order equivalence of their spectral densities at most frequencies with the exception of the frequencies extremely close to zero. However, the higher-order terms or near-zero frequencies may have varying impacts in different contexts. Therefore, to judge the relative merits of the two models, the application context should be considered and care in extrapolating lessons from other contexts is needed.

1. The Econometric Method

1.1. 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:

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

where α is the AR coefficient, u_t is an FI 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 in which $d \in (-0.5, 0.5)$ the fractional operator in (1) can be defined by binomial series expansion as

$$(1-L)^{-d} = \sum_{j=0}^{\infty} \binom{-d}{j} (-L)^j = \sum_{j=0}^{\infty} \frac{(d)_j}{j!} L^j \tag{2}$$

giving $u_t = \sigma(1-L)^{-d}\varepsilon_t = \sigma \sum_{j=0}^{\infty} \frac{(d)_j}{j!}\varepsilon_{t-j}$. In (2), $(d)_j = d(d+1)\ldots(d+j-1) = \frac{\Gamma(d+j)}{\Gamma(d)}$ is a forward factorial and $\Gamma(\cdot)$ is the gamma function. In nonstationary cases in which $d \geq 0.5$ 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 write the parameter of interest as $\theta = (\alpha, d)$, and the variance σ^2 is treated as a nuisance parameter.

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

$$\rho_u(k) = \frac{(-d)!(k+d-1)!}{(d-1)!(k-d)!} \sim_a \frac{(-d)!}{(d-1)!}\frac{1}{k^{1-2d}} \text{ as } k \to \infty, \tag{3}$$

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

$$f_\theta(\lambda) = \frac{\sigma^2}{2\pi}\frac{[2-2\cos(\lambda)]^{-d}}{1-2\alpha\cos(\lambda)+\alpha^2} \text{ for } -\pi \leq \lambda \leq \pi, \tag{4}$$

which encodes the dynamics of the observed series y_t.[12]

The sign of d determines whether the FI process u_t has long or short memory. McLeod and Hipel (1978) define a stationary process as having a long (resp., short) memory if its acf is not summable (resp., summable). Hence, u_t has long

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

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, Koul, and Surgailis 2012, chap. 3).[13] The Hurst index H controls the smoothness of the sample path of fBM and the process has 'rough' paths when $H \in (0, 1/2)$.

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). The long memory ($d > 0$) and rough ($d < 0$) sample path models can have different implications for volatility forecasting and option pricing. The conflicting empirical findings are surprising because the long memory property of volatility has been deemed a stylized fact. This in turn has stimulated an active area of research in the recent financial econometrics literature.

Besides its long-run implications, the distinction between long memory and rough-volatility models is also extremely important for the large literature on high-frequency-based nonparametric volatility estimation, as most of the existing work in that literature requires (in a stochastic sense) sufficient smoothness in the volatility path that is incompatible with the rough-volatility model. For instance, nonparametric estimation of spot volatility (e.g., over an event window before or after a critical news announcement) requires a small bandwidth to reduce bias when volatility is 'rough', which produces a slow optimal rate of convergence. In the boundary case with d approaching -0.5, the underlying fBM process is barely continuous and the convergence rate becomes arbitrarily close to zero even with optimal tuning.[14] This in turn would severely limit the use of the high-frequency identification strategy (Nakamura and Steinsson 2018a, b) based on heteroscedasticity (Rigobon 2003) or high-frequency regression-discontinuity designs (Bollerslev, Li, and Xue 2018).

Motivated by these considerations, we aim to shed new light on long memory versus rough-volatility empirical issues. While the existing empirical studies have focused on alternative ways of estimating d, we ask a more fundamental question: whether this parameter is strongly or weakly identified along with the companion AR 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.

1.2. The weak identification problem

Why are d and α weakly identified? To guide intuition, note that ARFIMA$(1, d, 0)$ with $\alpha = 1$ is observationally equivalent to

[13] 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, Xiao, and Yu 2023a).

[14] See, for example, Bollerslev, Li, and Liao (2021); Bollerslev, Li, and Li (2024).

ARFIMA$(1,d+1,0)$ with $\alpha=0$. That is,

$$(1-L)y_t=\sigma(1-L)^{-d}\varepsilon_t \iff y_t=\sigma(1-L)^{-(d+1)}\varepsilon_t. \tag{5}$$

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 operator $1-\alpha L$ reduces to $1-L$ and merges with the differencing filter $(1-L)^{-d}$ to yield $(1-L)^{-d-1}$. 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 AR 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 $d\approx0.5$, provided that 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 and manifests as a discrete 'phase transition' between $(1,d)$ and $(0,d+1)$. 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 AR component. This bias problem was shown in early simulations in Agiakloglou, Newbold, and Wohar (1993) and bias correction methods were considered in subsequent research (e.g., Andrews and Guggenberger 2003; and Poskitt, Martin, and Grose 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

$$\begin{cases} R_T = \{(\alpha_T, d_T) : |\alpha_T - 1| < \gamma_T, \left|d_T - d^*\right| < \eta_T\}, \\ \widetilde{R}_T = \{(\tilde{\alpha}_T, \tilde{d}_T) : |\tilde{\alpha}_T| < \tilde{\gamma}_T, |\tilde{d}_T - d^* - 1| < \eta_T\}. \end{cases} \tag{6}$$

Note that R_T and $\widetilde{R}_T$ are, respectively, near the identification-failure points $(1, d^*)$ and $(0, d^* + 1)$ in the AR 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. The formulation subsumes a large class of near-unity[15] and near-zero local parameterizations that have been used in the econometric literature.

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

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

where the lower bound $\underline{\lambda}_T > 0$ may possibly shrink to zero.[16] The idea under this definition is, for any $\theta \in R_T$ under the model $\mathcal{M}_T$, one can find $\tilde{\theta} \in \widetilde{R}_T$ under the other model $\widetilde{\mathcal{M}}_T$, such that the uniform distance between their spectral densities (in log form) is bounded by $\delta(\widetilde{\mathcal{M}}_T, \mathcal{M}_T)$. The deficiency measure thus quantifies the extent to which the dynamics generated by $\mathcal{M}_T$ cannot be captured by $\widetilde{\mathcal{M}}_T$. Symmetrizing the roles of $\mathcal{M}_T$ and $\widetilde{\mathcal{M}}_T$, we can then gauge the distance between the two local models using

$$\Delta(\mathcal{M}_T, \widetilde{\mathcal{M}}_T) = \max\left\{\delta(\widetilde{\mathcal{M}}_T, \mathcal{M}_T), \delta(\mathcal{M}_T, \widetilde{\mathcal{M}}_T)\right\},$$

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

Theorem 1. Let R_T and $\widetilde{R}_T$ be defined as (6) for some positive sequences $\gamma_T = o(1)$, $\tilde{\gamma}_T = o(1)$, and $\eta_T = O(1)$. Then, $\Delta(\mathcal{M}_T, \widetilde{\mathcal{M}}_T) = O((\underline{\lambda}_T^{-2}\gamma_T) \vee \tilde{\gamma}_T)$, where $\vee$ is the supremum operator.

---

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

[16] 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 $T \to \infty$.

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 AR dimension). 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 on the two isolated points $\left(1, d^*\right)$ and $\left(0, d^* + 1\right)$ in the (α, d) plane. As γ_T and $\tilde{\gamma}_T$ approach zero, these two regions become farther apart in the parameter space but, as shown in Theorem 1, the difference between their dynamic implications also vanishes, provided the lower frequency bound $\underline{\lambda}_T$ does not tend to zero too fast, that is, faster than $\sqrt{\gamma_T}$. As such, weak identification arises in a 'bimodal' form, with two distinct sets of parameters being observationally nearly equivalent.

To illustrate this point, Figure 2 plots the log spectral densities of the ARFIMA model under these two local models. In panel A, the configuration with $\alpha = 0$ and $d = 0.5$ belongs to $\widetilde{\mathcal{M}}_T$, while the others fall in $\mathcal{M}_T$ with α ranging from 0.8 to 0.999 and d fixed at -0.5. Similarly, in panel B, the configuration $\alpha = 0.995$ and $d = -0.5$ falls in $\mathcal{M}_T$, while the spectral densities for the remaining parameter settings are in $\widetilde{\mathcal{M}}_T$ with α between -0.2 and 0.2. Evidently, as $\alpha \to 1$ (resp., $\alpha \to 0$), the log spectral density generated from $\mathcal{M}_T$ (resp., $\widetilde{\mathcal{M}}_T$) approaches and eventually becomes virtually indistinguishable from that associated with $\widetilde{\mathcal{M}}_T$ (resp. $\mathcal{M}_T$), revealing the weak identification between them, subject to the lower frequency bound $\underline{\lambda}_T$ not passing to zero so fast that the different order of the fractional poles dominates the discrepancy in the spectral densities. To further appreciate the impact of $\underline{\lambda}_T$ on the discrepancy measure, we show in the inset in panel A an enlarged graphic of the log spectral densities focused at frequencies closer to zero, ranging between 0.001 and 0.02. For a given $\theta \in \tilde{R}_T$ (panel A), the uniform distance between the two spectral densities is affected by the lower bound of λ but diminishes rapidly provided $\underline{\lambda}_T$ does not pass to zero too fast. When θ is in R_T, the densities are very close (panel B). In this case, the impact of a small negative α under long memory with $d = 0.5$ also raises spectral power at high frequency.

This weak-identification perspective on ARFIMA specifications provides a plausible explanation for the conflicting empirical findings in the literature regarding long memory or roughness in volatility dynamics. In the empirical rough volatility literature when α_T is assumed to be unity or local to unity the estimated value of d is negative and often close to –0.5 (Gatheral, Jaisson, and Rosenbaum 2018; Fukasawa, Takabatake, and Westphal 2022; Wang, Xiao, and Yu 2023b; Bolko et al. 2023).[17] This corresponds to the R_T region in our analysis. As discussed above, such a parametric configuration is essentially indistinguishable from its counterpart in $\widetilde{R}_T$ with d around 0.5 and α_T

[17] 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.

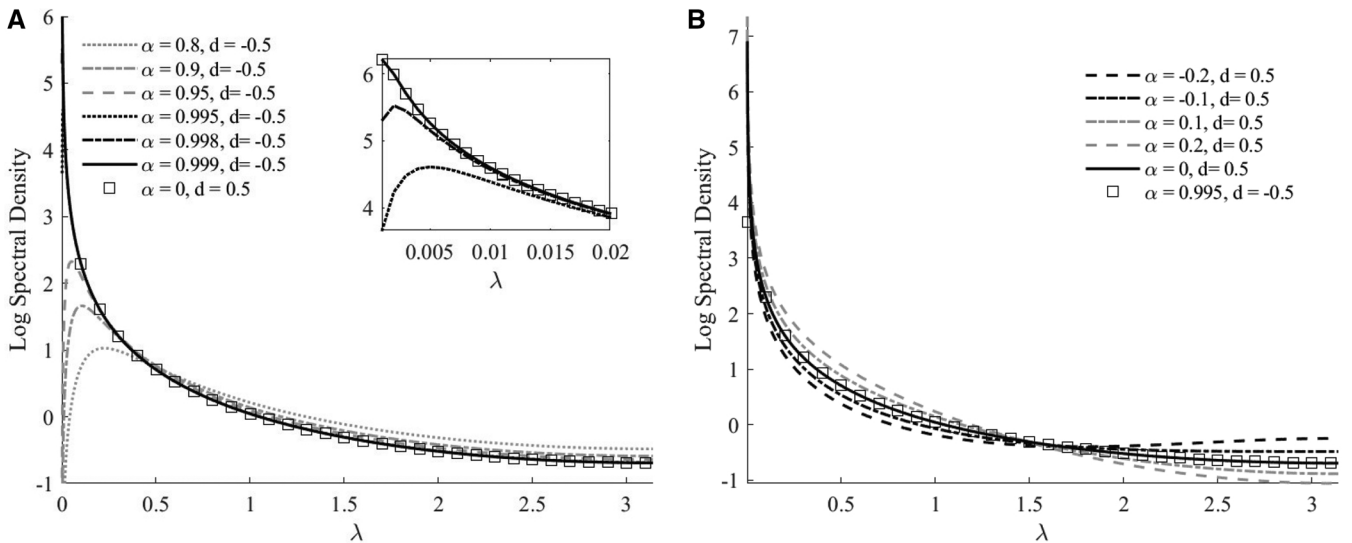

Figure 2
All illustration of log spectral densities of two local models
The spectral density of the ARFIMA$(1,d,0)$ model is given in (4). Panel A shows that when $(\alpha,d)\in\mathcal{R}_T$, as α moves closer to unity, the log spectral density of ARFIMA$(1,d,0)$ approaches that of $(0,1+d)\in\tilde{\mathcal{R}}_T$. Panel B illustrates the convergence in the opposite direction: when $(\alpha,d)\in\tilde{\mathcal{R}}_T$, the log spectral density of ARFIMA$(1,d,0)$ converges to that of $(\alpha=0.995,d-1)\in\mathcal{R}_T$ as $\alpha\to 0$.

near zero. The latter parameter values are actually in line with the estimates reported in the long memory RV literature reviewed in the Introduction.

If weak identification is indeed in force, conventional asymptotic inference based on strong identification may be unreliable. This explains why Shi and Yu (2023) find two disjoint intervals in the highest density set for the time-domain maximum likelihood estimators and frequency-domain maximum likelihood estimators. Similar effects have been extensively studied in the literature on weak instrumental variables (Staiger and Stock 1997; Moreira 2003), the setting of weak generalized method of moments (GMM) (Stock and Wright 2000; Andrews and Mikusheva 2022), and more generally in Andrews and Cheng (2012). 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.

1.3. 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 nonrejected parameter values, viz.,

$$\mathrm{CS}_{1-\beta}=\{\theta\in\Theta:\text{The null hypothesis } H_0:\theta_0=\theta \text{ is not rejected at level } \beta\}. \tag{7}$$

The remaining task is to construct a test that is robust to weak identification. Conventional tests derived from (quasi) maximum likelihood or GMM estimators are not suitable for this task, because the classical inferential theory relies heavily on strong identification. We instead consider a test that targets moment conditions implied by the null hypothesis $\theta_0=\theta$.[18] 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

$$H_0^{WN}:\gamma_j(\theta)=0 \text{ for all } j\geq 1, \tag{8}$$

[18] This idea is similar in spirit to the Anderson-Rubin test proposed by Chevillon, Massmann, and Mavroeidis (2010) to deal with inferences in a structural model with strongly persistent data.

where $\gamma_j(\theta)$ denotes the autocovariance of $\varepsilon_t(\theta)$ of order j.[19] We may test the original null hypothesis $H_0 : \theta_0 = \theta$ by testing the moment conditions in (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 toward 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 ε_t 'only' comprised white noise. We adopt the stationary MDS structure for technical convenience, which is common in the literature, because it simplifies the computation of the test statistic.

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

We adopt the Adaptive Portmanteau (AP) test proposed by Escanciano and Lobato (2009). It is designed to detect violations of the null hypothesis up to the pth lag. 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 a martingale difference sequence 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.

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,

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

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

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

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

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

$$Q_p(\theta) \equiv T \sum_{j=1}^{p} \frac{\hat{\gamma}_j(\theta)^2}{\hat{\tau}_j(\theta)}. \tag{9}$$

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

$$\pi(p,T) \equiv \begin{cases} p \log T & \text{if } \max\limits_{1 \leq j \leq \bar{p}} \sqrt{T} |\hat{\gamma}_j(\theta)| / \sqrt{\hat{\tau}_j(\theta)} \leq \sqrt{2.4 \log T}, \\ 2p & \text{otherwise.} \end{cases} \tag{10}$$

The lag order actually used in the AP test, denoted by $p^*(\theta)$, 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

$$Q^*(\theta) \equiv T \sum_{j=1}^{p^*(\theta)} \frac{\hat{\gamma}_j(\theta)^2}{\hat{\tau}_j(\theta)}. \tag{11}$$

Escanciano and Lobato (2009) show that the asymptotic distribution of this test statistic under the null hypothesis is χ^2_1, since the optimal lag order is one under the null hypothesis. Hence, we reject the null at the significance level β when $Q^*(\theta)$ exceeds the $1-\beta$ quantile of χ^2_1, denoted by $\chi^2_{1,1-\beta}$. Recalling (7), the $1-\beta$ level identification-robust confidence set that we propose can thus be written explicitly as follows

$$\mathrm{CS}_{1-\beta} = \left\{ \theta \in \Theta : Q^*(\theta) \leq \chi^2_{1,1-\beta} \right\}. \tag{12}$$

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

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

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 (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 (10).
Step 3: Compute the AP test statistic $Q^*(\theta)$ as in (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 (12), which collects all θ's such that $Q^*(\theta)$ is below the $1-\beta$ quantile of the χ^2_1 distribution. □

2. Empirical Applications

We now apply the proposed inference procedure to volatility measures in a manner akin to the role of an econometrician striving to identify the most suitable model for the data. This process illustrates the challenges encountered due to the striking similarity between the two models.

Daily realized volatility (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.[20] For the 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 (see Da and Xiu (2021) for the construction of these measures).[21] Table S1 of the internet appendix lists assets and summary statistics. We also conduct empirical analyses of international stock market indexes, for which the RV measures are obtained from the Realized Library and constructed as the sum of squared 5-minute intraday returns (see Table S2 in the Internet Appendix for a summary).

Following Andersen et al. (2003), we model each demeaned log RV series using the ARFIMA model in (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, we treat the RV measures as stand-alone time series and confine our attention to their properties. In principle it is possible to translate 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

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

[21] 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.

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

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

Figure 3 plots the one-dimensional confidence sets for the SPY and the nine industry ETFs, with panel A and panel B showing the results for α and d, respectively. Since the confidence set for (α, d) often contains two disjoint regions, we use two gray scales (dark and light) to signify them, so that the same-colored one-dimensional confidence sets of α and d are projected from the same parent two-dimensional confidence set. By convention, the dark-colored (resp., light-colored) confidence sets are associated with $d > 0$ (resp., $d < 0$). From Figure 3, we can see that 5 of the 10 ETFs (including SPY, XLP, XLU, XLV, and XLY) have confidence sets with disjoint regions. In dark-colored regions α is near-zero and the positive d 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 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 'coexist', 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.

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

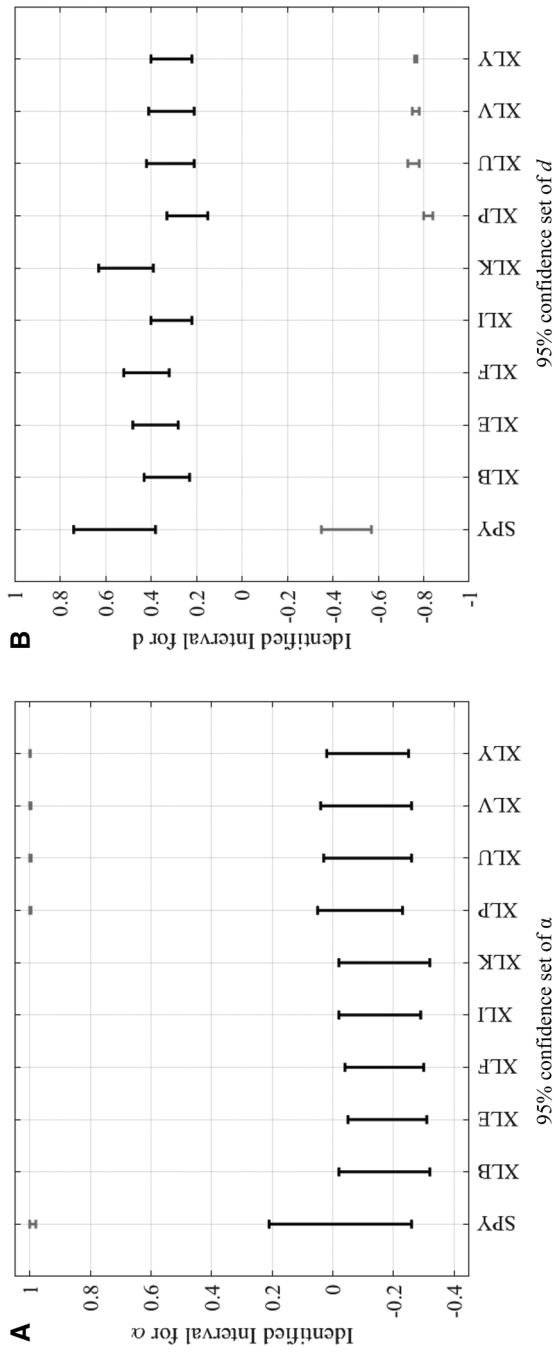

Figure 3
Confidence sets for selected ETFs
We compute 95% identification-robust confidence sets for the demeaned log RV of the 10 index ETFs from 1996 to 2021, and the projected confidence intervals of α and d are plotted in panels A and B. Labels on the x-axis are the tickers of the ETFs.

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

So far, the evidence from the 10 ETFs clearly demonstrates the empirical relevance of weak identification and the difficulty in robustly discriminating between long memory and rough dynamics. This phenomenon is not specific to ETFs. Similar analyses were conducted for each of the 30 constituent stocks of the Dow Jones Industrial Average, and Figure 4 plots the resultant 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.

Additional 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.[23] The same procedure is conducted for these market-level volatility measures, and Figure 5 plots the projection-based one-dimensional confidence sets. The confidence sets for nine of the 31 indices exhibit bifurcation, 18 of them are single-region, and the confidence sets for the remaining four indices (i.e., AEX, FCHI, FTMIB, and KSE) are empty.[24] These findings show that weak identification issues occur over a broad set of markets but that the overall evidence tends in favor of the long memory configuration, which is always present in the confidence set regardless 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 Internet Appendix provides additional subsample results that support the same conclusion.

The methodology is then applied to trading volume data, which holds independent economic interest and is expected to exhibit dynamics similar to price volatilities. This expectation is based on the mixture-of-distributions hypothesis (Clark 1973; Tauchen and Pitts 1983; Andersen 1996), which posits that both price volatility and trading volume are driven by underlying information flows. Additionally, trading volumes are somewhat easier to interpret than RV measures because volume series are directly observable

[23] Robustness checks based on alternative RV measures are provided in the Internet Appendix.

[24] 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).

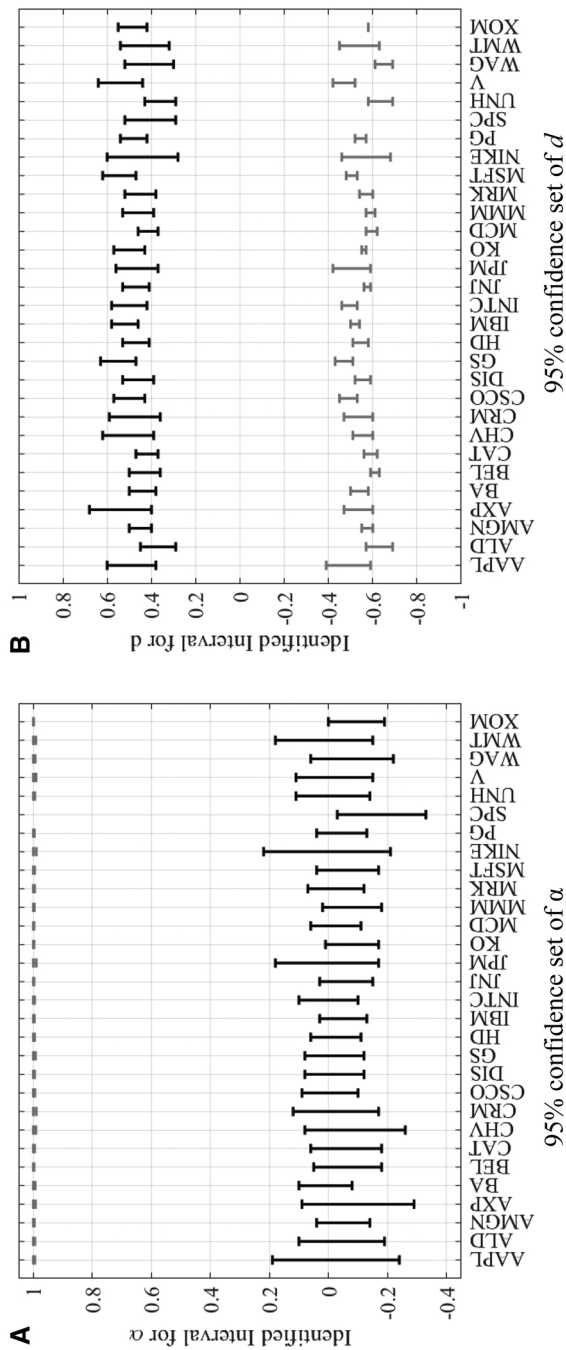

Figure 4
Confidence sets for Dow Jones Industrial Average component stocks
The panels show the 95% identification-robust confidence sets for the (demeaned) log RV of the 30 Dow Jones Industrial Average component stocks from 1996 to 2021. The left (right) panel provides projections of the confidence sets on the α-axis (d-axis) for each asset. Labels on the x-axis are the tickers of the 30 stocks.

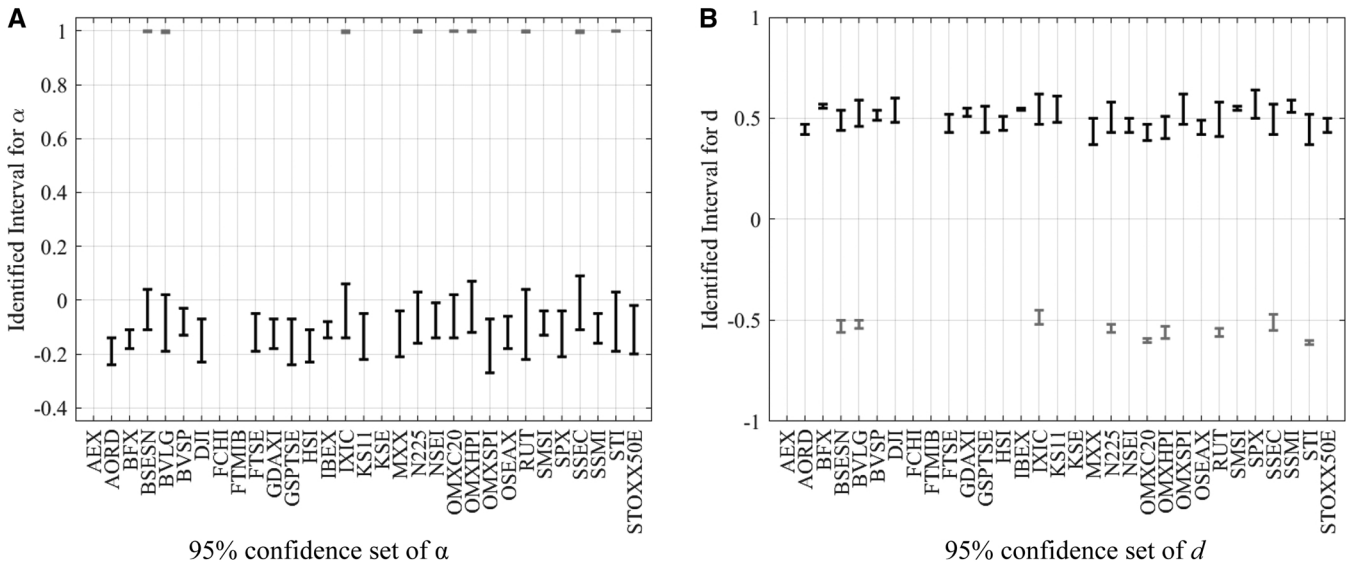

Figure 5
Confidence sets for international stock market indices
We compute 95% identification-robust confidence sets for the (demeaned) log RV of the 31 international stock market indices, and the confidence sets were projected onto the α- and d-axes. The left (right) panel displays the projected confidence intervals on the α-axis (d-axis) for each asset. Labels on the x-axis are the tickers of the 31 international stock market indices and their names can be found in Table S2 of the Internet Appendix.

(in contrast to RV measures which are often regarded as proxies for latent volatility functionals). Results are presented in Section S2 of the Internet Appendix. 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 and the other implying long memory with weak short-run dynamics.

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

3. Conclusion

In the early empirical finance literature the ARFIMA$(1,d,0)$ model was found to be an adequate model for log realized volatility with a fitted AR 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 AR parameters near-unity and memory parameters close to -0.5, providing evidence for 'rough volatility', reflecting a primary characteristic of antipersistent time series in contrast to long memory. This paper explains these coexisting, yet conflicting, empirical outcomes as a symptom of intrinsic weak identification within the ARFIMA model itself. Our theory suggests that, while the two parameter configurations appear very different, the distance between the corresponding models converges to zero when the AR parameter is localized to unity or zero. The resultant weak identification yields model ambiguities, explaining the emergence of two contrasting perspectives on volatility dynamics.[25]

To reveal the bifurcation in empirical data and address this potential ambiguity 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. This method not only serves as an effective tool for revealing the weak identification issue in empirical studies but also provides a robust framework for parameter inference in the presence of this issue. Extensive applications of this approach conducted on a broad range

[25] Weak identification does not equate to model equivalence. Evaluating the relative merits of the two models requires attention to the specific application context, and caution should be exercised when drawing conclusions from other settings.

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 weakness in ARFIMA modeling that this paper reveals 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 AR parameters vary according to news flow covariates that import information about memory in the data to assist in resolving ambiguities; another approach might incorporate such covariates in a multivariate system to jointly model news flows with relevant economic variables. Such extensions are left for future research.

Code Availability

The replication code is available in the Harvard Dataverse at https://dataverse.harvard.edu/api/access/datafile/10679995.

Appendix A. 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 $|\tilde{\alpha}_T|<\bar{\gamma}_T$ and set $\theta'=(\tilde{\alpha}_T,d_T+1)$. It is easy to see that $\theta'\in\widetilde{R}_T$. Hence,

$$\inf_{\tilde{\theta}\in\widetilde{R}_T}\sup_{\lambda}\left|\log f_{\theta}(\lambda)-\log f_{\tilde{\theta}}(\lambda)\right|\leq\sup_{\lambda}\left|\log f_{\theta}(\lambda)-\log f_{\theta'}(\lambda)\right|, \tag{A1}$$

where we have written $\sup_\lambda$ in place of $\sup_{\underline{\lambda}_T\leq|\lambda|\leq\pi}$ for brevity. By definition (recall (4)),

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

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

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

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

$$\leq\sup_{\lambda}\left|\log\left(\frac{2-2\cos(\lambda)}{1-2\alpha_T\cos(\lambda)+\alpha_T^2}\right)\right|+\sup_{\lambda}\left|\log\left(1-2\tilde{\alpha}_T\cos(\lambda)+\tilde{\alpha}_T^2\right)\right|. \tag{A2}$$

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

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

By the mean-value theorem, we further have

$$\sup_\lambda \left|\log\left(\frac{2-2\cos(\lambda)}{1-2\alpha_T\cos(\lambda)+\alpha_T^2}\right)\right| = \sup_\lambda \left|\log\left(\frac{2-2\cos(\lambda)}{1-2\alpha_T\cos(\lambda)+\alpha_T^2}\right) - \log(1)\right|$$

$$\leq \frac{9}{2-2\cos\left(\underline{\lambda}_T\right)} \sup_\lambda \left|1-2(1-\alpha_T)\cos(\lambda)-\alpha_T^2\right| \leq K\left(\frac{|1-\alpha_T|+|1-\alpha_T|^2}{\underline{\lambda}_T^2}\right). \tag{A3}$$

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

$$\sup_\lambda \left|\log\left(1-2\tilde{\alpha}_T\cos(\lambda)+\tilde{\alpha}_T^2\right)\right| \leq K\left(|\tilde{\alpha}_T|+\tilde{\alpha}_T^2\right). \tag{A4}$$

Combining (A1)–(A4) yields

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

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

Appendix B. Simulations: Identification-Robust Confidence Sets

We apply the proposed identification-robust inference method in a Monte Carlo setting that is designed to illustrate the intuition discussed in Section 1.2 and, at the same time, match some key patterns seen in empirical work, including our own study in Section 2. We generate the observed y_t series from the ARFIMA$(1,d,0)$ model (1) for different (α, d) parameter values, with the ε_t error terms simulated as i.i.d. standard normal variables.[26] 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.[27]

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 B1 plots the estimated confidence sets constructed

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

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

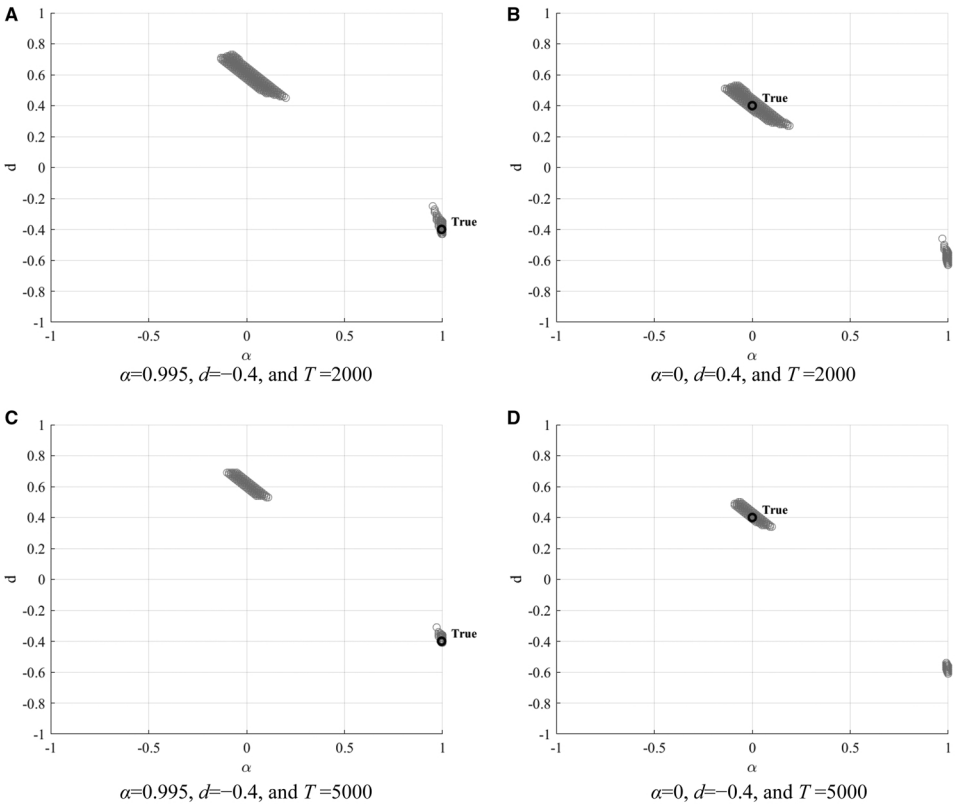

Figure B1
One-path illustration of identification-robust confidence sets
Data are simulated from the ARFIMA$(1,d,0)$ model (1) with different parameter settings. The figure displays the 95% identification-robust confidence set of the four data series, with α on the x-axis and d on the y-axis. The black circle represents the true model parameters.

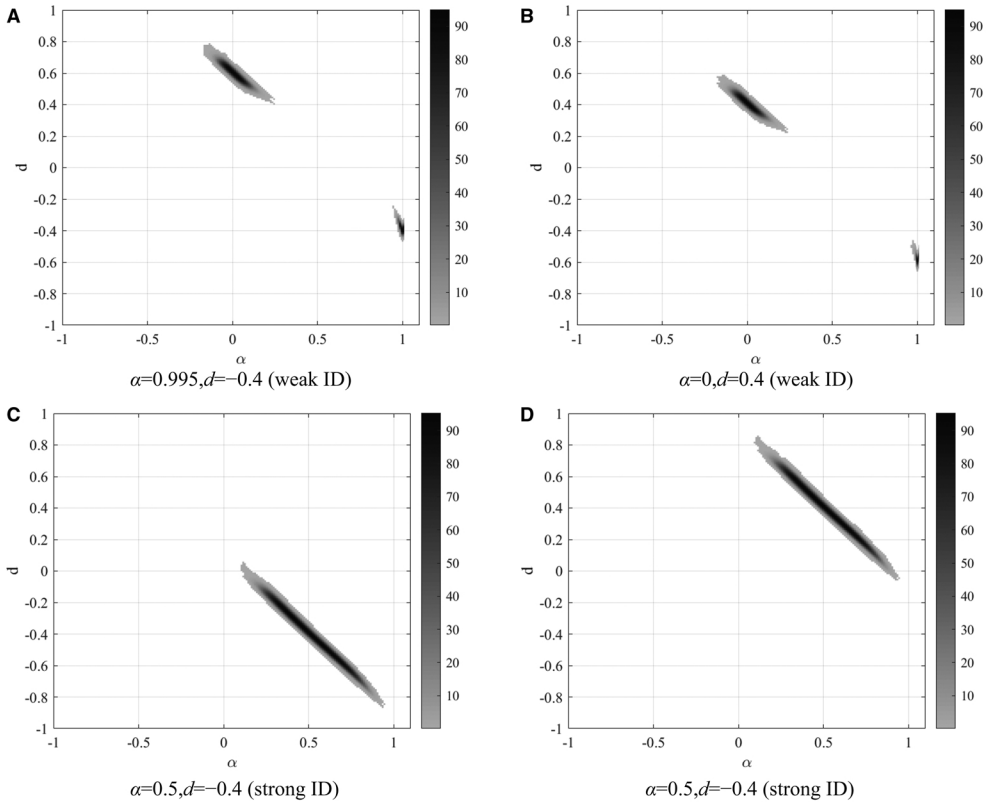

Figure B2
Coverage rates for identification-robust confidence sets
For each parameter setting, 1,000 data series were simulated from the ARFIMA(1, d, 0) model (1) and the 95% identification-robust confidence set computed for each series. The graphs show the frequency of a given (α, d) falling in the 95% identification-robust confidence set. Darker shading colors represent more frequent presence of the value in the confidence set. ID is short for identification.

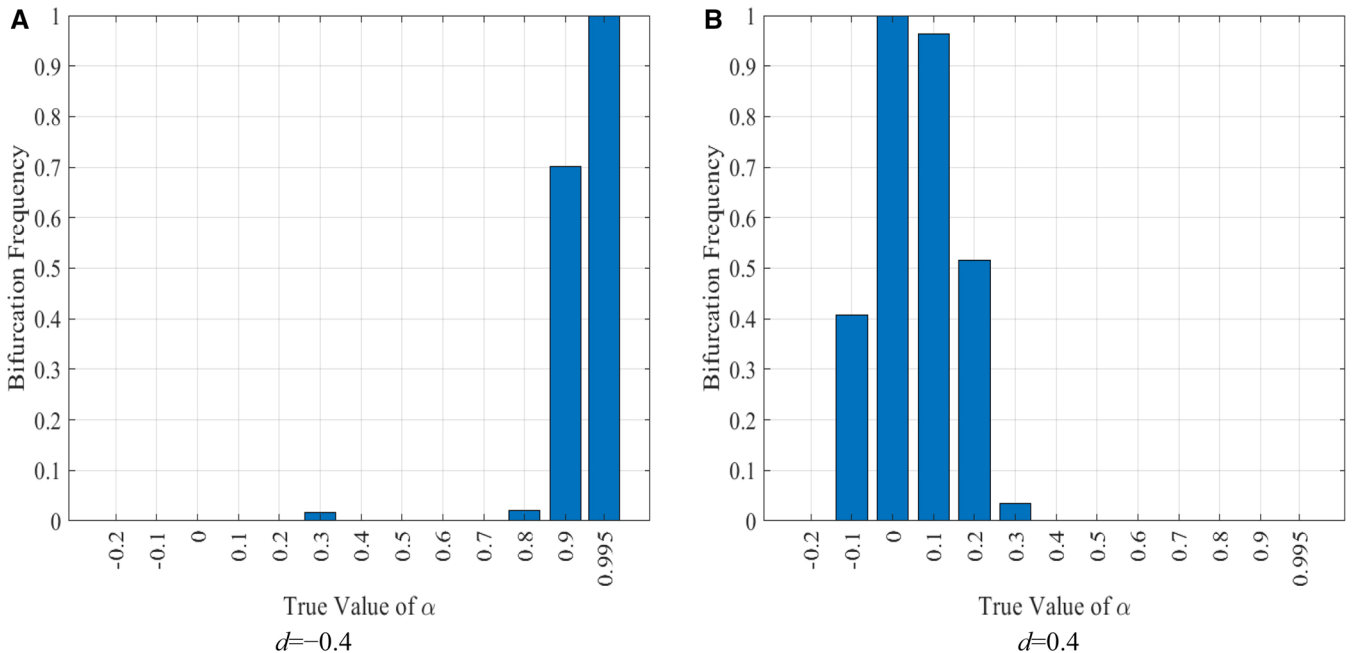

Figure B3
Bifurcation frequencies of the identification-robust confidence set
The autoregressive coefficient α takes a wide range of values (displayed on the horizontal axis). Under each parameter setting, we simulate 1,000 data series from ARFIMA$(1,d,0)$ and compute their 95% identification-robust confidence sets. The bifurcation frequency is the percentage of replications with two isolated areas in the confidence set. This figure plots the bifurcation frequency against the value of α when $d=-0.4$ (left panel) and $d=0.4$ (right panel).

for a single sample path in each of four Monte Carlo configurations. Specifically, we consider two sample sizes, $T=2,000$ or $T=5,000$, that are in line with the real data sets used in our empirical study. For each sample size, we consider two parameter configurations $(\alpha,d)=(0.995,-0.4)$ or $(0,0.4)$, which are representative of the empirical estimates in prior studies that support rough or long-memory volatility dynamics, respectively (see, e.g., Gatheral, Jaisson, and Rosenbaum (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 B1 reveals that they share a common 'bifurcation' pattern in which there are two disjoint regions regardless of which region actually contains the true parameters. One region features near-unity α and $d<0$ (signifying roughness), while the other features near-zero α and $d>0$ (signifying long-memory). Viewed through the lens of robust confidence set inference, neither of these two possibilities can be ruled out at the given confidence level, despite the fact that the parameter values seem highly disjoint and individually very different between the two regions. Although this indeterminacy may be disconcerting and unsatisfying from a practical viewpoint, it reflects the intrinsic difficulty arising from the near indistinguishability of the two parameter schemes, given the available data and the focus on autocovariances. Increasing the sample size from 2,000 to 5,000 sharpens the individual regions in the confidence sets as may be expected but it does not eliminate the bifurcation phenomenon.

The pattern depicted in these illustrations is representative. This may be shown by overlaying the plots in Figure B1 across all 1,000 Monte Carlo trials. More precisely, for each candidate parameter value on the (α,d) plane we compute the frequency that it falls in the confidence set; we then plot these coverage rates as a heatmap, where darker colors represent higher frequencies. For brevity, we focus on the case $T=5,000$, which is roughly the average sample size for data sets used in our empirical work. The top row of Figure B2 plots the coverage rate heatmaps for $(\alpha,d)=(0.995,-0.4)$ and $(\alpha,d)=(0,0.4)$, as in the illustrative examples. The bifurcation pattern is again self-evident, suggesting that the identification-robust confidence sets generally contain those two disjoint regions. While our approach does not estimate any parameter, our findings reinforce what Shi and Yu (2023) 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 B2. From the analysis in Section 1.2, 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 on the true parameter value, precisely what is expected in classical likelihood or moment-based inference. On the other hand, confidence sets under strong identification are not necessarily small. Instead, weak identification is revealed through nonstandard shapes, such as the bifurcation pattern seen here in the robust confidence sets, rather than by the size of the confidence set. Readers are referred to the literature for a deeper discussion of these differences (Staiger and Stock 1997; Stock and Wright 2000; Stock and Yogo 2005; Andrews and Cheng 2012; Andrews, Stock, and Sun 2019).

To reveal the incidence of bifurcation, Figure B3 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 B2. When $\alpha=0.9$, the bifurcation frequency drops to approximately 70%, suggesting that weak identification is still largely in play. As the true value of α moves farther 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 B3 shows that when $d>0$, weak identification is more severe when α is close to zero, complementing intuition and Theorem 1.

In cases of weak identification, the long-memory region is generally larger than the rough region, a result that is expected for the following reasons. First, from Figure B3, in the case of rough volatility (panel A) the weak identification issue disappears almost completely when $\gamma_T = 0.2$ (i.e., $\alpha = 0.8$), whereas in the case of long-memory (panel B) the bifurcation frequency remains more than 50% when $\tilde{\gamma}_T = 0.2$ (i.e., $\alpha = 0.2$). The asymmetric impact of γ_T and $\tilde{\gamma}_T$ on weak identification is also manifest in Theorem 1, where γ_T is shown to have a more significant impact on the order magnitude of the distance $O((\lambda_T^{-2}\gamma_T) \vee \tilde{\gamma}_T))$ than $\tilde{\gamma}_T$. It is therefore unsurprising that the long-memory region is larger than the rough volatility region. Second, we impose the restriction that the autoregressive coefficient $|\alpha| < 1$ and so the rough volatility region is truncated at $\alpha = 1$. Third, the power of the adaptive Portmanteau test may well differ in different regions of the parameter space. Alternative omnibus tests for serial correlation might be considered but are left to future work.

These Monte Carlo findings corroborate theory and intuition. In ARFIMA model simulations the (α, d) parameters show strong evidence of joint weak identification when α is near-unity or near-zero. In such cases, identification-robust confidence sets typically contain two distinct regions that exhibit the bifurcation pattern predicted by theory and guide the interpretation of our empirical findings presented in Section 2.

References

Agiakloglou, C., P. Newbold, and M. Wohar. 1993. Bias in an estimator of the fractional difference parameter. *Journal of Time Series Analysis* 14:235–46.

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

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

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

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

———. 2003. Modeling and forecasting realized volatility. *Econometrica* 71:579–625.

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

Andrews, D. W. K., and X. Cheng. 2012. Estimation and inference with weak, semi-strong, and strong identification. *Econometrica* 80:2153–211.

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

Andrews, I., and A. Mikusheva. 2022. Optimal decision rules for weak GMM. *Econometrica* 90:715–48.

Andrews, I., J. Stock, and L. Sun. 2019. Weak instruments in IV regression: Theory and practice. *Annual Review of Economics* 11:727–53.

Ansley, C. F., and P. Newbold. 1980. Finite sample properties of estimators for autoregressive moving average models. *Journal of Econometrics* 13:159–83.

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

Bolko, A. E., K. Christensen, M. S. Pakkanen, and B. Veliyev. 2023. A GMM approach to estimate the roughness of stochastic volatility. *Journal of Econometrics* 235:745–78.

Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31:307–27.

Bollerslev, T., J. Li, and Q. Li. 2024. Optimal nonparametric range-based volatility estimation. *Journal of Econometrics* 238:105548–.

Bollerslev, T., J. Li, and Z. Liao. 2021. Fixed-k inference for volatility. *Quantitative Economics* 12:1053–84.

Bollerslev, T., J. Li, and Y. Xue. 2018. Volume, volatility, and public news announcements. *Review of Economic Studies* 85:2005–41.

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

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

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

Chevillon, G., M. Massmann, and S. Mavroeidis. 2010. Inference in models with adaptive learning. *Journal of Monetary Economics* 57:341–51.

Clark, P. K. 1973. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* 41:135–55.

Corradi, V., and W. Distaso. 2006. Semi-parametric comparison of stochastic volatility models using realized measures. *Review of Economic Studies* 73:635–67.

Da, R., and D. Xiu. 2021. When moving-average models meet high-frequency data: Uniform inference on volatility. *Econometrica* 89:2787–825.

Dufour, J.-M. 1997. Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica* 65:1365–87.

El Euch, O., M. Fukasawa, and M. Rosenbaum. 2018. The microstructual foundations of leverage effect and rough volatility. *Finance and Stochastics* 22:241–80.

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

Escanciano, J. C., and I. N. Lobato. 2009. An automatic portmanteau test for serial correlation. *Journal of Econometrics* 151:140–9.

Fukasawa, M., T. Takabatake, and R. Westphal. 2022. Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics. *Mathematical Finance* 32:1086–132.

Garnier, J., and K. Sølna. 2017. Correction to Black-Scholes formula due to fractional stochastic volatility. *SIAM Journal Financial Mathematics* 8:560–88.

———. 2018. Option pricing under fast-varying and rough stochastic volatility. *Ann Finance* 14:489–516–.

Gatheral, J., T. Jaisson, and M. Rosenbaum. 2018. Volatility is rough. *Quantitative Finance* 18:933–49.

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

Geweke, J., and S. Porter-Hudak. 1983. The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4:221–38.

Giraitis, L., H. L. Koul, and D. Surgailis. 2012. *Large sample inference for long memory processes*. Imperial College Press. doi:10.1142/p591.

Granger, C. W. J. 1980. Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* 14:227–38.

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

Harvey, A. C. 2007. Long memory in stochastic volatility. In *Forecasting Volatility in the Financial Markets*, 351–63. Elsevier.

Hosking, J. R. 1981. Fractional differencing. *Biometrika* 68:165–76.

Jusselin, P., and M. Rosenbaum. 2020. No-arbitrage implies power-law market impact and rough volatility. *Mathematical Finance* 30:1309–36.

Li, J., and A. J. Patton. 2018. Asymptotic inference about predictive accuracy using high frequency data. *Journal of Econometrics* 203:223–40.

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

Moreira, M. J. 2003. A conditional likelihood ratio test for structural models. *Econometrica* 71:1027–48.

Nakamura, E., and J. Steinsson. 2018a. High-frequency identification of monetary non-neutrality: The information effect. *Quarterly Journal of Economics* 133:1283–330.

———. 2018b. Identification in macroeconomics. *Journal of Economic Perspectives* 32:59–86.

Phillips, P. C. B. 1987. Towards a unified asymptotic theory for autoregression. *Biometrika* 74:535–47.

———. 1989. Partially identified econometric models. *Econometric Theory* 5:181–240.

———. 1999. Discrete fourier transforms of fractional processes. *Cowles Foundation Discussion Paper No.1243 Yale University*.

———. 2023. Estimation and inference with near unit roots. *Econometric Theory* 39:221–63.

Phillips, P. C. B., and T. Magdalinos. 2007. Limit theory for moderate deviations from a unit root. *Journal of Econometrics* 136:115–30.

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* 33:578–609.

Rigobon, R. 2003. Identification through heteroskedasticity. *Review of Economics and Statistics* 85:777–92.

Risk Staff. 2021. Quants of the Year – Jim Gatheral and Mathieu Rosenbaum. *Risk.net* .

Robinson, P. M. 1978. Statistical inference for a random coefficient autoregressive model. *Scandinavian Journal of Statistics* 5:163–8.

———. 1995a. Gaussian semiparametric estimation of long range dependence. *Annals of Statistics* 23:1630–61.

———. 1995b. Log-periodogram regression of time series with long range dependence. *Annals of Statistics* 23:1048–1072.

Schennach, S. M. 2018. Long memory via networking. *Econometrica* 86:2221–48.

Shi, S., X. Liu, and J. Yu. 2025. Fractional stochastic volatility model. *Journal of Time Series Analysis* 46:378–97.

Shi, S., and J. Yu. 2023. Volatility puzzle: Long memory or anti-persistency. *Management Science* 7:3861–83.

Shimotsu, K., and P. C. B. Phillips. 2005. Exact local whittle estimation of fractional integration. *Annals of Statistics* 33:1890–933.

Solo, V. 1992. Intrinsic random functions and the paradox of 1/f noise. *SIAM Journal on Applied Mathematics* 52:270–91.

Staiger, D., and J. H. Stock. 1997. Instrumental variables regression with weak instruments. *Econometrica* 65:557–86.

Stock, J. H., and J. H. Wright. 2000. GMM with weak identification. *Econometrica* 68:1055–96.

Stock, J. H., and M. Yogo. 2005. *Testing for weak instruments in linear IV regression*, 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* 16:173–92.

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

Velasco, C., and P. M. Robinson. 2000. Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association* 95:1229–43.

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

———. 2023b. Modeling and forecasting realized volatility with the fractional Ornstein-Uhlenbeck process. *Journal of Econometrics* 232:389–415.