

A justification of conditional confidence intervals

Eric Beutner

*Department of Econometrics and Data Science, Vrije Universiteit Amsterdam,
The Netherlands*

e-mail: e.a.beutner@vu.nl

and

Alexander Heinemann*

Aegon N.V., Aegonplein 50, 2591 TV Den Haag

e-mail: Alexander.Heinemann@aegon.com

and

Stephan Smeekes

*Department of Quantitative Economics, Maastricht University, 6211LM Maastricht,
The Netherlands*

e-mail: s.smeekes@maastrichtuniversity.nl

Abstract: To quantify uncertainty around point estimates of conditional objects such as conditional means or variances, parameter uncertainty has to be taken into account. Attempts to incorporate parameter uncertainty are typically based on the unrealistic assumption of observing two independent processes, where one is used for parameter estimation, and the other for conditioning upon. Such unrealistic foundation raises the question whether these intervals are theoretically justified in a realistic setting. This paper presents an asymptotic justification for this type of intervals that does not require such an unrealistic assumption, but relies on a sample-split approach instead. By showing that our sample-split intervals coincide asymptotically with the standard intervals, we provide a novel, and realistic, justification for confidence intervals of conditional objects. The analysis is carried out for a rich class of time series models. We also present the results of a simulation study to evaluate the performance of the sample-split approach. The results indicate that also in practice sample-split intervals might be more appropriate than the standard intervals.

MSC2020 subject classifications: Primary 62M10, 62M20; secondary 60G10, 60G25.

Keywords and phrases: Conditional confidence intervals, parameter uncertainty, sample-splitting, prediction, merging.

Received December 2019.

Contents

1	Introduction	2518
---	------------------------	------

*The views are personal from the author and do not necessarily reflect the views of Aegon.

2	General setup	2520
2.1	The general prediction function	2520
2.1.1	Estimating the prediction function	2523
2.2	Argument of two independent processes	2526
2.3	Sample-split estimation	2527
3	Asymptotic justification	2528
3.1	Merging	2528
3.2	Merging of 2IP and SPL in probability	2530
3.3	Interval construction	2533
3.4	Interval construction under normality	2535
4	Simulation-based evaluation of sample-split intervals	2537
4.1	Conditional coverage	2537
4.2	Overall coverage	2539
5	Prediction intervals	2540
6	Conclusion	2542
A	Main asymptotic results	2543
A.1	Lemmas	2543
A.2	Proofs of theorems	2544
B	Additional proofs	2550
B.1	Proofs of lemmas	2550
B.2	Proofs of corollaries	2557
C	Graphs and tables	2559
	Acknowledgments	2562
	References	2562

1. Introduction

One of the open questions in time series is how to quantify uncertainty around point estimates of conditional objects such as conditional means or conditional variances. A fundamental issue arises in the construction of confidence intervals that ought to capture the parameter estimation uncertainty contained in these objects. This fundamental issue stems from the fact that on one hand one must *condition* on the sample as the past informs about the present, yet on the other hand one must allow the data up to now to be treated as *random* to account for estimation uncertainty. The issue is well-recognized in the literature, however in practice confidence intervals are commonly constructed by treating the sample *simultaneously* as fixed and random. Frequently, such approach is motivated by presuming to have two independent processes. Assuming two independent processes with the same stochastic structure, using one for conditioning and one for the estimation of the parameters, bypasses the issue. It is a mathematically convenient assumption as in such case the uncertainty quantification reduces to an ordinary inferential problem. However, practitioners rarely have a replicate, independent of the original series, at hand with the exception of perhaps some experimental settings. As such, the intervals commonly constructed by practitioners lack a satisfactory theoretical justification. Therefore it is the objective

of the present paper to develop a realistic justification for such confidence intervals around point estimates of conditional objects, and to study the statistical properties of the proposed justification through a simulation study.

In the literature the fundamental issue described above is encountered in various ways. In the specific case of a first-order autoregressive (AR) process with Gaussian innovations, [40] investigates the statistical dependence between the ordinary least squares (OLS) estimator and the endogenous variable conditioned upon. He obtains an Edgeworth-type expansion for the distribution of the conditional mean and, further, studies forecasting, where the fundamental issue equally arises.¹ [33, p. 95] explicitly states a two-independent-processes assumption in connection with vector AR models. He postulates that such assumption is asymptotically equivalent to using only data not conditioned upon for estimation. Other studies investigate parameter uncertainty by using resampling methods, that typically mimic a distribution in which the sample, or at least a subsample, is treated as fixed and random at the same time (cf. [37], [37], [38], [35], [35], [36]). Aware of this paradox, [29] points out that conditioning on observing specific *in-sample* values affects the parameter estimator, but the effect is often erroneously disregarded. Deviating from the various bootstrap approaches, [23] examines parameter uncertainty in interval forecasts in a classical statistical framework. Similar to a general regression framework, he conditions on an arbitrary fixed *out-of-sample* value to avoid the issue. However, conditioning on arbitrary fixed out-of-sample values appears incompatible with the usual setup of dynamics in which we condition on the final value(s) of the sample. Acknowledging the issue while avoiding the two independent processes argument bears careful statements as in [19] who write in view of this issue “the delta method ... suggests” (p. 162). Similarly, [39] notices that although such intervals “have been discussed in the econometrics literature, the particular assumptions that underlie them are not fully recognized” (p. 389).

This paper provides a novel, and realistic, justification for commonly constructed confidence intervals around point estimates of conditional objects. Our solution is based on a simple sample-split approach and a weak dependence condition, which allows to partition our sample into two asymptotically independent subsamples. For a rich class of time series models we construct asymptotically valid sample-split intervals, without relying on the assumption of observing two independent processes, and show that these intervals coincide asymptotically with the intervals commonly constructed by practitioners. As will be argued below, an appropriate concept to study conditional confidence intervals is merging, a concept that generalizes weak convergence. To the best of our knowledge, except for [3] and [32] who looked at this concept independently of [4], this paper is the only one to study merging in the context of conditional distributions. Moreover, our paper seems to be the first to employ merging of conditional distributions in time series. By employing this concept we avoid unnatural assumptions such as observing $X_T = x$ (in dynamic models), losing the time index T , and instead explicitly acknowledge that the conditional objects vary over time.

¹For prediction intervals some solutions have been discussed. We refer to Section 5.

The rest of the paper is organized as follows. Section 2 specifies the general setup and describes the argument of two independent processes as well as our sample-split approach. In Section 3 we establish merging among the proposed and the two-independent-processes estimator in probability under mild conditions. Further, we construct asymptotically valid sample-split intervals and show that these coincide asymptotically with the standard intervals. A simulation-based comparison of the performance of sample-split intervals and standard intervals in finite samples is provided in Section 4. The extension to prediction is discussed in Section 5. Section 6 concludes. The main proofs are collected in Appendix A, while Appendix B provides additional proofs of intermediate results.

2. General setup

2.1. The general prediction function

Let $\{X_t\}$ be a real-valued stochastic process defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. θ denotes a generic parameter vector of length $r \in \mathbb{N}$ and θ_0 the true value, unknown to the researcher.² Let $\Theta \subseteq \mathbb{R}^r$ be the corresponding parameter space.

Our general setup involves inference on an object that we call the *prediction function*, which is a function of both the process $\{X_t\}$ and of the parameter θ . It represents the random object of interest, and will typically express quantities such as a conditional mean or conditional variance (without conditioning on a specific value) as a function of the sample.

Definition 1. *The prediction function $\psi : \mathbb{R}^\infty \times \Theta \rightarrow \mathbb{R}$ is depending both on the parameter θ and the entire history of the process $\{X_t\}$, such that we can write the prediction of the quantity at time $T + 1$, using data up to time T , as*

$$\psi_{T+1} := \psi(X_T, X_{T-1}, \dots; \theta). \quad (2.1)$$

With this setup we can describe most of the possible applications of interest. We now provide three examples to illustrate the prediction function.

Example 1. Suppose the time series $\{X_t\}$ follows an AR(1) process given by

$$X_t = \beta X_{t-1} + \varepsilon_t, \quad (2.2)$$

where $|\beta| < 1$ and $\{\varepsilon_t\}$ are independent and identically distributed (*i.i.d.*) with $\mathbb{E}[\varepsilon_t] = 0$. The conditional mean of X_{T+1} given X_T is given by

$$\mu_{T+1} := \mathbb{E}[X_{T+1}|X_T] = \beta X_T. \quad (2.3)$$

Using the prediction function we can then write $\mu_{T+1} = \psi(X_T, X_{T-1}, \dots; \theta) = \beta X_T$ with $\theta = \beta$.

²Generally, in particular throughout Section 2, we do not distinguish between θ and θ_0 if there is no cause for confusion. In Section 3 we explicitly use θ_0 to avoid confusion.

A more precarious example, due to its large popularity, is the conditional variance in a generalized autoregressive conditional heteroskedasticity (GARCH) model [18, 9]. Whereas in the previous AR(1) case it suffices to condition on the terminal observation, the subsequent Example 2 is more extreme as the entire sample contains information about the object of interest.

Example 2. Suppose $\{X_t\}$ follows a GARCH(1, 1) process given by $X_t = \sigma_t \varepsilon_t$ with

$$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (2.4)$$

where $\omega > 0$, $\alpha \geq 0$, $1 > \beta \geq 0$ and $\{\varepsilon_t\}$ are *i.i.d.* with $\mathbb{E}[\varepsilon_t] = 0$ and $\mathbb{E}[\varepsilon_t^2] = 1$. The model's recursive structure implies

$$\sigma_{T+1}^2 = \frac{\omega}{1-\beta} + \alpha \sum_{k=0}^{\infty} \beta^k X_{T-k}^2. \quad (2.5)$$

It follows directly from (2.5) that

$$\sigma_{T+1}^2 = \psi(X_T, X_{T-1}, \dots; \theta) = \frac{\omega}{1-\beta} + \alpha \sum_{k=0}^{\infty} \beta^k X_{T-k}^2,$$

with $\theta = (\omega, \alpha, \beta)'$, and $\Theta \subset (0, \infty) \times [0, \infty) \times [0, 1)$.

The next example shows that for a large class of models the prediction function can be written in the form of Definition 1.

Example 3. Following [10], consider a Markov chain of the form

$$S_t = \varphi(S_{t-1}, X_t; \theta), \quad t = 1, 2, \dots \quad (2.6)$$

where φ is some map $\varphi : \mathbb{R}^a \times \mathbb{R} \times \Theta \rightarrow \mathbb{R}^a$. Whereas X_t is observable by the researcher at time t , S_t may be unobservable or only partially observable. The object of interest ψ_{T+1} is typically a function of the state of the Markov chain S_T , such that $\psi_{T+1} = \pi(S_T; \theta)$ for some function $\pi(\cdot)$. Through the recursion in (2.6), this is in turn a function of the past of X_T , such that we may write

$$\psi_{T+1} = \pi(S_T; \theta) = \psi_{T+1}(X_T, X_{T-1}, \dots; \theta).$$

Many stochastic processes are in fact Markov processes, including ARMA and GARCH models, several GARCH extensions such as Zakoian's (1994) threshold GARCH, and the set of *observation driven models* considered by [8]. For further details we refer to [4] and [5].

Note that in many cases, such as the GARCH(1,1) of Example 2, the prediction function actually depends on the infinite past of the series. In order to express (an approximation of) the prediction function in terms of observable variables only, we would need to replace X_t by s_t for all $t < 1$, where $\{s_t\}$ is

a sequence of (arbitrary) constants to which we refer as starting or initial values. For a fixed T , we accordingly define an approximate prediction function $\psi_{T+1}^s : \mathbb{R}^T \times \Theta \rightarrow \mathbb{R}$ that is only a function of observable variables as

$$\psi_{T+1}^s(\mathbf{X}_{1:T}; \theta) := \psi(X_T, X_{T-1}, \dots, X_1, s_0, s_{-1}, \dots; \theta), \quad (2.7)$$

where $\mathbf{X}_{1:T} = (X_1, \dots, X_T)'$. Note that, given the varying input of the left-hand side in (2.7), we now actually have a sequence of (varying) functions for $T \in \mathbb{N}$.

In many cases the values far in the past are negligible for a wide range of values for $\{s_t\}$. Consequently, ψ_{T+1}^s will be close to ψ_{T+1} . This property can be shown to hold for many different processes including the ones in the examples. We formalize the exact condition we need regarding the negligibility of the starting values in Assumption 1.b.

Although the prediction function typically represents a conditional object, we have not conditioned on anything yet in the definition. We therefore now extend the analysis by formally conditioning on observing a particular sample. Let $\mathbf{x}_{1:T} = (x_1, \dots, x_T)'$ denote a specific sample path of $\mathbf{X}_{1:T}$. Throughout the paper, we will discriminate between random variables and their realized counterparts by writing the former in capital and the latter in lowercase letters to avoid ambiguity.

As we will consider sample splitting later on, we define notation that also allows for conditioning on only a subsample. For that purpose, let $t_1 : t_2$ denote the (sub-)period from t_1 up to t_2 , and correspondingly $\mathbf{X}_{t_1:t_2} = (X_{t_1}, \dots, X_{t_2})'$ for any integers $1 \leq t_1 \leq t_2 \leq T$, with a corresponding definition for the observed subsample $\mathbf{x}_{t_1:t_2}$. Furthermore, let $\mathbf{X}_{t_1:T}^c = (c_1, \dots, c_{t_1-1}, X_{t_1}, \dots, X_T)'$ denote the vector where all non-considered subsamples are replaced by a sequence of constants $\{c_t\}$, in a similar way as we did for the starting values. We can now formally define the conditional prediction function.

Definition 2. *The prediction function conditional on observing $\mathbf{X}_{t_1:T} = \mathbf{x}_{t_1:T}$ is defined as*

$$\psi_{T+1|t_1:T} := \psi_{T+1}^s(\mathbf{x}_{t_1:T}^c; \theta). \quad (2.8)$$

Note that we phrase the conditional prediction function directly in terms of the approximate prediction function ψ_{T+1}^s rather than the true prediction function. We take this “shortcut” because we cannot observe x_0, x_{-1}, \dots , so we cannot condition on those values anyway. Therefore, the “true” conditional object (which we might represent as $\psi_{T+1|-\infty:T}$), is, from an applied point of view, only the theoretical benchmark.

Example 1. *(continued)* For the conditional mean of an AR(1) process, conditioning only on the terminal observation $X_T = x_T$ suffices; that is, for any $t_1 \geq 1$ and any sequence $\{c_t\}$, we have that

$$\psi_{T+1|t_1:T} = \psi_{T+1}^s(\mathbf{x}_{t_1:T}^c; \theta) = \beta x_T = \psi_{T+1}^s(\mathbf{x}_{T:T}^c; \theta) = \psi_{T+1|T}. \quad (2.9)$$

Example 2. *(continued)* For objects such as the conditional variance for the

GARCH(1,1), the conditioning set and the sequence $\{c_t\}$ make a difference, as

$$\begin{aligned} \psi_{T+1|t_1:T} = \psi_{T+1}^s(\mathbf{x}_{t_1:T}^c; \theta) &= \frac{\omega}{1-\beta} + \alpha \sum_{k=0}^{T-t_1} \beta^k x_{T-k}^2 \\ &+ \alpha\beta^{T-t_1} \sum_{k=1}^{t_1-1} \beta^k c_{t_1-k}^2 + \alpha\beta^T \sum_{k=0}^{\infty} \beta^k s_{-k}^2, \end{aligned} \tag{2.10}$$

which differs depending on the choice of t_1 . However, as will be shown later, with an appropriate choice of t_1 , our Assumption 1.b on the negligibility of the initial condition, also implies that the difference between $\psi_{T+1|t_1:T}$ and $\psi_{T+1|1:T}$ becomes negligibly asymptotically.

Before introducing estimators for θ let us discuss the objects we want to construct inference for. In principle there are two unknown objects one could develop statistical intervals for: $\psi_{T+1}^s(\mathbf{X}_{1:T}; \theta)$ (or slightly more generally $\psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)$) and $\psi_{T+1}^s(\mathbf{x}_{t_1:T}^c; \theta)$. For a GARCH(1,1), for instance, the first would read as

$$\psi_{T+1}^s(\mathbf{X}_{1:T}; \theta) = \frac{\omega}{1-\beta} + \alpha \sum_{k=0}^{T-1} \beta^k X_{T-k}^2 + \alpha\beta^T \sum_{k=0}^{\infty} \beta^k s_{-k}^2$$

whereas the second with $t_1 = 1$ reads as

$$\psi_{T+1}^s(\mathbf{x}_{1:T}; \theta) = \frac{\omega}{1-\beta} + \alpha \sum_{k=0}^{T-1} \beta^k x_{T-k}^2 + \alpha\beta^T \sum_{k=0}^{\infty} \beta^k s_{-k}^2$$

or more generally, if t_1 is not taken to be equal to one, as in (2.10). While statistical intervals for both objects can be constructed we focus here on conditional inference, i.e. on intervals for $\psi_{T+1|t_1:T} = \psi_{T+1}^s(\mathbf{x}_{t_1:T}^c; \theta)$. In a time series context intervals for $\psi_{T+1|t_1:T}$ are motivated by the *relevance property* of [25] which postulates that intervals should relate to what actually happened during the sample period opposed to what might have happened. Indeed, intervals for $\psi_{T+1|t_1:T}$ can theoretically be shown to be considerably shorter than the intervals for their unconditional counterparts. While the unconditional objects might lead to a conceptually easier analysis, our focus on the conditional objects is therefore not only theoretically but also empirically relevant.

2.1.1. Estimating the prediction function

As θ is unobserved, we need to estimate it. We assume that the estimator is based on a subsample $1 : T_E$ (with $1 \leq T_E \leq T$) of the process $\{X_t^E\}$ which is potentially a *different* sample than $\{X_t\}$ that arises in the prediction function. The estimator of θ based on $\mathbf{X}_{1:T_E}^E = (X_1^E, \dots, X_{T_E}^E)'$ will be denoted by $\hat{\theta}(\mathbf{X}_{1:T_E}^E)$. The introduction of $\{X_t^E\}$ serves three purposes: first, using a different process allows us to formulate the two-independent-processes argument

where $X_t^E = Y_t$, with $\{Y_t\}$ independent of $\{X_t\}$, $T_E = T$ and an interval is constructed for $\psi_{T+1|1:T}$. Second, it will allow us to discuss the standard approach where $X_t^E = X_t$, $T_E = T$, and an interval is constructed for $\psi_{T+1|1:T}$. Please note already here that this means that the same variables that arise in the prediction function are also used for estimating θ . Third, it allows us to define the sample splitting approach which we study here. In this approach $X_t^E = X_t$ and an interval is constructed for $\psi_{T+1|T_P:T}$ with $T_E < T_P$ (with $1 < T_P \leq T$) such that in contrast to the standard approach different subsamples are used for prediction and estimation.

Before we illustrate why the standard approach is problematic for constructing and evaluating conditional intervals, we need to define the final building block of prediction function estimation: the conditional prediction function estimator.

Definition 3. Let $1 \leq T_P \leq T$. Define the prediction function estimator conditional on observing $\mathbf{X}_{T_P:T} = \mathbf{x}_{T_P:T}$ as

$$\widehat{\psi}_{T+1|T_P:T} := \widehat{\psi}_{T+1}(\mathbf{x}_{T_P:T}, \mathbf{X}_{1:T_E}^E) = \psi_{T+1}^s(\mathbf{x}_{T_P:T}, \widehat{\theta}(\mathbf{X}_{1:T_E}^E)). \quad (2.11)$$

Note that in the above definition we do not condition on the sample $\mathbf{X}_{1:T_E}^E = \mathbf{x}_{1:T_E}^E$ that is used to estimate θ . The reason for not conditioning on $\mathbf{X}_{1:T_E}^E = \mathbf{x}_{1:T_E}^E$ is that the goal is to preserve the randomness in the second argument of ψ_{T+1}^s , i.e. in $\widehat{\theta}(\mathbf{X}_{1:T_E}^E)$, and consequently in $\widehat{\psi}_{T+1|T_P:T}$. Hence, if this goal is achieved we can use the (non-degenerate) conditional (on $\mathbf{X}_{T_P:T} = \mathbf{x}_{T_P:T}$) distribution of $\widehat{\psi}_{T+1|T_P:T}$ to construct confidence intervals for $\psi_{T+1|T_P:T}$. Having this said let us have a closer look at the standard approach. As mentioned above in the standard approach one has $X_t^E = X_t$, $T_P = 1$ and $T_E = T$. Hence, denoting by $\widehat{\psi}_{T+1|1:T}^{STA}$ the “standard” estimator of the prediction function conditional on observing $\mathbf{X}_{1:T} = \mathbf{x}_{1:T}$, it becomes

$$\widehat{\psi}_{T+1|1:T}^{STA} := \widehat{\psi}_{T+1}(\mathbf{x}_{1:T}, \mathbf{x}_{1:T}) = \psi_{T+1}^s(\mathbf{x}_{1:T}, \widehat{\theta}(\mathbf{x}_{1:T})). \quad (2.12)$$

Notice that there is no capital \mathbf{X} in (2.12) because there is only one sample and one typically conditions on all values of this sample. Hence, (2.12) is non-random and thus does not have a distribution that could be used to construct intervals. Instead, to still be able to construct a “standard-looking” interval in practice, researchers typically implicitly rely on the (approximate) quantiles of the estimator

$$\widehat{\psi}_{T+1|1:T}^{STA*} := \widehat{\psi}_{T+1}(\mathbf{x}_{1:T}, \mathbf{X}_{1:T}) = \psi_{T+1}^s(\mathbf{x}_{1:T}; \widehat{\theta}(\mathbf{X}_{1:T})). \quad (2.13)$$

It is well understood in the literature that considering the sample as random and non-random at the same time as in (2.13) does not provide a fully satisfactory justification of the intervals used in practice. For the readers not so familiar with the problem just discussed we provide two examples that both illustrate the problem arising from (2.13). The examples illustrate that the severity of the problem may vary; ranging from only complicating the analysis (Example 1) to making the analysis impossible (Example 2).

Example 1. (continued) For the AR(1), we know from (2.9) that $\psi_{T+1|1:T} = \psi_{T+1}^s(\mathbf{x}_{1:T}, \theta) = \beta x_T$. Estimating β by OLS, say $\hat{\beta}(\mathbf{X}_{1:T})$, the estimator in (2.13) becomes

$$\hat{\psi}_{T+1|1:T}^{STA*} = \psi_{T+1}(\mathbf{x}_{1:T}; \hat{\theta}(\mathbf{X}_{1:T})) = \psi_{T+1}^s(\mathbf{x}_{1:T}, \hat{\beta}(\mathbf{X}_{1:T})) = \hat{\beta}(\mathbf{X}_{1:T})x_T. \quad (2.14)$$

Note the discrepancy in treating the terminal observation as random in the estimation sample, yet fixed for the prediction sample. To construct an interval for βx_T , one uses

$$\sqrt{T}(\hat{\beta}(\mathbf{X}_{1:T}) - \beta) \xrightarrow{d} N(0, \sigma_\beta^2) \quad (2.15)$$

with $\sigma_\beta^2 = 1 - \beta^2$ (cf. [22, p. 215]) and that one can estimate the variance of this normal distribution by $\hat{\sigma}_\beta^2(\mathbf{X}_{1:T}) = 1 - \hat{\beta}(\mathbf{X}_{1:T})^2$. Then an interval for βx_T is typically constructed the following way:

$$\hat{\beta}(\mathbf{X}_{1:T})x_T \pm \Phi^{-1}(\gamma/2) x_T \hat{\sigma}_\beta(\mathbf{X}_{1:T})/\sqrt{T}, \quad (2.16)$$

where Φ^{-1} denotes the standard normal quantile function. However, the interval in (2.16) is hard to interpret as the terminal observation is treated simultaneously as fixed and random. In essence, researchers typically approximate the distribution of $\sqrt{T}(\hat{\beta}(\mathbf{X}_{1:T}) - \beta)x_T$ instead of the conditional distribution of $\sqrt{T}(\hat{\beta}(\mathbf{X}_{1:T}) - \beta)X_T$ given $X_T = x_T$. The approximation of the latter appears rather cumbersome because even the rather simple condition $X_T = x_T$ has an influence on the whole series $\mathbf{X}_{1:T}$; see, for instance, the discussion [29]. Despite the challenge, [40] obtains such approximation based on Edgeworth expansions in the case of $\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$.

Example 2. (continued) For the conditional variance of the GARCH(1,1), the standard estimator of the prediction function conditional on $\mathbf{X}_{1:T} = \mathbf{x}_{1:T}$, is given by

$$\begin{aligned} \hat{\sigma}_{T+1|1:T}^{2STA} = \psi_{T+1}^s(\mathbf{x}_{1:T}; \hat{\theta}(\mathbf{x}_{1:T})) &= \frac{\hat{\omega}(\mathbf{x}_{1:T})}{1 - \hat{\beta}(\mathbf{x}_{1:T})} + \hat{\alpha}(\mathbf{x}_{1:T}) \sum_{k=0}^{T-1} \hat{\beta}(\mathbf{x}_{1:T})^k x_{T-k}^2 \\ &+ \hat{\alpha}(\mathbf{x}_{1:T}) \hat{\beta}(\mathbf{x}_{1:T})^T \sum_{k=0}^{\infty} \hat{\beta}(\mathbf{x}_{1:T})^k s_{-k}^2, \end{aligned} \quad (2.17)$$

where $\hat{\theta}(\mathbf{x}_{1:T}) = (\hat{\omega}(\mathbf{x}_{1:T}), \hat{\alpha}(\mathbf{x}_{1:T}), \hat{\beta}(\mathbf{x}_{1:T}))'$ is some estimate for θ depending on $\mathbf{x}_{1:T}$. Clearly, (2.17) illustrates for the GARCH(1,1) the above mentioned problem that the standard estimator is not random (after conditioning). For the GARCH(1,1) the estimator in (2.13) whose quantiles are used for an interval

reads as

$$\begin{aligned} \hat{\sigma}_{T+1}^{2STA*} = \psi_{T+1}(\mathbf{x}_{1:T}; \hat{\theta}(\mathbf{X}_{1:T})) &= \frac{\hat{\omega}(\mathbf{X}_{1:T})}{1 - \hat{\beta}(\mathbf{X}_{1:T})} + \hat{\alpha}(\mathbf{X}_{1:T}) \sum_{k=0}^{T-1} \hat{\beta}(\mathbf{X}_{1:T})^k x_{T-k}^2 \\ &+ \hat{\alpha}(\mathbf{X}_{1:T}) \hat{\beta}(\mathbf{X}_{1:T})^T \sum_{k=0}^{\infty} \hat{\beta}(\mathbf{X}_{1:T})^k s_{-k}^2. \end{aligned} \quad (2.18)$$

This quantity exemplifies for the GARCH(1,1) that the complete sample is regarded as random and non-random at the same time. While for the AR(1) this complicated the analysis, yet not made it impossible, the dependence on the complete sample here makes it difficult to use this quantity to make meaningful probabilistic statements.

2.2. Argument of two independent processes

The argument of two independent processes can at least be traced back to [1], who studies the prediction of AR time series. It reoccurs in [31, p. 394]: “...the series used for estimation of parameters and the series used for prediction are generated from two independent processes which have the same stochastic structure.” The same argument also appears in [33, p. 95] and in [17]. Let $\{Y_t\}$ be a process independent of $\{X_t\}$ defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\{Y_t\}$ having the same stochastic structure as $\{X_t\}$. In addition to the sample $\mathbf{X}_{1:T}$ of the process $\{X_t\}$, suppose there is a sample $\mathbf{Y}_{1:T} = (Y_1, \dots, Y_T)'$ of the process $\{Y_t\}$ that we use as estimation sample, that is $\mathbf{X}_{1:T}^E = \mathbf{Y}_{1:T}$. In this situation we denote the conditional prediction function estimator of Definition 3 by $\hat{\psi}_{T+1|1:T}^{2IP}$ and it equals

$$\hat{\psi}_{T+1|1:T}^{2IP} := \hat{\psi}_{T+1}(\mathbf{x}_{1:T}, \mathbf{Y}_{1:T}) = \psi_{T+1}^s(\mathbf{x}_{1:T}; \hat{\theta}(\mathbf{Y}_{1:T})). \quad (2.19)$$

Notice that (2.19) does not have the same shortcoming as (2.13) because even if we consider $\mathbf{x}_{1:T}$ to be known we can nevertheless consider $\hat{\theta}(\mathbf{Y}_{1:T})$ to be random and can hence use its distribution to construct intervals. Throughout this paper, we call (2.19), the 2IP (two independent processes) estimator. Then, a conditional interval $I_\gamma^{2IP}(\mathbf{x}_{1:T}, \mathbf{Y}_{1:T})$ can be based on the (approximate) quantiles of $\hat{\psi}_{T+1|1:T}^{2IP}$. It satisfies

$$\mathbb{P} \left[I_\gamma^{2IP}(\mathbf{x}_{1:T}, \mathbf{Y}_{1:T}) \ni \psi_{T+1|1:T} \mid \mathbf{X}_{1:T} = \mathbf{x}_{1:T} \right] \stackrel{(\approx)}{=} 1 - \gamma, \quad (2.20)$$

with the approximate sign indicating asymptotic equivalence. Note that the independence implies that the distribution of $\mathbf{Y}_{1:T}$ in (2.20) does not depend on the realization $\mathbf{x}_{1:T}$, yet the statement does depend on $\mathbf{x}_{1:T}$ because the interval depends on it (for the AR(1) this can be directly seen from (2.16) when replacing $\mathbf{X}_{1:T}$ by $\mathbf{Y}_{1:T}$). Although the 2IP approach is statistically sound, it assumes two

independent processes with the same stochastic structure. [40] points out that the assumption “is quite unrealistic in practical situations” (p. 241). Indeed, it is difficult to imagine this assumption to be satisfied in any real-life application beyond experimental settings. Moreover, as only one sample realization is available, to compute the estimate of the interval $I_\gamma^{2IP}(\mathbf{x}_{1:T}, \mathbf{Y}_{1:T})$ it is frequently suggested to take $\mathbf{Y}_{1:T} = \mathbf{x}_{1:T}$, violating the independence assumption. Thus, the 2IP approach appears to be a rather questionable justification for the usual interval, and as such, in this paper we provide an alternative, realistic, justification of (asymptotically equivalent) intervals based on sample splitting.

2.3. Sample-split estimation

An intuitive motivation for the sample-split approach is the successive decline of the influence of past observations present in a substantial class of time series models. This property permits to split our sample into two (asymptotically) independent subsamples. Consider the end point of the estimation sample, T_E , and the starting point of the prediction sample, T_P satisfying $1 < T_E < T_P \leq T$, such that the two samples are non-overlapping. In this situation we denote the conditional prediction function estimator of Definition 3 by $\widehat{\psi}_{T+1|T_P:T}^{SPL}$ and it is given by

$$\widehat{\psi}_{T+1|T_P:T}^{SPL} := \widehat{\psi}_{T+1}(\mathbf{x}_{T_P:T}^c, \mathbf{X}_{1:T_E}) = \psi_{T+1}^s(\mathbf{x}_{T_P:T}^c; \hat{\theta}(\mathbf{X}_{1:T_E})). \tag{2.21}$$

Throughout the paper, we call (2.21) the SPL estimator (due to SPLitting). Similar to the two sample approach, we can consider the first argument of ψ_{T+1}^s in (2.21) as given and the second argument as random since the subsamples are non-overlapping. A conditional interval $I_\gamma^{SPL}(\mathbf{x}_{T_P:T}, \mathbf{X}_{1:T_E})$ can be constructed such that

$$\mathbb{P}\left[I_\gamma^{SPL}(\mathbf{x}_{T_P:T}, \mathbf{X}_{1:T_E}) \ni \psi_{T+1|T_P:T} \mid \mathbf{X}_{T_P:T} = \mathbf{x}_{T_P:T}\right] \underset{(\approx)}{=} 1 - \gamma. \tag{2.22}$$

This statement does make sense as there is still randomness in $\hat{\theta}(\mathbf{X}_{1:T_E})$ since $\mathbf{X}_{1:T_E}$ is not conditioned upon, yet the last $T - T_P + 1$ values of $\{X_t\}_{t=1}^T$ are fixed such that their randomness is not taken into account. Similar to (2.20), the statement in (2.22) does depend on $\mathbf{x}_{T_P:T}$ and in contrast to $\mathbf{x}_{1:T}$ in (2.20) the realization $\mathbf{x}_{T_P:T}$ may influence the distribution of $\mathbf{X}_{1:T_E}$. However, as said at the beginning of this subsection the idea of the sample split approach is that this dependence will vanish asymptotically.

Remark 1. In Section 3 we will discuss how T_E and T_P should be chosen from an asymptotic perspective to ensure that our regularity conditions are fulfilled. As we only consider sample splitting as a theoretical approach to validate commonly constructed conditional confidence intervals, these asymptotic guidelines are sufficient for our purposes and we do not have to consider how to choose T_E and T_P in practice. Of course, one could use the sample-split approach in practice to construct confidence intervals; we present some simulation results in Section 4

which indicate that the split-approach may lead to better intervals in practice. Of course, there is a trade-off between obtaining less dependence between the two subsamples and estimation precision. For the Gaussian AR(1) setting, [40] derives asymptotic expansions for the case where, in our notation, $T_E = T - l$ and $T_P = T$ for some $l \geq 0$, showing that even in this simple case there is indeed a trade-off as described above and the optimal choice of l is unclear. An interesting extension of our analysis, in particular in view of the results of Section 4, would therefore be to investigate the optimal choices of T_E and T_P to achieve the most accurate confidence intervals in small samples. However, determining the optimal choice is likely to require other tools than used here. This is therefore outside the scope of the current paper.

3. Asymptotic justification

In this section, we connect the sample-split procedure of Section 2.3 with the two-independent-samples approach of Section 2.2. First, in Section 3.1, we show that the notion of weak convergence is inadequate to study asymptotic closeness for objects that vary over time and discuss the concept of merging. Then, in Section 3.2 we link the 2IP and the SPL estimator by proving that their conditional distributions merge in probability (Theorem 1). Thereafter, in Section 3.3, we construct asymptotically valid intervals (Theorem 2) and show that the sample-split intervals coincide asymptotically with the intervals commonly constructed by practitioners (Theorem 3). Last, in Section 3.4, we state intervals of reduced form and simplified theoretical results under asymptotic normality of the parameter estimator.

3.1. Merging

To illustrate the inappropriateness of weak convergence in the context considered here, we revisit Example 1 for the 2IP approach and the SPL approach, which shows that studying asymptotic closeness between conditional distributions is often complicated by the absence of a limiting distribution.

Example 1. (*continued*) For the 2IP approach (2.15) implies that $\sqrt{T}(\hat{\beta}(\mathbf{Y}_{1:T}) - \beta) \xrightarrow{d} N(0, \sigma_\beta^2)$ and it entails that $\sqrt{T}(\hat{\beta}(\mathbf{Y}_{1:T}) - \beta)x$ converges weakly to $N(0, \sigma_\beta^2 x^2)$ for any fixed $x \neq 0$. Further, the result suggests that the conditional distribution of $\sqrt{T}(\hat{\beta}(\mathbf{Y}_{1:T}) - \beta)X_T$ given $X_T = x_T$, which is just the distribution of $\sqrt{T}(\hat{\beta}(\mathbf{Y}_{1:T}) - \beta)x_T$, is asymptotically close to $N(0, \sigma_\beta^2 x_T^2)$. Similarly, for the SPL-approach with $T_E/T \rightarrow 1$ we have $\sqrt{T}(\hat{\beta}(\mathbf{X}_{1:T_E}) - \beta) \xrightarrow{d} N(0, \sigma_\beta^2)$ which suggests as well (if the gap between T_E and T is large enough which will be formally specified below) that the conditional distribution of $\sqrt{T_E}(\hat{\beta}(\mathbf{X}_{1:T_E}) - \beta)X_T$ given $X_T = x_T$ is also close to $N(0, \sigma_\beta^2 x_T^2)$. For both approaches the approximating distribution $N(0, \sigma_\beta^2 x_T^2)$ varies with T through the terminal realization

x_T . Note that the concept of weak convergence is not applicable in this context to characterize this asymptotic closeness, as it requires a (fixed) limiting distribution, which is absent here.

Next, we discuss what closeness means in the absence of a limiting distribution. To do so, first recall that weak convergence of a sequence of cdfs $\{F_T\}$ on \mathbb{R}^k with $k \in \mathbb{N}$, i.e. $F_T(x) \rightarrow F(x)$ for all continuity points of F , can alternatively be defined by $d_{BL}(F_T, F) \rightarrow 0$. Here d_{BL} denotes the bounded Lipschitz metric defined by

$$d_{BL}(F, G) = \sup \left\{ \left| \int f d(F - G) \right| : \|f\|_{BL} \leq 1 \right\}, \quad (3.1)$$

where for any real-valued function f on \mathbb{R}^k one puts $\|f\|_{BL} = \sup_x |f(x)| + \sup_{x \neq y} \frac{|f(x) - f(y)|}{\|x - y\|}$, with $\|\cdot\|$ denoting the Euclidean norm, i.e. $\|A\| = \sqrt{\text{tr}(A'A)}$ for any vector or matrix A . We will employ the bounded Lipschitz metric in Definition 4 below to define merging for sequences of distribution functions. Our definition follows [16] who uses the bounded Lipschitz metric with probability measures replacing the distribution functions in Equation (3.1) to define merging for sequences of probability measures on separable metric spaces. Because we are working on the real line we prefer to define merging for distribution functions instead of probability measures. Before giving the definition of merging for distribution functions it is worth mentioning that [16] shows that defining merging with respect to the bounded Lipschitz metric is equivalent to defining merging with respect to the Lévy-Prokhorov metric; for discussions on further possibilities to define merging see [13] and [14].

Definition 4 (Merging). *Two sequences of cdfs $\{F_T\}$ and $\{G_T\}$ are said to merge if and only if $d_{BL}(F_T, G_T) \rightarrow 0$ as $T \rightarrow \infty$.*

Note that weak convergence can be seen as a special case of merging with $G_T = G$ for all $T \in \mathbb{N}$.

While merging is appropriate to capture the asymptotic closeness of the conditional distribution of $\sqrt{T}(\hat{\beta}(\mathbf{Y}_{1:T}) - \beta)X_T$ and $N(0, \sigma_\beta^2 x_T^2)$ for a given sample $X_T = x_T$, we now extend the concept in a way that allows us to deal with asymptotic closeness when we do not condition on a particular sample. The necessity of this definition can again be exemplified by the AR(1), which also illustrates how we will deal the dependence of the statements in (2.20) and in (2.22) on the sample that we mentioned below these equations. For instance, in Example 1 as described at the beginning of this section, the goal would be to formalize a statement like ‘when T is large, the probability of all x_T such that the distribution of $\sqrt{T}(\hat{\beta}(\mathbf{X}_{1:T}) - \beta)x_T$ merges with that of $\sqrt{T_E}(\hat{\beta}(\mathbf{X}_{1:T_E}) - \beta)x_T$ is approximately equal to one’. We now first introduce the conditional distributions of the 2IP and the SPL estimator in the general case and then give the definition capturing what we just illustrated for the AR(1).

Let m_T and m_{T_E} be two sequences of normalizing constants with $m_T, m_{T_E} \rightarrow \infty$ (e.g. $m_T = \sqrt{T}$ and $m_{T_E} = \sqrt{T_E}$). For any $t_1 \geq 1$, we define the sub σ -algebra $\mathcal{I}_{t_1:T} = \sigma(X_t : t_1 \leq t \leq T)$. We denote the *conditional cdfs* of the 2IP and SPL

estimator by

$$F_T^{2IP}(\tau|\mathcal{I}_{1:T}) := \mathbb{P}[m_T(\hat{\psi}_{T+1}(\mathbf{X}_{1:T}, \mathbf{Y}_{1:T}) - \psi_{T+1}) \leq \tau | \mathcal{I}_{1:T}] \quad (3.2)$$

$$F_T^{SPL}(\tau|\mathcal{I}_{T_P:T}) := \mathbb{P}[m_{T_E}(\hat{\psi}_{T+1}(\mathbf{X}_{T_P:T}, \mathbf{X}_{1:T_E}) - \psi_{T+1}) \leq \tau | \mathcal{I}_{T_P:T}], \quad (3.3)$$

respectively, so that by specifying an event of $\mathcal{I}_{1:T}$ and $\mathcal{I}_{T_P:T}$, we see that (3.2) and (3.3) are just the centered and scaled distributions of (2.19) and (2.21), respectively. Please note that (3.3) actually also depends on c , see (2.21), but since our assumptions will ensure that this dependence vanishes asymptotically we prefer to suppress the dependence on c here.

Remark 2. Although not explicitly mentioned above we consider (3.2) and (3.3) to be regular conditional cdfs, which indicates that we assume that $F_T^{2IP}(\cdot|\mathcal{I}_{1:T})(\omega)$ and $F_T^{SPL}(\cdot|\mathcal{I}_{T_P:T})(\omega)$ are cdfs for every $\omega \in \Omega$; for the exact definition and the existence see [16, Section 10.2].

We can now define merging in probability (we do so without explicitly using the conditional cdfs of the 2IP and the SPL approach).

Definition 5 (Merging in Probability). *Two sequences of conditional cdfs $\{F_T\}$ and $\{G_T\}$ are said to merge in probability if and only if $d_{BL}(F_T, G_T) \xrightarrow{P} 0$ as $T \rightarrow \infty$, where “ \xrightarrow{P} ” denotes “convergence in probability”.*

Remark 3. As alluded to in the introduction [3] also consider merging and merging in probability. In contrast to Definitions 4 and 5, respectively, they use the set of bounded continuous functions on the real line to define merging and also to define merging in probability; see their Definitions 1 and 2, respectively. The connection between this approach to define merging and the Lévy-Prokhorov metric is considered in [14].

3.2. Merging of 2IP and SPL in probability

Here, we give conditions such that the conditional cdfs of the 2IP and SPL estimator merge. Clearly, the conditional confidence intervals are functions of these distributions so that their merging is a building block for the study of the conditional confidence intervals based on them. The conditions we give are divided into three parts. Roughly speaking, the first part (general assumptions) makes sure that the function we want to predict is well behaved and that we can estimate the parameter it depends on. The second part (two independent processes) and third part (SPL estimator) guarantee that these assumptions are met by the 2IP and the SPL method. To write the conditions in compact form we employ the usual stochastic order symbols O_p and o_p . We assume that θ_0 belongs to the interior of Θ , i.e. $\theta_0 \in \overset{\circ}{\Theta}$, and we denote the set of all bounded, real-valued Lipschitz functions on \mathbb{R}^r by $BL = \{h : \mathbb{R}^r \rightarrow \mathbb{R} : \|h\|_{BL} < \infty\}$. We start with the general assumptions.

Assumption 1 (General Assumptions).

- a (*Estimator*) $m_T(\hat{\theta}(\mathbf{X}_{1:T}) - \theta_0) \xrightarrow{d} G_\infty$ as $T \rightarrow \infty$ for some cdf $G_\infty : \mathbb{R}^r \rightarrow [0, 1]$;

- b (*Differentiability*) $\psi(\cdot; \theta)$ is continuous on Θ and twice differentiable on $\mathring{\Theta}$;
c (*Gradient*) $\left\| \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \theta} \right\| = O_p(1)$;
d (*Hessian*) $\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \theta \partial \theta'} \right\| = O_p(1)$ for some open neighborhood $\mathcal{V}(\theta_0)$ around θ_0 ;
e (*Initial Condition*) Given sequences $\{s_t\}$ and $\{c_t\}$, we have

$$\begin{aligned} m_T(\psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0) - \psi(X_T, X_{T-1}, \dots; \theta_0)) &= o_p(1), \\ \left\| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta_0)}{\partial \theta} - \frac{\partial \psi(X_T, X_{T-1}, \dots; \theta_0)}{\partial \theta} \right\| &= o_p(1), \\ \sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{t_1:T}^c; \theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \psi(X_T, X_{T-1}, \dots; \theta)}{\partial \theta \partial \theta'} \right\| &= o_p(1) \end{aligned}$$

for any $t_1 \geq 1$ such that $(T - t_1)/l_T \rightarrow \infty$ as $T \rightarrow \infty$ and for some model-specific l_T with $l_T \rightarrow \infty$.

Assumption 1.a implies the existence of a limiting distribution for the parameter estimator. The differentiability assumption in 1.b plus the boundedness Assumptions 1.c ensure that the scaled prediction function estimators can accurately be approximated by a Taylor expansion; see Lemma 1 for details. Assumption 1.e with $t_1 = 1$ ensures the negligibility of the starting values when using the full-sample for prediction, while taking $t_1 = T_P$ ensures that this extends to the case where additionally X_1, \dots, X_{T_P-1} are replaced by constants, i.e. where only the subsample (X_{T_P}, \dots, X_T) is used for prediction. This assumption implicitly limits the choice of T_P ; as replacing past values of X_t for $t < T_P$ by arbitrary constants should have a negligible effect, $T - T_P$ needs to increase faster than some lower bound l_T . For models exhibiting an exponential decay in memory, it typically suffices to take $l_T = \log T$ (see e.g. [6, Eq. (4.6)]).

For the 2IP estimator, we additionally need the two-independent-processes assumption, which is formalized in Assumption 2.

Assumption 2 (Two Independent Processes).

- a (*Existence*) $\{Y_t\}$ is a process defined on $(\Omega, \mathcal{F}, \mathbb{P})$, distributed as $\{X_t\}$;
b (*Independence*) $\{Y_t\}$ is independent of $\{X_t\}$.

For the SPL estimator we replace the two-independent-processes assumption by a stationarity and a weak dependence condition, which allows to split our sample into two (asymptotically) independent and identical subsamples. In addition we need an assumption on T_P and T_E as functions of T , that is $T_E(T)$ and $T_P(T)$.

Assumption 3 (SPL Estimator).

- a (*Rates*) The functions $T_P : \mathbb{N} \rightarrow \mathbb{N}$ and $T_E : \mathbb{N} \rightarrow \mathbb{N}$ satisfy $T_E(T) < T_P(T)$ for all T , while $\frac{T - T_P(T)}{l_T} \rightarrow \infty$ and $m_{T_E(T)}/m_T \rightarrow 1$ as $T \rightarrow \infty$;
b (*Strict Stationarity*) $\{X_t\}$ is a strictly stationary process;

c (Weak Dependence) $\{X_t\}$ satisfies for each $h \in BL$

$$\int h d\left(G_{T_E}^{SPL}(\cdot|\mathcal{I}_{T_P:T}) - G_{T_E}^{SPL}\right) \xrightarrow{p} 0 \quad \text{as} \quad T \rightarrow \infty,$$

where $G_{T_E}^{SPL}$ denotes the unconditional cdf of $m_{T_E}(\hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0)$ and $G_{T_E}^{SPL}(\cdot|\mathcal{I}_{T_P:T})$ the corresponding conditional cdf given $\mathcal{I}_{T_P:T}$.

The subsample size assumption in 3.a ensures that the number of observations used for conditioning is increasing, which along with the negligibility of the initial conditions implies that the truncation of the prediction function is negligible. Furthermore, the sample size used for estimation should increase fast enough that the respective scaling of the 2IP and SPL estimators, m_T and m_{T_E} respectively, are asymptotically identical. If m_T increases no faster than a polynomial rate, which is generally the case, it is sufficient that $T_E/T \rightarrow 1$ for $m_{T_E}/m_T \rightarrow 1$ to hold.

The stationarity assumption in 3.b can actually be relaxed; what matters is that the conditions in Assumption 1 are still true if only a subsample is considered. In particular, we need that $m_{T_E}(\hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0) \xrightarrow{d} G_\infty$, which – along with the assumptions on gradient and Hessian – is certainly satisfied under stationarity. However, in general the assumption will be far too strict; here we use it simply to have a clear, interpretable assumption rather than a list of high-level assumptions that are difficult to interpret. The weak dependence condition in 3.c is met by numerous Markov processes. Intuitively, (X_1, \dots, X_{T_E}) and (X_{T_P}, \dots, X_T) approach independence as their temporal distance $T_P - T_E$ increases. We illustrate a particular case in the Remark 4.

Remark 4. Suppose $\{X_t\}$ is strong mixing (cf. [15]) and let α denote the strong mixing coefficient. For $h \in BL$ and for all $\epsilon > 0$, Markov's and Ibragimov's inequality (cf. [21, Theorem A.5]) imply

$$\begin{aligned} & \mathbb{P}\left[\left|\int h d\left(G_{T_E}^{SPL}(\cdot|\mathcal{I}_{T_P:T}) - G_{T_E}^{SPL}\right)\right| \geq \epsilon\right] \\ & \leq \frac{1}{\epsilon} \mathbb{E}\left[\left|\int h d\left(G_{T_E}^{SPL}(\cdot|\mathcal{I}_{T_P:T}) - G_{T_E}^{SPL}\right)\right|\right] \\ & = \frac{1}{\epsilon} \text{Cov}\left[h\left(m_{T_E}(\hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0)\right), \text{sign}\left\{\int h d\left(G_{T_E}^{SPL}(\cdot|\mathcal{I}_{T_P:T}) - G_{T_E}^{SPL}\right)\right\}\right] \\ & \leq \frac{4\|h\|_{BL}}{\epsilon} \alpha(T_P - T_E). \end{aligned}$$

Taking $T_P - T_E \rightarrow \infty$ such that $\alpha(T_P - T_E) \rightarrow 0$ verifies Assumption 3.c.

Assumptions 1 to 3 are met by the AR and GARCH processes considered in Examples 1 and 2 (with 1.e holding for bounded sequences). A detailed verification of each assumption under mild conditions is provided in [6]. We state the following theorem.

Theorem 1 (Merging of 2IP and SPL). *Under Assumptions 1 to 3, $F_T^{2IP}(\cdot|\mathcal{I}_{1:T})$ and $F_T^{SPL}(\cdot|\mathcal{I}_{T_P:T})$ merge in probability.*

Having established asymptotic closeness between the conditional cdfs $F_T^{2IP}(\cdot|\mathcal{I}_{1:T})$ and $F_T^{SPL}(\cdot|\mathcal{I}_{T_P:T})$, we now turn to the construction of asymptotic intervals.

3.3. Interval construction

Henceforth, for any cdf F we write F^{-1} to denote its generalized inverse given by $F^{-1}(u) = \inf \{ \tau \in \mathbb{R} : F(\tau) \geq u \}$. A confidence interval for ψ_{T+1} based on quantiles of (3.2) or (3.3) is typically infeasible as these cumulative distribution functions are unknown for finite T . Here, they are infeasible because roughly they are the distribution functions of some weights which induce merging multiplied by $m_T(\hat{\theta}(\mathbf{X}_{1:T}) - \theta_0)$ and $m_{T_E}(\hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0)$, respectively, where, in general, the distributions of $m_T(\hat{\theta}(\mathbf{X}_{1:T}) - \theta_0)$ and $m_{T_E}(\hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0)$ are unknown in finite samples. Since these are the only unknown distributions an asymptotic approximation can be based on G_∞ with merging induced by the non-convergent weights. In general, we also need to estimate G_∞ ; see Examples 4 and 5 below for common approaches. We denote estimators of (3.2) and (3.3) resulting from this approximation by $\widehat{F}_T^{2IP}(\cdot)$ and $\widehat{F}_T^{SPL}(\cdot)$, respectively. In Section 3.4, we provide explicit expressions when G_∞ is multivariate normal. For the general construction, we refer to relations (A.6) and (A.7) in Appendix A and the explanations preceding these relations. Based on the 2IP approach, we consider an interval of the form

$$\begin{aligned}
 & I_\gamma^{2IP}(\mathbf{x}_{1:T}, \mathbf{Y}_{1:T}) \\
 &= \left[\hat{\psi}_{T+1}(\mathbf{x}_{1:T}, \mathbf{Y}_{1:T}) - \frac{\widehat{F}_T^{2IP}{}^{-1}(1 - \gamma_2)}{m_T}, \hat{\psi}_{T+1}(\mathbf{x}_{1:T}, \mathbf{Y}_{1:T}) - \frac{\widehat{F}_T^{2IP}{}^{-1}(\gamma_1)}{m_T} \right] \\
 (2.19) \quad & \stackrel{=}{=} \left[\hat{\psi}_{T+1|1:T}^{2IP} - \frac{\widehat{F}_T^{2IP}{}^{-1}(1 - \gamma_2)}{m_T}, \hat{\psi}_{T+1|1:T}^{2IP} - \frac{\widehat{F}_T^{2IP}{}^{-1}(\gamma_1)}{m_T} \right], \tag{3.4}
 \end{aligned}$$

where $\gamma_1, \gamma_2 \in [0, 1)$ satisfy $\gamma = \gamma_1 + \gamma_2$. We typically take $\gamma_1 = \gamma_2 = \gamma/2$ such that the interval is equal-tailed. Similarly, we construct the following sample split interval:

$$\begin{aligned}
 & I_\gamma^{SPL}(\mathbf{x}_{T_P:T}^c, \mathbf{X}_{1:T_E}) \\
 &= \left[\hat{\psi}_{T+1}(\mathbf{x}_{T_P:T}^c, \mathbf{X}_{1:T_E}) - \frac{\widehat{F}_T^{SPL}{}^{-1}(1 - \gamma_2)}{m_{T_E}}, \hat{\psi}_{T+1}(\mathbf{x}_{T_P:T}^c, \mathbf{X}_{1:T_E}) - \frac{\widehat{F}_T^{SPL}{}^{-1}(\gamma_1)}{m_{T_E}} \right] \\
 (2.21) \quad & \stackrel{=}{=} \left[\hat{\psi}_{T+1|T_P:T}^{SPL} - \frac{\widehat{F}_T^{SPL}{}^{-1}(1 - \gamma_2)}{m_{T_E}}, \hat{\psi}_{T+1|T_P:T}^{SPL} - \frac{\widehat{F}_T^{SPL}{}^{-1}(\gamma_1)}{m_{T_E}} \right]. \tag{3.5}
 \end{aligned}$$

To achieve correct coverage, we need that $\widehat{F}_T^{2IP}(\cdot)$ and $F_T^{2IP}(\cdot|\mathcal{I}_{1:T})$ merge in probability and likewise for SPL. A sufficient condition for this in our setting is that we can consistently estimate the asymptotic distribution of the parameter

estimator, G_∞ , by an appropriate estimator. This is formulated in Assumption 4 below.

Assumption 4 (CDF Estimator). Let $\widehat{G}_T(\cdot)$ denote a random (r -dimensional) cdf as a function of $\mathbf{X}_{1:T}$, used to estimate G_∞ . Then $\int h d\widehat{G}_T(\cdot) \xrightarrow{P} \int h dG_\infty$ as $T \rightarrow \infty$ for all $h \in BL$.

Although we did not explicitly specify in Assumption 4 the dependence of \widehat{G}_T on $\mathbf{X}_{1:T}$, it should be understood to hold for any subsample of $\mathbf{X}_{1:T}$ whose size goes to infinity. The verification of Assumption 4 is a standard step in asymptotic analysis. The two examples below provide common methods for verifying Assumption 4.

Example 4. Suppose that G_∞ belongs to some parametric family $\{G_{\theta,\xi} | \theta \in \Theta, \xi \in \Xi\}$. Then, given some consistent estimators $\hat{\theta}(\mathbf{X}_{1:T})$ and $\hat{\xi}(\mathbf{X}_{1:T})$ for θ_0 and ξ_0 respectively, it follows from the continuous mapping theorem that $\widehat{G}_T = G_{\hat{\theta}(\mathbf{X}_{1:T}), \hat{\xi}(\mathbf{X}_{1:T})}$ satisfies Assumption 4 if $G_{\theta,\xi}$ is continuous in θ and ξ .

Example 5. If \widehat{G}_T is based on a consistent bootstrap procedure for G_∞ then Assumption 4 clearly holds.

The following theorem states the intervals' asymptotic validity.

Theorem 2 (Asymptotic Coverage).

1. (a) Under Assumption 1, 2 and 4, $F_T^{2IP}(\cdot | \mathcal{I}_{1:T})$ and $\widehat{F}_T^{2IP}(\cdot)$ merge in probability.
- (b) If in addition $\widehat{F}_T^{2IP}(\cdot)$ is stochastically uniformly equicontinuous, then

$$\mathbb{P}\left[I_\gamma^{2IP}(\mathbf{x}_{1:T}, \mathbf{Y}_{1:T}) \ni \psi_{T+1} \mid \mathcal{I}_{1:T}\right] \xrightarrow{P} 1 - \gamma. \tag{3.6}$$

2. (a) Under Assumption 1, 3 and 4, $F_T^{SPL}(\cdot | \mathcal{I}_{T_P:T})$ and $\widehat{F}_T^{SPL}(\cdot)$ merge in probability.
- (b) If in addition $\widehat{F}_T^{SPL}(\cdot)$ is stochastically uniformly equicontinuous, then

$$\mathbb{P}\left[I_\gamma^{SPL}(\mathbf{x}_{T_P:T}^c, \mathbf{X}_{1:T_E}) \ni \psi_{T+1} \mid \mathcal{I}_{T_P:T}\right] \xrightarrow{P} 1 - \gamma. \tag{3.7}$$

However, the standard approach, motivated by I_γ^{2IP} as in (3.6), computes an interval of the form $I_\gamma^{STA}(\mathbf{x}_{1:T}, \mathbf{x}_{1:T}) = I_\gamma^{2IP}(\mathbf{x}_{1:T}, \mathbf{x}_{1:T})$ as only one sample realization is available. This, of course, strongly violates the independence assumption of $\{X_t\}$ and $\{Y_t\}$. Specifically, replacing $\mathbf{Y}_{1:T}$ by $\mathbf{X}_{1:T}$ in equation (3.4), leads to

$$\begin{aligned} & I_\gamma^{STA*}(\mathbf{x}_{1:T}, \mathbf{X}_{1:T}) \\ &= \left[\hat{\psi}_{T+1}(\mathbf{x}_{1:T}, \mathbf{X}_{1:T}) - \frac{\widehat{F}_T^{STA* -1}(1 - \gamma_2)}{m_T}, \hat{\psi}_{T+1}(\mathbf{x}_{1:T}, \mathbf{X}_{1:T}) - \frac{\widehat{F}_T^{STA* -1}(\gamma_1)}{m_T} \right] \end{aligned}$$

$$(2.13) \left[\hat{\psi}_{T+1|1:T}^{STA*} - \frac{\widehat{F}_T^{STA^{-1}}(1 - \gamma_2)}{m_T}, \hat{\psi}_{T+1|1:T}^{STA*} - \frac{\widehat{F}_T^{STA^{-1}}(\gamma_1)}{m_T} \right], \tag{3.8}$$

where $\widehat{F}_T^{STA}(\cdot)$ is defined in relation (A.8) in Appendix A and the text preceding it. Whereas it is difficult to justify a conditional confidence interval like $I_\gamma^{STA}(\mathbf{x}_{1:T}, \mathbf{X}_{1:T})$ directly due to the lack of randomness, we can provide a justification by characterizing how closely the interval resembles the SPL interval. We establish the asymptotic equivalence, defined in terms of location and (scaled) length, between the two intervals in the following theorem. Note that, as our characterization of equivalence is probabilistic, we need to introduce the “doubly random” versions of the STA and SPL estimators, where the sample we condition on is considered random. These estimators are denoted as $\hat{\psi}_{T+1}(\mathbf{X}_{1:T}, \mathbf{X}_{1:T})$ and $\hat{\psi}_{T+1}(\mathbf{X}_{T_P:T}^c, \mathbf{X}_{1:T_E})$ respectively.

Theorem 3 (Asymptotic Equivalence Confidence Intervals).

1. (Location) If Assumptions 1–3 hold, then

$$\hat{\psi}_{T+1}(\mathbf{X}_{1:T}, \mathbf{X}_{1:T}) - \hat{\psi}_{T+1}(\mathbf{X}_{T_P:T}^c, \mathbf{X}_{1:T_E}) \xrightarrow{P} 0.$$

2. (Length) Under the assumptions of Theorem 1 and 2 and $\widehat{F}_T^{SPL^{-1}}(\cdot)$ being stochastically pointwise continuous at $u = \gamma_1, 1 - \gamma_2$, we have

$$\widehat{F}_T^{STA^{-1}}(u) - \widehat{F}_T^{SPL^{-1}}(u) \xrightarrow{P} 0. \tag{3.9}$$

The first implication states that the locations of the two intervals coincide asymptotically. The second statement establishes asymptotic closeness of the selected quantiles such that the scaled lengths of the intervals in (3.5) and (3.8) coincide asymptotically; recall that m_T/m_{T_E} converges to 1. As such, our sample-split interval coincides asymptotically with the standard interval, meaning that the standard interval can be substituted for an (asymptotically) equivalent interval which has a formal justification in terms of conditional coverage. As such, this provides a justification for the intervals commonly constructed in practice without having to rely on the two-independent-processes assumption.

3.4. Interval construction under normality

In this subsection we present intervals of reduced form and simplified theoretical results under asymptotic normality of the parameter estimator.

Assumption 5 (Normality). Let G_∞ be the cdf of the $N(0, \Upsilon_0)$ distribution with $\Upsilon_0 = \Upsilon(\theta_0, \xi_0)$ and assume there exist $\hat{\Upsilon}(\mathbf{X}_{1:T})$ converging in probability to Υ_0 .

Usually, the covariance estimator is obtained by inserting consistent estimators for θ_0 and ξ_0 into Υ_0 . Following the plug-in principle, we estimate

$F_T^{2IP}(\cdot|\mathcal{I}_{1:T})$ by a normal distribution with expected value 0 and variance

$$\hat{v}_T^{2IP} = \frac{\partial \psi_{T+1}(\mathbf{x}_{1:T}; \hat{\theta}(\mathbf{Y}_{1:T}))}{\partial \theta'} \hat{\Upsilon}(\mathbf{Y}_{1:T}) \frac{\partial \psi_{T+1}(x_{1:T}; \hat{\theta}(\mathbf{Y}_{1:T}))}{\partial \theta}$$

such that $\widehat{F}_T^{2IP}(\cdot) = \Phi(\cdot / \sqrt{\hat{v}_T^{2IP}})$. Then, the interval in (3.4) simplifies to

$$\begin{aligned} & I_\gamma^{2IP}(\mathbf{x}_{1:T}, \mathbf{Y}_{1:T}) \\ &= \left[\hat{\psi}_{T+1|1:T}^{2IP} - \frac{\sqrt{\hat{v}_T^{2IP}} \Phi^{-1}(1 - \gamma_2)}{m_T}, \hat{\psi}_{T+1|1:T}^{2IP} - \frac{\sqrt{\hat{v}_T^{2IP}} \Phi^{-1}(\gamma_1)}{m_T} \right]. \end{aligned} \quad (3.10)$$

Similarly, for the sample-split approach we consider $\widehat{F}_T^{SPL}(\cdot) = \Phi(\cdot / \sqrt{\hat{v}_T^{SPL}})$ with $\hat{v}_T^{SPL} = \frac{\partial \psi_{T+1}(\mathbf{x}_{T_P:T}^c; \hat{\theta}(\mathbf{X}_{1:T_E}))}{\partial \theta'} \hat{\Upsilon}(\mathbf{X}_{1:T_E}) \frac{\partial \psi_{T+1}(\mathbf{x}_{T_P:T}^c; \hat{\theta}(\mathbf{X}_{1:T_E}))}{\partial \theta'}$ such that (3.5) reduces to

$$\begin{aligned} & I_\gamma^{SPL}(\mathbf{x}_{T_P:T}, \mathbf{X}_{1:T_E}) \\ &= \left[\hat{\psi}_{T+1|T_P:T}^{SPL} - \frac{\sqrt{\hat{v}_T^{SPL}} \Phi^{-1}(1 - \gamma_2)}{m_{T_E}}, \hat{\psi}_{T+1|T_P:T}^{SPL} - \frac{\sqrt{\hat{v}_T^{SPL}} \Phi^{-1}(\gamma_1)}{m_{T_E}} \right]. \end{aligned} \quad (3.11)$$

In Appendix B we show that if the variance estimator is bounded away from zero in probability, e.g. $1/\hat{v}_T^{2IP} = O_p(1)$, then $\widehat{F}_T^{2IP}(\cdot)$ is stochastically uniform equicontinuous. Therefore, the asymptotic validity of both intervals can be deduced from Theorem 2.

Corollary 1 (Asymptotic Coverage under Normality).

1. (a) Under Assumption 1, 2 and 5, $F_T^{2IP}(\cdot|\mathcal{I}_{1:T})$ and $\Phi(\cdot / \sqrt{\hat{v}_T^{2IP}})$ merge in probability.
- (b) If in addition $1/\hat{v}_T^{2IP} = O_p(1)$, then

$$\mathbb{P} \left[I_\gamma^{2IP}(\mathbf{x}_{1:T}, \mathbf{Y}_{1:T}) \ni \psi_{T+1} \Big| \mathcal{I}_{1:T} \right] \xrightarrow{P} 1 - \gamma.$$

2. (a) Under Assumption 1, 3 and 5, $F_T^{SPL}(\cdot|\mathcal{I}_{T_P:T})$ and $\Phi(\cdot / \sqrt{\hat{v}_T^{SPL}})$ merge in probability.
- (b) If in addition $1/\hat{v}_T^{SPL} = O_p(1)$, then

$$\mathbb{P} \left[I_\gamma^{SPL}(\mathbf{x}_{T_P:T}^c, \mathbf{X}_{1:T_E}) \ni \psi_{T+1} \Big| \mathcal{I}_{T_P:T} \right] \xrightarrow{P} 1 - \gamma.$$

Bounding the variance estimator away from zero in probability to establish that the conditional coverage probability converges to $1 - \gamma$ in probability has intuitive appeal: as \hat{v}_T^{2IP} approaches zero, $N(0, \hat{v}_T^{2IP})$ becomes degenerate while the interval in (3.10) collapses (similar for SPL).

TABLE 1
AR models considered in the simulation study

DGP	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
A	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B	0.20	-0.50	0.40	0.40	0.00	0.00	0.00	0.00
C	1.20	-0.96	0.77	-0.61	0.49	-0.39	0.31	-0.25
D	0.80	-0.64	0.51	-0.41	0.33	-0.26	0.21	-0.17

For the standard interval, replacing $\mathbf{Y}_{1:T}$ by $\mathbf{X}_{1:T}$ in (3.10) leads to

$$I_\gamma^{STA*}(\mathbf{x}_{1:T}, \mathbf{X}_{1:T}) = \left[\hat{\psi}_{T+1|T_P:T}^{STA*} - \frac{\sqrt{\hat{v}_T^{STA} \Phi^{-1}(1-\gamma_2)}}{m_T}, \hat{\psi}_{T+1|T_P:T}^{STA*} - \frac{\sqrt{\hat{v}_T^{STA} \Phi^{-1}(\gamma_1)}}{m_T} \right] \quad (3.12)$$

with $\hat{v}_T^{STA} = \frac{\partial \psi_{T+1}(x_{1:T}; \hat{\theta}(\mathbf{X}_{1:T}))}{\partial \theta'} \hat{\Upsilon}(\mathbf{X}_n) \frac{\partial \psi_{T+1}(x_{1:T}; \hat{\theta}(\mathbf{X}_{1:T}))}{\partial \theta}$. In Appendix B we prove that \hat{v}_T^{SPL} is bounded in probability, which in turn implies that the quantile function $\widehat{F}_T^{SPL^{-1}}(\cdot) = \sqrt{\hat{v}_T^{SPL}} \Phi^{-1}(\cdot)$ is stochastically pointwise equicontinuous at any $u \in \mathbb{R}$. Hence, Theorem 3 applies. Whereas the first statement of the theorem remains unaffected, its second statement reads as follows under normality.

Corollary 2 (Length under Normality). *Under the assumptions of Theorem 1 and Corollary 1, we have $\sqrt{\hat{v}_T^{STA} \Phi^{-1}(u)} - \sqrt{\hat{v}_T^{SPL} \Phi^{-1}(u)} \xrightarrow{P} 0$ for $u = \gamma_1, 1-\gamma_2$.*

4. Simulation-based evaluation of sample-split intervals

In this section we consider the actual implementation of the sample-split intervals in practice, by comparing their performance to the standard interval (based on the 2IP justification) in a Monte Carlo simulation study.

4.1. Conditional coverage

We first investigate conditional coverage. We consider the DGP

$$x_t = \mu_t + \varepsilon_t, \quad \mu_t = \sum_{j=1}^p \beta_j x_{t-j}, \quad t = 1, \dots, T, \quad (4.1)$$

where we take $T = 50, 75, 100, 150, 200$. For the AR parameters we consider a large set of models, most which are taken from Ng and Perron [34]. Several of these models are truncated AR approximations to ARMA(1,1) models, see Section 3 in [34]. We consider a selection of those models here as given in Table 1 (full results are available on request).

Let $\hat{\beta}(\mathbf{x}_{1:T_E}) = (\hat{\beta}_1(\mathbf{x}_{1:T_E}), \dots, \hat{\beta}_p(\mathbf{x}_{1:T_E}))'$ denote the least squares estimate of (4.1) based on the sample $\mathbf{x}_{1:T_E}$. We then consider

$$\hat{\mu}_{T+1} = \sum_{j=1}^p \hat{\beta}_j(\mathbf{X}_{1:T_E}) x_{T-j+1} = \mathbf{x}'_{T_p:T} \hat{\beta}(\mathbf{x}_{1:T_E})$$

giving rise to the confidence interval

$$C_{T+1} = \left[\hat{\mu}_{T+1} - \frac{\sqrt{\mathbf{x}'_{T_p:T} \hat{\Upsilon}(\mathbf{x}_{1:T_E}) \mathbf{x}_{T_p:T} \Phi^{-1}(1 - \gamma/2)}}{\sqrt{T_E}}, \quad (4.2) \right.$$

$$\left. \hat{\mu}_{T+1} - \frac{\sqrt{\mathbf{x}'_{T_p:T} \hat{\Upsilon}(\mathbf{x}_{1:T_E}) \mathbf{x}_{T_p:T} \Phi^{-1}(\gamma/2)}}{\sqrt{T_E}} \right], \quad (4.3)$$

where $\hat{\Upsilon}(\mathbf{x}_{1:T_E})$ is the standard covariance matrix estimate of the least squares estimator.

Given the finite autoregressive order, we always take $T_P = T - p + 1$, but we vary T_E . For the standard interval we take $T_E = T$, while for the sample split intervals we vary T_E as $T_E = T - p - x\sqrt{T/50}$ for a grid of values $x = 0, 2, \dots, 20$. This setup allows us to evaluate the trade-off between using more observations for estimation (T_E) and creating a large enough gap between estimation and conditioning samples ($T_P - T_E$). We increase the gap for larger sample sizes by multiplying x with $\sqrt{T/50}$, where the factor 50 is simply chosen such that $\sqrt{T/50} = 1$ for our smallest sample size $T = 50$. The \sqrt{T} factor incorporates that with increasing sample size one automatically has more observations available for estimation, and thus the gap $T_P - T_E$ can be widened at reduced cost. While the square root is not guided by theory, we consider it as it implies a compromise between using the additional observations for either estimation or increasing the gap. Moreover, as we consider a grid of values for x anyway, the results are informative for more general choices as well.

To simulate conditional coverage, we simulate (4.1) backwards, fixing the final p observations by drawing them from the stationary distribution once, and then keeping these values over all 10,000 simulations. As we take $\varepsilon_t \sim i.i.d. N(0, 1)$, backwards simulation yields the same process as standard forward simulation. To investigate the sensitivity to the values of final p observations, we repeat the simulations 10 times, thus in total considering $10 \times 10,000$ simulations per DGP. We take $\gamma = 0.05$.

Figure 1 reports the mean coverage, averaged over the 10 sets of simulations. On the horizontal axis we report the different intervals. The standard interval is denoted as STA while the SPL methods for the grid of x are denoted as SPLx. Overall we can see that for low order AR models, as expected the standard method performs well. The split method becomes less accurate when increasing $T_p - T_E$, presumably because using fewer observations decreases the accuracy of the asymptotic Gaussian approximation. Indeed, the deterioration in coverage

is mostly visible for the small sample sizes. However, for AR models of higher order, where the “overlap” between estimation and prediction samples becomes larger, the SPL intervals outperform the standard interval. The performance is still sensitive to the split point, with small x the preferred choice, in particular for small sample sizes.

Figure 2 goes into more detail about the role of the observations conditioned on by displaying the median as well as minimum and maximum coverage over the 10 sets of simulated conditioned observations. While in general the patterns agree with the graphs for the mean coverage, these figures uncover some interesting insights. We see that in particular for higher order DGPs, the conditional coverage of the standard interval can be off by quite a large amount, dropping to as far as 0.4 for DGP D. While this only happens for an occasional value of $\mathbf{x}_{T_P:T}$, this is a remarkable result in the light of conditional coverage. In addition, the SPL intervals do not suffer from the same flaw for any value of x , making them considerably more reliable.

Finally, we investigate the average length of the intervals in Figures 3.³ As would be expected, the smaller T_E , the wider the interval. The increase is rather moderate however. Moreover, there are a few instances where SPL with $x = 0$ actually gives shorter intervals than the standard interval.

Concluding, we find that the sample split intervals are more reliable than the standard interval, in particular when conditional coverage is concerned, provided T_E is kept large (such that $T_P - T_E$ is small). As moreover the split intervals are not much wider than the standard intervals, if at all, their use in practice is recommended.

4.2. Overall coverage

In this section we report results obtained from simulating (4.1) forwards to estimate

$$\begin{aligned} \mathbb{P}(C_{T+1} \ni \mu_{T+1}) &= \int \mathbb{P}(C_{T+1} \ni \mu_{T+1} | X_t = x_T, \dots, X_{T-p+1} = x_{T-p+1}) \\ &\quad \times f^{X_T, \dots, X_{T-p+1}}(x_T, \dots, x_p) dx_T \dots dx_{T-p+1}, \end{aligned}$$

where μ_{T+1} and C_{T+1} are as in (4.1) and (4.2), respectively, and $f^{X_T, \dots, X_{T-p+1}}$ denotes the joint density of X_T, \dots, X_{T-p+1} . When estimating this quantity we are not bound to consider normally distributed errors as in the previous section. Therefore, in our simulations we used gamma distributed errors shifted such that their expectations equal zero. Additionally, to see if the results are sensitive to the distribution of the error terms we ran some simulations with errors that follow a mixture distribution. Moreover, to make sure that the results are not driven by the estimation method we estimated the parameters for the results of this section by MLE. All results are based on 100,000 simulations and the level

³Medians, minima and maxima show the same patterns as the means, and are therefore not shown here.

always equals 95%. As in the previous subsection we always took $T_P = T - p + 1$. For an AR(1) the coverage is even in small samples for both *STA* and sample-split close to the level with a mild dependency on the AR parameter; see Table 2 in Appendix C. On average, for $T = 50$ as well as $T_E = 40$ and $T_E = 45$ the *STA* intervals perform slightly better with the advantage of *STA* being smaller for $T_E = 45$. These conclusions also hold for AR(2) processes; see Table 3 in Appendix C. The picture changes already for an AR(4). For all combinations of T and T_E with $T - T_E$ ranging between 5, 10 and 15 the sample-split intervals perform better. Clearly, their advantage gets smaller as $T - T_E$ increases; see Table 4 in Appendix C. The same pattern but with a larger numerical advantage in terms of coverage for SPL intervals is observed in Tables 5 and 6 (also in Appendix C) for the AR(8) processes from Table 1. Presumably, the better performance of the SPL intervals for the AR(4) and the AR(8) processes is a result of multiplying the estimators of the AR coefficients which depend on the entire sample by the most recent observations which may move the true finite sample distribution away from the approximating normal distribution of the AR coefficient estimators. This is not the case for the sample split approach so that the normal approximation may work better although we use a smaller sample to estimate.

5. Prediction intervals

The preceding sections have focused purely on the construction of conditional confidence intervals to account for parameter uncertainty. Regarding prediction, a second source of uncertainty arises, that corresponds to the model's innovation process. In this setting, parameter estimation is typically disregarded in textbooks as the stochastic fluctuation stemming from the estimation procedure is generally dominated by the stochastic fluctuation of the innovations. Although the resulting prediction intervals may be asymptotically valid, they are typically characterized by under-coverage in finite samples. In response, [35] introduce the concept of *asymptotic pertinence* to evaluate distribution approximations that account for the two sources of randomness, innovation and parameter estimation uncertainty, according to their general orders of magnitude. Whereas [30] and [41] study properties of the *unconditional* law of the forecast error, we focus on its *conditional* distribution to conform with the relevance property of [25]. The fundamental issue also arises when considering prediction if one attempts to account for parameter uncertainty. To illustrate this point, we revisit the introductory examples and write $*$ to denote the convolution operator.⁴

Example 1. (*continued*) Prediction intervals for the AR are often constructed around the point estimate for the conditional mean. The conditional distribution

⁴For independent variables X and Y with $X \sim F_X$, $Y \sim F_Y$ and $Z = X + Y \sim F_Z$, we write $F_Z = F_X * F_Y$ to denote $F_Z(z) = \int_{-\infty}^z F_X(z - y) dF_Y(y)$.

of the forecast error decomposes into

$$\begin{aligned} \mathbb{P}[X_{T+1} - \hat{\beta}(\mathbf{X}_{1:T})X_T \leq \cdot | X_T = x_T] &= \mathbb{P}[\beta X_T - \hat{\beta}(\mathbf{X}_{1:T})X_T \leq \cdot | X_T = x_T] \\ &\quad * \mathbb{P}[\varepsilon_{T+1} \leq \cdot], \end{aligned} \tag{5.1}$$

corresponding to estimation and innovation uncertainty, respectively. As argued above, an approximation of $\mathbb{P}[\beta X_T - \hat{\beta}(\mathbf{X}_{1:T})X_T \leq \cdot | X_T = x_T]$ appears rather cumbersome. In the special case of $\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$, [40, Theorem 3] gives an approximation for (5.1) based on Edgeworth expansions.

Example 2. (*continued*) Suppose we are interested in providing a prediction interval for X_{T+1}^2 in the GARCH(1,1). Conditioning on $\mathbf{X}_{1:T} = \mathbf{x}_{1:T}$, a natural estimate of X_{T+1}^2 is $\hat{\sigma}_{T+1|1:T}^{2,STA}$ as defined in (2.17), since $\sigma_{T+1|1:T}^2$ is its expected value given information up to time T . As

$$\begin{aligned} \mathbb{P}[X_{T+1}^2 - \hat{\sigma}_{T+1}^{2,STA*} \leq \cdot | \mathbf{X}_{1:T} = \mathbf{x}_{1:T}] &= \mathbb{P}[\sigma_{T+1}^2 - \hat{\sigma}_{T+1}^{2,STA*} \leq \cdot | \mathbf{X}_{1:T} = \mathbf{x}_{1:T}] \\ &\quad * \mathbb{P}[\sigma_{T+1|1:T}^2(\varepsilon_{T+1}^2 - 1) \leq \cdot], \end{aligned} \tag{5.2}$$

where $\hat{\sigma}_{T+1}^{2,STA*}$ is defined in (2.18), the desired prediction interval, say J_γ^{STA} , leads to a sensible probabilistic statement due to variability in ε_{T+1}^2 :

$$\mathbb{P}\left[X_{T+1}^2 \in J_\gamma^{STA}(\mathbf{X}_{1:T}, \mathbf{X}_{1:T}) \mid \mathbf{X}_{1:T} = \mathbf{x}_{1:T}\right] \underset{(\approx)}{=} 1 - \gamma. \tag{5.3}$$

However, it cannot incorporate parameter uncertainty either, since the conditional distribution $\mathbb{P}[\sigma_{T+1}^2 - \hat{\sigma}_{T+1}^{2,STA*} \leq \cdot | \mathbf{X}_{1:T} = \mathbf{x}_{1:T}] = \mathbb{P}[\sigma_{T+1|1:T}^2 - \hat{\sigma}_{T+1|1:T}^{2,STA} \leq \cdot]$ is degenerate.

In his textbook [39] resorts to a Bayesian-akin approach to avoid the fundamental issue in forecasting. He argues that θ , although “fixed at the estimation stage ... is viewed best as a random variable at the forecasting stage” (p. 389). Consequently, he assigns some posterior distribution to θ motivated by an uninformed prior. Treating θ not fixed but random, the fundamental issue does not arise, however combining a frequentist view with a Bayesian method does not seem to be coherent.

[2] require the existence of a transitive statistic $U = U(\mathbf{X}_{1:T})$ of fixed low dimension to establish conditional independence between the sample $\mathbf{X}_{1:T}$ and their considered future random variable given $U = u$. [43], [44], [45]), [25], [26], [27], and [28]) extend their approach and derive improved prediction intervals. Although these methods absorb an additional $O(T^{-1})$ term in the associated conditional coverage probability, there are several drawbacks associated with them: the innovation distribution needs typically be specified (e.g. Gaussian), the results apply only to a limited set of estimators (e.g. maximum likelihood) and their framework can only incorporate finite autoregressive components (e.g. AR(p)).

Assuming two independent processes with the same stochastic structure, using one for prediction and one for the estimation of the parameters, alleviates

the fundamental issue faced in the continued Examples 1 and 2. As the conditional distributions of the 2IP and SPL estimators merge in probability by Theorem 1, the 2IP assumption can be avoided by following a sample-split approach as described in Section 2.3.

6. Conclusion

In the paper at hand, we study the construction of confidence intervals for conditional objects such as conditional means or conditional variances, focusing on the conceptual issue that arises in the process of taking parameter uncertainty into account. It stems from the fact that on one hand one must condition on the sample as the past informs about the present and future, yet on the other hand one must allow the data up to now to be treated as random to account for estimation uncertainty. Assuming two independent processes with the same stochastic structure, where one is used for conditioning and one for the estimation of the parameters, bypasses this issue, but the assumption itself can generally not be justified in applications. To avoid this assumption, we propose a solution based on a simple sample-split approach, that requires a much more realistic weak dependence condition instead. To acknowledge that the conditional quantities vary over time, we employ a merging concept generalizing the notion of weak convergence. The conditional distributions of the sample-split estimator and the estimator based on the two-independent-processes assumption are shown to merge in probability under mild conditions. The corresponding sample-split intervals are shown to coincide asymptotically with the intervals commonly constructed by practitioners, which provides a novel and theoretically satisfactory justification for commonly constructed confidence intervals for conditional objects, applicable to a wide class of time series models, including ARMA and GARCH-type models. Next to the theoretical justification sample-split intervals provide for the commonly used intervals, the simulation results in Section 4 suggest that sample-split intervals can also be superior in finite samples.

One limitation to our approach is that we restrict ourselves to univariate time series and objects of interests. At the expense of more involved notation this could be readily extended to multivariate time series and objects of interests. A second, and more restrictive, limitation is our weak dependence assumption needed to achieve asymptotic independence between the two subsamples, which for instance rules out application to integrated processes. Given the fundamental role of this assumption in our setup, it appears difficult to generalize this. However, this also casts further doubt on the two-independent-processes assumption as validation for confidence intervals constructed for such persistent processes. A case-by-case treatment, as for instance done by [20] for near unit root processes and [41] for explosive processes, appears to be necessary in such cases, and standard confidence intervals should be treated with caution.

Appendix A: Main asymptotic results

We only present the proofs of the leading results here. The proofs of the lemmas and corollaries can be found in Appendix B. Before going to the proofs, we first introduce the following auxiliary metrics that will be encountered in the proofs. For arbitrary cdfs F and G on \mathbb{R} , the Kolmogorov and Lévy metric are

$$\begin{aligned} d_K(F, G) &= \sup_{\tau \in \mathbb{R}} |F(\tau) - G(\tau)| \\ d_L(F, G) &= \inf \{ \xi > 0 : G(\tau - \xi) - \xi \leq F(\tau) \leq G(\tau + \xi) + \xi \quad \forall \tau \in \mathbb{R} \}. \end{aligned} \quad (\text{A.1})$$

Moreover, let $Z_\infty \sim G_\infty$ (with G_∞ given in Assumption 1.a) be defined on some probability space $(\check{\Omega}, \check{\mathcal{F}}, \check{\mathbb{P}})$ and define the product measure $\check{\mathbb{P}} = \mathbb{P} \times \check{\mathbb{P}}$ on the space $\Omega \times \check{\Omega}$ with σ -field, generated by the *measurable rectangles* $A \times \check{A}$ with $A \in \mathcal{F}$ and $\check{A} \in \check{\mathcal{F}}$ (cf. [7, Theorem 18.2]). Notice that Z_∞ is independent of $\{X_t\}$ and $\{Y_t\}$ by construction.

In the proofs we often consider the “doubly random” versions of estimators and intervals, where the first function argument, the sample to condition on, is considered random. Instead, we account for the conditioning in the probability statements. This notation is more convenient for proving the results, as we need them to hold for “all sequences $\mathbf{x}_{1:T}$ occurring with high probability”, which is much easier to quantify by treating these sequences as random. Therefore we use notations such as the unconditional estimators $\hat{\psi}_{T+1}^{2IP} = \hat{\psi}_{T+1}(\mathbf{X}_{1:T}, \mathbf{Y}_{1:T})$ and $\hat{\psi}_{T+1}^{SPL} = \hat{\psi}_{T+1}(\mathbf{X}_{T_P:T}^c, \mathbf{X}_{1:T_E})$, as well as their corresponding intervals $I^{2IP}(\mathbf{X}_{1:T}, \mathbf{Y}_{1:T})$ and $I^{SPL}(\mathbf{X}_{T_P:T}^c, \mathbf{X}_{1:T_E})$; also see (3.2)–(3.3) and the remarks above Theorem 3.

A.1. Lemmas

Lemma 1. *Let*

$$R_T^{2IP} := m_T \left(\hat{\psi}_{T+1}(\mathbf{X}_{1:T}, \mathbf{Y}_{1:T}) - \psi_{T+1} \right) - \frac{\partial \psi_{T+1}^s(\mathbf{X}_{1:T}, \theta_0)}{\partial \theta'} m_T (\hat{\theta}(\mathbf{Y}_{1:T}) - \theta_0) \quad (\text{A.2})$$

$$\begin{aligned} R_T^{SPL} &:= m_{T_E} \left(\hat{\psi}_{T+1}(\mathbf{X}_{T_P:T}^c, \mathbf{X}_{1:T_E}) - \psi_{T+1} \right) \\ &\quad - \frac{\partial \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c, \theta_0)}{\partial \theta'} m_{T_E} (\hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0). \end{aligned} \quad (\text{A.3})$$

1. If Assumptions 1.a, 1.b, 1.d, 1.e and 2.a hold, then R_T^{2IP} is $o_p(1)$;
2. if Assumptions 1.a, 1.b, 1.d, 1.e and 3.a hold, then R_T^{SPL} is $o_p(1)$.

Lemma 2. *Let $G_{T_E}^{SPL}(\cdot | \mathcal{I}_{T_P:T})$ be as given in Assumption 3.c and denote the conditional cdf of $Z_T^{2IP} := m_T (\hat{\theta}(\mathbf{Y}_{1:T}) - \theta_0)$ given $\mathcal{I}_{1:T}$ by $G_T^{2IP}(\cdot | \mathcal{I}_{1:T})$.*

1. Under Assumptions 1.a and 2, $\int h dG_T^{2IP}(\cdot | \mathcal{I}_{1:T}) \xrightarrow{P} \int h dG_\infty \quad \forall h \in BL$;
2. Under Assumptions 1.a and 3, $\int h dG_{T_E}^{SPL}(\cdot | \mathcal{I}_{T_P:T}) \xrightarrow{P} \int h dG_\infty \quad \forall h \in BL$.

Lemma 3. Let $\{G_T\}$ be a sequence of cdfs and G be a (non-random) cdf on $(\mathbb{R}^r, \|\cdot\|)$. If $\int h dG_T \xrightarrow{P} \int h dG$ for all $h \in BL$, then $\sup_{h \in \mathcal{H}} \left| \int h d(G_T - G) \right| \xrightarrow{P} 0$, where $\mathcal{H} = \{h : \mathbb{R}^r \rightarrow \mathbb{R} : \|h\|_{BL} \leq 1\}$.

Lemma 4. Assume that the \mathbb{R}^r -valued random variable w_T is $O_p(1)$ and \mathcal{I}_T -measurable. Further, suppose the real-valued random variable R_T is $o_p(1)$ and the \mathbb{R}^r -valued random variable Z_T satisfies $\mathbb{P}[Z_T \leq \cdot | \mathcal{I}_T] \xrightarrow{P} G_\infty$. Then, the two sequences of conditional cdfs $\mathbb{P}[w'_T Z_T + R_T \leq \cdot | \mathcal{I}_T]$ and $\mathbb{P}[w'_T Z_\infty \leq \cdot | \mathcal{I}_T]$ merge in probability.

Lemma 5. Let $\epsilon > 0$ and F and G be cdfs on \mathbb{R} with $G(\tau - \epsilon) - \epsilon \leq F(\tau) \leq G(\tau + \epsilon) + \epsilon$ for all $\tau \in \mathbb{R}$. Then $F^{-1}(u - \epsilon) - \epsilon \leq G^{-1}(u) \leq F^{-1}(u + \epsilon) + \epsilon$ for all $u \in (\epsilon, 1 - \epsilon)$.

Lemma 6. Suppose $\{F_T\}$ and $\{G_T\}$ are sequences of conditional cdfs with $d_L(F_T, G_T) \xrightarrow{P} 0$ as $T \rightarrow \infty$. Further, assume that G_T is stochastically uniformly equicontinuous: for every $\epsilon, \eta > 0$, there exist $\delta = \delta(\epsilon, \eta) > 0$ and $\bar{T} = \bar{T}(\epsilon, \eta)$ such that $\mathbb{P}\left[\sup_{\tau \in \mathbb{R}} \sup_{\tau' \in \mathbb{R}: |\tau - \tau'| < \delta} |G_T(\tau') - G_T(\tau)| > \epsilon\right] < \eta$ for all $T \geq \bar{T}$. Then, $d_K(F_T, G_T) \xrightarrow{P} 0$.

Lemma 7. If the sequences of conditional cdfs $\{F_T\}$ and $\{G_T\}$ merge in probability and G_T is stochastically uniformly equicontinuous, then $\mathbb{P}[G_T^{-1}(\gamma_1) \leq M_T \leq G_T^{-1}(1 - \gamma_2) | \mathcal{I}_T] \xrightarrow{P} 1 - \gamma_1 - \gamma_2$ whenever $0 \leq \gamma_1 \leq 1 - \gamma_2 \leq 1$, where the random variable M_T given the σ -algebra \mathcal{I}_T has the cdf F_T .

Lemma 8. Suppose $\{F_T\}$ and $\{G_T\}$ are sequences of conditional cdfs with $d_L(F_T, G_T) \xrightarrow{P} 0$ as $T \rightarrow \infty$. Further, assume that G_T^{-1} is stochastically point-wise equicontinuous at $u \in (0, 1)$: for all $\epsilon, \eta > 0$, there exist $\delta = \delta(\epsilon, \eta, u) > 0$ and $\bar{T} = \bar{T}(\epsilon, \eta, u)$ such that $\mathbb{P}\left[\sup_{|u-v| < \delta} |G_T^{-1}(v) - G_T^{-1}(u)| > \epsilon\right] < \eta$ for all $T \geq \bar{T}$. Then $|F_T^{-1}(u) - G_T^{-1}(u)| \xrightarrow{P} 0$.

A.2. Proofs of theorems

Proof of Theorem 1. According to our Definition 5 we need to show that

$$d_{BL}\left(F_T^{2IP}(\cdot | \mathcal{I}_{1:T}), F_T^{SPL}(\cdot | \mathcal{I}_{T_P:T})\right) \xrightarrow{P} 0.$$

In order to so, we now first introduce some auxiliary conditional distribution functions. Let w_T^{2IP} equal the transpose of $\frac{\partial \psi_{T+1}^s(\mathbf{X}_{1:T}; \theta_0)}{\partial \theta}$ and w_T^{SPL} the transpose of $\frac{\partial \psi_{T+1}^c(\mathbf{X}_{T_P:T}^c; \theta_0)}{\partial \theta}$ and set with Z_∞ as defined just below Equation (A.1)

$$F_{\infty,T}^{2IP}(\tau | \mathcal{I}_{1:T}) := \bar{\mathbb{P}}[w_T^{2IP} Z_\infty \leq \tau | \mathcal{I}_{1:T}] \tag{A.4}$$

$$F_{\infty,T}^{SPL}(\tau | \mathcal{I}_{T_P:T}) := \bar{\mathbb{P}}[w_T^{SPL} Z_\infty \leq \tau | \mathcal{I}_{T_P:T}], \tag{A.5}$$

the conditional cdfs of $w_T^{2IP}Z_\infty$ given $\mathcal{I}_{1:T}$ and $w_T^{SPL}Z_\infty$ given $\mathcal{I}_{T_P:T}$, respectively. Note that (A.4) and (A.5) can be considered to be the ‘conditional merging limits’ of $w_T^{2IP}m_T(\hat{\theta}(\mathbf{X}_{1:T}) - \theta_0)$ and $w_T^{SPL}m_{T_E}(\hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0)$ given $\mathcal{I}_{1:T}$ and $\mathcal{I}_{T_P:T}$ respectively, because the distributions of $m_T(\hat{\theta}(\mathbf{X}_{1:T}) - \theta_0)$ and $m_{T_E}(\hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0)$ will converge to the distribution of Z_∞ which has the cumulative distribution function G_∞ . On the other hand, w_T^{2IP} and w_T^{SPL} do not converge. We therefore indexed the ‘merging limits’ in (A.4) and (A.5) by ∞ and T where the former subscript should remind the reader of the fixed component, i.e. Z_∞ , of the ‘merging limit’ and the latter subscript of the components that do not converge at all. The triangle inequality implies

$$\begin{aligned} d_{BL}\left(F_T^{2IP}(\cdot|\mathcal{I}_{1:T}), F_T^{SPL}(\cdot|\mathcal{I}_{T_P:T})\right) &\leq \underbrace{d_{BL}\left(F_T^{2IP}(\cdot|\mathcal{I}_{1:T}), F_{\infty,T}^{2IP}(\cdot|\mathcal{I}_{1:T})\right)}_I \\ &\quad + \underbrace{d_{BL}\left(F_{\infty,T}^{2IP}(\cdot|\mathcal{I}_{1:T}), F_{\infty,T}^{SPL}(\cdot|\mathcal{I}_{T_P:T})\right)}_{II} \\ &\quad + \underbrace{d_{BL}\left(F_{\infty,T}^{SPL}(\cdot|\mathcal{I}_{T_P:T}), F_T^{SPL}(\cdot|\mathcal{I}_{T_P:T})\right)}_{III}. \end{aligned}$$

We now prove that I , II and III converge to zero in probability. To do so we will make use of the auxiliary lemmas given above. More precisely, for each term we will verify that the assumptions of Lemma 4 hold which will either be seen to be a consequence of our assumptions or can be proven to hold by showing that the conditions of Lemma 1 and Lemma 2 are met.

We start with I and note that $m_T(\hat{\psi}_{T+1}^{2IP} - \psi_{T+1}) = w_T^{2IP}Z_T^{2IP} + R_T^{2IP}$, where $Z_T^{2IP} = m_T(\hat{\theta}(\mathbf{Y}_{1:T}) - \theta_0)$ and where R_T^{2IP} was defined in (A.2). The weight w_T^{2IP} is $\mathcal{I}_{1:T}$ measurable and $O_p(1)$ by Assumptions 1.c and 1.e, R_T^{2IP} is $o_p(1)$ by Lemma 1 and $\int h dG_T^{2IP}(\cdot|\mathcal{I}_{1:T}) \xrightarrow{P} \int h dG_\infty$ for each $h \in BL$ by Lemma 2. Replacing Z_T , R_T , w_T and \mathcal{I}_T in Lemma 4 by Z_T^{2IP} , R_T^{2IP} , w_T^{2IP} and $\mathcal{I}_{1:T}$ implies that $F_T^{2IP}(\tau|\mathcal{I}_{1:T}) = \mathbb{P}[w_T^{2IP}Z_T^{2IP} + R_T^{2IP} \leq \tau|\mathcal{I}_{1:T}]$ and $F_{\infty,T}^{2IP}(\tau|\mathcal{I}_{1:T}) = \mathbb{P}[w_T^{2IP}Z_\infty \leq \tau|\mathcal{I}_{1:T}]$ merge in probability, i.e. $I \xrightarrow{P} 0$.

Consider II and rewrite $w_T^{SPL}Z_\infty = w_T^{2IP}Z_\infty + (w_T^{SPL} - w_T^{2IP})Z_\infty$. The weight w_T^{2IP} is $\mathcal{I}_{1:T}$ measurable and $O_p(1)$ by Assumptions 1.c and 1.e, $(w_T^{SPL} - w_T^{2IP})Z_\infty = o_p(1)O_p(1) = o_p(1)$ by Assumption 1.e and $\int h d\mathbb{P}[Z_\infty \leq \cdot|\mathcal{I}_{1:T}] = \int h dG_\infty$ for each $h \in BL$. Replacing Z_T , R_T , w_T and \mathcal{I}_T in Lemma 4 by Z_∞ , $(w_T^{SPL} - w_T^{2IP})Z_\infty$, w_T^{2IP} and $\mathcal{I}_{1:T}$ implies that $F_{\infty,T}^{2IP}(\tau|\mathcal{I}_{1:T}) = \mathbb{P}[w_T^{2IP}Z_\infty \leq \tau|\mathcal{I}_{1:T}]$ and $F_{\infty,T}^{SPL}(\tau|\mathcal{I}_{T_P:T}) = \mathbb{P}[w_T^{SPL}Z_\infty \leq \tau|\mathcal{I}_{1:T}]$ merge in probability, i.e. $II \xrightarrow{P} 0$.

Consider III and note that $m_{T_E}(\hat{\psi}_{T+1}^{SPL} - \psi_{T+1}) = w_T^{SPL}Z_T^{SPL} + R_T^{SPL} + S_T^{SPL}$, where $Z_T^{SPL} = m_{T_E}(\hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0)$, and R_T^{SPL} is as defined in (A.3). The weight w_T^{SPL} is $\mathcal{I}_{T_P:T}$ measurable and $O_p(1)$ by Assumptions 1.c and 1.e, R_T^{SPL} is $o_p(1)$ by Lemma 1 and $\int h dG_{T_E}^{SPL}(\cdot|\mathcal{I}_{T_P:T}) \xrightarrow{P} \int h dG_\infty$ for each $h \in BL$ by Lemma 2. Replacing Z_T , R_T , w_T and \mathcal{I}_T in Lemma 4 by Z_T^{SPL} , R_T^{SPL} ,

w_T^{SPL} and $\mathcal{I}_{T_P:T}$ implies that $F_T^{SPL}(\tau|\mathcal{I}_{T_P:T}) = \bar{\mathbb{P}}[w_T^{SPL}Z_T^{SPL} + R_T^{SPL} + S_T^{SPL} \leq \tau|\mathcal{I}_{T_P:T}]$ and $F_{\infty,T}^{SPL}(\tau|\mathcal{I}_{T_P:T}) = \mathbb{P}[w_T^{SPL}Z_\infty \leq \tau|\mathcal{I}_{T_P:T}]$ merge in probability, i.e. $III \xrightarrow{p} 0$. □

Proof of Theorem 2. Consider statement 1(a). This time according to our Definition 5 we have to show that

$$d_{BL}\left(F_T^{2IP}(\cdot|\mathcal{I}_{1:T}), \widehat{F_T^{2IP}}(\cdot)\right) \xrightarrow{\mathbb{P}} 0.$$

For this, as mentioned a few lines above Equation (3.4), we now first specify $\widehat{F_T^{2IP}}(\cdot)$. For this specification it is convenient to introduce the random variable \hat{Z}_T^{2IP} whose distribution given the sample approximates the (unknown) distribution of Z_T^{2IP} in the proof of Theorem 1. That is, \hat{Z}_T^{2IP} follows the mixture distribution $\widehat{G}_T(\cdot)$ as a function of $\mathbf{Y}_{1:T}$ such that given $\mathbf{Y}_{1:T}$ the conditional distribution of the random variable \hat{Z}_T^{2IP} is $\widehat{G}_T(\cdot)$. With the help of \hat{Z}_T^{2IP} we now let

$$\widehat{F_T^{2IP}}(\cdot) \text{ be the conditional cdf of } \hat{w}_T^{2IP} \hat{Z}_T^{2IP} \text{ given } \mathcal{H}_{1:T} \tag{A.6}$$

where \hat{w}_T^{2IP} equals the transpose of $\frac{\partial \psi_T^s(\mathbf{X}_{1:T}; \hat{\theta}(\mathbf{Y}_{1:T}))}{\partial \theta}$ and $\mathcal{H}_{1:T} = \sigma(X_1, \dots, X_T, Y_1, \dots, Y_T)$. Then

$$\begin{aligned} d_{BL}\left(F_T^{2IP}(\cdot|\mathcal{I}_{1:T}), \widehat{F_T^{2IP}}(\cdot)\right) &\leq d_{BL}\left(F_T^{2IP}(\cdot|\mathcal{I}_{1:T}), F_{\infty,T}^{2IP}(\cdot|\mathcal{I}_{1:T})\right) \\ &\quad + d_{BL}\left(F_{\infty,T}^{2IP}(\cdot|\mathcal{I}_{1:T}), \widehat{F_T^{2IP}}(\cdot)\right) \end{aligned}$$

by the triangle inequality, where $F_{\infty,T}^{2IP}(\cdot|\mathcal{I}_{1:T})$ is defined in equation (A.4). In the proof of Theorem 1, we have shown that $F_T^{2IP}(\cdot|\mathcal{I}_{1:T})$ and $F_{\infty,T}^{2IP}(\cdot|\mathcal{I}_{1:T})$ merge in probability under Assumptions 1 and 2. It suffices to prove that $F_{\infty,T}^{2IP}(\cdot|\mathcal{I}_{1:T})$ and $\widehat{F_T^{2IP}}(\cdot)$ merge in probability which will again be shown by verifying the assumptions of Lemma 4. Write $\hat{w}_T^{2IP} \hat{Z}_T^{2IP} = w_T^{2IP} \hat{Z}_T^{2IP} + \hat{R}_T^{2IP}$ with $\hat{R}_T^{2IP} = (\hat{w}_T^{2IP} - w_T^{2IP}) \hat{Z}_T^{2IP}$. First, we show $\hat{R}_T^{2IP} = o_p(1)$. Take an arbitrary $\epsilon > 0$. We obtain

$$\begin{aligned} \mathbb{P}\left[|\hat{R}_T^{2IP}| \geq \epsilon\right] &\leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{1:T}; \dot{\theta}_T)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{Y}_{1:T}) - \theta_0\right\| \left\|\hat{Z}_T^{2IP}\right\| \geq \epsilon\right] \\ &\leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{1:T}; \dot{\theta}_T)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{Y}_{1:T}) - \theta_0\right\| \left\|\hat{Z}_T^{2IP}\right\| \geq \epsilon \cap \dot{\theta}_T \in \mathcal{V}(\theta_0)\right] \\ &\quad + \mathbb{P}\left[\dot{\theta}_T \notin \mathcal{V}(\theta_0)\right] \\ &\leq \mathbb{P}\left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left\|\frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{1:T}; \theta)}{\partial \theta \partial \theta'}\right\| \left\|\hat{\theta}(\mathbf{Y}_{1:T}) - \theta_0\right\| \left\|\hat{Z}_T^{2IP}\right\| \geq \epsilon\right] \\ &\quad + \mathbb{P}\left[\dot{\theta}_T \notin \mathcal{V}(\theta_0)\right], \end{aligned}$$

where $\hat{\theta}_T$ lies between $\hat{\theta}(\mathbf{Y}_{1:T})$ and θ_0 . The first term vanishes as $\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi_{T+1}(\mathbf{X}_{1:T}; \theta)}{\partial \theta \partial \theta'} \right\| = O_p(1)$ by Assumptions 1.d and 1.e, $\|\hat{\theta}(\mathbf{Y}_{1:T}) - \theta_0\| = O_p(m_T^{-1})$ by Assumptions 1.a and 2.a and $\|\hat{Z}_T^{2IP}\| = O_p(1)$ as $\hat{Z}_T^{2IP} \sim \hat{G}_T(\cdot)$ and $\int h d\hat{G}_T \xrightarrow{p} \int h dG_\infty$ for all $h \in BL$ by Assumptions 2.a and 4. Further, as $\hat{\theta}(\mathbf{Y}_{1:T}) \xrightarrow{p} \theta_0 \in \mathcal{V}(\theta_0)$ and $\mathcal{V}(\theta_0)$ is open, we have $\mathbb{P}[\hat{\theta}_T \notin \mathcal{V}(\theta_0)] \rightarrow 0$ and $\hat{R}_T^{2IP} = o_p(1)$ follows. Moreover, w_T^{2IP} is \mathcal{H}_T -measurable and $O_p(1)$ by Assumptions 1.c and 1.e and $\int h d\bar{\mathbb{P}}[\hat{Z}_T^{2IP} \leq \cdot | \mathcal{H}_{1:T}] = \int h d\hat{G}_T \xrightarrow{p} \int h dG_\infty$ for each $h \in BL$. Replacing Z_T, R_T, w_T and \mathcal{I}_T in Lemma 4 by $\hat{Z}_T^{2IP}, \hat{R}_T^{2IP}, w_T^{2IP}$ and \mathcal{H}_T implies that $\widehat{F}_T^{2IP}(\cdot) = \bar{\mathbb{P}}[w_T^{2IP} \hat{Z}_T^{2IP} + \hat{R}_T^{2IP} \leq \cdot | \mathcal{H}_{1:T}]$ and $\bar{\mathbb{P}}[w_T^{2IP} Z_\infty \leq \cdot | \mathcal{H}_{1:T}] = F_{\infty, T}^{2IP}(\cdot | \mathcal{I}_{1:T})$ merge in probability.

Consider statement 1(b). As $F_T^{2IP}(\cdot | \mathcal{I}_{1:T})$ and $\widehat{F}_T^{2IP}(\cdot)$ merge in probability and $\widehat{F}_T^{2IP}(\cdot)$ is assumed to be stochastically uniformly continuous, Lemma 7 applies. Replacing F_T, G_T, M_T and \mathcal{I}_T by $F_T^{2IP}(\cdot | \mathcal{I}_{1:T}), \widehat{F}_T^{2IP}(\cdot), m_T(\hat{\psi}_{T+1}^{2IP} - \psi_{T+1})$ and $\mathcal{I}_{1:T}$, respectively, it follows that

$$\begin{aligned} & \mathbb{P}\left[I_\gamma^{2IP}(\mathbf{X}_{1:T}, \mathbf{Y}_{1:T}) \ni \psi_{T+1} \mid \mathcal{I}_{1:T} \right] \\ &= \mathbb{P}\left[\widehat{F}_T^{2IP}{}^{-1}(\gamma_1) \leq m_T(\hat{\psi}_{T+1}^{2IP} - \psi_{T+1}) \leq \widehat{F}_T^{2IP}{}^{-1}(1 - \gamma_2) \mid \mathcal{I}_{1:T} \right] \xrightarrow{p} 1 - \gamma. \end{aligned}$$

Claim 2(a) is similarly proven as 1(a). For the same reasons as in the proof of part 1(a) we begin by constructing the split analogue of \hat{Z}_T^{IP} . Let \hat{Z}_T^{SPL} follow the mixture distribution $\hat{G}_T(\cdot)$ as a function of $\mathbf{X}_{1:T_E}$ such that given $\mathbf{X}_{1:T_E}$ the conditional distribution of the random variable \hat{Z}_T^{SPL} is $\hat{G}_T(\cdot)$. Further, let

$$\widehat{F}_T^{SPL}(\cdot) \text{ be the conditional cdf of } \hat{w}_T^{SPL} \hat{Z}_T^{SPL} \text{ given } \mathcal{I}_{1:T} \tag{A.7}$$

where \hat{w}_T^{SPL} equals the transpose of $\frac{\partial \psi_T^s(\mathbf{X}_{T_P:T}; \hat{\theta}(\mathbf{X}_{1:T_E}))}{\partial \theta}$. Then

$$\begin{aligned} & d_{BL}\left(F_T^{SPL}(\cdot | \mathcal{I}_{T_P:T}), \widehat{F}_T^{SPL}(\cdot)\right) \\ & \leq d_{BL}\left(F_T^{SPL}(\cdot | \mathcal{I}_{T_P:T}), F_{\infty, T}^{SPL}(\cdot | \mathcal{I}_{T_P:T})\right) + d_{BL}\left(F_{\infty, T}^{SPL}(\cdot | \mathcal{I}_{T_P:T}), \widehat{F}_T^{SPL}(\cdot)\right), \end{aligned}$$

where $F_{\infty, T}^{SPL}(\cdot | \mathcal{I}_{T_P:T})$ is defined in equation (A.5). In the proof of Theorem 1, we have shown that $F_T^{SPL}(\cdot | \mathcal{I}_{T_P:T})$ and $F_{\infty, T}^{SPL}(\cdot | \mathcal{I}_{T_P:T})$ merge in probability under Assumptions 1 and 3. It suffices to show that $F_{\infty, T}^{SPL}(\cdot | \mathcal{I}_{T_P:T})$ and $\widehat{F}_T^{SPL}(\cdot)$ merge in probability. Once more we use Lemma 4 for this. Write $\hat{w}_T^{SPL} \hat{Z}_T^{SPL} = w_T^{SPL} \hat{Z}_T^{SPL} + \hat{R}_T^{SPL}$ with $\hat{R}_T^{SPL} = (\hat{w}_T^{SPL} - w_T^{SPL}) \hat{Z}_T^{SPL}$. First, we show $\hat{R}_T^{SPL} = o_p(1)$. Take an arbitrary $\epsilon > 0$. We obtain

$$\mathbb{P}\left[|\hat{R}_T^{SPL}| \geq \epsilon \right] \leq \mathbb{P}\left[\left\| \frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{T_P:T}; \hat{\theta}_T)}{\partial \theta \partial \theta'} \right\| \|\hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0\| \|\hat{Z}_T^{SPL}\| \geq \epsilon \right]$$

$$\begin{aligned} &\leq \mathbb{P} \left[\left\| \frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \ddot{\theta}_T)}{\partial \theta \partial \theta'} \right\| \left\| \hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0 \right\| \left\| \hat{Z}_T^{SPL} \right\| \geq \epsilon \cap \ddot{\theta}_T \in \mathcal{V}(\theta_0) \right] \\ &\quad + \mathbb{P} \left[\ddot{\theta}_T \notin \mathcal{V}(\theta_0) \right] \\ &\leq \mathbb{P} \left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \theta)}{\partial \theta \partial \theta'} \right\| \left\| \hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0 \right\| \left\| \hat{Z}_T^{SPL} \right\| \geq \epsilon \right] \\ &\quad + \mathbb{P} \left[\ddot{\theta}_T \notin \mathcal{V}(\theta_0) \right], \end{aligned}$$

where $\ddot{\theta}_T$ lies between $\hat{\theta}(\mathbf{X}_{1:T_E})$ and θ_0 . The first term vanishes as

$$\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \theta)}{\partial \theta \partial \theta'} \right\| = O_p(1)$$

by Assumptions 1.d and 1.e and $\|\hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0\| = O_p(m_{T_E}^{-1})$ by Assumptions 1.a and 3.a, $\|\hat{Z}_T^{SPL}\| = O_p(1)$ as $\hat{Z}_T^{SPL} \sim \hat{G}_T(\cdot)$ and $\int h d\hat{G}_T \xrightarrow{P} \int h dG_\infty$ for all $h \in BL$ by Assumptions 3.a, 3.b and 4. Further, as $\hat{\theta}(\mathbf{X}_{1:T_E}) \xrightarrow{P} \theta_0 \in \mathcal{V}(\theta_0)$ and $\mathcal{V}(\theta_0)$ is open, we have $\mathbb{P}[\ddot{\theta}_T \notin \mathcal{V}(\theta_0)] \rightarrow 0$ and $\hat{R}_T^{SPL} = o_p(1)$ follows. Moreover, w_T^{SPL} is $\mathcal{I}_{1:T}$ -measurable and $O_p(1)$ by Assumptions 1.c and 1.e and $\int h d\bar{\mathbb{P}}[\hat{Z}_T^{SPL} \leq \cdot | \mathcal{I}_{1:T}] = \int h d\hat{G}_T(\cdot) \xrightarrow{P} \int h dG_\infty$ for each $h \in BL$. Replacing Z_T , R_T , w_T and \mathcal{I}_T in Lemma 4 by \hat{Z}_T^{SPL} , \hat{R}_T^{SPL} , w_T^{SPL} and $\mathcal{I}_{1:T}$ implies that $\widehat{F}_T^{SPL}(\cdot) = \bar{\mathbb{P}}[w_T^{SPL} \hat{Z}_T^{SPL} + \hat{R}_T^{SPL} \leq \cdot | \mathcal{I}_{1:T}]$ and $\bar{\mathbb{P}}[w_T^{SPL} Z_\infty \leq \cdot | \mathcal{I}_{1:T}] = F_{\infty,T}^{SPL}(\cdot | \mathcal{I}_{T_P:T})$ merge in probability.

The proof of statement 2(b) is similar to the proof of claim 1(b). Because $F_T^{SPL}(\cdot | \mathcal{I}_{T_P:T})$ and $\widehat{F}_T^{SPL}(\cdot)$ merge in probability and $\widehat{F}_T^{SPL}(\cdot)$ is assumed to be stochastically uniformly continuous, Lemma 7 applies. Replacing F_T , G_T , M_T and \mathcal{I}_T by $F_T^{SPL}(\cdot | \mathcal{I}_{T_P:T})$, $\widehat{F}_T^{SPL}(\cdot)$, $m_{T_E}(\hat{\psi}_{T+1}^{SPL} - \psi_{T+1})$ and $\mathcal{I}_{T_P:T}$, respectively, it follows that

$$\begin{aligned} &\mathbb{P} \left[I_\gamma^{SPL}(\mathbf{X}_{T_P:T}, \mathbf{X}_{1:T_E}) \ni \psi_{T+1} | \mathcal{I}_{T_P:T} \right] \\ &= \mathbb{P} \left[\widehat{F}_T^{SPL}(\gamma_1) \leq m_{T_E}(\hat{\psi}_{T+1}^{SPL} - \psi_{T+1}) \leq \widehat{F}_T^{SPL}(\gamma_1) | \mathcal{I}_{T_P:T} \right] \\ &\xrightarrow{P} 1 - \gamma. \quad \square \end{aligned}$$

Proof of Theorem 3. Consider the first statement and expand

$$\begin{aligned} &\hat{\psi}_{T+1}(\mathbf{X}_{1:T}, \mathbf{X}_{1:T}) - \hat{\psi}_{T+1}(\mathbf{X}_{T_P:T}^c, \mathbf{X}_{1:T_E}) \\ &= (\hat{\psi}_{T+1}(\mathbf{X}_{1:T}, \mathbf{X}_{1:T}) - \psi_{T+1}) - (\hat{\psi}_{T+1}(\mathbf{X}_{T_P:T}^c, \mathbf{X}_{1:T_E}) - \psi_{T+1}). \end{aligned}$$

We show that both terms are $o_p(1)$. Using (A.3), we have

$$\hat{\psi}_{T+1}(\mathbf{X}_{T_P:T}^c, \mathbf{X}_{1:T_E}) - \psi_{T+1} = \frac{\partial \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \theta_0)}{\partial \theta'} (\hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0) + m_T^{-1} R_T^{SPL},$$

where $\frac{\partial \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \theta_0)}{\partial \theta'} = O_p(1)$ by Assumptions 1.c and 1.e and $\hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0 = o_p(1)$ by Assumptions 1.a, 3.a and 3.b. Together with $R_T^{SPL} = o_p(1)$ by Lemma 1 and $m_T^{-1} = o(1)$, it implies that $\hat{\psi}_{T+1}(\mathbf{X}_{T_P:T}^c, \mathbf{X}_{1:T_E}) - \psi_{T+1} = o_p(1)$. In addition, replacing $\mathbf{Y}_{1:T}$ by $\mathbf{X}_{1:T}$ in equation (A.2), we get

$$\hat{\psi}_{T+1}(\mathbf{X}_{1:T}, \mathbf{X}_{1:T}) - \psi_{T+1} = \frac{\partial \psi_{T+1}^s(\mathbf{X}_{1:T}, \theta_0)}{\partial \theta'} (\hat{\theta}(\mathbf{X}_{1:T}) - \theta_0) + m_T^{-1} R_T^{STA},$$

where R_T^{STA} is obtained by replacing $\mathbf{Y}_{1:T}$ by $\mathbf{X}_{1:T}$ in R_T^{2IP} . Note that we have $\left\| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{1:T}, \theta_0)}{\partial \theta} \right\| = O_p(1)$ by Assumptions 1.c and 1.e and $\hat{\theta}(\mathbf{X}_{1:T}) - \theta_0 = o_p(1)$ by Assumption 1.a. Since $R_T^{2IP} = o_p(1)$ has been shown in Lemma 1 without using Assumption 2.b, we have $R_T^{STA} = o_p(1)$. Together with $m_T^{-1} = o(1)$, it follows that $\hat{\psi}_{T+1}(\mathbf{X}_{1:T}, \mathbf{X}_{1:T}) - \psi_{T+1} = o_p(1)$ completing the claim.

To prove the second statement we want to apply Lemma 8. The stochastic continuity assumption of this lemma holds by the assumptions imposed in Theorem 3. It therefore only remains to show that the first assumption of Lemma 8 holds, i.e. $\widehat{F}_T^{STA}(\cdot)$ and $\widehat{F}_T^{SPL}(\cdot)$ merge in probability. To show the latter, similar to the proof of Theorem 2 we let \hat{Z}_T^{STA} follow the mixture distribution $\widehat{G}_T(\cdot)$ as a function of $\mathbf{X}_{1:T}$ such that given $\mathbf{X}_{1:T}$ the conditional cdf of the random variable \hat{Z}_T^{STA} is $\widehat{G}_T(\cdot)$. Further, let

$$\widehat{F}_T^{STA}(\cdot) \text{ be the conditional cdf of } \hat{w}_T^{STA} \hat{Z}_T^{STA} \text{ given } \mathcal{I}_{1:T} \quad (\text{A.8})$$

where \hat{w}_T^{STA} equals the transpose of $\frac{\partial \psi_T^s(\mathbf{X}_{1:T}; \hat{\theta}(\mathbf{X}_{1:T}))}{\partial \theta}$. Now we show that $\widehat{F}_T^{STA}(\cdot)$ and $\widehat{F}_T^{SPL}(\cdot)$, defined in (A.7), merge in probability. The triangle inequality implies

$$\begin{aligned} d_{BL} \left(\widehat{F}_T^{SPL}(\cdot), \widehat{F}_T^{STA}(\cdot) \right) &\leq d_{BL} \left(\widehat{F}_T^{SPL}(\cdot), F_T^{SPL}(\cdot | \mathcal{I}_{T_P:T}) \right) \\ &\quad + d_{BL} \left(F_T^{SPL}(\cdot | \mathcal{I}_{T_P:T}), F_T^{2IP}(\cdot | \mathcal{I}_{1:T}) \right) + d_{BL} \left(F_T^{2IP}(\cdot | \mathcal{I}_{1:T}), \widehat{F}_T^{STA}(\cdot) \right), \end{aligned}$$

where the first two terms on the right hand side converge in probability to zero by Theorem 2.2(a) and Theorem 1, respectively. We are left to show that $F_T^{2IP}(\cdot | \mathcal{I}_{1:T})$ and $\widehat{F}_T^{STA}(\cdot)$ merge in probability. The triangle inequality implies that

$$\begin{aligned} d_{BL} \left(F_T^{2IP}(\cdot | \mathcal{I}_{1:T}), \widehat{F}_T^{STA}(\cdot) \right) &\leq d_{BL} \left(F_T^{2IP}(\cdot | \mathcal{I}_{1:T}), F_{\infty,T}^{2IP}(\cdot | \mathcal{I}_{1:T}) \right) \\ &\quad + d_{BL} \left(F_{\infty,T}^{2IP}(\cdot | \mathcal{I}_{1:T}), \widehat{F}_T^{STA}(\cdot) \right), \end{aligned}$$

where $F_{\infty,T}^{2IP}(\cdot | \mathcal{I}_{1:T})$ is defined in equation (A.4). In the proof of Theorem 1, we have shown that $F_T^{2IP}(\cdot | \mathcal{I}_{1:T})$ and $F_{\infty,T}^{2IP}(\cdot | \mathcal{I}_{1:T})$ merge in probability under Assumptions 1 and 2. It suffices to show that $F_{\infty,T}^{2IP}(\cdot | \mathcal{I}_{1:T})$ and $\widehat{F}_T^{STA}(\cdot)$ merge in probability. Once more we use Lemma 4 for this. Write $\hat{w}_T^{STA} \hat{Z}_T^{STA} =$

$w_T^{2IP} \hat{Z}_T^{STA} + \hat{R}_T^{STA}$ with $\hat{R}_T^{STA} = (\hat{w}_T^{STA} - w_T^{2IP}) \hat{Z}_T^{STA}$ (note that, in contrast to \hat{w}_T^{STA} , there is no need to introduce w_T^{STA} as it equals w_T^{2IP}). First, we show $\hat{R}_T^{STA} = o_p(1)$. Take an arbitrary $\epsilon > 0$. We obtain

$$\begin{aligned} \mathbb{P}\left[|\hat{R}_T^{STA}| \geq \epsilon\right] &\leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{1:T}; \ddot{\theta}_T)}{\partial \theta \partial \theta'}\right\| \|\hat{\theta}(\mathbf{X}_{1:T}) - \theta_0\| \|\hat{Z}_T^{STA}\| \geq \epsilon\right] \\ &\leq \mathbb{P}\left[\left\|\frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{1:T}; \ddot{\theta}_T)}{\partial \theta \partial \theta'}\right\| \|\hat{\theta}(\mathbf{X}_{1:T}) - \theta_0\| \|\hat{Z}_T^{STA}\| \geq \epsilon \cap \ddot{\theta}_T \in \mathcal{V}(\theta_0)\right] \\ &\quad + \mathbb{P}\left[\ddot{\theta}_T \notin \mathcal{V}(\theta_0)\right] \\ &\leq \mathbb{P}\left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left\|\frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{1:T}; \theta)}{\partial \theta \partial \theta'}\right\| \|\hat{\theta}(\mathbf{X}_{1:T}) - \theta_0\| \|\hat{Z}_T^{STA}\| \geq \epsilon\right] \\ &\quad + \mathbb{P}\left[\ddot{\theta}_T \notin \mathcal{V}(\theta_0)\right], \end{aligned}$$

where $\ddot{\theta}_T$ lies between $\hat{\theta}(\mathbf{X}_{1:T})$ and θ_0 . The first term vanishes because we have that $\sup_{\theta \in \mathcal{V}(\theta_0)} \left\|\frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{1:T}; \theta)}{\partial \theta \partial \theta'}\right\|$ is $O_p(1)$ by Assumptions 1.d and 1.e, $\|\hat{\theta}(\mathbf{X}_{1:T}) - \theta_0\| = O_p(m_T^{-1})$ by Assumption 1.a and $\|\hat{Z}_T^{STA}\| = O_p(1)$ as $\hat{Z}_T^{STA} \sim \hat{G}_T(\cdot) \xrightarrow{P} G_\infty$ by Assumption 4. Further, as $\hat{\theta}(\mathbf{X}_{1:T}) \xrightarrow{P} \theta_0 \in \mathcal{V}(\theta_0)$ and $\mathcal{V}(\theta_0)$ is open, we have $\mathbb{P}[\ddot{\theta}_T \notin \mathcal{V}(\theta_0)] \rightarrow 0$ and $\hat{R}_T^{STA} = o_p(1)$ follows. Moreover, w_T^{2IP} is $\mathcal{I}_{1:T}$ -measurable and $O_p(1)$ by Assumptions 1.c and 1.e and $\int h d\bar{\mathbb{P}}[\hat{Z}_T^{STA} \leq \cdot | \mathcal{I}_{1:T}] = \int h d\hat{G}_T(\cdot) \xrightarrow{P} \int h dG_\infty$ for each $h \in BL$. Replacing Z_T , R_T , w_T and \mathcal{I}_T in Lemma 4 by \hat{Z}_T^{STA} , \hat{R}_T^{STA} , w_T^{2IP} and $\mathcal{I}_{1:T}$ implies that $\widehat{F}_T^{STA}(\cdot) = \bar{\mathbb{P}}[w_T^{2IP} \hat{Z}_T^{STA} + \hat{R}_T^{STA} \leq \cdot | \mathcal{I}_{1:T}]$ and $\bar{\mathbb{P}}[w_T^{2IP} Z_\infty \leq \cdot | \mathcal{I}_{1:T}] = F_{\infty,T}^{2IP}(\cdot | \mathcal{I}_{1:T})$ merge in probability. Thus, $\widehat{F}_T^{SPL}(\cdot)$ and $\widehat{F}_T^{STA}(\cdot)$ merge in probability. As mentioned earlier in the proof under the assumption that $\widehat{F}_T^{SPL}(\cdot)$ is stochastically pointwise continuous at γ_1 and $1 - \gamma_2$, assertion (3.9) follows by Lemma 8, which completes the proof. \square

Appendix B: Additional proofs

B.1. Proofs of lemmas

Proof of Lemma 1. Consider (i). By Assumption 1.b one can write R_T^{2IP} as follows:

$$\begin{aligned} R_T^{2IP} &= \underbrace{m_T(\psi_{T+1}^s(\mathbf{X}_{1:T}; \theta_0) - \psi_{T+1}(X_T, X_{T-1}, \dots; \theta_0))}_{=R_{1,T}^{2IP}} \\ &\quad + \underbrace{(\hat{\theta}(\mathbf{Y}_{1:T}) - \theta_0)' \frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{1:T}; \dot{\theta}_T)}{\partial \theta \partial \theta'} m_T(\hat{\theta}(\mathbf{Y}_{1:T}) - \theta_0)}_{=R_{2,T}^{2IP}}, \end{aligned}$$

where $\dot{\theta}_T$ lies between θ_0 and $\hat{\theta}(\mathbf{Y}_{1:T})$. By Assumption 1.e, $R_{1,T}^{2IP}$ is $o_p(1)$; hence we are left to show that $R_{2,T}^{2IP} = o_p(1)$. Take an arbitrary $\epsilon > 0$. We obtain

$$\begin{aligned} \mathbb{P}\left[|R_{2,T}^{2IP}| \geq \epsilon\right] &\leq \mathbb{P}\left[\left|\frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{1:T}; \dot{\theta}_T)}{\partial \theta \partial \theta'}\right| m_T \left\|\hat{\theta}(\mathbf{Y}_{1:T}) - \theta_0\right\|^2 \geq \epsilon\right] \\ &\leq \mathbb{P}\left[\left|\frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{1:T}; \dot{\theta}_T)}{\partial \theta \partial \theta'}\right| m_T \left\|\hat{\theta}(\mathbf{Y}_{1:T}) - \theta_0\right\|^2 \geq \epsilon \cap \dot{\theta}_T \in \mathcal{V}(\theta_0)\right] \\ &\quad + \mathbb{P}\left[\dot{\theta}_T \notin \mathcal{V}(\theta_0)\right] \\ &\leq \mathbb{P}\left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left|\frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{1:T}; \theta)}{\partial \theta \partial \theta'}\right| m_T \left\|\hat{\theta}(\mathbf{Y}_{1:T}) - \theta_0\right\|^2 \geq \epsilon\right] \\ &\quad + \mathbb{P}\left[\dot{\theta}_T \notin \mathcal{V}(\theta_0)\right]. \end{aligned}$$

The first term vanishes since $\sup_{\theta \in \mathcal{V}(\theta_0)} \left|\frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{1:T}; \theta)}{\partial \theta \partial \theta'}\right| = O_p(1)$ by Assumptions 1.d and 1.e and $m_T \left\|\hat{\theta}(\mathbf{Y}_{1:T}) - \theta_0\right\|^2 = O_p(m_T^{-1})$ by Assumptions 1.a and 2.a. Further, as $\hat{\theta}(\mathbf{Y}_{1:T}) \xrightarrow{p} \theta_0 \in \mathcal{V}(\theta_0)$ and $\mathcal{V}(\theta_0)$ is open, we have $\mathbb{P}[\dot{\theta}_T \notin \mathcal{V}(\theta_0)] \rightarrow 0$ and $R_{2,T}^{2IP} = o_p(1)$ follows.

The proof of (ii) is analogous; by Assumption 1.b one can express R_T^{SPL} as follows:

$$\begin{aligned} R_T^{SPL} &= \underbrace{m_{T_E} (\psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \theta_0) - \psi_{T+1}(X_T, X_{T-1}, \dots; \theta_0))}_{=R_{1,T}^{SPL}} \\ &\quad + \underbrace{(\hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0)' \frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \ddot{\theta}_T)}{\partial \theta \partial \theta'} m_{T_E} (\hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0)}_{=R_{2,T}^{SPL}} \end{aligned}$$

with $R_{1,T}^{SPL} = o_p(1)$ by Assumption 1.e and $\ddot{\theta}_T$ lying between θ_0 and $\hat{\theta}(\mathbf{X}_{1:T_E})$. For an arbitrary $\epsilon > 0$, we obtain

$$\begin{aligned} \mathbb{P}\left[|R_{2,T}^{SPL}| \geq \epsilon\right] &\leq \mathbb{P}\left[\left|\frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \ddot{\theta}_T)}{\partial \theta \partial \theta'}\right| m_{T_E} \left\|\hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0\right\|^2 \geq \epsilon\right] \\ &\leq \mathbb{P}\left[\left|\frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \ddot{\theta}_T)}{\partial \theta \partial \theta'}\right| m_{T_E} \left\|\hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0\right\|^2 \geq \epsilon \cap \ddot{\theta}_T \in \mathcal{V}(\theta_0)\right] \\ &\quad + \mathbb{P}\left[\ddot{\theta}_T \notin \mathcal{V}(\theta_0)\right] \\ &\leq \mathbb{P}\left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left|\frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \theta)}{\partial \theta \partial \theta'}\right| m_{T_E} \left\|\hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0\right\|^2 \geq \epsilon\right] \\ &\quad + \mathbb{P}\left[\ddot{\theta}_T \notin \mathcal{V}(\theta_0)\right]. \end{aligned}$$

The first term vanishes as $\sup_{\theta \in \mathcal{V}(\theta_0)} \left|\frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \theta)}{\partial \theta \partial \theta'}\right| = O_p(1)$ by Assumptions 1.d

and 1.e and $m_{T_E} \left| \hat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0 \right|^2 = O_p(m_{T_E}^{-1})$ by Assumptions 1.a and 3.a. Further, as $\hat{\theta}(\mathbf{X}_{1:T_E}) \xrightarrow{p} \theta_0 \in \mathcal{V}(\theta_0)$ and $\mathcal{V}(\theta_0)$ is open, we have $\mathbb{P}[\hat{\theta}_T \notin \mathcal{V}(\theta_0)] \rightarrow 0$ and $R_{2,T}^{SPL} = o_p(1)$ follows. \square

Proof of Lemma 2. Consider (i) and let G_T^{2IP} denote the unconditional distribution of $m_T(\hat{\theta}(\mathbf{Y}_{1:T}) - \theta_0)$. By Assumption 2.b, we have for each $h \in BL$

$$\int h dG_T^{2IP}(\cdot | \mathcal{I}_{1:T}) = \int h dG_T^{2IP} \rightarrow \int h dG_\infty,$$

where the last assertion comes from Assumptions 1.a and 2.a and Portmanteau’s Lemma (cf. [42, Lemma 2.2]). Consider (ii); for each $h \in BL$ we obtain

$$\begin{aligned} & \int h d(G_{T_E}^{SPL}(\cdot | \mathcal{I}_{T_P:T}) - G_\infty) \\ &= \underbrace{\int h d(G_{T_E}^{SPL} - G_\infty)}_I + \underbrace{\int h d(G_{T_E}^{SPL}(\cdot | \mathcal{I}_{T_P:T}) - G_{T_E}^{SPL})}_{II}, \end{aligned}$$

where $I \rightarrow 0$ by Assumptions 1.a, 3.a and 3.b and Portmanteau’s Lemma and $II \xrightarrow{p} 0$ by Assumption 3.c. \square

Proof of Lemma 3. For $r = 1$ Lemma 3 appears as Lemma 2 of the supplemental material to [11]. Extending their result to $r > 1$ we closely follow the proof of [16, Theorem 11.3.3] and write Q_T and Q to denote the probability measures corresponding to G_T and G , respectively. Let $\epsilon > 0$ and take a compact set $K \subset \mathbb{R}^r$ such that $Q(K) > 1 - \epsilon$. The set of functions $h \in \mathcal{H}$, restricted to K , form a compact set of functions for the supremum norm by the Arzela-Ascoli theorem (cf. [16, Theorem 2.4.7]). Thus for some finite $J = J(\epsilon)$ there are $h_1, \dots, h_J \in \mathcal{H}$ such that for any $h \in \mathcal{H}$, there is a $j \leq J$ with $\sup_{y \in K} |h(y) - h_j(y)| < \epsilon$. Let $K^\epsilon = \{y \in \mathbb{R}^r : \|x - y\| < \epsilon \text{ for some } x \in K\}$. One has $\sup_{x \in K^\epsilon} |h(x) - h_j(x)| < 3\epsilon$, since if $y \in K$ and $\|x - y\| < \epsilon$, then

$$\begin{aligned} |h(x) - h_j(x)| &\leq |h(x) - h(y)| + |h(y) - h_j(y)| + |h_j(y) - h_j(x)| \\ &\leq \|h\|_{BL} \|x - y\| + \epsilon + \|h_j\|_{BL} \|x - y\| < 3\epsilon. \end{aligned}$$

Let $g(x) = \max\{0, 1 - \|x - K\|/\epsilon\}$, where $\|x - K\| = \inf\{\|x - y\| : y \in K\}$ for all $x \in \mathbb{R}^r$. Then $g \in BL$ and $I\{x \in K\} \leq g \leq I\{x \in K^\epsilon\}$, where $I\{\cdot\}$ denotes the indicator function. It follows that

$$Q_T(\mathbb{R}^r \setminus K^\epsilon) = 1 - Q_T(K^\epsilon) \leq 1 - \int g dQ_T \xrightarrow{p} 1 - \int g dQ \leq 1 - Q(K) < \epsilon$$

or equivalently $\mathbb{P}[Q_T(\mathbb{R}^r \setminus K^\epsilon) \geq \epsilon] \rightarrow 0$. Thus, for each $h \in \mathcal{H}$ and h_j as above

$$\begin{aligned} \sup_{h \in \mathcal{H}} \left| \int h d(Q_T - Q) \right| &\leq \sup_{h \in \mathcal{H}} \int |h - h_j| d(Q_T + Q) + \left| \int h_j d(Q_T - Q) \right| \\ &\leq 2(Q_T + Q)(\mathbb{R}^r \setminus K^\epsilon) + 6\epsilon + \max_{1 \leq j \leq J} \left| \int h_j d(Q_T - Q) \right| \\ &\leq 8\epsilon + 2Q_T(\mathbb{R}^r \setminus K^\epsilon) + \max_{1 \leq j \leq J} \left| \int h_j d(Q_T - Q) \right|. \end{aligned}$$

Hence,

$$\begin{aligned} & \mathbb{P} \left[\sup_{h \in \mathcal{H}} \left| \int h d(Q_T - Q) \right| \geq 11\epsilon \right] \\ & \leq \mathbb{P} \left[2Q_T(\mathbb{R}^r \setminus K^\epsilon) + \max_{1 \leq j \leq J} \left| \int h_j d(Q_T - Q) \right| \geq 3\epsilon \right] \\ & \leq \mathbb{P} \left[Q_T(\mathbb{R}^r \setminus K^\epsilon) \geq \epsilon \right] + \mathbb{P} \left[\max_{1 \leq j \leq J} \left| \int h_j d(Q_T - Q) \right| \geq \epsilon \right] \\ & \leq \mathbb{P} \left[Q_T(\mathbb{R}^r \setminus K^\epsilon) \geq \epsilon \right] + \sum_{j=1}^J \mathbb{P} \left[\left| \int h_j d(Q_T - Q) \right| \geq \epsilon \right], \end{aligned}$$

where the last two terms are converging to 0 for finite J noting that $\int h_j d(Q_T - Q) = \int h_j d(G_T - G) \xrightarrow{P} 0$. Observing that $\sup_{h \in \mathcal{H}} \left| \int h d(Q_T - Q) \right| = \sup_{h \in \mathcal{H}} \left| \int h d(G_T - G) \right|$ completes the proof. \square

Proof of Lemma 4. This lemma is related to Lemma 8 in [3] where the quantity corresponding to $\bar{\mathbb{P}}[w'_T Z_\infty \leq \cdot | \mathcal{I}_T]$ is non-random. Set $\mathcal{F} = \{f : \mathbb{R} \rightarrow \mathbb{R} : \|f\|_{BL} \leq 1\}$. The triangle inequality implies

$$\begin{aligned} & \sup_{f \in \mathcal{F}} \left| \int [f(w'_T Z_T + R_T) - f(w'_T Z_\infty)] d\bar{\mathbb{P}}[\cdot | \mathcal{I}_T] \right| \\ & \leq \underbrace{\sup_{f \in \mathcal{F}} \left| \int [f(w'_T Z_T + R_T) - f(w'_T Z_T)] d\bar{\mathbb{P}}[\cdot | \mathcal{I}_T] \right|}_{=I} \\ & \quad + \underbrace{\sup_{f \in \mathcal{F}} \left| \int [f(w'_T Z_T) - f(w'_T Z_\infty)] d\bar{\mathbb{P}}[\cdot | \mathcal{I}_T] \right|}_{=II}. \end{aligned}$$

We show that $I \xrightarrow{P} 0$ and $II \xrightarrow{P} 0$. Let $\epsilon > 0$; as $\|f\|_{BL} \leq 1$ for all $f \in \mathcal{F}$ we obtain

$$\begin{aligned} I & \leq \sup_{f \in \mathcal{F}} \int \left| f(w'_T Z_T + R_T) - f(w'_T Z_T) \right| d\mathbb{P}[\cdot | \mathcal{I}_T] \\ & = \sup_{f \in \mathcal{F}} \int_{|R_T| \leq \epsilon} \left| f(w'_T Z_T + R_T) - f(w'_T Z_T) \right| d\bar{\mathbb{P}}[\cdot | \mathcal{I}_T] \\ & \quad + \sup_{f \in \mathcal{F}} \int_{|R_T| > \epsilon} \left| f(w'_T Z_T + R_T) - f(w'_T Z_T) \right| d\bar{\mathbb{P}}[\cdot | \mathcal{I}_T] \\ & \leq \sup_{f \in \mathcal{F}} \int_{|R_T| \leq \epsilon} \|f\|_{BL} |w'_T Z_T + R_T - w'_T Z_T| d\bar{\mathbb{P}}[\cdot | \mathcal{I}_T] \\ & \quad + \sup_{f \in \mathcal{F}} \int_{|R_T| > \epsilon} \left(|f(w'_T Z_T + R_T)| + |f(w'_T Z_T)| \right) d\bar{\mathbb{P}}[\cdot | \mathcal{I}_T] \end{aligned}$$

$$\begin{aligned} &\leq \sup_{f \in \mathcal{F}} \|f\|_{BL} \int_{|R_T| \leq \epsilon} |R_T| d\mathbb{P}[\cdot | \mathcal{I}_T] + 2 \sup_{f \in \mathcal{F}} \|f\|_{BL} \int_{|R_T| > \epsilon} d\bar{\mathbb{P}}[\cdot | \mathcal{I}_T] \\ &\leq \int_{|R_T| \leq \epsilon} \epsilon d\bar{\mathbb{P}}[\cdot | \mathcal{I}_T] + 2 \bar{\mathbb{P}}[|R_T| > \epsilon | \mathcal{I}_T] \leq \epsilon + 2 \bar{\mathbb{P}}[|R_T| > \epsilon | \mathcal{I}_T]. \end{aligned}$$

In line with [46, Theorem 3.3], employing Markov’s inequality we have

$$\bar{\mathbb{P}}[I \geq 2\epsilon] \leq \bar{\mathbb{P}}\left[\bar{\mathbb{P}}[|R_T| > \epsilon | \mathcal{I}_T] \geq \epsilon/2\right] \leq \frac{2}{\epsilon} \bar{\mathbb{P}}[|R_T| > \epsilon] \rightarrow 0$$

as $R_T = o_p(1)$ and hence $I = o_p(1)$. Consider II and let $K \geq 1$. We obtain

$$\bar{\mathbb{P}}[II \geq \epsilon] \leq \bar{\mathbb{P}}[\|w_T\| \geq K] + \bar{\mathbb{P}}[II \geq \epsilon \cap \|w_T\| \leq K].$$

As $\|w_T\| = O_p(1)$ the first term can be made arbitrarily small by choosing K large. For such K , consider the second term and note that

$$\begin{aligned} &\bar{\mathbb{P}}[II \geq \epsilon \cap \|w_T\| \leq K] \\ &= \bar{\mathbb{P}}\left[\sup_{f \in \mathcal{F}} \left| \int [f(w'_T Z_T) - f(w'_T Z_\infty)] d\bar{\mathbb{P}}[\cdot | \mathcal{I}_T] \right| \geq \epsilon \cap \|w_T\| \leq K\right] \\ &\leq \bar{\mathbb{P}}\left[\sup_{g \in \mathcal{G}} \left| \int [g(Z_T) - g(Z_\infty)] d\bar{\mathbb{P}}[\cdot | \mathcal{I}_T] \right| \geq \epsilon \cap \|w_T\| \leq K\right] \\ &\leq \bar{\mathbb{P}}\left[\sup_{g \in \mathcal{G}} \left| \int [g(Z_T) - g(Z_\infty)] d\bar{\mathbb{P}}[\cdot | \mathcal{I}_T] \right| \geq \epsilon\right] \\ &= \bar{\mathbb{P}}\left[\sup_{g \in \mathcal{G}} \left| \int g d\left(\mathbb{P}[Z_T \leq \cdot | \mathcal{I}_T] - G_\infty\right) \right| \geq \epsilon\right], \end{aligned}$$

where $\mathcal{G} = \{g : \mathbb{R}^r \rightarrow \mathbb{R} \mid g(x) = f(w'x), \text{ for some } f \in \mathcal{F} \text{ and some } w \in \mathbb{R}^r \text{ with } \|w\| \leq K\}$. We have that $\|\cdot\|_{BL}$ is uniformly bounded for \mathcal{G} since for every $g \in \mathcal{G}$

$$\begin{aligned} \|g\|_{BL} &= \sup_x |f(w'x)| + \sup_{x \neq y} \frac{|f(w'x) - f(w'y)|}{|w'x - w'y|} \frac{|w'x - w'y|}{\|x - y\|} \\ &\leq \sup_x |f(w'x)| + \sup_{x \neq y} \frac{|f(w'x) - f(w'y)|}{|w'x - w'y|} \|w\| \leq \|f\|_{BL} K \leq K. \end{aligned}$$

Thus, $\|g/K\|_{BL} \leq 1$ and it follows by $\bar{\mathbb{P}}[Z_T \leq \cdot | \mathcal{I}_T] \xrightarrow{p} G_\infty$ and Lemma 3 that

$$\begin{aligned} \bar{\mathbb{P}}[II \geq \epsilon \cap \|w_T\| \leq K] &\leq \bar{\mathbb{P}}\left[\sup_{g \in \mathcal{G}} \left| \int \frac{g}{K} d\left(\mathbb{P}[Z_T \leq \cdot | \mathcal{I}_T] - G_\infty\right) \right| \geq \frac{\epsilon}{K}\right] \\ &\leq \bar{\mathbb{P}}\left[\sup_{h \in \mathcal{H}} \left| \int h d\left(\mathbb{P}[Z_T \leq \cdot | \mathcal{I}_T] - G_\infty\right) \right| \geq \frac{\epsilon}{K}\right] \rightarrow 0, \end{aligned}$$

where \mathcal{H} is defined in Lemma 3. Thus, II is $o_p(1)$, which completes the proof. \square

Proof of Lemma 5. Take $\epsilon > 0$ and let F and G be cdfs on \mathbb{R} with $G(\tau - \epsilon) - \epsilon \leq F(\tau) \leq G(\tau + \epsilon) + \epsilon$ for all $\tau \in \mathbb{R}$. Fixing $u \in (\epsilon, 1 - \epsilon)$, we obtain

$$\inf \{ \tau \in \mathbb{R} : F(\tau) \geq u + \epsilon \} + \epsilon \tag{B.1}$$

$$\begin{aligned} &\geq \inf \{ \tau \in \mathbb{R} : G(\tau + \epsilon) + \epsilon \geq u + \epsilon \} + \epsilon \\ &= \inf \{ \tau \in \mathbb{R} : G(\tau + \epsilon) \geq u \} + \epsilon \\ &= \inf \{ \tau + \epsilon, \tau \in \mathbb{R} : G(\tau + \epsilon) \geq u \} \\ &= \inf \{ \tau \in \mathbb{R} : G(\tau) \geq u \} \end{aligned} \tag{B.2}$$

$$\begin{aligned} &= \inf \{ \tau + \epsilon, \tau \in \mathbb{R} : G(\tau) \geq u \} - \epsilon \\ &= \inf \{ \tau \in \mathbb{R} : G(\tau - \epsilon) \geq u \} - \epsilon \\ &= \inf \{ \tau \in \mathbb{R} : G(\tau - \epsilon) - \epsilon \geq u - \epsilon \} - \epsilon \\ &\geq \inf \{ \tau \in \mathbb{R} : F(\tau) \geq u - \epsilon \} - \epsilon. \end{aligned} \tag{B.3}$$

Identifying (B.1), (B.2) and (B.3) as $F^{-1}(u + \epsilon) + \epsilon$, $G^{-1}(u)$ and $F^{-1}(u - \epsilon) - \epsilon$, respectively, completes the proof. \square

Proof of Lemma 6. Let $\epsilon, \eta > 0$. As G_T is stochastically uniformly equicontinuous, there exists a $\delta > 0$ and an $\bar{T}_1 \in \mathbb{N}$ such that $\mathbb{P} \left[\sup_{\tau \in \mathbb{R}} \sup_{\tau' \in \mathbb{R} : |\tau - \tau'| < \delta} |G_T(\tau') - G_T(\tau)| > \epsilon \right] < \eta$ for all $T \geq \bar{T}_1$. Take $\varkappa = \min(\delta/2, \epsilon)$. As $d_L(F_T, G_T) \xrightarrow{P} 0$ as $T \rightarrow \infty$, there exists an \bar{T}_2 such that $\mathbb{P}[d_L(F_T, G_T) > \varkappa] < \eta$ for all $T \geq \bar{T}_2$. Let $\bar{T} = \max(\bar{T}_1, \bar{T}_2)$.

$$\begin{aligned} &\mathbb{P} \left[\sup_{\tau \in \mathbb{R}} |F_T(\tau) - G_T(\tau)| > 2\epsilon \right] \\ &\leq \mathbb{P} \left[\sup_{\tau \in \mathbb{R}} |F_T(\tau) - G_T(\tau)| > 2\epsilon \cap d_L(F_T, G_T) \leq \varkappa \right] + \mathbb{P} \left[d_L(F_T, G_T) > \varkappa \right] \\ &\leq \mathbb{P} \left[\varkappa + \sup_{\tau \in \mathbb{R}} |G_T(\tau \pm \varkappa) - G_T(\tau)| > 2\epsilon \right] + \mathbb{P} \left[d_L(F_T, G_T) > \varkappa \right] \\ &\leq \mathbb{P} \left[\sup_{\tau \in \mathbb{R}} \sup_{\tau' \in \mathbb{R} : |\tau - \tau'| < \delta} |G_T(\tau') - G_T(\tau)| > \epsilon \right] + \mathbb{P} \left[d_L(F_T, G_T) > \varkappa \right] < 2\eta \end{aligned}$$

for all $T \geq \bar{T}$. Since the choice of ϵ and η was arbitrary, the desired result follows. \square

Proof of Lemma 7. Since F_T and G_T merge in probability and $d_L \leq 2d_{BL}^{1/2}$ (cf. [24, p. 36]; [16, Theorem 11.3.3]), we have $d_L(F_T, G_T) \xrightarrow{P} 0$. Let $u \in (0, 1)$ and take $\epsilon > 0$ sufficiently small satisfying $u \in (\epsilon, 1 - \epsilon)$. $\mathbb{P}[d_L(F_T, G_T) > \epsilon | \mathcal{I}_T]$ is $o_p(1)$ since for every $\delta > 0$ the Markov inequality implies $\mathbb{P} \left[\mathbb{P}[d_L(F_T, G_T) > \epsilon | \mathcal{I}_T] \geq \delta \right] \leq \frac{1}{\delta} \mathbb{P}[d_L(F_T, G_T) > \epsilon] \rightarrow 0$. Employing Lemma 5 we derive the following bounds:

$$\mathbb{P}[M_T \leq G_T^{-1}(u) | \mathcal{I}_T]$$

$$\begin{aligned}
&\leq \mathbb{P}[M_T \leq G_T^{-1}(u) \cap d_L(F_T, G_T) \leq \epsilon | \mathcal{I}_T] + \mathbb{P}[d_L(F_T, G_T) > \epsilon | \mathcal{I}_T] \\
&\leq \mathbb{P}[M_T \leq F_T^{-1}(u + \epsilon) + \epsilon \cap d_L(F_T, G_T) \leq \epsilon | \mathcal{I}_T] + o_p(1) \\
&\leq \mathbb{P}[M_T \leq F_T^{-1}(u + \epsilon) + \epsilon | \mathcal{I}_T] + o_p(1) \\
&= F_T(F_T^{-1}(u + \epsilon) + \epsilon) + o_p(1) = U_T \\
\mathbb{P}[M_T < G_T^{-1}(u) | \mathcal{I}_T] \\
&\geq \mathbb{P}[M_T < G_T^{-1}(u) \cap d_L(F_T, G_T) \leq \epsilon | \mathcal{I}_T] \\
&\geq \mathbb{P}[M_T < F_T^{-1}(u - \epsilon) - \epsilon \cap d_L(F_T, G_T) \leq \epsilon | \mathcal{I}_T] \\
&\geq \mathbb{P}[M_T < F_T^{-1}(u - \epsilon) - \epsilon | \mathcal{I}_T] - \mathbb{P}[d_L(F_T, G_T) > \epsilon | \mathcal{I}_T] \\
&= F_T(F_T^{-1}(u - \epsilon) - \epsilon -) - o_p(1) = L_T,
\end{aligned}$$

where $F_T(\cdot -)$ denotes the left limit of $F_T(\cdot)$. We show that L_T and U_T converge in probability to u . Regarding the lower bound L_T we have

$$\begin{aligned}
&\left| F_T(F_T^{-1}(u - \epsilon) - \epsilon -) - u \right| \\
&\leq \left| F_T(F_T^{-1}(u - \epsilon) - \epsilon -) - F_T(F_T^{-1}(u - \epsilon) -) \right| \\
&\quad + \left| F_T(F_T^{-1}(u - \epsilon) -) - (u - \epsilon) \right| + \epsilon \\
&\leq \sup_{\tau \in \mathbb{R}} \left| F_T(\tau - \epsilon -) - F_T(\tau -) \right| + \left| F_T(F_T^{-1}(u - \epsilon) -) - (u - \epsilon) \right| + \epsilon \\
&\leq \sup_{\tau \in \mathbb{R}} \left| F_T(\tau - \epsilon -) - F_T(\tau -) \right| + \sup_{\tau \in \mathbb{R}} \left| F_T(\tau) - F_T(\tau -) \right| + \epsilon \\
&\leq 4d_K(F_T, G_T) + \sup_{\tau \in \mathbb{R}} \left| G_T(\tau - \epsilon -) - G_T(\tau -) \right| + \sup_{\tau \in \mathbb{R}} \left| G_T(\tau) - G_T(\tau -) \right| + \epsilon,
\end{aligned}$$

where the third inequality is due to [12, p. 217]. As $d_L(F_T, G_T) \xrightarrow{P} 0$ and G_T is stochastically uniformly equicontinuous, Lemma 6 implies $d_K(F_T, G_T) \xrightarrow{P} 0$. Further, $\sup_{\tau \in \mathbb{R}} |G_T(\tau - \epsilon -) - G_T(\tau -)| = o_p(1)$ and $\sup_{\tau \in \mathbb{R}} |G_T(\tau) - G_T(\tau -)| = o_p(1)$

by stochastic uniform equicontinuity completing $L_T \xrightarrow{P} u$. Regarding the upper bound U_T we have

$$\begin{aligned}
&\left| F_T(F_T^{-1}(u + \epsilon) + \epsilon) - u \right| \leq \left| F_T(F_T^{-1}(u + \epsilon) + \epsilon) - F_T(F_T^{-1}(u + \epsilon)) \right| \\
&\quad + \left| F_T(F_T^{-1}(u + \epsilon)) - F_T(F_T^{-1}(u + \epsilon) -) \right| \\
&\quad + \left| F_T(F_T^{-1}(u + \epsilon) -) - (u + \epsilon) \right| + \epsilon \\
&\leq \sup_{\tau \in \mathbb{R}} \left| F_T(\tau + \epsilon) - F_T(\tau) \right| + 2 \sup_{\tau \in \mathbb{R}} \left| F_T(\tau) - F_T(\tau -) \right| + \epsilon \\
&\leq 6d_K(F_T, G_T) + \sup_{\tau \in \mathbb{R}} \left| G_T(\tau + \epsilon) - G_T(\tau) \right| + 2 \sup_{\tau \in \mathbb{R}} \left| G_T(\tau) - G_T(\tau -) \right| + \epsilon,
\end{aligned}$$

where all terms on the right hand side are $o_p(1)$ such that $U_T \xrightarrow{p} u$. We obtain

$$\underbrace{L_T(u)}_{\xrightarrow{p} u} \leq \mathbb{P}[M_T < G_T^{-1}(u)|\mathcal{I}_T] \leq \mathbb{P}[M_T \leq G_T^{-1}(u)|\mathcal{I}_T] \leq \underbrace{U_T(u)}_{\xrightarrow{p} u},$$

which implies that $\mathbb{P}[M_T < G_T^{-1}(u)|\mathcal{I}_T]$ and $\mathbb{P}[M_T \leq G_T^{-1}(u)|\mathcal{I}_T]$ converge in probability to u for arbitrary $u \in (0, 1)$; in particular γ_1 and $1 - \gamma_2$. It follows that

$$\begin{aligned} &\mathbb{P}[G_T^{-1}(\gamma_1) \leq M_T \leq G_T^{-1}(1 - \gamma_2)|\mathcal{I}_T] \\ &= \mathbb{P}[M_T \leq G_T^{-1}(1 - \gamma_2)|\mathcal{I}_T] - \mathbb{P}[M_T < G_T^{-1}(\gamma_1)|\mathcal{I}_T] \xrightarrow{p} 1 - \gamma_2 - \gamma_1. \quad \square \end{aligned}$$

Proof of Lemma 8. Let $\epsilon, \eta > 0$ and set $\bar{\epsilon} = \min\{\epsilon, u, 1 - u\}/2$. Since G_T^{-1} is pointwise equicontinuous at u , there exist a $\delta > 0$ and an \bar{T}_1 such that $\mathbb{P}[\sup_{|u-v|<\delta} |G_T^{-1}(v) - G_T^{-1}(u)| > \bar{\epsilon}] < \eta$ for all $T \geq \bar{T}_1$. Take $\varkappa = \min\{\delta/2, \bar{\epsilon}\}$. As

$d_L(F_T, G_T) \xrightarrow{p} 0$ as $T \rightarrow \infty$, there exists an \bar{T}_2 such that $\mathbb{P}[d_L(F_T, G_T) > \varkappa] < \eta$ for all $T \geq \bar{T}_2$.

$$\begin{aligned} &\mathbb{P}\left[|F_T^{-1}(u) - G_T^{-1}(u)| > 2\epsilon\right] \leq \mathbb{P}\left[|F_T^{-1}(u) - G_T^{-1}(u)| > 2\bar{\epsilon}\right] \\ &\leq \mathbb{P}\left[|F_T^{-1}(u) - G_T^{-1}(u)| > 2\bar{\epsilon} \cap d_L(F_T, G_T) \leq \varkappa\right] + \mathbb{P}\left[d_L(F_T, G_T) > \varkappa\right] \\ &\leq \mathbb{P}\left[\varkappa + |G_T^{-1}(u \pm \varkappa) - G_T^{-1}(u)| > 2\bar{\epsilon}\right] + \mathbb{P}\left[d_L(F_T, G_T) > \varkappa\right] \\ &\leq \mathbb{P}\left[\sup_{|u-v|<\delta} |G_T^{-1}(v) - G_T^{-1}(u)| > \bar{\epsilon}\right] + \mathbb{P}\left[d_L(F_T, G_T) > \varkappa\right] < 2\eta \end{aligned}$$

for all $T \geq \bar{T} = \max(\bar{T}_1, \bar{T}_2)$, where the third inequality follows from Lemma 5 and $u \in (\bar{\epsilon}, 1 - \bar{\epsilon}) \subseteq (\varkappa, 1 - \varkappa)$. As ϵ and η were arbitrarily chosen, this completes the proof. \square

B.2. Proofs of corollaries

Proof of Corollary 1. Statement 1(a) follows immediately from Theorem 2.1(a) and $\widehat{F}_T^{2IP}(\cdot)$ equals to $\Phi(\cdot / \sqrt{\hat{v}_T^{2IP}})$. Regarding claim 1(b), it is sufficient to show that $1/\hat{v}_T^{2IP} = O_p(1)$ implies that $\Phi(\cdot / \sqrt{\hat{v}_T^{2IP}})$ is stochastically uniformly equicontinuous by Theorem 2.1(b). Since $1/\hat{v}_T^{2IP} = O_p(1)$ by assumption, we have for all $\varkappa > 0$, there exist $K = K(\varkappa)$ and $\bar{T} = \bar{T}(\varkappa)$ such that $\mathbb{P}[1/\hat{v}_T^{2IP} > K] < \varkappa$ for all $T > \bar{T}$. Let ϕ denote the standard normal density. Taking $\delta = \frac{\epsilon}{\phi(0)\sqrt{K}}$, we obtain

$$\begin{aligned} &\mathbb{P}\left[\sup_{\tau \in \mathbb{R}} \sup_{\tau': |\tau - \tau'| < \delta} |\Phi(\tau' / \sqrt{\hat{v}_T^{2IP}}) - \Phi(\tau / \sqrt{\hat{v}_T^{2IP}})| > \epsilon\right] \\ &\leq \mathbb{P}\left[\phi(0)\delta / \sqrt{\hat{v}_T^{2IP}} > \epsilon\right] = \mathbb{P}\left[1/\hat{v}_T^{2IP} > K\right] < \varkappa \end{aligned}$$

for all $T > \bar{T}$ such that the stochastic uniform equicontinuity condition holds.

Statement 2(a) follows from Theorem 2.2(a) and $\widehat{F}_T^{SPL}(\cdot)$ equals to $\Phi(\cdot / \sqrt{\widehat{v}_T^{SPL}})$. Claim 2(b) is proven analogously to the claim of 1(b) replacing \widehat{v}_T^{2IP} with \widehat{v}_T^{SPL} . \square

Proof of Corollary 2. In the proof of Theorem 3 we have shown that $\widehat{F}_T^{SPL}(\cdot)$ and $\widehat{F}_T^{STA}(\cdot)$ merge in probability, which simplify to $\Phi(\cdot / \sqrt{\widehat{v}_T^{SPL}})$ and $\Phi(\cdot / \sqrt{\widehat{v}_T^{STA}})$, respectively, under Assumption 5.

It remains to show that $\widehat{F}_T^{SPL}{}^{-1}(u) = \sqrt{\widehat{v}_T^{SPL}}\Phi^{-1}(u)$ is stochastically pointwise equicontinuous at $u = \gamma_1, 1 - \gamma_2$. First, we show that $\widehat{v}_T^{SPL} = O_p(1)$. The triangle inequality implies $\widehat{v}_T^{SPL} \leq v_T^{SPL} + |\widehat{v}_T^{SPL} - v_T^{SPL}|$, where $v_T^{SPL} = \frac{\partial \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \theta_0)}{\partial \theta'} \Upsilon_0 \frac{\partial \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \theta_0)}{\partial \theta}$ is $O_p(1)$ by Assumptions 1.c and 1.e. Moreover, for an arbitrary $\varepsilon > 0$, we have

$$\begin{aligned} & \mathbb{P} \left[\left\| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \theta_0)}{\partial \theta} - \frac{\partial \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \widehat{\theta}(\mathbf{X}_{1:T_E}))}{\partial \theta} \right\| \geq \varepsilon \right] \\ & \leq \mathbb{P} \left[\left\| \frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \tilde{\theta}_T)}{\partial \theta \partial \theta'} \right\| \left\| \widehat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0 \right\| \geq \varepsilon \cap \tilde{\theta}_T \in \mathcal{V}(\theta_0) \right] \\ & \quad + \mathbb{P}[\tilde{\theta}_T \notin \mathcal{V}(\theta_0)] \\ & \leq \mathbb{P} \left[\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \theta)}{\partial \theta \partial \theta'} \right\| \left\| \widehat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0 \right\| \geq \varepsilon \right] + \mathbb{P}[\tilde{\theta}_T \notin \mathcal{V}(\theta_0)], \end{aligned}$$

where $\tilde{\theta}_T$ lies between $\widehat{\theta}(\mathbf{X}_{1:T_E})$ and θ_0 . The first term vanishes as $\sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{\partial^2 \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \theta)}{\partial \theta \partial \theta'} \right\| = O_p(1)$ by Assumptions 1.d and 1.e and $\left\| \widehat{\theta}(\mathbf{X}_{1:T_E}) - \theta_0 \right\| = O_p(m_T^{-1})$ by Assumptions 1.a and 2.a. Further, since $\widehat{\theta}(\mathbf{X}_{1:T_E}) \xrightarrow{p} \theta_0 \in \mathcal{V}(\theta_0)$ and $\mathcal{V}(\theta_0)$ is open, we have $\mathbb{P}[\tilde{\theta}_T \notin \mathcal{V}(\theta_0)] \rightarrow 0$ and $\left\| \frac{\partial \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \theta_0)}{\partial \theta} - \frac{\partial \psi_{T+1}^s(\mathbf{X}_{T_P:T}^c; \widehat{\theta}(\mathbf{X}_{1:T_E}))}{\partial \theta} \right\| = o_p(1)$ follows. Together with $\widehat{\Upsilon}(\mathbf{X}_{1:T_E}) \xrightarrow{p} \Upsilon_0$, it implies $|\widehat{v}_T^{SPL} - v_T^{SPL}| = o_p(1)$ and hence $\widehat{v}_T^{SPL} = O_p(1)$. Next, we show that the stochastic pointwise equicontinuity condition is satisfied. For $K > 0$, we get

$$\begin{aligned} & \mathbb{P} \left[\sqrt{\widehat{v}_T^{SPL}} \sup_{v:|u-v|<\delta} |\Phi^{-1}(u) - \Phi^{-1}(v)| > \varepsilon \right] \\ & \leq \mathbb{P} \left[\sqrt{K} \sup_{v:|u-v|<\delta} |\Phi^{-1}(u) - \Phi^{-1}(v)| > \varepsilon \right] + \mathbb{P}[\widehat{v}_T^{SPL} > K]. \end{aligned}$$

K can be chosen such that the last term is arbitrary small for large T as $\widehat{v}_T^{SPL} = O_p(1)$. Given K , the first term is 0 by the choice of δ and continuity of Φ^{-1} . \square

Appendix C: Graphs and tables

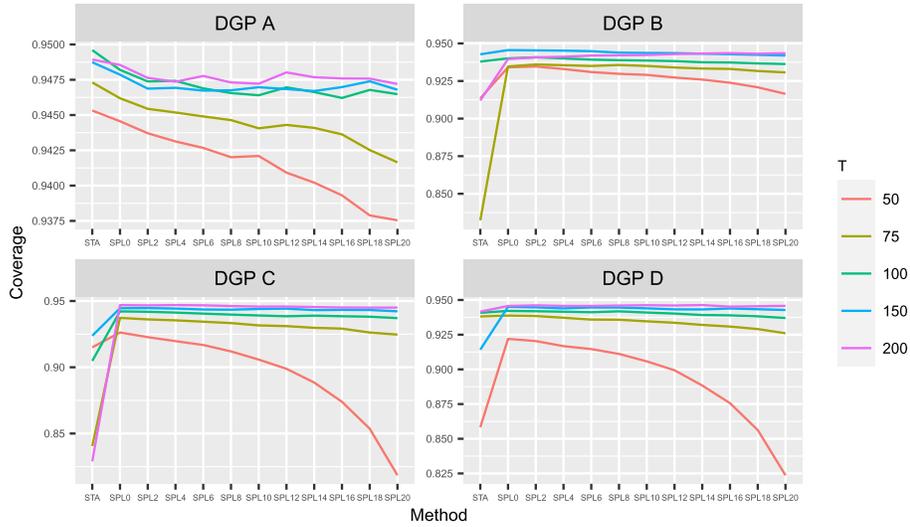


FIG 1. Mean coverage

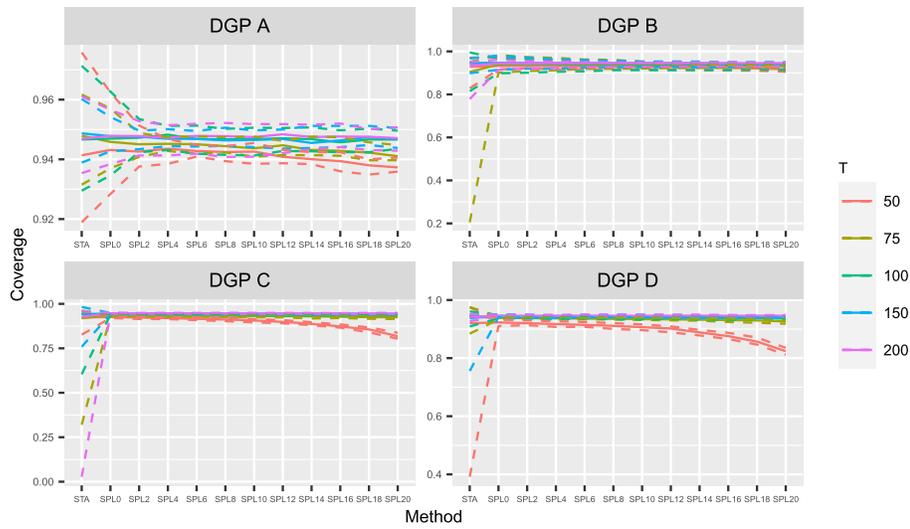


FIG 2. Median (solid), minimum and maximum (dashed) coverage

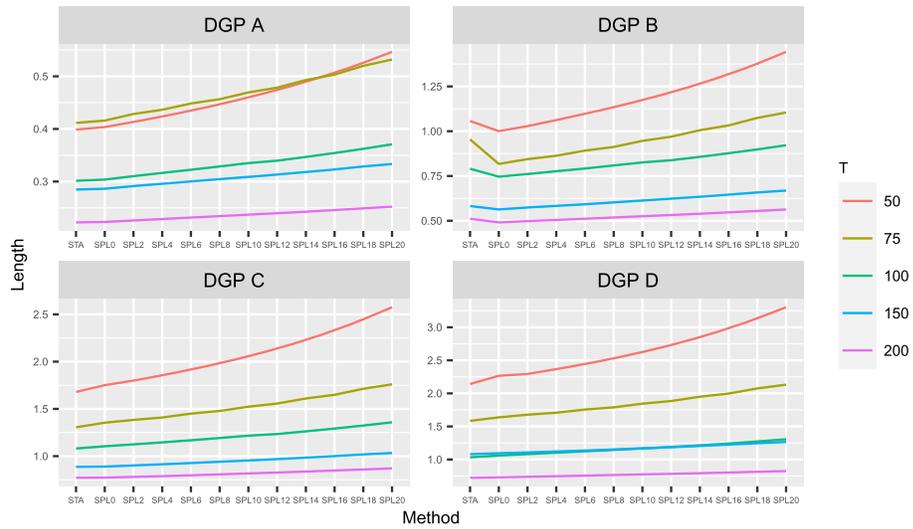


FIG 3. Mean interval length

TABLE 2

Coverage probability of STA and SPL for an $AR(\beta_1)$ process. DGP 1 refers to gamma distributed errors with shape 10 and scale 20 and shifted such that they have expectation zero. DGP 2 refers to a mixture distribution of this gamma distribution with a normal distribution with standard deviation equal to 10 and the weight of both distributions equals 0.5. Please note that there are small numerical differences for STA for $T = 50, T_E = 45$ and $T = 50, T_E = 40$, respectively. This is because when changing T_E we recalculated STA although T was kept fixed. We did so to compare STA and SPL based on the same sample.

β_1	AR(1)							
	DGP 1				DGP 2			
	$T = 50, T_E = 40$		$T = 50, T_E = 45$		$T = 50, T_E = 40$		$T = 50, T_E = 45$	
	STA	SPL	STA	SPL	STA	SPL	STA	SPL
0.1	94.8	94.7	94.5	94.4	95.2	95.1	94.4	99.4
0.2	94.9	94.8	94.6	94.6	95.3	95.1	94.5	94.3
0.3	94.7	94.5	94.7	94.7	95.0	95.0	94.4	94.4
0.4	94.7	94.7	94.4	94.4	95.0	95.0	94.4	94.4
0.5	94.7	94.5	94.7	94.7	95.0	95.0	94.7	4.6
0.6	94.7	94.5	94.6	94.8	95.0	94.8	94.8	94.7
0.7	94.5	94.4	94.8	94.7	94.7	94.5	94.9	94.8
0.8	94.1	94.0	94.9	94.8	94.5	94.3	94.9	94.9
0.9	93.5	93.2	94.7	94.5	93.8	93.6	94.6	94.6

TABLE 3

Coverage probability of STA and SPL for an AR(2) process with $\beta_1 = \beta_2$. DGP 1 and DGP 2 are as in Table 2. Small numerical differences between STA intervals for fixed T are as in Table 1 due to recalculating.

		AR(2)							
		DGP 1				DGP 2			
$\beta_1 = \beta_2$		$T = 50, T_E = 40$		$T = 50, T_E = 45$		$T = 50, T_E = 40$		$T = 50, T_E = 45$	
		STA	SPL	STA	SPL	STA	SPL	STA	SPL
0.1		94.3	94.2	93.9	93.9	94.6	94.8	93.7	93.8
0.15		94.3	94.2	93.9	94.0	94.8	94.8	93.8	94.0
0.2		94.2	94.1	93.8	94.0	94.8	94.8	94.0	94.0
0.25		94.2	94.1	94.0	94.0	94.7	94.5	94.0	94.0
0.3		94.2	94.0	94.1	94.0	94.4	94.4	94.0	94.0
0.35		94.2	93.7	94.1	94.0	94.4	94.2	94.0	94.0
0.4		94.1	93.7	94.2	94.0	94.3	93.9	94.0	93.8
0.45		93.6	92.9	94.1	93.6	93.7	93.0	94.2	93.7

TABLE 4

Coverage probability of STA and SPL for the AR(4) process in row B of Table 1. DGP 1 and DGP 2 are again as in Table 2. Small numerical differences between STA intervals for fixed T are as in Table 1 due to recalculating.

		AR(4)			
		DGP 1		DGP 2	
		STA	SPL	STA	SPL
$T_E = 45, T = 50$		92.8	93.2	92.7	93.2
$T_E = 55, T = 60$		93.3	93.7	93.2	93.5
$T_E = 65, T = 70$		93.6	93.8	93.7	93.9
$T_E = 75, T = 80$		93.7	94.0	93.6	93.9
$T_E = 85, T = 90$		94.0	94.2	93.9	94.2
$T_E = 40, T = 50$		92.9	93.2	92.7	93.0
$T_E = 50, T = 60$		93.3	93.6	93.3	93.5
$T_E = 60, T = 70$		93.6	94.0	93.6	93.9
$T_E = 70, T = 80$		93.8	94.0	93.8	94.1
$T_E = 80, T = 90$		93.9	94.2	93.8	94.1
$T_E = 40, T = 55$		93.0	93.1	93.1	93.0
$T_E = 50, T = 65$		93.5	93.5	93.4	93.4
$T_E = 60, T = 75$		93.6	93.8	93.7	93.9
$T_E = 70, T = 85$		93.9	94.0	93.8	94.0
$T_E = 80, T = 95$		94.0	94.1	93.9	94.2

TABLE 5

Coverage probability of STA and SPL for the AR(8) process in row C of Table 1. DGP 1 and DGP 2 are again as in Table 2. Small numerical differences between STA intervals for fixed T are as in Table 1 due to recalculating.

	AR(8), C			
	DGP 1		DGP 2	
	STA	SPL	STA	SPL
$T_E = 40, T = 50$	89.8	91.1	90.2	91.2
$T_E = 50, T = 60$	90.8	92.1	91.3	92.5
$T_E = 60, T = 70$	91.5	92.8	91.9	92.9
$T_E = 70, T = 80$	92.1	93.2	92.4	93.5
$T_E = 80, T = 90$	92.6	93.6	92.7	93.5
$T_E = 40, T = 55$	90.4	91.2	90.3	91.1
$T_E = 50, T = 65$	91.1	92.3	91.1	92.2
$T_E = 60, T = 75$	91.9	92.9	91.9	92.8
$T_E = 70, T = 85$	92.3	93.2	92.3	93.4
$T_E = 80, T = 95$	92.8	93.6	92.5	93.5

TABLE 6

Coverage probability of STA and SPL for the AR(8) process in row D of Table 1. DGP 1 and DGP 2 are again as in Table 2. Small numerical differences between STA intervals for fixed T are as in Table 1 due to recalculating.

	AR(8), D			
	DGP 1		DGP 2	
	STA	SPL	STA	SPL
$T_E = 40, T = 50$	89.6	91.3	89.5	91.3
$T_E = 50, T = 60$	90.9	92.4	90.6	92.1
$T_E = 60, T = 70$	91.5	92.8	91.5	92.9
$T_E = 70, T = 80$	92.0	93.4	92.0	93.1
$T_E = 80, T = 90$	92.6	93.7	92.5	93.5
$T_E = 40, T = 55$	90.2	91.2	90.2	91.2
$T_E = 50, T = 65$	91.4	92.4	91.0	92.2
$T_E = 60, T = 75$	91.8	92.9	91.7	92.8
$T_E = 70, T = 85$	92.2	93.2	92.2	93.2
$T_E = 80, T = 95$	92.7	93.5	92.6	93.6

Acknowledgments

The authors' work was supported by the Netherlands Organization for Scientific Research (NWO), grant number: 406-15-020.

References

- [1] Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* 21(1), 243–247. [MR0246476](#)
- [2] Barndorff-Nielsen, O. E. and D. R. Cox (1996). Prediction and asymptotics. *Bernoulli* 2(4), 319–340. [MR1440272](#)

- [3] Belyaev, Y. and S. Sjöstedt-De Luna (2000). Weakly approaching sequences of random distributions. *Journal of Applied Probability* 37(3), 807–822. [MR1782455](#)
- [4] Beutner, E., A. Heinemann, and S. Smeekes (2017a). A justification of conditional confidence intervals. GSBE Research Memorandum RM/17/023, Maastricht University.
- [5] Beutner, E., A. Heinemann, and S. Smeekes (2017b). Technical report on “a justification of conditional confidence intervals”. Technical report, Maastricht University, <http://researchers-sbe.unimaas.nl/stephansmeekes/research/>.
- [6] Beutner, E., A. Heinemann, and S. Smeekes (2019). A general framework for prediction in time series models. Working paper, Maastricht University, <http://researchers-sbe.unimaas.nl/stephansmeekes/research/>.
- [7] Billingsley, P. (1986). *Probability and Measure*. New York: John Wiley & Sons. [MR0830424](#)
- [8] Blasques, F., S. J. Koopman, K. Łasak, and A. Lucas (2016). In-sample confidence bands and out-of-sample forecast bands for time-varying parameters in observation-driven models. *International Journal of Forecasting* 32(3), 875–887. [MR2518405](#)
- [9] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31(3), 307–327. [MR0853051](#)
- [10] Boussama, F., F. Fuchs, and R. Stelzer (2011). Stationarity and geometric ergodicity of BEKK multivariate GARCH models. *Stochastic Processes and Their Applications* 121(10), 2331–2360. [MR2822779](#)
- [11] Castillo, I. and J. Rousseau (2015). A Bernstein–von Mises theorem for smooth functionals in semiparametric models. *Annals of Statistics* 43(6), 2353–2383. [MR3405597](#)
- [12] Cavaliere, G., I. Georgiev, and A. M. R. Taylor (2013). Wild bootstrap of the sample mean in the infinite variance case. *Econometric Reviews* 32(2), 204–219. [MR2988925](#)
- [13] D’Aristotile, A., P. Diaconis, and D. Freedman (1988). On merging of probabilities. *Sankhyā: The Indian Journal of Statistics, Series A* 50(3), 363–380. [MR1065549](#)
- [14] Davydov, Y. and V. Rotar (2009). On asymptotic proximity of distributions. *Journal of Theoretical Probability* 22(1), 82–98. [MR2472006](#)
- [15] Doukhan, P. (1994). *Mixing: Properties and Examples*. New York: Springer. [MR1312160](#)
- [16] Dudley, R. M. (2002). *Real Analysis and Probability*. Cambridge: Cambridge University Press. [MR1932358](#)
- [17] Dufour, J.-M. and A. Taamouti (2010). Short and long run causality measures: theory and inference. *Journal of Econometrics* 154(1), 42–58. [MR2558950](#)
- [18] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50(4), 987–1007. [MR0666121](#)

- [19] Francq, C. and J.-M. Zakoïan (2015). Risk-parameter estimation in volatility models. *Journal of Econometrics* 184(1), 158–173. [MR3277559](#)
- [20] Gospodinov, N. (2002). Median unbiased forecasts for highly persistent autoregressive processes. *Journal of Econometrics* 111(1), 85–101. [MR1926538](#)
- [21] Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and Its Application*. London: Academic Press. [MR0624435](#)
- [22] Hamilton, J. D. (1994). *Time Series Analysis*. Princeton: Princeton University Press. [MR1278033](#)
- [23] Hansen, B. E. (2006). Interval forecasts and parameter uncertainty. *Journal of Econometrics* 135(1), 377–398. [MR2328406](#)
- [24] Huber, P. J. (2009). *Robust Statistics* (2nd ed.). New York: John Wiley & Sons. [MR0606374](#)
- [25] Kabaila, P. (1999). The relevance property for prediction intervals. *Journal of Time Series Analysis* 20(6), 655–662.
- [26] Kabaila, P. and Z. He (2004). The adjustment of prediction intervals to account for errors in parameter estimation. *Journal of Time Series Analysis* 25(3), 351–358. [MR2062678](#)
- [27] Kabaila, P. and K. Syuhada (2008). Improved prediction limits for AR(p) and ARCH(p) processes. *Journal of Time Series Analysis* 29(2), 213–223. [MR2423111](#)
- [28] Kabaila, P. and K. Syuhada (2010). The asymptotic efficiency of improved prediction intervals. *Statistics & Probability Letters* 80(17), 1348–1353. [MR2669772](#)
- [29] Kreiss, J.-P. (2016). Discussion: bootstrap prediction intervals for linear, nonlinear and nonparametric autoregressions. *Journal of Statistical Planning and Inference* 177, 28–30. [MR3511319](#)
- [30] Kunitomo, N. and T. Yamamoto (1985). Properties of predictors in misspecified autoregressive time series models. *Journal of the American Statistical Association* 80(392), 941–950. [MR0819598](#)
- [31] Lewis, R. and G. C. Reinsel (1985). Prediction of multivariate time series by autoregressive model fitting. *Journal of Multivariate Analysis* 16(3), 393–411. [MR0793499](#)
- [32] Lunde, R. (2019, 2). Sample splitting and weak assumption inference for time series. [arXiv:1902.07425](#).
- [33] Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Berlin: Springer. [MR2172368](#)
- [34] Ng, S. and P. Perron (2005). A note on the selection of time series models. *Oxford Bulletin of Economics and Statistics* 67(1), 115–134.
- [35] Pan, L. and D. N. Politis (2016a). Bootstrap prediction intervals for linear, nonlinear and nonparametric autoregressions. *Journal of Statistical Planning and Inference* 177, 1–27. [MR3511318](#)
- [36] Pan, L. and D. N. Politis (2016b). Bootstrap prediction intervals for Markov processes. *Computational Statistics & Data Analysis* 100, 467–494. [MR3505813](#)

- [37] Pascual, L., J. Romo, and E. Ruiz (2004). Bootstrap predictive inference for ARIMA processes. *Journal of Time Series Analysis* 25(4), 449–465. [MR2073232](#)
- [38] Pascual, L., J. Romo, and E. Ruiz (2006). Bootstrap prediction for returns and volatilities in GARCH models. *Computational Statistics & Data Analysis* 50(9), 2293–2312. [MR2225571](#)
- [39] Pesaran, M. H. (2015). *Time Series and Panel Data Econometrics*. Oxford: Oxford University Press.
- [40] Phillips, P. C. B. (1979). The sampling distribution of forecasts from a first-order autoregression. *Journal of Econometrics* 9(3), 241–261. [MR0567058](#)
- [41] Samaranyake, V. A. and D. P. Hasza (1988). Properties of predictors for multivariate autoregressive models with estimated parameters. *Journal of Time Series Analysis* 9(4), 361–383. [MR0995614](#)
- [42] van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge: Cambridge University Press. [MR1652247](#)
- [43] Vidoni, P. (2004). Improved prediction intervals for stochastic process models. *Journal of Time Series Analysis* 25(1), 137–154. [MR2042115](#)
- [44] Vidoni, P. (2009). Improved prediction intervals and distribution functions. *Scandinavian Journal of Statistics* 36(4), 735–748. [MR2573305](#)
- [45] Vidoni, P. (2017). Improved multivariate prediction regions for Markov process models. *Statistical Methods & Applications* 26(1), 1–18. [MR3610179](#)
- [46] Xiong, S. and G. Li (2008). Some results on the convergence of conditional distributions. *Statistics & Probability Letters* 78(18), 3249–3253. [MR2479485](#)
- [47] Zakoïan, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18(5), 931–955.