

7540098-when-stress-tests-miss-the-risk-statistical-scenario-analysis-for-financial-portfolio

SUMMARY: This paper discusses the limitations of conventional scenario analysis in financial risk management, where stress tests often summarize scenario gains using point estimates that overlook the dependence between stressed and unstressed risk factors. The authors propose a machine learning framework, Adaptive Conformal Scenario Analysis (ACSA) and Kernel Scenario Analysis (KSA), to better quantify uncertainty in realized next-day scenario gains. ACSA calibrates scenario-specific quantile predictions and provides a long-run empirical coverage guarantee over realized scenarios, while KSA estimates scenario-conditional quantiles directly from specified stress and current market information. The methods are designed to provide better calibration and sharper prediction intervals than empirical-quantile baselines, and they move scenario analysis beyond point estimates by quantifying predictive uncertainty and offering tools for statistically validating scenario gains.

METHOD: The authors use experiments designed to reflect real-world markets to demonstrate the limitations of conventional scenario analysis, which can substantially understate risk, including for portfolios that appear safe under standard stress tests. They develop ACSA and KSA methods, which are built on machine learning frameworks. ACSA updates the quantile level based on whether the preceding interval covered the realized P&L in the realized scenario, while KSA reweights historical observations according to the specified stress and current market information.

KEY FINDINGS:

- ACSA provides online-calibrated prediction intervals that offer a long-run empirical coverage guarantee.
- KSA estimates scenario-conditional quantiles directly from specified stress and current market information.
- The proposed methods achieve better calibration and sharper prediction intervals than empirical-quantile baselines.
- The methods move scenario analysis beyond point estimates by quantifying predictive uncertainty and providing tools for statistically validating scenario gains.

IMPLICATION FOR TRADING: The proposed methods can improve risk management in financial portfolios by providing better quantification of predictive uncertainty. This can help practitioners avoid underestimating risk and make more informed decisions. By offering tools for statistically validating scenario gains, the methods can also help risk managers have a principled notion of the accuracy of their scenario numbers. This can lead to more reliable and robust risk management practices, ultimately improving the performance and stability of financial portfolios.

When Stress Tests Miss the Risk: Statistical Scenario Analysis for Financial Portfolios

Zhongze Cai¹ Martin B. Haugh¹ Xiaocheng Li²

¹Department of Analytics, Marketing and Operations
Imperial Business School, Imperial College
London, SW7 2AZ, UK

² Independent

This version: September 28, 2026

Abstract

Scenario analysis is widely used to stress test financial portfolios, yet conventional approaches often summarize scenario gains using point estimates that overlook dependence between stressed and unstressed risk factors. We develop a machine learning framework for quantifying uncertainty in realized next-day scenario gains through prediction intervals whose widths can be calibrated online. Adaptive conformal scenario analysis (ACSA) calibrates scenario-specific quantile predictions and provides a long-run empirical coverage guarantee over realized scenarios. Our main method, kernel scenario analysis (KSA), estimates scenario-conditional quantiles directly from the specified stress and current market information. KSA can also be combined with ACSA to obtain online-calibrated prediction intervals. In experiments designed to reflect real-world markets, conventional scenario analysis can substantially understate risk, including for portfolios that appear safe under standard stress tests. Our proposed methods achieve better calibration and sharper intervals than empirical-quantile baselines. Overall, the framework moves scenario analysis beyond point estimates by quantifying predictive uncertainty and providing tools for statistically validating scenario gains.

1 Introduction

Scenario analysis is a cornerstone of financial risk management. Banks, asset managers, and regulators routinely ask: *What happens to the portfolio if a small set of market drivers moves in an adverse but plausible way?* The output is typically a “scenario table” that reports an estimated profit-and-loss (P&L) for each prescribed stress or scenario. For an equity-derivatives portfolio, for example, a scenario might specify that the S&P 500 falls by 3% while the implied volatility surface shifts upward by several volatility points, and the scenario analysis report might state that the portfolio would gain (or lose) a particular dollar amount in that scenario.

Despite its operational importance, standard scenario analysis is often treated as a purely deterministic calculation. A small subset of risk-factor changes is “shocked”, while all other risk factors are implicitly held fixed. The resulting scenario P&L is then interpreted—often informally, but pervasively—as an *expected* P&L conditional on the scenario. This interpretation is appealing because it turns a high-dimensional and uncertain forecasting problem into a simple table of numbers. Unfortunately, it is also frequently wrong. There are at least three structural reasons for this.

First, financial risk factors often exhibit strong statistical dependence. Conditioning on a stress in one part of the market typically shifts the distribution of other risk factors, often materially. Treating the unstressed factors as unchanged can therefore yield a scenario P&L that is badly miscalibrated even as a conditional mean, let alone as a full description of risk. Second, for nonlinear portfolios, especially those containing options or other convex payoffs, even replacing unstressed factors by their conditional means can miss important *residual uncertainty* effects: convexity can cause the conditional expectation of the portfolio value to differ substantially from the portfolio value computed at the conditional expectation of the risk factors. Third, scenario analysis outputs are rarely accompanied by any statistical validation. Although the prescribed stress scenarios are typically low-probability events, *some* scenario (possibly mild) is realized every day. This makes it possible to backtest scenario analysis in an operationally relevant way: roll the clock back one day, run the scenario analysis for the scenario that actually occurred, and compare the predicted and realized P&L. Yet such backtesting is not standard practice, and, as a result, risk managers often lack a principled notion of how accurate their scenario numbers really are.

This paper argues that scenario analysis should be treated as a statistical prediction problem, and that risk managers should report *uncertainty* alongside point estimates. Concretely, let G_{t+1} denote the one-step portfolio gain (daily P&L in our experiments). Let $\mathbf{z}$ denote a stress scenario, defined in terms of shocks to selected common factors, and interpret the object of interest as the *conditional distribution* of G_{t+1} given the scenario and today’s information. Our goal is to produce, for each scenario $\mathbf{z}$ under consideration, a prediction interval $\mathcal{C}_t(\mathbf{z})$ for G_{t+1} that is (i) informative in practice (not overly conservative) and (ii) accompanied by provable coverage guarantees.

Our first method, *Adaptive Conformal Scenario Analysis* (ACSA), wraps a scenario-wise quantile predictor in an online calibration procedure. It updates the quantile level according to whether

the preceding interval covered the realized P&L in the realized scenario, widening intervals when empirical coverage is too low and shrinking them when it is too high. The resulting intervals enjoy a long-run *marginal* coverage guarantee across the realized scenario sequence, under mild requirements on the predictor and without a correct parametric specification. This provides a practical feedback mechanism for calibrating scenario uncertainty when the underlying prediction model is imperfect.

Marginal calibration does not ensure adequate coverage in every scenario. A procedure can achieve the correct average coverage while under-covering extreme regions, where losses may be large and nonlinear exposures are particularly important. Our second method, *Kernel Scenario Analysis* (KSA), estimates scenario-conditional quantiles by reweighting historical observations according to their similarity to the current market state and hypothetical factor stress, using a kernel-based notion of distance. Scenarios are defined in terms of a low-dimensional set of common-factor moves, while today’s market state supplies additional conditioning information about tomorrow’s distribution.

ACSA and KSA offer two complementary ways of improving scenario analysis. ACSA adjusts for systematic miscalibration as new outcomes are observed, while KSA uses the structure of the current market state and the proposed stress to distinguish between scenarios with different levels of uncertainty. They can be applied independently, or combined when both forms of adaptation are desirable.

The paper makes four contributions:

- We formulate scenario analysis as a statistical prediction problem and augment scenario tables with prediction intervals and backtestable coverage guarantees.
- We adapt conformal calibration to scenario tables through ACSA and obtain long-run marginal coverage across realized scenarios without requiring a correctly specified parametric model.
- We develop KSA, which estimates scenario-conditional quantiles by jointly conditioning on the specified scenario and the current market state. We establish a theoretical bound on its conditional coverage error under some regularity assumptions.
- Through numerical studies based on S&P 500 option portfolios, we demonstrate more reliable and informative scenario assessments than standard scenario analysis, including settings where dependence, convexity, and adversarial exposures make its point estimates misleading.

The remainder of this paper is organized as follows. Section 2 reviews related work. Sections 3–4 formulate the problem and develop the three-factor motivating example, which serves as a controlled testbed for the methods. Sections 5 and 6 present ACSA and KSA, respectively, while Section 7 reports numerical experiments. The E-Companion contains proofs, implementation details, and a summary of our notation may be found in E-Companion EC.1.1.

2 Literature Review

Much of financial risk management concerns scalar measures such as value-at-risk (VaR), conditional value-at-risk (CVaR), and coherent risk measures; see [McNeil et al. \(2015\)](#). [Wang and Ziegel \(2021\)](#) study risk measures based on loss distributions under multiple scenarios whereas our focus is on prediction intervals for portfolio gains within each specified stress scenario. Following the [European Central Bank \(2006\)](#), scenarios may be *historical* by replaying realized market moves on today’s portfolio, *hypothetical* by applying a specified narrative, *probabilistic* in that they are generated from a joint model fitted to historical data, or *reverse-engineered* where we work backwards from a portfolio-outcome threshold to find scenarios likely to exceed it. These categories often overlap. The empirical Bayes approach of [Glasserman, Kang, and Kang \(2015\)](#), for example, combines reverse engineering with probabilistic structure calibrated to historical data, while hypothetical narratives are commonly informed by historical frequencies and co-movement patterns. Our work also uses historical data to learn scenario-wise behavior, but we apply nonparametric methods rather than fitting a parametric joint distribution, and we provide asymptotic coverage guarantees for scenario-gain uncertainty statements.

The [Basel Committee on Banking Supervision \(2005\)](#) guidance emphasizes that scenarios should be *plausible*, *severe*, and *suggestive* of risk-mitigating actions. One way to balance plausibility and severity is to define a region of plausible risk-factor moves and search for the most adverse portfolio outcome within it. [Breuer, Jandacka, Rheinberger, and Summer \(2009\)](#) develop this idea and study *partial-scenario* stressing: they maximize plausibility by setting unstressed systematic factors to their conditional means given the stressed factors. [Haugh and Ruiz Lacedelli \(2020\)](#) propagate the full conditional *distribution* of unstressed factors, which matters for derivatives portfolios because of convexity. A related asset-allocation perspective appears in Black–Litterman ([Black and Litterman, 1991](#)) and its extensions ([Meucci, 2008](#)), which compute risk-factor distributions conditional on a specified view, typically a linear constraint in a Gaussian setting.

Other approaches to scenario selection and structuring include graphical models for scenario selection and conditional propagation ([Rebonato, 2010, 2019](#)), systematic search over a plausible risk-factor distribution ([Flood and Korenko, 2015](#)), hypothetical scenario design ([Golub et al., 2018](#)), and scenario aggregation under model uncertainty ([Cambou and Filipović, 2017](#)). Additional contributions include [Alfaro and Drehmann \(2009\)](#); [Quagliariello \(2009\)](#); [Bonti et al. \(2006\)](#), while [Borio et al. \(2012\)](#) review practice and policy considerations. Recent work also uses deep generative models for financial scenario simulation. [Cont, Cucuringu, Xu, and Zhang \(2026\)](#) train a generative adversarial network to reproduce value-at-risk and expected-shortfall characteristics of static portfolios and dynamic trading strategies, while [Chen, Xu, Xu, and Zhang \(2025\)](#) incorporate latent factor structure into diffusion models for high-dimensional asset-return generation.

Factor structure underlies many probabilistic scenario models. Equity examples include the capital asset pricing model (CAPM) and Fama–French three-factor model [Fama and French \(1993\)](#). For fixed income, [Diebold and Li \(2006\)](#) propose a three-factor yield-curve model, while [Saroka](#)

and Rebonato (2014) use principal component analysis (PCA) to identify likely yield-curve moves consistent with a specified view. For equity derivatives, Cont and Fonseca (2002) develop a PCA model of implied volatility surfaces. Since financial factors exhibit temporal dependence, time-series components are also needed; see e.g. Haugh and Ruiz Lacedelli (2020). Pelger and Xiong (2022) combine nonparametric estimation with PCA to estimate factor models whose loadings vary with the economic state. We refer to McNeil et al. (2015), Ruppert and Matteson (2015) and Tsay (2010) for textbook treatments and further references.

Our methodology draws on quantile regression (Koenker, 2005), uncertainty quantification, and nonparametric statistics. Conditional quantiles provide prediction intervals whose coverage can be studied conditionally or averaged over realizations of the conditioning variables. Early analyses provide conditional consistency under restrictive assumptions (Vovk et al., 2005; Takeuchi et al., 2006; Steinwart and Christmann, 2011). More recent work, motivated partly by flexible predictors such as deep networks, emphasizes marginal guarantees and evaluates conditional performance empirically. Major streams include conformal prediction (Angelopoulos and Bates, 2021; Papadopoulos et al., 2002; Lei et al., 2018) and uncertainty calibration (Roth, 2022). For non-exchangeable panel data, Tu and Giesecke (2026) combine history-based similarity weights with adaptive miscoverage updates, using contemporaneous outcomes from other units for calibration.

Kernel methods have also been used to learn dynamic portfolio values from simulated cash flows (Boudabsa and Filipović, 2022). Kernel conditional-quantile estimation for dependent data is well established. Cai (2002) estimates time-series regression quantiles by inverting a weighted Nadaraya–Watson conditional CDF estimator, while Hansen (2008) develops uniform convergence theory for kernel estimators under strong mixing. More recently, Lee et al. (2026) use reweighted Nadaraya–Watson estimation of nonconformity-score quantiles for time-series prediction and establish a conditional-coverage-gap bound under mixing assumptions. Our analysis provides an explicit finite-sample bound on scenario-conditional coverage error, uniformly over a compact operational feature region, under the stated dependence and predictive-sufficiency assumptions.

Our contribution is the combination of these statistical ingredients with a financial scenario-analysis formulation. KSA conditions jointly on a hypothetical stressed-factor scenario and the current market state to produce an entire scenario table. Its residual-centering construction is compatible with security-level conditional-mean estimation and practical aggregation across changing portfolios. The latter is an implementation advantage, whereas Theorem 6.1 concerns a fixed portfolio and deterministic centering function. The theorem gives a uniform conditional-coverage bound over a compact operational feature region under its stated assumptions. ACSA adapts conformal calibration for time series (Gibbs and Candes, 2021; Gibbs and Candès, 2024) to scenario tables, and KSA can supply its quantile predictor. The financial-risk interpretation and adversarial-portfolio application further distinguish the resulting framework from ordinary one-step time-series prediction.

Evaluation and ongoing validation provide a further connection to financial risk forecasting. Proper scoring rules reward calibration and sharpness (Gneiting and Raftery, 2007); quantile forecasts can

be assessed using pinball loss, while joint scoring of VaR and expected shortfall (ES) enables comparison of tail-risk forecasts (Fissler et al., 2016; He et al., 2022). Nolde and Ziegel (2017) advocate comparative backtests alongside traditional pass/fail tests to encourage accurate risk forecasts. Classic VaR backtests examine coverage and independence of exceedances, often in fixed samples. Sequential monitoring instead seeks prompt detection of forecast deterioration as observations arrive (Hoga and Demetrescu, 2023) while e-backtesting provides sequentially valid tests (Wang et al., 2026). Related work develops model-independent and conditional backtests for ES (Acerbi and Szekely, 2014; Du and Escanciano, 2017). Although these approaches concern scalar risk forecasts, their emphasis on statistical validation also applies to scenario tables. Our target is scenario-wise predictive uncertainty for portfolio gains, evaluated through interval scores and backtesting, with an online perspective suited to forecast accuracy that varies over time.

3 Preliminaries and “Standard” Scenario Analysis

In this section, we formulate the scenario analysis problem. We begin in Section 3.1 with a motivating portfolio setting where scenario analysis can be applied. In Section 3.2 we introduce the factor model structure that typically underlies a scenario analysis. Then in Section 3.3 we describe how a so-called standard scenario analysis works. Finally, in Section 3.4 we describe the main weaknesses of standard scenario analysis.

3.1 A Motivating Portfolio

We assume we have a fixed portfolio of securities that will not change over the horizon under consideration, e.g., 1 day. In principle the portfolio could include any combination of securities from any asset class, although in practice we are typically limited to reasonably liquid securities for which historical price data are available. Specifically, consider a U.S. equity and options portfolio. The time t portfolio value V_t satisfies

$$V_t = \sum_{i=1}^n u_i S_t^{(i)} + \sum_{j=1}^M u_j^d C_t^{(j)}, \quad (1)$$

where the first sum in (1) represents holdings (the u_i ’s) in domestic stocks (with time t prices $S_t^{(i)}$), and the second sum represents holdings (u_j^d ’s) in M options (with prices $C_t^{(j)}$) of various strikes, maturities and underlying securities. We might assume a simple CAPM-style factor model for the domestic market so that

$$R_{t+1}^{(i)} := \log(S_{t+1}^{(i)}/S_t^{(i)}) = r_f + \beta_i(R_{t+1}^m - r_f) + \epsilon_{t+1}^{(i)}, \quad (2)$$

where R_{t+1}^m is the domestic market log return between dates t and $t+1$, r_f the domestic (and assumed constant) risk-free rate, and the $\epsilon_{t+1}^{(i)}$ ’s are independent random noise. This makes R_{t+1}^m

the only common equity factor return. We may also assume that there are three volatility factors for the market, and that the market volatility model also drives the volatility surface for the individual stocks. This means there are a total of four common factors. In our numerical experiments in Section 7 we will use this setting with just a single underlying security.

3.2 The Factor Model Structure

We now provide a more formal description of the market structure and present the notation that we will use throughout the remainder of the paper. This market structure serves only as a vehicle for describing how a (standard) scenario analysis proceeds in our numerical experiments, but not for the presentation of our proposed algorithms. Knowing the correct market structure, including the factor model and dynamics, is *not required* for the ACSA algorithm of Section 5. Similarly, while the KSA algorithm of Section 6 makes stronger assumptions than the ACSA algorithm and incorporates some knowledge of market factors, it does *not require* knowledge of the precise factor dynamics.

Let V_t denote the time t value of the portfolio. The *portfolio gain* (also referred to as the portfolio profit and loss, P&L for short) at time $t + 1$ is $G_{t+1} := V_{t+1} - V_t$. In our financial context, time is measured in days so that the portfolio gain would then be a daily P&L. Assume V_t is $\mathcal{F}_t$ -adapted where $\mathbb{F} := \{\mathcal{F}_t\}_{t \geq 0}$ denotes the filtration generated by all relevant information (e.g. security prices, risk factors, etc) in the market. We use bold font and lower case to denote vectors, bold font and upper case to denote matrices, and non-bold font lower case to denote scalars. All vectors are assumed to be column vectors unless otherwise stated. A superscript (i) denotes the i^{th} component of a vector, and a superscript $(i:j)$ refers to the sub-vector comprising elements from index i to index j . For a matrix, a superscript $(i:j,:)$ denotes the sub-matrix containing rows i through j , and $(:,i:j)$ denotes the sub-matrix containing columns i through j . More generally, a superscript $(i_1:i_2,j_1:j_2)$ denotes the sub-matrix formed by rows i_1 through i_2 and columns j_1 through j_2 .

As is standard in the risk management literature (see e.g. McNeil et al. (2015)), we will assume the portfolio value V_t is a function of $\mathcal{F}_t$ -adapted **risk factors**, whose time t values we denote by the vector $\mathbf{x}_t \in \mathbb{R}^n$ for some n . For the equity and options portfolio in Section 3.1, we take $\mathbf{x}_t$ to consist of the stock prices $S_t^{(i)}$'s and the M implied volatilities that are required to compute the option prices $C_t^{(j)}$'s. In any case it follows that $V_t = v(\mathbf{x}_t)$ for some known pricing function $v : \mathbb{R}^n \rightarrow \mathbb{R}$, with the portfolio gain given by

$$G_{t+1} = v(\mathbf{x}_{t+1}) - v(\mathbf{x}_t) = \sum_{i=1}^{n_s} u_i (P_{t+1}^{(i)} - P_t^{(i)}) \quad (3)$$

where n_s is the total number of securities in the portfolio, u_i is the number of units of the i^{th}

security¹, and $P_t^{(i)} = P_t^{(i)}(\mathbf{x}_t)$ is² the time t price of the i^{th} security. We generally suppress the dependence of $P_t^{(i)}$ on $\mathbf{x}_t$ but note that it typically depends on just a few (and often just one) components of $\mathbf{x}_t$.

In financial risk management, it is common to employ a **common factor model**, in which a set of possibly latent factors $\mathbf{f}_t$ drives the dynamics of observed risk factors $\mathbf{x}_t$. A typical specification is

$$\Delta \mathbf{x}_{t+1} = \boldsymbol{\mu} + \mathbf{B} \Delta \mathbf{f}_{t+1} + \boldsymbol{\xi}_{t+1}, \quad t = 1, 2, \dots, \quad (4)$$

where $\mathbf{x}_t \in \mathbb{R}^n$ is the vector of **risk factors**, with $\Delta \mathbf{x}_{t+1} = \mathbf{x}_{t+1} - \mathbf{x}_t$, and $\boldsymbol{\mu} \in \mathbb{R}^n$ is its per-period drift. The vector $\mathbf{f}_t \in \mathbb{R}^m$ contains the **common factors**, with $\Delta \mathbf{f}_{t+1} = \mathbf{f}_{t+1} - \mathbf{f}_t$. The factor loading matrix is $\mathbf{B} = [\mathbf{b}_1 \cdots \mathbf{b}_m] \in \mathbb{R}^{n \times m}$, with $\mathbf{b}_i \in \mathbb{R}^n$ denoting its i^{th} column. Finally, $\{\boldsymbol{\xi}_t\}_{t=1,2,\dots}$ is an i.i.d. sequence of zero-mean random vectors in $\mathbb{R}^n$, representing idiosyncratic shocks that are assumed to be independent of all other variables. In what follows, for any time series-based variable, we always use Δ to denote its change from time t to $t + 1$. The equity and options setting above can be modeled according to (4).

3.3 Standard Scenario Analysis

A fundamental goal of risk management is to understand the distribution of the portfolio gain G_{t+1} or portfolio return R_{t+1} conditional on $\mathcal{F}_t$. Since computing or estimating the distribution is often too challenging in practice, we often aim instead to compute some functional of the distribution, such as the portfolio's value-at-risk (VaR) or conditional value-at-risk (CVaR). Alternatively, we may wish to conduct a scenario analysis (the topic of this paper) where the goal is to understand the distribution of G_{t+1} conditional on $\mathcal{F}_t$ and specific time $t + 1$ risk scenarios. In a standard scenario analysis (hereafter SSA), the risk manager would identify various stresses to apply to the changes (i.e., the returns) in the common factors $\mathbf{f}_t$. We can define a scenario by jointly stressing any number $k \leq m$ of the common factor returns. In the case of a portfolio of futures and options on the S&P 500 in the setting of Section 3.1 (with a single underlying security), these stresses might include shifts to the value of the underlying, i.e. the S&P 500, as well as some combination of parallel³ shifts to the implied volatility surface and a steepening or flattening of the skew (or term) structure of the implied volatilities. For example, suppose the first component of $\mathbf{x}_t$ (i.e., $x_t^{(1)}$) represents the log S&P 500 spot price, and the other components of $\mathbf{x}_t$ represent the implied

¹We omit the dependence of the time t portfolio holdings on t since they are assumed to be constant in any scenario analysis backtest. For the same reason, we also omit the dependence of the portfolio gain in (3) on the portfolio holdings.

²It is tempting to write $P_t^{(i)} = P^{(i)}(\mathbf{x}_t)$ for some fixed pricing function $P^{(i)}$ but we maintain the dependence on t as many securities have a fixed maturity and so given $\mathbf{x}_t$, security prices will still in general depend on t .

³A parallel shift may not be the best choice of stress for an implied volatility surface as it is well known that shorter maturity implied volatilities tend to move a lot more than longer maturity ones. It would, of course, be easy to define a common factor return that handles this phenomenon. It is perhaps worth emphasizing at this point, however, that we are agnostic as to how common factors and associated stresses / scenarios are defined. Rather, our goal in standard scenario analysis is to estimate the (expected) gains *given* these scenarios.

volatilities for the options that appear in the portfolio with various strike-maturity combinations. If $f_t^{(1)}$ also represents the log S&P 500 spot price, then $\mu^{(1)} = 0$, $\mathbf{b}_1 = (1, 0, \dots, 0)$ and $\xi_t^{(1)} = 0$ w.p. 1. Similarly, if the changes in $f_t^{(2)}$ correspond to the parallel shifts to the implied volatility surface, then the second column of $\mathbf{B}$ would be $\mathbf{b}_2 = (0, 1, \dots, 1)$. We can now consider a scenario where the returns of $f_t^{(1)}$ and $f_t^{(2)}$ are simultaneously stressed. For example, a scenario of interest might be one where $\Delta f_{t+1}^{(1)} = -5\%$ and $\Delta f_{t+1}^{(2)} = +10$, corresponding to a 5% fall in the S&P 500 spot price and a 10 volatility point increase across its entire implied volatility surface. The SSA approach computes the portfolio gain via (3) with $\mathbf{x}_{t+1}$ determined by (4) and the scenario setting: $\Delta f_{t+1}^{(1)} = -5\%$, $\Delta f_{t+1}^{(2)} = +10$, $\Delta f_{t+1}^{(3)} = \Delta f_{t+1}^{(4)} = 0$ and $\boldsymbol{\xi}_{t+1} = \mathbf{0}$.

In practice, multiple two- or even three-dimensional tables of the estimated portfolio gain can be computed, corresponding to the simultaneous stressing of $k = 2$ or $k = 3$ different common factor returns. It is important to emphasize, especially in the context of portfolios containing derivatives, that the typical risk / portfolio manager employing SSA does not have an explicit model like (4) at hand nor does she need one. Indeed, most of the modeling effort in the derivatives space goes into developing dynamic pricing models under an equivalent martingale measure $\mathcal{Q}$ and then developing numerical procedures for computing prices and the so-called Greeks within these models. Furthermore, it may be the case that only a subset of factors, say the first $l \leq m$, are ever considered for stressing. In that case, SSA works with a “model” of the form

$$\Delta \mathbf{x}_{t+1} = \boldsymbol{\mu} + \mathbf{B}^{(:,1:l)} \Delta \mathbf{f}_{t+1}^{(1:l)}. \quad (5)$$

The important feature of (5) is that noise terms, i.e., the $\boldsymbol{\xi}_{t+1}$ ’s, and probability distributions / factor dynamics are not specified and in fact play no role in it. It therefore follows that (5) is *not* a probabilistic model for the risk factor $\mathbf{x}_{t+1}$, and this is typically the case in SSA applications.

3.3.1 The Observability of Common Factor Returns

At this point, we emphasize that throughout the paper, both $\mathbf{x}_t$ and $\mathbf{f}_t$ are assumed to be $\mathcal{F}_t$ -adapted and therefore observable at time t . This is true for $\mathbf{x}_t$ since we can view $\mathbf{x}_t$ as simply an alternative way to quote security prices, e.g., in terms of bond yields, or log prices and implied volatilities. We also assume that $\mathbf{f}_t$ is observable since the stress scenarios will be defined in terms of $\mathbf{f}_t$, and it is necessary to define scenarios and also recognize what actual scenario transpired on any given day. That said, in some applications, a possible objection to this assumption is that some common factors $\mathbf{f}_t$ might be latent. For example, in so-called cross-sectional or statistical factor models, e.g. [Ruppert and Matteson \(2015\)](#), the factors are assumed to be latent and need to be estimated. Clearly, it would be very difficult to evaluate a scenario analysis if some or all of the common factors that are used to define scenarios are latent, since then one can never tell for sure which scenario actually transpired.

An easy solution to this problem is to work with observable estimates or proxies of the common

factors and to define scenarios by stressing the returns of these estimates / proxies. For the implied volatility surface of the S&P 500 index, we can easily define observable proxies for any latent common factors. For example, we can easily use the observed implied volatilities to define the average level of implied volatility, the skew, and the steepness of the term-structure of volatilities, respectively. Hence, the changes in these common factor proxies between t and $t + 1$ will be observable. In the remainder of this paper we will therefore assume all common risk factors are observable.

3.4 The Problems with Standard Scenario Analysis

Before proceeding, we point out several weaknesses of SSA which of course relate to the fact that (5) is not a probabilistic model for the risk factors $\mathbf{x}_{t+1}$. Some of these weaknesses were also highlighted by [Haugh and Ruiz Lacedelli \(2020\)](#).

1. SSA only produces a point estimate - typically interpreted as an expectation - of the portfolio gain in a given scenario. But an expectation only gives a partial view on the gain risk in a given scenario and it would be far more informative to provide a prediction interval for the realized scenario gain.
2. Let $\mathbf{s} = \{s_1, \dots, s_k\}$ where $1 \leq s_1 < s_2 \leq \dots < s_k \leq l$, and let $\mathbf{f}_{t+1}^{(\mathbf{s})} := (f_{t+1}^{(s_1)}, \dots, f_{t+1}^{(s_k)})$ denote the subset of common factors whose returns are stressed in a given scenario. SSA implicitly assumes that the unstressed common factors are unchanged between t and $t+1$, i.e., $\Delta \mathbf{f}_{t+1}^{(\mathbf{s}^c)} = \mathbf{0}$ (we use $\mathbf{s}^c$ to denote the complement of $\mathbf{s}$, i.e. the indices of the non-stressed common factors in the scenario). But this is not justified and can often lead to a very inaccurate estimation for the given scenario.
3. An obvious solution to the problem raised in the previous point would be to set the unstressed factors $\mathbf{f}_{t+1}^{(\mathbf{s}^c)}$ equal to their conditional expectation $\mathbb{E}_t[\mathbf{f}_{t+1}^{(\mathbf{s}^c)} \mid \mathbf{f}_{t+1}^{(\mathbf{s})}]$ ⁴ when estimating the portfolio gain. This should certainly be an improvement over SSA, e.g. see [Haugh and Ruiz Lacedelli \(2020\)](#), but it ignores the uncertainty in ξ_{t+1} and $\mathbf{f}_{t+1}^{(\mathbf{s}^c)} \mid (\mathcal{F}_t, \mathbf{f}_{t+1}^{(\mathbf{s})})$. This uncertainty may be significant, particularly for portfolios containing securities whose values depend non-linearly on $\mathbf{x}_{t+1}$. But even if we ignore this uncertainty, setting $\mathbf{f}_{t+1}^{(\mathbf{s}^c)} = \mathbb{E}_t[\mathbf{f}_{t+1}^{(\mathbf{s}^c)} \mid \mathbf{f}_{t+1}^{(\mathbf{s})}]$ is not a straightforward task as it requires a model for estimating these conditional expectations.
4. SSA is rarely if ever accompanied by any sort of backtesting or statistical validation. Perhaps the main reason for this is that each of the scenarios considered by an SSA is (typically) a zero probability event, and none of the considered scenarios will have actually occurred on day $t + 1$. However, this problem is easy to resolve. Specifically, on day $t + 1$ we could “see” exactly what scenario has just transpired. For example, we could see what the return in the

⁴We use $\mathbb{E}_t[\cdot]$ throughout to denote expectations that are conditional on $\mathcal{F}_t$. Hence $\mathbb{E}_t[\cdot \mid \mathbf{f}_{t+1}^{(\mathbf{s})}]$ denotes expectations that are conditional on $\mathcal{F}_t$ and $\mathbf{f}_{t+1}^{(\mathbf{s})}$.

S&P 500 and what the parallel change in the implied volatility surface were over the period $[t, t + 1]$. We could then roll time back one day, rerun the scenario analysis for exactly this realized scenario and then compare the estimated and realized portfolio gains.

5. Another weakness with SSA is that it relies on a reduced-form model such as (5). But if a portfolio is constructed with a view to (i) being neutral relative to all the common factor returns that are ever considered for stressing but (ii) highly exposed to non-stressed common factor returns, then an SSA will drastically underestimate the risk of the portfolio. This, of course, is related to other points raised above but the problem here is that the risk manager may not even be aware of the non-stressed common factors, whereas the trader or portfolio manager is aware of them. We refer to this as *adversarial portfolio selection*. A more benign possibility is simply that one or more of the common factors are misspecified, which can also lead to reported expected scenario gains that are misleading.

Haugh and Ruiz Lacedelli (2020) addressed some of these weaknesses by embedding the scenario analysis within a dynamic factor model for the underlying risk factors. Their approach required multivariate state-space models for modeling the time-series behavior of (sometimes latent) risk factors, and that were sufficiently tractable as to enable the computation (possibly via simulation) of the conditional distribution of unstressed risk factors. They showed how SSA and their approach could lead to dramatically different results, particularly in the case of adversarial portfolio selection. They adopted a parametric approach, however, which meant their reported expected scenario gains were only accurate to the extent that their modeling was correct. In contrast, in this paper we propose a more holistic approach to scenario analysis that does not rely on the correct specification of a parametric model defining the risk factor and common risk factor dynamics. In addition, a key feature of our framework is the provision of coverage guarantees for the prediction intervals we report for our portfolio gains.

4 A Motivating Example: A 3-Factor Log-Linear Equity Model

In this section, we present a simple motivating example that illustrates how inaccurate SSA can be. We then revisit this example in Sections 5 and 6, where we show how ACSA and KSA produce calibrated prediction intervals for the realized gain under each scenario.

Market Dynamics We assume the market consists of a single non-dividend paying equity index, e.g. the S&P 500, with time t price S_t and daily return dynamics that are IID and satisfy

$$\log \left(\frac{S_{t+1}}{S_t} \right) = \delta \cdot \left(\mu_\delta - \frac{\sigma_\epsilon^2}{2} \right) + \beta^\top \mathbf{F}_{t+1} + \epsilon_{t+1}, \quad (6)$$

where $\delta = 1/252$ denotes the length in years of one trading day, $\mathbf{F}_{t+1} \sim \mathcal{N}(\mathbf{0}, \delta \boldsymbol{\Sigma})$ is a 3-dimensional observable factor return, with $\boldsymbol{\Sigma}$ denoting its annualized covariance matrix, and $\epsilon_{t+1} \sim \mathcal{N}(0, \delta \sigma_\epsilon^2)$

is an idiosyncratic random noise variable with annualized volatility σ_ϵ . The $-\sigma_\epsilon^2/2$ term in (6) reflects the usual Itô correction associated with a geometric Brownian motion type specification.⁵ The common-factor contribution $\boldsymbol{\beta}^\top \mathbf{F}_{t+1}$ is specified in log-return form. Accordingly, we interpret μ_δ as the annualized drift parameter appearing in (6). We interpret the three factor returns as *oil*, *interest rate*, and *credit spread* shocks. We take $\mu_\delta = 0.03$, and the factor loadings $\boldsymbol{\beta} =: (\beta_1, \beta_2, \beta_3)$, together with their interpretations, are given in Table 1.

Factor	Daily observable shock	β_k
Oil	% daily change in Brent/WTI spot	-0.20
Rate	Daily change in 10y U.S. yield (bp)	-0.15
Credit	Daily change in ICE/BofA US HY OAS (bp)	-0.30

Table 1: The three factor loadings and their interpretation.

We note that (6) agrees with the model structure specified in (4), with $\mathbf{x}_t = \log S_t$, $\boldsymbol{\mu} = \delta \cdot \left(\mu_\delta - \frac{\sigma_\epsilon^2}{2} \right)$, $\mathbf{B} = \boldsymbol{\beta}^\top$, $\Delta \mathbf{f}_{t+1} = \mathbf{F}_{t+1}$ and $\boldsymbol{\xi}_{t+1} = \epsilon_{t+1}$. It follows from (6) that the T -day return satisfies

$$\log \left(\frac{S_T}{S_0} \right) = T \cdot \delta \left(\mu_\delta - \frac{\sigma_\epsilon^2}{2} \right) + \boldsymbol{\beta}^\top \mathbf{F}_{1:T} + \epsilon_{1:T} \quad (7)$$

where $\mathbf{F}_{1:T} = \sum_{t=1}^T F_t \sim \mathcal{N}(\mathbf{0}, T \cdot \delta \boldsymbol{\Sigma})$ is the T -day factor return and $\epsilon_{1:T} = \sum_{t=1}^T \epsilon_t \sim \mathcal{N}(0, T \cdot \delta \sigma_\epsilon^2)$. We assume $S_0 = 5000$ and the annual covariance matrix (corresponding to taking $T = 252$

days so that $T \cdot \delta = 1$) is given by $\boldsymbol{\Sigma} = \begin{bmatrix} 0.0900 & 0.0009 & -0.0240 \\ 0.0009 & 0.0001 & -0.0016 \\ -0.0240 & -0.0016 & 0.1600 \end{bmatrix}$. This in turn implies

correlations $\rho_{\text{Oil,Rate}} = 0.30$, $\rho_{\text{Oil,Credit}} = -0.20$ and $\rho_{\text{Rate,Credit}} = -0.40$. Finally, we take $\sigma_\epsilon = 18\%$ so that the total annual volatility is

$$\sigma_{\text{tot}} = \sqrt{\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta} + \sigma_\epsilon^2} \approx 21.8\%. \quad (8)$$

Throughout this example, $\boldsymbol{\Sigma}$ and σ_ϵ^2 denote annualized covariance and variance parameters, respectively. One-day variances are obtained by multiplying the corresponding annualized quantities by $\delta = 1/252$.

Portfolio Our portfolio consists of a single short position in a European call option on the index with a time to maturity of one month (i.e., $\tau = 21/252$) and a fixed moneyness (the strike price divided by the underlying asset's price) $m = 1.05$, i.e., the call option is 5% out-of-the-money. Given that $S_0 = 5000$, the strike price would be $K = 1.05 \times S_0 = 5,250$. We take the continuously compounded interest rate to be $r = 4\%$ so that the initial value of the option is then given by the

⁵For a geometric Brownian motion satisfying $\frac{dS_t}{S_t} = \mu dt + \sigma dW_t$, Itô's lemma gives $d \log S_t = (\mu - \sigma^2/2) dt + \sigma dW_t$.

Black-Scholes formula and equals \$44.2. (The use of the Black-Scholes formula to price options is consistent with the log-normal dynamics of S_T as evidenced by (7).)

Scenario Analysis We consider stress testing the portfolio over a single trading day by considering joint stresses to the oil and rate factors. In particular, we consider percentage shocks of size $\{-20, -12, -6, 0, +6, +12, +20\}$ to the oil factor and of size $\{-30, -20, -10, 0, +10, +20, +30\}$ b.p.'s to the rate factor. We use $\mathbf{z} = (z_{\text{Oil}}, z_{\text{Rate}})$ to denote a scenario so, for example, $\mathbf{z} = (+12\%, -20 \text{ b.p.})$ denotes a stress of +12% to the oil factor return and -20 b.p. to the rate factor return. While some of these stresses are considerably more extreme than would be implied by their 1-day volatilities, scenario analysis stresses in practice are often much larger than those implied by the size of typical daily moves. For example, oil price shocks of -20% occurred in 2020 during the COVID crash, and +15% days occurred on days of OPEC supply cuts. Similarly, ± 30 b.p. moves in 10-year yields can occur when the U.S. Federal Reserve surprises the market, and ± 10 b.p. moves are quite common on FOMC days.

Computing the Conditional 1-Day Log Return Distribution The credit factor is not stressed in any scenario, and its conditional mean and annualized conditional variance in a given scenario $\mathbf{z}$ are

$$\begin{aligned}\mu_{\text{Cred}|s} &= \Sigma_{\text{Cred},s} \Sigma_{ss}^{-1} \mathbf{z} \\ \sigma_{\text{Cred}|s}^2 &= \Sigma_{\text{Cred},\text{Cred}} - \Sigma_{\text{Cred},s} \Sigma_{ss}^{-1} \Sigma_{s,\text{Cred}}\end{aligned}\tag{9}$$

where Σ is partitioned according to $\Sigma = \begin{bmatrix} \Sigma_{ss} & \Sigma_{s,\text{Cred}} \\ \Sigma_{\text{Cred},s} & \Sigma_{\text{Cred},\text{Cred}} \end{bmatrix}$. We note that $\mu_{\text{Cred}|s}$ and $\sigma_{\text{Cred}|s}$ depend on the stress in distinct ways: $\sigma_{\text{Cred}|s}$ is determined by the *choice* of stressed factors (i.e., the subset s of common factors), whereas $\mu_{\text{Cred}|s}$ also reflects the specific stress vector $\mathbf{z}$ applied to those factors. Let $\mathbf{F}_s = (F_{\text{Oil}}, F_{\text{Rate}})$ be the stressed pair⁶ of factors and F_{Cred} the unstressed credit-spread shock. Then

$$\log(S_{t+1}/S_t) \Big| (\mathbf{F}_s = \mathbf{z}) \sim \mathcal{N}(m_{\mathbf{z}}, v_{\text{Cred}}),\tag{10}$$

where

$$m_{\mathbf{z}} = \delta \left(\mu_{\delta} - \frac{\sigma_{\epsilon}^2}{2} \right) + \beta_{\text{Oil}} z_{\text{Oil}} + \beta_{\text{Rate}} z_{\text{Rate}} + \beta_{\text{Cred}} \mu_{\text{Cred}|s}, \quad v_{\text{Cred}} = \delta \cdot \left(\beta_{\text{Cred}}^2 \sigma_{\text{Cred}|s}^2 + \sigma_{\epsilon}^2 \right).\tag{11}$$

Repricing the Short Option Position We can price the portfolio option using the Black-Scholes formula in every scenario, i.e., assuming a continuously compounded risk-free interest rate

⁶We allow for a slight abuse of notation here. In particular, $\mathbf{F}_s \in \mathbb{R}^2$ denotes the two stressed factor returns over a single day while we have also used $\mathbf{F}_t \in \mathbb{R}^3$ to denote the random factor return vector between times $t-1$ and t .

of $r = 4\%$ and volatility σ_{tot} . Because the scenario stresses roll time forward by 1 day, the time to maturity⁷ when repricing the options in each scenario is now 20/252.

Results Table 2 reports the estimated scenario P&L or portfolio gain (in dollars) for SSA and conditional-mean scenario analysis (CSA). The CSA entries are obtained by setting the unstressed credit factor to its mean conditional on the scenario, i.e. $\mu_{\text{Cred}|s}$ from (9). Because the idiosyncratic noise ϵ is uncorrelated with the factor noise, we set it to its mean of zero in each scenario. Finally, in Table 3 we present results for the oracle scenario analysis where the portfolio gains are obtained via (10) by integrating the Black-Scholes price of the option over the residual uncertainty of the credit factor conditional on the scenario.

Rate (bp)	SSA							CSA						
	Oil (%)							Oil (%)						
	-20	-12	-6	0	+6	+12	+20	-20	-12	-6	0	+6	+12	+20
-30	-70.7	-35.3	-14.5	1.9	14.6	24.0	32.6	-27.0	-6.2	6.1	15.9	23.5	29.4	35.0
-20	-70.3	-35.0	-14.3	2.1	14.7	24.1	32.6	-35.0	-12.5	1.0	11.9	20.4	27.0	33.4
-10	-69.9	-34.8	-14.0	2.3	14.8	24.2	32.7	-43.7	-19.3	-4.6	7.4	16.9	24.4	31.6
0	-69.6	-34.5	-13.8	2.5	15.0	24.3	32.8	-53.0	-26.7	-10.6	2.5	13.0	21.3	29.5
+10	-69.2	-34.2	-13.6	2.7	15.1	24.4	32.8	-63.0	-34.7	-17.3	-2.9	8.7	17.9	27.1
+20	-68.8	-33.9	-13.4	2.8	15.2	24.5	32.9	-73.7	-43.3	-24.5	-8.9	3.9	14.2	24.5
+30	-68.5	-33.6	-13.1	3.0	15.4	24.6	32.9	-85.0	-52.6	-32.4	-15.4	-1.4	10.0	21.5

Table 2: Scenario gains in dollars, rounded to one decimal place. The left panel reports standard scenario analysis (SSA), which sets the credit factor to zero. The right panel reports conditional-mean scenario analysis (CSA), which sets it to its scenario-conditional mean. Both set independent idiosyncratic noise to zero. Rate stresses are in basis points (bp) and oil stresses are in percent.

Rate\Oil	-20%	-12%	-6%	0%	+6%	+12%	+20%
-30 bp	-29.85	-8.73	3.84	13.94	21.91	28.08	34.03
-20 bp	-37.98	-15.10	-1.35	9.80	18.66	25.58	32.32
-10 bp	-46.73	-22.02	-7.04	5.21	15.04	22.77	30.38
0 bp	-56.11	-29.52	-13.25	0.16	11.01	19.62	28.18
+10 bp	-66.14	-37.62	-20.01	-5.38	6.55	16.10	25.69
+20 bp	-76.84	-46.34	-27.34	-11.44	1.64	12.19	22.90
+30 bp	-88.20	-55.69	-35.27	-18.04	-3.76	7.86	19.76

Table 3: Oracle Scenario Analysis. Expected value of option is computed in each scenario by integrating the Black-Scholes price of the option over the residual uncertainty of the credit factor and the idiosyncratic noise ϵ in that scenario.

Several observations are in order. First, we note there is a considerable difference in the scenario P&Ls across the three methods, illustrating just how inaccurate an SSA can be. We can interpret

⁷The short option position in the portfolio means the portfolio gains approximately 2.55 dollars at the zero-stress spot level from the one-day theta effect.

the improvement from the SSA entries to the CSA entries in Table 2 as being due to a correlation correction induced by setting the credit factor F_{Cred} to its conditional mean $\mu_{\text{Cred}|s}$ in each scenario. This correction can swing the P&L by up to approximately \$44, and generally results in less extreme scenario P&L’s because the credit factor is negatively correlated with the oil and rate factors and therefore provides a partial hedge in extreme scenarios. The differences (approximately \$1.0–\$3.2) between the CSA entries in Table 2 and Table 3 can be interpreted as a convexity correction. In particular, while the CSA entries contain the correct “delta” correction (in contrast to the SSA entries), Table 3 may be viewed as having an additional “gamma” correction by explicitly recognizing the uncertainty in the index price in each scenario. Because the portfolio is short a European option and therefore short gamma, we see that the P&L’s in Table 3 are always smaller than the corresponding CSA P&L’s in Table 2. We also mention that the (0, 0) entries are not zero for either SSA or CSA in Table 2. This is due to the “theta” effect whereby the portfolio gains when the short option position has 1 day less to maturity.

It’s worth remarking that the difference between the (inaccurate) SSA entries in Table 2 and (the accurate) Table 3 (or even between the CSA entries in Table 2 and Table 3) can be much greater depending on the portfolio in question. For example, an adversarial portfolio (as discussed in weakness # 5 of SSA in Section 3.4) could yield wildly inaccurate estimates of scenario gains when these estimates are conducted via SSA. Table 3 requires knowing the conditional distribution in each scenario (and integrating over residual uncertainty), which is infeasible in realistic markets. Sections 5-6 show how ACSA and KSA learn scenario-wise prediction intervals from data and then calibrate them via backtesting, avoiding reliance on a correct parametric specification.

5 Adaptive Conformal Scenario Analysis

In this section, we introduce our **adaptive conformal scenario analysis (ACSA)** algorithm. This algorithm resembles the adaptive conformal prediction algorithm recently proposed by [Gibbs and Candes \(2021\)](#), and we apply it with some minor modifications to our scenario analysis setting. The major advantages of the ACSA algorithm are its simplicity and its wide applicability. The main downside of the algorithm is that its theoretical guarantee is only a marginal guarantee, which is not entirely satisfactory in a risk management setting where conditional guarantees would be more desirable. This desire for an algorithm with conditional guarantees motivated our development of the **kernel scenario analysis (KSA)** algorithm in Section 6. Nevertheless, as we shall explain later, we view the two algorithms as complementary to each other.

We use $\mathbf{z} \in \mathbb{R}^k$ to represent a specific stress scenario so that, for example, if a stress $\mathbf{z}$ is applied to $\mathbf{f}_t^{(s)}$, then the resulting common factors after being stressed satisfy $\Delta \mathbf{f}_{t+1}^{(s)} = \mathbf{z}$ (i.e., $\mathbf{f}_{t+1}^{(s)} = \mathbf{f}_t^{(s)} + \mathbf{z}$). Take the equity and options portfolio in Section 3.1, for instance. If we apply the stress as specified in Section 3.3, then $\mathbf{z} = (-5\%, +10)$. We also use $\mathbf{z}_{t+1}$ for the *actual scenario* that transpired at time $t + 1$: once the realization of $\mathbf{f}_{t+1}$ is observed, the corresponding value of $\mathbf{z}_{t+1}$ is also revealed

as $\mathbf{z}_{t+1} = \mathbf{f}_{t+1}^{(s)} - \mathbf{f}_t^{(s)} = \Delta \mathbf{f}_{t+1}^{(s)}$. The ACSA algorithm (and indeed the KSA algorithm of Section 6) will take as input a miscoverage rate $\alpha \in (0, 1)$ and then aim to provide a $100 \times (1 - \alpha)\%$ prediction interval for $G_{t+1} \mid (\Delta \mathbf{f}_{t+1}^{(s)} = \mathbf{z}, \mathcal{F}_t)$ (which we generally abbreviate as $G_{t+1} \mid \mathbf{z}$) for each scenario $\mathbf{z}$ under consideration.

5.1 The ACSA Algorithm

We present ACSA in Algorithm 1. At each time t , a quantile predictor $g_t(\mathbf{z}; q)$ is used to construct a prediction interval for the next-period portfolio gain G_{t+1} under each scenario $\mathbf{z}$. The target miscoverage rate is α , but ACSA does not keep the quantile level fixed at α . Instead, it maintains an adjusted level $\tilde{\alpha}_t$ and updates it according to whether the most recent interval covered the realized portfolio gain: a miss decreases $\tilde{\alpha}_t$ and therefore tends to widen subsequent intervals, whereas successful coverage increases $\tilde{\alpha}_t$ and tends to shrink them.

For any adjusted level $\tilde{\alpha} \in \mathbb{R}$, define the prediction set by

$$\mathcal{C}_t(\mathbf{z}; \tilde{\alpha}) = \begin{cases} \mathbb{R}, & \tilde{\alpha} \leq 0, \\ [\min\{g_t(\mathbf{z}; \tilde{\alpha}/2), g_t(\mathbf{z}; 1 - \tilde{\alpha}/2)\}, \max\{g_t(\mathbf{z}; \tilde{\alpha}/2), g_t(\mathbf{z}; 1 - \tilde{\alpha}/2)\}], & 0 < \tilde{\alpha} < 1, \\ \emptyset, & \tilde{\alpha} \geq 1. \end{cases} \quad (12)$$

The boundary cases in (12) ensure certain coverage when $\tilde{\alpha} \leq 0$ and certain miscoverage when $\tilde{\alpha} \geq 1$, as required by the proof of Theorem 5.1. If the response is known to lie almost surely in a finite interval, that interval can replace $\mathbb{R}$ without changing the coverage indicator. In our numerical experiments, when $\tilde{\alpha}_t \leq 0$, we replace $\mathbb{R}$ by the empirical range of historical portfolio gains recomputed under the current portfolio configuration⁸.

Apart from the mild regularity conditions stated in Assumption 1 below, the quantile predictor g_t in (12) is otherwise unrestricted. It may be obtained from linear or nonlinear quantile regression, a neural network, or the KSA procedure introduced later in Section 6. The full procedure is summarized in Algorithm 1.

When $0 < \tilde{\alpha}_t < 1$, the two endpoints in (12) correspond to estimated conditional quantiles of G_{t+1} under scenario $\mathbf{z}$. If g_t were perfectly calibrated, the resulting interval would therefore have approximately $100(1 - \tilde{\alpha}_t)\%$ conditional coverage. ACSA does not require this perfect calibration to be true. Its purpose is precisely to compensate for systematic miscalibration in the underlying predictor. To see the feedback mechanism, suppose the interval reported at time $t - 1$ fails to cover G_t . Then $\text{err}_t = 1$ and $\tilde{\alpha}_t = \tilde{\alpha}_{t-1} - \gamma(1 - \alpha)$, so the adjusted miscoverage level decreases and the next interval will typically be wider. Conversely, if the previous interval covers G_t , then $\text{err}_t = 0$ and $\tilde{\alpha}_t = \tilde{\alpha}_{t-1} + \gamma\alpha$, which tends to make the next interval narrower. Thus ACSA continually uses realized coverage errors to correct the quantile level supplied to g_t . For example, suppose intervals

⁸Since this empirical range is not an almost-sure bound on future gains, this finite-interval substitution is an implementation approximation of (12). Additional implementation details are provided in E-Companion EC.2.3

Algorithm 1 Adaptive Conformal Scenario Analysis (ACSA)

Input: Target coverage rate $1 - \alpha$, step size $\gamma > 0$, initial quantile predictor of the portfolio gain $g_1(\cdot)$, current time index T

1: Set $\tilde{\alpha}_1 = \alpha$

2: **for** $t = 2, \dots, T$ **do**

3: Observe the realized portfolio gain G_t and realized stress scenario $\mathbf{z}_t$. Compute

$$\text{err}_t \leftarrow \mathbb{1}\{G_t \notin \mathcal{C}_{t-1}(\mathbf{z}_t; \tilde{\alpha}_{t-1})\} \quad (13)$$

4: Update $\tilde{\alpha}_t$ according to

$$\tilde{\alpha}_t \leftarrow \tilde{\alpha}_{t-1} + \gamma \cdot (\alpha - \text{err}_t) \quad (14)$$

5: Update / retrain the quantile predictor g_t , if desired

6: **Output:** For any stress scenario $\mathbf{z}$, produce the prediction set following (12)

$$\mathcal{C}_t(\mathbf{z}) \leftarrow \mathcal{C}_t(\mathbf{z}; \tilde{\alpha}_t) \quad (15)$$

7: **end for**

constructed using the nominal miscoverage level $\alpha = 10\%$ achieve only 80% coverage rather than the desired 90%. Repeated misses will push $\tilde{\alpha}_t$ below 10%, causing the procedure to request more extreme lower and upper quantiles from g_t . An adjusted level near 5%, for instance, may produce intervals whose nominal coverage is 95% but whose realized coverage is closer to the desired 90%. The adjusted level should therefore be interpreted as a calibration parameter rather than as the true miscoverage probability of the underlying quantile model. A more accurate predictor g_t will generally yield sharper intervals, so the quality of g_t remains important in practice. The coverage guarantee of ACSA, however, does not rely on g_t being an accurate conditional-quantile estimator. We require only the following basic regularity condition.

Assumption 1. *For each $t \geq 1$, the predictor $g_t : \mathbb{R}^k \times (0, 1) \rightarrow \mathbb{R}$ is finite-valued and measurable with respect to the information available at time t and its scenario and quantile-level arguments. The realized portfolio gains G_t are finite almost surely.*

Before stating the coverage result, we give two simple examples that illustrate the dynamics of the adjusted level. A third example comparing conditional and stationary quantile predictors at a common adjusted level is provided in E-Companion EC.1.2. For clarity, we do not condition on any scenario in these examples.

Example 1 (Exact Conditional Quantiles). Suppose the conditional distribution of G_t given $\mathcal{F}_{t-1}$ is continuous and g_{t-1} returns its exact conditional quantiles at every level in $(0, 1)$. Define $\pi(a) := \min\{1, \max\{0, a\}\}$. For the prediction set defined in (12), the conditional miscoverage probability is $\pi(\tilde{\alpha}_{t-1})$: it is zero when $\tilde{\alpha}_{t-1} \leq 0$ and one when $\tilde{\alpha}_{t-1} \geq 1$. Hence

$$\mathbb{E}_{t-1}[\tilde{\alpha}_t] = \tilde{\alpha}_{t-1} + \gamma(\alpha - \pi(\tilde{\alpha}_{t-1})). \quad (16)$$

When $0 \leq \tilde{\alpha}_{t-1} \leq 1$, (16) gives $\mathbb{E}_{t-1}[\tilde{\alpha}_t] = (1 - \gamma)\tilde{\alpha}_{t-1} + \gamma\alpha$. Thus, under the exact conditional-quantile predictor assumed here, the adjusted level exhibits mean reversion toward the nominal level α : its conditional expected change is positive when $\tilde{\alpha}_{t-1} < \alpha$ and negative when $\tilde{\alpha}_{t-1} > \alpha$. If $\tilde{\alpha}_{t-1} \leq 0$, the next update increases the level by $\gamma\alpha$; if $\tilde{\alpha}_{t-1} \geq 1$, it decreases the level by $\gamma(1 - \alpha)$. Thus the boundary behavior naturally pushes the adjusted level back toward $(0, 1)$.

Example 2 (Constant Quantile Predictor). Suppose $g_t(\mathbf{z}; u) = C$ for every t , $\mathbf{z}$, and $u \in (0, 1)$, where $C \in \mathbb{R}$ is fixed, and suppose each G_t has a continuous distribution. Whenever $0 < \tilde{\alpha}_t < 1$, the prediction set is the singleton $\{C\}$ and therefore misses almost surely. Repeated misses decrease $\tilde{\alpha}_t$ until it becomes nonpositive, at which point the prediction set becomes $\mathbb{R}$ and the next observation is covered with probability one. The subsequent update then moves $\tilde{\alpha}_t$ upward again. The adjusted levels $\tilde{\alpha}_t$ remain bounded, so the telescoping identity (EC.20) from the proof of Theorem 5.1 implies that the empirical coverage rate converges almost surely to $1 - \alpha$, as stated in Theorem 5.1. This example also shows that coverage alone does not imply informative prediction intervals: a poor predictor can satisfy the marginal guarantee only through highly conservative boundary behavior.

We are now ready to state the main theoretical guarantee. The result is an application of the adaptive conformal argument of Gibbs and Candes (2021); the contribution here is its scenario-analysis formulation and interpretation rather than a new conformal theorem. At each time t , ACSA reports a collection of intervals, one for every scenario $\mathbf{z}$ under consideration. The coverage result, however, is evaluated only along the sequence of scenarios that are actually realized.

Theorem 5.1. *Let $\alpha \in (0, 1)$ and $\gamma > 0$. Under Assumption 1, Algorithm 1, initialized with $\tilde{\alpha}_1 = \alpha$ and using the prediction sets in (12), satisfies*

$$\left| \frac{1}{T} \sum_{t=1}^T \mathbb{1}\{G_{t+1} \in \mathcal{C}_t(\mathbf{z}_{t+1})\} - (1 - \alpha) \right| \leq \frac{\max\{\alpha, 1 - \alpha\} + \gamma}{T\gamma}$$

almost surely. In particular, as $T \rightarrow \infty$, the empirical coverage rate converges to $1 - \alpha$ almost surely.

The proof follows from the update rule in (14). Summing the updates over time expresses the empirical miscoverage rate in terms of the difference between $\tilde{\alpha}_{T+1}$ and its initial value. The boundary convention in (12) keeps $\tilde{\alpha}_t$ bounded, so this remainder vanishes after division by T . We give the full argument in E-Companion EC.3.1.

The step size γ controls how quickly ACSA responds to recent coverage errors. A larger value adapts more rapidly but may cause $\tilde{\alpha}_t$ and the corresponding interval widths to fluctuate more strongly. A smaller value produces smoother adaptation but reacts more slowly to changes in calibration. The coverage conclusion above holds for every fixed $\gamma > 0$, so in practice γ can be chosen using secondary criteria such as interval width or validation performance. Zaffran et al.

(2022), for example, consider adaptive selection based on historical performance, while Gibbs and Candès (2024) aggregate over several candidate step sizes.

The guarantee in Theorem 5.1 is marginal over the sequence of realized scenarios. It does not imply correct coverage within every part of the scenario space. In particular, a procedure could attain the target average coverage while performing poorly in rare or extreme scenarios, which are often precisely the scenarios of greatest interest in risk management. One possible refinement is to partition the scenario space into a small number of groups and maintain a separate ACSA update within each group. We discuss this group-balanced variant in E-Companion EC.1.4. Section 6 introduces KSA, which instead targets scenario-conditional coverage under stronger assumptions. We therefore view ACSA and KSA as complementary: ACSA provides a simple and robust on-line calibration mechanism, while KSA uses additional structure to obtain more scenario-specific uncertainty estimates.

5.2 The 3-Factor Log-Linear Model Revisited

In order to illustrate the advantages of ACSA, we now revisit the three-factor log-linear model introduced in Section 4 where our portfolio consists of a single short position in a European call option on the index with a time to maturity of one month and a fixed moneyness $m = 1.05$ so that the call option is 5% out-of-the-money.

Parameter Misspecification We assume the true data-generating process is as specified in Section 4. We assume all model parameters are known by the risk manager except for σ_ϵ , the standard deviation of the random noise component. Recall the true value of σ_ϵ is 0.18 but here the risk manager assumes (incorrectly) its value is $\hat{\sigma}_\epsilon \neq \sigma_\epsilon$. Consequently, several quantities that depend on σ_ϵ are also incorrectly evaluated for each scenario $\mathbf{z}$, including σ_{tot} , $m_{\mathbf{z}}$ and v_{Cred} . Their estimates are obtained by replacing σ_ϵ with $\hat{\sigma}_\epsilon$ in expressions (8) and (11). The target coverage probability of the prediction intervals is set to $1 - \alpha = 50\%$. For a given risk scenario $\mathbf{z}$, the benchmark quantile predictor g_t uses $\hat{\sigma}_\epsilon$ to construct a prediction interval based on the estimated $\alpha/2$ and $1 - \alpha/2$ quantiles of the portfolio gain. In contrast, the ACSA approach instead uses the $\tilde{\alpha}_t/2$ and $1 - \tilde{\alpha}_t/2$ quantiles computed by Algorithm 1 using as input the benchmark g_t . We denote the two corresponding intervals by $\mathcal{C}_{\text{Mis}}(\mathbf{z}, \hat{\sigma}_\epsilon)$ (constructed using the benchmark g_t) and $\mathcal{C}_{\text{ACSA}}(\mathbf{z}, \hat{\sigma}_\epsilon, \tilde{\alpha}_t)$ (constructed using the ACSA algorithm). For reference, we let $\mathcal{C}_{\text{True}}(\mathbf{z})$ denote the oracle prediction interval based on the true $\alpha/2$ and $1 - \alpha/2$ quantiles. The explicit forms of $\mathcal{C}_{\text{Mis}}(\mathbf{z}, \hat{\sigma}_\epsilon)$, $\mathcal{C}_{\text{ACSA}}(\mathbf{z}, \hat{\sigma}_\epsilon, \tilde{\alpha}_t)$, and $\mathcal{C}_{\text{True}}(\mathbf{z})$ are provided in E-Companion EC.1.3.

Scenario Analysis Table In Table 4, we present⁹ the three aforementioned prediction intervals. Due to space considerations we omit the most extreme scenarios where the rate factor is shocked by

⁹Table 4 also presents the prediction interval $\mathcal{C}_{\text{KSA}}(\mathbf{z})$ produced by the KSA algorithm. We defer a discussion of this until Section 6.

± 30 bp's and the oil factor is shocked by $\pm 20\%$. In each cell, the first four rows respectively display: (1) the oracle interval $\mathcal{C}_{\text{True}}(\mathbf{z})$; (2) the misspecified interval $\mathcal{C}_{\text{Mis}}(\mathbf{z}, \hat{\sigma}_\epsilon)$ based on the benchmark g_t with $\hat{\sigma}_\epsilon = 0.1$; (3) the ACSA interval $\mathcal{C}_{\text{ACSA}}(\mathbf{z}, \hat{\sigma}_\epsilon, \bar{\alpha})$ with¹⁰ $\bar{\alpha} = 0.03$ and using g_t with $\hat{\sigma}_\epsilon = 0.1$ as the quantile predictor input and (4) the interval $\mathcal{C}_{\text{KSA}}(\mathbf{z})$ generated by the KSA algorithm of Section 6.3. The remaining two rows contain the corresponding scenario point estimates from Table 3 and the SSA/CSA entries in Table 2. The true coverage rate of each interval is shown as a percentage beside it, except for the oracle intervals whose true coverage is always 50%. We use red, blue and green asterisks as superscripts on the Mis, ACSA and KSA intervals, respectively, when those intervals fail to contain $\mathbb{E}[G_{t+1} | \mathbf{z}]$, the true conditional expected gain in that scenario. To be clear, this is only a diagnostic for interval location relative to the conditional mean $\mathbb{E}[G_{t+1} | \mathbf{z}]$. In particular, exclusion of $\mathbb{E}[G_{t+1} | \mathbf{z}]$ is not a validity failure, since the intervals target realized $G_{t+1} | \mathbf{z}$ and not $\mathbb{E}[G_{t+1} | \mathbf{z}]$.

The following observations are in order. The true coverage rate of $\mathcal{C}_{\text{Mis}}(\mathbf{z}, \hat{\sigma}_\epsilon)$ can fall significantly below the target 50%, and is even less than 10% in some extreme scenarios. But ACSA mitigates this and its prediction intervals have coverage rates that are much closer to the target coverage rate of 50%. This is especially true for relatively “common” scenarios where z_{Oil} and z_{Rate} are in the ranges $\{-6\%, 0, 6\%\}$ and $\{-10 \text{ bp}, 0, 10 \text{ bp}\}$, respectively. This illustrates how ACSA can compensate for an overly narrow quantile predictor by reducing the adjusted miscoverage level and widening the resulting intervals. Theorem 5.1 concerns coverage averaged over the realized sequence under the online update; it does not imply that the ACSA intervals in this table, constructed using the fixed value $\bar{\alpha} = 0.03$, have marginal coverage exactly equal to 50%. The table exhibits both undercoverage and overcoverage across scenarios.

Additionally, we see that the true coverage rates of both $\mathcal{C}_{\text{Mis}}(\mathbf{z}, \hat{\sigma}_\epsilon)$ and $\mathcal{C}_{\text{ACSA}}(\mathbf{z}, \hat{\sigma}_\epsilon, \bar{\alpha})$ decrease from left to right within each row, and increase from top to bottom within each column. This pattern mirrors the variation of the conditional expected log-return of the underlying asset (i.e., the $m_{\mathbf{z}}$ term in (11)) across the different scenarios. In particular, (10) and (11) yield $\log(S_{t+1}/S_t) \Big| (\mathbf{F}_s = \mathbf{z}) \sim \mathcal{N}(m_{\mathbf{z}}, v_{\text{Cred}})$ where

$$m_{\mathbf{z}} \approx 1.19 \times 10^{-4} - 0.16z_{\text{Oil}} + 4.33z_{\text{Rate}} - \frac{\sigma_\epsilon^2}{504} \quad (17)$$

and v_{Cred} does not depend on the scenario $\mathbf{z}$. The key observation is that $m_{\mathbf{z}}$ decreases with z_{Oil} and increases with z_{Rate} and this observation ultimately explains the monotonic behavior of the coverage rates of $\mathcal{C}_{\text{Mis}}(\mathbf{z}, \hat{\sigma}_\epsilon)$ and $\mathcal{C}_{\text{ACSA}}(\mathbf{z}, \hat{\sigma}_\epsilon, \bar{\alpha})$ as we decrease from left to right within each row, and increase from top to bottom within each column of Table 4. Further details are provided in E-Companion EC.1.3.

¹⁰The value $\bar{\alpha} = .03$ was the mean value of $\tilde{\alpha}_t$ across 2000 steps of ACSA on a simulated dataset and hence is a “typical” value of $\tilde{\alpha}_t$. See Figure 1 below for an illustration of the dynamics of $\tilde{\alpha}_t$.

Rate\Oil		-12%	-6%	0	+6%	+12%
-20 bp	$\mathcal{C}_{\text{True}}$	(-27.2, -0.2)	(-11.1, 10.9)	(2.1, 19.7)	(12.8, 26.5)	(21.1, 31.6)
	$\mathcal{C}_{\text{Mis}}$	(7.3, 21.8)*,12%	(19.1, 29.6)*,8%	(27.7, 35.1)*,5%	(33.8, 38.7)*,3%	(37.8, 41.0)*,2%
	$\mathcal{C}_{\text{ACSA}}$	(-16.1, 32.1),57%	(1.3, 36.7)*,49%	(14.6, 39.8)*,41%	(24.5, 41.6)*,33%	(31.6, 42.8)*,25%
	$\mathcal{C}_{\text{KSA}}$	(-20.6, 1.8),44%	(-10.1, 10.7),48%	(-3.2, 17.6),53%	(1.3, 22.6),52%	(3.2, 26.8),47%
	$\mathbb{E}[G_{t+1} \mathbf{z}]$	-15.10	-1.35	9.80	18.66	25.58
	SSA/CSA	-35.0, -12.5	-14.3, 1.0	2.1, 11.9	14.7, 20.4	24.1, 27.0
-10 bp	$\mathcal{C}_{\text{True}}$	(-35.2, -5.8)	(-17.8, 6.4)	(-3.3, 16.1)	(8.4, 23.7)	(17.7, 29.6)
	$\mathcal{C}_{\text{Mis}}$	(0.9, 17.4)*,13%	(14.4, 26.5)*,10%	(24.3, 32.9)*,7%	(31.4, 37.3)*,4%	(36.3, 40.1)*,3%
	$\mathcal{C}_{\text{ACSA}}$	(-25.1, 29.5),60%	(-5.8, 34.9)*,53%	(9.3, 38.6)*,45%	(20.6, 40.9)*,37%	(28.8, 42.4)*,29%
	$\mathcal{C}_{\text{KSA}}$	(-26.3, -3.2),42%	(-14.4, 7.3),47%	(-6.1, 15.1),52%	(-0.3, 20.5),52%	(3.2, 24.5),47%
	$\mathbb{E}[G_{t+1} \mathbf{z}]$	-22.02	-7.04	5.21	15.04	22.77
	SSA/CSA	-34.8, -19.3	-14.0, -4.6	2.3, 7.4	14.8, 16.9	24.2, 24.4
0	$\mathcal{C}_{\text{True}}$	(-43.9, -12.0)	(-25.0, 1.4)	(-9.3, 12.1)	(3.6, 20.6)	(13.9, 27.2)
	$\mathcal{C}_{\text{Mis}}$	(-6.2, 12.4)*,15%	(8.9, 22.9)*,11%	(20.3, 30.4)*,8%	(28.6, 35.6)*,5%	(34.4, 39.0)*,3%
	$\mathcal{C}_{\text{ACSA}}$	(-35.1, 26.3),63%	(-13.7, 32.8),56%	(3.2, 37.2)*,48%	(16.0, 40.0)*,40%	(25.5, 41.8)*,32%
	$\mathcal{C}_{\text{KSA}}$	(-34.0, -7.5),44%	(-19.5, 3.4),46%	(-9.1, 11.7),49%	(-2.5, 18.3),52%	(2.0, 22.7),50%
	$\mathbb{E}[G_{t+1} \mathbf{z}]$	-29.52	-13.25	0.16	11.01	19.62
	SSA/CSA	-34.5, -26.7	-13.8, -10.6	2.5, 2.5	15.0, 13.0	24.3, 21.3
10 bp	$\mathcal{C}_{\text{True}}$	(-53.3, -18.8)	(-32.9, -4.2)	(-15.8, 7.7)	(-1.7, 17.2)	(9.7, 24.5)
	$\mathcal{C}_{\text{Mis}}$	(-14.2, 6.7)*,17%	(2.8, 18.7)*,13%	(15.8, 27.4)*,9%	(25.3, 33.6)*,6%	(32.1, 37.7)*,4%
	$\mathcal{C}_{\text{ACSA}}$	(-45.9, 22.7),66%	(-22.5, 30.2),59%	(-3.7, 35.5)*,52%	(10.9, 38.9)*,44%	(21.8, 41.1)*,36%
	$\mathcal{C}_{\text{KSA}}$	(-40.8, -14.3),41%	(-25.3, -1.0),45%	(-12.8, 7.9),46%	(-5.1, 15.7),52%	(0.3, 21.0),52%
	$\mathbb{E}[G_{t+1} \mathbf{z}]$	-37.62	-20.01	-5.38	6.55	16.10
	SSA/CSA	-34.2, -34.7	-13.6, -17.3	2.7, -2.9	15.1, 8.7	24.4, 17.9
20 bp	$\mathcal{C}_{\text{True}}$	(-63.2, -26.2)	(-41.4, -10.2)	(-22.9, 2.8)	(-7.5, 13.3)	(5.0, 21.5)
	$\mathcal{C}_{\text{Mis}}$	(-23.0, 0.3)*,19%	(-4.1, 13.9)*,14%	(10.5, 24.0)*,11%	(21.5, 31.2)*,7%	(29.4, 36.1)*,5%
	$\mathcal{C}_{\text{ACSA}}$	(-57.7, 18.4),69%	(-32.2, 27.3),62%	(-11.4, 33.4)*,55%	(5.0, 37.6)*,47%	(17.4, 40.3)*,39%
	$\mathcal{C}_{\text{KSA}}$	(-49.1, -21.2),40%	(-31.6, -6.3),43%	(-17.1, 3.7),42%	(-7.9, 12.4),48%	(-1.8, 18.9),52%
	$\mathbb{E}[G_{t+1} \mathbf{z}]$	-46.34	-27.34	-11.44	1.64	12.19
	SSA/CSA	-33.9, -43.3	-13.4, -24.5	2.8, -8.9	15.2, 3.9	24.5, 14.2

Table 4: Portfolio gains across the (rate, oil) grid. The first row of each cell reports the true 50% prediction interval $\mathcal{C}_{\text{True}}$. Rows two and three report the estimated 50% prediction intervals $\mathcal{C}_{\text{Mis}}$ and $\mathcal{C}_{\text{ACSA}}$, respectively. The fourth row reports the KSA interval, $\mathcal{C}_{\text{KSA}}$. The fifth row reports the true conditional expected scenario gain $\mathbb{E}[G_{t+1} | \mathbf{z}]$, and the sixth row reports the SSA and CSA point estimates.

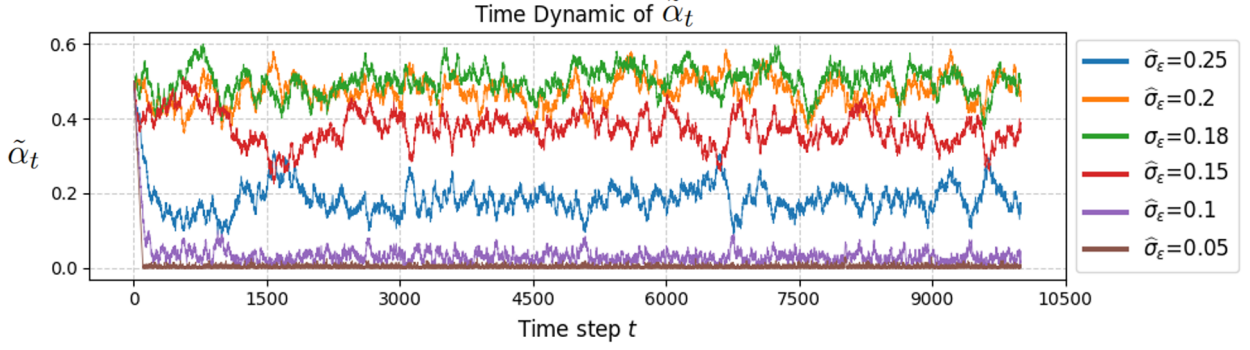


Figure 1: Dynamics of ACSA’s adjusted miscoverage level $\tilde{\alpha}_t$ under different $\hat{\sigma}_\epsilon$. For each estimation, the $\tilde{\alpha}_t$ value fluctuates around distinct values. Specifically, for $\hat{\sigma}_\epsilon = \{0.25, 0.2, 0.18, 0.15, 0.1, 0.05\}$, the stabilized values of $\tilde{\alpha}_t$ are approximately $\{0.18, 0.47, 0.5, 0.37, 0.03, 0.006\}$, respectively. These stabilized values summarize the dynamic ACSA trajectories targeting 50% empirical coverage.

The Dynamics of $\tilde{\alpha}_t$ The evolution of ACSA’s $\tilde{\alpha}_t$ parameter is illustrated in Figure 1. We test several values of the misspecified noise level $\hat{\sigma}_\epsilon$, ranging from 0.05 (an underestimate) to 0.25 (an overestimate). Across all cases, $\tilde{\alpha}_t$ starts at $\tilde{\alpha}_1 = 0.5$ and then fluctuates around a stable level $\bar{\alpha}$ that depends on $\hat{\sigma}_\epsilon$. This reflects ACSA’s process of learning the appropriate adjusted miscoverage level such that the realized prediction intervals achieve an empirical coverage rate of approximately 50%.

An interesting observation is that $\bar{\alpha}$ does not vary monotonically with $\hat{\sigma}_\epsilon$: when $\hat{\sigma}_\epsilon = 0.05$, the stabilized $\bar{\alpha}$ is 0.006, suggesting that using the nominal quantile level α would yield an empirical coverage below 50%; as $\hat{\sigma}_\epsilon$ approaches the true value $\sigma_\epsilon = 0.18$ from below, $\bar{\alpha}$ moves toward 0.5 as expected. When $\hat{\sigma}_\epsilon$ increases beyond σ_ϵ to 0.25, $\bar{\alpha}$ drops again, to around 0.18. Intuitively, this *non-monotonic* pattern arises from an underlying *monotonic* dynamic behavior of the benchmark interval $\mathcal{C}_{\text{Mis}}(\mathbf{z}, \hat{\sigma}_\epsilon)$. As $\hat{\sigma}_\epsilon$ increases, the estimated price (and its quantiles) for the call option increase, so the estimated portfolio gain decreases because the portfolio is short a call. Since the two endpoints of $\mathcal{C}_{\text{Mis}}(\mathbf{z}, \hat{\sigma}_\epsilon)$ are estimated quantiles of the portfolio gain, both endpoints move monotonically in the opposite direction of $\hat{\sigma}_\epsilon$. The resulting true coverage rate is nevertheless non-monotonic in $\hat{\sigma}_\epsilon$. As shown using the explicit coverage function derived in E-Companion EC.1.3 and illustrated in Figure EC.2, the coverage first increases and then decreases as $\hat{\sigma}_\epsilon$ varies over the range considered in our experiment. This non-monotonicity in turn explains the non-monotonic dependence of the stabilized value $\bar{\alpha}$ on $\hat{\sigma}_\epsilon$.

6 Kernel Scenario Analysis

ACSA provides an online calibration mechanism without prescribing how scenario-wise quantiles should be estimated. We now introduce kernel scenario analysis (KSA), which constructs these

quantiles by combining a predictor of the scenario gain with a kernel estimate of the remaining uncertainty. Historical observations receive greater weight when their stressed-factor moves and market conditions resemble those of the scenario being evaluated. In this way, KSA uses factor-based features to distinguish between scenarios with different gain distributions, without specifying a parametric model for the joint factor dynamics.

We first introduce the loss function, kernel, and feature vectors used by KSA. Section 6.1 presents the core algorithm with a fixed centering function and then discusses adaptive centering and changing portfolios as practical extensions. Section 6.2 establishes the conditional-coverage guarantee for the core procedure under its stated assumptions, and Section 6.3 revisits the three-factor running example.

Pinball Loss Function The pinball loss provides a convenient way to estimate quantiles through optimization. KSA uses a kernel-weighted version of this loss; it can also be used to train the quantile predictor g_t in ACSA.

Definition 6.1. *The pinball loss function for $y, q \in \mathbb{R}$ and a quantile level $u \in [0, 1]$ is*

$$\ell_u(y, q) := u(y - q)^+ + (1 - u)(y - q)^-, \quad (18)$$

where $y^+ := \max\{y, 0\}$ and $y^- := -\min\{y, 0\}$.

The pinball loss is piecewise linear and convex in q , and is a proper scoring rule for quantiles (Gneiting and Raftery, 2007). For a real-valued random variable Y with CDF F_Y , define its u -quantile by $Q_Y(u) := \inf\{y : F_Y(y) \geq u\}$ for $u \in (0, 1)$. If Y is integrable, then

$$Q_Y(u) \in \operatorname{argmin}_{q \in \mathbb{R}} \mathbb{E}[\ell_u(Y, q)]. \quad (19)$$

Replacing the expectation in (19) by a sample average gives an empirical quantile estimator. KSA extends this construction by weighting observations according to their relevance to the target scenario. We measure similarity between feature vectors using the following class of kernels.

Definition 6.2. *A function $\kappa : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ is a **similarity radial kernel** if there exists a nonnegative, nonincreasing function $\tilde{\kappa} : [0, \infty) \rightarrow [0, \infty)$ such that $\kappa(\mathbf{w}, \mathbf{w}') = \tilde{\kappa}(\|\mathbf{w} - \mathbf{w}'\|)$ for all $\mathbf{w}, \mathbf{w}' \in \mathbb{R}^d$.*

Thus, observations closer to the target feature vector receive at least as much weight as observations farther away. The kernel determines how broadly KSA pools historical observations around a given target; its bandwidth is made explicit in Section 6.2.

Historical Context Vector The historical context vector $\mathbf{H}_t$ collects the information available at time t that we use to predict the next-period portfolio gain. For example, we may take $\mathbf{H}_t =$

$(\Delta \mathbf{f}_t, \Delta \mathbf{f}_{t-1}, \dots, \Delta \mathbf{f}_{t-L_f+1})$ when L_f lags of common-factor returns capture the relevant history. Risk-factor changes, current factor levels, or other market-state information can be included when they provide additional predictive information. Conversely, redundant components can be omitted: if the relevant risk-factor changes are known functions of the included common-factor returns, they need not be included separately.

A Markov structure can further reduce the required history. When a current state contains all historical information relevant to the next-period gain distribution, $\mathbf{H}_t$ can be taken as that state rather than a longer vector of lagged observations. The important requirement is predictive sufficiency for the portfolio gain, not merely the Markov property of one component of the market model. The precise condition used by our analysis is stated in Assumption 2.

Feature Vector We combine the historical context with the stress scenario to form the **feature vector** $\mathbf{W}_t(\mathbf{z}) := (\mathbf{z}, \mathbf{H}_{t-1})$. At forecast date t , the target feature vector is therefore $\mathbf{W}_{t+1}(\mathbf{z}) = (\mathbf{z}, \mathbf{H}_t)$, where $\mathbf{z}$ is the hypothesized stressed-factor move on the next day. For a historical date $s \leq t$, the scenario $\mathbf{z}_s$ has already been observed, and we write $\mathbf{W}_s := \mathbf{W}_s(\mathbf{z}_s) = (\mathbf{z}_s, \mathbf{H}_{s-1})$. These definitions put historical observations and hypothetical scenarios in the same feature space. If $\mathbf{H}_t$ retains all relevant historical information, conditioning on $\mathcal{F}_t$ and $\mathbf{z}_{t+1} = \mathbf{z}$ gives the same gain distribution as conditioning on $\mathbf{H}_t$ and $\mathbf{z}_{t+1} = \mathbf{z}$. KSA then estimates this distribution locally by comparing $\mathbf{W}_{t+1}(\mathbf{z})$ with the observed feature vectors $\mathbf{W}_s$.

6.1 The Kernel Scenario Analysis Algorithm

The key idea of the Kernel Scenario Analysis (KSA) algorithm is to assign an *importance weight* to each historical date $s \leq t$, based on the similarity between the historical feature vector $\mathbf{W}_s$ and the scenario feature vector $\mathbf{W}_{t+1}(\mathbf{z})$. Specifically, we use a similarity radial kernel $\kappa(\mathbf{w}, \mathbf{w}')$ to encode similarities between feature vectors $\mathbf{w}$ and $\mathbf{w}'$. These importance weights are then used to form a reweighted empirical pinball loss, whose minimizer yields an estimate of the conditional quantile of the portfolio gain under a given scenario $\mathbf{z}$. The prediction interval is then constructed using the estimated quantiles as its endpoints. The core KSA procedure is summarized in Algorithm 2.

Algorithm 2 uses a predictor $\varphi(\mathbf{w})$ to represent the predictable component of the portfolio gain given the stress scenario and historical context vector. Its role is to separate the portfolio gain into a predictable part and a residual. In Step 4, the residual term Y_s is calculated for $s \leq t$ by subtracting $\varphi(\mathbf{W}_s)$ from G_s . The Y_s 's are then used in place of the G_s 's as the target variable whose conditional quantiles are estimated.

The reason that we work with the Y_s 's instead of the G_s 's is that if $\varphi(\mathbf{W}_s)$ predicts $\mathbb{E}[G_s \mid \mathbf{W}_s]$ well, then Y_s should contain less predictable variation and it should therefore be easier to estimate its conditional quantiles than those of the original G_s , which may be strongly covariate-dependent. Importantly, however, φ need not equal the true conditional expectation for the construction to be

Algorithm 2 Kernel Scenario Analysis

Input: Target coverage rate $1 - \alpha$, kernel $\kappa(\cdot, \cdot)$, predictor of expected portfolio gain $\varphi(\cdot)$

- 1: Set $\mathbf{W}_1 \leftarrow (\mathbf{z}_1, \mathbf{H}_0)$ and $Y_1 \leftarrow G_1 - \varphi(\mathbf{W}_1)$
- 2: **for** $t = 2, \dots, T$ **do**
- 3: Observe portfolio gain G_t and realized stress scenario $\mathbf{z}_t$
- 4: Set $\mathbf{W}_t \leftarrow (\mathbf{z}_t, \mathbf{H}_{t-1})$ and $Y_t \leftarrow G_t - \varphi(\mathbf{W}_t)$
- 5: For any $u \in [0, 1]$, $q \in \mathbb{R}$ and $\mathbf{w}$, define the reweighted residual pinball loss $\hat{\ell}_t$

$$\hat{\ell}_t(u; q, \mathbf{w}) := \frac{\sum_{s=1}^t \kappa(\mathbf{W}_s, \mathbf{w}) \cdot \ell_u(Y_s, q)}{\sum_{s=1}^t \kappa(\mathbf{W}_s, \mathbf{w})} \quad (20)$$

- 6: **Output:** For each scenario $\mathbf{z}$, set $\mathbf{W}_{t+1}(\mathbf{z}) \leftarrow (\mathbf{z}, \mathbf{H}_t)$ and

$$\mathcal{C}_t(\mathbf{z}) \leftarrow \left[\varphi(\mathbf{W}_{t+1}(\mathbf{z})) + \hat{Q}_t(\alpha/2; \mathbf{W}_{t+1}(\mathbf{z})), \varphi(\mathbf{W}_{t+1}(\mathbf{z})) + \hat{Q}_t(1 - \alpha/2; \mathbf{W}_{t+1}(\mathbf{z})) \right] \quad (21)$$

where

$$\hat{Q}_t(u; \mathbf{W}_{t+1}(\mathbf{z})) := \underset{q \in \mathbb{R}}{\operatorname{argmin}} \hat{\ell}_t(u; q, \mathbf{W}_{t+1}(\mathbf{z})), \quad u \in \{\alpha/2, 1 - \alpha/2\}. \quad (22)$$

- 7: **end for**
-

valid. For any fixed deterministic φ , the conditional quantile of $G_{t+1} \mid \mathbf{W}_{t+1}(\mathbf{z})$ can be recovered from the conditional quantile of $Y_{t+1} \mid \mathbf{W}_{t+1}(\mathbf{z})$ via

$$Q_{G_{t+1}}(u \mid \mathbf{W}_{t+1}(\mathbf{z})) = \varphi(\mathbf{W}_{t+1}(\mathbf{z})) + Q_{Y_{t+1}}(u \mid \mathbf{W}_{t+1}(\mathbf{z})). \quad (23)$$

Thus, a more accurate choice of φ is useful primarily because it can make the residual quantile surface easier to estimate, rather than because the validity of the quantile-shift identity requires φ to be the true conditional mean. We demonstrate this “smoother landscape” phenomenon in the following example.

Example 3 (Centering Removes Predictable Variation). *Suppose $G = W + \epsilon$, where $\epsilon \sim \mathcal{N}(0, 1)$ denotes a noise term independent of the random variable $W \in \mathbb{R}$. The conditional u -quantile of $G \mid W$ is $Q_G(u \mid W)$ and takes the form $Q_G(u \mid W = w) = w + \Phi^{-1}(u)$, where Φ is the standard normal CDF. If we have a perfect expectation predictor φ of the conditional expectation, i.e., $\varphi(w) = \mathbb{E}[G \mid W = w] = w$, then the constructed residual $Y := G - \varphi(w)$ satisfies $Y = \epsilon$. The conditional u -quantile of $Y \mid W$ is then $Q_Y(u \mid W = w) = \Phi^{-1}(u)$, which is independent of w .*

The implication of a “smoother landscape” is that the distributions, and consequently the quantiles, of $Y \mid W = w$ may vary less with w . In this particular example, $\Phi^{-1}(u)$ and therefore the conditional quantile of $Y \mid W = w$ does not depend on w at all. This property makes estimating $Q_Y(u \mid W)$ easier than estimating $Q_G(u \mid W)$.

Returning to Algorithm 2, in Steps 5 and 6, the conditional quantile of $Y_{t+1} \mid \mathbf{W}_{t+1}(\mathbf{z})$ is estimated by minimizing a reweighted version of the residual pinball loss function. To see what this minimizer

gives us, observe that $\hat{F}_t(y \mid \mathbf{W}_{t+1}(\mathbf{z})) := \frac{\sum_{s=1}^t \kappa(\mathbf{W}_s, \mathbf{W}_{t+1}(\mathbf{z})) \cdot \mathbb{1}\{y \geq Y_s\}}{\sum_{s=1}^t \kappa(\mathbf{W}_s, \mathbf{W}_{t+1}(\mathbf{z}))}$ is an estimator of the CDF of $Y_{t+1} \mid \mathbf{W}_{t+1}(\mathbf{z})$. Use $\hat{\mathcal{P}}_t$ to denote the probability distribution induced by this CDF. The u -quantile of $\hat{\mathcal{P}}_t$ is given by

$$\begin{aligned} Q_{\hat{\mathcal{P}}_t}(u) &= \operatorname{argmin}_{q \in \mathbb{R}} \mathbb{E}_{Y \sim \hat{\mathcal{P}}_t}[\ell_u(Y, q)] \\ &= \operatorname{argmin}_{q \in \mathbb{R}} \frac{\sum_{s=1}^t \kappa(\mathbf{W}_s, \mathbf{W}_{t+1}(\mathbf{z})) \cdot \ell_u(Y_s, q)}{\sum_{s=1}^t \kappa(\mathbf{W}_s, \mathbf{W}_{t+1}(\mathbf{z}))}. \end{aligned} \quad (24)$$

The estimator $\hat{Q}_t(u; \mathbf{W}_{t+1}(\mathbf{z}))$ in (22) is therefore exactly the u -quantile of $\hat{\mathcal{P}}_t$. The minimization of $\hat{\ell}_t(u; q, \mathbf{W}_{t+1}(\mathbf{z}))$ in Step 6 can be solved in $O(t)$ iterations. First, permute the indices with a permutation operator $\Pi : \{1, \dots, t\} \rightarrow \{1, \dots, t\}$ so that $Y_{\Pi(1)} \leq Y_{\Pi(2)} \leq \dots \leq Y_{\Pi(t)}$. Then,

$$\hat{Q}_t(u; \mathbf{W}_{t+1}(\mathbf{z})) = Y_{\Pi(j)}, \quad j = \min \left\{ i \in \{1, \dots, t\} \mid \frac{\sum_{s=1}^i \kappa(\mathbf{W}_{\Pi(s)}, \mathbf{W}_{t+1}(\mathbf{z}))}{\sum_{s=1}^t \kappa(\mathbf{W}_{\Pi(s)}, \mathbf{W}_{t+1}(\mathbf{z}))} \geq u \right\}.$$

A brief comparison between the input requirements for ACSA and KSA is useful. For each time t , the ACSA algorithm requires a quantile predictor $g_t(\mathbf{z})$ of the conditional portfolio gain, whereas the core KSA algorithm requires a centering predictor $\varphi(\cdot)$. It is not immediately clear which requirement is more demanding. For example, the median of a distribution is less sensitive to outliers and may be easier to estimate than the mean, whereas extreme quantiles, such as 1% quantiles, are generally harder to estimate because far fewer observations are available in the tails. If one has an excellent quantile predictor g_t at hand, then a downside of Algorithm 2 is that it does not directly use g_t , but instead constructs its own quantile predictor through (20) and (22). Quantile predictors, however, are often not readily available, so this aspect of KSA can be viewed as either a bug or a feature depending on the application.

In practice, it can be useful to replace the fixed centering function φ by a time-varying predictor φ_t that is updated as new data arrive. This adaptive-centering version follows the same kernel reweighting principle, but introduces additional dependence because the residuals themselves now depend on an estimated predictor. To avoid using the same observations both to update φ_t and to calibrate the residual distribution, one can maintain separate fitting and calibration samples; in particular, once an observation is used to update the predictor, it is excluded from subsequent residual calibration.

A practical advantage of KSA over ACSA is the convenience of estimating conditional expectations under frequently changing portfolios. Because conditional expectations aggregate linearly across positions, one can estimate security-level expectation predictors and then combine them using the current portfolio weights to obtain a portfolio-level expectation predictor. The scenario-wise quantiles are then obtained through the kernel reweighting step, without training a new portfolio-level quantile model. By contrast, ACSA trains a portfolio-level quantile predictor, and quantiles do not in general aggregate linearly across securities, so changes in portfolio composition typically require retraining that predictor.

6.2 Analysis of Algorithm 2

This section establishes a conditional-coverage guarantee for the core KSA procedure in Algorithm 2. The main distinction from Theorem 5.1 is the conditioning information: rather than controlling coverage averaged over the realized sequence of scenarios, we study next-period coverage conditional on the information currently available and on a specified scenario. Throughout this section, the portfolio is fixed. We retain the definitions $\mathbf{W}_t = (\mathbf{z}_t, \mathbf{H}_{t-1})$ and $\mathbf{W}_{t+1}(\mathbf{z}) = (\mathbf{z}, \mathbf{H}_t)$, and write $d_{\mathbf{W}}$ for the dimension of the feature vector. To allow for temporal dependence in the historical feature–residual pairs, we use the notion of β -mixing. More details are provided in Bradley (2005), and a brief definition is given below.

Definition 6.3 (β -mixing). *For a stationary process $\{X_t\}_{t \geq 0}$, let $\sigma(\mathcal{X})$ denote the σ -algebra generated by a collection of random variables $\mathcal{X}$. Define*

$$\beta_X(k) = \sup_{t \geq 0} \mathbb{E} \left[\sup_{B \in \sigma(\{X_s : s \geq t+k\})} \left| \mathbb{P}(B \mid \sigma(\{X_s : s \leq t\})) - \mathbb{P}(B) \right| \right].$$

The process is β -mixing if $\beta_X(k) \rightarrow 0$ as $k \rightarrow \infty$. It is geometrically β -mixing if $\beta_X(k) \leq \beta_0 \rho^k$ for some $\beta_0 < \infty$ and $0 < \rho < 1$.

Intuitively, $\beta_X(k)$ measures the dependence between the past of the process and its future after a gap of length k . A small value of $\beta_X(k)$ means that sufficiently separated observations are nearly independent, while geometric β -mixing requires this dependence to decay exponentially fast. In particular, $\beta_X(k) = 0$ for all $k \geq 1$ corresponds to the independent case.

We now specify the regularity conditions under which KSA achieves a conditional-coverage guarantee. Let $\mathcal{D} \subseteq \mathbb{R}^{d_{\mathbf{W}}}$ be a compact region containing the feature vectors $\mathbf{W}_{t+1}(\mathbf{z})$ corresponding to the scenarios and historical contexts of interest. For example, a scenario table typically considers stress scenarios $\mathbf{z}$ within a prescribed range together with historical contexts $\mathbf{H}_t$ that are not excessively extreme. We restrict the theoretical guarantee to targets satisfying $\mathbf{W}_{t+1}(\mathbf{z}) \in \mathcal{D}$. For some fixed $h_0 > 0$, define the enlarged region $\mathcal{D}^+ = \{\mathbf{w} : \text{dist}(\mathbf{w}, \mathcal{D}) \leq h_0\}$, where $\text{dist}(\mathbf{w}, \mathcal{D}) := \inf_{\mathbf{w}' \in \mathcal{D}} \|\mathbf{w} - \mathbf{w}'\|$ denotes the Euclidean distance¹¹ from $\mathbf{w}$ to $\mathcal{D}$. Since kernel estimation at a target in $\mathcal{D}$ uses observations in its neighborhood, the regularity conditions below are imposed on $\mathcal{D}^+$.

Assumption 2. *The following conditions hold.*

- (a) *The process $\{(\mathbf{W}_t, Y_t)\}_{t \geq 0}$ is stationary and geometrically β -mixing, with $\beta(k) \leq \beta_0 \rho^k$ for some $\beta_0 < \infty$ and $0 < \rho < 1$.*

¹¹Throughout this work, an unsubscripted $\|\cdot\|$ denotes the Euclidean norm.

- (b) The feature vector $\mathbf{W}_{t+1}(\mathbf{z})$ contains all information in $\mathcal{F}_t$ that is relevant to the conditional distribution of Y_{t+1} under scenario $\mathbf{z}$. Specifically,

$$\mathbb{P}(Y_{t+1} \leq y \mid \mathcal{F}_t, \mathbf{z}_{t+1} = \mathbf{z}) = F_{Y \mid \mathbf{W}}(y \mid \mathbf{W}_{t+1}(\mathbf{z})) \quad (25)$$

almost surely, simultaneously for every $y \in \mathbb{R}$ and every $\mathbf{z}$ such that $\mathbf{W}_{t+1}(\mathbf{z}) \in \mathcal{D}$.

- (c) The marginal distribution of $\mathbf{W}_t$ admits a density $\pi_{\mathbf{W}}$ that is bounded above and bounded away from zero on $\mathcal{D}^+$.
- (d) The conditional distribution of Y_t given $\mathbf{W}_t = \mathbf{w}$ is continuous for every $\mathbf{w} \in \mathcal{D}^+$. Moreover, the conditional CDF varies uniformly continuously with $\mathbf{w}$:

$$\lim_{\epsilon \downarrow 0} \sup_{\mathbf{w}, \mathbf{w}' \in \mathcal{D}^+} \sup_{\substack{y \in \mathbb{R} \\ \|\mathbf{w} - \mathbf{w}'\| \leq \epsilon}} |F_{Y \mid \mathbf{W}}(y \mid \mathbf{w}) - F_{Y \mid \mathbf{W}}(y \mid \mathbf{w}')| = 0.$$

- (e) The kernel radial function $\tilde{\kappa}$ in Definition 6.2 is bounded and has support either $[0, R_\kappa]$ for some $0 < R_\kappa < \infty$ or $[0, \infty)$. It is Lipschitz continuous on the interior of its support. Moreover, $0 < \int_{\mathbb{R}^d} \tilde{\kappa}(\|\mathbf{v}\|) d\mathbf{v} < \infty$ and $\int_{\mathbb{R}^d} \|\mathbf{v}\|^2 \tilde{\kappa}(\|\mathbf{v}\|) d\mathbf{v} < \infty$.

Assumption 2(a) provides the stationarity and weak dependence needed to control the kernel-weighted empirical averages. Stationarity is a working assumption for the KSA analysis; unlike the pathwise guarantee for ACSA, the result below relies on distributional regularity over time. Assumption 2(b) is a predictive-sufficiency condition: once $\mathbf{W}_{t+1}(\mathbf{z})$ is specified, earlier information in $\mathcal{F}_t$ provides no additional information about the conditional distribution of the next residual. Assumption 2(c) guarantees adequate local support for the feature distribution over the operational region. Assumption 2(d) ensures that nearby feature vectors have similar conditional residual distributions, thereby controlling the approximation error introduced by local smoothing. Finally, Assumption 2(e) imposes standard regularity conditions on the kernel. The RBF ($\tilde{\kappa}(r) = e^{-r^2/2}$), Laplace ($\tilde{\kappa}(r) = e^{-r}$), and naive kernels ($\tilde{\kappa}(r) = \mathbb{1}\{r \leq 1\}$) all satisfy these conditions.

To make the dependence on the bandwidth explicit, for $h > 0$, write $\kappa_h(\mathbf{w}, \mathbf{w}') = \tilde{\kappa}\left(\frac{\|\mathbf{w} - \mathbf{w}'\|}{h}\right)$. Algorithm 2 takes κ_h as a fixed input. The following theorem shows that, for a sufficiently small fixed bandwidth and a sufficiently large historical sample, the prediction intervals produced by KSA achieve scenario-wise conditional coverage close to the target level $1 - \alpha$.

Theorem 6.1. *Suppose Assumption 2 holds. For every $\epsilon > 0$, there exists $h_\epsilon > 0$ such that, for every fixed $h \in (0, h_\epsilon]$,*

$$\sup_{\mathbf{z}: \mathbf{W}_{t+1}(\mathbf{z}) \in \mathcal{D}} |\mathbb{P}(G_{t+1} \notin \mathcal{C}_t(\mathbf{z}) \mid \mathcal{F}_t, \mathbf{z}_{t+1} = \mathbf{z}) - \alpha| \leq \epsilon + O_p\left(\sqrt{\frac{\log t}{t}}\right). \quad (26)$$

Here, the O_p term is with respect to the randomness of the historical data in $\mathcal{F}_t$, and its implicit constant may depend on the fixed bandwidth h .

We use the notation $O_p(a_t)$ to denote a stochastic term whose magnitude, after division by a_t , is bounded in probability. An explicit finite-sample version of Theorem 6.1 is given as Theorem EC.3.4 in E-Companion EC.3.2, together with its proof.

Unlike Theorem 5.1, which controls coverage averaged over the realized sequence of scenarios, Theorem 6.1 controls next-period coverage conditional on the current information and the specified scenario. The two terms in (26) have a natural interpretation. The tolerance ϵ represents the approximation error from pooling observations whose feature vectors are close, but not identical, to $\mathbf{W}_{t+1}(\mathbf{z})$. This error can be made arbitrarily small by choosing a sufficiently small bandwidth. The stochastic term represents finite-sample estimation error and vanishes as t increases. A smaller bandwidth therefore reduces local approximation error but also places substantial weight on fewer nearby observations, so more historical data may be needed before the asymptotic behavior becomes apparent.

The guarantee applies only when the target feature vector lies in $\mathcal{D}$. This restriction is natural for a local-smoothing procedure: the target must have adequate nearby historical support, and the conditional residual distribution must vary regularly over its neighborhood. These conditions may fail for scenarios far outside the historically supported region. In addition, $\mathcal{D}$ cannot be taken to be all of $\mathbb{R}^{d_{\mathbf{w}}}$ under Assumption 2(c), since a probability density cannot be bounded away from zero over an unbounded space. In applications, $\mathcal{D}$ should therefore be chosen to cover the scenarios and historical contexts within the intended domain of use.

Under Assumption 2 alone, continuity of the conditional CDF controls the kernel-smoothing bias only qualitatively. The following stronger smoothness conditions give explicit orders in the bandwidth.

Corollary 6.2. *Suppose Assumption 2 holds.*

(i) *If, for some $L < \infty$,*

$$\sup_{y \in \mathbb{R}} |F_{Y|\mathbf{W}}(y | \mathbf{w}) - F_{Y|\mathbf{W}}(y | \mathbf{w}')| \leq L \|\mathbf{w} - \mathbf{w}'\|$$

for all $\mathbf{w}, \mathbf{w}' \in \mathcal{D}^+$, then for every sufficiently small h ,

$$\sup_{\mathbf{z}: \mathbf{W}_{t+1}(\mathbf{z}) \in \mathcal{D}} |\mathbb{P}(G_{t+1} \notin \mathcal{C}_t(\mathbf{z}) | \mathcal{F}_t, \mathbf{z}_{t+1} = \mathbf{z}) - \alpha| = O(h) + O_p\left(\sqrt{\frac{\log t}{t}}\right).$$

(ii) *If $\pi_{\mathbf{W}}$ is twice continuously differentiable with Hessian uniformly bounded over $\mathbf{w} \in \mathcal{D}^+$, and for every $y \in \mathbb{R}$, $\pi_{\mathbf{W}}(\cdot)F_{Y|\mathbf{W}}(y | \cdot)$ is twice continuously differentiable on an open neighborhood of $\mathcal{D}^+$, with Hessian uniformly bounded over $\mathbf{w} \in \mathcal{D}^+$ and $y \in \mathbb{R}$, then for every sufficiently*

small h ,

$$\sup_{z: \mathbf{W}_{t+1}(z) \in \mathcal{D}} |\mathbb{P}(G_{t+1} \notin \mathcal{C}_t(z) \mid \mathcal{F}_t, \mathbf{z}_{t+1} = z) - \alpha| = O(h^2) + O_p\left(\sqrt{\frac{\log t}{t}}\right).$$

The deterministic orders are understood as $h \downarrow 0$, with their constants uniform over all sufficiently small h . For each fixed h , the stochastic orders are understood as $t \rightarrow \infty$, and their implicit constants may depend on h .

An explicit finite-sample version of Corollary 6.2 is given as Corollary EC.3.5 in E-Companion EC.3.2, together with its proof. Corollary 6.2 makes the kernel-approximation term in Theorem 6.1 explicit. Under the Lipschitz condition in part (i), the approximation error is of order h . Under the stronger second-order condition in part (ii), radial symmetry of the kernel cancels the first-order contribution and improves this error to order h^2 . In both cases, the remaining stochastic term reflects estimation from a finite historical sample.

Although Algorithm 2 is formulated with a fixed bandwidth, the theoretical construction can also use a deterministic sequence of bandwidths h_t . A fixed bandwidth leaves a non-vanishing kernel-approximation error as $t \rightarrow \infty$, whereas a suitably shrinking bandwidth can drive both approximation and estimation errors to zero. The bandwidth-dependent bounds in Corollary 6.2 and its explicit restatement, Corollary EC.3.5, provide guidance for this choice.

Under the Lipschitz condition in part (i), taking $h_t \asymp (\log t/t)^{1/(2d\mathbf{w}+2)}$ gives a coverage-error rate of $O_p((\log t/t)^{1/(2d\mathbf{w}+2)})$. Under the stronger second-order condition in part (ii), taking $h_t \asymp (\log t/t)^{1/(2d\mathbf{w}+4)}$ gives the faster rate $O_p((\log t/t)^{1/(d\mathbf{w}+2)})$. The dependence of these rates on $d\mathbf{w}$ reflects the usual curse of dimensionality in nonparametric estimation: higher-dimensional feature vectors require more data to obtain the same degree of local accuracy. Corollary EC.3.6 in E-Companion EC.3.2 gives the explicit bandwidths that minimize the corresponding finite-sample bounds. These statements concern deterministic bandwidth choices; data-dependent bandwidth selection is a separate practical extension.

6.3 Applying KSA to the 3-Factor Log-Linear Model

We now revisit our running example, in which the log security price follows a three-factor log-linear model. The portfolio is short a call option with fixed time to maturity τ and moneyness (defined as the strike price divided by the underlying price) m . In the simulated historical dataset, for each date $s < t$, we assume that the option has the same moneyness m and time-to-maturity τ on day s , with the portfolio gain G_{s+1} then computed on day $s + 1$. As detailed in E-Companion EC.1.5, KSA is applied to the normalized gain $G'_{s+1} := \frac{G_{s+1}}{S_s}$. This normalization removes the changing index-price scale across historical dates. Since S_s is positive and known at time s , a prediction interval for G'_{s+1} can be converted into an equivalent prediction interval for G_{s+1} by multiplying both endpoints

by S_s . We evaluate the KSA algorithm using 2000 time steps of historical data simulated under the model parameters specified in Section 4. We set the feature vector to $\mathbf{W}_t(\mathbf{z}) = \mathbf{z}$, so no historical context $\mathbf{H}_t$ is needed and the two stressed factor returns are sufficient as the feature vector for KSA. In E-Companion EC.1.5, we show that this choice of $\mathbf{W}_t(\mathbf{z})$ satisfies Assumption 2. Additional implementation details for KSA, including the choice of the kernel κ and the expectation predictor φ , are also deferred to E-Companion EC.1.5.

The KSA prediction intervals for the various stress scenarios are reported in Table 4. Among all methods considered, the KSA intervals are closest to the oracle intervals. For relatively “common” scenarios with $z_{\text{Oil}} \in \{-6\%, 0, 6\%\}$ and $z_{\text{Rate}} \in \{-10 \text{ bp}, 0, 10 \text{ bp}\}$, the conditional coverage rates all lie within $\pm 5\%$ of the target 50% level. For the other, more extreme scenarios, KSA still generally attains better coverage than ACSA, with the conditional coverage remaining within $\pm 10\%$ of the target 50% level. The numerical results show that KSA performs exceptionally well in this simple and idealized setting.

The strong performance in the running example reflects several idealizations. The centering predictor is a fixed oracle conditional mean, evaluated numerically as described in E-Companion EC.1.5, so the experiment abstracts from the estimation error that would arise when φ must be learned from data. We also assume that the relevant factor structure is known. Most importantly, the feature vector is only two-dimensional; in a more realistic Markov setting with a two-dimensional stress scenario and a d -dimensional state vector, its dimension would typically increase to $2 + d$, making nearby historical observations increasingly sparse and local kernel estimation more difficult.

Thus, the running example is best viewed as an illustration of KSA’s mechanics and potential performance under favorable conditions. In practice, high-dimensional conditioning is likely to be a key limitation, although good estimates of the mean predictor and factor structure can mitigate this issue when sufficient historical data are available. Section 7 considers a more realistic setting in which φ must be estimated and richer historical context vectors $\mathbf{H}_t$ are used.

7 Further Numerical Experiments

The 3-factor log-linear model introduced earlier served primarily as a simple running example for demonstrating the ACSA and KSA algorithms. In this section, we turn to a more complex and realistic setting, where we take the perspective of a portfolio manager and stress test the portfolio returns on a daily basis. We again adopt a factor model structure as described in Section 3.2 in order to simulate the market environment. The model parameters are calibrated with real world data to ensure realism and we present a complete pipeline for implementing our proposed methods in practice. We also conduct a systematic analysis of performance across different algorithmic configurations, including the standard scenario analysis as a benchmark.

7.1 Portfolio, Model and Experimental Setup

Available Securities The portfolio consists of out-of-the-money European call and put options written on the S&P 500 index, along with a possible position in the underlying index itself. This corresponds to the equity and options setting of Section 3.1, except that we now restrict attention to a single underlying security. We assume there are n tradable securities. The first is the S&P 500 index, with time t price S_t , and the remaining $n - 1$ securities are European options of varying strikes and maturities. Each option is characterized by its fixed time-to-maturity $\{\tau_j\}_{j=1}^{n-1}$ and fixed moneyness (the strike price K divided by S_t) $\{m_j\}_{j=1}^{n-1}$. The portfolio is rebalanced daily and we assume the option maturities and possible moneyness values remain constant over time. While these characteristics are held fixed, the implied volatility of each option $\{\sigma_t^{(j)}\}_{j=1}^{n-1}$ evolves over time. For simplicity, we assume the index pays no dividends, and the risk-free continuously compounded interest rate is fixed at $r = 4\%$. The number of units of each of the n securities in the portfolio is denoted by $\{u_j\}_{j=1}^n$. Although the portfolio is rebalanced over time, at each evaluation date t we fix the current holdings and recompute historical gains as if those same holdings had been held at all earlier dates. This produces a pseudo-history for the current portfolio, and the procedure is repeated after each rebalance. Theorem 6.1 concerns a fixed portfolio; recomputing historical gains does not by itself extend its guarantee to holdings selected using the observed history. We therefore evaluate this changing-portfolio implementation empirically.

Common Factor Model We apply the factor model framework introduced in Section 3.2 to the specific setting under consideration. We take the risk factors $\mathbf{x}_t$ to be

$$\mathbf{x}_t := \left(\log S_t, \log \sigma_t^{(1)}, \dots, \log \sigma_t^{(n-1)} \right).$$

Because the option parameters m_j , τ_j , and portfolio positions u_j are fixed at a given time t , there exists a deterministic mapping v such that the portfolio value $V_t = v(\mathbf{x}_t)$. We assume a 4-dimensional vector $\mathbf{f}_t = \left(f_t^{(1)}, f_t^{(2)}, f_t^{(3)}, f_t^{(4)} \right)$ of common risk factors. The first factor $f_t^{(1)} = \log(S_t)$ is the log S&P 500 index price, and the remaining three factors drive the log-implied volatility surface $\mathbf{I}_t := (\log \sigma_t^{(1)}, \dots, \log \sigma_t^{(n-1)})$. In particular, following Cont and Fonseca (2002) we assume

$$\mathbf{I}_{t+1} = \mathbf{I}_0 + f_{t+1}^{(2)} \cdot \mathbf{M}_2 + f_{t+1}^{(3)} \cdot \mathbf{M}_3 + f_{t+1}^{(4)} \cdot \mathbf{M}_4 \quad (27)$$

or equivalently,

$$\log S_{t+1}/S_t = \Delta f_{t+1}^{(1)}, \quad \Delta \mathbf{I}_{t+1} = \Delta f_{t+1}^{(2)} \cdot \mathbf{M}_2 + \Delta f_{t+1}^{(3)} \cdot \mathbf{M}_3 + \Delta f_{t+1}^{(4)} \cdot \mathbf{M}_4, \quad (28)$$

where $\Delta \mathbf{I}_{t+1} = \mathbf{I}_{t+1} - \mathbf{I}_t$ and $\mathbf{M}_2, \mathbf{M}_3, \mathbf{M}_4$ and $\mathbf{I}_0$ are constant vectors in $\mathbb{R}^{n-1}$. We note that (28) agrees with the model structure specified in (4), e.g., by setting $\boldsymbol{\mu}, \boldsymbol{\xi}_t \equiv \mathbf{0}$ and

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \mathbf{0} & \mathbf{M}_2 & \mathbf{M}_3 & \mathbf{M}_4 \end{bmatrix}.$$

Further details are provided in E-Companion EC.2.1.

Dynamics of Common Factors The first factor follows a Gaussian random walk with drift¹², and following Cont and Fonseca (2002), we assume the remaining three factors each follow an Ornstein-Uhlenbeck process. In particular, we assume

$$\begin{aligned} \Delta f_{t+1}^{(1)} &= \lambda_1 + \epsilon_{t+1}^{(1)}, \quad \epsilon_{t+1}^{(1)} \sim \mathcal{N}(0, \sigma_1^2) \\ \Delta f_{t+1}^{(i)} &= \lambda_i + \gamma_i f_t^{(i)} + \epsilon_{t+1}^{(i)}, \quad \epsilon_{t+1}^{(i)} \sim \mathcal{N}(0, \sigma_i^2) \quad \text{for } i = 2, 3, 4 \end{aligned} \quad (29)$$

where $\{\lambda_i\}_{i=1}^4, \{\gamma_i\}_{i=2}^4$ and $\{\sigma_i\}_{i=1}^4$ are model parameters. The random vector of the noise terms $(\epsilon_t^{(1)}, \epsilon_t^{(2)}, \epsilon_t^{(3)}, \epsilon_t^{(4)})$ is independent of everything else and across time t . We further assume that for each t , $\epsilon_t^{(1)}$ is correlated with $\epsilon_t^{(2)}, \epsilon_t^{(3)}$ and $\epsilon_t^{(4)}$ through the following relationship

$$\epsilon_t^{(1)} = \eta_2 \epsilon_t^{(2)} + \eta_3 \epsilon_t^{(3)} + \eta_4 \epsilon_t^{(4)} + \epsilon_t^G, \quad (30)$$

where $\epsilon_t^G \sim \mathcal{N}(0, \sigma_G^2)$ and is independent of everything else. The variance σ_G^2 and the constants η_2, η_3, η_4 must satisfy

$$\eta_2^2 \sigma_2^2 + \eta_3^2 \sigma_3^2 + \eta_4^2 \sigma_4^2 + \sigma_G^2 = \sigma_1^2 \quad (31)$$

to ensure that the variance of $\epsilon_t^{(1)}$ is equal to σ_1^2 .

Backtesting Process and Evaluation Metrics We backtest several algorithms on synthetic data generated from the common factor model. At each time step t , each algorithm constructs a prediction interval for the scenario realized on the following day. The algorithm's performance is then evaluated based on the portfolio gain realized on that next day. We report four evaluation metrics: (1) the **realized coverage rate**, i.e. the fraction of time steps the realized portfolio gain falls within the prediction interval; (2) the **prediction interval (PI) width**, i.e. the width of the prediction intervals averaged across time steps; (3) the **interval (I) score**, which is a proper scoring rule (Gneiting and Raftery, 2007) that jointly assesses coverage rate and sharpness (further details are provided in E-Companion EC.2.2); and (4) the **root mean squared error (RMSE)**, i.e. the square root of the average (over all evaluation dates) of the squared difference between the realized portfolio gain and their point predictions¹³. Lower PI widths, I scores, and RMSE values

¹²This setting corresponds to the underlying index price S_t following a geometric Brownian motion.

¹³For ACSA, which does not produce an explicit point prediction for the expected portfolio gain, we use the midpoint of the prediction interval as its point prediction when computing RMSE.

indicate better performance.

7.2 The Skew Steepener Portfolio

We consider a **skew steepener** portfolio that is long out-of-the-money (OTM) put options and short OTM call options at 1-month, 3-month, and 6-month maturities. The corresponding moneyness levels are 5%, 10%, and 20% OTM, respectively. For example, if the underlying index level is \$5000, the put strikes are \$4750, \$4500, and \$4000, while the call strikes are \$5250, \$5500, and \$6000. The common option position size u^{skew} is chosen so that the combined market value of the long put positions equals \$100. An additional underlying-index position u^{hedge} is used to maintain delta neutrality. The resulting seven-position portfolio is rebalanced daily.

We stress the first two common factor returns, $\Delta \mathbf{f}_{t+1}^{(1:2)}$. The experiment generates 6000 observations from the dynamic factor model, with the first 2000 used as burn-in and the remaining 4000 used for evaluation. All methods construct 90% prediction intervals.

Table 5 compares ACSA and KSA with two empirical-quantile benchmarks and a Monte Carlo oracle. Implementation details are deferred to E-Companion EC.2.3, and computational times are reported in E-Companion EC.2.4. The “MC Oracle” uses Monte Carlo approximations to the oracle conditional quantiles under the known data-generating process. The two “Emp-quantile” benchmarks instead use historical portfolio gains recomputed under the current portfolio configuration. The “Vanilla” version uses their empirical quantiles directly, while the “SSA” version adds the current SSA point prediction to empirical quantiles of historical SSA residuals. Neither benchmark conditions the interval width on the current scenario or market state.

For ACSA, we consider three choices of quantile predictor g_t . “Linear- g_t ” uses linear quantile regression, “NN- g_t ” uses a two-layer neural network with a 20-dimensional hidden layer, and “MC Oracle- g_t ” uses Monte Carlo approximations to the oracle conditional quantiles. The learned predictors use $\Delta \mathbf{f}_{t+1}^{(1:2)}$ as features and are trained using the pinball loss. Because the portfolio evolves smoothly over time, the learned predictor is fine-tuned from the previous day’s model rather than retrained from scratch. All three ACSA variants attain coverage close to the 90% target. The nonlinear and oracle predictors also produce interval widths and scores close to the MC Oracle benchmark, whereas the linear predictor yields noticeably wider intervals. This illustrates the importance of the quality of the underlying quantile predictor for interval width and score.

For KSA, we study separately the effects of the historical context $\mathbf{H}_t$ and the expectation predictor φ . In the “KSA NN- φ ” block, φ is a two-layer neural network and we compare $\mathbf{H}_t = \mathbf{f}_t^{(2:4)}$, $\mathbf{f}_t^{(2:3)}$, and $\mathbf{f}_t^{(2)}$. The index-price level $f_t^{(1)}$ is excluded because, under our model, the next-day index return is independent of the current index level and the normalized option gain depends on the return rather than the price level itself; see E-Companion EC.2.5. In the “KSA $\mathbf{H}_t = \mathbf{f}_t^{(2:4)}$ ” block, we instead fix the full historical context and compare no centering ($\varphi \equiv 0$), the SSA predictor, a neural-network predictor, and a Monte Carlo oracle predictor.

Algorithm	Variants	Coverage	PI Width ↓	I Score ↓	RMSE ↓
MC Oracle		90.0%	27.2	1.72	8.59
Emp-quantile	Vanilla	99.8%	163.4	8.17	13.61
	SSA	92.1%	53.4	2.98	9.25
ACSA	Linear- g_t	89.8%	30.7	2.27	11.05
	NN- g_t	89.9%	28.5	1.89	9.27
	MC Oracle- g_t	89.9%	28.1	1.78	8.72
KSA NN- φ	$\mathbf{H}_t = \mathbf{f}_t^{(2:4)}$	90.0%	28.0	1.75	8.76
	$\mathbf{H}_t = \mathbf{f}_t^{(2:3)}$	91.7%	31.3	1.81	8.76
	$\mathbf{H}_t = \mathbf{f}_t^{(2)}$	92.6%	37.4	1.92	8.76
KSA $\mathbf{H}_t = \mathbf{f}_t^{(2:4)}$	no- φ	93.3%	50.6	3.87	16.6
	SSA- φ	92.6%	34.6	1.94	9.25
	NN- φ	90.0%	28.0	1.75	8.76
	MC Oracle- φ	90.1%	28.2	1.74	8.73
ACSA + KSA	$\mathbf{H}_t = \mathbf{f}_t^{(2:4)}$	90.5%	32.4	1.94	8.86
	$\mathbf{H}_t = \mathbf{f}_t^{(2:3)}$	90.7%	32.2	2.18	8.86
	$\mathbf{H}_t = \mathbf{f}_t^{(2)}$	90.3%	31.8	2.27	8.86

Table 5: MC Oracle denotes a Monte Carlo approximation to the oracle conditional quantities. Backtesting performance of the prediction intervals generated by the ACSA and KSA algorithms. The evaluation metrics include the realized coverage rate (Coverage), the prediction interval width (PI Width), the interval score (I Score), and the root mean squared error (RMSE). The target coverage for the prediction intervals is 90%. Smaller values of PI Width, I Score and RMSE all indicate better performance. Each main row represents a major class of algorithms, and each subrow corresponds to a different variant.

The results show that KSA improves as either the historical context or the centering predictor becomes more informative. With $\mathbf{H}_t = \mathbf{f}_t^{(2:4)}$ and either NN- φ or MC Oracle- φ , coverage is essentially at the 90% target and the remaining performance metrics are close to the MC Oracle benchmark. Performance deteriorates moderately when either $\mathbf{H}_t$ or φ is misspecified, illustrating the value of both informative conditioning variables and an effective centering predictor.

Finally, the “ACSA + KSA” variants use the KSA conditional-quantile estimator as the quantile predictor within ACSA, while retaining the online update of $\tilde{\alpha}_t$. A neural network is used for φ under the same three choices of $\mathbf{H}_t$. The hybrid achieves coverage close to the nominal level and performance comparable to standard ACSA. It can also be computationally attractive when portfolios change frequently, since the learned component is an expectation predictor rather than a portfolio-level quantile model.

7.3 An Adversarial Portfolio

We now consider a portfolio manager operating under the oversight of a risk management team. The manager forms opinions regarding future market conditions and aims to maximize the expected portfolio returns while simultaneously complying with risk management requirements. We assume the risk management requirements are specified via SSA across a range of predefined stress scenarios.

An *adversarial action* arises when the portfolio manager has access to information that is not available to the risk management team and exploits this informational asymmetry to construct portfolios that appear “safe” under the prescribed stress tests, yet contain directional exposure to risk factors that are not monitored by the risk management team. We assume the portfolio manager knows the true four-factor model, whereas the risk management team monitors only the first two factors. Under this setup, the portfolio manager can construct a portfolio that is neutral or safe with respect to the monitored factors according to SSA, but effectively represents a bet on upward movements of the third factor. We refer to such a portfolio as an **adversarial portfolio**. We will show that SSA fails to detect the hidden risks undertaken by the adversarial portfolio, whereas the ACSA and KSA methods are able to reveal them.

The adversarial portfolio consists of the S&P 500 index together with 12 options. The options have maturities of 1-month, 3-month, and 6-month, with moneyness levels of 5% and 10% OTM for both call and put options. The portfolio manager is allowed to take short positions in any of these securities. The total notional value of long positions is constrained to be at most \$100, and the sum of absolute position sizes across all securities is required to remain below \$500. The adversarial portfolio is constructed by solving a linear programming (LP) problem. The LP maximizes a linear approximation of the expected portfolio gain under the betting movement (i.e., an upward movement of the third common factor). The constraints enforce that all SSA evaluations remain within the range $\pm\$3$, along with additional constraints on the total notional value of long positions and the sum of absolute position sizes. Further details of the portfolio construction are contained

Algorithm	Variants	Coverage	PI Width ↓	I Score ↓	RMSE ↓
MC Oracle		90.0%	15.8	1.00	5.42
Emp-quantile	Vanilla	99.6%	2645.4	132.3	4129.4
	SSA	91.7%	1113.3	55.9	5.86
ACSA	Linear- g_t	89.7%	56.2	3.54	7.56
	NN- g_t	89.9%	19.3	1.69	5.67
	MC Oracle- g_t	90.0%	16.5	1.06	5.44
KSA $\mathbf{H}_t = \mathbf{f}_t^{(2)}$	no- φ	93.8%	32.6	2.35	8.83
	SSA- φ	87.3%	19.1	1.59	5.86
	NN- φ	92.6%	20.9	1.42	5.43
	MC Oracle- φ	90.2%	16.9	1.14	5.40
ACSA + KSA	$\mathbf{H}_t = \mathbf{f}_t^{(2)}$	90.2%	29.1	1.85	5.87

Table 6: Backtesting performance of the prediction intervals under the adversarial portfolio. The table structure mirrors that of Table 5, but we only report results for $\mathbf{H}_t = \mathbf{f}_t^{(2)}$ in the KSA setting, since the risk management team has access only to the first two common factors. For the KSA SSA- φ row, RMSE is computed from the SSA point predictor, matching the Emp-quantile SSA row.

in E-Companion EC.2.7. The backtesting results for the adversarial portfolio are reported in Table 6, which is organized analogously to Table 5. Because the risk management team has access only to the first two common factors, information on $f_t^{(3)}$ and $f_t^{(4)}$ cannot be included in the feature inputs when training the neural networks. This restriction does not affect ACSA, since it relies only on the first two factors (recall that “Linear- g_t ” and “NN- g_t ” correspond to training a linear model and a neural network as the quantile predictor, using $\Delta \mathbf{f}_{t+1}^{(1:2)}$ as the feature input). The KSA, however, is evaluated only under the configuration $\mathbf{H}_t = \mathbf{f}_t^{(2)}$, rather than the full historical context $\mathbf{H}_t = \mathbf{f}_t^{(2:4)}$. Accordingly, the expectation predictor φ used by KSA is also trained under the restricted information set, with $(\Delta \mathbf{f}_{t+1}^{(1:2)}, \mathbf{f}_t^{(2)})$ as the feature input.

We observe that the performance patterns across different algorithms in Table 6 closely mirror those observed in Table 5. The two “Emp-quantile”-based benchmarks produce exceptionally wide intervals. In contrast, our proposed algorithms demonstrate large and consistent advantages across different portfolio constructions. One caveat is that, due to the absence of a reliable φ , the overall performance of KSA still shows a noticeable gap relative to the Monte Carlo oracle benchmark.

The next experiment demonstrates the scenario analysis table at a specific date. We show that, although the adversarial portfolio appears neutral to risk across a range of scenarios under the SSA estimation, it nonetheless produces notably wide prediction intervals using the ACSA and KSA algorithms that we propose. The width of the prediction interval serves as a warning signal for the risk management team. For comparison, we analyze a **first-order four-factor-neutral benchmark portfolio**.¹⁴ Its optimization imposes additional constraints requiring the linearized

¹⁴We use “factor-neutral” here in a relative sense. It does not imply sensitivity to each factor is exactly zero. Rather, relative to the adversarial portfolio, this benchmark is more neutral with respect to the factor directions.

one-day gain to vanish under unit movements in each of the four common-factor directions. Thus, relative to the adversarial construction, the benchmark is more tightly controlled with respect to the modeled factor directions. The construction adds constraints for the third and fourth factor directions to the adversarial-portfolio formulation; see E-Companion [EC.2.7](#).

In Table 7, we report the SSA estimates and the Monte Carlo approximations to the oracle conditional expectations of portfolio gains under various stress scenarios. By construction, the SSA estimates for both portfolios are tightly controlled, remaining within $\pm\$3$ relative to the initial long position of \$100. However, the Monte Carlo approximations to the oracle conditional expectations $\mathbb{E}[G_{t+1} \mid \mathbf{W}_{t+1}(\mathbf{z})]$ under each scenario reveal substantial differences between the two portfolios. In particular, the expected gains of the factor-neutral benchmark are roughly bounded within $\pm\$6.5$, whereas the adversarial portfolio can incur expected losses as large as $-\$20$. This discrepancy highlights the risk of relying solely on SSA-based assessments. Our proposed ACSA and KSA methods address this issue, as shown below.

SSA: Adversarial Portfolio								$\mathbb{E}[G_{t+1} \mid \mathbf{W}_{t+1}(\mathbf{z})]$: Adversarial Portfolio							
$\Delta f_{t+1}^{(1)}$ ($\times 10^{-2}$)	$\Delta f_{t+1}^{(2)}$ ($\times 10^{-1}$)							$\Delta f_{t+1}^{(1)}$ ($\times 10^{-2}$)	$\Delta f_{t+1}^{(2)}$ ($\times 10^{-1}$)						
	-3	-2	-1	0	1	2	3		-3	-2	-1	0	1	2	3
-1.5	0.86	0.10	-0.62	-1.29	-1.91	-2.48	-3.00	-1.5	2.66	3.00	3.39	3.80	4.11	4.54	4.94
-1.0	1.57	1.09	0.65	0.25	-0.10	-0.41	-0.68	-1.0	1.43	2.42	3.28	4.13	5.00	5.83	6.61
-0.5	1.81	1.60	1.43	1.29	1.20	1.15	1.13	-0.5	-0.77	0.78	2.17	3.56	4.92	6.16	7.39
0.0	1.52	1.58	1.67	1.80	1.96	2.15	2.37	0.0	-3.98	-1.91	0.06	1.98	3.73	5.53	7.20
0.5	0.65	0.97	1.32	1.70	2.11	2.54	3.00	0.5	-8.29	-5.75	-3.11	-0.78	1.57	3.83	5.91
1.0	-0.84	-0.26	0.34	0.96	1.60	2.27	2.95	1.0	-13.87	-10.61	-7.54	-4.58	-1.68	1.00	3.65
1.5	-3.00	-2.18	-1.34	-0.49	0.39	1.28	2.18	1.5	-20.47	-16.72	-13.14	-9.57	-6.16	-2.98	0.10
SSA: Factor-neutral benchmark								$\mathbb{E}[G_{t+1} \mid \mathbf{W}_{t+1}(\mathbf{z})]$: Factor-neutral benchmark							
$\Delta f_{t+1}^{(1)}$ ($\times 10^{-2}$)	$\Delta f_{t+1}^{(2)}$ ($\times 10^{-1}$)							$\Delta f_{t+1}^{(1)}$ ($\times 10^{-2}$)	$\Delta f_{t+1}^{(2)}$ ($\times 10^{-1}$)						
	-3	-2	-1	0	1	2	3		-3	-2	-1	0	1	2	3
-1.5	-0.11	-0.58	-1.02	-1.42	-1.79	-2.14	-2.45	-1.5	-0.23	-1.34	-2.40	-3.43	-4.40	-5.31	-6.17
-1.0	-0.48	-0.77	-1.04	-1.27	-1.48	-1.66	-1.82	-1.0	0.01	-0.72	-1.44	-2.12	-2.77	-3.38	-3.94
-0.5	-0.73	-0.85	-0.94	-1.01	-1.05	-1.07	-1.06	-0.5	-0.16	-0.51	-0.85	-1.17	-1.47	-1.74	-2.01
0.0	-0.93	-0.87	-0.78	-0.68	-0.55	-0.41	-0.24	0.0	-0.84	-0.78	-0.72	-0.66	-0.59	-0.51	-0.42
0.5	-1.12	-0.88	-0.62	-0.34	-0.05	0.26	0.58	0.5	-2.06	-1.59	-1.13	-0.67	-0.21	0.24	0.68
1.0	-1.36	-0.94	-0.51	-0.06	0.39	0.87	1.35	1.0	-3.92	-3.02	-2.14	-1.27	-0.42	0.42	1.24
1.5	-1.70	-1.11	-0.51	0.10	0.73	1.36	2.00	1.5	-6.43	-5.11	-3.82	-2.53	-1.27	-0.04	1.16

Table 7: SSA point estimates and the Monte Carlo oracle conditional mean gains (measured in \$) across the scenario grid defined by the first two common factor returns $(\Delta f_{t+1}^{(1)}, \Delta f_{t+1}^{(2)})$. The two tables in the top row correspond to the adversarial portfolio and report the SSA estimates and the Monte Carlo oracle conditional expected scenario gain $\mathbb{E}[G_{t+1} \mid \mathbf{W}_{t+1}(\mathbf{z})]$. The bottom row reports the corresponding quantities for the factor-neutral benchmark. Although the SSA estimates for the two portfolios appear similar (with values contained within $\pm\$3$), the Monte Carlo approximations to the oracle conditional expectations differ substantially, in particular the adversarial portfolio can experience scenario losses as large as $-\$20$.

Table 8 reports the scenario tables of prediction intervals generated by our proposed algorithms,

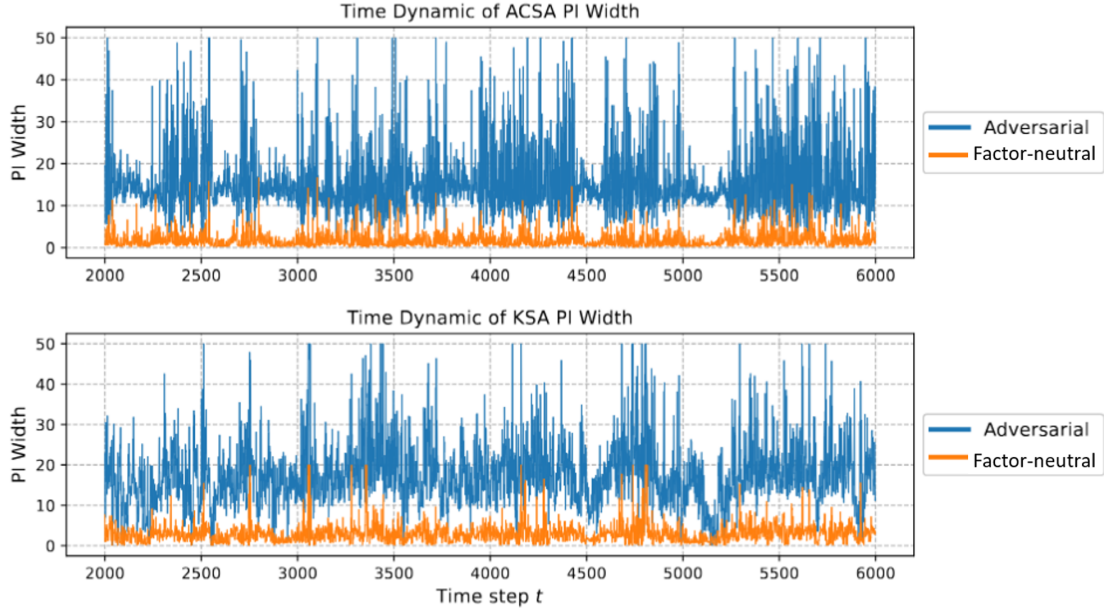


Figure 2: Dynamics of ACSA and KSA’s prediction interval (PI) widths, for the adversarial portfolio and the factor-neutral benchmark. On average, the adversarial portfolio has notably wider prediction intervals than the factor-neutral benchmark.

evaluated at a specific day (additional results for a different day are provided in the E-Companion). In the ACSA algorithm, a two-layer neural network is trained as the quantile predictor g_t . In the KSA algorithm, the SSA estimate is used as the expectation predictor φ , since the tables are produced from the perspective of a risk manager and we assume that this represents the best available estimator in practice. Comparing the prediction intervals for the two portfolios, we observe that, under both algorithms, the intervals for the adversarial portfolio are substantially wider than those for the factor-neutral benchmark. This finding suggests that the width of the prediction intervals can serve as an additional indicator of portfolio risk, beyond relying on a single SSA estimate alone. A clearer illustration of the time dynamics of the prediction interval width (under the realized scenarios) is provided in Figure 2, which shows that the prediction intervals of the adversarial portfolio are consistently wider than those of the factor-neutral benchmark.

ACSA 90% PI ($\tilde{\alpha}_t = 0.08$): Adversarial Portfolio							
$\Delta f_{t+1}^{(1)}$ ($\times 10^{-2}$)	$\Delta f_{t+1}^{(2)} (\times 10^{-1})$						
	-3	-2	-1	0	1	2	3
-1.5	(-7.1, 12.8)	(-6.0, 12.5)	(-4.9, 12.2)	(-3.8, 11.9)	(-2.8, 11.6)	(-1.8, 11.4)	(-0.9, 11.2)
-1.0	(-8.4, 11.8)	(-6.7, 12.0)	(-5.1, 12.2)	(-3.6, 12.3)	(-2.1, 12.5)	(-0.7, 12.6)	(0.7, 12.8)
-0.5	(-10.9, 9.8)	(-8.6, 10.5)	(-6.4, 11.2)	(-4.4, 11.9)	(-2.4, 12.5)	(-0.5, 13.1)	(1.3, 13.7)
0.0	(-14.4, 6.8)	(-11.6, 8.1)	(-8.9, 9.3)	(-6.2, 10.5)	(-3.7, 11.6)	(-1.4, 12.7)	(0.9, 13.7)
0.5	(-19.1, 2.9)	(-15.7, 4.7)	(-12.4, 6.4)	(-9.3, 8.1)	(-6.2, 9.7)	(-3.4, 11.2)	(-0.6, 12.7)
1.0	(-25.0, -2.3)	(-21.0, 0.1)	(-17.2, 2.4)	(-13.5, 4.6)	(-9.9, 6.7)	(-6.5, 8.8)	(-3.3, 10.7)
1.5	(-32.1, -8.6)	(-27.6, -5.6)	(-23.2, -2.8)	(-18.9, -0.0)	(-14.8, 2.6)	(-10.9, 5.2)	(-7.2, 7.6)
ACSA 90% PI ($\tilde{\alpha}_t = 0.09$): Factor-neutral benchmark							
$\Delta f_{t+1}^{(1)}$ ($\times 10^{-2}$)	$\Delta f_{t+1}^{(2)} (\times 10^{-1})$						
	-3	-2	-1	0	1	2	3
-1.5	(-1.3, 1.2)	(-2.3, -0.2)	(-3.5, -1.4)	(-4.7, -2.3)	(-5.9, -3.1)	(-7.1, -3.7)	(-8.2, -4.2)
-1.0	(-0.9, 1.3)	(-1.4, 0.2)	(-2.2, -0.7)	(-3.1, -1.3)	(-4.0, -1.7)	(-4.9, -1.9)	(-5.8, -2.1)
-0.5	(-1.2, 1.2)	(-1.0, 0.4)	(-1.3, -0.3)	(-1.8, -0.7)	(-2.4, -0.6)	(-3.0, -0.5)	(-3.6, -0.3)
0.0	(-2.1, 0.6)	(-1.6, 0.2)	(-1.1, -0.2)	(-0.9, -0.3)	(-1.2, 0.1)	(-1.5, 0.6)	(-1.8, 1.1)
0.5	(-3.7, -0.4)	(-2.8, -0.4)	(-1.9, -0.5)	(-1.1, -0.3)	(-0.6, 0.4)	(-0.5, 1.2)	(-0.5, 2.0)
1.0	(-6.0, -1.9)	(-4.6, -1.5)	(-3.4, -1.1)	(-2.1, -0.6)	(-1.1, 0.3)	(-0.2, 1.4)	(0.2, 2.5)
1.5	(-9.0, -4.0)	(-7.2, -3.2)	(-5.5, -2.3)	(-3.9, -1.4)	(-2.3, -0.3)	(-1.0, 1.0)	(0.1, 2.4)
KSA 90% PI: Adversarial Portfolio							
$\Delta f_{t+1}^{(1)}$ ($\times 10^{-2}$)	$\Delta f_{t+1}^{(2)} (\times 10^{-1})$						
	-3	-2	-1	0	1	2	3
-1.5	(-7.3, 11.4)	(-5.9, 11.7)	(-5.0, 13.5)	(-4.3, 19.4)	(-2.7, 18.7)	(-2.8, 18.2)	(-2.6, 17.6)
-1.0	(-7.7, 11.8)	(-7.1, 11.3)	(-5.9, 11.2)	(-4.8, 14.4)	(-3.2, 20.5)	(-1.2, 20.2)	(-1.0, 20.0)
-0.5	(-10.8, 11.0)	(-7.7, 11.2)	(-6.8, 11.6)	(-5.5, 11.8)	(-4.3, 13.9)	(-2.0, 21.8)	(0.2, 21.8)
0.0	(-12.3, 9.3)	(-10.5, 9.8)	(-8.2, 11.1)	(-6.8, 11.4)	(-5.0, 12.1)	(-3.8, 14.3)	(-0.7, 21.9)
0.5	(-13.2, 8.5)	(-12.4, 8.8)	(-10.0, 9.2)	(-8.7, 9.5)	(-6.9, 10.4)	(-5.0, 12.1)	(-3.7, 13.5)
1.0	(-11.2, 7.0)*	(-13.0, 7.6)	(-12.4, 8.2)	(-10.4, 8.6)	(-9.0, 9.2)	(-7.1, 9.9)	(-5.4, 11.2)
1.5	(-13.0, 4.8)*	(-12.5, 5.6)*	(-11.7, 6.5)*	(-11.9, 7.3)	(-11.0, 7.3)	(-9.3, 8.7)	(-8.1, 9.8)
KSA 90% PI: Factor-neutral benchmark							
$\Delta f_{t+1}^{(1)}$ ($\times 10^{-2}$)	$\Delta f_{t+1}^{(2)} (\times 10^{-1})$						
	-3	-2	-1	0	1	2	3
-1.5	(-3.1, 3.1)	(-3.9, 2.3)	(-4.5, 1.3)	(-5.2, 0.9)	(-5.9, 0.5)	(-6.2, 0.2)	(-6.5, -0.4)
-1.0	(-3.3, 3.0)	(-3.8, 2.4)	(-4.3, 1.8)	(-4.8, 1.3)	(-5.2, 0.8)	(-5.7, 0.7)	(-5.9, 0.5)
-0.5	(-3.7, 2.9)	(-3.7, 2.6)	(-3.9, 2.2)	(-4.1, 1.9)	(-4.3, 1.7)	(-4.7, 1.3)	(-5.1, 1.3)
0.0	(-4.1, 4.6)	(-4.0, 2.6)	(-3.7, 2.6)	(-3.7, 2.5)	(-3.6, 2.3)	(-3.6, 2.4)	(-3.8, 2.2)
0.5	(-4.7, 6.9)	(-4.0, 4.8)	(-3.8, 2.6)	(-3.3, 2.8)	(-3.0, 3.1)	(-2.8, 3.2)	(-2.5, 3.5)
1.0	(-4.5, 6.7)	(-4.1, 7.1)	(-3.7, 4.9)	(-3.2, 3.1)	(-2.8, 3.5)	(-2.2, 3.9)	(-1.7, 4.3)
1.5	(-4.3, 6.4)*	(-4.3, 6.9)*	(-3.7, 7.5)*	(-3.1, 4.0)	(-2.4, 3.7)	(-1.8, 4.3)	(-1.2, 5.0)

Table 8: ACSA and KSA 90% prediction intervals for the adversarial portfolio and the factor-neutral benchmark. In ACSA, the two portfolios follow different trajectories in the time dynamics of $\tilde{\alpha}_t$, and therefore the quantile arguments differ at the evaluated time step for the two tables ($\tilde{\alpha}_t = 0.08$ for the adversarial portfolio and $\tilde{\alpha}_t = 0.09$ for the factor-neutral benchmark). The KSA uses SSA estimation as the mean predictor. A red * superscript indicates scenarios where the prediction interval *does not contain* the Monte Carlo oracle conditional expected gain $\mathbb{E}[G_{t+1} \mid \mathbf{W}_{t+1}(\mathbf{z})]$ reported in Table 7. As was the case with Table 4, this is only a diagnostic for interval location relative to the conditional mean $\mathbb{E}[G_{t+1} \mid \mathbf{W}_{t+1}(\mathbf{z})]$. In particular, exclusion of $\mathbb{E}[G_{t+1} \mid \mathbf{z}]$ is not a validity failure since the intervals target realized $G_{t+1} \mid \mathbf{W}_{t+1}(\mathbf{z})$ and not $\mathbb{E}[G_{t+1} \mid \mathbf{W}_{t+1}(\mathbf{z})]$.

8 Conclusions and Further Research

Scenario analysis is widely used in financial risk management, but standard implementations typically report only point estimates of scenario P&L with little statistical validation. We instead formulate scenario analysis as a sequential prediction problem and augment scenario tables with prediction intervals for realized scenario gains. ACSA provides online marginal calibration, while KSA uses the stressed scenario and current market state to estimate scenario-conditional uncertainty.

Our experiments show that these methods can provide more informative assessments than standard scenario analysis, particularly when dependence, nonlinear exposures, or unmonitored risk factors make conventional scenario P&L estimates misleading. The two approaches are complementary: ACSA offers a broadly applicable calibration mechanism, whereas KSA provides more scenario-specific uncertainty estimates when sufficient structure and historical support are available. KSA can also be computationally attractive for changing portfolios when conditional expectations are estimated at the security level and aggregated across positions.

Several directions for future research remain. One is to improve the scalability of KSA when the conditioning information H_t is high-dimensional. Promising approaches include structured or additive kernels, sparsity assumptions, and learned low-dimensional representations $r(H_t)$ designed to preserve the information relevant for scenario-loss quantiles. More generally, this amounts to learning an efficient notion of similarity between market states while retaining useful conditional-coverage behavior.

A second direction is to integrate scenario design more closely with statistical calibration. Rather than specifying a scenario set first and validating it only afterwards, one could incorporate plausibility and severity directly into the learning and calibration process, for example by allocating more resolution or weight to plausible but severe scenarios. This could lead to scenario tables that are jointly designed to reflect risk-management objectives and to support systematic statistical validation.

References

- Acerbi, Carlo, Balazs Szekely. 2014. Backtesting expected shortfall: Introducing three model-independent, non-parametric back-test methodologies for expected shortfall. Tech. rep., MSCI Research Insights. URL https://www.msci.com/resources/research/articles/2014/Research_Insight_Backtesting_Expected_Shortfall_December_2014.pdf.
- Alfaro, R.A., M. Drehmann. 2009. Macro stress tests and crises: What can we learn. *BIS Quarterly Review* 29–41.
- Angelopoulos, Anastasios N, Stephen Bates. 2021. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511* .
- Basel Committee on Banking Supervision. 2005. International convergence of capital measurement and capital standards: A revised framework. Tech. rep.
- Black, Fischer, Robert B Litterman. 1991. Asset allocation. *The Journal of Fixed Income* **1**(2) 7–18.
- Bonti, G., M. Kalkbrener, C. Lotz, G. Stahl. 2006. Credit risk concentrations under stress. *Journal of Credit Risk* **2** 115–136.
- Borio, C., M. Drehmann, K. Tsatsaronis. 2012. Stress-testing macro stress testing: Does it live up to expectations? Tech. rep., Bank for International Settlements.
- Boudabsa, Lotfi, Damir Filipović. 2022. Machine learning with kernels for portfolio valuation and risk management. *Finance and Stochastics* **26**(2) 131–172. doi:10.1007/s00780-021-00465-4.
- Bradley, Richard C. 2005. Basic properties of strong mixing conditions. a survey and some open questions .
- Breuer, T., M. Jandacka, K. Rheinberger, M. Summer. 2009. How to find plausible, severe, and useful stress scenarios. *International Journal of Central Banking* **5** 205–224.
- Cai, Zongwu. 2002. Regression quantiles for time series. *Econometric Theory* **18**(1) 169–192. doi:10.1017/S0266466602181096.
- Cambou, Mathieu, Damir Filipović. 2017. Model uncertainty and scenario aggregation. *Mathematical Finance* **27**(2) 534–567. doi:10.1111/mafi.12097.
- Chen, Minshuo, Renyuan Xu, Yumin Xu, Ruixun Zhang. 2025. Diffusion factor models: Generating high-dimensional returns with factor structure. doi:10.48550/arXiv.2504.06566. URL <https://arxiv.org/abs/2504.06566>. Working paper, arXiv:2504.06566. Revised January 2026.
- Cont, R., J. Da Fonseca. 2002. Dynamics of implied volatility surfaces. *Quantitative Finance* **2**(1) 45–60.

- Cont, Rama, Mihai Cucuringu, Renyuan Xu, Chao Zhang. 2026. Tail-GAN: Learning to simulate tail risk scenarios. *Management Science* **72**(4) 2917–2936. doi:10.1287/mnsc.2023.00936.
- Diebold, F. X., C. Li. 2006. Forecasting the term structure of government bond yields. *Journal of Econometrics* **130** 337–364.
- Du, Zaichao, Juan Carlos Escanciano. 2017. Backtesting expected shortfall: Accounting for tail risk. *Management Science* **63**(4) 940–958. doi:10.1287/mnsc.2015.2342.
- European Central Bank. 2006. Country-level macro stressing practices. *Financial Stability Review* 147–154.
- Fama, Eugene F., Kenneth R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* **33**(1) 3–56. doi:10.1016/0304-405X(93)90023-5.
- Fissler, Tobias, Johanna F. Ziegel, Tilmann Gneiting. 2016. Expected shortfall is jointly elicitable with value-at-risk: Implications for backtesting. *Risk* **2016**(January) 58–61. URL <https://arxiv.org/abs/1507.00244>. Working note in Risk Magazine.
- Flood, M.D., G. Korenko. 2015. Systematic scenario selection: stress testing and the nature of uncertainty. *Quantitative Finance* **15**(1) 43–59.
- Gibbs, Isaac, Emmanuel Candes. 2021. Adaptive conformal inference under distribution shift. *Advances in Neural Information Processing Systems* **34** 1660–1672.
- Gibbs, Isaac, Emmanuel J Candès. 2024. Conformal inference for online prediction with arbitrary distribution shifts. *Journal of Machine Learning Research* **25**(162) 1–36.
- Glasserman, P., C. Kang, W. Kang. 2015. Stress scenario selection by empirical likelihood. *Quantitative Finance* **15**(1) 25–41.
- Gneiting, Tilmann, Adrian E Raftery. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* **102**(477) 359–378.
- Golub, B., D. Greenberg, R. Ratcliffe. 2018. Market-driven scenarios: An approach for plausible scenario construction. *The Journal of Portfolio Management* .
- Hansen, Bruce E. 2008. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24**(3) 726–748. URL <http://www.jstor.org/stable/20142515>.
- Haugh, Martin B., Octavio Ruiz Lacedelli. 2020. Scenario analysis for derivative portfolios via dynamic factor models. *Quantitative Finance* **20**(4) 547–571.
- He, Xue Dong, Steven Kou, Xianhua Peng. 2022. Risk measures: Robustness, elicibility, and backtesting. *Annual Review of Statistics and Its Application* **9**(Volume 9, 2022) 141–166.

- Hoga, Yannick, Matei Demetrescu. 2023. Monitoring value-at-risk and expected shortfall forecasts. *Management Science* **69**(5) 2954–2971. doi:10.1287/mnsc.2022.4460.
- Koenker, Roger. 2005. *Quantile regression*, vol. 38. Cambridge university press.
- Lee, Jonghyeok, Chen Xu, Yao Xie. 2026. Kernel-based optimally weighted conformal time-series prediction. URL <https://arxiv.org/abs/2405.16828>.
- Lei, Jing, Max G'Sell, Alessandro Rinaldo, Ryan J Tibshirani, Larry Wasserman. 2018. Distribution-free predictive inference for regression. *Journal of the American Statistical Association* **113**(523) 1094–1111.
- McNeil, A. J., R. Frey, P. Embrechts. 2015. *Quantitative risk management: concepts, techniques, and tools*. Princeton university press.
- Meucci, A. 2008. The Black-Litterman approach: Original model and extensions. Tech. rep., SSRN. URL <https://ssrn.com/abstract=1117574> or <http://dx.doi.org/10.2139/ssrn.1117574>.
- Nolde, Natalia, Johanna F. Ziegel. 2017. Elicitability and backtesting: Perspectives for banking regulation. *Annals of Applied Statistics* **11**(4) 1833–1874. doi:10.1214/17-AOAS1041.
- Papadopoulos, Harris, Kostas Proedrou, Volodya Vovk, Alex Gammerman. 2002. Inductive confidence machines for regression. *Machine Learning: ECML 2002: 13th European Conference on Machine Learning Helsinki, Finland, August 19–23, 2002 Proceedings 13*. Springer, 345–356.
- Pelger, Markus, Ruoxuan Xiong. 2022. State-varying factor models of large dimensions. *Journal of Business & Economic Statistics* **40**(3) 1315–1333. doi:10.1080/07350015.2021.1927744.
- Quagliarello, M. 2009. *Stress-Testing the Banking System*. Cambridge University Press, Cambridge, U.K.
- Rebonato, R. 2010. *Coherent Stress Testing: A Bayesian Approach to the Analysis of Financial Stress*. John Wiley & Sons.
- Rebonato, R. 2019. A financially justifiable and practically implementable approach to coherent stress testing. *Quantitative Finance* **19**(5) 827–842.
- Roth, Aaron. 2022. Uncertain: Modern topics in uncertainty estimation .
- Ruppert, D., S. Matteson. 2015. *Statistics and Data Analysis for Financial Engineering*. Springer.
- Saroka, Ivan, Riccardo Rebonato. 2014. How to deform a yield curve: The most likely deformation of a yield curve consistent with subjective views. *The Journal of Fixed Income* **24**(3) 30–38.
- Steinwart, Ingo, Andreas Christmann. 2011. Estimating conditional quantiles with the help of the pinball loss .

- Takeuchi, Ichiro, Quoc V Le, Timothy D Sears, Alexander J Smola. 2006. Nonparametric quantile estimation. *Journal of Machine Learning Research* **7**(45) 1231–1264.
- Tsay, R. S. 2010. *Analysis of Financial Time Series*. John Wiley & Sons.
- Tu, Daohong, Kay Giesecke. 2026. Online conformal prediction for non-exchangeable panel data. doi:10.48550/arXiv.2605.17705. URL <https://arxiv.org/abs/2605.17705>. Working paper, arXiv:2605.17705.
- Vovk, Vladimir, Alexander Gammerman, Glenn Shafer. 2005. *Algorithmic learning in a random world*, vol. 29. Springer.
- Wang, Qiuqi, Ruodu Wang, Johanna Ziegel. 2026. E-backtesting. *Management Science* **72**(6) 4952–4973. doi:10.1287/mnsc.2023.01659.
- Wang, Ruodu, Johanna F. Ziegel. 2021. Scenario-based risk evaluation. *Finance and Stochastics* **25**(4) 725–756. doi:10.1007/s00780-021-00460-9.
- Zaffran, Margaux, Olivier Féron, Yannig Goude, Julie Josse, Aymeric Dieuleveut. 2022. Adaptive conformal predictions for time series. *International Conference on Machine Learning*. PMLR, 25834–25866.

E-Companion

Stress Testing Financial Portfolios with Coverage Guarantees

EC.1 Additional Supporting Material

EC.1.1 Notation Summary

Table EC.1 summarizes the main symbols used throughout the paper.

Table EC.1: Notation summary.

Symbol	Meaning
General market and scenario notation	
t	Time index.
V_t	Portfolio value at time t .
G_{t+1}	One-step portfolio gain / P&L.
$\mathcal{F}_t$	Information available at time t .
$\mathbf{x}_t$	Vector of risk factors.
$\mathbf{f}_t$	Vector of common factors.
$\Delta \mathbf{x}_{t+1}, \Delta \mathbf{f}_{t+1}$	One-step changes in the risk factors and common factors.
$\mathbf{B}$	Factor loading matrix in the factor model.
$\mathbf{z}$	Stress scenario applied to selected common factor returns.
$\mathbf{z}_t$	Realized stress scenario on date t .
$v(\cdot)$	Pricing map from risk factors to portfolio value.
ACSA notation	
α	Target miscoverage level.
$\tilde{\alpha}_t$	Time- t adjusted miscoverage level used by ACSA.
$\bar{\alpha}$	Average adjusted miscoverage level used in the ACSA simulations.
γ	ACSA step size.
$g_t(\mathbf{z}; \tilde{\alpha})$	Scenario-wise quantile predictor evaluated at quantile level $\tilde{\alpha}$.
$\mathcal{C}_t(\mathbf{z})$	Prediction interval reported for scenario $\mathbf{z}$.
err_t	Indicator that the previously reported interval misses the realized gain.
$B_1, \dots, B_m$	Scenario groups in group-balanced ACSA.
$\tilde{\alpha}_{t,k}$	Group-specific adjusted miscoverage level.

Continued on next page

Table EC.1 (continued)

Symbol	Meaning
KSA notation	
u	Quantile level arguments used in pinball-loss and quantile expressions.
$\ell_u(y, q)$	Pinball loss at quantile level u .
$Q_Y(u)$	u -quantile of a random variable Y .
$\mathbf{H}_t$	Historical context vector.
$\mathbf{W}_t(\mathbf{z})$	Feature vector $(\mathbf{z}, \mathbf{H}_{t-1})$.
$\varphi(\mathbf{w})$	Expectation predictor of the scenario gain.
Y_t	Residual of the portfolio gain after subtracting the predicted expectation.
$\kappa(\mathbf{w}, \mathbf{w}')$	Kernel similarity function.
$\hat{\ell}_t(u; q, \mathbf{w})$	Kernel-reweighted empirical pinball loss.
$\hat{Q}_t(u; \mathbf{W}_{t+1}(\mathbf{z}))$	KSA estimator of the residual u -quantile.
$\mathcal{S}$	Set of scenarios in the scenario table.
h	Kernel bandwidth.
Running example and experiments	
β	Factor loadings in the three-factor log-linear example of Section 4.
Σ	Annualized covariance matrix of the common factor returns.
$\sigma_{\text{Cred} s}^2$	Annualized conditional variance of the unstressed credit factor given the stressed factors.
m_z	One-day conditional mean of the log return under scenario z .
v_{Cred}	One-day conditional variance of the log return under scenario z .
$\sigma_\epsilon, \hat{\sigma}_\epsilon$	True and misspecified annualized idiosyncratic volatilities in the running example.
σ_{tot}	Total annualized volatility in the running example.
δ	One-trading-day time increment in years, set to $1/252$.
u_j	Number of units held in security j .
$u^{\text{skew}}, u^{\text{hedge}}$	Common option position size and underlying-index hedge position in the skew-steepener portfolio of Section 7.2.
m_j, τ_j	Option moneyness and time to maturity of security j .

Continued on next page

Table EC.1 (continued)

Symbol	Meaning
Dependence and theory	
$\beta(k)$	β -mixing coefficient of the feature-residual process $\{(\mathbf{W}_t, Y_t)\}$.

EC.1.2 A More Complex Example of ACSA Dynamics

Following on from Examples 1 and 2, we now consider a more complex ACSA example in which G_t follows an autoregressive process. The example compares conditional and stationary quantile predictors at a common adjusted level. As in Examples 1 and 2, we do not condition on any scenario argument in this example.

Example EC.1. We assume the dynamics of G_t satisfy $G_{t+1} = \mu + \theta G_t + \epsilon_{t+1}$, where $\epsilon_{t+1} \sim \mathcal{N}(0, \sigma^2)$ is an exogenous noise random variable. Then $G_{t+1} \mid G_t \sim \mathcal{N}(\mu + \theta G_t, \sigma^2)$ and the true $\tilde{\alpha}$ -quantile of $G_{t+1} \mid G_t$ (denoted by $Q_{G_{t+1}}(\tilde{\alpha} \mid G_t)$) is $Q_{G_{t+1}}(\tilde{\alpha} \mid G_t) = \mu + \theta G_t + \sigma \Phi^{-1}(\tilde{\alpha})$, where $\Phi^{-1}(\tilde{\alpha})$ is the $\tilde{\alpha}$ -quantile of the standard normal distribution. If we take $g_t(\tilde{\alpha}) \equiv Q_{G_{t+1}}(\tilde{\alpha} \mid G_t)$, then the prediction interval produced in (15) is $\mathcal{C}_t \equiv [\mu + \theta G_t + \sigma \Phi^{-1}(\tilde{\alpha}_t/2), \mu + \theta G_t + \sigma \Phi^{-1}(1 - \tilde{\alpha}_t/2)]$. This prediction interval depends only on G_t and has width $\text{WI}(\mathcal{C}_t)$ given by

$$\text{WI}(\mathcal{C}_t) = \sigma \cdot (\Phi^{-1}(1 - \tilde{\alpha}_t/2) - \Phi^{-1}(\tilde{\alpha}_t/2)). \quad (\text{EC.1})$$

As g_t is correctly calibrated, we know from Example 1 how the $\tilde{\alpha}_t$'s will behave for $0 < \gamma < 2$ and so this tells us (loosely speaking) how $\mathcal{C}_t$ and $\text{WI}(\mathcal{C}_t)$ behave.

Now consider an alternative quantile predictor. Assuming $-1 < \theta < 1$, then it is easy to see that the stochastic process $\{G_t\}_{t \geq 0}$ has an $\mathcal{N}(\mu/(1 - \theta), \sigma^2/(1 - \theta^2))$ stationary distribution and its $\tilde{\alpha}$ -quantile is $Q_{G_\infty}(\tilde{\alpha}) = \mu/(1 - \theta) + \sigma/\sqrt{1 - \theta^2} \Phi^{-1}(\tilde{\alpha})$. Suppose we take $g_t(\tilde{\alpha}) \equiv Q_{G_\infty}(\tilde{\alpha})$. In this case the prediction interval produced in (15) is $\mathcal{C}_t \equiv \left[\frac{\mu}{1 - \theta} + \frac{\sigma}{\sqrt{1 - \theta^2}} \Phi^{-1}(\tilde{\alpha}_t/2), \frac{\mu}{1 - \theta} + \frac{\sigma}{\sqrt{1 - \theta^2}} \Phi^{-1}(1 - \tilde{\alpha}_t/2) \right]$. This prediction interval does not depend on G_t and has width

$$\text{WI}(\mathcal{C}_t) = \frac{\sigma}{\sqrt{1 - \theta^2}} \cdot (\Phi^{-1}(1 - \tilde{\alpha}_t/2) - \Phi^{-1}(\tilde{\alpha}_t/2)). \quad (\text{EC.2})$$

At a common adjusted level $\tilde{\alpha} \in (0, 1)$, the stationary-quantile interval has width equal to the conditional-quantile interval width multiplied by $1/\sqrt{1 - \theta^2}$. Thus, whenever $\theta \neq 0$, conditioning on G_t produces a strictly sharper interval at the same quantile level.

The two ACSA procedures generally generate different adaptive level sequences, however, and the stationary quantiles need not be calibrated conditionally on G_t . Therefore, the common-level comparison above does not imply an ordering of the realized interval widths along the two adaptive trajectories. Rather, it isolates the potential gain in sharpness from using a predictor that conditions on the information relevant to the next-period distribution.

EC.1.3 Additional Details for Section 5.2

This appendix (i) derives closed form expressions for the intervals $\mathcal{C}_{\text{True}}(\mathbf{z})$, $\mathcal{C}_{\text{Mis}}(\mathbf{z}, \hat{\sigma}_\epsilon)$ and $\mathcal{C}_{\text{ACSA}}(\mathbf{z}, \hat{\sigma}_\epsilon, \tilde{\alpha}_t)$ of Section 5.2 and (ii) provides further explanations for the behavior we observed in Table 4 and Figure 1. We first recall that all model parameters are known except for σ_ϵ , for which only an estimator $\hat{\sigma}_\epsilon$ is available. As a result, $m_{\mathbf{z}}$, v_{Cred} , and σ_{tot} and their estimated counterparts are

$$\begin{aligned}\hat{m}_{\mathbf{z}} &:= \delta \left(\mu_\delta - \frac{\hat{\sigma}_\epsilon^2}{2} \right) + \beta_{\text{Oil}} z_{\text{Oil}} + \beta_{\text{Rate}} z_{\text{Rate}} + \beta_{\text{Cred}} \mu_{\text{Cred}|s} \\ \hat{v}_{\text{Cred}} &:= \delta \cdot \left(\beta_{\text{Cred}}^2 \sigma_{\text{Cred}|s}^2 + \hat{\sigma}_\epsilon^2 \right), \quad \hat{\sigma}_{\text{tot}} := \sqrt{\beta^\top \Sigma \beta + \hat{\sigma}_\epsilon^2}.\end{aligned}\tag{EC.3}$$

Explicit Forms of Prediction Intervals In order to provide the explicit forms of the prediction intervals $\mathcal{C}_{\text{True}}(\mathbf{z})$, $\mathcal{C}_{\text{Mis}}(\mathbf{z}, \hat{\sigma}_\epsilon)$ and $\mathcal{C}_{\text{ACSA}}(\mathbf{z}, \hat{\sigma}_\epsilon, \tilde{\alpha}_t)$, recall that V_0 and V_1 are the $t = 0$ and $t = 1$ portfolio values, respectively, so that the portfolio gain at $t = 1$ is $G_1 := V_1 - V_0$. Then recalling that portfolio consists of a single short call position, we have $V_0 = -C(S_0, K, \tau, \sigma_{\text{tot}})$ and $V_1 = -C(S_1, K, \tau - \delta, \sigma_{\text{tot}})$ where S_0 and S_1 denote the $t = 0$ and $t = 1$ prices, respectively, of the S&P 500, δ represents 1 day (in years), and we use $C(S, K, \tau, \sigma)$ to denote¹⁵ the Black-Scholes price of a call option with underlying price S , time to maturity τ and strike K . We write G_1 as $G_1(S_1)$ since the randomness of G_1 arises entirely from the randomness of S_1 . The explicit dependency of G_1 on S_1 is given by:

$$G_1(S_1) = -V_0 - C(S_1, K, \tau - \delta, \sigma_{\text{tot}}).\tag{EC.4}$$

Based on (10), for any given scenario $\mathbf{z}$, the conditional distribution $G_1 \mid \mathbf{z}$ is determined by (EC.4) together with $S_1 = S_0 \cdot \exp(m_{\mathbf{z}} + \sqrt{v_{\text{Cred}}} \cdot \mathcal{Z})$, where $\mathcal{Z}$ is a standard normal random variable. To derive the expression for $\mathcal{C}_{\text{True}}(\mathbf{z})$, recall that its endpoints correspond to the $\alpha/2$ and $1 - \alpha/2$ conditional quantiles of G_1 . Since $G_1(S_1)$ is strictly decreasing in S_1 , for any fixed $\mathbf{z}$ the conditional quantiles of G_1 are obtained by evaluating $G_1(\cdot)$ at the corresponding conditional quantiles of S_1 . That is, $Q_{G_1}(\tilde{\alpha} \mid \mathbf{z}) = G_1(Q_{S_1}(1 - \tilde{\alpha} \mid \mathbf{z})) = G_1(S_0 \cdot \exp(m_{\mathbf{z}} + \sqrt{v_{\text{Cred}}} \cdot \Phi^{-1}(1 - \tilde{\alpha})))$ where we use $Q_Y(\tilde{\alpha} \mid \mathbf{z})$ to denote the $\tilde{\alpha}$ quantile of the random variable Y conditional on $\mathbf{z}$. The explicit form of $\mathcal{C}_{\text{True}}(\mathbf{z})$ is therefore

$$\mathcal{C}_{\text{True}}(\mathbf{z}) = [G_1(S_0 \cdot \exp(m_{\mathbf{z}} + \sqrt{v_{\text{Cred}}} \cdot \Phi^{-1}(1 - \alpha/2))), G_1(S_0 \cdot \exp(m_{\mathbf{z}} + \sqrt{v_{\text{Cred}}} \cdot \Phi^{-1}(\alpha/2)))] .\tag{EC.5}$$

The expressions of $\mathcal{C}_{\text{Mis}}(\mathbf{z}, \hat{\sigma}_\epsilon)$ and $\mathcal{C}_{\text{ACSA}}(\mathbf{z}, \hat{\sigma}_\epsilon, \tilde{\alpha}_t)$ can be derived similarly. In particular, let $\hat{G}_1(S_1) = -V_0 - C(S_1, K, \tau - \delta, \hat{\sigma}_{\text{tot}})$ denote the *estimated* portfolio gain as a function of S_1 but with $\hat{\sigma}_\epsilon$ used in place of σ_ϵ , and $\hat{m}_{\mathbf{z}}$, $\hat{v}_{\text{Cred}}$ and $\hat{\sigma}_{\text{tot}}$ determined by $\hat{\sigma}_\epsilon$ (as well as $\mathbf{z}$ in the case of $\hat{m}_{\mathbf{z}}$; see (EC.3). For $0 < \tilde{\alpha}_t < 1$, the explicit forms of $\mathcal{C}_{\text{Mis}}(\mathbf{z}, \hat{\sigma}_\epsilon)$ and $\mathcal{C}_{\text{ACSA}}(\mathbf{z}, \hat{\sigma}_\epsilon, \tilde{\alpha}_t)$ are then given by

$$\begin{aligned}\mathcal{C}_{\text{Mis}}(\mathbf{z}, \hat{\sigma}_\epsilon) &= \left[\hat{G}_1(S_0 \cdot \exp(\hat{m}_{\mathbf{z}} + \sqrt{\hat{v}_{\text{Cred}}} \cdot \Phi^{-1}(1 - \alpha/2))), \hat{G}_1(S_0 \cdot \exp(\hat{m}_{\mathbf{z}} + \sqrt{\hat{v}_{\text{Cred}}} \cdot \Phi^{-1}(\alpha/2))) \right] \\ \mathcal{C}_{\text{ACSA}}(\mathbf{z}, \hat{\sigma}_\epsilon, \tilde{\alpha}_t) &= \left[\hat{G}_1(S_0 \cdot \exp(\hat{m}_{\mathbf{z}} + \sqrt{\hat{v}_{\text{Cred}}} \cdot \Phi^{-1}(1 - \tilde{\alpha}_t/2))), \hat{G}_1(S_0 \cdot \exp(\hat{m}_{\mathbf{z}} + \sqrt{\hat{v}_{\text{Cred}}} \cdot \Phi^{-1}(\tilde{\alpha}_t/2))) \right]\end{aligned}$$

¹⁵We omit the dependence of the Black-Scholes price on the interest rate and dividend yield.

where $\tilde{\alpha}_t$ is the time t quantile argument of the ACSA algorithm. We remark that when $\tilde{\alpha}_t = \alpha$, we have exactly $\mathcal{C}_{\text{ACSA}}(\mathbf{z}, \hat{\sigma}_\epsilon, \tilde{\alpha}_t) = \mathcal{C}_{\text{Mis}}(\mathbf{z}, \hat{\sigma}_\epsilon)$.

True Coverage Rates of Various Prediction Intervals For any stress scenario $\mathbf{z}$, we derive an explicit expression for the true coverage rate of the true scenario gain G_1 associated with a given prediction interval $[A, B]$. As is clear from (EC.4), the distribution of G_1 is determined by the normal distribution of $\log S_1/S_0$. Consequently, by transforming the prediction interval for G_1 into an “equivalent” prediction interval for $\log S_1/S_0$, we can obtain a closed-form expression for its true coverage rate. In particular, for any interval $[A, B]$, we have

$$\begin{aligned} \mathbb{P}(G_1 \in [A, B]) &= \mathbb{P}\left(\log S_1/S_0 \in [\log G_1^{-1}(B) - \log S_0, \log G_1^{-1}(A) - \log S_0]\right) \\ &= \mathbb{P}\left(\mathcal{Z} \in \left[\frac{1}{\sqrt{v_{\text{Cred}}}} (\log G_1^{-1}(B) - \log S_0 - m_{\mathbf{z}}), \frac{1}{\sqrt{v_{\text{Cred}}}} (\log G_1^{-1}(A) - \log S_0 - m_{\mathbf{z}})\right]\right) \\ &= \Phi\left(\frac{1}{\sqrt{v_{\text{Cred}}}} (\log G_1^{-1}(A) - \log S_0 - m_{\mathbf{z}})\right) - \Phi\left(\frac{1}{\sqrt{v_{\text{Cred}}}} (\log G_1^{-1}(B) - \log S_0 - m_{\mathbf{z}})\right) \end{aligned} \quad (\text{EC.6})$$

where $\Phi(\cdot)$ denotes the standard normal CDF. For $0 < \tilde{\alpha} < 1$, it follows from (EC.6) and our explicit expression for $\mathcal{C}_{\text{ACSA}}(\mathbf{z}, \hat{\sigma}_\epsilon, \tilde{\alpha})$ above that

$$\begin{aligned} \psi(m_{\mathbf{z}}, \hat{\sigma}_\epsilon, \tilde{\alpha}) &:= \mathbb{P}(G_1 \in \mathcal{C}_{\text{ACSA}}(\mathbf{z}, \hat{\sigma}_\epsilon, \tilde{\alpha})) \\ &= \Phi\left(\frac{1}{\sqrt{v_{\text{Cred}}}} \left(\log G_1^{-1}(\hat{G}_1(S_0 \cdot \exp(\hat{m}_{\mathbf{z}} + \sqrt{\hat{v}_{\text{Cred}}} \cdot \Phi^{-1}(1 - \tilde{\alpha}/2)))) - \log S_0 - m_{\mathbf{z}}\right)\right) \\ &\quad - \Phi\left(\frac{1}{\sqrt{v_{\text{Cred}}}} \left(\log G_1^{-1}(\hat{G}_1(S_0 \cdot \exp(\hat{m}_{\mathbf{z}} + \sqrt{\hat{v}_{\text{Cred}}} \cdot \Phi^{-1}(\tilde{\alpha}/2)))) - \log S_0 - m_{\mathbf{z}}\right)\right) \end{aligned} \quad (\text{EC.7})$$

Outside the interior, define $\psi(m_{\mathbf{z}}, \hat{\sigma}_\epsilon, \tilde{\alpha}) = 1$ for $\tilde{\alpha} \leq 0$ and 0 for $\tilde{\alpha} \geq 1$. The expression above is the probability of the scenario gain G_1 being covered by the ACSA interval $\mathcal{C}_{\text{ACSA}}(\mathbf{z}, \hat{\sigma}_\epsilon, \tilde{\alpha})$ as a function of $m_{\mathbf{z}}$, $\hat{\sigma}_\epsilon$ and $\tilde{\alpha}$. Although the function ψ provides the true coverage rate for any stress scenario $\mathbf{z}$, it depends on $\mathbf{z}$ only through $m_{\mathbf{z}}$. In particular, $\hat{m}_{\mathbf{z}}$ is determined by $m_{\mathbf{z}}$ through $\hat{m}_{\mathbf{z}} = m_{\mathbf{z}} + \frac{\delta}{2}(\sigma_\epsilon^2 - \hat{\sigma}_\epsilon^2)$, while v_{Cred} and $\hat{v}_{\text{Cred}}$ do not depend on the explicit value of $\mathbf{z}$. We also note that (EC.7) may be used to compute true coverage rate of $\mathcal{C}_{\text{Mis}}$ since $\mathbb{P}(G_1 \in \mathcal{C}_{\text{Mis}}(\mathbf{z}, \hat{\sigma}_\epsilon)) = \psi(m_{\mathbf{z}}, \hat{\sigma}_\epsilon, \alpha)$. We can now use $\psi(m_{\mathbf{z}}, \hat{\sigma}_\epsilon, \tilde{\alpha})$ to explain the monotonic pattern of coverage rates that we observed in Table 4 as well as the non-monotonic (in $\hat{\sigma}_\epsilon$) behavior of $\tilde{\alpha}_t$ we observed in Figure 1.

Monotonic Pattern of Table 4 In Table 4, we observed that for both $\mathcal{C}_{\text{Mis}}(\mathbf{z}, \hat{\sigma}_\epsilon)$ and $\mathcal{C}_{\text{ACSA}}(\mathbf{z}, \hat{\sigma}_\epsilon, \tilde{\alpha})$, the true coverage rates decreased from left to right within each row and increased from top to bottom within each column. Since both $\hat{\sigma}_\epsilon$ and $\tilde{\alpha}$ are constant across scenarios $\mathbf{z}$, $m_{\mathbf{z}}$ is the only argument of ψ that drives differences in the real coverage rate across scenarios. As demonstrated in Figure EC.1, $\psi(m_{\mathbf{z}}, \hat{\sigma}_\epsilon, \tilde{\alpha})$ consistently increases with $m_{\mathbf{z}}$ under all combinations of $\hat{\sigma}_\epsilon$ and $\tilde{\alpha}$.

considered in the experiments of Section 5.2. As we recall from (17), however, $m_{\mathbf{z}}$ decreases with z_{Oil} and increases with z_{Rate} and together these observations therefore explain the monotonicity patterns of Table 4.

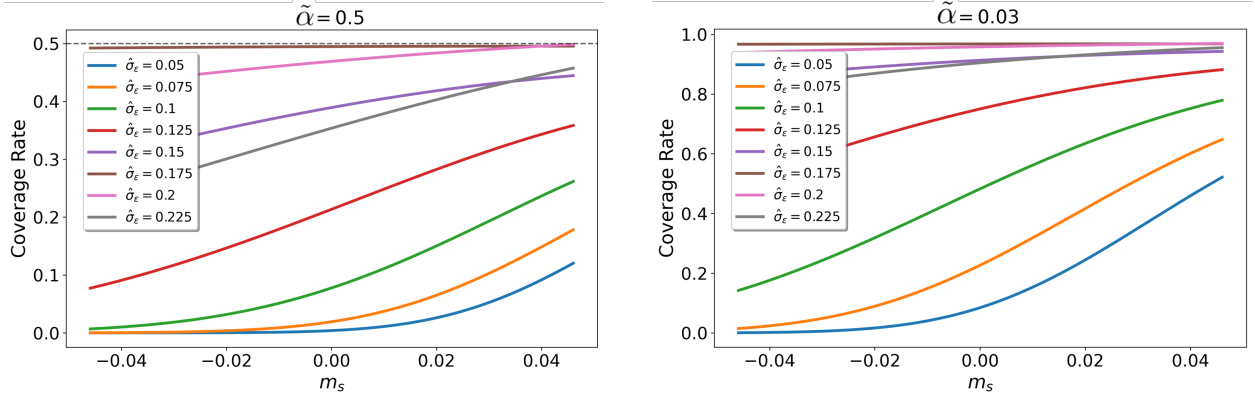


Figure EC.1: The variation of $\psi(m_{\mathbf{z}}, \hat{\sigma}_\epsilon, \tilde{\alpha})$ with respect to $m_{\mathbf{z}}$. The left figure sets $\tilde{\alpha} = 0.5$ and the right figure sets $\tilde{\alpha} = 0.03$. In each figure, different values of $\hat{\sigma}_\epsilon$ are considered, ranging from 0.05 to 0.225. In each of these curves, $\psi(m_{\mathbf{z}}, \hat{\sigma}_\epsilon, \tilde{\alpha})$ monotonically increases with $m_{\mathbf{z}}$.

Non-monotonicity Pattern of Figure 1 In order to understand the non-monotonicity pattern of Figure 1, we need to consider $\psi(m_{\mathbf{z}}, \hat{\sigma}_\epsilon, \tilde{\alpha})$ as a function of $\hat{\sigma}_\epsilon$. As demonstrated in Figure EC.2, the function $\psi(m_{\mathbf{z}}, \hat{\sigma}_\epsilon, \tilde{\alpha})$ is non-monotonic in $\hat{\sigma}_\epsilon$ for all combinations of $m_{\mathbf{z}}$ and $\tilde{\alpha}$ that we consider. Indeed, it first increases and then decreases as $\hat{\sigma}_\epsilon$ moves from 0 to 0.25¹⁶. But $\psi(m_{\mathbf{z}}, \hat{\sigma}_\epsilon, \tilde{\alpha})$ decreases in $\tilde{\alpha}$ since $\mathcal{C}_{\text{ACSA}}(\mathbf{z}, \hat{\sigma}_\epsilon, \tilde{\alpha})$ is the region between the estimated $\tilde{\alpha}/2$ - and $(1 - \tilde{\alpha}/2)$ - quantiles. Hence, increasing $\tilde{\alpha}$ narrows the prediction interval and thus decreases $\psi(m_{\mathbf{z}}, \hat{\sigma}_\epsilon, \tilde{\alpha})$. Finally, recall that the ACSA algorithm selects $\tilde{\alpha}_t$ to target a 50% empirical coverage rate. Consequently, when $\hat{\sigma}_\epsilon$ becomes either too small or too large relative to the true value $\sigma_\epsilon = .18$, $\psi(m_{\mathbf{z}}, \hat{\sigma}_\epsilon, \alpha)$ decreases and achieves a coverage rate that is less than the target rate of 50% in our experiments. Thus, a smaller $\tilde{\alpha}_t$ is required to bring $\psi(m_{\mathbf{z}}, \hat{\sigma}_\epsilon, \tilde{\alpha})$ back towards 50%. This explains the non-monotonic behavior in Figure 1 where $\tilde{\alpha}$ (the long-term average of $\tilde{\alpha}_t$) decreases as $\hat{\sigma}_\epsilon$ moves away from its true value σ_ϵ .

EC.1.4 The Group-Balanced ACSA Algorithm

As discussed at the end of Section 5.1, the ACSA algorithm provides only a marginal coverage guarantee. If we analyze the prediction intervals under each *fixed scenario*, then the coverage rate is often far from the marginal coverage rate. This limitation can be observed in Table 4, where the true coverage rates can deviate substantially from the nominal level of 50% in more

¹⁶Specifically, ψ increases for smaller values of $\hat{\sigma}_\epsilon$ (e.g., $\hat{\sigma}_\epsilon \in (0, 0.15]$) and decreases for larger values (e.g., $\hat{\sigma}_\epsilon \in [0.20, 0.25]$). At the correctly specified value $\hat{\sigma}_\epsilon = \sigma_\epsilon = 0.18$, the coverage equals $1 - \tilde{\alpha}$: this is 50% in the left panel ($\tilde{\alpha} = 0.5$) and 97% in the right panel ($\tilde{\alpha} = 0.03$).

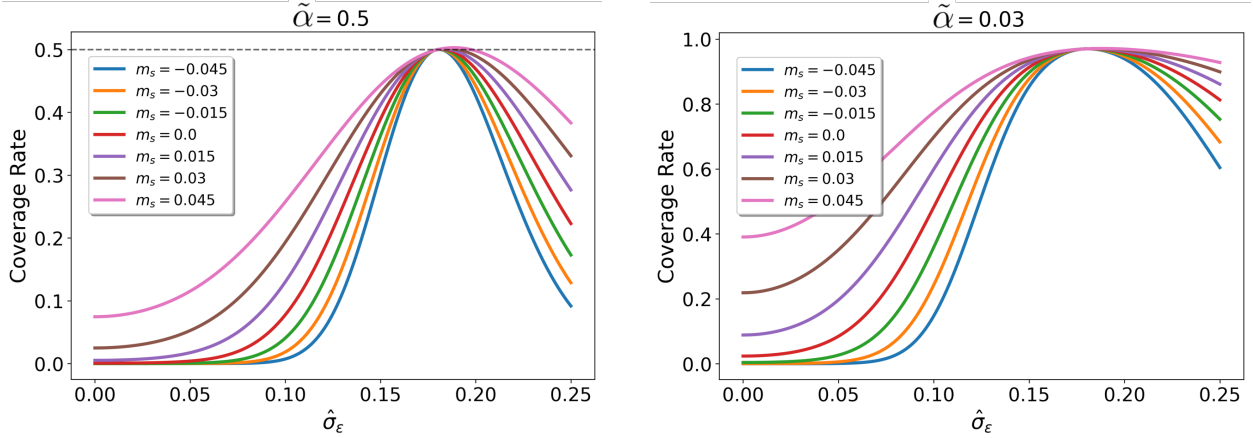


Figure EC.2: The variation of $\psi(m_{\mathbf{z}}, \hat{\sigma}_\epsilon, \tilde{\alpha})$ with respect to $\hat{\sigma}_\epsilon$. The left panel sets $\tilde{\alpha} = 0.5$ and the right panel sets $\tilde{\alpha} = 0.03$. In each sub-figure, different values of $m_{\mathbf{z}}$ are considered, ranging from -0.045 to 0.045 . Across all curves, $\psi(m_{\mathbf{z}}, \hat{\sigma}_\epsilon, \tilde{\alpha})$ exhibits a consistent pattern with respect to $\hat{\sigma}_\epsilon$: it first increases and then decreases as $\hat{\sigma}_\epsilon$ passes through some value near the true value $\sigma_\epsilon = 0.18$.

extreme scenarios. For example, when $(z_{\text{Oil}}, z_{\text{Rate}}) = (12\%, -20 \text{ bp})$ and $(-12\%, 20 \text{ bp})$, the true coverage rates are 25% and 69%, respectively. One of the primary causes of this discrepancy is that ACSA updates its quantile argument $\tilde{\alpha}_t$ iteratively based on *realized* observations. As a result, rare or extreme scenarios are underrepresented¹⁷ in the updates due to their disproportionately low frequency in the observed data. One approach for tackling this problem is to incorporate a group-balancing¹⁸ adjustment into the vanilla ACSA procedure.

The idea behind group-balancing is to partition the space of stress scenarios into several groups, which we denote by $B_1, \dots, B_m$, and to run ACSA separately within each group. Each group maintains its own quantile parameter, denoted by $\tilde{\alpha}_{t,1}, \dots, \tilde{\alpha}_{t,m}$. For a concrete example, suppose we want to balance between the extreme and non-extreme scenarios in Table 4. Then we may partition the stress scenarios into two groups: the “extreme scenario group” B_1 and the “non-extreme scenario group” B_2 , with

$$B_1 := \{\mathbf{z} : |z_{\text{Oil}}| > 6\% \text{ or } |z_{\text{Rate}}| > 10 \text{ bp}\}, \quad B_2 := \{\mathbf{z} : |z_{\text{Oil}}| \leq 6\% \text{ and } |z_{\text{Rate}}| \leq 10 \text{ bp}\}.$$

Separate quantiles arguments $\tilde{\alpha}_{t,1}$ and $\tilde{\alpha}_{t,2}$ are maintained for these two groups, since the quantile predictor g_t may exhibit different estimation accuracies across them and thus require different adjustments to the quantile argument $\tilde{\alpha}_t$. At each time step, when the realized scenario $\mathbf{z}_t$ is observed, the group-balanced ACSA first identifies its corresponding group (B_1 or B_2) and then updates only the quantile parameter of that group. For instance, if $\mathbf{z}_t \in B_k$ for some $k \in \{1, \dots, m\}$, then only $\tilde{\alpha}_{t,k}$ is updated according to (14), and the quantile arguments for the remaining groups

¹⁷Loosely speaking, the marginal coverage rate (which we control via the ACSA algorithm) can be viewed as a weighted average of the coverage rates within each scenario with weights given by the probability of each scenario.

¹⁸See Section 4.1 of [Angelopoulos and Bates \(2021\)](#) for detailed discussions on the group-balanced conformal prediction technique.

are left unchanged, i.e.

$$\begin{aligned}\tilde{\alpha}_{t,k} &\leftarrow \tilde{\alpha}_{t-1,k} + \gamma \cdot (\alpha - \text{err}_t), \quad \text{with } \text{err}_t \leftarrow \mathbb{1}\{G_t \notin \mathcal{C}_{t-1}(\mathbf{z}_t; \tilde{\alpha}_{t-1,k})\} \\ \tilde{\alpha}_{t,k'} &\leftarrow \tilde{\alpha}_{t-1,k'}, \quad \text{for } k' = 1, \dots, m \text{ with } k' \neq k\end{aligned}$$

The group-balanced variant of our ACSA algorithm is summarized in Algorithm EC.1, where each group effectively runs Algorithm 1 independently. The corresponding coverage guarantee for Group-Balanced ACSA is stated in Theorem EC.1.1, which follows almost directly from the original coverage result in Theorem 5.1.

Theorem EC.1.1. *Under Assumption 1, with $\alpha \in (0, 1)$, $\gamma > 0$, and the boundary construction (12) applied within each group, for each group B_k , the empirical coverage of the prediction interval $\mathcal{C}_t(\cdot)$ generated by group-balanced ACSA (Algorithm EC.1) satisfies the coverage guarantee*

$$\left| \frac{1}{T_k} \sum_{t=1}^T \mathbb{1}\{G_{t+1} \in \mathcal{C}_t(\mathbf{z}_{t+1})\} \cdot \mathbb{1}\{\mathbf{z}_{t+1} \in B_k\} - (1 - \alpha) \right| \leq \frac{\max\{\alpha, 1 - \alpha\} + \gamma}{T_k \gamma}$$

almost surely, with $T_k := \sum_{t=1}^T \mathbb{1}\{\mathbf{z}_{t+1} \in B_k\}$ and is assumed > 0 .

Algorithm EC.1 Group-Balanced ACSA

Input: Partition of stress scenarios into groups $B_1, \dots, B_m$, plus the inputs of Algorithm 1

- 1: Set $\tilde{\alpha}_{1,k} = \alpha$ and report $\mathcal{C}_1(\mathbf{z}) = \mathcal{C}_1(\mathbf{z}; \alpha)$ for $\mathbf{z} \in B_k$, $k = 1, \dots, m$
- 2: **for** $t = 2, \dots, T$ **do**
- 3: Observe the realized portfolio gain G_t and stress scenario $\mathbf{z}_t$
- 4: Identify the group to which $\mathbf{z}_t$ belongs, for instance $\mathbf{z}_t \in B_k$, then compute

$$\text{err}_t \leftarrow \mathbb{1}\{G_t \notin \mathcal{C}_{t-1}(\mathbf{z}_t; \tilde{\alpha}_{t-1,k})\}$$

- 5: Keep $\tilde{\alpha}_{t,k'} \leftarrow \tilde{\alpha}_{t-1,k'}$ for all $k' = 1, \dots, m$ with $k' \neq k$, and update

$$\tilde{\alpha}_{t,k} \leftarrow \tilde{\alpha}_{t-1,k} + \gamma \cdot (\alpha - \text{err}_t)$$

- 6: Update / retrain quantile predictor g_t if so desired
- 7: **Output:** For all $k' = 1, \dots, m$, and any stress scenario $\mathbf{z} \in B_{k'}$, produce prediction interval

$$\mathcal{C}_t(\mathbf{z}) \leftarrow \mathcal{C}_t(\mathbf{z}; \tilde{\alpha}_{t,k'})$$

- 8: **end for**
-

The group-balanced ACSA helps mitigate the lack of theoretical guarantees in extreme scenarios. However, the method remains limited by data scarcity in the more extreme scenario groups. For such scenario groups, Theorem EC.1.1 will provide only limited comfort. Hence, there is a resolution vs sample size tradeoff in how finely you partition scenarios. That said, the group-balanced ACSA can provide theoretical guarantees in moderate scenario groups where there are sufficiently many realizations for calibrating the corresponding $\tilde{\alpha}_t$ values. Moreover, as may be seen from Table 4,

even the non-group-balanced ACSA algorithm can yield a substantial improvement over a scenario analysis that relies on an inaccurate model.

EC.1.5 Additional Details for Section 6.3

We provide additional implementation details for the KSA experiments in Section 6.3.

Feature Vector and Assumption 2 One of the key preliminary steps before implementing the KSA algorithm is to specify the historical context $\mathbf{H}_t$ and the feature vector $\mathbf{W}_t(\mathbf{z})$. Intuitively, $\mathbf{H}_t$ should contain all historical information relevant to the portfolio gain on day $t + 1$. In our setting, the portfolio is short a single call option with fixed time to maturity τ and moneyness (strike price divided by the underlying price) m . We show below that *no historical context is required* once the portfolio gain is normalized by the current underlying price. Thus, in the original factor coordinates, the information content of the feature vector can be represented by $\mathbf{W}_t(\mathbf{z}) = \mathbf{z}$. The portfolio gain at time $t + 1$ is given by

$$G_{t+1} = -\text{BS}(S_{t+1}, m \cdot S_t, \tau - \delta, \sigma_{\text{tot}}, 1) + \text{BS}(S_t, m \cdot S_t, \tau, \sigma_{\text{tot}}, 1), \quad (\text{EC.8})$$

where $\text{BS}(S_t, K, \tau, \sigma, c)$ denotes the Black–Scholes price of an option, with S_t , K , τ , and σ denoting the underlying price, strike, time to maturity, and implied volatility, respectively. $c \in \{0, 1\}$ indicating the option type, with $c = 1$ for a call and $c = 0$ for a put. The risk-free rate and dividend yield are treated as fixed and are therefore omitted from the arguments of BS. Since the portfolio maintains a fixed moneyness m , its strike price is $m \cdot S_t$, while the remaining time to maturity decreases from τ to $\tau - \delta$ between times t and $t + 1$. After normalizing by S_t , the portfolio gain depends on the past only through the return S_{t+1}/S_t . Specifically,

$$G'_{t+1} := \frac{G_{t+1}}{S_t} = -\frac{S_{t+1}}{S_t} \widetilde{\text{BS}}\left(m \cdot \frac{S_t}{S_{t+1}}, \tau - \delta, \sigma_{\text{tot}}, 1\right) + \widetilde{\text{BS}}(m, \tau, \sigma_{\text{tot}}, 1), \quad (\text{EC.9})$$

where $\widetilde{\text{BS}}(m, \tau, \sigma, c)$ denotes the Black–Scholes price normalized by the underlying price:

$$\widetilde{\text{BS}}(m, \tau, \sigma, c) := \frac{1}{S_t} \text{BS}(S_t, K, \tau, \sigma, c), \quad m = \frac{K}{S_t}. \quad (\text{EC.10})$$

The only source of randomness in (EC.9) is therefore the return S_{t+1}/S_t . Any prediction interval for G'_{t+1} can be mapped to one for G_{t+1} by multiplying both endpoints by the positive, $\mathcal{F}_t$ -measurable quantity S_t , which preserves its conditional coverage probability. Hence, when applying the theory in Section 6.2, we regard G'_t as the response and set $Y_t = G'_t - \varphi(\mathbf{W}_t)$ where φ is the fixed centering function specified below.

We now verify parts (a)–(d) of Assumption 2; the kernel condition in part (e) is verified in the next paragraph. Under (6), the factor vectors $\mathbf{F}_t$ and idiosyncratic noises are independent

across time. Since $\mathbf{W}_t = \mathbf{F}_t^{(1:2)}$ and G'_t is a fixed function of the contemporaneous factor vector and idiosyncratic noise, $\{(\mathbf{W}_t, Y_t)\}_{t=1}^\infty$ is i.i.d. when φ is fixed. It is therefore stationary and geometrically β -mixing, with $\beta(k) = 0$ for every $k \geq 1$, which verifies part (a). Conditional on the stressed factors $\mathbf{F}_{t+1}^{(1:2)} = \mathbf{z}$, the remaining credit factor has the Gaussian conditional distribution used in (10), and the idiosyncratic noise remains independent of $\mathcal{F}_t$. Consequently, the conditional distribution of G'_{t+1} , and hence that of Y_{t+1} , depends on $\mathcal{F}_t$ only through the specified scenario $\mathbf{z} = \mathbf{W}_{t+1}(\mathbf{z})$. This gives (25) and verifies part (b). Moreover, $\mathbf{W}_t = \mathbf{F}_t^{(1:2)} \sim \mathcal{N}(\mathbf{0}, \delta\Sigma_{ss})$, where Σ_{ss} is positive definite. Its density is continuous and strictly positive, and is therefore bounded above and away from zero on every compact enlarged operational region $\mathcal{D}^+$. Thus, part (c) holds. Finally, by (10), the conditional log return is Gaussian with strictly positive variance independent of $\mathbf{w}$, while its mean varies affinely with $\mathbf{w}$. The normalized option gain in (EC.9) and the oracle centering function defined below are smooth functions of their arguments. It follows that the conditional distribution of Y_t is continuous and that its conditional CDF varies uniformly continuously with $\mathbf{w}$ over the compact region $\mathcal{D}^+$. This verifies part (d).

Kernel and Bandwidth Tuning For the kernel κ , we use the weighted RBF kernel

$$\kappa_h(\mathbf{w}, \mathbf{w}') := \exp\left(-\frac{(\mathbf{w} - \mathbf{w}')^\top W(\mathbf{w} - \mathbf{w}')}{2h^2}\right), \quad (\text{EC.11})$$

where h is the bandwidth and $W = \delta^{-1}\Sigma_{ss}^{-1}$. Since $\mathbf{W}_t \sim \mathcal{N}(\mathbf{0}, \delta\Sigma_{ss})$, this choice of W standardizes the two stressed factors. Although (EC.11) is not radial with respect to the Euclidean distance in the original coordinates unless W is proportional to the identity, it becomes an ordinary radial RBF kernel under the fixed invertible transformation $\bar{\mathbf{w}} = W^{1/2}\mathbf{w}$, i.e., $\kappa_h(\mathbf{w}, \mathbf{w}') = \exp\left(-\frac{\|\bar{\mathbf{w}} - \bar{\mathbf{w}}'\|^2}{2h^2}\right)$. Thus, applying KSA to the original features with the weighted kernel is equivalent to applying it to the standardized features with radial function $\tilde{\kappa}(r) = e^{-r^2/2}$. This function is bounded and Lipschitz, has support $[0, \infty)$, has a positive finite integral, and has a finite second moment. It therefore satisfies Assumption 2(e). Hence all parts of Assumption 2 hold in the standardized representation, and Theorem 6.1 applies to the corresponding fixed-bandwidth implementation.

For the numerical experiment, we select the bandwidth adaptively during backtesting. At each time t and for each scenario z , we choose a scenario-specific bandwidth $h_t(z)$ from the candidate set $\{5, 1, 0.5, 0.1, 0.05\}$. For each candidate bandwidth, we evaluate its historical empirical coverage using observations available prior to time t , excluding the observation being evaluated from the construction of its own prediction interval, and select the bandwidth whose coverage is closest to the target level $1 - \alpha = 50\%$. The bandwidth is reselected at every time step and may therefore vary across both time and scenarios.

Expectation Predictor φ We use the oracle expectation predictor φ defined as the true conditional expectation of G'_{t+1} . For any stress scenario $\mathbf{z} := (z_{\text{Oil}}, z_{\text{Rate}})$, it takes the form

$$\varphi(\mathbf{z}) = \widetilde{\text{BS}}(m, \tau, \sigma_{\text{tot}}, 1) - \mathbb{E}_{\mathcal{Z}} \left[e^{m_{\mathbf{z}} + \sqrt{v_{\text{Cred}}} \cdot \mathcal{Z}} \cdot \widetilde{\text{BS}} \left(m \cdot e^{-m_{\mathbf{z}} - \sqrt{v_{\text{Cred}}} \cdot \mathcal{Z}}, \tau - \delta, \sigma_{\text{tot}}, 1 \right) \right], \quad (\text{EC.12})$$

where $\mathcal{Z} \sim \mathcal{N}(0, 1)$, and $m_{\mathbf{z}}$ and v_{Cred} are determined by $\mathbf{z}$ through (11). We approximate the right-hand side of (EC.12) using Monte Carlo simulation by drawing i.i.d. samples of $\mathcal{Z}$ and replacing $\mathbb{E}_{\mathcal{Z}}[\cdot]$ with the corresponding empirical average. The theoretical verification above concerns the exact fixed oracle function φ ; the Monte Carlo approximation is an implementation device used to evaluate this function numerically.

EC.2 Experimental Details for Section 7

EC.2.1 Model Calibration

The parameters of the common factor model include $\mathbf{I}_0$, λ_i for $i = 1, 2, 3, 4$, $\mathbf{M}_i, \gamma_i, \sigma_i, \eta_i$ for $i = 2, 3, 4$, and σ_G . We calibrate the model on a real financial dataset. We obtain the implied volatility data on the S&P 500 index from the OptionMetrics IVY database, during a selected period from January 2006 to August 2013. The data is transferred to the moneyness-maturity coordinates. We select 90 different options to construct the log-implied volatility surface, with their moneyness-maturity pairs from $\{(m_j, \tau_j)\}_{j=1}^{90} = \{(m, \tau) : m \in \Omega_m, \tau \in \Omega_\tau\}$, where $\Omega_m = \{0.5, 0.6, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 0.975, 1, 1.025, 1.05, 1.1, 1.15, 1.2, 1.25, 1.3, 1.4\}$ and $\Omega_\tau = \{1/12, 1/6, 1/4, 1/2, 1\}$. The dataset consists of $T = 1890$ time steps. Applying PCA to the log-implied volatility records $\{\mathbf{I}_t\}_{t=0}^{T-1}$, we obtain the first three principal component directions, denoted as $\mathbf{M}_2, \mathbf{M}_3, \mathbf{M}_4$, along with the corresponding principal component scores $\{\mathbf{f}_t\}_{t=1}^T$. For $i = 2, 3, 4$, the model parameters λ_i, γ_i and σ_i are estimated by fitting the autoregressive process specified in (29) to $\{f_t^{(i)}\}_{t=1}^T$. For the first common factor, λ_1 and σ_1 are estimated by fitting the autoregressive process in (29) to the log-underlying prices $\{\log S_t\}_{t=0}^{T-1}$. To compute η_2, η_3 and η_4 , we use the relation $\text{Cov}(\epsilon_t^{(1)}, \epsilon_t^{(i)}) = \eta_i \sigma_i^2$ for $i = 2, 3, 4$, where $\epsilon_t^{(1)}$ equals the log returns of S_t and $\epsilon_t^{(i)}$ are the residuals from fitting $\{f_t^{(i)}\}_{t=1}^T$ to the autoregressive process. Finally, σ_G is computed based on (31) as follows $\sigma_G = \sqrt{\sigma_1^2 - \eta_2^2 \sigma_2^2 - \eta_3^2 \sigma_3^2 - \eta_4^2 \sigma_4^2}$. The calibration results are as follows: $\lambda_1 = 1.9 \times 10^{-4}$, $\lambda_2 = 5.7 \times 10^{-4}$, $\lambda_3 = 1.9 \times 10^{-5}$, $\lambda_4 = -3.3 \times 10^{-5}$, $\gamma_2 = -9.6 \times 10^{-3}$, $\gamma_3 = -4.2 \times 10^{-2}$, $\gamma_4 = -6.1 \times 10^{-2}$, $\eta_2 = 3.8 \times 10^{-2}$, $\eta_3 = -1.4 \times 10^{-2}$, $\eta_4 = 4.0 \times 10^{-2}$, $\sigma_2 = 0.30$, $\sigma_3 = 0.11$, $\sigma_4 = 0.075$, $\sigma_G = 0.0028$, $\mathbf{I}_0 = [-1.47, -1.50, -1.57, -1.82, -1.94, -1.93, -1.95, -1.42, \dots]^\top$, $\mathbf{M}_2 = [-0.15, -0.15, -0.13, -0.18, -0.17, -0.17, -0.17, -0.14, \dots]^\top$, $\mathbf{M}_3 = [0.18, 0.16, 0.08, -0.07, -0.27, -0.30, -0.27, 0.19, \dots]^\top$, $\mathbf{M}_4 = [0.25, 0.30, 0.26, 0.28, 0.19, 0.17, 0.13, 0.16, \dots]^\top$. For readability, we report only the leading entries of $\mathbf{I}_0$ and $\mathbf{M}_2, \mathbf{M}_3, \mathbf{M}_4$ here. The complete calibrated arrays, together with their exact moneyness-maturity ordering, are provided in the accom-

panying code repository¹⁹.

EC.2.2 Proper Scoring Rules: the CRPS and Interval Score

We now provide a brief introduction to proper scoring rules, focusing particularly on the CRPS and interval score. For a comprehensive review of proper scoring rules, we refer the interested reader to [Gneiting and Raftery \(2007\)](#). A proper scoring rule S is a mechanism that is used to evaluate the quality of a probabilistic forecast or distribution $\hat{P}$ by assigning a numerical score that is a function of both $\hat{P}$ and the realized outcome. The expected score is then given by $\mathcal{S}(\hat{P}, P) := \int S(\hat{P}, \omega) dP(\omega)$ where we use ω to denote a generic outcome and P is the true data-generating distribution. Under the loss convention used here, a scoring rule is *proper* if $\hat{P} = P$ minimizes the expected score $\mathcal{S}(\hat{P}, P)$, and *strictly proper* if it is the unique minimizer. Smaller CRPS and interval-score values are better. The primary goal of proper scoring rules is to encourage the honest reporting of probabilities. They aid in the assessment and comparison of predictive models within statistical inference and decision theory and have been widely applied in various fields including, for example, weather forecasting and risk management.

The most commonly used score function for the probabilistic prediction of continuous random variables (scenario gains in our context) is the *continuous ranked probability score (CRPS)*. The score is defined as

$$S(\hat{P}, y) := \int_{x \in \mathbb{R}} (\hat{F}(x) - \mathbb{1}\{x \geq y\})^2 dx \quad (\text{EC.13})$$

with $\hat{F}$ denoting the cumulative distribution function of distribution $\hat{P}$. The CRPS is therefore the integral of the so-called Brier scores for the associated binary probability forecasts at all real-valued thresholds. The CRPS is proper relative to the class $\mathcal{P}$ of Borel probability measures on $\mathbb{R}$ and strictly proper relative to the subclass $\mathcal{P}_1$ of $\mathcal{P}$ whose distributions have finite first moment. Of particular interest is when $\hat{P}$ is an empirical distribution. In particular, let δ_y denote a point mass distribution at y and let $\hat{P}$ take the form $\hat{P} = \frac{1}{m} \sum_{j=1}^m \delta_{y'_j}$, then $\hat{F}(y) = \frac{1}{m} \sup\{j : y \geq y'_j\}$, with $-\infty =: y'_0 < y'_1 \leq \dots \leq y'_m$. In this case, the CRPS can be simplified to $S(\hat{P}, y) = \frac{1}{m} \sum_{j=1}^m |y - y'_j| - \frac{1}{2m^2} \sum_{i=1}^m \sum_{j=1}^m |y'_i - y'_j|$. In many practical settings (including the focus of this work), we do not estimate an entire predictive distribution $\hat{P}$. Instead, we report a prediction interval $[L, U]$. We may still want to assess whether this interval is honest. For example, if $[L, U]$ is intended to approximate the $\alpha/2$ and $1 - \alpha/2$ predictive quantiles, we would like to verify that these endpoint estimates are accurate. A convenient proper scoring rule in this setting is the *interval score*, defined

¹⁹Upon publication, we will make the code required to reproduce the numerical experiments publicly available in a GitHub repository.

by

$$S_\alpha([L, U], y) = \begin{cases} (L - y) + \frac{\alpha}{2}(U - L), & y < L, \\ \frac{\alpha}{2}(U - L), & L \leq y \leq U, \\ (y - U) + \frac{\alpha}{2}(U - L), & y > U. \end{cases}$$

This is a scaled version of the original interval score introduced in (Gneiting and Raftery, 2007). Intuitively, the score trades off sharpness and calibration: it penalizes wide intervals through the term proportional to $(U - L)$, and it additionally penalizes observations that fall outside the interval in proportion to their distance from the nearest endpoint.

The CRPS and interval score are connected in the following way. The CRPS score defined in (EC.13) can be represented as an average of the pinball losses over all levels: $S(\hat{P}, y) = 2 \int_{u=0}^1 \ell_u(y, \hat{Q}(u)) du$, where ℓ_u is the pinball loss defined in (18) and $\hat{Q}(u)$ is the u quantile of $\hat{P}$. The interval score defined here is the sum of the pinball losses at its two endpoint quantile levels: $S_\alpha([L, U], y) = \ell_{\alpha/2}(y, L) + \ell_{1-\alpha/2}(y, U)$. This makes the relationship clear that CRPS aggregates quantile performance across all quantile levels, whereas the interval score samples that same quantile-based scoring idea at only two levels $\alpha/2$ and $(1 - \alpha/2)$.

EC.2.3 Experiment and Algorithm Details

In this section, we provide detailed descriptions of the algorithms evaluated in our experiments. These implementation details apply to both experimental settings considered in Sections 7.2 and 7.3. Moreover, because the portfolio composition evolves over time, for each backtesting date t , we assume that, for all prior dates $s < t$, the portfolio holds the same asset units as at time t , and we recompute all historical statistics accordingly. This approach follows a standard backtesting convention commonly adopted in production settings. All methods use the same normalization convention in the numerical experiments. Portfolio gains are first expressed relative to the contemporaneous underlying-price scale and are then rescaled to the common reference level $S_0 = 5000$. The same normalization and rescaling are applied to the empirical-quantile benchmarks, ACSA, KSA, and the Monte Carlo oracle. All prediction intervals and all evaluation metrics reported in Tables 5 and 6 are therefore expressed on the same monetary scale. For notational simplicity, we continue to write G_s for these consistently rescaled gains in the algorithm descriptions below.

- The “MC Oracle” algorithm constructs a Monte Carlo approximation to the oracle conditional prediction interval for G_{t+1} under the stress scenario $\mathbf{z}$ and current market state. It generates synthetic draws from the known conditional data-generating distribution and uses their empirical quantiles as endpoints; these are numerical approximations, rather than exact oracle quantiles.
- The “Emp-quantile” algorithm uses the empirical quantiles of different quantities to construct the prediction intervals.

- The “Vanilla” variant of the algorithm first computes the empirical $\alpha/2$ and $1 - \alpha/2$ quantiles of $\{G_s\}_{s \leq t}$, and then uses the two empirical quantiles as the endpoints of the prediction interval for G_{t+1} .
- The “SSA” variant of the algorithm leverages the empirical quantiles of the residuals of $\{G_s\}_{s \leq t}$ subtracted by their *SSA predictions*. Concretely, let $\hat{G}_t$ denote the SSA prediction of G_t under the realized stress scenario $\mathbf{z}_t$, then the algorithm computes the empirical $\alpha/2$ and $1 - \alpha/2$ quantiles of $\{\hat{G}_{t+1} + G_s - \hat{G}_s\}_{s \leq t}$, and then uses the two empirical quantiles as the endpoints of the prediction interval for G_{t+1} .

The “ACSA” algorithm has 3 variants that differ in their choices of the quantile predictor g_t . The step size γ is fixed to 0.005.

- The “Linear- g_t ” variant implements $g_t(\mathbf{z}; \tilde{\alpha})$ as a linear function whose parameters are estimated via quantile regression.
 - The “NN- g_t ” variant implements $g_t(\mathbf{z}; \tilde{\alpha})$ as a 2-layer feedforward neural network. The network is trained by minimizing the pinball loss.
 - The “MC Oracle- g_t ” variant supplies g_t with Monte Carlo approximations to conditional quantiles of G_{t+1} given the stress scenario $\mathbf{z}$ and current state, evaluated at the interior quantile levels required by the corrected ACSA construction in (12). We emphasize that this quantile predictor is not attainable in practice, as the true conditional quantile depends not only on the stress scenario but also on historical information, such as the common factors observed on the previous day.
- The “KSA NN- φ ” row presents three variants of the KSA algorithm. The three variants differ in their choices of the historical context vector $\mathbf{H}_t$. For all the algorithm variants we implement the expectation predictor φ as a 2-layer feedforward neural network. Expectation predictors φ are fitted once using the burn-in observations and are then held fixed throughout the evaluation period.
 - The “ $\mathbf{H}_t = \mathbf{f}_t^{(2:4)}$ ” variant incorporates the full set of relevant historical information into the historical context vector. As shown in E-Companion EC.2.5, the components $\mathbf{f}_t^{(2:4)}$ contain all historical information that is relevant for the distribution of the portfolio gain G_{t+1} at time $t + 1$.
 - The “ $\mathbf{H}_t = \mathbf{f}_t^{(2:3)}$ ” and “ $\mathbf{H}_t = \mathbf{f}_t^{(2)}$ ” variants both use incomplete historical information.

For each variant of the KSA algorithm, we use the weighted RBF kernel κ_h defined in (EC.11). The explicit form of the weight matrix W is given in E-Companion EC.2.5. The bandwidth parameter h is selected as described in E-Companion EC.1.5: at each time t , we choose h from the candidate set 5, 1, 0.5, 0.1, 0.05 based on empirical coverage evaluated using historical data up to time $t - 1$.

- The “KSA $\mathbf{H}_t = \mathbf{f}_t^{(2:4)}$ ” row presents four variants of the KSA algorithm, which differ in the choices of the expectation predictor φ . For all the algorithm variants we set $\mathbf{H}_t = \mathbf{f}_t^{(2:4)}$ as the historical context vector.
 - The “no- φ ” variant sets $\varphi \equiv 0$.
 - The “SSA- φ ” variant uses the SSA estimation of the portfolio gain.
 - The “NN- φ ” variant implements φ as a 2-layer feedforward neural network.
 - The “MC Oracle- φ ” variant replaces φ with a Monte Carlo approximation to the oracle conditional expectation.
- The “ACSA + KSA” row corresponds to a new implementation that integrates the KSA quantile predictor into the ACSA framework. Specifically, the KSA quantile estimator is embedded inside ACSA while the adjusted miscoverage level $\tilde{\alpha}_t$ is updated online. The three variants differ in the choices of the historical context vector $\mathbf{H}_t$ that the KSA quantile predictor uses.

Some benchmark methods and algorithm variants we evaluate assume access to *oracle* knowledge of the relevant conditional distributions. In particular, conditional on a given stress scenario $\Delta \mathbf{f}_{t+1}^{(1:2)}$, the distribution of the remaining common factor returns $\Delta \mathbf{f}_{t+1}^{(3:4)}$ admits the following conditional mean:

$$\mathbb{E} \left[\Delta \mathbf{f}_{t+1}^{(3:4)} \mid \Delta \mathbf{f}_{t+1}^{(1:2)}, \mathbf{f}_t^{(2:4)} \right] = \begin{pmatrix} \lambda_3 + \gamma_3 f_t^{(3)} \\ \lambda_4 + \gamma_4 f_t^{(4)} \end{pmatrix} + (\epsilon_{t+1}^{(1)} - \eta_2 \epsilon_{t+1}^{(2)}) \begin{pmatrix} \frac{\eta_3 \sigma_3^2}{\eta_3^2 \sigma_3^2 + \eta_4^2 \sigma_4^2 + \sigma_G^2} \\ \frac{\eta_4 \sigma_4^2}{\eta_3^2 \sigma_3^2 + \eta_4^2 \sigma_4^2 + \sigma_G^2} \end{pmatrix}.$$

The corresponding conditional covariance matrix is given by

$$\text{Cov} \left[\Delta \mathbf{f}_{t+1}^{(3:4)} \mid \Delta \mathbf{f}_{t+1}^{(1:2)}, \mathbf{f}_t^{(2:4)} \right] = \begin{pmatrix} \frac{\sigma_3^2(\eta_4^2 \sigma_4^2 + \sigma_G^2)}{\eta_3^2 \sigma_3^2 + \eta_4^2 \sigma_4^2 + \sigma_G^2} & -\frac{\eta_3 \eta_4 \sigma_3^2 \sigma_4^2}{\eta_3^2 \sigma_3^2 + \eta_4^2 \sigma_4^2 + \sigma_G^2} \\ -\frac{\eta_3 \eta_4 \sigma_3^2 \sigma_4^2}{\eta_3^2 \sigma_3^2 + \eta_4^2 \sigma_4^2 + \sigma_G^2} & \frac{\sigma_4^2(\eta_3^2 \sigma_3^2 + \sigma_G^2)}{\eta_3^2 \sigma_3^2 + \eta_4^2 \sigma_4^2 + \sigma_G^2} \end{pmatrix}.$$

Here, $\epsilon_{t+1}^{(1)} = \Delta f_{t+1}^{(1)} - \lambda_1$ and $\epsilon_{t+1}^{(2)} = \Delta f_{t+1}^{(2)} - \lambda_2 - \gamma_2 f_t^{(2)}$. Given this conditional Gaussian specification, we use Monte Carlo simulation to generate a large number of samples of $\Delta \mathbf{f}_{t+1}^{(3:4)}$, which are then used to compute Monte Carlo approximations to oracle conditional means, quantiles, and other functionals.

The complete implementation used for the numerical experiments, including the full calibrated parameter arrays, neural-network architectures, optimizers, training hyperparameters and bandwidth-selection configuration will be available in a public GitHub repository upon publication.

EC.2.4 Computational Efficiency

To provide a fine-grained assessment of computational efficiency, we decompose runtime into four components: (1) the *initial training time*, i.e., the time required to train the models over the burn-in period (the first 2000 time steps); (2) the average daily *method-update time*, i.e., the time required at each evaluation date to update portfolio-dependent quantities used by the method; when a learned component is adaptively updated, this quantity also includes the corresponding model-update cost; (3) the average daily interval-construction time, i.e., the time required to construct the prediction interval for the target scenario after the method’s daily portfolio-dependent quantities have been updated; and (4) the *total backtest time* over a 3000-step benchmark subset of the 4000-step evaluation period. All timings are recorded on an Ubuntu system with an A100 GPU. Computational time is measured for the skew-steepener portfolio experiment reported in Table 5 of Section 7.2, and the results are summarized in Table EC.2.

Algorithm	Initial training	Avg. daily method-update	Avg. daily interval construction	Total backtest
Emp-quantile (SSA)	0	0	$< 10^{-3}$ s	< 3 s
ACSA Linear- g_t	4.8 s	0.93 s	$< 10^{-3}$ s	0.78 h
ACSA NN- g_t	40.7 s	16.2 s	$< 10^{-3}$ s	13.51 h
KSA NN- φ	25.3 s	3.76 s	3×10^{-3} s	3.14 h
ACSA + KSA	26.5 s	3.79 s	3×10^{-3} s	3.17 h

Table EC.2: Runtime decomposition for representative methods from Table 5. Initial training time is measured over the 2000-step burn-in period. The last column is the 3000-step total implied by the reported component times, computed as initial training plus 3000 times the sum of the average daily method-update and interval-construction times; inequalities in the component times are propagated to the total. The statistical experiments themselves use the 4000-step evaluation period.

The empirical-quantile baselines are computationally cheapest because they avoid model fitting altogether. Among the proposed methods, ACSA has the highest daily update cost, since the quantile predictor g_t must be retrained or fine-tuned as the portfolio evolves over time. By contrast, KSA fits the expectation predictor φ on the burn-in sample and holds it fixed throughout the evaluation period. Its daily method-update cost therefore reflects the recomputation of portfolio-dependent quantities and historical residuals under the current portfolio, rather than refitting φ . This is computationally attractive because expectation estimates can adapt to portfolio changes through weighted linear combinations across securities, whereas quantiles do not aggregate linearly across securities. KSA then obtains scenario-wise quantiles through kernel reweighting of the resulting historical residuals. The hybrid ACSA+KSA implementation inherits much of this reuse advantage while retaining ACSA’s online calibration update.

EC.2.5 Historical Context and Kernel for KSA

In this section, we discuss the choice of the historical context vector $\mathbf{H}_t$ and the kernel κ used in the KSA algorithm. The structure of this section, as well as several definitions, follow E-Companion [EC.1.5](#). Our discussion serves two main purposes. First, we show that setting $\mathbf{H}_t = \mathbf{f}_t^{(2:4)}$ is sufficient to capture all relevant historical information. Second, we provide the explicit form of the kernel that our experiment implements.

Historical Context The portfolio gain G_{t+1} has the explicit representation

$$G_{t+1} = u_1(S_{t+1} - S_t) + \sum_{j=1}^{n-1} u_{j+1} \left(\text{BS}(S_{t+1}, m_j S_t, \tau_j - \delta, \sigma_{t+1}^{(j)}, c_j) - \text{BS}(S_t, m_j S_t, \tau_j, \sigma_t^{(j)}, c_j) \right),$$

where $\text{BS}(\cdot)$ denotes the Black-Scholes pricing function, and c_j is an indicator of the option type ($c_j = 1$ for a call and $c_j = 0$ for a put). As discussed in E-Companion [EC.1.5](#), constructing a prediction interval for G_{t+1} is equivalent to constructing one for the normalized portfolio gain $G'_{t+1} := G_{t+1}/S_t$, which admits the explicit form

$$G'_{t+1} = u_1 \left(\frac{S_{t+1}}{S_t} - 1 \right) + \sum_{j=1}^{n-1} u_{j+1} \left(\frac{S_{t+1}}{S_t} \widetilde{\text{BS}} \left(\frac{m_j S_t}{S_{t+1}}, \tau_j - \delta, \sigma_{t+1}^{(j)}, c_j \right) - \widetilde{\text{BS}}(m_j, \tau_j, \sigma_t^{(j)}, c_j) \right). \quad (\text{EC.14})$$

Here, $\widetilde{\text{BS}}(\cdot)$ is defined in [\(EC.10\)](#). From [\(EC.14\)](#), the components that constitute the *historical information* at time t are $\{\sigma_t^{(j)}\}_{j=1}^{n-1}$, while the quantities that remain random at time t are the underlying return S_{t+1}/S_t and the future volatilities $\{\sigma_{t+1}^{(j)}\}_{j=1}^{n-1}$. Therefore, for a historical context vector $\mathbf{H}_t$ to be sufficiently informative, it suffices that $\mathbf{H}_t$ both determines $\{\sigma_t^{(j)}\}_{j=1}^{n-1}$ and contains all factors that drive the distributions of S_{t+1}/S_t and $\{\sigma_{t+1}^{(j)}\}_{j=1}^{n-1}$. From the model construction, choosing $\mathbf{H}_t = \mathbf{f}_t^{(2:4)}$ satisfies all of the above requirements. In particular, from [\(27\)](#), we have $(\sigma_t^{(1)}, \dots, \sigma_t^{(n-1)}) = \exp(\mathbf{I}_0 + f_t^{(2)} \cdot \mathbf{M}_2 + f_t^{(3)} \cdot \mathbf{M}_3 + f_t^{(4)} \cdot \mathbf{M}_4)$, which shows that $\mathbf{f}_t^{(2:4)}$ fully determines the current volatilities $\{\sigma_t^{(j)}\}_{j=1}^{n-1}$. Moreover, by [\(28\)](#), the distributions of S_{t+1}/S_t and $\{\sigma_{t+1}^{(j)}\}_{j=1}^{n-1}$ are driven by $\Delta \mathbf{f}_{t+1}$. As further specified in [\(29\)](#), the distribution of $\Delta \mathbf{f}_{t+1}^{(1)}$ is history-independent, while the distribution of $\mathbf{f}_{t+1}^{(2:4)}$ is entirely determined by $\mathbf{f}_t^{(2:4)}$.

We can further show that Assumption [2](#) is satisfied with $\mathbf{H}_t = \mathbf{f}_t^{(2:4)}$. As in the analysis of Section [6.2](#), fix a deterministic continuous centering function φ and set $Y_t = G'_t - \varphi(\mathbf{W}_t)$. The discussion above establishes Part [\(b\)](#), because subtracting a function of $\mathbf{W}_t$ does not change predictive sufficiency. It remains to verify Parts [\(a\)](#), [\(c\)](#), and [\(d\)](#). To begin, rewrite $\mathbf{W}_t$ as follows:

$$\mathbf{W}_t = (\Delta f_t^{(1)}, \Delta f_t^{(2)}, f_{t-1}^{(2)}, f_{t-1}^{(3)}, f_{t-1}^{(4)}) = \boldsymbol{\lambda} + \mathbf{A}_1 \mathbf{f}_t^{(2:4)} + \mathbf{A}_0 \mathbf{f}_{t-1}^{(2:4)} + \boldsymbol{\epsilon}_t^G$$

where

$$\boldsymbol{\lambda} = \begin{pmatrix} \lambda_1 - \eta_2\lambda_2 - \eta_3\lambda_3 - \eta_4\lambda_4 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{A}_1 = \begin{pmatrix} \eta_2 & \eta_3 & \eta_4 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{A}_0 = \begin{pmatrix} -\eta_2(1 + \gamma_2) & -\eta_3(1 + \gamma_3) & -\eta_4(1 + \gamma_4) \\ -1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and $\boldsymbol{\epsilon}_t^G = [\epsilon_t^G, 0, 0, 0, 0]^\top$. For the calibrated parameters, $|1 + \gamma_i| < 1$ for $i = 2, 3, 4$. Taking the Gaussian OU factors in their stationary regime, $\{\mathbf{f}_t^{(2:4)}\}_{t=1}^\infty$ is stationary and geometrically β -mixing. The corresponding properties of $\{\mathbf{W}_t\}_{t=1}^\infty$ then follow from [Bradley \(2005\)](#) and the following two facts:

- (1) If $V_t = f(X_t, X_{t-1})$ for any measurable function f , then the β -mixing coefficients of $\{X_t\}_{t=1}^\infty$ and $\{V_t\}_{t=1}^\infty$, denoted by $\beta_X(k)$ and $\beta_V(k)$, satisfy $\beta_V(k) \leq \beta_X(k-1)$.
- (2) If $\{\epsilon_t\}_{t=1}^\infty$ is i.i.d. and independent of $\{X_t\}_{t=1}^\infty$, then the augmented process $\{(X_t, \epsilon_t)\}_{t=1}^\infty$ has β -mixing coefficients no larger than those of $\{X_t\}_{t=1}^\infty$.

By [\(EC.14\)](#) and the factor dynamics, $(\mathbf{W}_t, G'_t)$, and hence $(\mathbf{W}_t, Y_t)$, is a measurable function of $(\mathbf{f}_t^{(2:4)}, \epsilon_t^G)$ and $(\mathbf{f}_{t-1}^{(2:4)}, \epsilon_{t-1}^G)$. The two facts above therefore imply that $\{(\mathbf{W}_t, Y_t)\}_{t=1}^\infty$ is stationary and geometrically β -mixing. This verifies Assumption [2\(a\)](#). The vector $\mathbf{W}_t$ is a nondegenerate stationary Gaussian vector, so its density is continuous and strictly positive. It is therefore bounded above and bounded away from zero on the compact region $\mathcal{D}^+$, as required by Assumption [2\(c\)](#). Finally, conditional on $\mathbf{W}_t = \mathbf{w}$, the remaining factor innovations have a nondegenerate Gaussian distribution whose parameters vary continuously with $\mathbf{w}$. For the nondegenerate portfolios considered here, [\(EC.14\)](#) is a smooth, nonconstant function of these innovations. Since φ is continuous, the conditional distribution of Y_t is therefore continuous and its CDF varies continuously with $\mathbf{w}$. Compactness of $\mathcal{D}^+$ makes this continuity uniform, verifying Assumption [2\(d\)](#).

The Kernel We use the weighted RBF kernel κ_h defined in [\(EC.11\)](#). The bandwidth h is adaptively tuned, and the weight matrix W standardizes the feature vector $\mathbf{W}_t = (\Delta f_t^{(1)}, \Delta f_t^{(2)},$

$f_{t-1}^{(2)}, f_{t-1}^{(3)}, f_{t-1}^{(4)}$. Specifically, the covariance matrix of $\mathbf{W}_t$ takes the form

$$\text{Cov}(\mathbf{W}_t) = \begin{pmatrix} \sigma_1^2 & \eta_2 \sigma_2^2 & 0 & 0 & 0 \\ \eta_2 \sigma_2^2 & \frac{2\sigma_2^2}{2+\gamma_2} & \frac{\gamma_2 \sigma_2^2}{1-(1+\gamma_2)^2} & 0 & 0 \\ 0 & \frac{\gamma_2 \sigma_2^2}{1-(1+\gamma_2)^2} & \frac{\sigma_2^2}{1-(1+\gamma_2)^2} & 0 & 0 \\ 0 & 0 & 0 & \frac{\sigma_3^2}{1-(1+\gamma_3)^2} & 0 \\ 0 & 0 & 0 & 0 & \frac{\sigma_4^2}{1-(1+\gamma_4)^2} \end{pmatrix}$$

and we set $W = \text{Cov}(\mathbf{W}_t)^{-1}$. Equivalently, under the fixed invertible standardization $\bar{\mathbf{w}} = W^{1/2} \mathbf{w}$, the kernel becomes $\kappa_h(\mathbf{w}, \mathbf{w}') = \exp\left(-\frac{\|\bar{\mathbf{w}} - \bar{\mathbf{w}}'\|^2}{2h^2}\right)$. Thus, in the standardized coordinates, its radial function is $\tilde{\kappa}(r) = e^{-r^2/2}$, which satisfies Assumption 2(e). The fixed invertible transformation does not affect the preceding stationarity, mixing, density, predictive-sufficiency, or continuity conclusions.

In some experiments, we additionally consider incomplete feature vectors. For instance, we may set $\mathbf{H}_t = \mathbf{f}_t^{(2:3)}$ instead of $\mathbf{f}_t^{(2:4)}$. In such cases, the weight matrix is taken to be the corresponding submatrix of W associated with the truncated feature vector. The verification of Assumption 2 above applies to the full context $\mathbf{H}_t = \mathbf{f}_t^{(2:4)}$; the incomplete-context specifications are included as deliberately misspecified empirical comparisons and need not satisfy Part (b).

EC.2.6 Additional Experiment Results

Tables EC.3 and EC.4 present results under the same settings as Tables 7 and 8, but evaluated at a different time step.

EC.2.7 Construction of the Adversarial Portfolio

The portfolio consists of $n = 13$ securities. The first asset is the S&P 500 index. The remaining $n-1$ assets are European options characterized by moneyness $\{m_j\}_{j=1}^{n-1}$, time to maturity $\{\tau_j\}_{j=1}^{n-1}$, implied volatility $\{\sigma_t^{(j)}\}_{j=1}^{n-1}$, and a call-put indicator $\{c_j\}_{j=1}^{n-1}$. Let $\{C_t^{(j)}\}_{j=1}^{n-1}$ denote the corresponding option prices, given by $C_t^{(j)} = \text{BS}\left(S_t, m_j S_t, \tau_j, \sigma_t^{(j)}, c_j\right)$, where $\text{BS}(\cdot)$ denotes the Black-Scholes pricing formula. The option universe is constructed by forming all combinations of moneyness and maturity levels: $\{(m_j, \tau_j)\}_{j=1}^{12} = \{(m, \tau) : m \in \Omega_m, \tau \in \Omega_\tau\}$, where $\Omega_m = \{0.9, 0.95, 1.05, 1.1\}$ and $\Omega_\tau = \{1/12, 1/4, 1/2\}$. For each option, the call-put indicator is $c_j = 1$ when $m_j \geq 1$ (call option) and $c_j = 0$ when $m_j < 1$ (put option). The adversarial portfolio $\mathbf{u}$ (represented by the

SSA: Adversarial Portfolio								Cond Exp: Adversarial Portfolio							
$\Delta f_{t+1}^{(1)}$ ($\times 10^{-2}$)	$\Delta f_{t+1}^{(2)}$ ($\times 10^{-1}$)							$\Delta f_{t+1}^{(1)}$ ($\times 10^{-2}$)	$\Delta f_{t+1}^{(2)}$ ($\times 10^{-1}$)						
	-3	-2	-1	0	1	2	3		-3	-2	-1	0	1	2	3
-1.5	2.93	1.43	0.17	-0.89	-1.76	-2.45	-3.00	-1.5	6.64	7.07	8.03	9.37	10.86	12.64	14.50
-1.0	3.00	2.30	1.78	1.42	1.20	1.09	1.09	-1.0	4.13	5.19	6.56	8.29	10.27	12.36	14.46
-0.5	2.04	1.90	1.91	2.05	2.29	2.61	3.00	-0.5	1.43	2.44	3.87	5.73	7.84	10.02	12.28
0.0	0.27	0.45	0.78	1.22	1.75	2.34	2.98	0.0	-0.17	-0.00	0.94	2.44	4.23	6.32	8.47
0.5	-1.75	-1.50	-1.10	-0.56	0.07	0.78	1.54	0.5	1.28	-0.30	-0.68	-0.31	0.75	2.24	3.91
1.0	-3.00	-3.00	-2.79	-2.41	-1.88	-1.25	-0.54	1.0	9.00	4.26	1.28	-0.32	-0.88	-0.70	0.08
1.5	-1.94	-2.57	-2.90	-2.98	-2.85	-2.55	-2.11	1.5	27.46	17.76	10.57	5.41	1.99	-0.10	-1.16

SSA: Factor-neutral benchmark								Cond Exp: Factor-neutral benchmark							
$\Delta f_{t+1}^{(1)}$ ($\times 10^{-2}$)	$\Delta f_{t+1}^{(2)}$ ($\times 10^{-1}$)							$\Delta f_{t+1}^{(1)}$ ($\times 10^{-2}$)	$\Delta f_{t+1}^{(2)}$ ($\times 10^{-1}$)						
	-3	-2	-1	0	1	2	3		-3	-2	-1	0	1	2	3
-1.5	0.83	0.46	0.13	-0.15	-0.39	-0.59	-0.75	-1.5	0.43	-0.28	-0.90	-1.43	-1.88	-2.25	-2.56
-1.0	0.31	0.06	-0.15	-0.34	-0.49	-0.62	-0.71	-1.0	0.37	-0.14	-0.60	-1.00	-1.35	-1.65	-1.88
-0.5	0.07	-0.05	-0.17	-0.26	-0.34	-0.41	-0.45	-0.5	0.25	0.01	-0.25	-0.50	-0.74	-0.95	-1.14
0.0	-0.07	-0.06	-0.06	-0.07	-0.07	-0.07	-0.07	0.0	-0.29	-0.12	-0.07	-0.09	-0.17	-0.26	-0.37
0.5	-0.34	-0.17	-0.03	0.09	0.18	0.26	0.32	0.5	-1.82	-0.98	-0.42	-0.05	0.16	0.27	0.30
1.0	-1.07	-0.66	-0.31	-0.02	0.22	0.42	0.58	1.0	-5.30	-3.38	-1.94	-0.90	-0.17	0.33	0.66
1.5	-2.67	-1.92	-1.28	-0.73	-0.26	0.13	0.46	1.5	-12.01	-8.46	-5.67	-3.49	-1.84	-0.64	0.23

Table EC.3: Portfolio gains across the scenario grid corresponding to the first two common factor returns ($\Delta f_{t+1}^{(1)}, \Delta f_{t+1}^{(2)}$). The table entries are computed at a different time step from Table 7. Cond Exp denotes Monte Carlo approximations to the oracle conditional mean gains.

ACSA 90% PI ($\tilde{\alpha}_t = 0.09$): Adversarial Portfolio

$\Delta f_{t+1}^{(1)}$ ($\times 10^{-2}$)	$\Delta f_{t+1}^{(2)}$ ($\times 10^{-1}$)						
	-3	-2	-1	0	1	2	3
-1.5	(-4.9, 19.4)	(-3.3, 18.8)	(-1.5, 18.6)	(0.7, 19.0)	(2.9, 19.7)	(5.2, 20.7)	(7.6, 21.9)
-1.0	(-5.9, 15.5)	(-4.2, 15.6)	(-2.1, 16.1)	(0.3, 17.1)	(2.8, 18.4)	(5.3, 19.9)	(7.9, 21.5)
-0.5	(-6.9, 11.2)	(-5.6, 11.6)	(-3.8, 12.5)	(-1.6, 13.8)	(0.9, 15.4)	(3.4, 17.1)	(6.0, 19.0)
0.0	(-6.5, 7.5)	(-6.4, 7.6)	(-5.5, 8.3)	(-4.0, 9.6)	(-2.0, 11.1)	(0.2, 13.0)	(2.5, 14.8)
0.5	(-3.0, 6.4)	(-4.7, 5.2)	(-5.4, 5.0)	(-5.3, 5.6)	(-4.5, 6.7)	(-3.1, 8.2)	(-1.5, 9.8)
1.0	(4.1, 14.5)	(0.3, 8.5)	(-2.3, 5.2)	(-3.9, 3.7)	(-4.8, 3.6)	(-4.9, 4.1)	(-4.5, 5.1)
1.5	(17.6, 39.8)	(10.4, 26.8)	(5.0, 17.1)	(1.1, 10.1)	(-1.6, 5.7)	(-3.6, 3.5)	(-4.8, 2.7)

ACSA 90% PI ($\tilde{\alpha}_t = 0.1$): Factor-neutral benchmark

$\Delta f_{t+1}^{(1)}$ ($\times 10^{-2}$)	$\Delta f_{t+1}^{(2)}$ ($\times 10^{-1}$)						
	-3	-2	-1	0	1	2	3
-1.5	(-0.3, 1.4)	(-0.9, 0.5)	(-1.4, -0.3)	(-1.9, -0.9)	(-2.3, -1.4)	(-2.7, -1.8)	(-3.0, -2.1)
-1.0	(-0.1, 1.1)	(-0.5, 0.4)	(-1.0, -0.2)	(-1.4, -0.6)	(-1.8, -0.9)	(-2.1, -1.2)	(-2.4, -1.4)
-0.5	(-0.1, 0.8)	(-0.2, 0.3)	(-0.5, -0.0)	(-0.8, -0.2)	(-1.2, -0.3)	(-1.4, -0.5)	(-1.7, -0.6)
0.0	(-1.0, 0.3)	(-0.5, 0.2)	(-0.2, 0.0)	(-0.3, 0.1)	(-0.5, 0.2)	(-0.7, 0.2)	(-0.9, 0.2)
0.5	(-3.5, -0.6)	(-2.0, -0.2)	(-1.0, -0.0)	(-0.3, 0.2)	(-0.1, 0.4)	(-0.1, 0.7)	(-0.2, 0.8)
1.0	(-8.4, -2.8)	(-5.7, -1.6)	(-3.5, -0.8)	(-2.0, -0.2)	(-0.8, 0.3)	(-0.0, 0.7)	(0.3, 1.0)
1.5	(-17.6, -7.4)	(-12.8, -4.9)	(-8.9, -3.1)	(-5.9, -1.6)	(-3.6, -0.6)	(-1.8, 0.2)	(-0.5, 0.8)

KSA 90% PI: Adversarial Portfolio

$\Delta f_{t+1}^{(1)}$ ($\times 10^{-2}$)	$\Delta f_{t+1}^{(2)}$ ($\times 10^{-1}$)						
	-3	-2	-1	0	1	2	3
-1.5	(-48.4, 37.7)	(-58.7, 32.2)	(-64.3, 27.1)	(-83.7, 26.0)	(-99.9, 25.2)	(-100.6, 24.5)	(-101.1, 23.9)
-1.0	(-35.3, 45.2)	(-42.2, 36.4)	(-51.9, 30.3)	(-59.6, 28.4)	(-64.2, 28.1)	(-97.0, 28.0)	(-97.0, 28.0)
-0.5	(-26.6, 55.6)	(-31.9, 45.4)	(-39.7, 36.8)	(-45.5, 30.5)	(-54.7, 26.1)	(-58.8, 27.3)	(-77.5, 29.9)
0.0	(-20.4, 74.8)	(-24.5, 58.0)	(-29.9, 45.9)	(-35.8, 36.1)	(-41.9, 30.0)	(-49.9, 24.1)	(-56.0, 21.7)
0.5	(-15.9, 94.7)	(-17.4, 77.8)	(-22.3, 61.5)	(-29.0, 48.0)	(-33.8, 36.3)	(-41.3, 29.0)	(-47.1, 21.5)
1.0	(-12.1, 146.6)	(-15.4, 108.2)	(-16.9, 77.8)	(-18.9, 65.9)	(-25.7, 51.0)	(-32.7, 37.0)	(-40.5, 27.8)
1.5	(-6.9, 234.2)	(-8.4, 218.1)	(-12.0, 115.5)	(-13.5, 79.5)	(-17.1, 71.7)	(-23.6, 52.8)	(-31.0, 38.9)

KSA 90% PI: Factor-neutral benchmark

$\Delta f_{t+1}^{(1)}$ ($\times 10^{-2}$)	$\Delta f_{t+1}^{(2)}$ ($\times 10^{-1}$)						
	-3	-2	-1	0	1	2	3
-1.5	(-9.0, 12.2)	(-7.6, 12.4)	(-7.1, 13.1)	(-6.7, 15.0)	(-6.4, 15.8)	(-5.1, 14.6)	(-4.1, 10.6)
-1.0	(-11.2, 11.4)	(-9.8, 11.5)	(-8.0, 11.9)	(-7.2, 13.3)	(-6.5, 14.7)	(-6.3, 15.9)	(-5.2, 14.5)
-0.5	(-15.8, 9.9)	(-12.0, 10.9)	(-10.0, 11.4)	(-8.0, 12.0)	(-6.9, 13.7)	(-6.2, 14.8)	(-5.4, 16.1)
0.0	(-20.4, 7.2)	(-17.9, 8.7)	(-12.5, 10.5)	(-10.4, 11.5)	(-7.8, 12.7)	(-6.1, 14.2)	(-5.0, 15.1)
0.5	(-27.7, 3.3)	(-20.7, 6.1)	(-18.9, 7.7)	(-14.2, 10.1)	(-10.7, 11.7)	(-8.1, 13.6)	(-5.7, 14.8)
1.0	(-37.8, 1.8)	(-30.0, 2.7)	(-21.8, 4.9)	(-19.8, 7.3)	(-15.9, 9.5)	(-11.1, 11.9)	(-8.3, 14.3)
1.5	(-69.7, -0.8)	(-52.9, 0.7)	(-34.9, 2.0)	(-26.4, 3.4)	(-20.7, 6.6)	(-17.9, 8.6)	(-12.4, 11.7)

Table EC.4: ACSA and KSA 90% prediction intervals. The prediction intervals are computed at a different time step from Table 8.

number of units held in each security) is obtained by solving the following optimization problem:

$$\begin{aligned}
& \max_{\mathbf{u}^+, \mathbf{u}^- \geq 0} \quad \tilde{G}_{t+1}(\mathbf{u}^+ - \mathbf{u}^-, \mathbf{e}_3), \\
& \text{s.t.} \quad |\hat{G}_{t+1,SSA}(\mathbf{u}^+ - \mathbf{u}^-, \mathbf{z}^i)| \leq 3, \quad i = 1, \dots, m \\
& \quad \mathbf{C}_t^\top \mathbf{u}^+ \leq 100, \quad \mathbf{C}_t^\top \mathbf{u}^- + \mathbf{C}_t^\top \mathbf{u}^+ \leq 500,
\end{aligned} \tag{EC.15}$$

where

- $\tilde{G}_{t+1}(\mathbf{u}, \Delta \mathbf{f}_{t+1})$: the approximated portfolio gain as a linear function of the portfolio position vector $\mathbf{u}$ and the common factor return $\Delta \mathbf{f}_{t+1}$.
- $\mathbf{u}^+, \mathbf{u}^-$: nonnegative auxiliary position variables, with the net portfolio position given by $\mathbf{u} = \mathbf{u}^+ - \mathbf{u}^-$.
- $\mathbf{e}_i = (0, \dots, 1, \dots, 0) \in \mathbb{R}^4$: the i -th canonical basis vector.
- $\hat{G}_{t+1,SSA}(\mathbf{u}, \mathbf{z})$: the SSA estimate of the portfolio gain as a function of the portfolio position vector $\mathbf{u}$ and the stress scenario $\mathbf{z}$.
- $\{\mathbf{z}^1, \dots, \mathbf{z}^m\}$: the collection of candidate stress scenarios used in the scenario analysis table. In this experiment, we set $m = 49$ with $\{\mathbf{z}^i\}_{i=1}^{49} = \{(z_1, z_2) : z_1 \in \Omega_1, z_2 \in \Omega_2\}$, where $\Omega_1 = \{-0.015, -0.01, -0.005, 0, 0.005, 0.01, 0.015\}$ and $\Omega_2 = \{-0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3\}$.
- $\mathbf{C}_t = (S_t, C_t^{(1)}, \dots, C_t^{(n-1)})$: the vector of current security values.

The objective of program (EC.15) maximizes the portfolio gain under an upward movement of the third factor, represented by $\mathbf{e}_3$. The first constraint requires the SSA estimate of the portfolio gain to lie within $\pm 3\%$ of the total capital. The second constraint ensures that long positions are fully financed by the \$100 capital. Finally, the last constraint places a cap on the total notional represented by the positive and negative auxiliary position variables in the split-position formulation. The explicit forms of $\tilde{G}_{t+1}(\mathbf{u}, \Delta \mathbf{f}_{t+1})$ and $\hat{G}_{t+1,SSA}(\mathbf{u}, \mathbf{z})$ are presented next.

First Order Approximation of G_{t+1} A first-order approximation to the portfolio gain gives:

$$\begin{aligned}
G_{t+1} &= u_1 \cdot (S_{t+1} - S_t) + \sum_{j=1}^{n-1} u_{j+1} \cdot (C_{t+1}^{(j)} - C_t^{(j)}) \\
&\approx u_1 \cdot dS_{t+1} + \sum_{j=1}^{n-1} u_{j+1} \cdot \left(\delta_t^{(j)} dS_{t+1} + \nu_t^{(j)} \cdot d\sigma_{t+1}^{(j)} - \frac{1}{252} \Theta_t^{(j)} \right)
\end{aligned} \tag{EC.16}$$

where the Greek letters are: $\delta_t^{(j)} := \frac{\partial}{\partial S} \text{BS}(S, m_j S_t, \tau_j, \sigma_t^{(j)}, c_j) \Big|_{S=S_t}$, $\nu_t^{(j)} := \frac{\partial}{\partial \sigma} \text{BS}(S_t, m_j S_t, \tau_j, \sigma, c_j) \Big|_{\sigma=\sigma_t^{(j)}}$, and $\Theta_t^{(j)} := \frac{\partial}{\partial \tau} \text{BS}(S_t, m_j S_t, \tau, \sigma_t^{(j)}, c_j) \Big|_{\tau=\tau_j}$. Equation (EC.16) decomposes fluctuations in the portfolio value into three components: (i) movements in the underlying index, (ii) the passage of time,

which reduces the remaining maturity, and (iii) shifts in the implied volatility surface. Moreover, (EC.16) implies that the portfolio gain can be approximated by a linear function of the common factor returns $\Delta \mathbf{f}_{t+1}$. To see this, recall from (28) that

$$\begin{aligned} dS_{t+1} &\approx S_t \cdot \Delta f_{t+1}^{(1)} \\ (d\sigma_{t+1}^{(1)}, \dots, d\sigma_{t+1}^{(n-1)}) &\approx (\sigma_t^{(1)}, \dots, \sigma_t^{(n-1)}) \odot \Delta \mathbf{I}_{t+1} \\ &= (\sigma_t^{(1)}, \dots, \sigma_t^{(n-1)}) \odot \left(\Delta f_{t+1}^{(2)} \cdot \mathbf{M}_2 + \Delta f_{t+1}^{(3)} \cdot \mathbf{M}_3 + \Delta f_{t+1}^{(4)} \cdot \mathbf{M}_4 \right) \end{aligned} \quad (\text{EC.17})$$

where $\odot$ denotes the Hadamard (element-wise) product. Define the approximated portfolio gain $\tilde{G}_{t+1}(\mathbf{u}, \Delta \mathbf{f}_{t+1})$ as a linear function of the common factor returns $\Delta \mathbf{f}_{t+1}$ and the portfolio position vector $\mathbf{u}$. This approximation is obtained by substituting (EC.17) into the linearized gain expression in (EC.16). The explicit expression of $\tilde{G}_{t+1}$ is

$$\begin{aligned} \tilde{G}_{t+1}(\mathbf{u}, \Delta \mathbf{f}_{t+1}) &= u_1 S_t \Delta f_{t+1}^{(1)} + S_t \Delta f_{t+1}^{(1)} \cdot \boldsymbol{\delta}_t^\top \mathbf{u}^{(2:n)} \\ &\quad + \left(\boldsymbol{\nu}_t \odot \boldsymbol{\sigma}_t \odot \left(\Delta f_{t+1}^{(2)} \cdot \mathbf{M}_2 + \Delta f_{t+1}^{(3)} \cdot \mathbf{M}_3 + \Delta f_{t+1}^{(4)} \cdot \mathbf{M}_4 \right) - \frac{1}{252} \boldsymbol{\Theta}_t \right)^\top \mathbf{u}^{(2:n)} \end{aligned}$$

SSA Estimate of G_{t+1} Under a given stress scenario $\mathbf{z} = (z_1, z_2)$, the SSA estimate $\hat{G}_{t+1,SSA}(\mathbf{u}, \mathbf{z})$ is given by

$$\hat{G}_{t+1,SSA}(\mathbf{u}, \mathbf{z}) = u_1 S_t (e^{z_1} - 1) + \sum_{j=1}^{n-1} u_{j+1} S_t \left(e^{z_1} \widetilde{\text{BS}}(m_j e^{-z_1}, \tau_j - \delta, \sigma_{t+1}^{(j)}, c_j) - \widetilde{\text{BS}}(m_j, \tau_j, \sigma_t^{(j)}, c_j) \right)$$

where the post-stress implied volatilities satisfy $(\sigma_{t+1}^{(1)}, \dots, \sigma_{t+1}^{(n-1)}) = (\sigma_t^{(1)}, \dots, \sigma_t^{(n-1)}) \odot \exp(z_2 \cdot \mathbf{M}_2)$. This expression is derived analogously to (EC.14).

Factor-neutral benchmark. In a similar spirit to (EC.15), we construct a factor-neutral benchmark in a relative sense. The benchmark uses the same linearized one-day gain approximation as above, but imposes additional constraints requiring the approximated one-day gain to vanish under unit movements in each of the four modeled common-factor directions. Because the approximation also contains the one-day time-decay contribution, these constraints do not in general imply exactly zero marginal sensitivity to each factor. Rather, their role is to make the benchmark more tightly controlled than the adversarial portfolio along all four modeled factor directions.

$$\begin{aligned} \max_{\mathbf{u}^+, \mathbf{u}^- \geq 0} \quad & \mathbf{C}_t^\top \mathbf{u}^- + \mathbf{C}_t^\top \mathbf{u}^+, \\ \text{s.t.} \quad & |\hat{G}_{t+1,SSA}(\mathbf{u}^+ - \mathbf{u}^-, \mathbf{z}^i)| \leq 3, \quad i = 1, \dots, m \\ & \tilde{G}_{t+1}(\mathbf{u}^+ - \mathbf{u}^-, \mathbf{e}_j) = 0, \quad j = 1, \dots, 4 \\ & \mathbf{C}_t^\top \mathbf{u}^+ \leq 100, \quad \mathbf{C}_t^\top \mathbf{u}^- + \mathbf{C}_t^\top \mathbf{u}^+ \leq 500. \end{aligned} \quad (\text{EC.18})$$

The objective in (EC.18) serves as a selection criterion within the split-position LP representation, favoring a large total notional allocation subject to the scenario, directional-gain, and capital constraints. Since $\mathbf{u}^+$ and $\mathbf{u}^-$ are auxiliary variables and need not coincide with the positive and negative parts of the net position $\mathbf{u} = \mathbf{u}^+ - \mathbf{u}^-$, this objective should not be interpreted as the gross exposure computed from the absolute net holdings.

EC.3 Proofs

EC.3.1 Proof of Theorem 5.1

We include the proof of Theorem 5.1 here for the sake of completion. As mentioned in Section 5, the theorem and proof are due to Gibbs and Candes (2021). Throughout the proof, we use the theoretical prediction sets defined in (12). We need the following lemma.

Lemma EC.3.1. *Let $\alpha \in (0, 1)$, $\gamma > 0$, and $\tilde{\alpha}_1 = \alpha$. With the prediction sets in (12) and finite realized gains, the ACSA recursion satisfies $\tilde{\alpha}_t \in [-\gamma, 1 + \gamma]$ for every integer $t \geq 1$.*

Proof. The claim holds at $t = 1$. Suppose it holds at time t . The update is

$$\tilde{\alpha}_{t+1} = \tilde{\alpha}_t + \gamma(\alpha - \text{err}_{t+1}). \quad (\text{EC.19})$$

If $\tilde{\alpha}_t \leq 0$, the prediction set is $\mathbb{R}$, so $\text{err}_{t+1} = 0$ and $\tilde{\alpha}_{t+1} = \tilde{\alpha}_t + \gamma\alpha \in [-\gamma, 1 + \gamma]$. If $\tilde{\alpha}_t \geq 1$, the prediction set is $\emptyset$, so $\text{err}_{t+1} = 1$ and $\tilde{\alpha}_{t+1} = \tilde{\alpha}_t - \gamma(1 - \alpha) \in [-\gamma, 1 + \gamma]$. Finally, if $0 < \tilde{\alpha}_t < 1$, either value of the indicator gives $-\gamma < \tilde{\alpha}_{t+1} < 1 + \gamma$. Induction proves the result. $\square$

Theorem 5.1 now follows easily from Lemma EC.3.1.

Proof of Theorem 5.1. After observing G_{T+1} , define err_{T+1} and $\tilde{\alpha}_{T+1}$ by the same recursion. The update rule (EC.19) of $\tilde{\alpha}_t$, together with Lemma EC.3.1, implies

$$\begin{aligned} \tilde{\alpha}_{T+1} &= \tilde{\alpha}_T + \gamma \cdot (\alpha - \text{err}_{T+1}) = \tilde{\alpha}_{T-1} + \gamma \cdot (\alpha - \text{err}_{T+1}) + \gamma \cdot (\alpha - \text{err}_T) = \dots \\ &= \alpha + \gamma \cdot \sum_{t=2}^{T+1} (\alpha - \text{err}_t) \in [-\gamma, 1 + \gamma]. \end{aligned} \quad (\text{EC.20})$$

Rearranging yields $\left| \frac{1}{T} \sum_{t=2}^{T+1} \text{err}_t - \alpha \right| \leq \frac{\gamma + \max\{\alpha, 1 - \alpha\}}{T\gamma}$. Since $\text{err}_{t+1} = 1 - \mathbb{1}\{G_{t+1} \in \mathcal{C}_t(\mathbf{z}_{t+1})\}$, this is equivalent to the stated coverage bound. $\square$

EC.3.2 Proofs for Section 6.2

To simplify the notation, define $d := d_{\mathbf{W}}, K(\mathbf{v}) := \tilde{\kappa}(\|\mathbf{v}\|)$, $\bar{K} := \sup_{\mathbf{v} \in \mathbb{R}^d} K(\mathbf{v})$. For $\mathbf{w} \in \mathcal{D}$ and $y \in \mathbb{R}$, define $A_{t,h}(y, \mathbf{w}) := \frac{1}{t} \sum_{s=1}^t \kappa_h(\mathbf{W}_s, \mathbf{w}) \mathbb{1}\{Y_s \leq y\}$, $A_h(y, \mathbf{w}) := \mathbb{E}[\kappa_h(\mathbf{W}_1, \mathbf{w}) \mathbb{1}\{Y_1 \leq y\}]$,

$B_{t,h}(\mathbf{w}) := \frac{1}{t} \sum_{s=1}^t \kappa_h(\mathbf{W}_s, \mathbf{w})$, $B_h(\mathbf{w}) := \mathbb{E}[\kappa_h(\mathbf{W}_1, \mathbf{w})]$, and $F_h(y \mid \mathbf{w}) := \frac{A_h(y, \mathbf{w})}{B_h(\mathbf{w})}$. Whenever $B_{t,h}(\mathbf{w}) > 0$, we have $\hat{F}_t(y \mid \mathbf{w}) = \frac{A_{t,h}(y, \mathbf{w})}{B_{t,h}(\mathbf{w})}$. For $B_{t,h}(\mathbf{w}) > 0$ and $u \in (0, 1)$, define the u -quantile of $\hat{F}_t(y \mid \mathbf{w})$ as

$$\hat{Q}_t(u; \mathbf{w}) := \inf\{y \in \mathbb{R} : \hat{F}_t(y \mid \mathbf{w}) \geq u\}. \quad (\text{EC.21})$$

And if $B_{t,h}(\mathbf{w}) = 0$, set $\hat{F}_t(y \mid \mathbf{w}) = \mathbb{1}\{y \geq 0\}$ and $\hat{Q}_t(u; \mathbf{w}) = 0$.

The primary goal of this section is to show that $\hat{F}_t(y \mid \mathbf{w})$ accurately estimates $F_{Y|\mathbf{W}}(y \mid \mathbf{w})$ for all $\mathbf{w} \in \mathcal{D}$ and $y \in \mathbb{R}$. The following decomposition provides the basic structure for the proofs.

$$\sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} \left| \hat{F}_t(y \mid \mathbf{w}) - F_{Y|\mathbf{W}}(y \mid \mathbf{w}) \right| \leq \sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} \left| \hat{F}_t(y \mid \mathbf{w}) - F_h(y \mid \mathbf{w}) \right| + \sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} \left| F_h(y \mid \mathbf{w}) - F_{Y|\mathbf{W}}(y \mid \mathbf{w}) \right|. \quad (\text{EC.22})$$

The first term on the right-hand side is controlled by an empirical-process argument, and the second by a direct kernel-bias calculation. A key difficulty is the dependence among observations. The β -mixing condition allows sufficiently separated blocks of observations to be treated as “almost” independent for concentration purposes. The following lemma formalizes this comparison using a *blocking argument* based on the idea of [Yu \(1994\)](#).

Lemma EC.3.2. *Let $\{X_s\}_{s \geq 1}$ be a β -mixing process taking values in a standard Borel space. For $1 \leq j \leq m$, let $V_j = (X_{a_j}, \dots, X_{b_j})$ for $a_j \leq b_j < a_{j+1}$, and let $V_1^*, \dots, V_m^*$ be independent, with V_j^* distributed as V_j . If ψ is measurable and takes values in $[0, M]$, then*

$$|\mathbb{E}\psi(V_1, \dots, V_m) - \mathbb{E}\psi(V_1^*, \dots, V_m^*)| \leq M \sum_{j=2}^m \beta(a_j - b_{j-1}).$$

In particular, if $a_j - b_{j-1} \geq q$ for every $j \geq 2$, the right-hand side is at most $M(m-1)\beta(q)$.

Proof. For probability measures μ and ν , use the convention $\|\mu - \nu\|_{\text{TV}} = \sup_{0 \leq g \leq 1} \left| \int g d\mu - \int g d\nu \right|$. For $j \geq 2$, set $\mathcal{G}_{j-1} = \sigma(\{V_1, \dots, V_{j-1}\})$, $\mathcal{P}_{j-1} = \sigma(\{X_s : s \leq b_{j-1}\})$ and $\mathcal{A}_j = \sigma(V_j)$. Then $\mathcal{G}_{j-1} \subseteq \mathcal{P}_{j-1}$ and $\mathcal{A}_j \subseteq \sigma(\{X_s : s \geq a_j\})$. For $B \in \mathcal{A}_j$, the tower property gives $\mathbb{P}(B \mid \mathcal{G}_{j-1}) - \mathbb{P}(B) = \mathbb{E}[\mathbb{P}(B \mid \mathcal{P}_{j-1}) - \mathbb{P}(B) \mid \mathcal{G}_{j-1}]$. Consequently,

$$\mathbb{E} \left[\sup_{B \in \mathcal{A}_j} |\mathbb{P}(B \mid \mathcal{G}_{j-1}) - \mathbb{P}(B)| \right] \leq \mathbb{E} \left[\sup_{B \in \sigma(\{X_s : s \geq a_j\})} |\mathbb{P}(B \mid \mathcal{P}_{j-1}) - \mathbb{P}(B)| \right] \leq \beta(a_j - b_{j-1}).$$

Because the state space is standard Borel, the left-hand side equals the total-variation distance between the joint law of $(V_1, \dots, V_j)$ and the *product of the law* (denoted by $\otimes$) of $(V_1, \dots, V_{j-1})$ and the marginal law of V_j . Thus $\|\mathcal{L}(V_1, \dots, V_j) - \mathcal{L}(V_1, \dots, V_{j-1}) \otimes \mathcal{L}(V_j)\|_{\text{TV}} \leq \beta(a_j - b_{j-1})$. For $1 \leq j \leq m$, define the hybrid law $\nu_j = \mathcal{L}(V_1, \dots, V_j) \otimes \bigotimes_{\ell=j+1}^m \mathcal{L}(V_\ell)$. Then ν_1 is the product law of $(V_1^*, \dots, V_m^*)$, whereas ν_m is the joint law of $(V_1, \dots, V_m)$. Tensoring two probability measures with a common probability measure does not increase their total variation distance. Therefore, $\|\nu_j - \nu_{j-1}\|_{\text{TV}} \leq \beta(a_j - b_{j-1})$ for $2 \leq j \leq m$. Applying the triangle inequality to the sequence

$\nu_1, \dots, \nu_m$ yields

$$\|\nu_m - \nu_1\|_{\text{TV}} \leq \sum_{j=2}^m \|\nu_j - \nu_{j-1}\|_{\text{TV}} \leq \sum_{j=2}^m \beta(a_j - b_{j-1}).$$

Since $0 \leq \psi/M \leq 1$, the definition of total variation gives

$$\begin{aligned} |\mathbb{E}\psi(V_1, \dots, V_m) - \mathbb{E}\psi(V_1^*, \dots, V_m^*)| &= \left| \int \psi \, d\nu_m - \int \psi \, d\nu_1 \right| = M \left| \int \frac{\psi}{M} \, d\nu_m - \int \frac{\psi}{M} \, d\nu_1 \right| \\ &\leq M \|\nu_m - \nu_1\|_{\text{TV}} \leq M \sum_{j=2}^m \beta(a_j - b_{j-1}). \end{aligned}$$

Finally, if $a_j - b_{j-1} \geq q$ for every $j \geq 2$, the monotonicity of the mixing coefficients implies $\sum_{j=2}^m \beta(a_j - b_{j-1}) \leq (m-1)\beta(q)$, which proves the final assertion. $\square$

Lemma EC.3.2 provides the bridge from an empirical-process bound for independent observations to the dependent sample. For any statistic bounded by M , replacing m dependent blocks by independent copies with the same marginal distributions changes its expectation by at most $M \sum_{j=2}^m \beta(a_j - b_{j-1})$. Thus, the effect of temporal dependence is represented by an explicit additive error that decreases as the separation between the blocks increases.

We now establish a uniform concentration bound between $A_h(y, \mathbf{w})$, $B_h(\mathbf{w})$ and their empirical estimates $A_{t,h}(y, \mathbf{w})$, $B_{t,h}(\mathbf{w})$. The proof first derives a maximal inequality for independent samples and then transfers it to the original dependent sample using Lemma EC.3.2. The independent-sample argument is an empirical-process bound over a class of kernel superlevel sets and threshold events. To state this argument compactly, we introduce the VC and covering-number notation that will be used below.

We use standard notation from empirical process theory. For a class $\mathcal{F}$ of measurable functions and a probability measure P , let $N(\epsilon, \mathcal{F}, L_2(P))$ denote the ϵ -covering number of $\mathcal{F}$ under the $L_2(P)$ semimetric. We write $\text{VCdim}(\mathcal{A})$ for the VC dimension of a class $\mathcal{A}$ of measurable sets. Standard background on covering numbers and VC classes may be found, for example, in [van der Vaart and Wellner \(1996\)](#) and [Mohri et al. \(2018\)](#). Let $\mathcal{B}_d$ denote the class consisting of the empty set and all open or closed Euclidean balls in $\mathbb{R}^d$, and define $\mathcal{E}_d = \{B \times (-\infty, y] : B \in \mathcal{B}_d, y \in \mathbb{R} \cup \{\infty\}\}$. The proof leverages the standard facts that $\text{VCdim}(\mathcal{B}_d) = d+1$ and that, for any class $\mathcal{A}$ with $\text{VCdim}(\mathcal{A}) \leq v$ and any probability measure P ,

$$N(\epsilon, \{\mathbb{1}_A : A \in \mathcal{A}\}, L_2(P)) \leq e(v+1) \left(\frac{2e}{\epsilon^2} \right)^v, \quad 0 < \epsilon < 1. \quad (\text{EC.23})$$

The latter can be easily derived from [Haussler \(1995\)](#). We are now ready to state the lemma.

Lemma EC.3.3. *Suppose parts (a) and (e) of Assumption 2 hold. For any $\delta \in (0, 1)$, define $D_\delta := 6\sqrt{2 \{ \log(18e^2(d+3)^2) + 2(d+2) \log(32e) \}}$, $T_\rho := \max \left\{ 3, \left\lceil \left(1 + \frac{2}{\log(1/\rho)} \right)^2 \right\rceil \right\}$, and $R_\delta :=$*

$\frac{\bar{K}}{\delta} \left[D_d \sqrt{1 + \frac{2}{\log(1/\rho)}} + \beta_0 \right]$. Then, for every fixed $h > 0$ and every integer $t \geq T_\rho$,

$$\mathbb{P} \left\{ \sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |A_{t,h}(y, \mathbf{w}) - A_h(y, \mathbf{w})| \vee \sup_{\mathbf{w} \in \mathcal{D}} |B_{t,h}(\mathbf{w}) - B_h(\mathbf{w})| \leq R_\delta \sqrt{\frac{\log t}{t}} \right\} \geq 1 - \delta. \quad (\text{EC.24})$$

Proof. The two quantities in (EC.24) are kernel-weighted empirical processes indexed by the kernel center $\mathbf{w}$ and, for the numerator, by the threshold y . The proof proceeds in four steps. First, we identify a VC class that contains the indicator functions associated with the kernel superlevel sets. Second, we use its covering numbers to derive a maximal inequality for the corresponding indicator process under independent sampling. Third, the layer-cake representation transfers this bound to the kernel-weighted processes. Finally, we partition the dependent observations into well-separated subsequences and apply Lemma EC.3.2 to transfer the independent-sample bound to the original dependent sample.

Step 1: Let $\mathcal{B}_d$ denote the class consisting of the empty set and all open or closed Euclidean balls in $\mathbb{R}^d$, and define $\mathcal{E}_d = \{B \times (-\infty, y] : B \in \mathcal{B}_d, y \in \mathbb{R} \cup \{\infty\}\}$. The relevance of this class to the kernel representation follows from radial monotonicity. Specifically, for every $\mathbf{w} \in \mathcal{D}$ and $s \in (0, \bar{K})$, the superlevel set $\{\mathbf{w}' : \kappa_h(\mathbf{w}', \mathbf{w}) > s\}$ is either empty or an open or closed Euclidean ball. Hence, for every $y \in \mathbb{R} \cup \{\infty\}$, $\{(\mathbf{w}', y') : \kappa_h(\mathbf{w}', \mathbf{w}) > s, y' \leq y\} \in \mathcal{E}_d$. Setting $y = \infty$ also yields the indicators required for the denominator. Therefore, a uniform empirical-process bound over $\mathcal{E}_d$ simultaneously controls the kernel-weighted averages appearing in both the numerator and the denominator.

We next show that $\text{VCdim}(\mathcal{E}_d) \leq d + 2$. Since $\text{VCdim}(\mathcal{B}_d) = d + 1$, it suffices to show that imposing the additional lower-half-line restriction can increase the VC dimension by at most one. Suppose that the points $x_i = (\mathbf{w}_i, y_i)$, $1 \leq i \leq n$, are shattered by $\mathcal{E}_d$, and choose i^* such that $y_{i^*} = \max_{1 \leq i \leq n} y_i$. First, the feature vectors $\mathbf{w}_1, \dots, \mathbf{w}_n$ must be distinct. Indeed, if $\mathbf{w}_i = \mathbf{w}_j$ and $y_i \leq y_j$, then any set of the form $B \times (-\infty, c]$ that contains x_j must also contain x_i , contradicting the assumption that the points are shattered. Next, fix any $I \subseteq \{1, \dots, n\} \setminus \{i^*\}$. Since the points are shattered by $\mathcal{E}_d$, there exist $B_I \in \mathcal{B}_d$ and $c_I \in \mathbb{R} \cup \{\infty\}$ such that the trace of $B_I \times (-\infty, c_I]$ on $\{x_1, \dots, x_n\}$ is $\{x_{i^*}\} \cup \{x_i : i \in I\}$. Because this trace contains x_{i^*} , we must have $c_I \geq y_{i^*}$. By the choice of i^* , this implies $c_I \geq y_i$ for every i . Hence the lower-half-line restriction no longer excludes any of the points, and we must have $B_I \cap \{\mathbf{w}_i : i \neq i^*\} = \{\mathbf{w}_i : i \in I\}$. Since this holds for every $I \subseteq \{1, \dots, n\} \setminus \{i^*\}$, the class $\mathcal{B}_d$ shatters the $n - 1$ feature vectors $\{\mathbf{w}_i : i \neq i^*\}$. Therefore $n - 1 \leq \text{VCdim}(\mathcal{B}_d) = d + 1$, and hence $n \leq d + 2$, which proves the claim. Applying (EC.23) with $v = d + 2$ gives

$$N(\eta, \{\mathbb{1}_E : E \in \mathcal{E}_d\}, L_2(P_0)) \leq e(d + 3) \left(\frac{2e}{\eta^2} \right)^{d+2} \quad (\text{EC.25})$$

for every probability measure P_0 on $\mathbb{R}^d \times \mathbb{R}$ and every $0 < \eta < 1$,

Step 2: Let P denote the common marginal law of $(\mathbf{W}_s, Y_s)$, which does not depend on s by

stationarity. Fix an integer $m \geq 1$, let $X_i^* = (\mathbf{W}_i^*, Y_i^*)$, $1 \leq i \leq m$ be independent observations with law P , and write $P_m = \frac{1}{m} \sum_{i=1}^m \delta_{X_i^*}$ for their empirical measure. Let $\xi_1, \dots, \xi_m$ be Rademacher variables independent of the observations, and let $\mathbb{E}_\xi$ denote conditional expectation with respect to these variables given $X_1^*, \dots, X_m^*$.

The goal of this step is to establish the independent-sample maximal inequality that holds over $(\mathbf{w}, y) \in \mathcal{D} \times \mathbb{R}$. In view of the indicator representation identified in the previous step, we first prove

$$\mathbb{E} \left[\sup_{E \in \mathcal{E}_d} \left| \frac{1}{m} \sum_{i=1}^m \{\mathbb{1}\{X_i^* \in E\} - \mathbb{P}(X_1^* \in E)\} \right| \right] \leq \frac{D_d}{\sqrt{m}}. \quad (\text{EC.26})$$

We obtain this bound by first controlling the associated Rademacher process through chaining and then applying symmetrization. Start with a finite-class increment bound that will be applied at each level of the chaining argument. Fix $\lambda > 0$ and take $f, g \in \{\pm \mathbb{1}_E : E \in \mathcal{E}_d\} \cup \{0\}$. For every $u \in \mathbb{R}$, $\mathbb{E}_\xi [e^{u\xi_i}] = \frac{e^u + e^{-u}}{2} =: \cosh(u)$. Since $\cosh(u) \leq e^{u^2/2}$, we obtain

$$\begin{aligned} \mathbb{E}_\xi \exp \left\{ \frac{\lambda}{m} \sum_{i=1}^m \xi_i (f(X_i^*) - g(X_i^*)) \right\} &= \prod_{i=1}^m \cosh \left(\frac{\lambda \{f(X_i^*) - g(X_i^*)\}}{m} \right) \\ &\leq \exp \left\{ \frac{\lambda^2}{2m^2} \sum_{i=1}^m \{f(X_i^*) - g(X_i^*)\}^2 \right\} = \exp \left\{ \frac{\lambda^2 \|f - g\|_{P_m, 2}^2}{2m} \right\}. \end{aligned}$$

Now consider pairs (f_ℓ, g_ℓ) , $1 \leq \ell \leq J$, from the signed-indicator class, that satisfy $\|f_\ell - g_\ell\|_{P_m, 2} \leq r_0$. Applying the preceding bound to each of the J pairs, summing the resulting inequalities, and then using Jensen's inequality gives:

$$\begin{aligned} \mathbb{E}_\xi \max_{1 \leq \ell \leq J} \left| \frac{1}{m} \sum_{i=1}^m \xi_i \{f_\ell(X_i^*) - g_\ell(X_i^*)\} \right| &\leq \frac{1}{\lambda} \log \mathbb{E}_\xi \exp \left\{ \lambda \max_{1 \leq \ell \leq J} \left| \frac{1}{m} \sum_{i=1}^m \xi_i \{f_\ell(X_i^*) - g_\ell(X_i^*)\} \right| \right\} \\ &\leq \frac{\log(2J)}{\lambda} + \frac{\lambda r_0^2}{2m}. \end{aligned}$$

The right-hand side is minimized by $\lambda = \sqrt{2m \log(2J)}/r_0$, which gives

$$\mathbb{E}_\xi \max_{1 \leq \ell \leq J} \left| \frac{1}{m} \sum_{i=1}^m \xi_i \{f_\ell(X_i^*) - g_\ell(X_i^*)\} \right| \leq \frac{r_0}{\sqrt{m}} \sqrt{2 \log(2J)}.$$

We now extend the finite-class increment bound to the full indicator class by chaining through successively finer covers. Conditional on $X_1^*, \dots, X_m^*$, write $Z_m(f) = \frac{1}{m} \sum_{i=1}^m \xi_i f(X_i^*)$ and let $\mathcal{F}_d = \{\pm \mathbb{1}_E : E \in \mathcal{E}_d\} \cup \{0\}$. The target supremum can be written as $\sup_{E \in \mathcal{E}_d} |Z_m(\mathbb{1}_E)| = \sup_{f \in \mathcal{F}_d} |Z_m(f)|$. For every integer $j \geq 1$, apply (EC.25) with $P_0 = P_m$ and $\eta = 2^{-j}$. The positive-indicator class therefore admits a 2^{-j} -net with at most $N_j = e(d+3)(2e4^j)^{d+2}$ elements. Negating the elements of this net gives a net of the same size for the negative indicators. After also including the zero function, we obtain a 2^{-j} -net $\mathcal{F}_j$ for $\mathcal{F}_d$ satisfying $|\mathcal{F}_j| \leq 2N_j + 1 \leq 3N_j$. Set $\mathcal{F}_0 = \{0\}$. For every

$f \in \mathcal{F}_d$, choose $f_j \in \mathcal{F}_j$ such that $\|f - f_j\|_{P_{m,2}} \leq 2^{-j}$, and set $f_0 = 0$. Since $\|f - f_0\|_{P_{m,2}} \leq 1$, the triangle inequality gives, for every $j \geq 1$,

$$\|f_j - f_{j-1}\|_{P_{m,2}} \leq \|f_j - f\|_{P_{m,2}} + \|f - f_{j-1}\|_{P_{m,2}} \leq 2^{-j} + 2^{-(j-1)} = 3 \cdot 2^{-j}.$$

At level j , let $\mathcal{P}_j = \{(f_j, f_{j-1}) : f \in \mathcal{F}_d\}$ denote the collection of successive approximation pairs that can arise. For $j \geq 2$, $|\mathcal{P}_j| \leq |\mathcal{F}_j| |\mathcal{F}_{j-1}| \leq 9N_j N_{j-1} \leq 9N_j^2$. For $j = 1$, since $\mathcal{F}_0 = \{0\}$, $|\mathcal{P}_1| \leq |\mathcal{F}_1| \leq 3N_1 \leq 9N_1^2$. Thus, $|\mathcal{P}_j| \leq 9N_j^2$ for every $j \geq 1$. By the Cauchy-Schwarz inequality,

$$\begin{aligned} |Z_m(f) - Z_m(f_j)| &= \left| \frac{1}{m} \sum_{i=1}^m \xi_i \{f(X_i^*) - f_j(X_i^*)\} \right| \\ &\leq \left(\frac{1}{m} \sum_{i=1}^m \xi_i^2 \right)^{1/2} \|f - f_j\|_{P_{m,2}} \leq 2^{-j} \rightarrow 0. \end{aligned}$$

It follows that $Z_m(f) = \sum_{j=1}^{\infty} \{Z_m(f_j) - Z_m(f_{j-1})\}$. Taking absolute values and then the supremum over $f \in \mathcal{F}_d$ gives the key chaining inequality

$$\sup_{f \in \mathcal{F}_d} |Z_m(f)| \leq \sum_{j=1}^{\infty} \max_{(u,v) \in \mathcal{P}_j} |Z_m(u) - Z_m(v)|.$$

Every pair in $\mathcal{P}_j$ has $L_2(P_m)$ distance at most $3 \cdot 2^{-j}$, and $|\mathcal{P}_j| \leq 9N_j^2$. The preceding finite-class inequality therefore gives

$$\mathbb{E}_{\xi} \max_{(u,v) \in \mathcal{P}_j} |Z_m(u) - Z_m(v)| \leq \frac{3 \cdot 2^{-j}}{\sqrt{m}} \sqrt{2 \log(2|\mathcal{P}_j|)} \leq \frac{3\sqrt{2}}{\sqrt{m}} 2^{-j} \sqrt{\log(18N_j^2)}.$$

Summing this bound over j yields

$$\mathbb{E}_{\xi} \sup_{E \in \mathcal{E}_d} \left| \frac{1}{m} \sum_{i=1}^m \xi_i \mathbb{1}\{X_i^* \in E\} \right| \leq \frac{3\sqrt{2}}{\sqrt{m}} \sum_{j=1}^{\infty} 2^{-j} \sqrt{\log(18N_j^2)}.$$

From the definition of N_j , $\log(18N_j^2) = \log(18e^2(d+3)^2) + 2(d+2)\log(2e) + 4j(d+2)\log 2$. Since $\sum_{j=1}^{\infty} 2^{-j} = 1$ and $\sum_{j=1}^{\infty} j2^{-j} = 2$, the concavity of the square-root function gives

$$\sum_{j=1}^{\infty} 2^{-j} \sqrt{\log(18N_j^2)} \leq \sqrt{\sum_{j=1}^{\infty} 2^{-j} \log(18N_j^2)} = \sqrt{\log(18e^2(d+3)^2) + 2(d+2)\log(32e)}.$$

Recall the definition of D_d , thus

$$\mathbb{E}_{\xi} \sup_{E \in \mathcal{E}_d} \left| \frac{1}{m} \sum_{i=1}^m \xi_i \mathbb{1}\{X_i^* \in E\} \right| \leq \frac{D_d}{2\sqrt{m}}. \quad (\text{EC.27})$$

We now convert inequality (EC.27) into the required centered empirical-process bound by symmetrization. Let $X'_1, \dots, X'_m$ be an independent copy of $X_1^*, \dots, X_m^*$. For $E \in \mathcal{E}_d$, write $P(E) = \mathbb{P}(X_1^* \in E)$. Conditional Jensen's inequality replaces $P(E)$ by the empirical average over the copied sample. The exchangeability of each pair (X_i^*, X'_i) then permits the introduction of the Rademacher signs. Finally, the triangle inequality and the identical distributions of the two samples give

$$\begin{aligned} \mathbb{E} \sup_{E \in \mathcal{E}_d} \left| \frac{1}{m} \sum_{i=1}^m \{\mathbb{1}_E(X_i^*) - P(E)\} \right| &\leq \mathbb{E} \sup_{E \in \mathcal{E}_d} \left| \frac{1}{m} \sum_{i=1}^m \{\mathbb{1}_E(X_i^*) - \mathbb{1}_E(X'_i)\} \right| = \mathbb{E} \sup_{E \in \mathcal{E}_d} \left| \frac{1}{m} \sum_{i=1}^m \xi_i \{\mathbb{1}_E(X_i^*) - \mathbb{1}_E(X'_i)\} \right| \\ &\leq 2 \mathbb{E} \sup_{E \in \mathcal{E}_d} \left| \frac{1}{m} \sum_{i=1}^m \xi_i \mathbb{1}_E(X_i^*) \right| \leq \frac{D_d}{\sqrt{m}}. \end{aligned}$$

This proves the desired inequality (EC.26).

Step 3: We transfer the bound (EC.26) from indicators to kernel-weighted averages. For every $(\mathbf{w}', y') \in \mathbb{R}^d \times \mathbb{R}$, the “layer-cake” identities give $\kappa_h(\mathbf{w}', \mathbf{w}) = \int_0^{\bar{K}} \mathbb{1}\{\kappa_h(\mathbf{w}', \mathbf{w}) > s\} ds$ and $\kappa_h(\mathbf{w}', \mathbf{w}) \mathbb{1}\{y' \leq y\} = \int_0^{\bar{K}} \mathbb{1}\{\kappa_h(\mathbf{w}', \mathbf{w}) > s, y' \leq y\} ds$. Since every indicator appearing under these integrals belongs to $\mathcal{E}_d$, and $\left| \int_0^{\bar{K}} g(s) ds \right| \leq \int_0^{\bar{K}} |g(s)| ds$, we may apply the indicator bound (EC.26) at each level s and then integrate over $(0, \bar{K})$. This gives

$$\begin{aligned} \mathbb{E} \left[\sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} \left| \frac{1}{m} \sum_{i=1}^m \{\kappa_h(\mathbf{W}_i^*, \mathbf{w}) \mathbb{1}\{Y_i^* \leq y\} - \mathbb{E}[\kappa_h(\mathbf{W}_1, \mathbf{w}) \mathbb{1}\{Y_1 \leq y\}]\} \right| \right. \\ \left. \vee \sup_{\mathbf{w} \in \mathcal{D}} \left| \frac{1}{m} \sum_{i=1}^m \{\kappa_h(\mathbf{W}_i^*, \mathbf{w}) - \mathbb{E}\kappa_h(\mathbf{W}_1, \mathbf{w})\} \right| \right] \leq \frac{\bar{K} D_d}{\sqrt{m}}. \end{aligned} \tag{EC.28}$$

Step 4: The final step is to transfer the independent-sample bound in (EC.28) to the original β -mixing observations and thereby prove (EC.24). Let Δ_t denote the maximum of the two empirical deviations appearing in the lemma:

$$\begin{aligned} \Delta_t &= \sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} \left| \frac{1}{t} \sum_{s=1}^t \{\kappa_h(\mathbf{W}_s, \mathbf{w}) \mathbb{1}\{Y_s \leq y\} - \mathbb{E}[\kappa_h(\mathbf{W}_1, \mathbf{w}) \mathbb{1}\{Y_1 \leq y\}]\} \right| \\ &\quad \vee \sup_{\mathbf{w} \in \mathcal{D}} \left| \frac{1}{t} \sum_{s=1}^t \{\kappa_h(\mathbf{W}_s, \mathbf{w}) - \mathbb{E}\kappa_h(\mathbf{W}_1, \mathbf{w})\} \right| \\ &= \sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |A_{t,h}(y, \mathbf{w}) - A_h(y, \mathbf{w})| \vee \sup_{\mathbf{w} \in \mathcal{D}} |B_{t,h}(\mathbf{w}) - B_h(\mathbf{w})|. \end{aligned}$$

It therefore suffices to show that $\mathbb{E}\Delta_t \leq \delta R_\delta \sqrt{\frac{\log t}{t}}$ because the required probability bound will then follow immediately from Markov's inequality. The bound in (EC.28) cannot be applied directly, because the observations are dependent.

We first derive a comparison bound for an arbitrary separation length. Fix any integer $q \in$

$\{1, \dots, t\}$ and partition $\{1, \dots, t\}$ into the residue classes

$$I_\ell = \{\ell, \ell + q, \ell + 2q, \dots\} \cap \{1, \dots, t\}, \quad n_\ell = |I_\ell| = \left\lfloor \frac{t - \ell}{q} \right\rfloor + 1, \quad 1 \leq \ell \leq q.$$

These sets form a partition of $\{1, \dots, t\}$, and consecutive observations within each I_ℓ are separated by q time points. For each $\ell \in \{1, \dots, q\}$, define the corresponding subsequence deviation by

$$\Delta_\ell := \sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} \left| \frac{1}{n_\ell} \sum_{s \in I_\ell} \{\kappa_h(\mathbf{W}_s, \mathbf{w}) \mathbb{1}\{Y_s \leq y\} - \mathbb{E}[\kappa_h(\mathbf{W}_1, \mathbf{w}) \mathbb{1}\{Y_1 \leq y\}]\} \right| \\ \vee \sup_{\mathbf{w} \in \mathcal{D}} \left| \frac{1}{n_\ell} \sum_{s \in I_\ell} \{\kappa_h(\mathbf{W}_s, \mathbf{w}) - \mathbb{E}[\kappa_h(\mathbf{W}_1, \mathbf{w})]\} \right|.$$

Because the sets $I_1, \dots, I_q$ partition $\{1, \dots, t\}$ and $\sum_{\ell=1}^q n_\ell = t$, each centered full-sample average is the weighted sum of its corresponding subsequence averages, with weights n_ℓ/t . Applying the triangle inequality and then taking the corresponding suprema shows that each of the two terms defining Δ_t is bounded by $\sum_{\ell=1}^q (n_\ell/t) \Delta_\ell$. Therefore, $\Delta_t \leq \sum_{\ell=1}^q \frac{n_\ell}{t} \Delta_\ell$.

We next compare each separated subsequence with an independent sample. For a fixed ℓ , let $(\mathbf{W}_{\ell,1}^*, Y_{\ell,1}^*), \dots, (\mathbf{W}_{\ell,n_\ell}^*, Y_{\ell,n_\ell}^*)$ be independent observations with the common marginal law P , and let Δ_ℓ^* be obtained from Δ_ℓ by replacing the observations indexed by I_ℓ with these independent copies. Applying (EC.28) with $m = n_\ell$ gives $\mathbb{E}\Delta_\ell^* \leq \frac{\bar{K}D_d}{\sqrt{n_\ell}}$. Both Δ_ℓ and Δ_ℓ^* take values in $[0, \bar{K}]$, since they are maxima of absolute differences between averages of variables in $[0, \bar{K}]$ and their expectations. Moreover, consecutive observations in I_ℓ are separated by q . Lemma EC.3.2, applied to the corresponding singleton blocks, therefore yields

$$\mathbb{E}\Delta_\ell \leq \mathbb{E}\Delta_\ell^* + \bar{K}(n_\ell - 1)\beta(q) \leq \frac{\bar{K}D_d}{\sqrt{n_\ell}} + \bar{K}(n_\ell - 1)\beta(q).$$

We now aggregate these bounds over the residue classes. Since $\sum_{\ell=1}^q n_\ell = t$, Cauchy–Schwarz gives $\sum_{\ell=1}^q \sqrt{n_\ell} \leq \sqrt{q \sum_{\ell=1}^q n_\ell} = \sqrt{qt}$. Also, because the indices in each residue class are separated by q , $n_\ell - 1 \leq \frac{t}{q}$. Consequently,

$$\mathbb{E}\Delta_t \leq \sum_{\ell=1}^q \frac{n_\ell}{t} \mathbb{E}\Delta_\ell \leq \frac{\bar{K}D_d}{t} \sum_{\ell=1}^q \sqrt{n_\ell} + \frac{\bar{K}\beta(q)}{t} \sum_{\ell=1}^q n_\ell(n_\ell - 1) \leq \bar{K}D_d \sqrt{\frac{q}{t}} + \bar{K}\beta(q) \frac{t}{q}.$$

This inequality displays the role of the separation length q . Increasing q reduces the mixing term through $\beta(q)$, but also decreases the effective size of each subsequence and therefore enlarges the independent-sample term. Under geometric mixing, a logarithmic separation is sufficient to control these two effects.

We now make this choice precise. Set $q_t = \left\lceil \frac{2 \log t}{\log(1/\rho)} \right\rceil$. Before substituting $q = q_t$, we verify that

this is an admissible separation length. Since $t \geq \left(1 + \frac{2}{\log(1/\rho)}\right)^2$, we have $1 \leq \log t \leq \sqrt{t}$ and

$$q_t \leq \frac{2 \log t}{\log(1/\rho)} + 1 \leq \left(1 + \frac{2}{\log(1/\rho)}\right) \log t \leq \left(1 + \frac{2}{\log(1/\rho)}\right) \sqrt{t} \leq t.$$

Thus $q_t \in \{1, \dots, t\}$. Moreover, geometric mixing gives $\beta(q_t) \leq \beta_0 \rho^{q_t} \leq \beta_0 t^{-2}$. Substituting $q = q_t$ into the preceding general bound therefore yields

$$\begin{aligned} \mathbb{E} \Delta_t &\leq \bar{K} D_d \sqrt{\frac{q_t}{t}} + \bar{K} \beta(q_t) \frac{t}{q_t} \leq \bar{K} D_d \sqrt{\left(1 + \frac{2}{\log(1/\rho)}\right) \frac{\log t}{t}} + \frac{\bar{K} \beta_0}{q_t t} \\ &\leq \bar{K} \left[D_d \sqrt{1 + \frac{2}{\log(1/\rho)}} + \beta_0 \right] \sqrt{\frac{\log t}{t}} = \delta R_\delta \sqrt{\frac{\log t}{t}}. \end{aligned}$$

The final inequality uses $q_t \geq 1$ and $\frac{1}{t} \leq \sqrt{\frac{\log t}{t}}$ for $t \geq 3$. Finally, Markov's inequality gives $\mathbb{P} \left\{ \Delta_t > R_\delta \sqrt{\frac{\log t}{t}} \right\} \leq \frac{\mathbb{E} \Delta_t}{R_\delta \sqrt{\log t/t}} \leq \delta$. Since Δ_t is exactly the maximum of the two deviations in (EC.24), this proves the result. $\square$

We are now ready to give a more explicit restatement of Theorem 6.1 and provide its proof. First we introduce the new notations needed. Let $\mu_0 := \int_{\mathbb{R}^d} K(\mathbf{v}) d\mathbf{v}$, $\mu_2 := \int_{\mathbb{R}^d} \|\mathbf{v}\|^2 K(\mathbf{v}) d\mathbf{v}$, and $v_d := \frac{\pi^{d/2}}{\Gamma(1+d/2)}$. Assumption 2(e) gives $0 < \mu_0 < \infty$ and $0 < \mu_2 < \infty$. Also write $\underline{\pi} := \inf_{\mathbf{w} \in \mathcal{D}^+} \pi \mathbf{W}(\mathbf{w})$ and $\bar{\pi} := \sup_{\mathbf{w} \in \mathcal{D}^+} \pi \mathbf{W}(\mathbf{w})$.

Theorem EC.3.4 (Restatement of Theorem 6.1). *Under Assumption 2, for any $h > 0$ and $\delta \in (0, 1)$, define $C_{h,\delta} := \frac{16R_\delta}{\pi \mu_0 h^d}$ and $T_{h,\delta} := \left\lceil \max \left\{ T_\rho, \left(\frac{4R_\delta}{\pi \mu_0 h^d} \right)^4 \right\} \right\rceil$. Then, for every $\epsilon > 0$, there exists $h_\epsilon > 0$ such that, for every fixed $h \in (0, h_\epsilon]$, $\delta \in (0, 1)$, and $t \geq T_{h,\delta}$,*

$$\mathbb{P} \left(\sup_{\mathbf{z}: \mathbf{W}_{t+1}(\mathbf{z}) \in \mathcal{D}} |\mathbb{P}(G_{t+1} \notin \mathcal{C}_t(\mathbf{z}) \mid \mathcal{F}_t, \mathbf{z}_{t+1} = \mathbf{z}) - \alpha| \leq \epsilon + C_{h,\delta} \sqrt{\frac{\log t}{t}} \right) \geq 1 - \delta.$$

The outer probability is over the historical data that generate $\mathcal{F}_t$.

Proof. The proof has three steps. First, we show that the population kernel CDF F_h uniformly approximates $F_{Y|\mathbf{W}}$ and construct the bandwidth threshold h_ϵ needed for this approximation. Second, Lemma EC.3.3 is used to control the empirical kernel ratio around F_h . Finally, we convert the resulting uniform CDF error into a coverage error at the two estimated quantiles.

Step 1: In this step we construct $h_\epsilon > 0$ such that $\sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |F_h(y \mid \mathbf{w}) - F_{Y|\mathbf{W}}(y \mid \mathbf{w})| \leq \frac{\epsilon}{2}$ whenever $0 < h \leq h_\epsilon$. By stationarity and conditioning on $\mathbf{W}_1$, $A_h(y, \mathbf{w}) = \int \kappa_h(\mathbf{u}, \mathbf{w}) F_{Y|\mathbf{W}}(y \mid$

$\mathbf{u})\pi_{\mathbf{W}}(\mathbf{u})d\mathbf{u}$ and $B_h(\mathbf{w}) = \int \kappa_h(\mathbf{u}, \mathbf{w})\pi_{\mathbf{W}}(\mathbf{u})d\mathbf{u}$. Consequently,

$$F_h(y | \mathbf{w}) - F_{Y|\mathbf{W}}(y | \mathbf{w}) = \frac{1}{B_h(\mathbf{w})} \int \kappa_h(\mathbf{u}, \mathbf{w}) [F_{Y|\mathbf{W}}(y | \mathbf{u}) - F_{Y|\mathbf{W}}(y | \mathbf{w})] \pi_{\mathbf{W}}(\mathbf{u})d\mathbf{u}. \quad (\text{EC.29})$$

Thus, we first need a uniform lower bound for the denominator $B_h(\mathbf{w})$ and then an upper bound for the numerator in (EC.29). For the denominator, the ball $\{\mathbf{u} : \|\mathbf{u} - \mathbf{w}\| \leq h_0\}$ is contained in $\mathcal{D}^+$ when $\mathbf{w} \in \mathcal{D}$. Therefore,

$$B_h(\mathbf{w}) \geq \pi \int_{\|\mathbf{u}-\mathbf{w}\| \leq h_0} \kappa_h(\mathbf{u}, \mathbf{w})d\mathbf{u} = \pi h^d \int_{\|\mathbf{v}\| \leq h_0/h} K(\mathbf{v})d\mathbf{v} = \pi h^d \left[\mu_0 - \int_{\|\mathbf{v}\| > h_0/h} K(\mathbf{v})d\mathbf{v} \right].$$

By the definition of μ_2 ,

$$\int_{\|\mathbf{v}\| > h_0/h} K(\mathbf{v})d\mathbf{v} \leq \frac{h^2}{h_0^2} \int_{\|\mathbf{v}\| > h_0/h} \|\mathbf{v}\|^2 K(\mathbf{v})d\mathbf{v} \leq \frac{\mu_2 h^2}{h_0^2}.$$

It follows that, whenever $0 < h \leq \min \left\{ h_0, h_0 \sqrt{\frac{\mu_0}{2\mu_2}} \right\}$, we have

$$B_h(\mathbf{w}) \geq \pi h^d \left(\mu_0 - \frac{\mu_2 h^2}{h_0^2} \right) \geq \frac{\pi \mu_0}{2} h^d \quad (\text{EC.30})$$

for every $\mathbf{w} \in \mathcal{D}$. We now control the numerator in (EC.29). Fix $\epsilon > 0$. By Assumption 2(d), we may choose $a \in (0, h_0)$ such that $\sup_{\substack{\mathbf{w}, \mathbf{w}' \in \mathcal{D}^+ \\ \|\mathbf{w} - \mathbf{w}'\| \leq a}} \sup_{y \in \mathbb{R}} |F_{Y|\mathbf{W}}(y | \mathbf{w}) - F_{Y|\mathbf{W}}(y | \mathbf{w}')| \leq \frac{\epsilon}{4}$. To use this local continuity while accounting for observations farther from $\mathbf{w}$, split the integral in (EC.29) into the three regions: (1) $\|\mathbf{u} - \mathbf{w}\| \leq a$, (2) $a < \|\mathbf{u} - \mathbf{w}\| \leq h_0$, and (3) $\|\mathbf{u} - \mathbf{w}\| > h_0$.

On the first region, $\mathbf{u}$ and $\mathbf{w}$ both lie in $\mathcal{D}^+$, and the conditional CDFs differ by at most $\epsilon/4$. Hence, after division by $B_h(\mathbf{w})$, this region contributes at most $\epsilon/4$; On the second region, both points still lie in $\mathcal{D}^+$, so $\pi_{\mathbf{W}}(\mathbf{u}) \leq \bar{\pi}$. Since the difference between two CDF values is at most one, the corresponding part of the numerator is bounded by

$$\bar{\pi} \int_{a < \|\mathbf{u}-\mathbf{w}\| \leq h_0} \kappa_h(\mathbf{u}, \mathbf{w})d\mathbf{u} \leq \bar{\pi} h^d \int_{\|\mathbf{v}\| > a/h} K(\mathbf{v})d\mathbf{v} \leq \frac{\bar{\pi} \mu_2}{a^2} h^{d+2},$$

where the final inequality again follows from the definition of μ_2 ; And on the third region $\|\mathbf{u} - \mathbf{w}\| > h_0$, the density bound on $\mathcal{D}^+$ is no longer available there, so instead we use the pointwise decay implied by radial monotonicity. For every $r > 0$, by

$$\frac{v_d(1 - 2^{-d})}{4} r^{d+2} \tilde{\kappa}(r) \leq \int_{r/2 < \|\mathbf{v}\| \leq r} \|\mathbf{v}\|^2 K(\mathbf{v})d\mathbf{v} \leq \mu_2,$$

we have

$$\tilde{\kappa}(r) \leq \frac{4\mu_2}{v_d(1 - 2^{-d})r^{d+2}}. \quad (\text{EC.31})$$

Because $\tilde{\kappa}$ is nonincreasing and $\pi_{\mathbf{W}}$ integrates to one, the contribution of the third region to the numerator is at most

$$\int_{\|\mathbf{u}-\mathbf{w}\|>h_0} \kappa_h(\mathbf{u}, \mathbf{w}) \pi_{\mathbf{W}}(\mathbf{u}) d\mathbf{u} \leq \tilde{\kappa}(h_0/h) \leq \frac{4\mu_2}{v_d(1-2^{-d})h_0^{d+2}} h^{d+2}.$$

Combining the bounds for the three regions and using (EC.30), we obtain

$$\sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |F_h(y | \mathbf{w}) - F_{Y|\mathbf{W}}(y | \mathbf{w})| \leq \frac{\epsilon}{4} + \frac{2}{\pi\mu_0} \left(\frac{\pi\mu_2}{a^2} + \frac{4\mu_2}{v_d(1-2^{-d})h_0^{d+2}} \right) h^2,$$

where the first term is the local approximation error, while the second collects the kernel contributions from observations more than a away from $\mathbf{w}$. Define

$$h_\epsilon := \min \left\{ h_0, \quad h_0 \sqrt{\frac{\mu_0}{2\mu_2}}, \quad \left[\frac{8}{\epsilon \pi\mu_0} \left(\frac{\pi\mu_2}{a^2} + \frac{4\mu_2}{v_d(1-2^{-d})h_0^{d+2}} \right) \right]^{-1/2} \right\}. \quad (\text{EC.32})$$

This choice is strictly positive and ensures that, for every $0 < h \leq h_\epsilon$,

$$\sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |F_h(y | \mathbf{w}) - F_{Y|\mathbf{W}}(y | \mathbf{w})| \leq \frac{\epsilon}{2}. \quad (\text{EC.33})$$

Step 2: We now control the difference between the empirical CDF $\hat{F}_t$ and its population counterpart F_h . Since $\hat{F}_t(y | \mathbf{w}) = \frac{A_{t,h}(y, \mathbf{w})}{B_{t,h}(\mathbf{w})}$ and $F_h(y | \mathbf{w}) = \frac{A_h(y, \mathbf{w})}{B_h(\mathbf{w})}$, the first task is to show that the empirical denominator $B_{t,h}(\mathbf{w})$ remains uniformly away from zero. After that, the concentration bounds for the empirical numerator and denominator can be converted into a bound for the ratio. Fix $\delta \in (0, 1)$ and a bandwidth satisfying $0 < h \leq \min \left\{ h_0, h_0 \sqrt{\frac{\mu_0}{2\mu_2}} \right\}$ (this condition is automatically satisfied whenever $h \leq h_\epsilon$ from (EC.32)). Also fix $t \geq T_{h,\delta}$. Since $T_{h,\delta} \geq T_\rho$, Lemma EC.3.3 implies that, with probability at least $1 - \delta$,

$$\sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |A_{t,h}(y, \mathbf{w}) - A_h(y, \mathbf{w})| \vee \sup_{\mathbf{w} \in \mathcal{D}} |B_{t,h}(\mathbf{w}) - B_h(\mathbf{w})| \leq R_\delta \sqrt{\frac{\log t}{t}}. \quad (\text{EC.34})$$

We work on this event for the remainder of the step. The second component in the definition of $T_{h,\delta}$ ensures that the deviation in (EC.34) is small relative to the population denominator. Indeed, $T_{h,\delta} \geq T_\rho \geq 3$, so $\log t \leq \sqrt{t}$, and therefore $R_\delta \sqrt{\frac{\log t}{t}} \leq R_\delta t^{-1/4}$. Moreover, the inequality $t \geq \left(\frac{4R_\delta}{\pi\mu_0 h^d} \right)^4$ implies $R_\delta t^{-1/4} \leq \frac{\pi\mu_0}{4} h^d$. Combining this bound with (EC.30) and (EC.34) gives

$$\begin{aligned} \inf_{\mathbf{w} \in \mathcal{D}} B_{t,h}(\mathbf{w}) &\geq \inf_{\mathbf{w} \in \mathcal{D}} B_h(\mathbf{w}) - \sup_{\mathbf{w} \in \mathcal{D}} |B_{t,h}(\mathbf{w}) - B_h(\mathbf{w})| \\ &\geq \frac{\pi\mu_0}{2} h^d - \frac{\pi\mu_0}{4} h^d = \frac{\pi\mu_0}{4} h^d > 0. \end{aligned} \quad (\text{EC.35})$$

We can now control the ratio error. For every $\mathbf{w} \in \mathcal{D}$ and $y \in \mathbb{R}$,

$$\widehat{F}_t(y | \mathbf{w}) - F_h(y | \mathbf{w}) = \frac{A_{t,h}(y, \mathbf{w}) - A_h(y, \mathbf{w})}{B_{t,h}(\mathbf{w})} + F_h(y | \mathbf{w}) \frac{B_h(\mathbf{w}) - B_{t,h}(\mathbf{w})}{B_{t,h}(\mathbf{w})}.$$

Because $0 \leq F_h(y | \mathbf{w}) \leq 1$, the two concentration bounds in (EC.34) and the lower bound for $B_{t,h}$ give

$$\sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |\widehat{F}_t(y | \mathbf{w}) - F_h(y | \mathbf{w})| \leq \frac{2R_\delta \sqrt{\log t/t}}{(\pi\mu_0/4)h^d} = \frac{8R_\delta}{\pi\mu_0 h^d} \sqrt{\frac{\log t}{t}}. \quad (\text{EC.36})$$

Finally, for $h \leq h_\epsilon$, (EC.22), (EC.33), and (EC.36) imply

$$\sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |\widehat{F}_t(y | \mathbf{w}) - F_{Y|\mathbf{W}}(y | \mathbf{w})| \leq \frac{\epsilon}{2} + \frac{8R_\delta}{\pi\mu_0 h^d} \sqrt{\frac{\log t}{t}} = \frac{1}{2} \left(\epsilon + C_{h,\delta} \sqrt{\frac{\log t}{t}} \right). \quad (\text{EC.37})$$

The final equality follows from the definition of $C_{h,\delta}$.

Step 3: The last step is to translate the uniform CDF bound (EC.37) into a bound on conditional coverage. This result follows from two facts: first, uniform closeness of $\widehat{F}_t$ and $F_{Y|\mathbf{W}}$ implies that an empirical quantile has approximately the correct probability under $F_{Y|\mathbf{W}}$; second, predictive sufficiency of $\mathbf{W}_{t+1}(\mathbf{z})$ on Y_{t+1} identifies these probabilities with the conditional tail probabilities of Y_{t+1} under each admissible scenario.

Fix $u \in (0, 1)$ and $\mathbf{w} \in \mathcal{D}$, and write $q = \widehat{Q}_t(u; \mathbf{w})$. By the generalized-inverse convention in (EC.21), $\widehat{F}_t(q | \mathbf{w}) \geq u$ and $\widehat{F}_t(x | \mathbf{w}) < u$ for every $x < q$. The first inequality gives

$$\begin{aligned} F_{Y|\mathbf{W}}(q | \mathbf{w}) &\geq \widehat{F}_t(q | \mathbf{w}) - \sup_{\substack{\mathbf{w}' \in \mathcal{D} \\ y \in \mathbb{R}}} |\widehat{F}_t(y | \mathbf{w}') - F_{Y|\mathbf{W}}(y | \mathbf{w}')| \\ &\geq u - \sup_{\substack{\mathbf{w}' \in \mathcal{D} \\ y \in \mathbb{R}}} |\widehat{F}_t(y | \mathbf{w}') - F_{Y|\mathbf{W}}(y | \mathbf{w}')|. \end{aligned}$$

Similarly, for every $x < q$, $F_{Y|\mathbf{W}}(x | \mathbf{w}) \leq u + \sup_{\substack{\mathbf{w}' \in \mathcal{D} \\ y \in \mathbb{R}}} |\widehat{F}_t(y | \mathbf{w}') - F_{Y|\mathbf{W}}(y | \mathbf{w}')|$. Letting $x \uparrow q$ and using the continuity in Assumption 2(d) yields

$$\sup_{\mathbf{w} \in \mathcal{D}} \left| F_{Y|\mathbf{W}}(\widehat{Q}_t(u; \mathbf{w}) | \mathbf{w}) - u \right| \leq \sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |\widehat{F}_t(y | \mathbf{w}) - F_{Y|\mathbf{W}}(y | \mathbf{w})|. \quad (\text{EC.38})$$

We next apply this result to the two endpoints of the KSA interval. Fix an admissible scenario $\mathbf{z}$, write $\mathbf{w} = \mathbf{W}_{t+1}(\mathbf{z})$, $q_- = \widehat{Q}_t(\alpha/2; \mathbf{w})$, and $q_+ = \widehat{Q}_t(1 - \alpha/2; \mathbf{w})$. The positivity of the empirical denominator established in (EC.35) ensures that these quantiles are well defined. They are $\mathcal{F}_t$ -measurable, and monotonicity of the generalized inverse gives $q_- \leq q_+$. By Assumption 2(b), the predictive-sufficiency identity holds almost surely, simultaneously for all thresholds and admissible scenarios. Since q_- and q_+ are $\mathcal{F}_t$ -measurable, we may therefore evaluate this identity at these two

thresholds.

$$\mathbb{P}(Y_{t+1} < q_- \mid \mathcal{F}_t, \mathbf{z}_{t+1} = \mathbf{z}) = F_{Y|\mathbf{W}}(q_- \mid \mathbf{w}), \quad \mathbb{P}(Y_{t+1} \leq q_+ \mid \mathcal{F}_t, \mathbf{z}_{t+1} = \mathbf{z}) = F_{Y|\mathbf{W}}(q_+ \mid \mathbf{w}).$$

Under the same conditional law, $Y_{t+1} = G_{t+1} - \varphi(\mathbf{w})$. Since the endpoints of $\mathcal{C}_t(\mathbf{z})$ are $\varphi(\mathbf{w}) + q_-$ and $\varphi(\mathbf{w}) + q_+$, respectively,

$$G_{t+1} \notin \mathcal{C}_t(\mathbf{z}) \iff Y_{t+1} < q_- \quad \text{or} \quad Y_{t+1} > q_+.$$

Consequently, $\mathbb{P}(G_{t+1} \notin \mathcal{C}_t(\mathbf{z}) \mid \mathcal{F}_t, \mathbf{z}_{t+1} = \mathbf{z}) = F_{Y|\mathbf{W}}(q_- \mid \mathbf{w}) + 1 - F_{Y|\mathbf{W}}(q_+ \mid \mathbf{w})$. Subtracting $\alpha = \alpha/2 + \{1 - (1 - \alpha/2)\}$ and applying the triangle inequality gives

$$|\mathbb{P}(G_{t+1} \notin \mathcal{C}_t(\mathbf{z}) \mid \mathcal{F}_t, \mathbf{z}_{t+1} = \mathbf{z}) - \alpha| \leq \left| F_{Y|\mathbf{W}}(q_- \mid \mathbf{w}) - \frac{\alpha}{2} \right| + \left| F_{Y|\mathbf{W}}(q_+ \mid \mathbf{w}) - \left(1 - \frac{\alpha}{2}\right) \right|.$$

Applying (EC.38) at the two quantile levels and taking the supremum over all admissible scenarios yields

$$\sup_{\mathbf{z}: \mathbf{W}_{t+1}(\mathbf{z}) \in \mathcal{D}} |\mathbb{P}(G_{t+1} \notin \mathcal{C}_t(\mathbf{z}) \mid \mathcal{F}_t, \mathbf{z}_{t+1} = \mathbf{z}) - \alpha| \leq 2 \sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |\hat{F}_t(y \mid \mathbf{w}) - F_{Y|\mathbf{W}}(y \mid \mathbf{w})|. \quad (\text{EC.39})$$

The factor 2 appears because the interval has two endpoints, each of which contributes at most the uniform CDF error. Finally, combining (EC.37) and (EC.39) gives, on the event from Lemma EC.3.3,

$$\sup_{\mathbf{z}: \mathbf{W}_{t+1}(\mathbf{z}) \in \mathcal{D}} |\mathbb{P}(G_{t+1} \notin \mathcal{C}_t(\mathbf{z}) \mid \mathcal{F}_t, \mathbf{z}_{t+1} = \mathbf{z}) - \alpha| \leq \epsilon + C_{h,\delta} \sqrt{\frac{\log t}{t}}.$$

This event has probability at least $1 - \delta$, which proves the probability-quantified restatement. Since $\delta \in (0, 1)$ is arbitrary and $C_{h,\delta} < \infty$ for every fixed $h > 0$, the same result yields the $O_p(\sqrt{\log t/t})$ conclusion in (26). $\square$

The next result makes the two bandwidth orders in Corollary 6.2 explicit.

Corollary EC.3.5 (Restatement of Corollary 6.2). *Suppose Assumption 2 holds. For $h > 0$ and $\delta \in (0, 1)$, define $C_{h,\delta}$ and $T_{h,\delta}$ as in Theorem EC.3.4. Further define $\mu_1 := \int_{\mathbb{R}^d} \|\mathbf{v}\| K(\mathbf{v}) d\mathbf{v}$, $\bar{h} := \min \left\{ h_0, h_0 \sqrt{\frac{\mu_0}{2\mu_2}} \right\}$, and $R_0 := \frac{4\mu_2}{v_d(1-2^{-d})h_0^{d+2}}$. Then the following statements hold.*

(i) *Suppose that, for some $L < \infty$,*

$$|F_{Y|\mathbf{W}}(y \mid \mathbf{w}) - F_{Y|\mathbf{W}}(y \mid \mathbf{w}')| \leq L \|\mathbf{w} - \mathbf{w}'\|, \quad \mathbf{w}, \mathbf{w}' \in \mathcal{D}^+, \quad y \in \mathbb{R}.$$

Let $D_L := \frac{4}{\pi\mu_0} (\pi L\mu_1 + R_0\bar{h})$. For every $h \in (0, \bar{h}]$, $\delta \in (0, 1)$, and $t \geq T_{h,\delta}$,

$$\mathbb{P} \left(\sup_{\mathbf{z}: \mathbf{W}_{t+1}(\mathbf{z}) \in \mathcal{D}} |\mathbb{P}(G_{t+1} \notin \mathcal{C}_t(\mathbf{z}) \mid \mathcal{F}_t, \mathbf{z}_{t+1} = \mathbf{z}) - \alpha| \leq D_L h + C_{h,\delta} \sqrt{\frac{\log t}{t}} \right) \geq 1 - \delta.$$

(ii) Suppose $\pi_{\mathbf{w}}$ is twice continuously differentiable and, for every $y \in \mathbb{R}$, $\pi_{\mathbf{w}}(\cdot)F_{Y|\mathbf{w}}(y|\cdot)$ is twice continuously differentiable on an open neighborhood of $\mathcal{D}^+$. Assume $M_\pi := \sup_{\mathbf{w} \in \mathcal{D}^+} \|\nabla_{\mathbf{w}}^2 \pi_{\mathbf{w}}(\mathbf{w})\|_{\text{op}}$ and $M_{\pi F} := \sup_{\substack{\mathbf{w} \in \mathcal{D}^+ \\ y \in \mathbb{R}}} \|\nabla_{\mathbf{w}}^2 [\pi_{\mathbf{w}}(\mathbf{w})F_{Y|\mathbf{w}}(y|\mathbf{w})]\|_{\text{op}}$ are both finite, where $\|\cdot\|_{\text{op}}$ is the operator norm induced by the Euclidean norm. Define $D_2 := \frac{2(M_{\pi F} + M_\pi)\mu_2 + 8R_0}{\pi\mu_0}$. For every $h \in (0, \bar{h}]$, $\delta \in (0, 1)$, and $t \geq T_{h,\delta}$,

$$\mathbb{P} \left(\sup_{\mathbf{z}: \mathbf{w}_{t+1}(\mathbf{z}) \in \mathcal{D}} |\mathbb{P}(G_{t+1} \notin \mathcal{C}_t(\mathbf{z}) | \mathcal{F}_t, \mathbf{z}_{t+1} = \mathbf{z}) - \alpha| \leq D_2 h^2 + C_{h,\delta} \sqrt{\frac{\log t}{t}} \right) \geq 1 - \delta.$$

In both statements, the outer probability is over the historical data that generate $\mathcal{F}_t$.

Proof. The proof refines the population-approximation argument in **Step 1** of the proof of Theorem EC.3.4. The empirical-concentration and coverage-conversion arguments remain unchanged. We first derive the two refined population bounds and then combine them with the remaining arguments.

Step 1: For every $0 < h \leq \bar{h}$, the denominator bound (EC.30) gives $\inf_{\mathbf{w} \in \mathcal{D}} B_h(\mathbf{w}) \geq \frac{\pi\mu_0}{2} h^d$. We now sharpen the numerator bound under each smoothness condition.

Under part (i), fix $\mathbf{w} \in \mathcal{D}$ and $y \in \mathbb{R}$. In the representation (EC.29), split the integral according to whether $\|\mathbf{u} - \mathbf{w}\| \leq h_0$. On this region, both $\mathbf{u}$ and $\mathbf{w}$ lie in $\mathcal{D}^+$. The Lipschitz condition and the upper bound on $\pi_{\mathbf{w}}$ therefore give

$$\begin{aligned} & \int_{\|\mathbf{u}-\mathbf{w}\| \leq h_0} \kappa_h(\mathbf{u}, \mathbf{w}) |F_{Y|\mathbf{w}}(y|\mathbf{u}) - F_{Y|\mathbf{w}}(y|\mathbf{w})| \pi_{\mathbf{w}}(\mathbf{u}) d\mathbf{u} \\ & \leq L\bar{\pi} \int_{\|\mathbf{u}-\mathbf{w}\| \leq h_0} \kappa_h(\mathbf{u}, \mathbf{w}) \|\mathbf{u} - \mathbf{w}\| d\mathbf{u} \leq L\bar{\pi}\mu_1 h^{d+1}. \end{aligned}$$

On the complementary region, the difference between the two CDF values is at most one. Since $\tilde{\kappa}$ is nonincreasing and $\pi_{\mathbf{w}}$ integrates to one,

$$\int_{\|\mathbf{u}-\mathbf{w}\| > h_0} \kappa_h(\mathbf{u}, \mathbf{w}) |F_{Y|\mathbf{w}}(y|\mathbf{u}) - F_{Y|\mathbf{w}}(y|\mathbf{w})| \pi_{\mathbf{w}}(\mathbf{u}) d\mathbf{u} \leq \tilde{\kappa}(h_0/h) \leq R_0 h^{d+2},$$

where the last inequality follows from (EC.31) and the definition of R_0 . Substituting these bounds into (EC.29) and using (EC.30), we obtain

$$\sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |F_h(y|\mathbf{w}) - F_{Y|\mathbf{w}}(y|\mathbf{w})| \leq \frac{2}{\pi\mu_0} (\bar{\pi}L\mu_1 h + R_0 h^2) \leq \frac{2}{\pi\mu_0} (\bar{\pi}L\mu_1 + R_0 \bar{h}) h = \frac{D_L}{2} h. \quad (\text{EC.40})$$

Under part (ii), define for this part of the proof $q_y(\mathbf{u}) := \pi_{\mathbf{w}}(\mathbf{u})F_{Y|\mathbf{w}}(y|\mathbf{u})$ and $J_h := \int_{\|\mathbf{v}\| \leq h_0/h} K(\mathbf{v}) d\mathbf{v}$. Fix $\mathbf{w} \in \mathcal{D}$ and $y \in \mathbb{R}$. If $\|\mathbf{u} - \mathbf{w}\| \leq h_0$, then the entire line segment joining $\mathbf{w}$

and $\mathbf{u}$ lies in $\mathcal{D}^+$. After the change of variables $\mathbf{u} = \mathbf{w} + h\mathbf{v}$, Taylor's theorem gives

$$q_y(\mathbf{w} + h\mathbf{v}) = q_y(\mathbf{w}) + h\nabla_{\mathbf{w}}q_y(\mathbf{w})^\top \mathbf{v} + r_y(\mathbf{v}), \quad |r_y(\mathbf{v})| \leq \frac{M_{\pi F}}{2}h^2\|\mathbf{v}\|^2, \quad \|\mathbf{v}\| \leq \frac{h_0}{h}.$$

Because K is radial and the ball $\{\mathbf{v} : \|\mathbf{v}\| \leq h_0/h\}$ is symmetric, $\int_{\|\mathbf{v}\| \leq h_0/h} \mathbf{v}K(\mathbf{v})d\mathbf{v} = \mathbf{0}$. The linear Taylor term therefore integrates to zero, while the remainder contributes at most $M_{\pi F}\mu_2 h^2/2$. The portion of $h^{-d}A_h(y, \mathbf{w})$ arising from $\|\mathbf{u} - \mathbf{w}\| > h_0$ is bounded by $h^{-d}\tilde{\kappa}(h_0/h) \int_{\|\mathbf{u}-\mathbf{w}\| > h_0} q_y(\mathbf{u})d\mathbf{u} \leq R_0 h^2$, because $0 \leq q_y \leq \pi\mathbf{w}$. Consequently,

$$\left| h^{-d}A_h(y, \mathbf{w}) - J_h q_y(\mathbf{w}) \right| \leq \left(\frac{M_{\pi F}\mu_2}{2} + R_0 \right) h^2. \quad (\text{EC.41})$$

The same argument, with q_y replaced by $\pi\mathbf{w}$, gives

$$\left| h^{-d}B_h(\mathbf{w}) - J_h \pi\mathbf{w}(\mathbf{w}) \right| \leq \left(\frac{M_{\pi}\mu_2}{2} + R_0 \right) h^2. \quad (\text{EC.42})$$

Moreover, $q_y(\mathbf{w})/\pi\mathbf{w}(\mathbf{w}) = F_{Y|\mathbf{W}}(y | \mathbf{w}) \in [0, 1]$. Using $F_h = A_h/B_h$, we may therefore write

$$\begin{aligned} |F_h(y | \mathbf{w}) - F_{Y|\mathbf{W}}(y | \mathbf{w})| &= \left| \frac{h^{-d}A_h(y, \mathbf{w}) - J_h q_y(\mathbf{w})}{h^{-d}B_h(\mathbf{w})} + \frac{q_y(\mathbf{w})}{\pi\mathbf{w}(\mathbf{w})} \frac{J_h \pi\mathbf{w}(\mathbf{w}) - h^{-d}B_h(\mathbf{w})}{h^{-d}B_h(\mathbf{w})} \right| \\ &\leq \frac{|h^{-d}A_h(y, \mathbf{w}) - J_h q_y(\mathbf{w})| + \frac{q_y(\mathbf{w})}{\pi\mathbf{w}(\mathbf{w})} |J_h \pi\mathbf{w}(\mathbf{w}) - h^{-d}B_h(\mathbf{w})|}{h^{-d}B_h(\mathbf{w})}. \end{aligned}$$

Combining (EC.30), (EC.41) and (EC.42) yields

$$\sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |F_h(y | \mathbf{w}) - F_{Y|\mathbf{W}}(y | \mathbf{w})| \leq \frac{(M_{\pi F} + M_{\pi})\mu_2 + 4R_0}{\pi\mu_0} h^2 = \frac{D_2}{2} h^2. \quad (\text{EC.43})$$

Step 2: The remaining arguments follow from **Steps 2–3** in the proof of Theorem EC.3.4. Fix $\delta \in (0, 1)$ and an integer $t \geq T_{h,\delta}$. On the event in (EC.24), the argument leading to (EC.35) and (EC.36) applies because $h \leq \bar{h}$. Thus,

$$\sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |\hat{F}_t(y | \mathbf{w}) - F_h(y | \mathbf{w})| \leq \frac{8R_\delta}{\pi\mu_0 h^d} \sqrt{\frac{\log t}{t}} = \frac{C_{h,\delta}}{2} \sqrt{\frac{\log t}{t}}.$$

Combining this inequality with (EC.40) gives, under part (i),

$$\sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |\hat{F}_t(y | \mathbf{w}) - F_{Y|\mathbf{W}}(y | \mathbf{w})| \leq \frac{1}{2} \left(D_L h + C_{h,\delta} \sqrt{\frac{\log t}{t}} \right). \quad (\text{EC.44})$$

Similarly, (EC.43) gives, under part (ii),

$$\sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |\hat{F}_t(y | \mathbf{w}) - F_{Y|\mathbf{W}}(y | \mathbf{w})| \leq \frac{1}{2} \left(D_2 h^2 + C_{h,\delta} \sqrt{\frac{\log t}{t}} \right). \quad (\text{EC.45})$$

By the quantile and predictive-sufficiency argument in **Step 3** of the proof of Theorem EC.3.4, (EC.39) gives $\sup_{\mathbf{z}: \mathbf{W}_{t+1}(\mathbf{z}) \in \mathcal{D}} |\mathbb{P}(G_{t+1} \notin \mathcal{C}_t(\mathbf{z}) | \mathcal{F}_t, \mathbf{z}_{t+1} = \mathbf{z}) - \alpha| \leq 2 \sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |\hat{F}_t(y | \mathbf{w}) - F_{Y|\mathbf{W}}(y | \mathbf{w})|$. Substituting (EC.44) yields $D_L h + C_{h,\delta} \sqrt{\log t/t}$ under part (i), and substituting (EC.45) yields $D_2 h^2 + C_{h,\delta} \sqrt{\log t/t}$ under part (ii). Finally, Lemma EC.3.3 shows that the event on which these bounds hold has probability at least $1 - \delta$. $\square$

In the next corollary, we allow the bandwidth to vary with t and choose it to minimize the explicit upper bounds in Corollary EC.3.5.

Corollary EC.3.6. *Suppose Assumption 2 holds, fix $\delta \in (0, 1)$, and retain the notation from Theorem EC.3.4 and Corollary EC.3.5.*

(i) *Suppose the Lipschitz condition in part (i) of Corollary EC.3.5 holds. At time t , use*

$$h_t := \left[\frac{4dR_\delta}{\bar{\pi}L\mu_1 + \frac{4\mu_2\bar{h}}{v_d(1-2^{-d})h_0^{d+2}}} \sqrt{\frac{\log t}{t}} \right]^{1/(d+1)}. \quad (\text{EC.46})$$

For every integer $t \geq T_p$ such that

$$h_t \leq \min \left\{ \bar{h}, \frac{d\bar{\pi}\mu_0}{\bar{\pi}L\mu_1 + \frac{4\mu_2\bar{h}}{v_d(1-2^{-d})h_0^{d+2}}} \right\}, \quad (\text{EC.47})$$

the KSA interval constructed using κ_{h_t} satisfies

$$\begin{aligned} & \mathbb{P} \left(\sup_{\mathbf{z}: \mathbf{W}_{t+1}(\mathbf{z}) \in \mathcal{D}} |\mathbb{P}(G_{t+1} \notin \mathcal{C}_t(\mathbf{z}) | \mathcal{F}_t, \mathbf{z}_{t+1} = \mathbf{z}) - \alpha| \right. \\ & \leq \frac{4(d+1)}{d\bar{\pi}\mu_0} \left(\bar{\pi}L\mu_1 + \frac{4\mu_2\bar{h}}{v_d(1-2^{-d})h_0^{d+2}} \right) \\ & \left. \times \left[\frac{4dR_\delta}{\bar{\pi}L\mu_1 + \frac{4\mu_2\bar{h}}{v_d(1-2^{-d})h_0^{d+2}}} \sqrt{\frac{\log t}{t}} \right]^{1/(d+1)} \right) \geq 1 - \delta. \end{aligned} \quad (\text{EC.48})$$

(ii) Suppose the second-order condition in part (ii) of Corollary EC.3.5 holds. At time t , use

$$h_t := \left[\frac{4dR_\delta}{(M_{\pi F} + M_\pi)\mu_2 + \frac{16\mu_2}{v_d(1-2^{-d})h_0^{d+2}}} \sqrt{\frac{\log t}{t}} \right]^{1/(d+2)}. \quad (\text{EC.49})$$

For every integer $t \geq T_\rho$ such that

$$h_t \leq \min \left\{ \bar{h}, \sqrt{\frac{d\pi\mu_0}{(M_{\pi F} + M_\pi)\mu_2 + \frac{16\mu_2}{v_d(1-2^{-d})h_0^{d+2}}}} \right\}, \quad (\text{EC.50})$$

the KSA interval constructed using κ_{h_t} satisfies

$$\begin{aligned} & \mathbb{P} \left(\sup_{\mathbf{z}: \mathbf{w}_{t+1}(\mathbf{z}) \in \mathcal{D}} |\mathbb{P}(G_{t+1} \notin \mathcal{C}_t(\mathbf{z}) \mid \mathcal{F}_t, \mathbf{z}_{t+1} = \mathbf{z}) - \alpha| \right. \\ & \leq \frac{2(d+2)}{d\pi\mu_0} \left((M_{\pi F} + M_\pi)\mu_2 + \frac{16\mu_2}{v_d(1-2^{-d})h_0^{d+2}} \right) \\ & \times \left[\frac{4dR_\delta}{(M_{\pi F} + M_\pi)\mu_2 + \frac{16\mu_2}{v_d(1-2^{-d})h_0^{d+2}}} \sqrt{\frac{\log t}{t}} \right]^{2/(d+2)} \Bigg) \geq 1 - \delta. \end{aligned} \quad (\text{EC.51})$$

The conditions in (EC.47) and (EC.50) are eventually satisfied because the corresponding bandwidths converge to zero. The two coverage-error bounds converge to zero at the rates $(\log t/t)^{1/(2d+2)}$ and $(\log t/t)^{1/(d+2)}$ respectively.

Proof. Fix $\delta \in (0, 1)$ and $t \geq T_\rho$, and take h_t to be the bandwidth specified in the relevant part of the corollary. Since h_t is deterministic, Lemma EC.3.3 may be applied at time t with $h = h_t$. On the event in (EC.24), which has probability at least $1 - \delta$,

$$\sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |A_{t,h_t}(y, \mathbf{w}) - A_{h_t}(y, \mathbf{w})| \vee \sup_{\mathbf{w} \in \mathcal{D}} |B_{t,h_t}(\mathbf{w}) - B_{h_t}(\mathbf{w})| \leq R_\delta \sqrt{\frac{\log t}{t}}.$$

We work on this event. The ratio argument in the proof of Theorem EC.3.4 gives

$$\sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |\hat{F}_t(y \mid \mathbf{w}) - F_{h_t}(y \mid \mathbf{w})| \leq \frac{8R_\delta}{\pi\mu_0 h_t^d} \sqrt{\frac{\log t}{t}}. \quad (\text{EC.52})$$

For part (i), expanding R_0 and D_L in (EC.40) gives

$$\sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |F_{h_t}(y | \mathbf{w}) - F_{Y|\mathbf{W}}(y | \mathbf{w})| \leq \frac{2}{\pi\mu_0} \left(\pi L\mu_1 + \frac{4\mu_2\bar{h}}{v_d(1-2^{-d})h_0^{d+2}} \right) h_t.$$

Combining this bound with (EC.52) and the coverage-conversion inequality (EC.39), we obtain

$$\begin{aligned} & \sup_{\mathbf{z}: \mathbf{W}_{t+1}(\mathbf{z}) \in \mathcal{D}} |\mathbb{P}(G_{t+1} \notin \mathcal{C}_t(\mathbf{z}) | \mathcal{F}_t, \mathbf{z}_{t+1} = \mathbf{z}) - \alpha| \\ & \leq \frac{4}{\pi\mu_0} \left(\pi L\mu_1 + \frac{4\mu_2\bar{h}}{v_d(1-2^{-d})h_0^{d+2}} \right) h_t + \frac{16R_\delta}{\pi\mu_0 h_t^d} \sqrt{\frac{\log t}{t}}. \end{aligned} \quad (\text{EC.53})$$

The bandwidth in (EC.46) is the unique minimizer of the right-hand side of (EC.53) over $h_t > 0$. At this bandwidth, the second term on the right-hand side equals $1/d$ times the first. Hence, the minimum is $\frac{4(d+1)}{d\pi\mu_0} \left(\pi L\mu_1 + \frac{4\mu_2\bar{h}}{v_d(1-2^{-d})h_0^{d+2}} \right) h_t$, which, after substituting (EC.46), is the right-hand side of (EC.48).

For part (ii), expanding R_0 and D_2 in (EC.43) gives

$$\sup_{\mathbf{w} \in \mathcal{D}} \sup_{y \in \mathbb{R}} |F_{h_t}(y | \mathbf{w}) - F_{Y|\mathbf{W}}(y | \mathbf{w})| \leq \frac{1}{\pi\mu_0} \left((M_{\pi F} + M_\pi)\mu_2 + \frac{16\mu_2}{v_d(1-2^{-d})h_0^{d+2}} \right) h_t^2.$$

Combining this bound with (EC.52) and (EC.39) yields

$$\begin{aligned} & \sup_{\mathbf{z}: \mathbf{W}_{t+1}(\mathbf{z}) \in \mathcal{D}} |\mathbb{P}(G_{t+1} \notin \mathcal{C}_t(\mathbf{z}) | \mathcal{F}_t, \mathbf{z}_{t+1} = \mathbf{z}) - \alpha| \\ & \leq \frac{2}{\pi\mu_0} \left((M_{\pi F} + M_\pi)\mu_2 + \frac{16\mu_2}{v_d(1-2^{-d})h_0^{d+2}} \right) h_t^2 + \frac{16R_\delta}{\pi\mu_0 h_t^d} \sqrt{\frac{\log t}{t}}. \end{aligned} \quad (\text{EC.54})$$

The bandwidth in (EC.49) is the unique minimizer of the right-hand side of (EC.54) over $h_t > 0$. At this bandwidth, the second term on the right-hand side equals $2/d$ times the first. Therefore, the minimum is

$$\frac{2(d+2)}{d\pi\mu_0} \left((M_{\pi F} + M_\pi)\mu_2 + \frac{16\mu_2}{v_d(1-2^{-d})h_0^{d+2}} \right) h_t^2,$$

which is the right-hand side of (EC.51) after substituting (EC.49). $\square$

References

- Angelopoulos, Anastasios N, Stephen Bates. 2021. A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511* .
- Bradley, Richard C. 2005. Basic properties of strong mixing conditions. a survey and some open questions .
- Gibbs, Isaac, Emmanuel Candes. 2021. Adaptive conformal inference under distribution shift. *Advances in Neural Information Processing Systems* **34** 1660–1672.
- Gneiting, Tilmann, Adrian E Raftery. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* **102**(477) 359–378.
- Haussler, David. 1995. Sphere packing numbers for subsets of the boolean n -cube with bounded vapnik–chervonenkis dimension. *Journal of Combinatorial Theory, Series A* **69**(2) 217–232. doi:10.1016/0097-3165(95)90052-7.
- Mohri, Mehryar, Afshin Rostamizadeh, Ameet Talwalkar. 2018. *Foundations of Machine Learning*. 2nd ed. MIT Press, Cambridge, MA.
- van der Vaart, Aad W., Jon A. Wellner. 1996. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York. doi:10.1007/978-1-4757-2545-2.
- Yu, Bin. 1994. Rates of convergence for empirical processes of stationary mixing sequences. *The Annals of Probability* 94–116.