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 latent conditional density functions by projecting a prior time-series estimator onto distributions that satisfy a set of conditional moment conditions with functional nuisance parameters. The derivation of limit theory coupled with information-theoretic results characterizes the estimator and its improvements over the prior estimator. Theoretically, CIPDE is shown to achieve a lower limiting relative entropy to the latent distribution, provided that the prior is inconsistent and the moment conditions are well specified. An application to stock index options is presented using conditional moment restrictions based on market prices and pricing restrictions for index options. CIPDE is shown to enhance index return density forecasts out-of-sample and improve the out-of-sample investment performance of index option strategies by better timing protective put purchases and covered call writing.
Keywords: Conditional density estimation, information projection, Stochastic Dominance, stock index returns, stock index options

1 Introduction

Financial decision making in risk management, derivative pricing, and portfolio optimization often requires conditional density estimates of asset returns. Approaches

[*]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

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. This results in a new density estimate that integrates information from both the prior estimate and the moment conditions.

The divergence-based projection approaches are rooted in a rich literature spanning econometrics, statistics, and information theory [1, 2, 3, 4, 5] 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 inequities and include functional nuisance parameters such as the ubiquitous Stochastic Discount Factor (SDF).

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 regularity conditions.

We apply CIPDE to estimate the conditional distribution of monthly S&P 500 index (SPX) returns and optimize combinations of Chicago Board Options Exchange (CBOE) SPX stock index 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. Furthermore, payoffs from concurrent option series are driven solely by index price returns. This natural single-factor structure avoids the curse of dimensionality of a model-free approach.

The conditional moment conditions are based on observed market prices for SPX options and general option pricing restrictions that exclude SD relations in the spirit of [6], henceforth CJP09. The optimization of option combinations is inspired by [7, 8], 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 (unless the moment conditions are imposed for a subset of tradable options).

The aforementioned index option studies employed relatively simple conditional density estimates that are inconsistent because they fix or restrict the variation of the shape of the distribution and its parameters and use a minimal 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.

CIPDE significantly improves the density forecasts out-of-sample (OOS) for a range of prior density forecasts. It tends to reduce downside risk and increase upside potential in the estimated distribution. These adjustments lead to significant improvements in the OOS forecasting ability.

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, [9] 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, [9] 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.[1]

[11] 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 improvements upon EL that can be achieved using CIPDE, 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 applies the methodology to the estimation of the conditional distribution of monthly S&P 500 index returns and optimization of SPX option combinations. Section 4 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 valid empirical conditional moment inequalities. The estimator belongs to the class of information projection-based methods, relying on the minimization of the KL divergence subject to moment constraints. Unlike classical Bayesian updating which uses a likelihood function, this procedure integrates information through conditional moment restrictions.

The setting is conditional, with the conditioning variables taking values in a Euclidean space. For notational economy, we suppress this dependence, though all results are interpreted as holding for almost every value of the conditioning variables.

Let F denote the true conditional distribution with compact support $\mathcal{S} \subset \mathbb{R}^p$, assumed independent of the conditioning variables. In our application, a univariate distribution ($p = 1$) is used, but the method also applies to multivariate cases

[1] In a related study, [10] pioneers the application of Bayesian econometrics for model selection and validation in empirical asset pricing. This study develops a formal statistical framework, but is not focused on estimating the risk neutral distribution implemented with the option or the physical distribution of index returns.

($p \geq 2$), although the curse of dimensionality presumably makes direct applicability in high-dimensional applications elusive; see the Conclusions.

The prior $\hat{F}$, for example, a kernel density or a truncated log-normal estimator, is assumed almost surely to be absolutely continuous with respect to the Lebesgue measure.

To incorporate the empirical moment inequalities, we introduce a functional parameter $m \in \mathcal{M}$, where $\mathcal{M} \subset W^{2,2}(\mathcal{S})$, a convex and closed subset of a Sobolev space of sufficiently smooth functions-see, for example, [12]. Moment inequalities take the form:

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

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 set of admissible densities that satisfy the moment inequalities and are absolutely continuous with respect to $\hat{F}$ at the functional parameter m is denoted:

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

where $\ll$ denotes absolute continuity. This set is convex and closed under both weak and total variation topology due to the linearity of the integral and compactness of $\mathcal{S}$.

The asymptotic properties of the CIPDE estimator $\hat{G}$ as a sample size $T \to \infty$ are also investigated in the following. This 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.

All results assume a probabilistic data-generating process (DGP) and a well-defined conditional distribution F, while the CIPDE acts as a correction mechanism for an initial, possibly inconsistent, estimator $\hat{F}$. The CIPDE optimizes over a set of plausible functional moment parameters $m \in \mathcal{M}$, allowing the incorporation of the economic structure in the form of SDFs in our application.

Some useful further assumptions and notation are introduced: all densities are assumed to be bounded; they are thus elements of $\ell^\infty(\mathcal{S})$ the space of bounded real functions defined on the compact support equipped with the uniform metric. Then $\rightsquigarrow$ denotes weak convergence in $\ell^\infty(\mathcal{S})$, and $\overset{\text{epi}}{\rightsquigarrow}$ denotes epi-convergence in distribution; see [13].

2.2 The CIPDE Estimator

We define the CIPDE estimator as the solution to the following variational problem:

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

where $\chi_{\mathbb{M}}$ is the characteristic function of the admissible set, taking the value 0 inside and $+\infty$ outside.

This formulation highlights a key novelty: the estimation involves joint optimization over the density G and the functional weight m, which in finance applications will correspond to SDFs. This dual structure reflects the idea that the moment functions themselves can vary within a plausible class, e.g., admissible SDFs in asset pricing.

Using duality and convex optimization arguments (e.g., [14], [15]), this 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:

$$\hat{G}(S) := \int_{-\infty}^{S} \hat{g}_{m_T}(x) dx,$$

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 can be thought of as playing 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 parameter.

Notice that if an additional prior distribution were available over the functional space $\mathcal{M}$-see, for example, [16], 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 a Bayesian I-projection averaging, contrasting the current one where the I-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)**.** *There exists a bounded non-sample-dependent density f_∞, such that $\hat{f} \rightsquigarrow f_\infty$ in $\ell^\infty(\mathcal{S})$.*

This condition ensures that the initial prior density estimator converges uniformly in probability to a deterministic limit f_∞, which may differ from the true DGP density f. It implies a Glivenko-Cantelli-type result; $\hat{F} \rightsquigarrow F_\infty$ in $\ell^\infty(\mathcal{S})$, where $F_\infty(x) := \int_{\mathcal{S}} 1_{(-\infty,x]}(S) f_\infty(S) dS$. F_∞ need not equal the DGP distribution F. In many applications, it can be expected that $F_\infty \notin \mathbb{M}(\mathbf{H}, F_\infty; m)$ for all $m \in \mathcal{M}$, 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 assumption 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 [17] 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, [18]. 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, [19]. Thus, Assumption 1 encompasses a wide range of estimation frameworks relevant to practitioners.

The second assumption is useful for asymptotic identification.

Assumption 2 (Functional Parameter Identification)**.** *Suppose 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*.*

Assumption 2 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* $\hat{G}_{m_T} \rightsquigarrow G_{m_\infty}$ *for a unique* $G_{m_\infty} \in \mathbb{M}(\mathbf{H}, F_\infty; m_\infty)$*. 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 [14]. Also due to Assumption 1, and via Skorokhod representations applicable due to Theorem 1 of [20], we have $\mathbb{M}(\mathbf{H}, \hat{F}, m) \overset{\mathrm{epi}}{\rightsquigarrow} \mathbb{M}(\mathbf{H}, F_\infty, m)$, 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 [13], 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 $r_T \to \infty$, $r_T(\hat{f} - f_\infty) \rightsquigarrow \mathcal{G}$ in $\ell^\infty(\mathbb{R})$, 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*

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

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

The assumption requires a Donsker-type result for $\hat{F}$. For the ECDF, it 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 [21]. The second part of the assumption holds whenever the J_∞ 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 along with the compactness of the support 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

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

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

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

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

where

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

while differentiability of the optimal limiting Lagrange multiplier follows from Assumptions 1-3, and the Implicit Function Theorem in Banach spaces, see for example

[22]. The previous expansion can be similarly seen to converge, locally uniformly w.r.t. $\boldsymbol{\zeta}$, to

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

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 $J_\infty = \emptyset$, 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.

Whenever $J_\infty \neq \emptyset$, then the CIPDE estimator not only may partially correct for the asymptotic KL inconsistency of $\hat{F}$ w.r.t. F, but it may also incorporate efficiency gains compared to $\hat{F}$ due to the information infused by the moment inequalities.

In the special case where $r_T = \sqrt{T}$, $\mathcal{G}$ is zero mean Gaussian, $J_\infty \neq \emptyset$, $\boldsymbol{\lambda}_{j,\infty} < 0, \forall j \in J_\infty$, 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 chi-square LR tests performed in Section 3.5, 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 [23] 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 [1] 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 Inference on SD

To motivate the application of CIPDE weights in SD analysis, the above statistical properties and improvements are translated into consequences for empirical decision errors regarding SD relations in the present section.

Here, $p = 1$, and $X = S$ and $Y = S + L(S)$, for L a measurable transformation, represent the continuous outcomes of two choice alternatives supported on the same compact $\mathcal{S}$ and $n \geq 1$ is the relevant integer order of SD. X dominates Y under F w.r.t. the n^{th}-order SD relation, denoted by $X \underset{F;n}{\succeq} Y$, iff, for any z in $\mathcal{S}$, $D(z, X, Y; n, F) \leq 0$, with $D(z, X, Y; n, F) := \int_{\mathcal{S}}((z-X(S))_+^{n-1} - (z-Y(S))_+^{n-1})f(S)dS$, the degree $(n-1)$ Lower Partial Moment (LPM) difference.

An empirically useful slack-augmented modification, $X \underset{F;n,\epsilon_T}{\succeq} Y$, is defined as $D(z, X, Y; n, F) \leq \epsilon_T$, for the slack variables $\epsilon_T > 0$, where $T \to \infty$, $\epsilon_T \to 0$, and $\sqrt{T}\epsilon_T \to \infty$.

The convergence of $\hat{G}$ to G implies the convergence of $D(z, X, Y; n, \hat{G})$ to the limiting $D(z, X, Y; n, G)$ uniformly in z due to Theorem 1, the compactness of $\mathcal{S}$ and the uniform (in z) boundedness of the intergrands $(z - X(S))_+^n - (z - Y(S))_+^n$ in the definition of the dominance relation:

$$\sup_{z\in\mathcal{S}} |D(z, X, Y; n, \hat{G}) - D(z, X, Y; n, G)| \rightsquigarrow 0.$$

Furthermore, if $X \underset{G;n}{\succeq} Y$, then Theorem 2, the compactness of $\mathcal{S}$, the uniform (in z) Lipschitz continuity of the intergrands $(z-X(S))_+^n - (z-Y(S))_+^n$, and the generalized Delta method-see [24] imply the following convergence result:

$$\sqrt{T}(\sup_{z\in\mathcal{S}} D(z, X, Y; n, \hat{G}) - \sup_{z\in\mathcal{S}} D(z, X, Y; n, G)) \rightsquigarrow \sup_{z\in\mathcal{S}_\infty} D(z, X, Y; n, \Lambda(z)),$$

where $\mathcal{S}_\infty = \arg\sup_{z\in\mathcal{S}} D(z, X, Y; n, G)$. This directly implies that, for the slack-augmented dominance relation,

$$\mathbb{P}(X \underset{\hat{G};n,\epsilon_T}{\succeq} Y | X \underset{G;n}{\succeq} Y) \to 1,$$

and

$$\mathbb{P}(X \underset{\hat{G};n,\epsilon_T}{\succeq} Y | X \underset{G;n}{\not\succeq} Y) \to 0.$$

As a consequence, the possible errors in deciding whether $X \underset{G;n}{\succeq} Y$ holds via the empirical slack-augmented relation $X \underset{\hat{G};n,\epsilon_T}{\succeq} Y$ become asymptotically negligible in probability. The same holds if, under Assumption 3, G is replaced by F_∞ and $\hat{G}$ by $\hat{F}$ in the above, or if G is replaced by F and $\hat{G}$ by any $\sqrt{T}$-consistent estimator of F.

As discussed in Section 3.1, it is often the case that $G \neq F$ due to the use of a statistically inconsistent prior density estimator. In this case, the research interest lies on whether the reduction of the Kullback-Liebler inconsistency obtained by CIPDE compared to $\hat{F}$, improves the empirical decision properties regarding the truth of $X \underset{F;n}{\succeq} Y$. This question is addressed via Theorem 3 and the comparison between $|D(z, X, Y; n, F) - D(z, X, Y; n, G)|$ and $|D(z, X, Y; n, F) - D(z, X, Y; n, F_\infty)|$.

First, using Theorem 3, Pinsker's inequality, as in [25], the uniform Lipschitz properties of the intergrands $(z - X(S))_+^n - (z - Y(S))_+^n$, and the representation of total variation as an integral probability metric, we obtain that:

$$\sup_{z\in\mathcal{S}} |D(z, X, Y; n, F) - D(z, X, Y; n, G)| \leq C_1 \sqrt{\mathrm{KL}(F\|F_\infty) - \mathrm{KL}(G\|F_\infty)},$$

where $C_1 := \sqrt{2}\mathrm{diam}(\mathcal{S})$. This can also be expressed as:

$$\sup_{z\in\mathcal{S}} |D(z, X, Y; n, F) - D(z, X, Y; n, G)| \leq C_2 \sqrt{\mathrm{KL}(F\|F_\infty)},$$

where $C_2 := \sqrt{2}\text{diam}(\mathcal{S})\sqrt{1 - \frac{\text{KL}(G\|F_\infty)}{\text{KL}(F\|F_\infty)}}$.

Second, under the assumption that $0 < \inf \frac{dF_\infty}{dF} < 1 < \sup \frac{dF_\infty}{dF}$, where the infimum and the supremum are taken over the support of F_∞, the inverse Pinsker inequality, as in Theorem 1 of [26], implies that:

$$\sqrt{\text{KL}(F\|F_\infty)} \leq C_3 |D(z, X, Y; n, F) - D(z, X, Y; n, F_\infty)|,$$

where $C_3 := \frac{2\sqrt{\sup_{g\in\mathcal{F}} |\mathbb{E}_F(g) - \mathbb{E}_{F_\infty}(g)|}}{|D(z,X,Y;n,F) - D(z,X,Y;n,F_\infty)|}$ and $\mathcal{F}$ is the set of Lipschitz continuous real functions with Lipschitz coefficient bounded above by one. Taking into account that $\sup_{g\in\mathcal{F}} |\mathbb{E}_F(g) - \mathbb{E}_{F_\infty}(g)| \leq 2\text{diam}(\mathcal{S})$ it is finally obtained that:

Lemma 2. *Under the premises of Theorem 3 and if*

$$0 < \inf \frac{dF_\infty}{dF} < 1 < \sup \frac{dF_\infty}{dF}$$

and $|D(z, X, Y; n, F) - D(z, X, Y; n, F_\infty)| \geq 2\text{diam}(\mathcal{S})\sqrt{1 - \frac{\text{KL}(G\|F_\infty)}{\text{KL}(F\|F_\infty)}}\left(\frac{\inf \frac{dF_\infty}{dF} \ln(\inf \frac{dF_\infty}{dF})}{1 - \inf \frac{dF_\infty}{dF}} + \frac{\sup \frac{dF_\infty}{dF} \ln(\sup \frac{dF_\infty}{dF})}{\sup \frac{dF_\infty}{dF} - 1}\right)$, *then*

$$|D(z, X, Y; n, F) - D(z, X, Y; n, G)| < |D(z, X, Y; n, F) - D(z, X, Y; n, F_\infty)|. \quad (8)$$

The previous imply that if $\hat{f}$ is bounded away from zero at the support w.h.p.- consider for example the truncated log-normal density, and the divergence of G from F_∞ is similar to that of the DGP F from F_∞, then utilizing the conditions that check whether $X \underset{\hat{G};n,\epsilon_T}{\succeq} Y$ instead of $X \underset{\hat{F};n,\epsilon_T}{\succeq} Y$ could introduce improvements in the probability of false dominance classifications. Extending these types of results to situations where $\alpha_F = 0$ could be achieved by suitably broadening the inverse Pinsker inequality to cover distributions with these densities; these more complex issues are reserved for future investigation.

Inequality (8) helps link the partial inconsistency correction of $\hat{G}$ to enhanced inference properties regarding SD. For example, if for some z, $D(z, X, Y; n, F) > c$, with c equal the strictly positive difference between the rhs and the lhs of (8), while $D(z, X, Y; n, F_\infty) < 0$, then $D(z, X, Y; n, G) > 0$ and thereby using the previous:

$$\mathbb{P}(X \underset{\hat{G};n,\epsilon_T}{\succeq} Y | X \underset{F;n}{\not\succeq} Y) \to 0.$$

Hence, if dominance does not hold due to a large enough violation of some of the LPM conditions, this violation will eventually be picked up by LPMs based on the partially inconsistency corrected CIPDE even if it is not asymptotically picked up by LPMs based on the initial inconsistent estimator $\hat{F}$.

3 Stock Index Returns and Options

The pricing and trading of index options is a promising application area for CIPDE and SD. 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. Furthermore, side information from option theory is available to translate market prices to conditional moment conditions.

Additionally, the payoffs at expiry of concurrent option series are driven solely by the value of the index at the option expiration date. This natural single-factor structure avoids the curse of dimensionality and allows for model-free estimation of the joint distribution and accurate approximation using discretized versions of the distribution for numerical purposes.

Since option payoffs are inherently asymmetric, mean-variance analysis is not appropriate and higher-order risk needs to be taken into account. A body of literature studies the pricing and trading of options in a SD framework. The monograph by [27] provides a comprehensive literature survey and research agenda.

CJP09 develop a system of model-free pricing restrictions for multiple concurrent options with different types (put or call) and different strikes. These restrictions are uncontroversial 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.

Violations of the pricing 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 option combinations. [7], and [8], develop and apply optimization problems to construct such option combinations when the option market is out of equilibrium.

From an econometric perspective, a limitation of these studies seems that they do not account for time-variation of higher-order risk and risk premiums which generally results in statistical inconsistency of the conditional distribution estimator. 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 analyzes the use of conditional information projection using option pricing restrictions as conditional moment conditions to achieve and exploit the statistical improvements that were discussed in the previous section. We analyze the statistical significance of the improvements in OOS forecasting ability. We analyze the economic significance using the OOS investment performance of option combinations that are optimized under the prior density estimates and the posterior density estimates.

We deviate from the original studies in two important ways. First, the option pricing system of [6] is tightened by requiring that the SDF is convex, which amounts to refining second-order SD to third-order SD. Second, the optimization problem is based on a notion of approximate dominance because the posterior estimates (which obey the pricing system) exclude exact dominance (given that 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.

3.1 Preliminaries

Stock index options are used in this study for extracting conditioning information from quotes market prices and to enhance the passive index using active positions in options that appear mispriced. The focus is on European-type options on the SPX. 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, where the price return R_S is treated as a random variable. The (annualized and maturity-matched) dividend yield to the index is R_D.

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 difficult 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. 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 expiration date is $(X_{i,T} - a_i(1 + R_B))$ for a bought option and $(b_i(1 + R_B) - X_{i,T})$ for a written option.

3.2 Data

Data for the one-month T-bill rate and dividend yield are obtained from Option Metrics IvyDB. S&P 500 Index and VIX values are downloaded from Thomson Reuters Eikon. We use two sources for SPX option price data: intra-day quotes captured at 15:45 ET from the iVolatility database and closing price quotes from the Option Metrics Ivy database. The intra-day data are available from Jan. 2004 through Jan. 2023. To extend the data sample, we backfill the intra-day quotes with closing price quotes from Option Metrics for the period Jan. 1996 - Dec. 2003. We include options with 29 calendar days to expiry ($T \approx 29/365$).[2]

We use various filters based on moneyness, option delta and option premium to select options that will be included in our analysis. Following the convention for empirical research on index options for purposes of estimating Implied Volatility (IV) surfaces and risk-neutral distributions, we exclude in-the-money (ITM) options, which tend to be less liquid than out-of-the money (OTM) options. We also require a minimum option delta of 0.05 for calls and a maximum delta of -0.05 for puts, to exclude deep OTM options. In addition, options with a bid price of less than 15 cents are excluded.

These option filters are applied both to define the moment conditions used to estimate the posterior density and to define the investment universe for optimizing option combinations. The analysis can be generalized by tightening the filters for estimation or loosening the filters for optimization if the objective is to detect mispriced options.

In four months, the sample trade dates fell on Thanksgiving Day, and the options exchange was closed (22nd Nov. 2001/2007/2012/2018). The absence of quotes on these months left $M = 321$ trade dates with valid option quotes for our analysis.

Table 1 presents summary statistics for the dataset of options that pass our 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 number of options (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, implied volatility (IV), and the bid-ask spread.

Consistent with the maturation of the options market, the number of quoted strikes increases and the spreads decrease over time, with the number of qualified options series rising above $N_P + N_C = 100$ and the quoted spread falling below 1% of

[2] Prior to 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.

the midpoint premium for both puts and calls in the most recent five years. These patterns suggest that the degree of option market efficiency appears to have increased and the profitability of active option trading appears to have fallen over time.

[Insert Table 1 about here.]

3.3 Prior density estimators

As prior density estimates, we consider three common methods for estimating the conditional distribution of stock index price returns: (i) a lognormal distribution with conditional mean and volatility estimates, (ii) a transformation of a historical unconditional Empirical Cumulative Distribution Function (ECDF), and (iii) a forecast density generated using GJR-GARCH with simulated returns from Filtered Historical Simulation (FHS).

The conditional means of all three forecast distributions and the conditional standard deviation of the lognormal and transformed ECDF are estimated using arguments and results from CJP09.

The conditional mean is based on the prevailing one-month Treasury bill rate (R_B), the estimated Market Risk Premium ($M\hat{R}P$) and the prevailing dividend yield (R_D): $\mathbb{E}_{\hat{F}}[R_S] = (R_B + \mathrm{M\hat{R}P} - R_D)$. Following CJP09, we employ an annualized $\mathrm{M\hat{R}P}$ of 4% and we verify that the results are highly robust to plausible variation of the $\mathrm{M\hat{R}P}$, presumably because option prices are more affected by the scale and shape of the distribution than the location.

The conditional volatility is estimated using the IV of the nearest-to-the-money call option, because IV-scaling was found to produce the best forecasts and least amount of option mispricing in CJP09. The results and conclusions in this study are robust to plausible variation in the prior conditional standard deviation estimator. However, it should be noted that the CIPDE problem may not have a solution if a very poor prior is used.

The conditional ECDF is based on the unconditional ECDF in a moving estimation window of 800 monthly returns prior to the trade date. A broad window is selected to capture as many historical tail events as possible. A regular grid for the price returns with 25 bps spacing in the range [-0.6, 0.6] is used. To obtain a conditional ECDF, translation and scale transformations are applied to the atoms of the unconditional ECDF to match the aforementioned target mean ($\mathbb{E}_{\hat{F}}[R_S]$) and target volatility (IV).

The GJR-GARCH FHS model is fitted following the procedure in [28].[3] On each trade date a model is fitted to the returns in a window spanning from 1990 to the trade date. A set of 10,000 simulated returns is then generated with the drift

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

set to the target mean described above, and an ECDF is estimated as the forecast distribution.

To allow for OOS evaluation, the discrete forecast distributions (the transformed unconditional ECDF and the FHS conditional ECDF) are smoothed using KDE to obtain continuous density estimates $p_h^*(S) := \frac{1}{h}\sum_{i=1}^{N}\pi_i^* k\left(\left(S_i - S\right)/h\right)$, using a Gaussian smoothing kernel function k and Silverman's rule for the bandwidth parameter h. The smoothing is essential, because the OOS realized returns are not atoms of the discrete distributions.

All three priors are consistent with our statistical theory. They converge uniformly in probability to non-stochastic pseudo-true limits under standard regularity conditions involving stationarity, ergodicity and mixing for the time series. For the lognormal, if the estimated conditional mean and volatility converge uniformly, it converges to a pseudo-true limiting lognormal. For the transformed and smoothed ECDF, convergence follows from the Glivenko–Cantelli theorem applied to the unconditional empirical distribution constructed from a fixed-size rolling window, with affine transformations preserving uniform convergence. Finally, for the FHS, as long as the GARCH model is pseudo-consistently estimated, and the empirical residuals approximate a well-defined ergodic process, then the smoothed ECDF converges uniformly to a pseudo-true limit. Hence, Assumption 1 is satisfied in all three cases.

As far as Assumption 3 is concerned, and under standard regularity conditions, the lognormal weakly converges at $\sqrt{T}$-rate via delta method involving arguments, as its underlying parameters (mean and variance) converge at standard parametric rates. The transformed and smoothed ECDF satisfies Donsker-type result under standard assumptions (e.g., strong mixing, Lipschitz continuous kernel) as discussed in [29]; the fixed bandwidth preserves the rate. For the FHS, the volatility model is estimated using quasi-maximum likelihood methods that ensure $\sqrt{T}$-rate convergence under regularity, and the simulation step with 10,000 paths effectively regularizes the empirical distribution; uniform-type empirical process weak convergence results would imply the validity of the assumption, under stationarity and mixing conditions; see for example [30].

Table 2 compares the three prior density estimators using the quartile boundaries (p25, p50, and p75), computed across the $M = 321$ months, for skewness, kurtosis, and the KL divergence between each pair of estimators. The ECDF and FHS tend to be significantly negatively skewed and fat-tailed. Since these features are most pronounced for FHS, it is not surprising that the largest divergence occurs between lognormal and lognormal. Additional properties of the prior estimates are discussed in the subsections below.

The prior density estimates fix or restrict the distributional shape and its parameters while using limited amounts of conditioning information (R_B, R_D and IV). These features likely make the priors biased and inefficient, allowing for improvements through conditional moment conditions based on index option market prices.

[Insert Table 2 about here.]

3.4 Posterior density estimation

CIPDE is based on 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 CJP09. 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):

$$\mathbb{E}_F[m(1 + R_S)] = 1; \tag{9}$$

$$\mathbb{E}_F[m(1 + R_B)] = 1; \tag{10}$$

$$b_i \leq \mathbb{E}_F[mX_{i,T}] \leq a_i;\ i = 1, \cdots, N. \tag{11}$$

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 only for options that pass our tight filters based on moneyness, option delta and option premium (see Section 3.1).

The SDF is generally not unique under these general conditions. It 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:

$$\left\{m \in \mathcal{C}^2 : m(S) \geq 0; m'(S) \leq 0; m''(S) \geq 0\right\} =: \mathcal{M} \ni m. \tag{12}$$

These SDFs are positive, increasing and convex, as required for the IMRS of standard utility functions. Although [6] do not require this property, convexity directly follows from the generally accepted property of 'prudence' or decreasing risk aversion for standard utility functions ([31]). Violations of the pricing system (9)-(12), or pricing errors, imply that the market index is dominated by third-degree SD by portfolios that enhance the index using certain option combinations.

If the convexity condition is not imposed, as in CJP09, the analysis allows for SDFs with a pathological shapes 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.

The pricing restrictions (9)-(12) are conditional moment conditions because the set of included options and the option prices are updated every month.

The CIPDE relies on the empirical counterparts of the moment conditions (9)-(12) 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 to a finite set of linear restrictions on the SDF values $\{m(S_{j,T})\}$, along the lines of [32] . 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, BETEL estimation becomes a Bi-convex Programming problems that is convex in the SDF values $\{m(S_{j,T})\}$ and convex in the probability weights $\{\hat{g}(S_{j,T})\}$.[4]

Table 3 summarizes the differences between the prior distributions and the posterior distributions using summary statistics (computed across the time series of monthly distribution estimates) for the boundaries of the nine deciles of the estimated (standardized) index return distribution. The boundaries are standardized using the prior mean and prior standard deviation, to facilitate comparison.

The posterior decile boundaries deviate from the prior decile boundaries because the prior distribution is inconsistent with the conditional moment conditions (9)-(12) based on market prices and SDF restrictions. The general pattern is that conditional information projection increases the lower and middle decile boundaries while decreasing the upper decile boundaries. As a result, the posterior has less

[4] The problems are solved using an Alternating Direction Method of Multiplier (ADMM) algorithm that alternates between an EL problem for estimating the probability weights given the SDF values and a Generalized Methods of Moments (GMM) problem for estimation the SDF values given the EL weights until convergence. The algorithm extends the two-step procedure by [33], which was designed for log-convex kernels $\left\{m \in \mathcal{M} : \ln(m(S))'' \geq 0\right\}$, and does not apply here, as $\mathcal{M}$ does not ensure the log-convex property which is required for the log-linearization of the kernel in that procedure.

downside risk and more upside potential than the prior distribution.

The bottom panel of the table displays the KL divergence from the prior distribution. Given the counterfactual skewness and kurtosis of the lognormal, it is not surprising that the divergence is largest for this prior. The small KL divergence for the ECDF suggests that this prior is has the highest consistency with the option quotes among the three specifications.

[Insert Table 3 about here.]

3.5 Goodness-of-fit tests

To evaluate the OOS forecasting ability, we use asymptotic chi-squared Goodness-of-Fit (GoF) tests for the proportions of realized returns that fall into the various deciles of the forecast distribution. We use $q_i = q := 1/10$ for the expected proportion and p_i for the realized proportion in a given decile, $i = 1, \ldots, 10$. For evaluating all 10 deciles jointly, the test statistic is:

$$\chi^2_{\text{All}} = -2T \sum_{i=1}^{10} q \log\left(\frac{q}{p_i}\right). \tag{13}$$

Separate tests were performed for the first decile (left tail) and last decile (right tail). In these cases, the test statistic reduces to

$$\chi^2_{LT} = -2T\left(q \log\left(\frac{p_1}{q}\right) + (1-q)\log\left(\frac{1-p_1}{1-q}\right)\right); \tag{14}$$

$$\chi^2_{RT} = -2T\left((1-q)\log\left(\frac{1-p_{10}}{1-q}\right) + q\log\left(\frac{p_{10}}{q}\right)\right). \tag{15}$$

We estimate critical values and p-values using a χ^2_9 distribution for χ^2_{All} and a χ^2_1 distribution for χ^2_{LT} and χ^2_{RT} . Very similar results were obtained using a re-centered block bootstrap or sub-sampling to account for possible serial dependence.

Table 4 summarizes the test results. The top panel shows the empirical percentage hit rates for each of the 10 deciles, where a percentage smaller than 10 indicates a deficiency relative to the expected frequency, and a percentage larger than 10 indicates an excess.

The most striking pattern is that the prior tends to show deficiency in the lower half and the top decile, while exhibiting excess in the upper-middle deciles. This reflects an overestimation of downside risk and an underestimation of upside potential. Consistent with the changes in the decile boundaries in Table 1, conditional information projection tends to reduce these local deficiencies and excesses, bringing the hit rates closer to the expected value of 10 percent per decile.

In the middle panel, the GoF test statistic is reported for the left tail ("Left"), the right tail ("Right"), and all deciles ("All"), with asterisks denoting the level of statistical significance achieved using the asymptotic chi-squared test. Similar significance levels are obtained using recentered bootstrapping and sub-sampling approaches.

The GoF test is rejected for all three priors. The lognormal prior stands out as a particularly poor density forecast. In all three cases, conditional information projection leads to a significant improvement in the goodness of fit. Among the three priors, the improvement from CIPDE is smallest for the ECDF, consistent with the low KL divergence in Table 3 and the low chi-square statistic for the prior.

The bottom panel includes the number of months for which the prior distribution passes the conditional, moment conditions (M_{Pass}). Also shown is the number of months for which the prior distribution provided such a poor fit to the option prices that a posterior distribution could not be found (M_{Inf}). In these cases, the prior could not be updated, and the posterior was equated with the prior.

Given the counterfactual shape of the lognormal, it is not surprising that insolvability occurs most frequently for this prior distribution. However, for the ECDF and GARCH priors, the prior can be updated in the vast majority (97–98%) of months.

[Insert Table 4 about here.]

3.6 Optimized option combinations

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

The analysis differs from [7] and [8] 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, the performance improvement under the posterior can stem from (i) avoiding spurious arbitrage opportunities that arise under the prior due to estimation error and (ii) improvement upon the general risk profile of 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 strongly risk-averse index investors when added as an overlay to the index.

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

$$\max \mathbb{E}_{\hat{G}} \left[u \left(S_T + \sum_{i=1}^{N} \alpha_i (X_{i,T} - a_i(1+R_B)) + \sum_{i=1}^{N} \beta_i (b_i(1+R_B) - X_{i,T}) \right) \right]; \quad (16)$$

$$\sum_{i=1}^{N} \alpha_i \leq 1; \quad (17)$$

$$\sum_{i=1}^{N} \beta_i \leq 1; \quad (18)$$

$$\alpha_i, \beta_i \in \{0,1\},\ i = 1, \cdots, N. \quad (19)$$

The binary variables $\{\alpha_i\}$ and $\{\beta_i\}$ represent the open long positions and open short 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 strongly risk-averse index investors. Therefore, they should lower expected return in equilibrium with a strongly risk-averse representative investor. Consequently, non-zero solutions for the option overlay are non-existent if the utility function is risk-neutral ($u(S_T) = S_T$) and the density estimate obeys the moment conditions (9)-(12) for qualified all options.[5] 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 strongly 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 ($\hat{G} \neq F$).

We use here the standard logarithmic utility function $u(S_T) = ln(S_T)$ which has a Relative Risk Aversion coefficient of RRA=1. The optimization can be seen as

[5] 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.

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 strongly risk-averse than assumed in the objective function. In this respect, an approximate dominance relation is obtained in the spirit of 'Almost Stochastic Dominance'.

For every month in our sample, we solve the Binary Convex Optimization problem for the various discretized prior and posterior density forecast models. For every model, we evaluate the consequences 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 effect of the forecast model using the option overlay composition and the OOS investment performance.

For verification purposes, we also solved the optimization problems for the risk-neutral objective function $u(S_T) = S_T$. Under all three prior distribution estimates, the in-sample optimal option combinations resulted in OOS underperformance for this specification. Furthermore, under all three posterior distribution estimates, no non-zero solutions were found. These results 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 non-zero solutions were found under the posterior).

3.7 Out-of-sample investment performance

Table 5 summarizes the composition and performance of the optimized option combinations for the various prior and posterior distributions. The top panel reports the frequency of buying protective puts and writing covered calls, as well as the mean and standard deviation of the associated moneyness and IV of the chosen options.

The optimal solution based on the prior distribution frequently buys protective puts and writes covered calls. By contrast, the posterior is more reluctant to take open option positions and tends to focus on options that are deeper OTM. This result is consistent with the reduction of downside risk and the increase in upside potential under the posterior. Among the three priors, the ECDF prior features the fewest open option positions, consistent with the small KL divergence in Table 3 and the low chi-square statistic in Table 4.

The middle panel summarizes OOS investment performance. Reported are the mean and standard deviation of realized returns, along with the computed Certainty Equivalent Return (CER) for a power utility function with RRA = 1, 2, 4.

Although the open option positions based on the prior reduce risk (as evidenced by the reduction in standard deviation), OOS performance deteriorates. Notably, the solutions trail the index in terms of CER1, which represents the logarithmic utility function, and the deterioration worsens as the number of open option positions increases. The performance decline is consistent with the overestimation of downside risk and underestimation of upside potential seen in Table 2.

By contrast, the optimal solution based on the posterior distribution outperforms the index and achieves improvements in CER across all three prior distributions and all three levels of risk aversion. These performance gains align with the improvements in goodness-of-fit observed in Table 4. They are smallest for the ECDF, as the number of open positions is smallest under this prior.

The bottom panel includes formal test results. We test for the existence of an SD relation using the statistical test by Linton, Song and Whang (2010; LSW10)[34]. The tests are expected to be powerful here, because they are based on a null hypothesis of second-order SD while an approximate third-order SD is expected here.

The null that the option-enhanced portfolio dominates the passive index is rejected with more than 90% confidence for the Lognormal and FHS priors. Not surprisingly, it is not rejected for the solution based on the ECDF prior, which features only a few open option positions. Encouragingly, no rejections occur for any of the three posteriors, reinforcing the economic significance of the forecast improvements from conditional information projection.

[Insert Table 5 about here.]

3.8 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 [35] 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 6 summarizes the results of the χ^2 GoF tests 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 6 about here.]

Table 7 summarizes the OOS investment performance of the option-enhanced strategies after the exclusions. The performance improvements are robust for the Lognormal and FHS priors. However, they are less robust for the ECDF prior, which is unsurprising given the low frequency of option position openings and low LWS10 p-value under this prior.

[Insert Table 7 about here.]

4 Conclusions

This study has introduced the CIPDE to estimate latent conditional distributions by integrating prior estimates with conditional moment conditions with functional nuisance parameters. CIPDE updates a prior density estimate through information projection onto the set of distributions that satisfy these moment conditions. Theoretical analysis demonstrates 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 compact supports and 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 nonparametric 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, [36] and the references therein). The consideration of such technologies in our framework seems like an interesting path for further research.

In applications to the pricing and trading of index options, conditional density estimates are essential due to the dynamic and non-Gaussian nature of index return distributions. 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. Robustness checks confirm the stability of these improvements across alternative subsamples that exclude extreme economic conditions and illiquid options.

While we have demonstrated the statistical and economic significance of these forecast improvements, assessing the general pricing efficiency of the index option market was not our objective. Our results are inconclusive on this issue, as we applied tight option filters based on moneyness, delta, and premium (see Section 3.1). These filters were used both to impose moment conditions and optimize option combinations. Evidence of mispricing and risk arbitrage opportunities may arise if the estimation filters used in the CIPDE are stricter than those used to select the range of options included in the optimization of investment decisions.

References

[1] Imre Csiszár. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3(1):146–158, 1975.

[2] Yuichi Kitamura and Michael Stutzer. An information-theoretic alternative to generalized method of moments estimation. *Econometrica*, 65(4):861–874, 1997.

[3] Guido Imbens, Richard Spady, and Phillip Johnson. Information theoretic approaches to inference in moment condition models. *Econometrica*, 66(2):333–357, 1998.

[4] Susanne M. Schennach. Bayesian exponentially tilted empirical likelihood. *Biometrika*, 92(1):31–46, 2005.

[5] Ivana Komunjer and Giuseppe Ragusa. Existence and characterization of conditional density projections. *Econometric Theory*, 32(4):947–987, 2016.

[6] George M. Constantinides, Jens Carsten Jackwerth, and Stylianos Perrakis. Mispricing of s&p 500 index options. *Review of Financial Studies*, 22(3):1247–1277, 2009.

[7] George M Constantinides, Michal Czerwonko, and Stylianos Perrakis. Mispriced index option portfolios. *Financial Management*, 49(2):297–330, 2020.

[8] Thierry Post and Iñaki Rodríguez Longarela. Risk arbitrage opportunities for stock index options. *Operations Research*, 69(1):100–113, 2021.

[9] Michael Stutzer. A simple nonparametric approach to derivative security valuation. *The Journal of Finance*, 51(5):1633–1652, 1996.

[10] Michael Stutzer. A bayesian approach to diagnosis of asset pricing models. *Journal of Econometrics*, 68(2):367–397, 1995.

[11] Thierry Post, Selçuk Karabatı, and Stelios Arvanitis. Portfolio optimization based on stochastic dominance and empirical likelihood. *Journal of Econometrics*, 206(1):167–186, 2018.

[12] Robert A. Adams and John J.F. Fournier. *Sobolev Spaces*, volume 140 of *Pure and Applied Mathematics*. Academic Press, 2nd edition, 2003.

[13] Keith Knight. Epi-convergence in distribution and stochastic equi-semicontinuity. *Unpublished manuscript*, 1999.

[14] Jean Feydy, Thibault Séjourné, François-Xavier Vialard, Shun-ichi Amari, Alain Trouvé, and Gabriel Peyré. Interpolating between optimal transport and mmd using sinkhorn divergences. In *The 22nd international conference on artificial intelligence and statistics*, pages 2681–2690. PMLR, 2019.

[15] Hasan Yilmaz. A generalization of multiplier rules for infinite-dimensional optimization problems. *Optimization*, 70(8):1825–1835, 2021.

[16] Xuefeng Li and Linda H Zhao. Bayesian nonparametric point estimation under a conjugate prior. *Statistics & Probability Letters*, 58(1):23–30, 2002.

[17] Francisco Blasques, Siem Jan Koopman, and André Lucas. Information-theoretic optimality of observation-driven time series models for likelihood inference. *Journal of Econometrics*, 202(1):112–131, 2018.

[18] Alexander B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, 2009.

[19] Richard Blundell, Joel L. Horowitz, and Matthias Parey. Nonparametric engel curves and revealed preference. *Econometrica*, 80(1):197–229, 2012.

[20] Jean Cortissoz. On the skorokhod representation theorem. *Proceedings of the American Mathematical Society*, 135(10):3995–4007, 2007.

[21] Aad W. Van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

[22] R. Tyrrell Rockafellar and Roger J-B Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften*. Springer, Berlin, Heidelberg, 1998.

[23] Frank Nielsen. An elementary introduction to information geometry. *Entropy*, 22(10), 2020.

[24] Zheng Fang and Andres Santos. Inference on directionally differentiable functions. *Review of Economic Studies*, 86(1):377–412, 2019.

[25] Imre Csiszár and János Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[26] Ke Gui and Yanjun Huang. Inverse pinsker inequality: A refined inequality for total variation and kullback–leibler divergence. *IEEE Transactions on Information Theory*, 68(10):6583–6592, 2022.

[27] Stylianos Perrakis. Stochastic dominance option pricing. *Springer Books*, 2019.

[28] Giovanni Barone-Adesi, Nicola Fusari, Antonietta Mira, and Carlo Sala. Option market trading activity and the estimation of the pricing kernel: A bayesian approach. *Journal of Econometrics*, 216(2):430–449, 2020.

[29] Majid Mojirsheibani. A note on the strong approximation of the smoothed empirical process of α-mixing sequences. *Statistical inference for stochastic processes*, 9:97–107, 2006.

[30] M Sudhakara Rao and YS Rama Krishnaiah. Weak convergence of empirical processes of strong mixing sequences under estimation of parameters. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 26–43, 1988.

[31] Miles S Kimball. Precautionary saving in the small and in the large.

[32] Thierry Post. Empirical tests for stochastic dominance efficiency. *The Journal of Finance*, 58(5):1905–1931, 2003.

[33] Thierry Post and Valerio Potì. Portfolio analysis using stochastic dominance, relative entropy, and empirical likelihood. *Management Science*, 63(1):153–165, 2017.

[34] Oliver Linton, Kyungchul Song, and Yoon-Jae Whang. An improved bootstrap test of stochastic dominance. *Journal of Econometrics*, 154(2):186–202, 2010.

[35] Kyle Jurado, Sydney C Ludvigson, and Serena Ng. Measuring uncertainty. *American Economic Review*, 105(3):1177–1216, 2015.

[36] Martin J Wainwright. Structured regularizers for high-dimensional problems: Statistical and computational issues. *Annual Review of Statistics and Its Application*, 1(1):233–253, 2014.

	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. 1: 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.

		Lognormal			ECDF			FHS		
		p25	p50	p75	p25	p50	p75	p25	p50	p75
	Skewness	0.00	0.00	0.00	-0.62	-0.59	-0.51	-1.07	-0.91	-0.80
	Kurtosis	3.00	3.00	3.00	5.40	5.68	5.96	5.32	6.03	6.96
KL	Lognormal	0.00	0.00	0.00	0.02	0.02	0.03	0.03	0.04	0.09
	ECDF	0.04	0.04	0.05	0.00	0.00	0.00	0.02	0.03	0.05
	FHS	0.04	0.06	0.09	0.02	0.03	0.05	0.00	0.00	0.00

Tab. 2: Forecast skewness, kurtosis and pairwise KL divergence between the three prior density estimates (reported as `KL(row, col)`). Quartile values are based on 321 forecast dates with valid option data (Jan. 1996–Jan. 2023).

		Lognormal		ECDF		FHS	
		Prior	Postr	Prior	Postr	Prior	Postr
	p25	-1.29	-1.19	-1.25	-1.20	-1.25	-1.33
D1	p50	-1.28	-1.12	-1.22	-1.17	-1.23	-1.24
	p75	-1.27	-1.06	-1.19	-1.12	-1.22	-1.15
	p25	-0.86	-0.72	-0.77	-0.72	-0.75	-0.77
D2	p50	-0.84	-0.65	-0.75	-0.66	-0.72	-0.70
	p75	-0.83	-0.60	-0.72	-0.61	-0.70	-0.62
	p25	-0.54	-0.43	-0.44	-0.40	-0.42	-0.42
D3	p50	-0.53	-0.35	-0.42	-0.35	-0.40	-0.38
	p75	-0.51	-0.30	-0.40	-0.29	-0.38	-0.30
	p25	-0.27	-0.17	-0.17	-0.15	-0.16	-0.16
D4	p50	-0.25	-0.11	-0.16	-0.11	-0.14	-0.12
	p75	-0.24	-0.05	-0.14	-0.05	-0.12	-0.07
	p25	-0.01	0.04	0.06	0.08	0.07	0.07
D5	p50	0.01	0.11	0.08	0.11	0.09	0.10
	p75	0.02	0.17	0.09	0.16	0.11	0.15
	p25	0.24	0.28	0.28	0.29	0.30	0.29
D6	p50	0.25	0.33	0.30	0.32	0.31	0.32
	p75	0.26	0.37	0.32	0.35	0.33	0.36
	p25	0.51	0.52	0.51	0.50	0.52	0.48
D7	p50	0.52	0.54	0.53	0.52	0.54	0.54
	p75	0.54	0.57	0.54	0.55	0.55	0.58
	p25	0.83	0.75	0.75	0.69	0.78	0.69
D8	p50	0.84	0.78	0.77	0.73	0.80	0.77
	p75	0.85	0.81	0.79	0.76	0.81	0.83
	p25	1.27	1.02	1.07	0.96	1.13	0.94
D9	p50	1.28	1.08	1.10	1.00	1.15	1.08
	p75	1.29	1.15	1.12	1.05	1.17	1.20
	p25		0.016		0.005		0.006
KL	p50		0.039		0.014		0.019
	p75		0.064		0.030		0.045

Tab. 3: Decile Boundaries for Standardized Returns. Boundaries of the deciles for each of the prior and posterior forecasts are standardized by the prior mean and standard deviation. Kullback Leibler Divergence percentiles illustrate the magnitude of the adjustments to the prior.

	Lognormal		ECDF		FHS	
	Prior	Postr	Prior	Postr	Prior	Postr
Hit Rates (%)						
D1	8.10	11.21	8.72	9.35	10.28	9.97
D2	6.23	6.85	7.48	8.10	6.54	7.79
D3	6.85	10.28	8.41	11.53	8.72	10.90
D4	9.66	6.23	7.48	6.85	7.79	8.10
D5	8.10	12.15	10.59	11.21	10.59	9.66
D6	13.08	10.28	10.90	10.59	8.72	9.35
D7	13.40	13.08	13.40	10.59	14.33	10.59
D8	18.07	12.15	13.71	12.77	12.15	13.08
D9	12.77	10.90	12.15	10.28	13.40	11.53
D10	3.74	6.85	7.17	8.72	7.48	9.03
χ^2 GoF Tests						
Left Tail	1.46	0.49	0.63	0.16	0.03	0.00
Right Tail	24.31***	4.40**	3.48*	0.63	2.69	0.35
All	57.38***	20.16**	17.86**	9.55	19.67**	7.26
M_{Pass}	0	278	8	310	12	313
M_{Inf}	321	43	313	11	309	8

Tab. 4: Forecasting ability: Hit Rate & Goodness of Fit Results. Hit rates show the percentage of realized returns that fall into specific deciles of the probability distribution for each forecast distribution (the ‘true’ percentage being 10 per decile). The χ^2 GoF tests are are reported for four different scenarios: Left Tail, with the emphasis on D1 (χ^2_1); Right Tail, with the emphasis on D10 (χ^2_1), and finally, All, covering all ten deciles (χ^2_9). Statistical significance of the scores is indicated in the standard convention using asterisks.

	Index	Lognormal Prior	Lognormal Postr	ECDF Prior	ECDF Postr	FHS Prior	FHS Postr
Prot. Puts		0.221	0.022	0.056	0.003	0.174	0.059
Avg Mnes		0.999	0.965	0.985	1.003	0.997	0.955
St. Dev. Mnes		0.005	0.091	0.038	0	0.017	0.076
Avg IV		0.210	0.249	0.203	0.155	0.200	0.311
St. Dev. IV		0.108	0.277	0.066	0	0.128	0.215
Cov. Calls		0.383	0.053	0.022	0.025	0.380	0.405
Avg Mnes		1.007	1.012	1.014	1.029	1.014	1.027
St. Dev. Mnes		0.008	0.030	0.014	0.026	0.015	0.022
Avg IV		0.174	0.146	0.304	0.304	0.201	0.206
St. Dev. IV		0.065	0.053	0.157	0.171	0.075	0.088
Ann. Mean (%)	7.88	6.50	7.88	7.58	7.85	7.04	7.99
Ann. St. Dev. (%)	17.58	14.47	17.40	17.04	17.35	14.21	15.30
Ann. CER1 (%)	6.24	5.39	6.27	6.03	6.25	5.95	6.73
Ann. CER2 (%)	4.52	4.21	4.58	4.39	4.56	4.79	5.39
Ann. CER4 (%)	0.58	1.45	0.69	0.60	0.67	2.05	2.26
LSW10 p-val		0.02	0.93	0.84	0.96	0.07	0.27

Tab. 5: 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 maximisation, 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 (delta range +/-0.05-0.5). Also reported is the LSW10 p-value for the null that the option-enhanced index dominates the passive index.

	Lognormal		ECDF		FHS	
	Prior	Postr	Prior	Postr	Prior	Postr
Excl. Recessions						
Left Tail	2.76*	0.12	1.43	1.43	0.57	0.95
Right Tail	20.81***	2.76*	2.03	0.12	1.43	0.02
All	58.05***	17.06**	20.45**	11.18	19.22**	6.64
Excl. High Fin. Unc.						
Left Tail	1.53	0.00	1.03	0.14	0.03	0.00
Right Tail	18.06***	2.15	1.53	0.03	1.03	0.00
All	50.64***	17.74**	15.84*	10.68	16.87*	8.28
Excl. High VIX						
Left Tail	0.64	1.28	0.14	0.00	0.35	0.16
Right Tail	21.21***	2.89*	2.15	0.14	1.53	0.03
All	44.86***	16.22*	9.65	6.81	10.11	4.98
Alt. Filters						
Left Tail	1.46	0.76	0.63	0.16	0.03	0.00
Right Tail	24.31***	1.46	3.48*	0.35	2.69	0.04
All	57.38***	12.86	17.86**	7.93	19.67**	8.79

Tab. 6: Goodness of Fit: Robustness Analysis. GoF 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 [0.15,0.5].

	Index	Lognormal		ECDF		FHS	
		Prior	Postr	Prior	Postr	Prior	Postr
Excl. Recessions							
Ann. Mean (%)	10.40	8.81	10.46	10.64	10.58	9.91	10.51
Ann. St. Dev. (%)	14.07	11.63	13.89	13.73	13.97	11.42	12.80
CER1 (%)	9.36	8.10	9.44	9.64	9.54	9.21	9.63
CER2 (%)	8.35	7.42	8.46	8.69	8.56	8.56	8.81
CER4 (%)	6.17	5.95	6.32	6.60	6.40	7.12	6.99
LSW10 p-val		0.02	0.87	0.87	0.96	0.06	0.27
Excl. High Fin. Unc.							
Ann. Mean (%)	9.02	7.83	9.07	9.06	9.08	8.69	9.26
Ann. St. Dev. (%)	14.72	12.22	14.55	14.40	14.63	12.22	13.55
CER1 (%)	7.88	7.05	7.96	7.97	7.96	7.90	8.29
CER2 (%)	6.76	6.28	6.86	6.90	6.85	7.13	7.34
CER4 (%)	4.33	4.62	4.49	4.57	4.45	5.44	5.28
LSW10 p-val		0.03	0.90	0.82	0.97	0.09	0.33
Excl. High VIX							
Ann. Mean (%)	4.16	3.79	4.21	4.53	4.44	4.34	4.63
Ann. St. Dev. (%)	15.99	13.91	15.83	15.72	15.95	13.91	14.93
CER1 (%)	2.80	2.75	2.87	3.21	3.08	3.30	3.43
CER2 (%)	1.32	1.61	1.42	1.78	1.61	2.15	2.13
CER4 (%)	-2.08	-1.09	-1.94	-1.54	-1.78	-0.57	-0.93
LSW10 p-val		0.06	0.91	0.93	0.98	0.15	0.37
Alt. Filters							
Ann. Mean (%)	7.88	6.50	7.74	7.58	8.00	7.04	8.60
Ann. St. Dev. (%)	17.58	14.47	17.47	17.04	17.32	14.21	15.09
CER1 (%)	6.24	5.39	6.12	6.03	6.41	5.95	7.36
CER2 (%)	4.52	4.21	4.41	4.39	4.72	4.79	6.07
CER4 (%)	0.58	1.45	0.49	0.60	0.84	2.05	3.01
LSW10 p-val		0.02	0.97	0.85	0.96	0.07	0.27

Tab. 7: 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 [0.15,0.5]. Also reported is the LSW10 p-value for the null that the option-enhanced index dominates the passive index.