ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS
ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
SCHOOL OF ECONOMIC SCIENCES
ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΗΣ ΕΠΙΣΤΗΜΗΣ
DEPARTMENT OF ECONOMICS

Department of Economics

Athens University of Economics and Business

WORKING PAPER no. 13-2025

Updating Density Estimates using Conditional Information Projection: Stock Index Returns and Stochastic Dominance Relations

Stelios Arvanitis, Richard McGee, and Thierry Post

August 2025

Πατησίων 76, 104 34 Αθήνα. Τηλ.: 210 8203303-5. E-mail: econ@aueb.gr / www.aueb.gr
76, Patission Street, Athens 104 34 Greece. Tel.: (+30) 210 8203303-5

Πατησίων 76, 104 34 Αθήνα. Τηλ.: 210 8203303-5. E-mail: econ@aueb.gr / www.aueb.gr
76, Patission Street, Athens 104 34 Greece. Tel.: (+30) 210 8203303-5

Updating Density Estimates using Conditional Information Projection: Stock Index Returns and Stochastic Dominance Relations

Stelios Arvanitis[*], *Richard McGee*[†], *and Thierry Post*[‡]

Abstract

We propose, analyze, and apply a Conditional Information Projection Density Estimator (CIPDE) that estimates the conditional density function by projecting a prior time-series estimator onto the set of distributions that satisfy given conditional moment conditions with a functional nuisance parameter. A statistical theory is developed based on asymptotic theory, information geometry and variational analysis. Theoretically, CIPDE is shown to reduce the limiting divergence to the latent population density, if the prior is inconsistent and the moment conditions are well specified. A Monte Carlo simulation experiment and an empirical analysis apply the proposed method to stock index returns using pricing restrictions for index options. The simulations show large potential improvements in the density estimates in a controlled but realistic environment. In the empirical analysis, CIPDE is shown to enhance monthly S&P500 index return density forecasts and improve the out-of-sample investment performance of one-month S&P500 index option strategies by better timing of protective put purchases and covered call writing.
Keywords: Conditional density estimation, information projection, Stochastic Dominance, stock index returns, stock index options

[*]Queen's University, Canada, and Athens University of Economics and Business, Greece. Email: stelios@aueb.gr

[†]University College Dublin, Ireland. Email: richard.mcgee@ucd.ie

[‡]Graduate School of Business of Nazarbayev University, Kazakhstan. E-mail: thierry-post@hotmail.com

1 Introduction

Financial decision making in risk management, derivative pricing, and portfolio optimization often requires conditional density estimates of asset returns. Approaches based on Expected Utility, Stochastic Dominance (SD), or Mean-Lower Partial Moment orders rely on a full probabilistic view that accounts for asymmetry and tail risk beyond point forecasts and variances. In addition, the return distribution should be updated to reflect prevailing market conditions. Common sources of conditioning information include market prices and investment yields of securities, and theoretical pricing restrictions.

A key challenge is incorporating new information to refine an existing density estimate. A large forecasting literature estimates conditional return distributions using GARCH-type models, quantile regressions, and nonparametric methods. These approaches rely primarily on historical data and often do not explicitly integrate real-time market conditions or theoretical pricing restrictions. How can an existing density estimate be updated to incorporate the available conditioning information?

We propose a new estimator for this task: the Conditional Information Projection Density Estimator (CIPDE). It starts with a prior estimate of the conditional return distribution and updates it to satisfy conditional moment conditions via information projection to minimize the divergence from the prior. This results in a new density estimate that integrates information from both the prior estimate and the moment conditions.

Although CIPDE is valid for multivariate densities in large samples, our focus is primarily on forecasting univariate densities to avoid the curse of dimensionality in finite samples.

The divergence-based projection approaches are rooted in a rich literature spanning econometrics, statistics, and information theory (Csiszár, 1975; Kitamura and Stutzer, 1997; Imbens et al., 1998; Schennach, 2005; Komunjer and Ragusa, 2016) and the references therein. Early work developed exponential tilting and Empirical Likelihood (EL) methods as tools to incorporate moment restrictions into estimation.

Although the structure resembles Bayesian updating, adjusting a prior via an exponential tilt, we depart from a full Bayesian interpretation. The CIPDE does not require a likelihood or parametric model for returns; instead, it uses the moment conditions to define an admissible set of distributions and projects the prior onto this set. This leads to a distribution that is closest to the prior in Kullback-Leibler (KL) divergence, subject to satisfying the economic constraints.

To acknowledge that side information is generally incomplete, the conditional moment conditions may take the form of inequalities and include functional nuisance parameters such as the ubiquitous Stochastic Discount Factor (SDF) in Finance.

In our empirical analysis, we anticipate that the initial density estimator may be statistically inconsistent (and thus may asymptotically violate moment conditions) because it is constructed under assumptions of debatable distributional shapes or erroneous conditioning information.

We establish a statistical theory for the CIPDE, including pseudo-consistency, convergence rates, and the limiting distribution, using a combination of information-theoretic and variational techniques. Among other things, the theory shows that CIPDE asymptotically reduces KL divergence to the latent true conditional distribution under general regularity conditions.

A Monte Carlo simulation experiment and an empirical analysis apply CIPDE to stock index returns using conditional moment restrictions based on market prices and pricing restrictions for index options. The experiment foreshadows the empirical analysis by forecasting and optimization based on repeated random samples drawn from a simplified but known simulation process. The empirical analysis focuses on density forecasting for monthly S&P 500 index (SPX) returns and optimization of combinations of one-month Chicago Board Options Exchange (CBOE) SPX options given the posterior density estimate.

For the pricing and trading of options, it is essential to account for the conditional and non-Gaussian nature of the return distribution. Most liquid index options have a short maturity of several days or weeks and their valuation and risk assessment critically requires depends on up-to-date market information. Since option payoffs are inherently asymmetric, mean-variance analysis is not appropriate and higher-order risk needs to be taken into account. Encouragingly, payoffs at expiry of concurrent option series are driven solely by by the value of the index at the option expiration date, so that the estimation can focus the one-dimensional density of SPX returns.

The conditional moment conditions in the simulation experiment and empirical analysis are based on observed market price quotes for index options and general option pricing restrictions that exclude SD relations in the spirit of Constantinides et al. (2009). In the simulation experiment, these conditions are known to be true for all options; in the empirical analysis, they are assumed to be true for the options that are selected based on a number of option characteristics (maturity, moneyness, delta and premium).

The optimization of option combinations is inspired by Constantinides et al. (2020); Post and Longarela (2021); Beare et al. (2025), but, in contrast to these studies, focuses on a type of approximate SD because, by construction, exact SD relations do not exist under the posterior density estimate (at least for the options that are used to define the conditional moment conditions).

The aforementioned index option studies employed conditional density estimates that are statistically inconsistent because they fix or restrict the variation of the shape of the distribution and its parameters and use a limited set of conditioning information (such as the risk-free rate and a market volatility index). In the present study, these estimates serve as the initial estimator, and we attempt to correct their biases using the conditional moment conditions.

The simulations demonstrate significant potential improvements in density estimates through information projection, particularly when the prior is biased or imprecise. The reduction in relative entropy increases with the number of qualified options and decreases in the bid-ask spread, highlighting the crucial role of conditioning information.

In the empirical analysis of SPX returns and options, CIPDE significantly improves the density forecasts out-of-sample (OOS) for a range of prior density forecasts. In optimizing index option combinations, the use of CIPDE yields significant OOS performance improvements, accounting for quoted bid-ask spreads using a simple buy-and-hold strategy. These improvements result from better timing in buying protective put options and writing covered call options.

Apart from the link with the aforementioned studies on the pricing and trading of index options using SD, our study is related to several earlier applications of EL and Bayesian econometrics in asset pricing and portfolio optimization.

Most notably, Stutzer (1996) applied Bayesian updating with moment conditions to option prices for option valuation. However, the study focuses on estimating the risk-neutral distribution rather than the physical (or real-world) distribution. The risk-neutral distribution is a biased estimator of the physical distribution because it conflates the physical distribution with the SDF. Although the risk-neutral distribution can be used for valuing options, it is less suitable for real-world decision making and portfolio optimization based on Expected Utility due to its biased nature. Furthermore, Stutzer (1996) does not provide a formal statistical theory for the estimated risk neutral distribution, instead focusing on practical application rather than developing a comprehensive statistical framework for the distribution itself.

In a related study, Stutzer (1995) pioneers the application of Bayesian econometrics for model selection and validation in empirical asset pricing. Similarly, Almeida and Garcia (2012) and Post and Potì (2017) develop and apply asset pricing tests based on general Minimum Discrepancy estimation and a combination of EL and SD, respectively. These studies however do not focus on density estimation and its application in forecasting and optimization.

Post et al. (2018) developed statistical theory and numerical algorithms for portfolio optimization based on EL and applied them to the rotation of the equity industry. The present study focuses on information projection instead of moment projection, considers conditional moment conditions, and it account for potentially inconsistent priors and the role of functional nuisance parameters. In addition, the index option combination problem seems more suitable for a model-free approach than the asset allocation problem, due to its lower dimensionality (one underlying stock index vs. a multitude of industries).

The remainder of the paper is structured as follows. Section 2 introduces the formal framework of the CIPDE, presents its variational representation, and establishes the statistical properties of the estimator. Section 3 introduces the principles of option pricing and option trading that will be used in our simulation study and empirical study. Section 4 conducts Monte Carlo experiments, and Section 5 applies the methodology to the estimation of the conditional distribution of monthly S&P 500 index returns and optimization of SPX option combinations. Section 5 concludes.

2 Statistical Theory

2.1 Preliminaries

We aim to estimate a latent conditional distribution F, using (i) an initial estimator $\hat{F}$ (the prior) and (ii) a set of J valid conditional moment inequalities. The proposed estimator belongs to the class of information projection-based methods, relying on the minimization of the KL divergence subject to the moment constraints. Unlike classical Bayesian updating which uses a likelihood function, this procedure integrates information through conditional moment restrictions.

The framework admits the possibility that the distribution depends on conditioning information summarized by observed and/or latent variables (e.g., covariates). While those are not explicitly modeled here, all results are to be understood as applying conditionally, for almost every fixed value of the conditioning variables.

The support $\mathcal{S} \subseteq \mathbb{R}^p$ is assumed independent of the conditioning variables. In our simulation experiment and empirical analysis, a univariate distribution ($p = 1$) is used, but the method also applies to multivariate cases ($p \geq 2$), although the curse of dimensionality presumably makes direct applicability in high-dimensional applications elusive; see the Conclusions.

This curse arises because the in high dimensions, the linear growth of the available data is not usually able to match the exponentially growing complexity of the feasible set, making it increasingly difficult to accurately estimate the high-dimensional conditional density or ensure feasibility of the moment inequalities without substantial regularization or structural assumptions.

The prior $\hat{F}$, which may be a kernel density estimator or a parametric model-based estimate, is assumed to be absolutely continuous with respect to the Lebesgue measure, almost surely. This means that $\hat{F}$ admits a density function $\hat{f}$ on its support. This assumption is standard in nonparametric and semiparametric density estimation and holds for most practical estimators used in empirical work. It ensures that the KL divergence between $\hat{F}$ and other distributions with densities is well-defined. $\hat{F}$ is allowed to be an inconsistent estimator of F.

The conditional moment conditions include a functional parameter $m \in \mathcal{M}$, where $\mathcal{M}$ is a convex and closed subset of a Sobolev space of smooth functions. More precisely, $\mathcal{M} \subset W^{2,2}(\mathcal{S})$, the space of square-integrable functions on $\mathcal{S}$ whose first and second weak derivatives are also square-integrable; see Adams and Fournier (2003) for a formal definition. In our simulation experiment and empirical analysis, the functional parameter is an SDF. The moment conditions take the form of inequalities:

$$\int_{\mathcal{S}} \mathbf{H}(S)m(S)g(S)dS \geq \mathbf{0},$$

where g is the density of a candidate conditional distribution G, and $\mathbf{H} : \mathcal{S} \to \mathbb{R}^J$ is a known vector-valued function.

The convex set of admissible densities that satisfy the moment inequalities at the functional parameter m is denoted by

$$\mathbb{M}(\mathbf{H}, \hat{F}; m) := \left\{ G \ll \hat{F} : \int_{\mathcal{S}} \mathbf{H}(S)m(S)g(S)dS \geq \mathbf{0} \right\},$$

where $\ll$ denotes absolute continuity.

The CIPDE estimator $\hat{G}$ is introduced in the next subsection. Its asymptotic properties as a sample size $T \to \infty$ are investigated in the subsequent subsections. The analysis builds upon variational representations, convexity properties, and information-theoretic tools to derive results concerning pseudo-consistency, convergence rates, robustness to prior inconsistency, and inference validity for SD comparisons.

In the following, convergence of densities is mostly assumed to be taking place in $\ell^1(S)$; the space of (equivalence classes of) integrable real functions on S equipped with the usual integral of absolute value norm. The symbol $\rightsquigarrow$ denotes weak convergence, and $\overset{\text{epi}}{\rightsquigarrow}$ denotes epi-convergence in distribution; see Knight (1999).

2.2 The CIPDE Estimator

To define the CIPDE estimator we use a variational problem formulation of the traditional constrained minimization of the relative entropy:

$$\hat{G} \in \arg\min_{G} \min_{m\in\mathcal{M}} \left\{ \mathrm{KL}(G\|\hat{F}) + \chi_{\mathbb{M}(\mathbf{H},\hat{F};m)}(G) \right\}, \tag{1}$$

where $\mathrm{KL}(G\|\hat{F})$ is the Kullback-Leibler divergence from G to $\hat{F}$ and $\chi_{\mathbb{M}}$ is the characteristic function of the admissible set, taking the value 0 for admissible and $+\infty$ for inadmissible. The estimation involves joint optimization over the density G and the functional parameter m.

Using duality and convex optimization arguments (e.g., Feydy et al. (2019), Yilmaz (2021)), the variational problem can be equivalently written in a saddle-point form:

$$\inf_{m\in\mathcal{M}} \inf_{G} \sup_{\boldsymbol{\lambda}\leq\mathbf{0}} \left\{ \mathrm{KL}(G\|\hat{F}) + \boldsymbol{\lambda}^{\top} \int_{\mathcal{S}} \mathbf{H}(S)m(S)g(S)dS - \mu \left(\int_{\mathcal{S}} g(S)dS - 1 \right) \right\}, \tag{2}$$

with Lagrange multipliers $\boldsymbol{\lambda} \in \mathbb{R}^{J}_{-}$ and $\mu \in \mathbb{R}$ enforcing the moment and normalization constraints, respectively. Solving the inner problem yields a closed-form expression for the optimal density $\hat{g}_m$ (for a fixed m) in the form of a Gibbs posterior:

$$\hat{g}_m(S) = \frac{\hat{f}(S)\exp\left(\boldsymbol{\lambda}_T^{\top}\mathbf{H}(S)m(S)\right)}{\int_{\mathcal{S}} \hat{f}(S)\exp\left(\boldsymbol{\lambda}_T^{\top}\mathbf{H}(S)m(S)\right)dS}, \tag{3}$$

where $\boldsymbol{\lambda}_T$ is the empirical Lagrange multiplier vector satisfying feasibility and complementary slackness. The estimator $\hat{G}$ is then given by:

$$\hat{G}(s) := \int_{S} 1_{\{\boldsymbol{x}\leq s\}}(\boldsymbol{x})\hat{g}_{m_T}(\boldsymbol{x})d\boldsymbol{x},$$

where m_T is the minimizer over $\mathcal{M}$.

The CIPDE can be viewed as a generalized Bayesian estimator: Moment inequalities replace the likelihood in Bayesian updating. The KL divergence plays the role of a regularizer that ensures proximity to the prior $\hat{F}$, given that the solution respects the moment conditions. The additional optimization over the functional parameter m creates further deviation from the classical Bayesian setting; it provides leeway in the projection procedure to choose optimally between the projection sets indexed by the functional parameter.

If an additional prior distribution were available over the functional space $\mathcal{M}$-see, for example, Li and Zhao (2002), then a density estimator could also be definable by integrating $\hat{g}_m$ w.r.t. the m-values for which $\mathbb{M}(\boldsymbol{H}, \hat{F}; m) \neq \emptyset$, via the aforementioned prior that must be supported on them. Such a procedure could then be characterized as Bayesian information projection averaging, contrasting the current one where the information projection additionally involves optimal selection of the parameter.

2.3 Pseudo-Consistency

We begin the investigation of the asymptotic properties, by establishing the pseudo-consistency of the CIPDE estimator; high-level assumptions about the convergence of the initial estimator $\hat{F}$ and the asymptotic identification of the optimal functional parameter are introduced.

Assumption 1 (Prior Convergence-Exponential Moments)**.** *There exists a density f_∞, such that $\hat{f} \rightsquigarrow f_\infty$ in $\ell^1(S)$. Furthermore, $\int_{\mathcal{S}} \exp(\boldsymbol{t}^T \boldsymbol{H}(s) m(s) f_\infty(s) ds < +\infty$ for some $\boldsymbol{t} > \mathbf{0}$, $m \in \mathcal{M}$.*

This first part of the assumption ensures that the initial prior density estimator converges in probability to a deterministic limit f_∞ inside $\ell^1(S)$. The limit may differ from the true DGP density f. $F_\infty(s) := \int_{\mathcal{S}} 1_{\{\boldsymbol{x} \leq s\}}(\boldsymbol{x}) f_\infty(\boldsymbol{x}) d\boldsymbol{x}$ need not equal the DGP cdf F. In many applications, it can be expected that $F_\infty \notin \mathbb{M}(\mathbf{H}, F_\infty; m)$ for all $m \in \mathcal{M}$ for a set of values for the conditioning variables of positive probability, even though $F \in \mathbb{M}(\mathbf{H}, F_\infty; m)$ for some SDF and that the DGP satisfies strictly the moment inequalities, because the estimator may use a counterfactual distribution shape or does not correctly take into account the conditioning information or it is incorrectly smoothed, etc.

In parametric models, the ℓ^1 convergence follows if the likelihood has a bounded derivative w.r.t. the parameter, and the parametric estimator has a unique pseudo-true value at which it converges weakly; see Blasques et al. (2018) for examples in the context of GARCH-type models. In nonparametric kernel-based density estimation, such convergence can be established under standard bandwidth and smoothness assumptions; see, for example, Tsybakov (2009). In semiparametric settings, such as sieve maximum likelihood estimation or partially linear models, uniform convergence results for $\hat{f}$ are also available when the parametric component is estimated at $\sqrt{n}$-rate and the nonparametric correction satisfies a Donsker-type condition; see, for example, Blundell et al. (2012). Thus, the first part of Assumption 1 encompasses a wide range of estimation frameworks relevant to practitioners.

The second part of Assumption 1 requires the existence of some f_∞-exponential moments for the product $\boldsymbol{H}m$, for almost every value of the conditioning variables. It is a strong condition that is, however, weaker than support boundedness. Along with the first part of the assumption, it implies the weak epi-continuity-see Knight (1999)-of the functionals $g \to \mathbb{M}(\boldsymbol{H}, g; m)$ for any g that converges to f_∞ in $\ell^1(S)$.

The second assumption is useful for asymptotic identification.

Assumption 2 (Functional Parameter Identification)**.** *If $f_\infty = f$, then there exists some $m_0 \in \mathcal{M}$, such that $\int_{\mathcal{S}} \mathbf{H}(S) m_0(S) f(S) dS > \mathbf{0}$. Otherwise, the following hold:*

1. *The set* $M^U := \cup_{m\in\mathcal{M}}\mathbb{M}(\boldsymbol{H}, F_\infty, m)$ *is convex.*
2. *The mapping* $m \mapsto G_m$*, where* $G_m := \arg\min_{G\in\mathbb{M}(\mathbf{H},F_\infty;m)} \mathrm{KL}(G\|F_\infty)$*, is injective on* M^U*.*

The first part of Assumption 2 requires the existence of a functional parameter value w.r.t. which the moment conditions are non-binding for the limiting f_∞ when this is actually the DGP. This is a mild restriction given that $\boldsymbol{H}$ can be always conveniently modified. The remaining part of the assumption ensures the identifiability of the limiting functional parameter m_∞ by leveraging the strict convexity of the KL divergence in its first argument and the convex structure of the feasible sets $\mathbb{M}(\mathbf{H}, F_\infty; m)$. The injectivity condition ensures that different m's lead to distinct projections, thereby preventing flat regions in the objective function.

The assumption implies that the composite map

$$m \to \inf_{G\in\mathbb{M}(\mathbf{H},F_\infty;m)} \mathrm{KL}(G\|F_\infty)$$

is strictly convex on M^U, and thus admits a unique minimizer m_∞. A sufficient condition for Assumption 2.1 it that there exists a $G \ll F_\infty$ which is a member of any non empty $\mathbb{M}(\boldsymbol{H}, F_\infty, m)$, A sufficient condition for the injectivity of the mapping $m \mapsto G_m := \arg\min_{G\in\mathbb{M}(\mathbf{H},F_\infty;m)} \mathrm{KL}(G\|F_\infty)$ over the set M^U is that the moment function $\mathbf{H}(S)m(S)$ is injective in m, for every S in a set of positive F_∞ probability.

The following result establishes then (pseudo) consistency for $\hat{G}$ if for some m, the set $\mathbb{M}(\mathbf{H}, F_\infty; m)$ is non empty. It derives the existence of a unique limiting and sample independent m_∞ at which the empirical parameter m_T weakly converges. The limiting density will then be an element of $\mathbb{M}(\mathbf{H}, F_\infty; m_\infty)$. If on the other hand emptyness is the case for any $\mathbb{M}(\mathbf{H}, F_\infty; m)$ then the optimization problem is asymptotically ill-posed; weak convergence to a limiting criterion that is identically equal to $+\infty$ is obtained:

Theorem 1. *Under Assumptions 1 and 2, as* $T \to \infty$*, then: a) If for some* m*,* $\mathbb{M}(\mathbf{H}, F_\infty; m) \neq \emptyset$*, then there exists a unique* $m_\infty \in \mathcal{M}$*, such that* $\hat{g}_{m_T} \rightsquigarrow g_{m_\infty}$ *in* $\ell^1(S)$*, for a unique* $G_{m_\infty} \in \mathbb{M}(\mathbf{H}, F_\infty; m_\infty)$*, with* $G_{m_\infty}(s) := \int_S 1_{\{\boldsymbol{x}\leq s}(\boldsymbol{x})g_{m_\infty}(\boldsymbol{x})d\boldsymbol{x}$*. b) If* $\mathbb{M}(\mathbf{H}, F_\infty; m) \neq \emptyset$ *for all* $m \in \mathcal{M}$*, then the optimization problem in (1) is ill-posed.*

Proof. First, notice that $\mathrm{KL}(\cdot\|\cdot)$ is lower semicontinuous due to Proposition 8 in Feydy et al. (2019) and the fact that ℓ^1 convergence of densities imply weak convergence of the underlying measures. Also due to Assumption 1, and via Skorokhod representations applicable due to Theorem 1 of Cortissoz (2007), we have $\mathbb{M}(\mathbf{H}, \hat{F}, m) \overset{\text{epi}}{\rightsquigarrow} \mathbb{M}(\mathbf{H}, F_\infty, m)$, continuously w.r.t. m, which is then directly translated to the weak epiconvergence of the indicators. Using again the Skorokhod representations it is obtained that:

$$\mathrm{KL}(\cdot\|\hat{F}) + \chi_{\mathbb{M}(\mathbf{H},\hat{F},m)}(\cdot) \overset{\mathrm{epi}}{\rightsquigarrow} \mathrm{KL}(\cdot\|F_\infty) + \chi_{\mathbb{M}(\mathbf{H},F_\infty,m)}(\cdot),$$

continuously w.r.t. m. If $\mathbb{M}(\mathbf{H}, F_\infty, m) \neq \emptyset$ for some $m \in \mathcal{M}$, then existence and uniqueness of m_∞ and subsequently of G_{m_∞} follows from the strict convexity of $\mathrm{KL}(\cdot\|F)$ on the convex $\mathbb{M}(\mathbf{H}, F_\infty, m)$ for each m for which this is non-empty as well as from the existence and the identifiability of the limiting parameter implied by Assumption 2. □

Thus, the CIPDE procedure guarantees the weak approximation of a unique limiting posterior distribution G_{m_∞} for which the moment conditions are satisfied, as long as there exists an m such that $\mathbb{M}(\boldsymbol{H}, F_\infty, m)$ is non-empty. When moreover F_∞ obeys the moment conditions, hence $F_\infty \in \mathbb{M}(\boldsymbol{H}, F_\infty, m)$, then the CIPDE procedure asymptotically recovers it. G_{m_∞} need not equal the DGP distribution F; this is the case only if $F \ll F_\infty$, and the functional parameter space $\mathcal{M}$ is well-specified.

2.4 Rate and Limiting Distribution

The results in case a) of Theorem 1 are refined to obtain standard convergence rates and the limiting distribution for the scaled discrepancy between the CIPDE density estimator and its limit.

The derivations are based on the convenient representation of the CIPDE estimator as the solution of the inner Kuhn-Tucker problem in (2) evaluated at the optimal SDF. This representation along with Theorem 1.a) already entails that $\boldsymbol{\lambda}_T(m_T)$ weakly converges to a non-sample dependent pointwise negative Lagrange multiplier $\boldsymbol{\lambda}_\infty(m_\infty)$; its j^{th} component is zero if $\int_{\mathcal{S}} \mathbf{H}_j(S)m(S)f_\infty(S)dS > 0$. Using an argument that is based on Skorokhod representations-see Knight (1999), if $\int_{\mathcal{S}} \mathbf{H}_j(S)m(S)f_\infty(S)dS > 0$, then $\boldsymbol{\lambda}_{j,T}(m)$ is eventually zero almost surely.

The following high-level assumption enables the refinement of the aforementioned limiting properties of the optimal empirical Lagrange multipliers:

Assumption 3 (Prior Rate and Weak Convergence). *For some $r_T \to \infty$, $r_T(\hat{f} - f_\infty) \rightsquigarrow \mathcal{G}$ pointwisely over S, where $\mathcal{G}$ is a process with almost surely continuous sample paths. If $J(\mathbf{H}, F_\infty; m) \neq \emptyset$, there exists a neighborhood of m_∞, such that for any m inside this, $M(\boldsymbol{H}, F_\infty, m)$ and $J(\mathbf{H}, F_\infty; m)$ are non-empty, $J(\mathbf{H}, F_\infty; m)$ is independent of m, and the matrix*

$$V_{J_m} := \int_{\mathbb{R}} \mathbf{H}_{J_m}(S)\mathbf{H}^{\mathrm{T}}_{J_m}(S)\exp(\boldsymbol{\lambda}^{\mathrm{T}}_\infty(m)m(S)\mathbf{H}(S))f_\infty(S)dS,$$

is well-defined, and has a minimum eigenvalue bounded below a positive constant that is independent of m*; there* $J_m := J(\mathbf{H}, F_\infty; m)$.

For the Empirical Cumulative Distribution Function (ECDF), the assumption can be verified via results like the ones in Doukhan et al. (1994) for stationary and strong mixing processes. For parametric models with differentiable likelihoods, it can be verified in cases where the derivative is bounded, the parameter estimator has a unique pseudo-true value and it is asymptotically Gaussian with standard rates; see again Blasques et al. for GARCH-type models. Other semi-non-parametric estimators, like the kernel-based estimators or the non-parametric MLEs may have more complicated rates that could depend on bandwidths, and/or non-Gaussian limiting distributions; see for example Ch. 24 of Van der Vaart (1998). The second part of the assumption holds whenever the J_{m_∞} components of $\mathbf{H}$ are linearly independent, and the number of binding constraints depends continuously on m locally around m_∞ at the limiting F_∞. This among others enables the invocation of implicit function arguments that ensure asymptotic smoothness of the associated Lagrange multiplier vectors w.r.t. the integral constraints.

Assumptions 1-3 and standard expansions then suffice for the derivation of the limiting behavior of the random element $(\boldsymbol{\xi}_T, \boldsymbol{\zeta}_t)$ comprised of the translated and rescaled by r_T components $\boldsymbol{\xi}_T := r_T(\boldsymbol{\lambda}_T(m_T) - \boldsymbol{\lambda}_\infty(m_\infty))$, and $\boldsymbol{\zeta}_T := r_T(m_T - m_\infty)$. This is summarized in the following auxiliary lemma, which is in turn useful for the derivation of the rates and the limiting distribution of the CIPDE estimator:

Lemma 1. *Suppose that Assumptions 1-3 hold. Then, on* $\ell_\infty(\mathbb{R}^{J_{m_\infty}} \times \mathcal{S})$,

$$(\boldsymbol{\xi}_T, \boldsymbol{\zeta}_T) \rightsquigarrow (\boldsymbol{\xi}_\infty, \boldsymbol{\zeta}_\infty), \tag{4}$$

for

$$\boldsymbol{\xi}_\infty := \begin{cases} \arg\min_{\boldsymbol{\xi} \in \mathcal{H}} \|\boldsymbol{\xi} + V_{J_{m_\infty}}^{-1} \boldsymbol{z}(\mathcal{G}, \boldsymbol{\lambda}_\infty, m_\infty)\|^2_{V_{J_{m_\infty}}}, \text{on } J_{m_\infty} \\ \mathbf{0}, \text{on } \{1, 2, \ldots, J\} - J_{m_\infty}, \end{cases}$$

$$\boldsymbol{z}(\mathcal{G}, \boldsymbol{\lambda}_\infty, m_\infty) := \int_{\mathbb{R}} \mathbf{H}_{J_{m_\infty}}(S) \mathcal{G}^\star(S) dS,$$

$\mathcal{G}^\star(S) := \exp(\boldsymbol{\lambda}_\infty^{\mathrm{T}}(m_\infty) m_\infty(S) \mathbf{H}(S)) \mathcal{G}(S)$, $\mathcal{H} := \prod_{j \in J_{m_\infty}} \mathcal{H}_j$, *with*

$\mathcal{H}_j := \begin{cases} \mathbb{R}_-, & \boldsymbol{\lambda}_{j,\infty}(m_\infty) = 0 \\ \mathbb{R}, & \boldsymbol{\lambda}_{j,\infty}(m_\infty) < 0 \end{cases}$, $\|\boldsymbol{u}\|^2_A := \boldsymbol{u}^{\mathrm{T}} A \boldsymbol{u}$, *while*

$$\boldsymbol{\zeta}_\infty := \arg \min_{\boldsymbol{\zeta} \in \mathcal{H}_{m_\infty}} \boldsymbol{Z}(\boldsymbol{\zeta})$$

where,

$$\boldsymbol{Z}(\boldsymbol{\zeta}) := \left(\boldsymbol{\lambda}_\infty(m_\infty) + D_{\lambda_\infty}\left[\int_{\mathcal{S}} \boldsymbol{H}(S)\,\boldsymbol{\zeta}(S) f_\infty(S)\,dS\right]\right)^T \int_{\mathcal{S}} \boldsymbol{H}(S)\,\boldsymbol{\zeta}(S)\,\mathcal{G}^\star(S)\,dS + \frac{1}{2}\int_{\mathcal{S}} \boldsymbol{\kappa}(S)^2 g_{m_\infty}(S)\,dS,$$

for $\boldsymbol{\kappa}(S) := \boldsymbol{\lambda}_\infty^T(m_\infty)\,\boldsymbol{H}(S)\,\boldsymbol{\zeta}(S) + \left(D_{\lambda_\infty}\left[\int_{\mathcal{S}} \boldsymbol{H}(S)\,\boldsymbol{\zeta}(S) f_\infty(S)\,dS\right]\right)^T \boldsymbol{H}(S)\,m_\infty(S) + \boldsymbol{\xi}_\infty^T \boldsymbol{H}(S)\,m_\infty(S)$, *with* D_{λ_∞} *denoting the derivative of the limiting vector of Lagrange multipliers w.r.t. the integral constarints, and* $\mathcal{H}_{m_\infty}$ *denotes the convex cone obtained as the Painleve-Kuratowski limit of* $r_T(\mathcal{M} - m^\star)$.

Proof. Duality implies that for each m, the optimal Lagrange multiplier solves the optimization problem $\min_{\boldsymbol{\lambda}\leq \mathbf{0}} \int_{\mathbb{R}} \exp(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{H}(S) m_T(S)) d\hat{F}$. Assumption 1 implies that the objective converges locally uniformly over the multiplier, and continuously w.r.t. m_∞ in probability to the limiting criterion $\int_{\mathbb{R}} \exp(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{H}(S) m_\infty(S)) d\hat{F}_\infty$, while Assumption 3 implies that the latter is strictly convex. The previous imply the existence of the limiting multiplier $\boldsymbol{\lambda}_\infty(m_\infty)$ as the unique solution to the asymptotic problem $\min_{\boldsymbol{\lambda}\leq \mathbf{0}} \int_{\mathbb{R}} \exp(\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{H}(S) m_\infty(S)) dF_\infty$. Consider now the rescaled and translated problem:

$$r_T^2\left(\int_{\mathcal{S}} \exp\left(\boldsymbol{\lambda}_T^{\mathrm{T}}\boldsymbol{H}(S)\,m_T(S)\right)\hat{f}(S)\,dS - \int_{\mathcal{S}} \exp\left(\boldsymbol{\lambda}_\infty^{\mathrm{T}}(m_\infty)\,\boldsymbol{H}(S)\,m_\infty(S)\right)\hat{f}(S)\,dS\right),$$

which can be expanded as

$$= r_T^2 \int_{\mathcal{S}} \exp\left(\boldsymbol{\lambda}_\infty^{\mathrm{T}}(m_\infty)\,\boldsymbol{H}(S)\,m_T(S)\right)\left(1 + \frac{1}{r_T}\boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{H}(S)\,m_T(S) + \frac{1}{r_T^2}\left(\boldsymbol{\xi}^{\mathrm{T}}\boldsymbol{H}(S)\,m_T(S)\right)^2\right)\hat{f}(S)\,dS$$

$$- r_T^2 \int_{\mathcal{S}} \exp\left(\boldsymbol{\lambda}_\infty^{\mathrm{T}}(m_\infty)\,\boldsymbol{H}(S)\,m_\infty(S)\right)\hat{f}(S)\,dS,$$

which weakly converges locally uniformly in $\boldsymbol{\xi}$ to the strictly convex $\boldsymbol{\xi}^{\mathrm{T}}\int_{\mathcal{S}} \boldsymbol{H}(S)\,m_\infty(S)\,\mathcal{G}^\star(S) dS + \frac{1}{2}\boldsymbol{\xi}^{\mathrm{T}}\int_{\mathcal{S}} \boldsymbol{H}(S)\,\boldsymbol{H}^{\mathrm{T}}(S)\,m_\infty^2(S)\,g_{m_\infty}(S)\,dS\boldsymbol{\xi}$ establishing the first result. An analogous expansion w.r.t. $\boldsymbol{\zeta}$ produces

$$\begin{gathered} r_T\boldsymbol{\lambda}_\infty^{\mathrm{T}}(m_\infty)\int_{\mathcal{S}} \exp\left(\boldsymbol{\lambda}_\infty^{\mathrm{T}}(m_\infty)\,\boldsymbol{H}(S)\,m_\infty(S)\right)\boldsymbol{H}(S)\,\boldsymbol{\zeta}(S)\,\hat{f}(S)\,dS + \\ r_T\int_{\mathcal{S}} \exp\left(\boldsymbol{\lambda}_\infty^{\mathrm{T}}(m_\infty)\,\boldsymbol{H}(S)\,m_\infty(S)\right) D_{\lambda_T}\int_{\mathcal{S}} \boldsymbol{H}(S)\,\boldsymbol{\zeta}(S)\,\hat{f}(S)\,dS\boldsymbol{H}(S)\,m_\infty(S)\,\hat{f}(S)\,dS + \\ \int_{\mathcal{S}} \exp\left(\boldsymbol{\lambda}_\infty^{\mathrm{T}}(m_\infty)\,\boldsymbol{H}(S)\,m_\infty(S)\right) D_{\lambda_T}\int_{\mathcal{S}} \boldsymbol{H}(S)\,\boldsymbol{\zeta}(S)\,\hat{f}(S)\,dS\boldsymbol{H}(S)\,\boldsymbol{\zeta}(S)\,\hat{f}(S)\,dS + \\ \int_{\mathcal{S}} \exp\left(\boldsymbol{\lambda}_\infty^{\mathrm{T}}(m_\infty)\,\boldsymbol{H}(S)\,m_\infty(S)\right)\boldsymbol{\xi}_T\boldsymbol{H}(S)\,\boldsymbol{\zeta}(S)\,\hat{f}(S)\,dS + \\ \int_{\mathcal{S}} \exp\left(\boldsymbol{\lambda}_\infty^{\mathrm{T}}(m_\infty)\,\boldsymbol{H}(S)\,m_\infty(S)\right)\frac{1}{2}\left(k_T(S)\right)^2\hat{f}(S)\,dS, \end{gathered}$$

where

$$\begin{gathered} k_T(S) := \boldsymbol{\lambda}_\infty^{\mathrm{T}}\left(m_\infty\right) \boldsymbol{H}\left(S\right) \boldsymbol{\zeta}\left(S\right) \\ +D_{\lambda_T} \int_{\mathcal{S}} \boldsymbol{H}\left(S\right) \boldsymbol{\zeta}\left(S\right) \hat{f}\left(S\right) dS \boldsymbol{H}\left(S\right) m_\infty\left(S\right) + a\boldsymbol{\xi}_T^{\mathrm{T}} \boldsymbol{H}\left(S\right) m_\infty\left(S\right), \end{gathered}$$

while differentiability of the optimal limiting Lagrange multiplier follows from Assumptions 1-3, and the Implicit Function Theorem in Banach spaces, see for example Rockafellar and Wets (1998). The previous expansion can be similarly seen to converge, locally uniformly w.r.t. $\boldsymbol{\zeta}$, to

$$\begin{gathered} \left(\boldsymbol{\lambda}_\infty\left(m_\infty\right) + K\right)^{\mathrm{T}} \int_{\mathcal{S}} \boldsymbol{H}\left(S\right) \boldsymbol{\zeta}\left(S\right) \mathcal{G}^\star\left(S\right) dS + \\ \tfrac{1}{2} \int_{\mathcal{S}} \left[\boldsymbol{\lambda}_\infty^{\mathrm{T}}\left(m_\infty\right) \boldsymbol{H}\left(S\right) \boldsymbol{\zeta}\left(S\right) + K^{\mathrm{T}} \boldsymbol{H}\left(S\right) m_\infty\left(S\right) + \boldsymbol{\xi}_\infty^{\mathrm{T}} \boldsymbol{H}\left(S\right) m_\infty\left(S\right)\right]^2 g_{m_\infty}\left(S\right) dS, \end{gathered}$$

where $K := D_{\lambda_\infty}\left[\int_{\mathcal{S}} \boldsymbol{H}\left(S\right) \boldsymbol{\zeta}\left(S\right) f_\infty\left(S\right) dS\right]$, due to that under Assumptions 1 and 3, the following term converges to zero in probability as $T \to \infty$:

$$\boldsymbol{\lambda}_\infty^{\mathrm{T}}(m_\infty) \int_{\mathcal{S}} \boldsymbol{H}(S) \boldsymbol{\zeta}(S) \exp\left(\boldsymbol{\lambda}_\infty^{\mathrm{T}} \boldsymbol{H}(S) m_\infty(S)\right) r_T \hat{f}(S) dS.$$

This last result holds because $\boldsymbol{\lambda}_\infty(m_\infty)$ satisfies the first-order condition of the limiting dual problem:

$$\int_{\mathcal{S}} \boldsymbol{H}(S) m_\infty(S) \exp\left(\boldsymbol{\lambda}_\infty^{\mathrm{T}} \boldsymbol{H}(S) m_\infty(S)\right) f_\infty(S) dS = \mathbf{0},$$

implying that first-order deviations in the moment function are asymptotically orthogonal to the gradient direction. The limiting criterion is then strictly convex due to that the integrand of the second term is a strictly convex quadratic form in $\boldsymbol{\zeta}$, provided that $g_{m_\infty}(S) > 0$ almost everywhere and $\boldsymbol{H}(S)$ has full rank on the support $\mathcal{S}$. The first term is linear in $\boldsymbol{\zeta}$. Therefore, $\boldsymbol{Z}(\boldsymbol{\zeta})$ is a strictly convex functional.

Also, the feasible set $\mathcal{H}_{m_\infty}$, defined as the Painlevé–Kuratowski limit of $r_T(\mathcal{M} - m_\infty)$, is a convex cone. Hence, the strict convexity of the objective over a convex domain implies the existence and uniqueness of the minimizer. The previous establish the second result. □

Then the required limit theory of the CIPDE density estimator is readily obtained from (3) and the Delta method; the following theorem describes it:

Theorem 2. *Suppose that Assumptions 1-3 hold, and that* $\mathbb{M}(\mathbf{H}, F_\infty; m) \neq \emptyset$ *for some* $m \in \mathcal{M}$. *a. If* $J_{m_\infty} = \emptyset$, *then in* $\ell^\infty(\mathcal{S})$,

$$r_T(\hat{g} - f_\infty) \rightsquigarrow \mathcal{G}. \tag{5}$$

b. If $J_{m_\infty} \neq \emptyset$, then in $\ell^\infty(\mathcal{S})$,

$$r_T(\hat{g} - g_{m_\infty}) \rightsquigarrow \mathcal{G}_\infty, \tag{6}$$

where

$$\mathcal{G}_\infty(S) := \frac{\boldsymbol{\Lambda}\left(S\right) + g_{m_\infty}\left(S\right)\int_{\mathcal{S}} \boldsymbol{\Lambda}\left(S\right) dS}{\int_{\mathcal{S}} \exp\left(\boldsymbol{\lambda}_\infty^T\left(m_\infty\right) \boldsymbol{H}\left(S\right) m_\infty\left(S\right)\right) f_\infty\left(S\right) dS},$$

$$\boldsymbol{\Lambda}\left(S\right) := \mathcal{G}^\star\left(S\right) + \left(\boldsymbol{\xi}_\infty^T\left(S\right) m_\infty\left(S\right) + \boldsymbol{\lambda}_\infty^T\left(m_\infty\right) \boldsymbol{\zeta}_\infty\left(S\right)\right) \boldsymbol{H}\left(S\right) g_{m_\infty}\left(S\right),$$

and $\boldsymbol{\xi}_\infty$, $\boldsymbol{\zeta}_\infty$, $\mathcal{G}^\star$, as in Lemma 1.

Proof. The first case follows from the eventual nullification of $\boldsymbol{\lambda}_T$ w.h.p. and Assumption 3. The second case follows from the fact that Assumptions 1-3 along with Lemma 1, imply that $\begin{pmatrix} r_T(f_T - f_\infty) \\ \boldsymbol{\xi}_T \\ \boldsymbol{\zeta}_T \end{pmatrix} \rightsquigarrow \begin{pmatrix} \mathcal{G} \\ \boldsymbol{\xi}_\infty \\ \boldsymbol{\zeta}_\infty \end{pmatrix}$, the fact that $\hat{g}_{m_T}(S) = \frac{\hat{f}(S)\exp\left(\boldsymbol{\lambda}_T^\top \mathbf{H}(S) m_T(S)\right)}{\int_{\mathcal{S}} \hat{f}(S)\exp\left(\boldsymbol{\lambda}_T^\top \mathbf{H}(S) m_T(S)\right) dS}$, $\hat{g}_{m_\infty}(S) = \frac{f(S)\exp\left(\boldsymbol{\lambda}_\infty^\top \mathbf{H}(S) m_\infty(S)\right)}{\int_{\mathcal{S}} f_\infty(S)\exp\left(\boldsymbol{\lambda}_\infty^\top \mathbf{H}(S) m_\infty(S)\right) dS}$, and two successive applications of the Delta method. □

Theorem 2 specifies a rate of convergence, inherited by the limit theory of the prior, and a limiting distribution for the translated conditional density estimator of F. In all the considered cases the CIPDE estimator is not asymptotically independent of the limiting prior; this result is not surprising because the KL inconsistency reduction obtained by the CIPDE is asymptotically defined by restrictions that depend on the limiting prior F_∞.

Whenever $J_{m_\infty} = \emptyset$, and thus F_∞ strictly satisfies the moment conditions, the eventual nullification of the multiplier implies that the limit theory of the initial estimator is recovered. Due to the first part of Assumption 2 this encompasses the case where the prior is consistent, i.e. $f_\infty = f$.

When the limiting Lagrange multipliers are non-zero, i.e., $J_{m_\infty} \neq \emptyset$, the CIPDE estimator incorporates binding moment inequality constraints. These constraints act as a source of side information, effectively restricting the estimator to a lower-dimensional feasible set. The resulting limiting distribution $\mathcal{G}_\infty$ is a projection of the unconstrained limit process $\mathcal{G}$ onto a constrained tangent cone, modulated by the Lagrange multipliers and the local geometry of the feasible set (via the influence functions $\boldsymbol{\xi}_\infty$ and $\boldsymbol{\zeta}_\infty$). As is well known from the theory of constrained M-estimation and semiparametric efficiency bounds, such projections reduce asymptotic variance in directions aligned with the active constraints. Hence, the CIPDE estimator may achieve variance reduction-i.e., efficiency gains-relative to the unconstrained estimator

$\hat{f}$, provided that the constraints are informative. This interpretation aligns with results in Andrews (1999), van der Vaart (1998), and Bickel et al. (1998), and formalizes the intuition that valid economic restrictions, when binding, yield more precise inference.

In the special case where $r_T = \sqrt{T}$, $\mathcal{G}$ is zero mean Gaussian, $J_{m_\infty} \neq \emptyset$, $\boldsymbol{\lambda}_{j,\infty} < 0$, $\forall j \in J_{m_\infty}$, and for any $S \in \mathcal{S}$, $\cup_{\boldsymbol{\zeta} \in \mathcal{H}_{m_\infty}} \{\boldsymbol{\zeta}(S)\} = \mathbb{R}$, then $\mathcal{G}_\infty$ is a zero mean Gaussian process, and the likelihood ratio tests performed in Section 5.5, can be proven to be asymptotically valid.

2.5 Partial Inconsistency Correction

Theorem 1 provides limited information regarding the relation between the limiting G and the true F. If $\mathbb{M}(\mathbf{H}, F_\infty; m) = \{F\}$ for some $m \in \mathcal{M}$, then $G = F$, and $\hat{G}$ is then a weakly consistent estimator of the DGP distribution. However, in many applications, F cannot be expected to be the single element of $\mathbb{M}(\mathbf{H}, F_\infty; m)$ for any premissible SDF. For instance, when F satisfies moment inequalities with strict inequality, there will typically be other elements nearby it.

The result below, utilizing the Pythagorean Theorem within information geometry (refer to Nielsen (2020) for instance), asserts that $\hat{G}$ effectively diminishes the asymptotic divergence to F in scenarios where F_∞ fails to meet the moment inequalities:

Theorem 3. *Under Assumption 1, and if $F \in \mathbb{M}(\mathbf{H}, F_\infty; m)$ for some $m \in \mathcal{M}$, then:*

$$\mathrm{KL}(F \| G) = \mathrm{KL}(F \| F_\infty) - \mathrm{KL}(G \| F_\infty). \tag{7}$$

Proof. If $F = F_\infty$ or $G = F_\infty$ the result follows trivially by Theorem 1. If $F \neq F_\infty$, and $G \neq F_\infty$, then Theorem 1 along with the CIPDE estimator representation in (3) implies that there exists a limiting non-sample dependent, pointwise non-positive and non identically zero multiplier $\boldsymbol{\lambda}_\infty(m_\infty)$, such that $g_{m_\infty}(S) = \frac{f_\infty(S)\exp(\boldsymbol{\lambda}_\infty^{\mathrm{T}}(m_\infty)m_\infty(S)\mathbf{H}(S))}{\int_{\mathcal{S}} f_\infty(S)\exp(\boldsymbol{\lambda}_\infty^{\mathrm{T}}(m_\infty)m_\infty(S)\mathbf{H}(S))dS}$. The second part of Theorem 3.1 in Csiszár (1975) then implies the result. □

Equation (7) has a nontrivial information-theoretic interpretation when $F_\infty \notin \mathbb{M}(\mathbf{H}, F_\infty; m_\infty)$: the information lost by approximating the DGP F by the limiting posterior G, compared to the information lost by approximating F by the limiting prior F_∞ is reduced by the non-negative quantity $\mathrm{KL}(G \| F_\infty) \geq 0$, that is, the information gained by using moment inequalities to update the inference to the limiting posterior G, instead of using the moment conditions ignorant limiting prior.

Hence, whenever F_∞ violates the moment conditions, the CIPDE density estimator performs asymptotically a (partial) KL inconsistency correction.

2.6 Partial Inconsistency Correction under Approximate Feasibility

In practice, F may not exactly satisfy the pricing-implied moment inequalities required by the feasible set; $F \notin \mathbb{M}(\boldsymbol{H}, F_\infty, m)$, $\forall m \in \mathcal{M}$. Notably, asset pricing restrictions may be violated if the stock market or option market is out of equilibrium or if the observed prices are not fully synchronized. This motivates the study of cases of approximate feasibility where F violates the constraints only by a small amount at least for some m. The question then becomes whether CIPDE still asymptotically yields a meaningful Kullback-Liebler correction towards F.

Suppose that the limiting prior F_∞ is both different from F and does not satisfy the moment inequalities, while additionally F is only approximately feasible:

$$\inf_{j \in J_{m_\infty}} \int_{\mathcal{S}} \boldsymbol{H}_j(S) m_\infty(S) f(S) dS = -\delta, \quad \text{for some small } \delta > 0,$$

thus the Pythagorean decomposition of KL divergence does not necessarily hold exactly. However, it remains plausible to ask whether the limiting CIPDE still satisfies:

$$\mathrm{KL}(F \| G) < \mathrm{KL}(F \| F_\infty),$$

thus constituting an asymptotic partial KL-correction on $\hat{F}$. This is formalized in the following result:

Proposition 1 (Partial KL Correction under Approximate Feasibility)**.** *Under Assumption 1, suppose also that:*

(a) $F \notin \mathbb{M}(\boldsymbol{H}, F_\infty, m)$ *for all* $m \in \mathcal{M}$,

(b) the following exponential moment condition holds:

$$\int_{\mathcal{S}} \exp\left(\boldsymbol{\lambda}_\infty^{\mathrm{T}} \boldsymbol{H}(S) m_\infty(S)\right) f_\infty(S)\, dS > \exp\left(|J_{m_\infty}| \cdot \sup_{j \in J^\star} |\boldsymbol{\lambda}_{\infty,j}| \cdot \delta\right),$$

where

$$\delta := -\inf_{j \in J_{m_\infty}} \int_{\mathcal{S}} \boldsymbol{H}_j(S) m_\infty(S) f(S)\, dS,$$

$$J^\star := \arg\min_{j \in J_{m_\infty}} \int_{\mathcal{S}} \boldsymbol{H}_j(S) m_\infty(S) f(S)\, dS.$$

Then,

$$\mathrm{KL}(F\|G) < \mathrm{KL}(F\|F_\infty).$$

Proof. Using elementary calculus, we obtain the KL decomposition:

$$\mathrm{KL}(F\|F_\infty) = \mathrm{KL}(F\|G) + \mathrm{KL}(G\|F_\infty) + \int_{\mathcal{S}} \left(\log g(S) - \log f_\infty(S)\right)\left(f(S) - g(S)\right)\, dS.$$

Rearranging the expression with respect to $\mathrm{KL}(F\|G)$, we get:

$$\begin{aligned}\mathrm{KL}(F\|G) = {} & \mathrm{KL}(F\|F_\infty) - \mathrm{KL}(G\|F_\infty) \\ & - \int_{\mathcal{S}} \left(\log g(S) - \log f_\infty(S)\right)\left(f(S) - g(S)\right)\, dS.\end{aligned}$$

By the first-order condition of the limiting dual problem, the cross term equals:

$$\begin{aligned}\int_{\mathcal{S}} \left(\log g(S) - \log f_\infty(S)\right)\left(f(S) - g(S)\right)\, dS = {} & \log\left(\int_{\mathcal{S}} \exp\left(\boldsymbol{\lambda}_\infty^{\mathrm{T}} \boldsymbol{H}(S) m_\infty(S)\right) f_\infty(S)\, dS\right) \\ & - \int_{\mathcal{S}} \boldsymbol{\lambda}_\infty^{\mathrm{T}} \boldsymbol{H}(S) m_\infty(S) f(S)\, dS.\end{aligned}$$

The second term above can be bounded using the worst-case component in the set J_{m_∞}:

$$\int_{\mathcal{S}} \boldsymbol{\lambda}_\infty^{\mathrm{T}} \boldsymbol{H}(S) m_\infty(S) f(S)\, dS \le |J_{m_\infty}| \cdot \sup_{j \in J^\star} |\boldsymbol{\lambda}_{\infty,j}| \cdot \delta.$$

Therefore, we obtain the inequality:

$$\mathrm{KL}(F\|G) \le \mathrm{KL}(F\|F_\infty) - \mathrm{KL}(G\|F_\infty) + |J_{m_\infty}| \cdot \sup_{j \in J^\star} |\boldsymbol{\lambda}_{\infty,j}| \cdot \delta.$$

Hence, partial correction is ensured whenever:

$$\mathrm{KL}(G\|F_\infty) > |J_{m_\infty}| \cdot \sup_{j \in J^\star} |\boldsymbol{\lambda}_{\infty,j}| \cdot \delta.$$

Finally, by the dual representation of the projection G, we have:

$$\mathrm{KL}(G\|F_\infty) = \log \int_{\mathcal{S}} \exp\left(\boldsymbol{\lambda}_\infty^{\mathrm{T}} \boldsymbol{H}(S) m_\infty(S)\right) f_\infty(S)\, dS,$$

which completes the proof. □

The above result shows that even when the moment conditions fail to hold exactly, the CIPDE projection may still produce a KL-partially corrected update, provided the dual potential (expressed through the exponential moment) outweighs the severity of the moment violations. Intuitively, this implies that the model's "pull" through the dual weighting structure can still dominate minor economic inconsistencies — a robustness property particularly valuable in contexts where asset pricing restrictions rely on data from disequilibrium or unsynchronized sources. The proposition formalizes a sufficiency threshold: if the prior is bad enough and feasibility violations are small enough, the update is justified.

The bound involving $|J_{m_\infty}| \cdot \sup_{j \in J^\star} |\boldsymbol{\lambda}_{\infty,j}| \cdot \delta$ is conservative. A refined inequality could use Jensen's inequality or concentration bounds to exploit curvature in the exponential term, especially if the vector $\boldsymbol{\lambda}_\infty$ is sparse or structured.

3 Option Pricing and Option Combinations

The CIPDE is applied in this study to stock index returns using moment conditions based on an option pricing system in the spirit of Constantinides et al. (2009), and optimization of combinations of index options in the spirit of Constantinides et al. (2020); Post and Longarela (2021); Beare et al. (2025). This approach is applied both to randomly generated data sets from a known DGP in a simulation experiment (Section 4) and to historical monthly SPX returns and price quotes of CBOE SPX options in an empirical application (Section 5).

Constantinides et al. (2009) develop a system of model-free pricing restrictions for multiple concurrent options with different option types (put or call) and different strike prices. These restrictions are quite general because they don't assume that the option market is complete or perfect, and they furthermore don't assume a specific functional form for the SDF. Still, the truth of these restrictions may depend on the characteristics of the selected option series and the prevailing market conditions.

Violations of the equilibrium system, or pricing errors, imply that the market index is inefficient in the sense of being dominated by second-degree SD by portfolios that are enhanced using properly constructed option combinations. Constantinides et al. (2020); Post and Longarela (2021); Beare et al. (2025) develop and apply optimization problems to construct such enhanced portfolios. These studies do not find robust evidence for mispricing in the form of significant out-of-sample out-performance for one-month options, consistent with the notion that the price quotes reflect the latent conditional return distribution.

A limitation of these studies is that they do not account for time-variation of the shape of the distribution and risk premiums which generally results in statistical inconsistency of the conditional distribution estimator. For instance, the empirical analysis by Constantinides et al. (2009) is based on a scaled version of the *unconditional* ECDF, which uses a fixed shape, fixed Market Risk Premium and fixed Variance Risk Premium. Estimation error for the conditional distribution in turn introduces the risk of false rejection of options market efficiency and poor OOS performance of optimized option combinations.

The present section discusses the use of CIPDE in this application area. We analyze the statistical significance of the improvements in OOS forecasting ability using a Likelihood Ratio Test (LRT). We analyze the economic significance using the OOS investment performance of option combinations that are optimized under the posterior density ($\hat{G}$).

We deliberately deviate from the aforementioned studies in two subtle ways. First, the equilibrium system of Constantinides et al. (2009) is tightened by requiring that the SDF is log-convex, which amounts to refining second-degree SD to Decreasing Absolute Risk Aversion SD (DSD) ((Vickson (1975), Bawa (1975))). Second, the optimization problem is based on a notion of approximate dominance because the posterior estimates (which by construction obey the equilibrium system) exclude exact dominance (if the pricing restrictions are imposed for all option series included in the optimization). These issues will be discussed more detail in the relevant subsections below.

To illustrate the importance of estimating the physical distribution and modeling the SDF, we also estimate the risk-neutral distribution by applying CIPDE to a constant SDF, which resembles the methodology by Stutzer (1996). Although the risk-neutral distribution by construction is well-suited for valuing options, it is less suitable for forecasting and optimization due to its biased nature.

3.1 Preliminaries

The focus is on European-type stock index options. A total of N distinct option series are considered with different option types (put or call) and different strike prices, $K_i, i = 1, \cdots, N$. All options have the same time to expiry (T) and their expiration date equals the forecast horizon.

The annualized risk-free rate on a maturity-matched Treasury bill is R_B. The current index value and the index value at expiry date are S_0 and $S_T = S_0(1 + R_S)$, respectively. The (annualized and maturity-matched) dividend yield to the index is R_D.

The price return R_S is treated as a random variable with latent distribution F that is estimated with prior $\hat{F}$ and posterior $\hat{G}$. The specification of the prior will be discussed in more detail in Section 4 (Monte Carlo simulation) and Section 5 (empirical study).

The option payoffs are $X_{i,T} = P_{i,T} := (K_i - S_T)_+$ for puts and $X_{i,T} = C_{i,T} := (S_T - K_i)_+$ for calls, $i = 1, \cdots, N$. The payoffs of all options are driven exclusively by the index price return R_S, so that a perfect single-factor structure arises for the payoffs of all options.

The Present Values are denoted by $X_{i,0} = P_{i,0}$ for puts and $X_{i,0} = C_{i,0}$ for calls, respectively. Whereas the Present Values are latent, the quoted ask price a_i and quoted bid price $b_i \geq a_i$ are observable. In addition, if the prices are efficient, then the ranking $b_i \geq X_{i,0} \geq a_i$ is obtained.

Dynamic replication or hedging of options using combinations of bills and stocks is not considered because it is challenging to estimate the entire dynamic process and the relevant transactions costs of portfolio rebalancing. Instead, the investment universe consists of combinations of bills, stocks and options that are held until the option expiry date.

Buying or selling bills and stocks is assumed to involve proportional transaction costs of 0.1%. Options can be bought at the quoted ask price a_i and written at the quoted bid price b_i. The profit or loss at the expiration date is given by $(X_{i,T} - a_i(1 + R_B))$ for a bought option and $(b_i(1 + R_B) - X_{i,T})$ for a written option.

The focus is on one-month options. For this maturity segment, a relatively long time-series, broad cross-section of liquid options and rich research literature is available for SPX options. In addition, the pricing restrictions that are used as moment conditions are more plausible for one-month options than for shorter-dated options, given the evidence in Constantinides et al. (2020); Post and Longarela (2021); Beare et al. (2025) about the performance of active trading strategies in various maturity segments.

3.2 Option pricing conditions

To update the prior $\hat{F}$ to a posterior $\hat{G}$, CIPDE uses conditioning information in the form of conditional moment conditions. It is important that the moment conditions are not controversial because they are not tested but used to incorporate side information, in this study. For this reason, we eschew parametric option pricing formulas, fully-specified asset pricing theories or models that assume a complete or perfect option market.

Instead, we employ and extend a system of model-free pricing restrictions based on Constantinides et al. (2009). In arbitrage-free equilibrium, an SDF for pricing cash flows at the option expiration date, $m : S_T \to \mathbb{R}$, exists that is consistent with the prevailing market prices of the securities (index, bill and options):

$$0.999 \leq \mathbb{E}_F[m(1+R_S)] \leq 1.001; \tag{8}$$

$$0.999 \leq \mathbb{E}_F[m(1+R_B)] \leq 1.001; \tag{9}$$

$$b_i \leq \mathbb{E}_F[mX_{i,T}] \leq a_i;\ i = 1, \cdots, N. \tag{10}$$

Attractively, these pricing restrictions do not assume that options can be replicated without costs using dynamic combinations of bills and stocks, and they furthermore accounts for the relatively large bid-ask spread for index options. In addition, the restrictions are imposed in the empirical application only for options that pass our tight filters based on maturity, moneyness, option delta and option premium (see Section 5.1).

The SDF is generally not unique under these conditions, and the estimation benefits from imposing structure on the SDF to prevent overfitting it to a misspecified prior $\hat{F}$. The SDF is assumed to represent the Intertemporal Marginal Rate of Substitution (IMRS) of index investors and it is partially identified by a set of functions that obey standard regularity conditions for the IMRS:

$$\left\{m \in \mathcal{C}^2 : m(S_T) \geq 0; m'(S_T) \leq 0; ln(m(S_T))'' \geq 0\right\} =: \mathcal{M} \ni m. \tag{11}$$

These SDFs are positive, increasing and log-convex, as required for the IMRS of standard utility functions. Although Constantinides et al. (2009) do not require this property, convexity directly follows from the generally accepted property of DARA for standard utility functions (Kimball (1990)). Violations of the equilibrium system (8)-(11), or pricing errors, imply that the market index is dominated by DSD by portfolios that enhance the index using certain option combinations.

If the log-convexity condition is not imposed, as in Constantinides et al. (2009), the analysis allows for SDFs with a pathological shape such an implausible reverse S-shape with increasing risk aversion followed by decreasing risk aversion. As a result, pricing errors occur less frequently under the prior estimate; the posterior density will diverge less from the prior density if pricing errors do occur; and the evidence for OOS forecasting success and investment outperformance weakens. Similarly, Basso and Pianca (1975) show that general n-th degree SD allow for financial option prices that are inconsistent with DARA.

The risk-neutral distribution (Q) can be estimated using the same pricing conditions (((8))-((10))) by restricting the SDF to be constant: $m(S) = (1 + R_B)^{-1}$. This approach naturally introduces a pessimistic bias for estimating the physical distribution which may weaken forecasting success and investment performance.

The pricing restrictions (8)-(11) are conditional moment conditions because the probability distribution and option prices are updated every month. The truth of the moment conditions is not explicitly tested in this study. In the simulation study, the moment conditions are known to be true. In the empirical study, we select options for which the moment conditions appear plausible. In addition, the truth of the moment conditions (for the selected options) can be tested indirectly based on the OOS forecasting and investment success.

The CIPDE relies on the empirical counterparts of the moment conditions (8)-(11) based on $\mathbb{E}_{\hat{G}}[m(1+R_S)]$, $\mathbb{E}_{\hat{G}}[m(1+R_B)]$, and $\mathbb{E}_{\hat{G}}[mX_{i,T}]$, $i = 1, \cdots, N$. These empirical conditions are expected to be effective because specification error and estimation error is likely to lead to violations of these conditions for at least some of the option series.

If the density estimator is discretized using a finite set of atoms $\{S_{j,T}\}$, the condition $m \in \mathcal{M}$ can be discretized and linearized to a finite set of linear restrictions on the SDF values $\{m(S_{j,T})\}$. In this study, we discretize the density estimators using an equally spaced grid in the range[-0.6,0.6] with 25 bps (0.0025) grid size, resulting in 481 atoms. After discretization and linearization, CIPDE becomes a Bi-convex Programming problem that is convex in the SDF values $m(S_{j,T})$ and convex in the probability weights $\hat{g}(S_{j,T})$. The present study solves the problem using the two-step optimization procedure of Post (2003).

3.3 Likelihood Ratio Test

To evaluate the OOS forecasting ability, we apply the LRT of Vuong (1989), as per Amisano and Giacomini (2007).

$$\mathrm{WLR}_{t+1}(\hat{F}, \hat{G}) := \Big(\log \hat{f}_t(S_{t+1}) - \log \hat{g}_t(S_{t+1})\Big)\,; \tag{12}$$

$$\mathrm{LRT}(\hat{F}, \hat{G}) := \frac{n^{-1}\sum_{t=1}^{n} \mathrm{WLR}_{t+1}(\hat{F}, \hat{G})}{\sqrt{n^{-1}\sum_{t=1}^{n} \mathrm{WLR}_{t+1}(\hat{F}, \hat{G})^2}/\sqrt{n}}. \tag{13}$$

where $t = 1$ refers to the month that the first one-step-ahead OOS forecast is produced and n is the number of forecasts evaluated.

As per Amisano and Giacomini (2007), the numerator does not include a correction for serial dependence because serial dependence is relatively weak for monthly stock index returns. Ljung-Box tests and autocorrelation function analysis of the log likelihood ratios, $\mathrm{WLR}_{t+1}(\hat{F}, \hat{G})$, do not detect significant serial autocorrelation over our sample of 321 months. Short truncation lags in correcting for serial dependence in the denominator of Eq. (13) have been shown to improve the finite-sample properties of tests of equal predictive ability (see Diebold and Mariano, 2002), this is used to motivate excluding the correction term in Amisano and Giacomini (2007).

3.4 Optimized option combinations

To analyze the economic significance of the improvements in the conditional density estimates from information projection with conditional moment conditions, we analyze the performance of option combinations that are constructed using optimization based on prior or posterior estimates.

The analysis differs from Constantinides et al. (2020), Post and Longarela (2021) and Beare et al. (2025) because it does not assume that pricing errors occur for the options that are included in the moment conditions used to construct the posterior. Instead, performance improvement under the posterior can stem from (i) avoiding unprofitable positions based on spurious arbitrage opportunities that arise under the prior due to estimation error and (ii) improvement upon the general risk profile of the index that are not arbitrage opportunities.

The optimization problem seeks to maximize expected utility by buying one (if any) protective put and/or writing one (if any) covered call for a given long position in the stock index. The combination of one protective put and one covered call is known as an option collar. All these positions (protective puts, covered calls and option collars) are defensive by nature and improve the risk profile of the index for strictly risk-averse index investors when added as an overlay to the index. The writing of puts and buying of calls is not considered because it increases downside risk and its appeal critically depends on the assumed level of risk aversion.

The optimization problem for a given density forecast $\hat{G}$ follows:

$$\max \mathbb{E}_{\hat{G}}\left[u\left(S_T+\sum_{i=1}^{N}\alpha_i X_{i,T}-\sum_{i=1}^{N}\beta_i X_{i,T})\right)\right]; \tag{14}$$

$$\sum_{i=1}^{N}\alpha_i a_i-\sum_{i=1}^{N}\beta_i b_i \le 0 \tag{15}$$

$$\sum_{i=1}^{N}\alpha_i \le 1; \tag{16}$$

$$\sum_{i=1}^{N}\beta_i \le 1; \tag{17}$$

$$\alpha_i = 0, \forall i : X_{i,T} = (S_T - K_i)_+; \tag{18}$$

$$\beta_i = 0, \forall i : X_{i,T} = (K_i - S_T)_+; \tag{19}$$

$$\alpha_i, \beta_i \in \{0,1\},\ i = 1, \cdots, N. \tag{20}$$

The binary variables $\{\alpha_i\}$ and $\{\beta_i\}$ represent the open long put positions and open short call positions, respectively, in the individual options. By activating one of the binary variables $\{\alpha_i\}$, the model selects the optimal strike for the protective put, and by activating one of the binary variables $\{\beta_i\}$, the model selects the strike for the covered call. For the discretized density estimates, a finite Binary Convex Optimization problem is obtained.

Protective puts, covered calls and option collars are defensive positions that are known to reduce risk for strictly risk-averse index investors. Therefore, they should lower expected return in equilibrium with a strictly risk-averse representative investor. Consequently, non-zero solutions for the option overlay ($\{\alpha_i, \beta_i\} \neq \{0, 0\}$) are non-existent if the utility function is risk-neutral ($u(S_T) = S_T$) and the density estimate obeys the moment conditions (8)-(11) for qualified all options.[1] In addition, even if such risk arbitrage opportunities do occur (in disequilibrium), they may get obscured by distributional estimation error ($\hat{G} \neq F$).

By contrast, if a strictly risk-averse utility function is selected for the objective function, buying fairly valued puts and writing fairly valued calls can become optimal to improve the risk profile of the index by reducing downside risk. This approach also allows for some slack for SD relations to account for the effect of estimation error ($\hat{G} \neq F$).

[1] Non-zero solutions may occur under the posterior if the option filters for imposing the moment conditions are more restrictive than the option filters for the inclusion in the option combinations.

We use here the standard logarithmic utility function $u(S_T) = ln(S_T)$ which has a Relative Risk Aversion (RRA) coefficient of one. The optimization can be seen as building a Growth Optimal Portfolio using options subject to the imposed investment constraints. The defensive nature of the option overlay implies that the solution improves upon the index also for all utility functions that are more risk-averse than assumed in the objective function. In this respect, an approximate dominance relation is obtained in the spirit of 'Almost Stochastic Dominance'.

At every formation date, we solve the Binary Convex Optimization problem for the various discretized prior and posterior density forecast models. For every model, we evaluate the consequences for an index investors of buying the chosen put (if any) at its ask price and writing the chosen call (if any) at its bid price, and holding the options until cash settlement at expiration. We evaluate the OOS investment performance of the index combined with the optimized option positions. In addition to the mean and the standard deviation of return, we evaluate the Certainty Equivalent Return (CER) for a power utility function with RRA = 1, 2, 4.

4 Monte Carlo Experiment

A Monte Carlo simulation experiment is used to analyze the pre-asymptotic behavior of the updated density estimator. We will simulate the divergence $\mathrm{KL}(\cdot||F)$ for $\hat{F}$, $\hat{G}$ and $\hat{Q}$, the test statistic $\mathrm{LRT}(\cdot, \hat{F})$ for F, $\hat{G}$ and $\hat{Q}$, and the investment performance of active investment based on $\hat{F}$, $\hat{G}$ and $\hat{Q}$.

4.1 Design

Suppose that $S_0 = 100$ and R_S obeys a log-normal distribution (F). The expected return is based on the classic mean-variance approximation $\mu_S = \mu(\gamma) := R_B + \gamma\sigma_S^2$, where γ is a latent aggregate RRA coefficient.

The variance is based on a second-order Taylor series approximation by Kang and Yoon (2010): $\sigma_S^2 = \sigma^2(\gamma) := \sigma_Q^2 + \gamma\sigma_Q^3\theta_Q + 0.5\gamma^2\sigma_Q^4(\kappa_Q - 3)$, where $(\sigma_Q^2, \theta_Q, \kappa_Q)$ are the risk-neutral variance, skewness and kurtosis.

By the probability calculus of the lognormal, the location parameter is $M(\gamma) = \ln(1 + \mu(\gamma)) - \frac{1}{2}V(\gamma)$ and the scale parameter is $V(\gamma) = \ln\left(1 + \sigma^2(\gamma)/(1 + \mu(\gamma))^2\right)$. The risk-neutral distribution Q arises here as the special case with $\gamma = 0$.

The prior $\hat{F}$ takes the correct, lognormal form but its parameters are based on the RRA estimator $\hat{\gamma}_T$: $M(\hat{\gamma}_T)$ and $V(\hat{\gamma}_T)$. It could be based on time-series estimates for the physical variance σ_S^2, combined with the extracted Q moments, in a GMM estimation procedure, as in Bakshi and Madan (2006) or Kang and Yoon (2010).

Instead of modeling the precise volatility dynamics and estimation method, we assume that $\hat{\gamma}_T \sim U(\gamma+\delta_0-\delta_1, \gamma+\delta_0+\delta_1)$, where δ_0 is the bias and $\delta_1 \geq 0$ is the semi-range.

One may think of δ_0 as capturing the non-vanishing effect of specification error for the volatility dynamics and δ_1 as capturing the vanishing effect of sampling variation reflecting the sample size (T). The limiting case $\delta_1 \to 0$ then represents $T \to \infty$.

The conditional information for constructing the posterior $\hat{G}$ includes the quotes b_i and a_i for N call options with strikes $K_i = S_0 + 0.01S_0(i - 0.5(N+1))$, $i = 1, \cdots, N$, as well as the equilibrium system (8)-(11).

The Present Value of the call options amounts to $X_{i,0} := \mathbb{E}_F[mX_{i,T}]$, where the latent SDF is assumed to take a power shape: $m \propto (1+R_S)^{-\gamma}$. The bid-ask spread has width $W_i := (a_i - b_i) = wX_{i,0}$, where $w \geq 0$ is the fractional spread; it is randomly located around the Present Value: $a_i \sim U(X_{i,0}, X_{i,0}(1+w))$, so that it includes the latent true Present Value: $a_i \leq X_{i,0} \leq b_i, 1 = 1, \cdots, N$.

We set $\gamma = 5$ and $(R_B, \sigma_Q^2, \theta_Q, \kappa_Q) = (0, 0.04/12, -1, 4.5)$, all representative values for monthly S&P500 returns and one-month SPX options. The RRA level $\gamma = 5$ is relatively high compared to estimates from risky choice experiments, to generate realistic values for the Market Risk Premium and Variance Risk Premium, as is also the case for typical option-implied RRA estimates in Bakshi and Madan (2006) or Kang and Yoon (2010). We abstract from dividends in the simulations: $R_D = 0$.

We vary the bias $\delta_0 = 0, 5$ and semi-range $\delta_1 = 1, 5$ to analyze the effect of the goodness of the prior, and we vary the cross-sectional breadth $N = 5, 25$ and the spread $w = 0.010, 0.002$ to analyze the effect of the information contents of the moment conditions. The assumed breadth is smaller than the number of qualified SPX options in our empirical analysis; the assumed spreads are representative for the quoted spreads under normal market conditions in recent years.

The improvements in density estimation are quantified by repeatedly drawing a random RRA estimate $\hat{\gamma}_t \sim U(\gamma+\delta_0-\delta_1, \gamma+\delta_0+\delta_1)$, $t = 1, \cdots, 300$, each representing a random underlying time series sample, and computing the parameters of $\hat{F}$ accordingly $(M(\hat{\gamma}_t), V(\hat{\gamma}_t))$. Next, we construct $\hat{G}$ using CIPDE based on system (8)-(11) and the simulated bid-ask prices $\{a_i, b_i\}$. We also construct the risk-neutral $\hat{Q}$ using the constant SDF $m(S_T) = (1+R_B)^{-1}$.

Having generated $\hat{F}$, $\hat{G}$ and $\hat{Q}$ for a given $\hat{\gamma}_t$, we proceed to compute the divergence measures $\mathrm{KL}(\hat{F}||F)$, $\mathrm{KL}(\hat{G}||F)$, and $\mathrm{KL}(\hat{Q}||F)$. In addition, we apply forecasting ability tests based on $\mathrm{LRT}(F, \hat{F})$, $\mathrm{LRT}(\hat{G}, \hat{F})$ or $\mathrm{LRT}(\hat{Q}, \hat{F})$ to 10,000 random samples of 300 return realizations from the true F. Finally, we construct the optimal active portfolio with open option positions and evaluate its economic goodness relative to the true F using various investment performance measures.

The simulation design aligns with the statistical theory outlined in Section 2.

The form of the SDF and the moment functions, together with the assumption of log-normal index returns, imply the existence of exponential moments. The choice of the bias and semi-range parameters is consistent with RRA estimates that converge, as $\delta_1 \to 0$ (or $T \to \infty$ in case of time-series estimation), to a deterministic value at standard rates and with asymptotic Gaussianity.

4.2 Divergence

Table 1 summarizes the results for the divergence measures $\mathrm{KL}(\hat{F}||F)$, $\mathrm{KL}(\hat{G}||F)$, and $\mathrm{KL}(\hat{Q}||F)$. Reported are the quartile boundaries (p25, p50 and p75) across the 300 random draws of $\hat{\gamma}_T$.

Not surprisingly, if $\hat{F}$ is unbiased and precise ($(\delta_0, \delta_1) = (0, 1)$), improvements are difficult to achieve, even if many tight moment inequalities are available ($(N, w) = (25, 0.002)$).

However, the divergence $\mathrm{KL}(\hat{F}||F)$ quickly increases as the bias (δ_0) or semi-range (δ_1) is increased. Whereas the bias (δ_0) mostly affects the median value of the divergence (p50), the semi-range mostly affects the inter-quartile range (p75-p25). Even a small number of loose moment inequalities ($(N, w) = (5, 0.010)$) now suffices for achieving material improvements.

The strongest results naturally are obtained if $\hat{F}$ is biased and imprecise, and many tight moment inequalities are available ($(\delta_0, \delta_1, N, w) = (5, 5, 25, 0.002)$). In this case, $\hat{G}$ generates more than 50% reduction of the $\mathrm{KL}(\cdot|F)$ levels.

The estimated risk-neutral $\hat{Q}$ strongly diverges from the true F, due to its false assumption that $\gamma = 0$, even in the benign case with $(\delta_0, \delta_1) = (0, 1)$. The bias is already severe if only a few loose moment conditions are imposed. If the prior is consistent ($\delta_0 = 0$) or precise ($\delta_1 = 1$), then $\hat{Q}$ is clearly inferior to the physical $\hat{G}$. Even in the least favorable case $(\delta_0, \delta_1) = (5, 5)$, $\hat{Q}$ shows greater divergence than $\hat{G}$ in most cases.

[Insert Table 1 about here.]

4.3 Predictive Ability

An alternative way to evaluate the effect of the information projection is to analyze the rejection rates for the forecasting ability test. For this purpose, $\mathrm{LRT}(F, \hat{F})$, $\mathrm{LRT}(\hat{G}, \hat{F})$, or $\mathrm{LRT}(\hat{Q}, \hat{F})$ are computed based on 10,000 random samples of 300 return realizations from the true F. Table 2 reports the rejection rates for nominal significance levels of 2.5%, 5%, and 10%. The rejection rate measures the relative frequency of samples in which F, $\hat{G}$, or $\hat{Q}$ is classified as significantly better than $\hat{F}$ at the assumed significance level.

The rejection rates for $\mathrm{LRT}(F, \hat{F})$ are slightly above the nominal significance level if $\hat{F}$ is unbiased and precise $((\delta_0, \delta_1) = (0, 1))$. This finding is reassuring about the statistical test size, given that F_∞ lies in the null for $\delta_0 = 0$.

The rejection rate increases significantly if $\hat{F}$ is estimated with lower accuracy. The high rejection rate for $\delta_0 = 5$ is reassuring about the statistical test power, given that F_∞ lies deep in the alternative in this case. The rejection rate for $\mathrm{LRT}(\hat{G}, \hat{F})$ is even higher than for $\mathrm{LRT}(F, \hat{F})$, as a result of the relatively high correlation between the likelihood scores of $\hat{F}$ and $\hat{G}$.

Consistent with the high $\mathrm{KL}(\hat{Q} \,||\, F)$ in Table 1, rejection rates are very low for $\mathrm{LRT}(\hat{Q}, \hat{F})$ across all specifications, reflecting the inherent pessimistic bias of $\hat{Q}$.

[Insert Table 2 about here.]

4.4 Investment Performance

For each random draw of $\hat{\gamma}_T$, the optimal option combination and active portfolio is constructed for each of the three estimators, $\hat{F}$, $\hat{G}$ and $\hat{Q}$.

To evaluate economic goodness, we draw 10,000 random one-year sub-samples of 12 monthly returns from the true F and compute the returns on the option combinations. The active positions are updated each month based on a random draw from the 300 simulated values of $\hat{\gamma}_T$. We then estimate the mean, standard deviation, and CER of the annual holding-period return by aggregating across the 12-month samples, and report percentile values across the resulting 10,000 annualized metrics in Table 3.

The CER for RRA=1 is arguably the most relevant statistic because it matches the objective function of the Growth Optimal Portfolio, but the CERs for RRA=2,4 are also reported to detect approximate SD relations.

Under F, the optimal solution is a passive position in the index without any open option positions; the risk premiums are based on $\gamma = 5$ which is too high for the Growth Optimal Portfolio to buy fairly priced puts or write fairly priced calls. Due to underestimation of RRA, unprofitable protective puts and covered calls can be opened based on $\hat{F}$ if $\hat{\gamma}_T \leq 1$. In this simulation, this situation can occur only for $\{\delta_0, \delta_1\} = \{0, 5\}$; hence, the focus is on this specification here.

Table 3 shows results for the simulation for $\{\delta_0, \delta_1\} = \{0, 5\}$, again using percentiles across the 300 random draws for $\hat{\gamma}_T$.

The estimation error based on the prior $\hat{F}$ translates into decision errors in the form of open option positions. Puts are bought in 28.5% of the months, so that almost every year includes several long put positions. These unprofitable positions worsen the risk–return profile. Notably, the CER_1—which represents the objective function for the optimization—drops by about 100 basis points per annum.

The forecasting improvements from the physical $\hat{G}$ translate into fewer and less consequential errors, and lead to substantial performance improvements relative to the performance based on $\hat{F}$. There is a reduction in the number of months in which puts are bought, and in those months, deeper OTM (and hence cheaper) puts are selected. When only a few loose moment conditions are used, the gains are limited to tens of basis points per annum. However, for $(N, w) = (25, 0.002)$, the performance based on $\hat{G}$ closely approximates the optimal performance based on the true F by mostly avoiding open option positions.

Using the estimated risk-neutral distribution $\hat{Q}$ leads to large decision errors in this simulation. Due to its pessimistic bias, the estimator makes fairly priced put options appear inexpensive at their ask price, and fairly priced call options appear expensive at their bid price. As a result, put options are bought and call options are written with high frequency, leading to reductions in the expected return and CERs of several hundred basis points per annum.

[Insert Table 3 about here.]

5 Empirical Analysis

The analysis now shifts to real-life stock index returns and stock index option price quotes. The focus is on density forecasting for monthly SPX returns and optimization of combinations of monthly CBOE SPX options.

5.1 Data

The one-month T-bill rate and dividend yield are obtained from OptionMetrics IvyDB. S&P 500 Index and VIX values are downloaded from Thomson Reuters Eikon.

The primary source for SPX option price data is intra-day quotes captured at 15:45 ET from the iVolatility database, available from January 2004 through January 2023. We backfill these intra-day quotes with closing price data from OptionMetrics IvyDB, covering the period from January 1996 to December 2003. We include options with 29 calendar days to expiry ($T \approx 29/365$).

Before January 2016, SPX option expiry dates were recorded as the Saturday following the third Friday of the month, although the Friday index level was used for settlement. We use the time to the settlement value in our time-to-expiry estimation. In four months, the sample trade dates fell on Thanksgiving Day, when the options exchange was closed (22 November 2001, 2007, 2012, and 2018). The absence of quotes on these days leaves $M = 321$ trade dates with valid option data for analysis.

We apply a series of filters based on moneyness, option delta, and option premium to select options for inclusion in the analysis. Following standard conventions in empirical research, we exclude in-the-money (ITM) options, which tend to be less liquid than out-of-the-money (OTM) options. We also impose a minimum delta of 0.15 for calls and a maximum delta of -0.15 for puts, in order to exclude deep OTM options. Additionally, we exclude options with bid prices below 15 cents.

These filters are applied both to define the moment conditions used in posterior density estimation and to determine the investment universe used in optimizing option combinations. The analysis can be generalized by tightening the filters for estimation purposes or loosening them for optimization, if the objective is to detect potential mispricing.

Table 4 presents summary statistics for the options that pass these filters. Separate results are provided for puts and calls. The table reports the quartile boundaries (p25, p50, and p75), computed across all $M = 321$ monthly strips, for the cross-sectional breadth (N_P for puts and N_C for calls), as well as the average and standard deviation (computed across all N_P or N_C qualified options) of moneyness, IV, and the bid-ask spread.

Consistent with the maturation of the options market, the cross-sectional breadth increases and quoted spreads decline over time. In the most recent five years, the number of qualified option series ($N_P + N_C$) exceeds 100, and the quoted spread falls below 1% of the midpoint premium for both puts and calls. These patterns suggest increasing market efficiency and declining profitability of active option trading over time. They also support the calibration of option breadth and bid-ask spreads used in the Monte Carlo simulations in Section 4.

[Insert Table 4 about here.]

5.2 Prior Density Estimators

To generate prior density estimates ($\hat{F}$), we apply three common methods for estimating the conditional distribution of stock index returns:

$\hat{F}_{LN}$: a lognormal distribution with conditional mean and volatility,

$\hat{F}_{ECDF}$: a transformation of the historical unconditional ECDF, and

$\hat{F}_{FHS}$: a forecast density generated using a GJR-GARCH model with returns simulated via Filtered Historical Simulation (FHS).

The conditional means for all three forecast distributions, and the conditional standard deviation for $\hat{F}_L N$ and $\hat{F}_{ECDF}$, are estimated using arguments from Constantinides et al. (2009). Specifically, the conditional mean is based on the prevailing one-month Treasury bill rate (R_B), the estimated market risk premium ($\hat{\text{MRP}}$), and the prevailing dividend yield (R_D), such that $\mathbb{E}_{\hat{F}}[R_S] = R_B + \hat{\text{MRP}} - R_D$.

We set the annualized $\hat{\text{MRP}}$ to 4%, and verify that that our results and conclusions are robust to plausible variations in this parameter—presumably because option prices are more sensitive to the scale and shape of the return distribution than to its location.

Conditional volatility is proxied by the implied volatility (IV) of the nearest-to-the-money call option. Our results and conclusions about information projection are robust to reasonable variations in the estimator of conditional standard deviation. However, it should be noted that the CIPDE problem may fail to admit a solution if an excessively poor prior is used.

The conditional ECDF $\hat{F}_{ECDF}$ is derived by transforming the unconditional ECDF computed over a rolling window of 800 monthly returns before each trade date. This broad window is chosen to capture sufficiently many historical tail events. The ECDF is build on a regular grid of return values in the range $[-0.6, 0.6]$, spaced at 25 basis points. Translation and scale transformations are applied to the atoms of the unconditional ECDF to match the target mean $\mathbb{E}_{\hat{F}}[R_S]$ and volatility (IV). Using this approach, the conditional distribution preserves the shape of the unconditional distribution.

The GJR-GARCH FHS model is implemented following Barone-Adesi et al. (2020).[2] On each trade date, the model is fitted using daily return data in an 800 month historical moving window up to the trade date updated to match 800 month window used for ECDF, note this weakened the FHS forecasting power slightly, but an assumption for the A&G test is that compared forecasts are formed on the same information set. A set of 10,000 return paths is then simulated, with drift adjusted to the target mean, and the conditional ECDF is estimated from the simulated series using a Gaussian kernel smoothing function with bandwidth set using Silverman's rule.

To facilitate OOS evaluation, the discrete forecast distributions—namely, the transformed unconditional ECDF and the conditional ECDF from FHS—are smoothed using kernel density estimation to yield continuous densities: $p_h^*(S) := \frac{1}{h} \sum_{i=1}^{N} \pi_i^* k\left(\frac{S_i - S}{h}\right)$, where k is a Gaussian kernel and h is the bandwidth chosen via Silverman's rule. Smoothing is essential because the OOS realized returns almost surely do not coincide with atoms in the discrete forecast distributions. In contrast, $\hat{F}_{LN}$ already defines a smooth density and thus requires no additional smoothing.

[2] We thank Carlo Sala for kindly providing the code to implement this method.

All three priors satisfy our statistical theory. Under standard regularity conditions—including stationarity, ergodicity, and mixing—their estimated densities converge uniformly in probability to pseudo-true limits. For $\hat{F}_{LN}$, convergence follows if the estimated mean and volatility converge uniformly. For $\hat{F}_{ECDF}$, convergence derives from the Glivenko–Cantelli theorem applied to the rolling-window empirical distribution, with affine transformations preserving uniform convergence. For $\hat{F}_{FHS}$, provided the GARCH model is pseudo-consistently estimated and the empirical residuals are ergodic, the smoothed ECDF also converges uniformly to a pseudo-true limit. Hence, Assumption 1 is satisfied in all three cases.

Regarding Assumption 3, and under regularity conditions, $\hat{F}_{LN}$ exhibits $\sqrt{T}$-rate weak convergence via the delta method, since its parameters (mean and variance) converge at parametric rates. $\hat{F}_{ECDF}$ satisfies a Donsker-type weak convergence result under standard assumptions (e.g., strong mixing and a Lipschitz-continuous kernel), as discussed in Mojirsheibani (2006); the use of a fixed bandwidth preserves this rate. For $\hat{F}_{FHS}$, the GARCH volatility model is estimated via quasi-maximum likelihood, which guarantees $\sqrt{T}$-consistency under suitable conditions, and the simulation of 10,000 paths regularizes the empirical forecast distribution. Uniform-type empirical process convergence results, under stationarity and mixing, justify the assumption; see, for example, Rao and Krishnaiah (1988).

Table 5 compares the three prior density estimators based on quartile boundaries (p25, p50, and p75), computed over $M = 321$ months, along with the first four central moments and pairwise KL divergences. $\hat{F}_{LN}$ and $\hat{F}_{ECDF}$ tend to exhibit pronounced negative skewness and fat tails, consistent with empirical stylized facts. These features are most pronounced in $\hat{F}_{LN}$, which helps explain why the largest divergence is observed between $\hat{F}_{LN}$ and $\hat{F}_{FHS}$.

The prior density estimates fix or constrain the shape and parameters of the return distribution using limited conditioning information (R_B, R_D, and IV). These simplifying features make the priors potentially biased and inefficient, allowing for improvements via additional conditional moment conditions based on index option market prices.

[Insert Table 5 about here.]

5.3 Posterior Density Estimation

The priors are updated to posteriors using the moment conditions (8)–(11), applied to all options that pass the data filters. For the risk-neutral posterior, the SDF is restricted to $m(S_T) = (1 + R_B)^{-1}$.

Table 6 summarizes the differences between the priors ($\hat{F}$) and posterior distributions ($\hat{G}$ and $\hat{Q}$), using quartile breakpoints for the central moments and KL divergence measures. It also reports the number of months in which the prior distribution passes the conditional moment conditions (M_{Pass}), and the number of months in which the prior could not be updated because it provided too poor a fit to option prices (M_{Inf}). In those months, the posterior is equated with the prior.

The posterior physical distribution ($\hat{G}$) tends to display more negative skewness and excess kurtosis than the prior $\hat{F}$, reflecting a more pessimistic view of future returns. Consequently, the posterior implies less downside risk and greater upside potential than the prior.

Given the counterfactual skewness and kurtosis of the lognormal, it is unsurprising that the divergence $\text{KL}(\hat{G}, \cdot)$ is largest for this prior. Moreover, insolvability arises most frequently for this prior. By contrast, for $\hat{F}_{ECDF}$ and $\hat{F}_{FHS}$, the updating is feasible in the vast majority (97–98%) of months, and the resulting posteriors diverge less from their priors.

The posterior risk-neutral distribution $\hat{Q}$ is typically even more pessimistic than the physical posterior, as reflected by its lower mean, higher standard deviation (SD), more negative skewness, and greater excess kurtosis. It also diverges more from the prior than the physical posterior does.

[Insert Table 6 about here.]

5.4 Predictive Ability

To evaluate the improvements in forecasting ability achieved by the posteriors, LRT statistics are computed based on the realized index return on the option expiry date. In addition comparing posteriors with underlying priors (e.g., $\text{LRT}(\hat{G}_{LN}, \hat{F}_{LN})$ and $\text{LRT}(\hat{Q}_{LN}, \hat{F}_{LN})$), comparison is also made across priors (e.g., $\text{LRT}(\hat{F}_{ECDF}, \hat{F}_{LN})$, $\text{LRT}(\hat{G}_{ECDF}, \hat{F}_{LN})$ and $\text{LRT}(\hat{Q}_{ECDF}, \hat{F}_{LN})$).

Table 7 summarizes the test results.

The posterior physical distribution $\hat{G}$ provides a significantly better density forecast than the prior $\hat{F}$ for all three specifications (LN, ECDF, FHS), as indicated by the low p-values for $\text{LRT}(\hat{G}_{LN}, \hat{F}_{LN})$, $\text{LRT}(\hat{G}_{ECDF}, \hat{F}_{ECDF})$, and $\text{LRT}(\hat{G}_{FHS}, \hat{F}_{FHS})$. These results are consistent with the theoretical benefits of information projection and the assumption that the option pricing system is true and informative.

The most pronounced improvements are observed for $\hat{F}_{LN}$, consistent with the counterfactual moments of the lognormal in Table 5 and the large divergence of the lognormal from the posterior in Table 6. The improvements remain statistically significant for $\hat{F}_{FHS}$, which exhibits the highest forecasting ability among the three priors, as evidenced by the low p-values for $\text{LRT}(\hat{F}_{FHS}, \hat{F}_{LN})$ and $\text{LRT}(\hat{F}_{FHS}, \hat{F}_{ECDF})$.

The improvements are less pronounced for the posterior risk-neutral distribution $\hat{Q}$, reflecting its intrinsic pessimistic bias. In fact, for FHS, the risk-neutral posterior does not significantly improve the forecast.

[Insert Table 7 about here.]

5.5 Investment Performance

Table 8 summarizes the composition and OOS performance of the optimized option combinations for the various prior and posterior distributions. The first two panels report the frequency of buying protective puts and writing covered calls, respectively, as well as the mean and standard deviation of the associated moneyness and implied volatility of the selected options.

The optimal solution based on the prior distribution frequently involves buying protective puts and writing covered calls. By contrast, open option positions are less frequent under the posterior physical distribution $\hat{G}$ and tend to concentrate on options that are deeper OTM. This result is consistent with the reduction in downside risk and the increase in upside potential under the posterior. Among the three priors, $\hat{F}_{ECDF}$ exhibits the fewest open option positions, consistent with the small KL divergence reported in Table 5 and the low LRT statistic in Table 7.

The third panel summarizes OOS investment performance. Although the open option positions based on the prior $\hat{F}$ reduce risk (as evidenced by the reduction in standard deviation), OOS performance deteriorates. Notably, these strategies underperform the index in terms of CER_1, which corresponds to the logarithmic utility function, and the underperformance worsens as the number of open option positions increases.

In contrast, the optimal solutions based on the posterior physical distribution $\hat{G}$ outperform the index and show improvements in CER across all three priors and all three levels of risk aversion. These gains align with the improvements in forecasting ability observed in Table 7. The gains are smallest for $\hat{F}_{ECDF}$, which features the fewest open positions under the posterior.

The bottom panel reports formal test results. We test for SD using the bootstrap test by Linton et al. (2010). This test is expected to be powerful in our context, as the null hypothesis is exact second-order SD, while approximate dominance is anticipated. Moreover, the bootstrap test is expected to be more powerful than the subsampling test of Linton et al. (2005), given the limited sample size.

The null that the option-enhanced portfolio dominates the passive index is rejected with more than 90% confidence for $\hat{F}_{LN}$ and $\hat{F}_{FHS}$. As expected, it is not rejected for the solution based on $\hat{F}_{ECDF}$, which includes only a few open positions. Encouragingly, no rejections occur for any of the three posterior physical distributions ($\hat{G}_{LN}$, $\hat{G}_{ECDF}$, and $\hat{G}_{FHS}$), reinforcing the economic significance of the forecast improvements under conditional information projection.

A very different picture emerges for the posterior risk-neutral distribution $\hat{Q}$. Due to its strong pessimistic bias, protective puts are bought and covered calls are written in most months. The optimized portfolio underperforms the index across all performance metrics, owing to reduced equity exposure and transaction costs due to bid-ask spreads. As expected, stochastic dominance is strongly rejected. These findings highlight that a pessimistic forecast bias can lead to economically significant decision errors.

For verification purposes, we also solved the optimization problems using the risk-neutral objective $u(S_T) = S_T$ instead of expected logarithmic utility. Under all three prior estimates, the resulting in-sample solutions led to OOS underperformance. Moreover, under all three posterior estimates, no non-zero optimal solutions were found. These outcomes support our choice of data filters (as no evidence of pricing errors was found for the selected options under the prior) and confirm the internal consistency of the estimation and optimization procedures (as no positions were opened under the posterior).

[Insert Table 8 about here.]

5.6 Robustness Analysis

To evaluate the robustness of our results and conclusions, we analyze the effects of excluding months with extreme economic circumstances from the time series and excluding illiquid deep OTM options from the monthly option strips.

We consider three alternative time-series subsamples: (i) a subsample excluding the 34 months where the US economy was classified as being in recession by the NBER; (ii) a subsample excluding the 32 months where the financial uncertainty index of Jurado et al. (2015) was above its 90th percentile level; (iii) a subsample excluding the 32 months where the VIX was above its 90th percentile level for the sample.

In addition, we consider a more restrictive option filter for the conditional moment conditions used to estimate the posterior forecasts. Specifically, the OTM delta filter is tightened to exclude OTM options with absolute delta in the range [0, 0.15] instead of [0, 0.05].

Table 9 summarizes the LRT results for the alternative samples. The improvements in forecasting accuracy from conditional information projection are highly robust to the exclusion of extreme months and deep OTM options. Across all four sub-samples and three prior forecast distributions, the posterior forecast distribution consistently enhances the OOS goodness of fit.

[Insert Table 9 about here.]

Table 10 summarizes the OOS investment performance of the option-enhanced strategies after the exclusions. The performance improvements are robust for $\hat{F}_{LN}$ and $\hat{F}_{FHS}$. However, they are less robust for $\hat{F}_{ECDF}$, which is unsurprising given the low frequency of option position openings and low LWS10 p-value under this prior.

[Insert Table 10 about here.]

6 Conclusions

This study has introduced CIPDE to estimate latent conditional distributions by integrating prior estimates with conditional moment conditions with functional nuisance parameters. CIPDE updates the prior density estimate through information projection onto the set of distributions that satisfy these moment conditions. Theoretical analysis and Monte Carlo simulations demonstrate that CIPDE achieves lower relative entropy to the latent conditional distribution when the prior is inconsistent and the moment conditions hold.

Our statistical theory is designed to work with conditioning variable ranges in arbitrary Euclidean spaces. In practice, the curse of dimensionality is expected to kick in for dimensions greater than two. Recent advances in non-parametric statistics have extended projection methods to high-dimensional settings (potentially diverging with the sample size) by leveraging structured regularization techniques based on Hilbert norms (see, for example, Wainwright (2014) and the references therein). The consideration of such technologies in our framework seems like an interesting path for further research.

In option market research, conditional density estimates are essential due to the dynamic and non-Gaussian nature of index return distributions and the short maturity if the most liquid series. Our empirical analysis shows that conditional information projection enhances the forecasting performance of standard conditional density estimators used in this domain.

The conditional moment conditions incorporate observed SPX option prices and general option pricing restrictions that rule out SD relations for qualified options. CIPDE, based on these moment conditions, significantly improves OOS forecasting accuracy and investment performance through better timing of protective put purchases and covered call writing. The improvements critically depend on incorporating the SDF as a parameter in the moment conditions to avoid the pessimistic bias of the risk-neutral density estimate. Robustness checks confirm the stability of the improvements across alternative subsamples that exclude extreme economic conditions and illiquid options.

The documented forecast improvements support the truth of the moment conditions for the selected options. Nevertheless, assessing the overall pricing efficiency of the index option market is not our objective, and our results are inconclusive on this broader issue. We applied strict option filters based on maturity, moneyness, delta, and premium (see Section 5.1). These filters were used both to impose moment conditions and to optimize option combinations. Evidence of mispricing or risk arbitrage opportunities may arise if the estimation filters used in the CIPDE are more restrictive than those applied to the options selected for investment optimization. Moreover, the documented forecast improvements are also consistent with the equilibrium system being only approximately—rather than exactly—correct, provided pricing errors are rare or small.

References

Adams, R. A. and J. J. Fournier (2003): *Sobolev Spaces*, vol. 140 of *Pure and Applied Mathematics*, Academic Press, 2nd ed.

Almeida, C. and R. Garcia (2012): "Assessing Misspecified Asset Pricing Models with Empirical Likelihood Estimators," *Journal of Econometrics*, 519–537.

Amisano, G. and R. Giacomini (2007): "Comparing density forecasts via weighted likelihood ratio tests," *Journal of Business & Economic Statistics*, 25, 177–190.

Andrews, D. W. (1999): "Estimation when a parameter is on a boundary," *Econometrica*, 67, 1341–1383.

Bakshi, G. and D. Madan (2006): "A theory of volatility spreads." *Management Science*, 52, 1945–1956.

Barone-Adesi, G., N. Fusari, A. Mira, and C. Sala (2020): "Option market trading activity and the estimation of the pricing kernel: A Bayesian approach," *Journal of Econometrics*, 216, 430–449.

Basso, A. and P. Pianca (1975): "Decreasing Absolute Risk Aversion And Option Pricing." *Management Science*, 43, 206—-216.

Bawa, V. (1975): "Optimal Rules for Ordering Uncertain Prospects." *Journal of Financial Economics*, 2, 95–121.

Beare, B. K., J. Seo, and Z. Zheng (2025): "Stochastic arbitrage with market index options," *Journal of Banking & Finance*, 173.

Bickel, P. J., C. A. Klaassen, Y. Ritov, and J. A. Wellner (1998): *Efficient and Adaptive Estimation for Semiparametric Models*, Springer.

Blasques, F., S. J. Koopman, and A. Lucas (2018): "Information-Theoretic Optimality of Observation-Driven Time Series Models for Likelihood Inference," *Journal of Econometrics*, 202, 112–131.

Blundell, R., J. L. Horowitz, and M. Parey (2012): "Nonparametric Engel curves and revealed preference," *Econometrica*, 80, 197–229.

Constantinides, G. M., M. Czerwonko, and S. Perrakis (2020): "Mispriced index option portfolios," *Financial Management*, 49, 297–330.

Constantinides, G. M., J. C. Jackwerth, and S. Perrakis (2009): "Mispricing of S&P 500 Index Options," *Review of Financial Studies*, 22, 1247–1277.

Cortissoz, J. (2007): "On the Skorokhod Representation Theorem," *Proceedings of the American Mathematical Society*, 135, 3995–4007.

Csiszár, I. (1975): "I-divergence Geometry of Probability Distributions and Minimization Problems," *The Annals of Probability*, 3, 146–158.

Diebold, F. X. and R. S. Mariano (2002): "Comparing predictive accuracy," *Journal of Business & economic statistics*, 20, 134–144.

Feydy, J., T. Séjourné, F.-X. Vialard, S.-i. Amari, A. Trouvé, and G. Peyré (2019): "Interpolating between optimal transport and mmd using sinkhorn divergences," in *The 22nd international conference on artificial intelligence and statistics*, PMLR, 2681–2690.

Imbens, G., R. Spady, and P. Johnson (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66, 333–357.

Jurado, K., S. C. Ludvigson, and S. Ng (2015): "Measuring uncertainty," *American Economic Review*, 105, 1177–1216.

Kang, B.J., T. K. and S.-J. Yoon (2010): "Information content of volatility spreads." *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 30, 533–558.

Kimball, M. S. (1990): "Precautionary Saving in the Small and in the Large," *Econometrica*, 58, 53–73.

Kitamura, Y. and M. Stutzer (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65, 861–874.

Knight, K. (1999): "Epi-convergence in distribution and stochastic equi-semicontinuity," *Unpublished manuscript*.

Komunjer, I. and G. Ragusa (2016): "Existence and characterization of conditional density projections," *Econometric Theory*, 32, 947–987.

Li, X. and L. H. Zhao (2002): "Bayesian nonparametric point estimation under a conjugate prior," *Statistics & Probability Letters*, 58, 23–30.

Linton, O., E. Maasoumi, and Y.-J. Whang (2005): "Consistent testing for stochastic dominance under general sampling schemes," *Review of Economic Studies*, 72, 735–765.

Linton, O., K. Song, and Y.-J. Whang (2010): "An improved bootstrap test of stochastic dominance," *Journal of Econometrics*, 154, 186–202.

Mojirsheibani, M. (2006): "A Note on the Strong Approximation of the Smoothed Empirical Process of α-mixing Sequences," *Statistical inference for stochastic processes*, 9, 97–107.

Nielsen, F. (2020): "An Elementary Introduction to Information Geometry," *Entropy*, 22.

Post, T. (2003): "Empirical tests for stochastic dominance efficiency," *The Journal of Finance*, 58, 1905–1931.

Post, T., S. Karabati, and S. Arvanitis (2018): "Portfolio optimization based on stochastic dominance and empirical likelihood," *Journal of Econometrics*, 206, 167–186.

Post, T. and I. R. Longarela (2021): "Risk arbitrage opportunities for stock index options," *Operations Research*, 69, 100–113.

Post, T. and V. Potì (2017): "Portfolio analysis using stochastic dominance, relative entropy, and empirical likelihood," *Management Science*, 63, 153–165.

RAO, M. S. AND Y. R. KRISHNAIAH (1988): "Weak convergence of empirical processes of strong mixing sequences under estimation of parameters," *Sankhyā: The Indian Journal of Statistics, Series A*, 26–43.

ROCKAFELLAR, R. T. AND R. J.-B. WETS (1998): *Variational Analysis*, vol. 317 of *Grundlehren der Mathematischen Wissenschaften*, Berlin, Heidelberg: Springer.

SCHENNACH, S. M. (2005): "Bayesian Exponentially Tilted Empirical Likelihood," *Biometrika*, 92, 31–46.

STUTZER, M. (1995): "A Bayesian approach to diagnosis of asset pricing models," *Journal of Econometrics*, 68, 367–397.

——— (1996): "A simple nonparametric approach to derivative security valuation," *The Journal of Finance*, 51, 1633–1652.

TSYBAKOV, A. B. (2009): *Introduction to Nonparametric Estimation*, Springer.

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*, Cambridge University Press.

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

VICKSON, R. (1975): "Stochastic Dominance Tests for Decreasing Absolute Risk Aversion. I. Discrete Random Variables." *Management Science*, 21, 1438—-1446.

VUONG, Q. H. (1989): "Likelihood ratio tests for model selection and non-nested hypotheses," *Econometrica: journal of the Econometric Society*, 307–333.

WAINWRIGHT, M. J. (2014): "Structured regularizers for high-dimensional problems: Statistical and computational issues," *Annual Review of Statistics and Its Application*, 1, 233–253.

YILMAZ, H. (2021): "A generalization of multiplier rules for infinite-dimensional optimization problems," *Optimization*, 70, 1825–1835.

δ_0	δ_1	N	w	KL($\hat{F}\|\|F$) p25	KL($\hat{F}\|\|F$) p50	KL($\hat{F}\|\|F$) p75	KL($\hat{G}\|\|F$) p25	KL($\hat{G}\|\|F$) p50	KL($\hat{G}\|\|F$) p75	KL($\hat{Q}\|\|F$) p25	KL($\hat{Q}\|\|F$) p50	KL($\hat{Q}\|\|F$) p75
0	1	5	0.010	0.000	0.000	0.001	0.001	0.001	0.001	0.027	0.027	0.028
			0.002	0.000	0.000	0.001	0.001	0.001	0.002	0.029	0.029	0.029
		25	0.010	0.000	0.000	0.001	0.001	0.001	0.002	0.029	0.030	0.031
			0.002	0.000	0.000	0.001	0.001	0.002	0.003	0.032	0.032	0.033
	5	5	0.010	0.003	0.012	0.023	0.003	0.011	0.020	0.028	0.032	0.035
			0.002	0.003	0.012	0.023	0.003	0.009	0.016	0.029	0.030	0.034
		25	0.010	0.003	0.012	0.023	0.003	0.008	0.016	0.030	0.031	0.033
			0.002	0.003	0.012	0.023	0.002	0.007	0.014	0.032	0.033	0.034
5	1	5	0.010	0.031	0.038	0.045	0.025	0.031	0.036	0.036	0.037	0.039
			0.002	0.031	0.038	0.045	0.016	0.020	0.023	0.030	0.030	0.030
		25	0.010	0.031	0.038	0.045	0.016	0.019	0.023	0.030	0.031	0.032
			0.002	0.031	0.038	0.045	0.013	0.016	0.020	0.033	0.033	0.033
	5	5	0.010	0.010	0.040	0.081	0.008	0.032	0.063	0.030	0.038	0.043
			0.002	0.010	0.040	0.081	0.005	0.021	0.042	0.030	0.030	0.030
		25	0.010	0.010	0.040	0.081	0.005	0.020	0.042	0.030	0.031	0.032
			0.002	0.010	0.040	0.081	0.004	0.017	0.037	0.033	0.033	0.033

Tab. 1: Simulated Divergence. For each combination of the bias ($\delta_0 = 0, 5$), semi-range ($\delta_1 = 1, 5$), number of option series ($N = 5, 25$) and fractional bid-ask spread ($w = 0.010, 0.002$), 300 prior forecasts are generated based on 300 simulated RRA estimates $\hat{\gamma}_T$. Given the simulated prior and simulated option prices, the posterior is estimated using the CIPDE procedure described in Section 2.2. Quartile boundary values (p25, p50, p75) are reported for the KL divergence from the prior, $\hat{F}$, and the posterior, $\hat{G}$, to the true distribution F. The posterior forecast can be seen to move significantly closer (in terms of KL divergence) than the prior forecast to the true distribution in each case, with the exception of $(\delta_0, \delta_1) = (0, 1)$, where the prior is already very close to the true distribution. The KL divergence is shown to reduce with increased number of option series (25 vs. 5) and reduced spread (20 bps vs. 100 bps) and, for given semi-range and bias.

				$\mathrm{LRT}(F,\hat{F})$			$\mathrm{LRT}(\hat{G},\hat{F})$			$\mathrm{LRT}(\hat{Q},\hat{F})$		
δ_0	δ_1	N	w	2.5%	5%	10%	2.5%	5%	10%	2.5%	5%	10%
0	1	5	0.010	0.040	0.060	0.140	0.020	0.040	0.070	0.000	0.000	0.000
			0.002	0.040	0.060	0.140	0.020	0.020	0.050	0.000	0.000	0.000
		25	0.010	0.040	0.060	0.140	0.020	0.030	0.060	0.000	0.000	0.000
			0.002	0.040	0.060	0.140	0.000	0.020	0.040	0.000	0.000	0.000
	5	5	0.010	0.290	0.410	0.590	0.100	0.200	0.280	0.000	0.000	0.010
			0.002	0.290	0.410	0.590	0.180	0.240	0.370	0.000	0.000	0.000
		25	0.010	0.290	0.410	0.590	0.180	0.290	0.370	0.000	0.000	0.000
			0.002	0.290	0.410	0.590	0.190	0.250	0.390	0.000	0.000	0.000
5	1	5	0.010	0.720	0.850	0.890	0.720	0.830	0.890	0.020	0.040	0.130
			0.002	0.720	0.850	0.890	0.880	0.910	0.960	0.050	0.080	0.170
		25	0.010	0.720	0.850	0.890	0.890	0.910	0.970	0.060	0.080	0.190
			0.002	0.720	0.850	0.890	0.890	0.930	0.980	0.050	0.080	0.150
	5	5	0.010	0.750	0.810	0.890	0.810	0.870	0.940	0.040	0.120	0.200
			0.002	0.750	0.810	0.890	0.940	0.950	0.970	0.070	0.160	0.280
		25	0.010	0.750	0.810	0.890	0.910	0.960	0.970	0.070	0.150	0.270
			0.002	0.750	0.810	0.890	0.940	0.960	0.970	0.070	0.110	0.250

Tab. 2: Simulated LRT Rejection Rates. For each combination of the bias ($\delta_0 = 0, 5$), semi-range ($\delta_1 = 1, 5$), number of option series ($N = 5, 25$) and fractional bid-ask spread ($w = 0.010, 0.002$), 300 random RRA estimates $\hat{\gamma}_T$ are generated from the uniform distribution. For each RRA estimate, the density estimators $\hat{F}$, $\hat{G}$ and $\hat{Q}$ constructed and their predictive ability is evaluated. $\mathrm{LRT}(F,\hat{F})$, $\mathrm{LRT}(\hat{G},\hat{F})$ and $\mathrm{LRT}(\hat{Q},\hat{F})$ are computed based on random samples of 10,000 return realizations from the true F. The Table reports the rejection rates for nominal significance levels of 2.5%, 5%, or 10%.

δ_0	δ_1	N	w	Perf. Metric	F	$\hat{F}$ p25	$\hat{F}$ p50	$\hat{F}$ p75	$\hat{G}$ p25	$\hat{G}$ p50	$\hat{G}$ p75	$\hat{Q}$ p25	$\hat{Q}$ p50	$\hat{Q}$ p75
0	5	5	0.01	Mean	18.56	17.36	17.75	18.07	17.77	18.01	18.22	11.94	12.73	13.47
				Std. Dev.	17.86	17.13	17.36	17.56	17.37	17.52	17.65	12.50	13.16	13.78
				CER_1	16.74	15.71	16.04	16.32	16.06	16.26	16.44	11.08	11.77	12.42
				CER_2	14.95	14.08	14.36	14.60	14.38	14.55	14.70	10.23	10.82	11.37
				CER_4	11.44	10.92	11.09	11.24	11.11	11.22	11.30	8.51	8.92	9.29
			0.002	Mean	18.56	17.36	17.75	18.07	17.93	18.11	18.30	6.86	7.30	7.77
				Std. Dev.	17.86	17.13	17.36	17.56	17.47	17.59	17.70	6.95	7.35	7.69
				CER_1	16.74	15.71	16.04	16.32	16.19	16.36	16.52	6.60	7.01	7.45
				CER_2	14.95	14.08	14.36	14.60	14.50	14.63	14.76	6.33	6.70	7.12
				CER_4	11.44	10.92	11.09	11.24	11.19	11.27	11.33	5.76	6.09	6.43
		25	0.01	Mean	18.56	17.36	17.75	18.07	17.95	18.14	18.30	12.42	13.18	13.92
				Std. Dev.	17.86	17.13	17.36	17.56	17.48	17.60	17.70	12.99	13.71	14.38
				CER_1	16.74	15.71	16.04	16.32	16.22	16.38	16.52	11.49	12.14	12.78
				CER_2	14.95	14.08	14.36	14.60	14.51	14.65	14.76	10.58	11.12	11.65
				CER_4	11.44	10.92	11.09	11.24	11.20	11.28	11.34	8.75	9.10	9.43
			0.002	Mean	18.56	17.36	17.75	18.07	18.43	18.47	18.50	7.72	7.95	8.15
				Std. Dev.	17.86	17.13	17.36	17.56	17.78	17.81	17.83	7.43	7.66	7.84
				CER_1	16.74	15.71	16.04	16.32	16.63	16.66	16.69	7.41	7.63	7.81
				CER_2	14.95	14.08	14.36	14.60	14.86	14.88	14.91	7.11	7.30	7.46
				CER_4	11.44	10.92	11.09	11.24	11.40	11.41	11.42	6.46	6.61	6.73

Tab. 3: Simulated Investment Performance. For (δ_0,δ_1)=(0,5), and each combination of the number of option series ($N = 5, 25$) and fractional bid-ask spread ($w = 0.010, 0.002$), 300 random RRA estimates $\hat{\gamma}_T$ are generated from the uniform distribution. For each RRA estimate, the density estimators $\hat{F}$, $\hat{G}$ and $\hat{Q}$ and associated optimal option combinations are constructed. For each of the three density estimators, 12 out of 300 option combinations are sampled with replacement 10,000 times and the mean, standard deviation and CERs (RRA=1,2,4) of annual holding-period return are computed across the 10,000 samples. The table shows the 25th, 50th and 75th percentile levels for the performance measures.

	Calls			Puts		
	p25	p50	p75	p25	p50	p75
# options	10	15	21	16	27	42
IV (Mean)	0.11	0.14	0.19	0.16	0.21	0.26
St. Dev.	0.01	0.01	0.01	0.02	0.03	0.04
MNES (Mean)	1.02	1.03	1.04	0.94	0.95	0.96
St. Dev.	0.01	0.02	0.02	0.03	0.03	0.04
Spread (Mean)	10.25	14.85	18.85	8.81	12.80	16.20
St. Dev.	5.16	9.02	12.43	3.48	6.10	8.37

Tab. 4: Option strip statistics, estimated over 251 monthly call and put option strips (Jan. 1996 - Jan 2023). The sample includes OTM options with abs(delta) >= 0.05 and bid price > $0.15. The mean and standard deviation of the annualised implied volatility of the options in each monthly strip are reported at the quartile boundaries in the sample; MNES reports the average and st. dev. of the ratio of the option strip strikes to the underlying spot price; Spread captures the average and standard deviation of the option spread in percentage terms (of the midpoint price) across the option strips.

	$\hat{F}_{LN}$			$\hat{F}_{ECDF}$			$\hat{F}_{FHS}$		
	p25	p50	p75	p25	p50	p75	p25	p50	p75
Mean	0.02	0.04	0.08	0.02	0.04	0.08	0.02	0.04	0.09
St. Dev.	0.12	0.16	0.21	0.12	0.17	0.22	0.11	0.15	0.20
Skewness	0.00	0.00	0.00	-0.62	-0.59	-0.51	-0.92	-0.81	-0.73
Kurtosis	3.00	3.00	3.00	5.40	5.68	5.96	5.59	6.33	7.32
VaR_{90}	-0.07	-0.06	-0.04	-0.07	-0.05	-0.04	-0.07	-0.05	-0.04
VaR_{10}	0.05	0.06	0.08	0.04	0.05	0.07	0.04	0.05	0.07
$\text{KL}(\cdot\|\|\hat{F}_{LN})$	0	0	0	0.0208	0.0233	0.0285	0.0216	0.0375	0.0771
$\text{KL}(\cdot\|\|\hat{F}_{ECDF})$	0.0361	0.0394	0.0470	0	0	0	0.0190	0.0411	0.1132
$\text{KL}(\cdot\|\|\hat{F}_{FHS})$	0.0356	0.0474	0.0786	0.0152	0.0223	0.0462	0	0	0

Tab. 5: Forecast mean, standard deviation, skewness, kurtosis and pairwise KL divergence (reported as `KL(col, row)`), between the three prior density estimates. Quartile values are based on 321 forecast dates with valid option data (Jan. 1996–Jan. 2023).

	$\hat{G}_{LN}$			$\hat{G}_{ECDF}$			$\hat{G}_{FHS}$		
	p25	p50	p75	p25	p50	p75	p25	p50	p75
Mean	0.05	0.08	0.10	0.03	0.05	0.07	0.02	0.04	0.07
St. Dev.	0.10	0.14	0.19	0.12	0.16	0.21	0.12	0.15	0.20
Skewness	-0.59	-0.41	-0.27	-1.33	-0.98	-0.73	-1.42	-1.01	-0.78
Kurtosis	3.35	3.52	3.77	6.26	7.27	10.00	5.75	6.80	9.36
VaR_{90}	-0.07	-0.05	-0.03	-0.07	-0.05	-0.04	-0.07	-0.05	-0.04
VaR_{10}	0.04	0.05	0.07	0.04	0.05	0.07	0.04	0.05	0.07
$\text{KL}(\cdot\|\|\hat{F}_{LN})$	0.0233	0.0414	0.0672	0.0163	0.0246	0.0380	0.0149	0.0326	0.0584
$\text{KL}(\cdot\|\|\hat{F}_{ECDF})$	0.0551	0.0751	0.1123	0.0035	0.0114	0.0237	0.0249	0.0507	0.1015
$\text{KL}(\cdot\|\|\hat{F}_{FHS})$	0.0473	0.0628	0.0970	0.0179	0.0277	0.0405	0.0035	0.0175	0.0404
# Infeasible		1			1			1	
	$\hat{Q}_{LN}$			$\hat{Q}_{ECDF}$			$\hat{Q}_{FHS}$		
	p25	p50	p75	p25	p50	p75	p25	p50	p75
Mean	-0.04	-0.02	0.01	-0.04	-0.02	0.01	-0.04	-0.02	0.01
St. Dev.	0.13	0.17	0.21	0.14	0.18	0.23	0.14	0.18	0.23
Skewness	-1.10	-0.85	-0.64	-1.97	-1.63	-1.23	-2.48	-1.90	-1.44
Kurtosis	3.51	3.86	4.30	7.49	9.35	10.76	8.47	11.32	15.17
VaR_{90}	-0.09	-0.07	-0.05	-0.08	-0.06	-0.05	-0.08	-0.06	-0.05
VaR_{10}	0.04	0.05	0.06	0.03	0.05	0.06	0.04	0.05	0.06
$\text{KL}(\cdot\|\|\hat{F}_{LN})$	0.0601	0.0931	0.1505	0.0354	0.0538	0.0774	0.0409	0.0679	0.1034
$\text{KL}(\cdot\|\|\hat{F}_{ECDF})$	0.1042	0.1441	0.1800	0.0131	0.0234	0.0404	0.0432	0.0724	0.1327
$\text{KL}(\cdot\|\|\hat{F}_{FHS})$	0.0810	0.1105	0.1518	0.0297	0.0431	0.0644	0.0197	0.0393	0.0657
# Infeasible		1			1			1	

Tab. 6: Forecast mean, standard deviation, skewness, kurtosis and pairwise KL divergence between the three posterior density estimates (reported as `KL(col, row)`), for the TSD ($\hat{G}$) and risk-neutral () kernels. Quartile values are based on 321 forecast dates with valid option data (Jan. 1996–Jan. 2023).

	$\hat{F}_{LN}$	$\hat{G}_{LN}$	$\hat{Q}_{LN}$	$\hat{F}_{ECDF}$	$\hat{G}_{ECDF}$	$\hat{Q}_{ECDF}$	$\hat{F}_{FHS}$	$\hat{G}_{FHS}$	$\hat{Q}_{FHS}$
LRT$(\cdot, \hat{F}_{LN})$	-	5.06	3.68	4.67	5.00	4.24	4.70	5.21	4.18
	-	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LRT$(\cdot, \hat{F}_{ECDF})$	-4.67	0.88	0.98	-	3.27	2.32	1.23	2.90	2.29
	(1.00)	(0.19)	(0.16)	-	(0.00)	(0.01)	(0.11)	(0.00)	(0.01)
LRT$(\cdot, \hat{F}_{FHS})$	-4.70	-0.21	-0.18	-1.23	0.69	0.37	-	1.82	1.67
	(1.00)	(0.58)	(0.57)	(0.89)	(0.24)	(0.36)	-	(0.03)	(0.05)

Tab. 7: LRT statistics: test statistics are reported for LRT(col, row), with p-values in brackets. For example, for the lognormal prior, the LRT of the posterior, $\hat{G}_{LN}$, to the prior, $\hat{F}_{LN}$, is 5.06.

		$\hat{F}_{LN}$			$\hat{F}_{ECDF}$			$\hat{F}_{FHS}$		
	Index	$\hat{F}$	$\hat{G}$	$\hat{Q}$	$\hat{F}$	$\hat{G}$	$\hat{Q}$	$\hat{F}$	$\hat{G}$	$\hat{Q}$
Prot. Puts		0.11	0.01	0.92	0.05	0.00	0.93	0.14	0.02	0.90
Avg Mnes		1.00	0.95	0.99	1.00	1.00	0.99	1.00	0.94	0.99
Std. Dev. Mnes		0.01	0.08	0.01	0.01	0.00	0.02	0.01	0.07	0.02
Avg IV		0.24	0.46	0.18	0.19	0.15	0.18	0.19	0.40	0.19
Std. Dev. IV		0.13	0.44	0.09	0.07	0.00	0.09	0.11	0.26	0.09
Cov. Calls		0.00	0.07	0.55	0.02	0.11	0.55	0.36	0.37	0.63
Avg Mnes		0.00	1.02	1.01	1.01	1.02	1.01	1.01	1.01	1.01
Std. Dev. Mnes		0.00	0.02	0.01	0.01	0.02	0.01	0.01	0.01	0.01
Avg IV		0.00	0.26	0.17	0.30	0.23	0.18	0.19	0.20	0.18
Std. Dev. IV		0.00	0.16	0.08	0.16	0.13	0.08	0.06	0.09	0.08
Ann. Mean (%)	7.88	6.19	7.13	1.38	7.60	7.80	2.56	7.69	7.79	3.06
Ann. St. Dev. (%)	17.58	16.52	16.80	7.37	17.04	16.81	7.65	14.82	15.66	7.54
Ann. CER_1(%)	6.24	4.73	5.61	1.11	6.05	6.28	2.26	6.51	6.46	2.77
Ann. CER_2(%)	4.52	3.17	4.00	0.84	4.41	4.67	1.98	5.27	5.05	2.49
Ann. CER_4(%)	0.58	-0.45	0.25	0.30	0.62	0.91	1.39	2.36	1.72	1.93
LSW10 p-val		0.48	0.89	0.00	0.84	0.86	0.00	0.07	0.24	0.00

Tab. 8: **Option Combinations: Composition and Performance.** Returns to a portfolio consisting of the index and max one protective put and max one covered call, where the included options (if any) are selected through expected utility maximization, under the relevant forecast. The Prot. Puts row records the fraction of months that a protective put is included. The Cov. Calls row records the fraction of months that a covered call is sold. CIPDEs are estimated using OTM Puts and Calls (absolute delta range 0.15-0.5). Also reported is the LSW10 p-value for the null that the option-enhanced index dominates the passive index.

	$\hat{F}_{LN}$	$\hat{G}_{LN}$	$\hat{Q}_{LN}$	$\hat{F}_{ECDF}$	$\hat{G}_{ECDF}$	$\hat{Q}_{ECDF}$	$\hat{F}_{FHS}$	$\hat{G}_{FHS}$	$\hat{Q}_{FHS}$
Excl. High VIX									
LRT$(\cdot, \hat{F}_{LN})$	-	3.89***	2.92***	3.84***	4.06***	3.46***	3.94***	4.34***	3.54***
LRT$(\cdot, \hat{F}_{ECDF})$	-3.84	0.29	0.61	-	2.53**	1.85**	1.05	2.37**	2.02**
LRT$(\cdot, \hat{F}_{FHS})$	-3.94	-0.53	-0.31	-1.05	0.45	0.24	-	1.49*	1.54*
Excl. Recessions									
LRT$(\cdot, \hat{F}_{LN})$	-	4.72***	3.33***	4.33***	4.60***	3.79***	4.76***	5.03***	4.11***
LRT$(\cdot, \hat{F}_{ECDF})$	-4.33	0.96	0.69	-	3.01***	1.80**	1.34*	3.24***	2.19**
LRT$(\cdot, \hat{F}_{FHS})$	-4.76	-0.02	-0.29	-1.34	0.86	0.26	-	1.46*	0.65
Excl. High Fin. Unc.									
LRT$(\cdot, \hat{F}_{LN})$	-	4.62***	3.45***	4.31***	4.54***	3.86***	4.60***	4.91***	4.14***
LRT$(\cdot, \hat{F}_{ECDF})$	-4.31	0.63	0.96	-	2.81***	2.04**	0.88	2.69***	2.44**
LRT$(\cdot, \hat{F}_{FHS})$	-4.60	-0.01	0.21	-0.88	1.04	0.70	-	1.50*	1.15
Alt. Filters									
LRT$(\cdot, \hat{F}_{LN})$	-	4.49***	3.49***	4.67***	4.76***	4.06***	4.70***	4.96***	4.04***
LRT$(\cdot, \hat{F}_{ECDF})$	-4.67	0.34	0.32	-	2.78***	1.78**	1.23	2.43**	2.06**
LRT$(\cdot, \hat{F}_{FHS})$	-4.70	-0.61	-0.66	-1.23	0.42	0.05	-	1.38*	1.43*

Tab. 9: LRT Robustness Analysis: LRT tests are repeated in three alternative samples, excluding (i) the 10% of months with the highest VIX on the trade date (ii) all months the US economy is classified as being in recession by the NBER and (iii) the 10% of months with the highest financial uncertainty on the trade date. The Alt. Filters results are generated using a more restrictive filter on OTM options where abs(delta) of included options must be in the range [0.2,0.5].

	Index	$\hat{F}_{LN}$ $\hat{F}$	$\hat{G}$	$\hat{Q}$	$\hat{F}_{ECDF}$ $\hat{F}$	$\hat{G}$	$\hat{Q}$	$\hat{F}_{FHS}$ $\hat{F}$	$\hat{G}$	$\hat{Q}$
Excl. High VIX										
Ann. Mean (%)	4.16	3.78	4.17	0.92	4.55	4.72	2.00	4.52	4.71	2.13
Ann. St. Dev. (%)	15.99	15.48	15.78	6.60	15.73	15.73	7.00	14.08	14.73	6.88
CER_1(%)	2.80	2.50	2.84	0.70	3.23	3.39	1.75	3.45	3.53	1.89
CER_2(%)	1.32	1.11	1.40	0.48	1.80	1.96	1.51	2.28	2.25	1.66
CER_4(%)	-2.08	-2.11	-1.95	0.05	-1.52	-1.38	1.01	-0.51	-0.77	1.18
LSW10 p-val		0.60	0.93	0.00	0.93	0.93	0.00	0.14	0.27	0.00
Excl. Recessions										
Ann. Mean (%)	10.40	9.73	10.47	2.17	10.66	11.00	3.62	10.07	10.70	3.90
Ann. St. Dev. (%)	14.07	13.42	13.85	6.51	13.73	13.79	6.82	11.75	12.60	6.77
CER_1(%)	9.36	8.79	9.46	1.95	9.66	9.99	3.39	9.33	9.85	3.67
CER_2(%)	8.35	7.88	8.48	1.75	8.71	9.03	3.16	8.64	9.05	3.44
CER_4(%)	6.17	5.90	6.36	1.33	6.62	6.92	2.70	7.11	7.27	2.99
LSW10 p-val		0.50	0.94	0.00	0.87	0.94	0.00	0.08	0.23	0.00
Excl. High Fin. Unc.										
Ann. Mean (%)	9.02	8.39	8.93	1.90	9.08	9.48	3.15	8.96	9.36	3.57
Ann. St. Dev. (%)	14.72	14.10	14.50	7.02	14.40	14.43	7.37	12.60	13.41	7.20
CER_1(%)	7.88	7.35	7.82	1.65	7.99	8.38	2.88	8.12	8.40	3.30
CER_2(%)	6.76	6.32	6.73	1.41	6.92	7.30	2.61	7.31	7.47	3.05
CER_4(%)	4.33	4.10	4.37	0.93	4.59	4.95	2.08	5.51	5.42	2.54
LSW10 p-val		0.51	0.93	0.00	0.83	0.94	0.00	0.11	0.29	0.00
Alt. Filters										
Ann. Mean (%)	7.88	6.19	7.11	2.14	7.60	8.11	2.67	7.65	7.89	2.87
Ann. St. Dev. (%)	17.58	16.52	16.69	7.15	17.04	16.75	7.58	14.80	16.11	7.61
CER_1(%)	6.24	4.73	5.61	1.88	6.05	6.60	2.38	6.47	6.49	2.58
CER_2(%)	4.52	3.17	4.01	1.63	4.41	5.00	2.10	5.24	5.00	2.29
CER_4(%)	0.58	-0.45	0.29	1.13	0.62	1.26	1.53	2.33	1.50	1.71
LSW10 p-val		0.48	0.81	0.00	0.85	0.87	0.00	0.07	0.26	0.00

Tab. 10: Investment Performance: Robustness Analysis. Option trading performance is evaluated in three alternative samples, excluding (i) the 10% of months with the highest VIX on the trade date (ii) all months the US economy is classified as being in recession by the NBER and (iii) the 10% of months with the highest financial uncertainty on the trade date. The Alt. Filters results are generated using a more restrictive filter on OTM options where abs(delta) of included options must be in the range [0.2,0.5]. Also reported is the LSW10 p-value for the null that the option-enhanced index dominates the passive index.

Department of Economics
Athens University of Economics and Business

List of Recent Working Papers

2023

01-23 Real interest rate and monetary policy in the post Bretton Woods United States, George C. Bitros and Mara Vidali
02-23 Debt targets and fiscal consolidation in a two-country HANK model: the case of Euro Area, Xiaoshan Chen, Spyridon Lazarakis and Petros Varthalitis
03-23 Central bank digital currencies: Foundational issues and prospects looking forward, George C. Bitros and Anastasios G. Malliaris
04-23 The State and the Economy of Modern Greece. Key Drivers from 1821 to the Present, George Alogoskoufis
05-23 Sparse spanning portfolios and under-diversification with second-order stochastic dominance, Stelios Arvanitis, Olivier Scaillet, Nikolas Topaloglou
06-23 What makes for survival? Key characteristics of Greek incubated early-stage startup(per)s during the Crisis: a multivariate and machine learning approach, Ioannis Besis, Ioanna Sapfo Pepelasis and Spiros Paraskevas
07-23 The Twin Deficits, Monetary Instability and Debt Crises in the History of Modern Greece, George Alogoskoufis
08-23 Dealing with endogenous regressors using copulas; on the problem of near multicollinearity, Dimitris Christopoulos, Dimitris Smyrnakis and Elias Tzavalis
09-23 A machine learning approach to construct quarterly data on intangible investment for Eurozone, Angelos Alexopoulos and Petros Varthalitis
10-23 Asymmetries in Post-War Monetary Arrangements in Europe: From Bretton Woods to the Euro Area, George Alogoskoufis, Konstantinos Gravas and Laurent Jacque
11-23 Unanticipated Inflation, Unemployment Persistence and the New Keynesian Phillips Curve, George Alogoskoufis and Stelios Giannoulakis
12-23 Threshold Endogeneity in Threshold VARs: An Application to Monetary State Dependence, Dimitris Christopoulos, Peter McAdam and Elias Tzavalis
13-23 A DSGE Model for the European Unemployment Persistence, Konstantinos Giakas
14-23 Binary public decisions with a status quo: undominated mechanisms without coercion, Efthymios Athanasiou and Giacomo Valletta
15-23 Does Agents' learning explain deviations in the Euro Area between the Core and the Periphery? George Economides, Konstantinos Mavrigiannakis and Vanghelis Vassilatos
16-23 Mild Explocivity, Persistent Homology and Cryptocurrencies' Bubbles: An Empirical Exercise, Stelios Arvanitis and Michalis Detsis
17-23 A network and machine learning approach to detect Value Added Tax fraud, Angelos Alexopoulos, Petros Dellaportas, Stanley Gyoshev, Christos Kotsogiannis, Sofia C. Olhede, Trifon Pavkov
18-23 Time Varying Three Pass Regression Filter, Yiannis Dendramis, George Kapetanios, Massimiliano Marcellino

19-23 From debt arithmetic to fiscal sustainability and fiscal rules: Taking stock, George Economides, Natasha Miouli and Apostolis Philippopoulos
20-23 Stochastic Arbitrage Opportunities: Set Estimation and Statistical Testing, Stelios Arvanitis and Thierry Post
21-23 Behavioral Personae, Stochastic Dominance, and the Cryptocurrency Market, Stelios Arvanitis, Nikolas Topaloglou, and Georgios Tsomidis
22-23 Block Empirical Likelihood Inference for Stochastic Bounding: Large Deviations Asymptotics Under *m*-Dependence, Stelios Arvanitis and Nikolas Topaloglou
23-23 A Consolidation of the Neoclassical Macroeconomic Competitive General Equilibrium Theory via Keynesianism (Part 1 and Part 2), Angelos Angelopoulos
24-23 Limit Theory for Martingale Transforms with Heavy-Tailed Noise, Stelios Arvanitis and Alexandros Louka

2024

01-24 Market Timing & Predictive Complexity, Stelios Arvanitis, Foteini Kyriazi, Dimitrios Thomakos
02-24 Multi-Objective Frequentistic Model Averaging with an Application to Economic Growth, Stelios Arvanitis, Mehmet Pinar, Thanasis Stengos, Nikolas Topaloglou
03-24 State dependent fiscal multipliers in a Small Open Economy, Xiaoshan Chen, Jilei Huang, Petros Varthalitis
04-24 Public debt consolidation: Aggregate and distributional implications in a small open economy of the Euro Area, Eleftherios-Theodoros Roumpanis
05-24 Intangible investment during the Global Financial Crisis in the EU, Vassiliki Dimakopoulou, Stelios Sakkas and Petros Varthalitis
06-24 Time will tell! Towards the construction of instantaneous indicators of different agent-types, Iordanis Kalaitzoglou, Stelios Arvanitis
07-24 Norm Constrained Empirical Portfolio Optimization with Stochastic Dominance: Robust Optimization Non-Asymptotics, Stelios Arvanitis
08-24 Asymptotics of a QLR-type test for optimal predictive ability, Stelios Arvanitis
09-24 Strongly Equitable General Equilibrium Allocations, Angelos Angelopoulos
10-24 The Greek macroeconomy: A note on the current situation and future outlook, Apostolis Philippopoulos
11-24 Evolution of Greek Tax System, A Survey of Legislated Tax Changes from 1974 to 2018, Panagiotis Asimakopoulos
12-24 Macroeconomic Impact of Tax Changes, The case of Greece from 1974 to 2018, Panagiotis Asimakopoulos
13-24 `Pareto, Edgeworth, Walras, Shapley' Equivalence in a Small Economy, Angelos Angelopoulos
14-24 Stimulating long-term growth and welfare in the U.S, James Malley and Apostolis Philippopoulos
15-24 Distributionally Conservative Stochastic Dominance via Subsampling, Stelios Arvanitis
16-24 Before and After the Political Transition of 1974. Institutions, Politics, and the Economy of Post-War Greece, George Alogoskoufis
17-24 Gaussian Stochastic Volatility, Misspecified Volatility Filters and Indirect Inference Estimation, Stelios Arvanitis, Antonis Demos
18-24 Endogenous Realtor Intermediation and Housing Market Liquidity, Miroslav Gabrovski, Ioannis Kospentaris, Victor Ortego-Marti
19-24 Universal Choice Spaces and Expected Utility: A Banach-type Functorial Fixed Point, Stelios Arvanitis, Pantelis Argyropoulos, Spyros Vassilakis

2025

01-25 Democracy, redistribution, and economic growth: Some evidence from post-1974 Greece, George C. Bitros

02-25 Assessing Downside Public Debt Risks in an Environment of Negative Interest Rates Growth Differentials, Yiannis Dendramis, Georgios Dimitrakopoulos and Elias Tzavalis

03-25 Wicksellian General Equilibrium, Angelos Angelopoulos

04-25 Institutions and Ethics, Angelos Angelopoulos

05-25 General Equilibrium in a Capitalist Economy, Angelos Angelopoulos

06-25 Public investment multipliers revisited: The role of production complementarities, Vasiliki Dimakopoulou, George Economides, and Apostolis Philippopoulos

07-25 The heterogeneous causal effects of the EU's Cohesion Fund, Angelos Alexopoulos, Ilias Kostarakos, Christos Mylonakis, and Petros Varthalitis

08-25 Gaussian Invariant Markov Chain Monte Carlo, Michalis K. Titsias, Angelos Alexopoulos , Siran Liu, and Petros Dellaportas

09-25 Financial Intermediation, Investment, and Monetary Policy, Ioannis Kospentaris and Florian Madison

10-25 Temperature as a premium factor. Insurance incorporating climate change. A comparative case study between Europe and North America, Nomikou-Lazarou Irene

11-25 How High will our Buildings Become? Angelos Angelopoulos

12-25 Emergency Room General Equilibrium, Angelos Angelopoulos

13-25 Updating Density Estimates using Conditional Information Projection: Stock Index Returns and Stochastic Dominance Relations, Stelios Arvanitis, Richard McGee, and Thierry Post

Department of Economics
Athens University of Economics and Business

The Department is the oldest Department of Economics in Greece with a pioneering role in organising postgraduate studies in Economics since 1978. Its priority has always been to bring together highly qualified academics and top quality students. Faculty members specialize in a wide range of topics in economics, with teaching and research experience in world-class universities and publications in top academic journals.

The Department constantly strives to maintain its high level of research and teaching standards. It covers a wide range of economic studies in micro-and macroeconomic analysis, banking and finance, public and monetary economics, international and rural economics, labour economics, industrial organization and strategy, economics of the environment and natural resources, economic history and relevant quantitative tools of mathematics, statistics and econometrics.

Its undergraduate program attracts high quality students who, after successful completion of their studies, have excellent prospects for employment in the private and public sector, including areas such as business, banking, finance and advisory services. Also, graduates of the program have solid foundations in economics and related tools and are regularly admitted to top graduate programs internationally. Three specializations are offered:1. Economic Theory and Policy, 2. Business Economics and Finance and 3. International and European Economics. The postgraduate programs of the Department (M.Sc and Ph.D) are highly regarded and attract a large number of quality candidates every year.

For more information:

https://www.dept.aueb.gr/en/econ/