Semiparametric Estimation of Probability Weighting Functions Implicit in Option Prices*

H. Peter Boswijk
Amsterdam School of Economics
University of Amsterdam
and Tinbergen Institute

Jeroen Dalderop†
Department of Economics
University of Notre Dame

Roger J. A. Laeven
Amsterdam School of Economics
University of Amsterdam, EURANDOM
and CentER

Niels Marijnen
Amsterdam School of Economics
University of Amsterdam
and Tinbergen Institute

This version: March 19, 2025

Abstract

This paper develops a semiparametric estimation method that jointly identifies the probability weighting and utility functions implicit in option prices. Our econometric method avoids direct specification of the objective conditional return distributions, which are instead obtained by transforming the options' implied risk-neutral distributions according to the posited rank-dependent utility model. We nonparametrically estimate the probability weighting function using the kernel density of suitable utility-adjusted probability integral transforms. The parameters of the utility function are estimated by maximizing the resulting profile likelihood. We establish the asymptotic properties of our estimation procedure, and demonstrate its good finite sample performance in Monte Carlo simulations. Empirical results based on S&P 500 index option prices and returns over the period 1996–2023 reveal the relevance of probability weighting, in particular at the monthly horizon where the weighting function is inverse-S shaped, which is robust to various specifications of the utility function.

Keywords: Semiparametric inference; Probability weighting function; Profile likelihood; Kernel estimation; Options.

JEL Classification: Primary: C14; C58; Secondary: G13.

*We are very grateful to Yacine Aït-Sahalia, Carsten Chong, Glenn Harrison, Oliver Linton, Olivier Scaillet, as well as seminar and conference participants at the Financial Econometrics Conference at Cambridge University, the 2024 Netherlands Econometric Study Group, the 2024 QFFE conference at Aix-Marseille Université, the Bernoulli-IMS World Congress in Probability and Statistics at Ruhr University Bochum, the 2024 EEA-ESEM at Erasmus University Rotterdam, the 2024 NBER Time Series conference at the University of Pennsylvania, the 2024 ES Meeting at Palma de Mallorca, the Tinbergen Institute, and the University of Amsterdam for comments and suggestions. This research was funded in part by the Netherlands Organization for Scientific Research under grant NWO Vici 2020–2027 (Laeven). *E-mail addresses*: H.P.Boswijk@uva.nl, jdalderop@nd.edu, R.J.A.Laeven@uva.nl, and N.Marijnen@uva.nl.

†Corresponding author.

1 Introduction

Probability weighting is at the heart of the leading non-expected utility models for describing and understanding risky choices. By distinguishing attitudes toward wealth, as captured by the utility function, and attitudes toward probabilities, as represented by the probability weighting function, the decision-theoretic models developed by Quiggin (1982) and Tversky and Kahneman (1992) can explain a wide variety of traits in economics and finance. Key examples include consumption-savings decisions, portfolio choice, gambling and insurance (Eeckhoudt et al., 2005; Barberis, 2013).

Existing evidence of probability weighting comes primarily from experimental studies, in which subjects choose from, predominantly fictitious, sets of lotteries designed and controlled by the researcher. Meanwhile, the scarcer real-world evidence of probability weighting is often based on market prices of contingent claims, such as betting odds (Jullien and Salanié, 2000) or financial option prices (Kliger and Levy, 2009; Polkovnichenko and Zhao, 2013). Option markets, in particular, provide detailed information on investors' attitudes toward risk by trading large numbers of contracts for a wide range of payoff thresholds. However, unlike in the laboratory, their objective, or physical, payoff probabilities at any point in time are not known by the researcher. Market-based measures of probability weighting therefore depend on the correct specification of conditional distributions. Their estimation based on historical returns requires imposing a distributional form and/or listing the set of conditioning variables. The latter is particularly challenging, as forward-looking investors may gather relevant information beyond the historical data available to the econometrician.[1]

This paper develops an estimation procedure that does not require direct specification of objective conditional distributions. Instead, we obtain physical distributions from option-implied risk-neutral distributions by transformations that follow from the asserted rank-dependent utility model. Our estimator only requires extracting cumulative risk-neutral distributions, which can be estimated more accurately than risk-neutral densities due to the curse of differentiation. The marginal utility and probability weighting functions are then estimated by maximizing the conditional likelihood of returns. Since economic theory puts few restrictions on the shape of the probability weighting function, we consider its nonparametric estimation. We do so based on the probability integral transforms (PITs) of marginal-utility adjusted risk-neutral distributions. Since these PITs are observed for a given utility function, their density can be directly estimated using nonparametric methods, for which we use a kernel density estimator. By considering parametric models for the utility function, we obtain a tractable semiparametric profile likelihood estimator that only requires low-dimensional optimization.[2]

Econometric theory to disentangle attitudes toward wealth and probabilities using option prices is limited. This paper fills this gap by formally establishing the identification and asymp-

[1] In models without probability weighting, missing out on conditioning variables has been attributed to cause the so-called 'pricing kernel puzzle' (Chabi-Yo et al., 2008; Song and Xiu, 2016; Linn et al., 2018).

[2] As we are primarily interested in the shape of the probability weighting function, we treat the probability weighting component nonparametrically. Our empirical analysis studies its sensitivity to the parametric form of the utility function.

totic properties of our profile likelihood estimator. First, we show that the utility function parameters are separately identified from the probability weighting function by considering the utility-adjusted PITs. At the true parameter vector, the latter ought to be independent of conditioning information, which includes the risk-neutral densities themselves. However, for any other parameter value, the conditional quantiles of the PITs are time-varying under mild regularity conditions. Since the weighting component for a given probability is fixed, this time variation can be attributed to an incorrect utility parameter vector, which strictly lowers the profile likelihood criterion. The probability weighting function is then identified from the distribution of the PITs evaluated at the identified utility parameters.

Second, we provide conditions under which the nonparametric estimation of the probability weighting function does not affect the asymptotic distribution of the utility parameter estimator. We verify these conditions for the kernel estimator of the probability weighting density based on the utility-adjusted PITs. A key challenge is that the densities of popular probability weighting functions diverge for probabilities near zero and one, which jeopardizes the required stochastic equicontinuity condition. We ensure the latter using a trimming rule that can be implemented prior to estimation, and by using a censored variant of the likelihood criterion.

Monte Carlo simulations demonstrate the good finite sample performance of our semiparametric estimator in realistic settings. When inverse-S shaped probability weighting of the form proposed by Tversky and Kahneman (1992) is present, our profile likelihood estimator of the risk aversion parameter of the utility function yields more than five times lower mean squared error (MSE) than the (then incorrectly specified) expected utility-based maximum likelihood estimator. Meanwhile, in the absence of probability weighting, the profile likelihood estimator typically has less than double the MSE of the (then correctly specified) maximum likelihood estimator. Furthermore, a nonparametric bootstrap procedure corresponding to our semiparametric estimator yields confidence intervals with accurate coverage rates. Finally, we propose a test for probability weighting based on the test for correct distributional specification by Bai (2003), and show that it has good size and power properties.

An empirical analysis of a panel of S&P 500 index option prices and returns spanning the period 1996–2023 reveals the importance of probability weighting, especially at the monthly horizon, where the Bai (2003) test rejects the expected utility model. At the monthly frequency, the probability weighting function takes an inverse S-shape, with larger concave (i.e., locally risk averse) than convex (i.e., locally risk seeking) regions. In a shorter sample at the weekly frequency, we find a globally concave probability weighting function. Furthermore, allowing for probability weighting leads to lower estimates of the constant relative risk aversion parameter than when probability weighting is ignored. The corresponding bootstrap confidence intervals suggest that linear utility cannot be rejected. Moreover, the shape of the probability weighting function is robust to the specification of the utility function. Using a flexible exponential-polynomial model, the estimated pricing kernel without probability weighting is U-shaped at the monthly horizon and non-decreasing in the right tail at the weekly horizon. However, when probability weighting is allowed, the estimated marginal utility functions decrease monotonically.

1.1 Related literature

A large number of experimental studies report evidence of probability weighting. Gonzalez and Wu (1999), Abdellaoui (2000), and Bruhin et al. (2010), to name a few, find inverse S-shaped probability weighting functions that display diminishing sensitivity when probabilities move away from zero and one. Other experimental papers find a globally concave weighting function, consistent with probability-driven risk aversion (Van de Kuilen and Wakker, 2011; Qiu and Steiger, 2011).

Our paper contributes to a growing literature on measuring probability weighting in real-world settings. Several studies have used individual-level field data, such as insurance choices (Cicchetti and Dubin, 1994; Barseghyan et al., 2013), to estimate rank-dependent utility models using micro-econometric methods. Like Barseghyan et al. (2013), we avoid parametric restrictions on the probability weighting function. However, our method only requires observing the market prices of risky payoffs, rather than individual holdings. The reliance on market prices is also common in studies that use betting odds (Jullien and Salanié, 2000; Snowberg and Wolfers, 2010) and transform them into predicted payoff probabilities by modeling risk preferences. However, betting odds are for discrete payoffs, whereas we consider financial assets with continuous payoffs.

Option prices are particularly informative about risk preferences, as their variation across strike prices reveals entire risk-neutral probability densities for the underlying return. Upon dividing these densities by estimated objective densities, Rosenberg and Engle (2002) and Aït-Sahalia and Lo (2000) estimate the pricing kernel, or the marginal utility as a function of the aggregate wealth return. Other pricing kernel estimators avoid modeling the objective distribution by using maximum likelihood (Bliss and Panigirtzoglou, 2004; Liu et al., 2007; Schreindorfer and Sichert, 2023), conditional density integration (Linn et al., 2018), or inverse density weighting (Dalderop and Linton, 2024). However, none of the above studies allow for a probability weighting component in the pricing kernel. Conversely, studies that measure probability weighting using option prices, such as Kliger and Levy (2009), Polkovnichenko and Zhao (2013), Dierkes (2013), and Chen et al. (2024), all rely on 'plug-in' methods for the objective return distributions. An exception is the GMM estimator of Baele et al. (2019), who specify a fully parametric model. To our best knowledge, our econometric method yields the first nonparametric estimator of the probability weighting function that is robust against misspecification of the conditional objective distributions.

The asymptotic analysis of our profile likelihood estimator builds on the general semiparametric estimation theory developed by Andrews (1994) and Newey (1994). Their key stochastic equicontinuity property has been established for other kernel-based profile likelihood estimators, such as for the single-index model in Ai (1997), the transformation model of Linton et al. (2008), and the independent component analysis in Hafner et al. (2024). Relative to these studies, our theoretical results are obtained under primitive conditions on structural objects, namely utility and probability weighting functions, that can be verified for specific economic models. Moreover, our theory allows for general time-series dependence, subject to stationarity and mixing conditions.

1.2 Outline

This paper is organized as follows. Section 2 introduces the model framework. Section 3 develops our semiparametric estimation theory. Section 4 describes the Monte Carlo simulations. Empirical results are in Section 5. Conclusions are in Section 6. Appendices provide the proofs, uniform convergence of the employed kernel estimator, and additional simulation results.

2 Model Framework

This section introduces the rank-dependent utility model and studies its asset pricing implications.

2.1 Rank-dependent utility and the risk-neutral density

Probability weighting is a pivotal ingredient of the rank-dependent utility (RDU) model of Quiggin (1982). The RDU model encompasses the expected utility (EU) model and the dual theory of Yaari (1987) as special cases and is the main building block in (cumulative) prospect theory of Tversky and Kahneman (1992). These extensions of EU have been developed to address its descriptive failures for decision under risk, notably Allais (1953)-type behavior.[3]

Consider the one-period stochastic (gross) return on wealth, R_1, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Denote by $F(r) := \mathbb{P}(R_1 \leq r)$ its cumulative distribution function and by $f(r) := F'(r)$ its probability density, assumed to exist. Consider a representative investor with utility function $u(\cdot;\gamma)$ for some parameter vector γ, henceforth assumed to be strictly increasing and twice differentiable, with initial wealth $w_0 > 0$ and next-period wealth $W_1 = R_1 w_0$. Under RDU, the investor maximizes

$$V := \int_0^\infty u\left(r w_0; \gamma\right) \mathrm{d}Z(F(r)), \tag{2.1}$$

with $Z : [0,1] \to [0,1]$ a non-decreasing function satisfying $Z(0) = 0$ and $Z(1) = 1$, referred to as the probability weighting function, and henceforth assumed to be strictly increasing and differentiable.[4] Also define the decumulative probability weighting function $\overline{Z}(P) := 1 - Z(1-P)$ for $0 \leq P \leq 1$, which is itself a probability weighting function with density $\overline{Z}'(P) = Z'(1-P)$.[5]

Any nonlinearity in Z implies that outcomes are no longer weighted linearly in probabilities. Effectively, this transforms objective probabilities into subjective decision weights, which helps describe empirical and experimental behavior at odds with the EU model. Experimental studies into the shape of the probability weighting function indeed find evidence of non-linear weighting, often with an inverse S-shaped function that is concave in the lower part of the domain and

[3] Based on experimental evidence, Harrison and Swarthout (2023) argue that RDU emerges as the most promising non-EU model for descriptive purposes.

[4] Non-decreasing utility and probability weighting functions ensure that RDU is compatible with first-order stochastic dominance (Quiggin, 1982; Yaari, 1987).

[5] In the microeconomic literature following Yaari (1987), the convention is often to distort decumulative probabilities instead of cumulative probabilities. Clearly, convexity of $\overline{Z}$ is equivalent to concavity of Z, and $\overline{Z}$ is inverse S-shaped if and only if Z is.

convex in the upper part of the domain (Gonzalez and Wu, 1999; Abdellaoui, 2000; Bruhin et al., 2010)[6] or a globally concave function corresponding to probability-driven risk aversion (Van de Kuilen and Wakker (2011); Bruggen et al. (2024)). Prelec (1998) has axiomatized a popular functional form of the probability weighting function that is capable of rationalizing inverse S-shaped probability attitudes.

Suppose the investor allocates a fraction α_0 of initial wealth to a risk-free asset with return R^0, and a fraction α_i to each of n risky assets with returns R^i, for $i = 1, \ldots, n$. Hence, under full allocation, $R_1 = \sum_{i=0}^{n} \alpha_i R^i$ with $\alpha_0 = 1 - \sum_{i=1}^{n} \alpha_i$. We assume that the R^i have absolutely continuous distributions. Provided both the utility and probability weighting functions are differentiable, the first-order conditions (FOCs) for optimal allocation are (e.g., Ai, 2005)

$$\frac{\partial}{\partial \alpha_i} V = \mathbb{E}\left(u'\left(W_1;\gamma\right) Z'(F(R_1))(R^i - R^0)\right) = 0, \qquad i = 1, \ldots, n. \tag{2.2}$$

The FOCs can be represented as $\mathbb{E}\left(m(R^i - R^0)\right) = 0$ in terms of the pricing kernel, or stochastic discount factor, $m := u'(W_1;\gamma)Z'(F(R_1))$. The pricing kernel is positive and induces a linear pricing rule, and hence is arbitrage-free. It defines an equivalent risk-neutral probability measure $\mathbb{Q}$ with probability density q given by

$$q(r) = c^{-1} u'(w;\gamma) Z'(F(r)) f(r), \tag{2.3}$$

where $c = \mathbb{E}\left(u'(W_1;\gamma)Z'(F(R_1))\right)$. For example, CRRA utility, i.e., $u(w;\gamma) = \frac{w^{1-\gamma}}{1-\gamma}$ for $\gamma \neq 1$, implies $q(r) \propto r^{-\gamma} Z'(F(r)) f(r)$.

2.2 From risk-neutral to physical distributions and densities

The physical distribution has a closed-form expression in terms of the risk-neutral density. Specifically, re-arranging (2.3) yields

$$Z'(F(r)) f(r) = c \frac{q(r)}{u'(w;\gamma)}. \tag{2.4}$$

This relation establishes the following string of identities:

$$Z(F(r)) = \int_0^{F(r)} Z'(F)\,\mathrm{d}F = \int_0^r Z'(F(s)) f(s)\,\mathrm{d}s = c \int_0^r \frac{q(s)}{u'(sw_0;\gamma)}\,\mathrm{d}s =: U(r;q,\gamma). \tag{2.5}$$

The function $U(r;q,\gamma)$ is itself a (*utility-adjusted*) distribution function, so that the normalization constant equals $c = \left(\int_0^\infty \frac{q(r)}{u'(rw_0;\gamma)}\,\mathrm{d}r\right)^{-1}$. Inverting (2.5) represents the physical distribution function as

$$F(r) = Z^{-1}(U(r;q,\gamma)). \tag{2.6}$$

[6] The inverse S-shape does not conform to second-order stochastic dominance: the RDU maximizer is (globally strongly) risk averse if and only if the utility and probability weighting functions are concave (Chew et al., 1987; Roëll, 1987; Eeckhoudt and Laeven, 2022).

Differentiating with respect to r then yields the physical density:

$$f(r) = c\frac{q(r)}{u'(rw_0;\gamma)}Z^{-1\prime}(U(r;q,\gamma)). \tag{2.7}$$

3 Estimation Theory

Observing financial markets over time results in a sample $\{R_{t+1}, q_t\}_{t=1}^T$ of stock return realizations and risk-neutral densities from option prices. Using a dynamic version of the model-based physical density of R_{t+1} given q_t derived in Section 2, this section develops a profile likelihood estimator to identify the utility parameters and the probability weighting function. First, we propose a nonparametric estimator of the probability weighting density based on the PITs of the utility-adjusted risk-neutral distributions (Sections 3.1–3.3). Next, we construct the profile likelihood function by substituting the density estimator for each utility parameter value into the semiparametric likelihood (Section 3.4). We then establish the joint identification of the utility and probability weighting functions (Sections 3.5–3.6), and derive relevant asymptotic properties (Sections 3.7–3.8). Throughout we allow for general time-series dependence in the observations, subject to stationarity. Before each result, we specify the required assumptions. All proofs are in Appendix A. Useful convergence results for the kernel estimator are in Appendix B.

3.1 Dynamic conditional density specification

The static version of the RDU model (2.1) naturally extends to a dynamic setting, in which investors re-balance their positions each period based on currently available information. Let $F_t(r) := \mathbb{P}(R_{t+1} \leq r|\mathcal{F}_t)$ and $f_t(r)$ be the conditional cumulative distribution function (CDF) and probability density function (PDF), respectively, of the return on wealth given the natural filtration $\{\mathcal{F}_t\}$. Similarly, denote the risk-neutral conditional CDF by Q_t and the corresponding PDF by q_t. Moreover, we re-define the utility u for the dynamic investor to be a function of the return on wealth, to avoid relying on the non-stationary wealth level. This holds automatically under CRRA utility, and can more generally be motivated by the multiplicative habit formation model of Abel (1990), with habit equal to current wealth. Additionally, it is consistent with the general pricing kernel specifications in Rosenberg and Engle (2002).

For any value of γ, define the utility-adjusted conditional risk-neutral CDF as follows:

$$U_t(r;\gamma) := c_t(\gamma)\int_0^r \frac{q_t(s)}{u'(s;\gamma)}\,\mathrm{d}s, \tag{3.1}$$

where $c_t(\gamma) = \left(\int_0^\infty \frac{q_t(r)}{u'(r;\gamma)}\,\mathrm{d}r\right)^{-1}$. The conditional physical CDF and PDF of the return are then modeled as counterparts of (2.6) and (2.7) by

$$F_t(r;\gamma,Z) = Z^{-1}(U_t(r;\gamma)), \tag{3.2}$$

$$f_t(r;\gamma,Z) = c_t(\gamma)\frac{q_t(r)}{u'(r;\gamma)}Z^{-1\prime}(U_t(r;\gamma)). \tag{3.3}$$

3.2 Estimating the probability weighting function via PITs

Define the utility-adjusted probability integral transform (PIT) as a function of γ

$$U_{t+1}(\gamma) := U_t(R_{t+1};\gamma) \in (0,1). \tag{3.4}$$

These PITs play a key role in measuring probability weighting. At γ_0, the true parameter value, $U_{t+1}(\gamma_0) = Z(F_t(R_{t+1}))$ by (3.2). Since $F_t(R_{t+1})$ are i.i.d. standard uniform, $U_{t+1}(\gamma_0)$ are i.i.d. with distribution function equal to the inverse probability weighting function Z^{-1}.[7]

For given γ, define the CDF of $U_{t+1}(\gamma)$ as $G(v;\gamma) := \mathbb{P}(U_{t+1}(\gamma) \leq v)$, or G_γ for short. At the true parameter value, $G(v;\gamma_0) = Z^{-1}(v)$. A natural estimator for $G_\gamma(v)$ is the empirical CDF

$$\widehat{G}(v;\gamma) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}(U_{t+1}(\gamma) \leq v). \tag{3.5}$$

Define the probability density $g(v;\gamma) := \frac{\partial}{\partial v}G(v;\gamma)$ and its shorthand g_γ. A similarly natural estimator for $g_\gamma(v)$ is the kernel estimator

$$\widehat{g}(v;\gamma) = \frac{1}{T}\sum_{t=1}^{T} K_h(U_{t+1}(\gamma) - v), \tag{3.6}$$

where $K_h(\cdot) = \frac{1}{h}K\left(\frac{\cdot}{h}\right)$ for some kernel K and bandwidth h. Since this estimator is generally consistent for any γ when $T \to \infty$ and $h \to 0$, the remaining challenge is estimating γ itself.[8]

3.3 Computing the PITs

The utility-adjusted PITs can be computed by applying integration by parts to (3.1)

$$\int_0^{R_{t+1}} \frac{q_t(r)}{u'(r;\gamma)}\,\mathrm{d}r = \frac{Q_t(R_{t+1})}{u'(R_{t+1};\gamma)} + \int_0^{R_{t+1}} Q_t(r)\frac{u''(r;\gamma)}{u'(r;\gamma)^2}\,\mathrm{d}r, \tag{3.7}$$

provided $\lim_{r\to 0}\frac{Q_t(r)}{u'(r;\gamma)} = 0$. The latter expression relies on the risk-neutral CDF, which can be estimated more efficiently from cross-sections of option prices than its density, as also exploited by the insightful approach of Polkovnichenko and Zhao (2013).

The normalization constants $c_t(\gamma)$ cannot be directly computed using the integration-by-parts formula (3.7) in the realistic case that $\lim_{r\to\infty}\frac{1}{u'(r;\gamma)} = \infty$. However, we can split the domain into two parts for some threshold κ:

$$\int_0^{\infty} \frac{q_t(r)}{u'(r;\gamma)}\,\mathrm{d}r = \int_0^{\kappa} \frac{q_t(r)}{u'(r;\gamma)}\,\mathrm{d}r + \int_\kappa^{\infty} \frac{q_t(r)}{u'(r;\gamma)}\,\mathrm{d}r. \tag{3.8}$$

[7] Henceforth, we often use the abbreviation "PITs" to refer to *utility-adjusted* probability integral transforms, recognizing that these are not standard uniform in the presence of probability weighting.

[8] In Appendix B, we establish the consistency of $\widehat{g}_\gamma$, uniformly over γ and v, allowing for general time-series dependence.

The first term can be computed using (3.7), and, provided $\lim_{r\to\infty} \frac{1-Q_t(r)}{u'(r;\gamma)} = 0$, the second as

$$\int_\kappa^\infty \frac{q_t(r)}{u'(r;\gamma)}\,\mathrm{d}r = \frac{1-Q_t(\kappa)}{u'(\kappa;\gamma)} - \int_\kappa^\infty (1-Q_t(r))\,\frac{u''(r;\gamma)}{(u'(r;\gamma))^2}\,\mathrm{d}r. \tag{3.9}$$

3.4 Profile likelihood estimator

Given the conditional density (3.3), the conditional log likelihood of the sample $\{R_{t+1} \mid q_t\}_{t=1}^T$ for any γ in some parameter space $\Theta \subseteq \mathbb{R}^k$ and density function g on $(0,1)$ equals

$$\begin{aligned}\ell_T(\gamma, g) &:= \frac{1}{T}\sum_{t=1}^T \log f_t(R_{t+1};\gamma,g) - \log q_t(R_{t+1}) \\ &= \frac{1}{T}\sum_{t=1}^T \log c_t(\gamma) - \log u'(R_{t+1};\gamma) + \log g(U_{t+1}(\gamma)),\end{aligned}$$

subtracting the parameter-independent $\log q_t(R_{t+1})$ for convenience. Rather than optimizing this criterion over the infinite-dimensional joint parameter space, we propose the estimator $\hat\gamma := \operatorname{argmax}_{\gamma\in\Theta} \ell_T(\gamma)$, maximizing the profile likelihood (PL) defined by substituting in the kernel density estimator (3.6):[9]

$$\ell_T(\gamma) := \ell_T(\gamma, \widehat{g}_\gamma). \tag{3.10}$$

In a general semiparametric estimation framework, Newey (1994) shows that the asymptotic variance of $\hat\gamma$ does not depend on the estimation error in $\widehat{g}_\gamma$ provided its probability limit maximizes the expected log-likelihood $\ell(\gamma, g) := \mathbb{E}\ell_T(\gamma, g)$ for any γ. The following result confirms that the density function g_γ has the latter property.

Lemma E. *For any γ, g_γ uniquely maximizes $\ell(\gamma, g)$ in the space of positive probability density functions on $(0,1)$.*

As a result, the infeasible PL estimator based on g_γ attains the semiparametric efficiency bound. Section 3.8 establishes that our feasible PL estimator $\hat\gamma$ asymptotically behaves *as if* g_γ is known, confirming its asymptotic efficiency.

3.5 Identification of utility parameters

The identification of the parameter vector γ depends on whether the profiled log-likelihood population criterion $\ell(\gamma) := \ell(\gamma, g_\gamma)$ is uniquely maximized at the true parameter γ_0. Jensen's inequality implies that $\ell(\gamma) \le \ell(\gamma_0)$ for any γ, with equality holding if and only if

$$f_t(R_{t+1};\gamma,g_\gamma) = f_t(R_{t+1};\gamma_0,g_{\gamma_0}) \quad \text{a.s.} \tag{3.11}$$

[9] While we focus on the kernel estimator, our estimation theory allows for other consistent nonparametric estimators for g_γ.

Thus, identification requires that no parameter value $\gamma \neq \gamma_0$ always yields the true conditional density $f_t(r) = f_t(r; \gamma_0, g_{\gamma_0})$. Let $ARA(r;\gamma) := -\frac{u''(r;\gamma)}{u'(r;\gamma)}$ denote the absolute risk aversion function. The following assumption assures that (3.11) cannot hold for any $\gamma \neq \gamma_0$.

Assumption I.

(i) *$Z'(P)$ is positive on $(0,1)$, $u'(r;\gamma)$ is positive and differentiable in r on $\mathbb{R}_{++} \times \Theta$, and $c_t(\gamma)$ is finite a.s. for all $\gamma \in \Theta$;*

(ii) *For any $\gamma \in \Theta \setminus \{\gamma_0\}$, there exists some $\bar{r} > 0$ such that $ARA(\bar{r};\gamma) \neq ARA(\bar{r};\gamma_0)$ and $u''(r;\gamma)$ is continuous at $\bar{r}$;*

(iii) *$f_t : \mathbb{R}_{++} \to \mathbb{R}_{++}$ is continuous a.s., and for some pair $(v_1, v_2) \in \text{supp}(F_t(\bar{r}))$ the conditional quantiles $F_t^{-1}(v_1)$ and $F_t^{-1}(v_2)$ are not one-to-one.*

Lemma I. *Under Assumption I, $\ell(\gamma)$ is uniquely maximized at γ_0.*

The intuition behind Lemma I is as follows. Condition I*(i)* ensures that the pricing kernel is positive and finite under model (3.3), and hence satisfies the fundamental theorem of asset pricing. This ensures that G_γ is strictly increasing for all $\gamma \in \Theta$. Therefore, (3.11) becomes equivalent to $U_{t+1}(\gamma) = U_{t+1}(\gamma_0)$ *a.s.*, which implies that $U_{t+1}(\gamma)$ must be independent of any information in $\mathcal{F}_t$, and thus of the risk-neutral density q_t by (3.3). However, we prove that for some (v_1, v_2), the conditional quantiles of $U_{t+1}(\gamma)$ vary over time unless either $ARA(r;\gamma) \equiv ARA(r;\gamma_0)$ or there is a particular one-factor structure among the corresponding physical quantiles. The former case is directly ruled out by condition I*(ii)*. The latter case is ruled out by condition I*(iii)*, which ensures that the physical densities are positive, continuous, and non-deterministic in a way that is satisfied for dynamic models with at least two state variables. Thus, $U_{t+1}(\gamma)$ must depend on time-t information for any $\gamma \neq \gamma_0$, so that resulting differences in the conditional densities of R_{t+1} given q_t result in a strictly lower value of the profile likelihood.

3.6 Trimming

The presence of probability weighting results in terms $\frac{\partial}{\partial\gamma} \log g_\gamma \left(U_{t+1}(\gamma)\right)$ in the scores of the likelihood function. For several popular probability weighting functions (see Section 3.10), the density $Z'(P)$ diverges to infinity for probabilities near zero and one. As a result, the density g_γ converges to zero at the boundaries, causing a small denominator issue.[10] Furthermore, $\frac{\partial}{\partial\gamma} g_\gamma$ may diverge at the boundaries for common utility functions, which prohibits the uniform convergence of its kernel estimator.

To overcome these issues, we propose the following trimming rule that bounds $U_{t+1}(\gamma)$ away from zero and one, uniformly over Θ. Let $R_l^*(q,v) := \max_{\gamma\in\Theta} U^{-1}(v;q,\gamma)$ and $R_u^*(q,v) := \min_{\gamma\in\Theta} U^{-1}(v;q,\gamma)$, and for some small $v^* > 0$ let $R_{t,l}^* = R_l^*(q_t, v^*)$ and $R_{t,u}^* = R_u^*(q_t, 1-v^*)$. By design, $R_{t,l}^* \leq R_{t+1} \leq R_{t,u}^*$ ensures that $v^* \leq U_{t+1}(\gamma) \leq 1 - v^*$ for all $\gamma \in \Theta$. We assume that v^* is chosen small enough to guarantee that $R_{t,l}^* < R_{t,u}^*$, or else the time-t observation is

[10] A similar issue occurs in semiparametric copula-based models (Chen and Fan, 2006), due to asymptotes in the copula densities.

removed. Since the thresholds $(R^*_{t,l}, R^*_{t,u})$ do not depend on the parameter, they only need to be computed once prior to estimation.

Define the censored profile log-likelihood function $\ell^*_T(\gamma) := \ell^*_T(\gamma, \widehat{g}_\gamma)$, where

$$\ell^*_T(\gamma, g) := \frac{1}{T}\sum_{t=1}^{T}\left[1^l_{t+1}\log G(U^l_t(\gamma)) + 1^m_{t+1}\log\left(\frac{c_t(\gamma)g(U_{t+1}(\gamma))}{u'(R_{t+1};\gamma)}\right) + 1^u_{t+1}\log\left(1 - G(U^u_t(\gamma))\right)\right], \tag{3.12}$$

with $U^i_t(\gamma) := U_t(R^*_{t,i};\gamma)$ for $i \in \{l, u\}$ and $1^l_{t+1} := 1(R_{t+1} \leq R^*_{t,l})$, $1^m_{t+1} := 1(R^*_{t,l} < R_{t+1} \leq R^*_{t,u})$, and $1^u_{t+1} := 1(R^*_{t,u} < R_{t+1})$ indicators for the lower, middle, and upper parts of the distribution, respectively. This method does not exclude the trimmed returns, but reduces them to binary tail event variables when estimating γ. The density estimator $\widehat{g}_\gamma$ still uses all observations.

The following lemma confirms that the censored profile likelihood population criterion $\ell^*(\gamma) := \mathbb{E}\ell^*_T(\gamma, g_\gamma)$ still identifies γ_0, as long as the non-censored intervals are 'wide' enough.

Lemma I*. *Under Assumption I, with* $\mathbb{P}(R^*_{t,l} \leq \bar{r} \leq R^*_{t,u}) > 0$ *and* $(v_1, v_2) \in \text{supp}(F_t(\bar{r}) \mid R^*_{t,l} \leq \bar{r} \leq R^*_{t,u})$, $\ell^*(\gamma)$ *is uniquely maximized at* γ_0.

3.7 Consistency

Besides identification, consistency of $\hat{\gamma}$ requires the uniform convergence of $\ell^*_T(\gamma)$ to $\ell^*(\gamma)$ over the parameter space Θ. If the profile density g_γ were known, this would follow from the uniform convergence of $\ell^*_T(\gamma, g_\gamma)$ to $\ell^*(\gamma, g_\gamma)$. As we estimate g_γ nonparametrically, we additionally require the uniform convergence of $\widehat{g}_\gamma$. Let $\|g\|_{\infty,v^*} = \sup_{v^* \leq v \leq 1-v^*} |g(v)|$ be the sup-norm over the trimmed support. We then establish the consistency of the profile-likelihood estimator under the following assumptions.

Assumption C.

(i) $Z'(P)$ *and* $u'(r;\gamma)$ *are continuous on* $(0,1)$ *and* $\mathbb{R}_{++} \times \Theta$, *respectively, with* Θ *compact;*

(ii) $\mathbb{E}\left(\sup_{\gamma\in\Theta} |\log u'(R_{t+1};\gamma)|\right) < \infty$ *and* $\mathbb{E}\left(\sup_{\gamma\in\Theta} |\log c_t(\gamma)|\right) < \infty$;

(iii) $\sup_{\gamma\in\Theta} \|\widehat{g}_\gamma - g_\gamma\|_{\infty,v^*} \xrightarrow{p} 0$ *and* $\sup_{\gamma\in\Theta} |\widehat{G}_\gamma(v^*) - G_\gamma(v^*)| \xrightarrow{p} 0$;

(iv) $f_t(r) = f(r|X_t)$ *is a positive, measurable function of the stationary ergodic process* (R_{t+1}, X_t), *where* X_t *is some state vector.*

Proposition C. *Suppose that Assumptions I and C hold. Then* $\hat{\gamma} := \arg\max_{\gamma\in\Theta} \ell^*_T(\gamma) \xrightarrow{p} \gamma_0$ *when* $T \to \infty$.

Assumption C contains sufficient conditions for those in Newey (1994, Lemma 5.2). Continuity of the probability weighting density and marginal utility function in C*(i)* is required for the continuity of the criterion function. Crucially, $Z'(P)$ only needs to be continuous on the open interval $(0,1)$, as under inverse S-shaped probability weighting $Z'(P)$ has asymptotes at the boundaries, as shown in Section 3.10. The moment conditions in C*(ii)* ensure the criterion

function is dominated by an integrable function. The first moment is readily verified for common utility functions. For example, for power utility it amounts to the existence of $\mathbb{E}|\log R_{t+1}|$.

The uniform convergence conditions C*(iii)* allow for a general nonparametric estimator, not necessarily a kernel estimator. This condition can in turn be implied from primitive conditions on specific estimators. Lemma K-1 in Appendix B provides sufficient conditions for the kernel density estimator (3.6), based on results in Kristensen (2009). Similarly, Lemma K-1* provides sufficient conditions for the kernel CDF estimator

$$\widehat{G}_h(v;\gamma) = \frac{1}{T}\sum_{t=1}^{T} F_K\left(\frac{v - U_{t+1}(\gamma)}{h}\right) = \int_{-\infty}^{v} \hat{g}(u;\gamma)\,\mathrm{d}u, \tag{3.13}$$

with $F_K(x) = \int_{-\infty}^{x} K(z)\,\mathrm{d}z$, where the second equality presumes K is symmetric around zero.

Importantly, the uniform convergence of the density estimator is only required on the compact subset $[v^*, 1-v^*]$, as the kernel estimator may not be consistent on $[0,1]$ when g or g' have asymptotes near the boundaries. For the CDF estimator this is not problematic, as its summands are uniformly bounded. Finally, condition C*(iv)* is used to establish a pointwise LLN for the criterion function, while the compact parameter space is used to extend this to a uniform law.

3.8 Asymptotic normality

The score of the profile likelihood function for each non-trimmed observation is given by

$$s(R_{t+1}, q_t, \gamma, g) := \frac{\partial}{\partial\gamma}\log f_t(R_{t+1};\gamma, g_\gamma) = \frac{c_t'(\gamma)}{c_t(\gamma)} - \frac{\partial u'(R_{t+1};\gamma)/\partial\gamma}{u'(R_{t+1};\gamma)} + \frac{\frac{\mathrm{d}}{\mathrm{d}\gamma}g_\gamma(U_{t+1}(\gamma))}{g_\gamma\left(U_{t+1}(\gamma)\right)},$$

for any family of density functions $g(v;\gamma)$. The scores for left and right censored observations equal $\frac{\partial}{\partial\gamma}\log G_\gamma(U_t(R^*_{t,l};\gamma))$ and $\frac{\partial}{\partial\gamma}\log\left(1 - G_\gamma(U_t(R^*_{t,u};\gamma))\right)$, respectively. The first order condition (FOC) for maximizing (3.10) then equals

$$\frac{\partial}{\partial\gamma}\ell^*_T(\gamma) = \frac{1}{T}\sum_{t=1}^{T} s(R_{t+1}, q_t, \gamma, \widehat{g}) = 0,$$

where $\widehat{g}$ is given in (3.6).[11] Hereafter we denote the true probability density and cumulative distribution functions of $U_{t+1}(\gamma)$ as $g_{0,\gamma}$ and $G_{0,\gamma}$ when distinguishing them from arbitrary families of distributions. The results in this section establish that when $\widehat{g}$ converges uniformly to its population counterpart g_0, the nonparametric estimation error is asymptotically irrelevant for the parameter estimator $\hat{\gamma}$.

First, note that $g(v;\gamma)$ only enters the profile likelihood score at γ_0 via the functional

$$S(r,q,g) := \frac{\frac{\mathrm{d}}{\mathrm{d}\gamma}g_\gamma(U(r;q,\gamma))\big|_{\gamma=\gamma_0}}{g_{\gamma_0}(U_0(r,q))} = \frac{\dot{g}_{\gamma_0}(U_0(r,q)) + g'_{\gamma_0}(U_0(r,q))\dot{U}_0(r,q)}{g_{\gamma_0}(U_0(r,q))},$$

[11] Trimming requires adding $\widehat{G}$ as an argument to s, unless the CDF estimator equals the integrated PDF estimator. The latter holds for the kernel CDF estimator in (3.13).

where $\dot{f}_\gamma(v) := \frac{\partial}{\partial\gamma} f_\gamma(v)$ and $f'_\gamma(v) := \frac{\partial}{\partial v} f_\gamma(v)$ for any function f, and $U_0(r,q) := U(r;q,\gamma_0)$. The functional S only depends on the $(k+1)$-dimensional function $g(u;\gamma)$ through the univariate functions g_{γ_0}, g'_{γ_0}, and $\dot{g}_{\gamma_0}$. Therefore, define the following norm for functions $g(v;\gamma)$:

$$\|g\| := \max\{\|g_{\gamma_0}\|_{\infty,v^*}, \|g'_{\gamma_0}\|_{\infty,v^*}, \|\dot{g}_{\gamma_0}\|_{\infty,v^*}\}.$$

The following conditions, based on Newey (1994, Assumptions 5.1–5.3), ensure that the non-parametric estimation of g does not affect the limiting distribution of $\frac{\partial}{\partial\gamma}\ell^*_T(\gamma_0)$.

Assumption A.

(i) *$Z'(P)$, $u'(r;\gamma)$, and $c_t(\gamma)$ are positive and continuously differentiable on $(0,1)$, $\mathbb{R}_{++}$ at γ_0, and at γ_0 a.s., respectively, and $\mathbb{E}\|\dot{U}_0\| < \infty$;*

(ii) *$\|\widehat{g} - g_0\| = o_p(T^{-\frac{1}{4}})$, and $(\widehat{G}_{\gamma_0}, \widehat{\dot{G}}_{\gamma_0})(v^*) = (G_{\gamma_0}, \dot{G}_{\gamma_0})(v^*) + o_p(T^{-\frac{1}{4}})$;*

(iii) *$\frac{1}{\sqrt{T}}\sum_{t=1}^T D_0(R_{t+1}, q_t, \widehat{g} - g_0) \xrightarrow{p} 0$, where $D_0(r,q,\bar{g})$ is the pathwise derivative of S at g_0 in direction $\bar{g}$.*

Lemma A. *Under Assumptions A and C(iv), when $T \to \infty$,*

$$\sqrt{T}\frac{\partial}{\partial\gamma}\ell^*_T(\gamma_0) = \frac{1}{\sqrt{T}}\sum_{t=1}^T s(R_{t+1}, q_t, \gamma_0, g_0) + o_p(1).$$

Condition A*(i)* ensures that the score functional is asymptotically linear in the unknown density g at the true value. The differentiability of $u'(r;\gamma_0)$ and $c_t(\gamma)$ ensures that $\dot{U}_0$ exists. The trimming avoids the small denominator problem, as the continuous and positive density g_{γ_0} is bounded away from zero on $[v^*, 1-v^*]$.

A*(ii)* requires that the estimators $\widehat{g}_\gamma$, $\widehat{\dot{g}}_\gamma$ and $\widehat{g}'_\gamma$ converge uniformly at at least a $T^{-\frac{1}{4}}$ rate. For our suggested kernel estimator (3.6), Lemma K-2 provides sufficient conditions for this rate, based on results in Andrews (1995). In particular, it requires at least fourth order kernels to ensure that the kernel density derivative estimator converges fast enough. Similarly, K-2* provides conditions for the sufficiently fast convergence of the kernel CDF estimator (3.13) at the trimming point v^*.

For kernel estimators, the following lemma establishes the stochastic equicontinuity condition A*(iii)* using a result for dependent U-statistics.

Lemma SE. *Suppose A(i), C(iv), and the following conditions hold for some $\delta' > 0$ and $\delta > \delta'$:*

(i) *(R_{t+1}, X_t) is β-mixing with $\beta(\tau) = O\left(\tau^{-\frac{2+\delta'}{\delta'}}\right)$ when $\tau \to \infty$;*

(ii) *$\mathbb{E}(\|\dot{U}_0\|^{2+\delta} \mid U_0 = v)$ is continuous for $v \in (0,1)$, $\mathbb{E}\left(\|\dot{U}_{0,s} - \dot{U}_{0,t}\|^{2+\delta} \mid U_{0,s} = u, U_{0,t} = v\right)$ is continuously differentiable at $u = v \in (0,1)$ for all $s \neq t$;*

(iii) *$\int K(z)\,\mathrm{d}z = 1$, $\int K(z)z\,\mathrm{d}z = 0$, $K(0) < \infty$, $\int |K(z)|^{2+\delta}\,\mathrm{d}z + \int |zK'(z)|^{2+\delta}\,\mathrm{d}z < \infty$;*

(iv) $T^{\frac{\delta-\delta'}{\delta'}}h^{1+\delta} \to \infty$ *and* $Th^2 \to \infty$.

Then Assumption A(iii) holds for the kernel density (3.6).

Assumption SE*(i)* stipulates that the mixing coefficients vanish at least as fast as $1/\tau$ when the time between observations τ becomes large. A larger value of δ' allows for a slower rate within this class, though by SE*(ii)*-*(iii)* requires higher moments of the PIT-derivative $\dot{U}_0$ and kernel function K to exist. Meanwhile, Assumption SE*(iv)* ensures that the bandwidth does not vanish too quickly when the sample period increases.

The asymptotic distribution of $\hat{\gamma}$ follows from the mean-value expansion of the FOC around γ_0,

$$\sqrt{T}\left(\hat{\gamma}-\gamma_0\right) = \left(\frac{\partial^2}{\partial\gamma\partial\gamma^\top}\ell_T^*(\tilde{\gamma})\right)^{-1}\sqrt{T}\frac{\partial}{\partial\gamma}\ell_T^*(\gamma_0), \tag{3.14}$$

for some $\tilde{\gamma}$ in between $\hat{\gamma}$ and γ_0. The following additional assumptions therefore yield the $\sqrt{T}$-convergence rate and asymptotic normality of our profile ML estimator.

Assumption N.

(i) $\mathbb{E}\|\frac{\partial}{\partial\gamma}\log c_t(\gamma)\big|_{\gamma=\gamma_0}\|^2$ *and* $\mathbb{E}\|\frac{\partial}{\partial\gamma}\log u'(R_{t+1};\gamma)\big|_{\gamma=\gamma_0}\|^2$ *are finite;*

(ii) $Z'(P)$, $u'(r,\gamma)$, $c_t(\gamma)$, *and* $F_t(r)$ *are twice continuously differentiable a.s. on* $(0,1)$, $\mathbb{R}_+\times\mathcal{N}_0$, $\mathcal{N}_0$, *and* $\mathbb{R}_+$, *respectively, where* $\mathcal{N}_0$ *is a neighborhood of* γ_0, γ_0 *is interior to* Θ, *and* $M := \partial\mathbb{E}\left(s(R_{t+1},q_t,\gamma,g_0)\right)/\partial\gamma\big|_{\gamma=\gamma_0}$ *is non-singular;*

(iii) $\sup_{\gamma\in\mathcal{N}_0}\|\frac{\partial^{i+|j|}}{\partial v^i\partial\gamma^j}(\widehat{g}_\gamma - g_\gamma)\|_{\infty,v^*} \xrightarrow{p} 0$ *for any non-negative integer* i *and multi-index* j *with* $i+|j|\le 2$, $\sup_{\gamma\in\mathcal{N}_0}\left|\frac{\partial^{|j|}}{\partial\gamma^j}(\widehat{G}_\gamma(v^*) - G_\gamma(v^*))\right| \xrightarrow{p} 0$ *for* $|j|\le 2$;

(iv) $\mathbb{E}\left(\sup_{\gamma\in\mathcal{N}_0}\|W_t(\gamma)\|\right) < \infty$ *for* $W_t \in \{\frac{\partial^2}{\partial\gamma\partial\gamma^\top}\log c_t(\gamma), \frac{\partial^2}{\partial\gamma\partial\gamma^\top}\log u'(R_{t+1};\gamma), \dot{U}_{t+1}, \ddot{U}_{t+1}\}$.

Proposition N. *Under Assumptions I, C, A, and N, when* $T\to\infty$,

$$\sqrt{T}\left(\hat{\gamma}-\gamma_0\right) \xrightarrow{d} N(0, M^{-1}VM^{-1}), \tag{3.15}$$

where $V := \mathbb{V}\mathrm{ar}\left(s(R_{t+1},q_t,\gamma_0,g_0)\right)$.

Assumption N*(i)* is required for the covariance matrix V of the score of the observations to exist. Meanwhile, the conditions on continuity in N*(ii)*, uniform convergence in N*(iii)*, and integrability in N*(iv)*, ensure the convergence in probability of the Hessian term in (3.14) to a non-singular matrix uniformly in a neighborhood of γ_0. Lemmas K-3 and K-3* in Appendix B provide sufficient conditions for the uniform convergence of the derivatives in N*(iii)* for the kernel density and CDF estimators (3.6) and (3.13), respectively.

3.9 Bootstrap estimation of PL confidence interval

Inference requires estimators of the variance. However, estimation of V and M in (3.15) via plug-in methods necessitates nonparametric estimation of higher order derivatives, which is unstable in small samples. We propose instead to conduct inference via a simple bootstrap algorithm: we

draw, with replacement, T pairs $(\tilde{R}_{t+1}, \tilde{q}_t)$ from $\{R_{t+1}, q_t\}_{t=1}^T$, and repeat the estimation with this bootstrap sample to find some $\tilde{\gamma}$. Doing so K times, for $K \in \mathbb{N}$ large, results in a sample $\{\tilde{\gamma}_k - \hat{\gamma}\}_{k=1}^K$, of which the distribution approximates that of $\hat{\gamma} - \gamma_0$. These bootstrap estimates can be used to construct quantile- or standard deviation-based confidence intervals.

The algorithm described above is a standard (nonparametric) percentile bootstrap, see e.g., Efron and Tibshirani (1993), developed for independent and identically distributed data. The independence assumption is violated here for $\{R_{t+1}, q_t\}_{t=1}^T$. However, because the transformation $s(R_{t+1}, q_t, \gamma_0, g_0)$ is free of serial correlation, the dependence in the original data does not affect the limiting distribution in Proposition N. This implies that a dependent bootstrap method (such as the block bootstrap) is not needed here: the proposed bootstrap algorithm replicates the dependence structure in the original data insofar as it appears in the asymptotic distribution of $\hat{\gamma}$, and hence is valid. A similar result for autoregressive models was obtained by Gonçalves and Kilian (2004), who refer to this as the pairwise bootstrap.

3.10 Verification of conditions for common models

Our assumptions impose certain regularity and smoothness conditions on the probability weighting function, such as positivity of the weighting density by Assumption I*(i)*, and continuity and differentiability by Assumptions C*(i)*, A*(i)* and N*(ii)*. Due to our trimming procedure, these conditions only need to hold on the open interval $(0, 1)$, which is easily verified for popular parametrizations of the probability weighting function. In particular, consider the one-parameter model by Tversky and Kahneman (1992), and the two-parameter model by Prelec (1998), which are defined by their decumulative probability weighting functions

$$\overline{Z}_{TK}(P) = \frac{P^\delta}{(P^\delta + (1-P)^\delta)^{1/\delta}}, \qquad \overline{Z}_P(P) = \exp(-(-\beta \log(P))^\alpha),$$

respectively, for some positive scalars δ, α, and β. Both specifications allow for the typical inverse-S shape observed in experiments, yet reduce to linear probability weighting for $\delta \equiv 1$ and $(\alpha, \beta) \equiv (1, 1)$, respectively. For the probability weighting function (PWF) of Tversky and Kahneman (1992),

$$\overline{Z}'_{TK}(P) = \left[\frac{\delta - 1}{P^{1-\delta}} + \delta \frac{(1-P)^\delta}{P} + \frac{1}{(1-P)^{1-\delta}}\right] \frac{\overline{Z}_{TK}(P)}{P^\delta + (1-P)^\delta},$$

which diverges to infinity at the boundary points for $\delta < 1$ and to zero when $\delta > 1$. Meanwhile, for the PWF of Prelec (1998),

$$\overline{Z}'_P(P) = \frac{\alpha\beta}{P}(-\beta \log(P))^{\alpha-1} \overline{Z}_P(P).$$

When $\alpha < 1$, $\overline{Z}'_P(P)$ is $O\left(\frac{1}{P}\right)$ when $P \to 0$ and $O\left((\log(P))^{\alpha-1}\right)$ when $P \to 1$, and thus diverges to infinity at both boundary points. On the other hand, when $\alpha > 1$, $\overline{Z}'_P(P)$ vanishes to zero at the boundary points. The same is true for the associated cumulative probability weighting functions. As we trim at the boundary, our conditions allow for such divergence to infinity

and/or zero at the boundary points $P = 0$ and $P = 1$. In particular, A*(i)* merely requires the probability weighting density $Z'(P)$ to be positive and continuously differentiable on the open interval $(0, 1)$, which both models satisfy. Moreover, our conditions allow for linear probability weighting $Z_0(P) = P$, which corresponds to the standard expected utility model.

Since $g_{\gamma_0}(v) = \frac{1}{Z'(Z^{-1}(v))}$ by the inverse function theorem, the density of $U_{t+1}(\gamma_0)$ goes to zero when v goes to zero and one when either PWF specification takes the inverse S-shape. Our Assumptions C*(iii)*, A*(ii)*, and N*(iii)* on the uniform convergence of $\hat{g}$ allow for this as $g_\gamma(v)$ remains bounded away from zero on the trimmed domain $[v^*, 1 - v^*]$, uniformly over γ, due to the positivity and continuity of $g(v; \gamma)$ on $(0, 1) \times \Theta$ established in Lemma D in the Appendix.

Similarly, our trimming procedure allows for asymptotes in the derivatives of $g_\gamma(v)$ near boundary points, which arise under common models. In particular, differentiating expression (A.2) for $G_\gamma(v)$ w.r.t. γ yields

$$\dot{G}_{\gamma_0}(v) = -\mathbb{E}\left(v \frac{c_t'(\gamma_0)}{c_t(\gamma_0)} - \int_0^v \frac{\dot{u}'(U_t^{-1}(u; \gamma_0); \gamma_0)}{u'(U_t^{-1}(u; \gamma_0); \gamma_0)} \, \mathrm{d}u \right) g_{\gamma_0}(v).$$

Under power utility, $\frac{\dot{u}'(r;\gamma)}{u'(r;\gamma)} = -\log r$. Since $\mathbb{E} \log U_t^{-1}(v; \gamma_0)$ diverges when $v \to 0$ or $v \to 1$ for any strictly increasing PWF, $\dot{g}_{\gamma_0}(u) = \frac{\partial}{\partial v} \dot{G}_{\gamma_0}(v)$ diverges at the boundaries. Still, $\dot{g}_\gamma$ remains positive and continuous on the trimmed domain, which enables stochastic equicontinuity for our kernel estimator by Lemma SE and the uniform convergence of its derivatives in Appendix B.

4 Monte Carlo Simulations

In this section, we describe the results of a Monte Carlo simulation study that analyzes the finite sample properties of our estimation procedure. We consider three economic data generating processes (DGPs), which differ in the shape of the probability weighting function. Additional simulation results, under a different data frequency, a larger risk aversion parameter, and different $\mathbb{P}$-dynamics, can be found in Appendix C.

4.1 Set-up

We simulate $N = 1000$ replications of 25 years of monthly data ($T = 300$). We model the price process of a futures contract, F_t, under the physical measure $\mathbb{P}$ by the following dynamics:

$$\begin{aligned}
\mathrm{d} \log F_t &= \left(-\tfrac{1}{2} V_t - \mu_J \lambda_t\right) \mathrm{d}t + \sqrt{V_t} \, \mathrm{d}W_{1,t} + \sqrt{H_t} \, \mathrm{d}W_{2,t} + J_t \, \mathrm{d}N_t, \\
\mathrm{d}V_t &= 12(0.015 - V_t) \, \mathrm{d}t + 0.5 \sqrt{V_t} \left(-0.9 \, \mathrm{d}W_{1,t} + \sqrt{1 - 0.9^2} \, \mathrm{d}W_{3,t}\right) + J_t^V \mathbb{1}_{\{J_t < 0\}} \, \mathrm{d}N_t, \\
\mathrm{d}H_t &= (0.01 - H_t) \, \mathrm{d}t + 0.125 \sqrt{H_t} \left(-0.5 \, \mathrm{d}W_{2,t} + \sqrt{1 - 0.5^2} \, \mathrm{d}W_{4,t}\right),
\end{aligned} \tag{4.1}$$

where W_t is a four-dimensional standard Brownian motion, and N_t is a Poisson jump process with intensity $\lambda_t = 60V_t + 30H_t$. The presence of at least two volatility factors, which together drive the time-varying jump intensity under the measure $\mathbb{P}$, is consistent with findings in Andersen et al. (2015). The jumps J_t in the price process follow a double-exponential distribution: at

a jump event, there is a probability $p^- = 0.7$ of a negative exponential jump with a mean of $\eta^- = 0.05$, and a probability $1-p^-$ of a positive exponential jump with a mean of $\eta^+ = 0.02$. In case of a negative jump in the price process, volatility co-jumps: J_t^V is exponentially distributed with an average jump size of 0.01. The parameter μ_J ensures the jumps are compensated, and equals the expected relative jump size in returns:

$$\mu_J := \mathbb{E}^{\mathbb{P}}\left(e^{J_t} - 1\right) = \frac{1-p^-}{1-\eta^+} + \frac{p^-}{1+\eta^-} - 1, \tag{4.2}$$

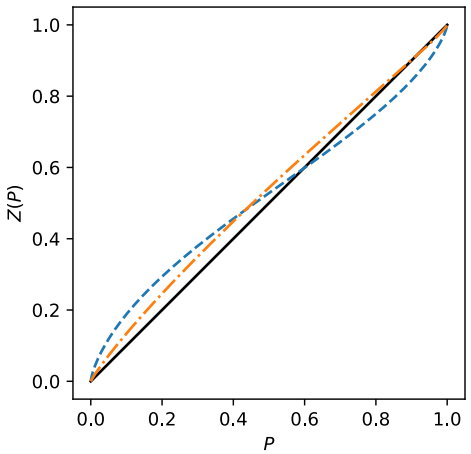
see also Boswijk et al. (2024). We compute sample paths using an Euler discretization, with initial volatility values drawn from its stationary distribution.[12] Given the volatility level at the start of each month, we compute the conditional characteristic function of the physical return distribution following Duffie et al. (2000). The physical density follows by Fourier inversion.

Constructing the risk-neutral densities requires modeling the conditional pricing kernel (cf. (2.2))

$$m_t(r) = u'(r;\gamma_0)Z'(F_t(r)). \tag{4.3}$$

We specify the utility function u as CRRA with $\gamma_0 = 2$. For the probability weighting function Z, we consider an inverse-S shape using Tversky and Kahneman (1992) with $\delta = 0.75$, a nearly globally concave weighting using Prelec (1998) with $(\alpha, \beta) = (0.9, 1.1)$, and linear weighting, as described in Section 3.10. We denote their PWFs by Z_{TK}, Z_P and Z_0, respectively. The resulting weighting functions are displayed in Figure 1.

Figure 1: The probability weighting functions



Note: This figure displays the probability weighting functions Z_{TK} (blue, dashed), Z_P (orange, dash-dotted), Z_0 (black, solid).

The weighting densities Z'_{TK} and Z'_P diverge at 0 and 1. For numerical stability we therefore truncate the physical CDF F_t at 0.0001 and 0.9999 when calculating the pricing kernel (4.3). As this occurs far in the tail, minimal bias is introduced. We calculate the risk-neutral density q_t by multiplying the conditional physical density and pricing kernel, and normalizing to integrate

[12] In practice, we simulate a single path of $100 + (T+5)N$ years using an Euler discretization, of which the first 100 years serve as a burn-in period. The remainder is then cut into N blocks of $T+5$ years, of which we drop the last 5 to reduce the dependence between the replications.

to one. Our simulations thus yield N draws of the sample path $\{R_{t+1}, q_t\}_{t=1}^T$.

We simulate the profile likelihood estimator developed in Section 3 using the kernel density (3.6) to estimate the unknown weighting function g_γ. We use a fourth-order Gaussian kernel $K(u) = \frac{1}{2}(3-u^2)\phi(u)$, with ϕ the standard normal density, and consider a variety of bandwidths. We consider the untrimmed ($v^* = 0$), and the trimmed estimator for the parameter space $\Theta = [-5, 10]$ and a range of v^*. The trimmed estimators depend on the kernel CDF estimator (3.13), for which we consider a smaller bandwidth $h^* = h/2$ than used for the density estimation.[13]

4.2 Simulation results

Table 1 summarizes the properties of the profile likelihood estimator (PLE) based on the $N = 1000$ replications. The bottom row of each panel also shows results for the parametric maximum likelihood estimator (MLE), which fixes the weighting function as the identity map.

This MLE performs well under the expected utility model Z_0, in which case it is correctly specified. However, its misspecification under the probability weighting models Z_{TK} and Z_P results in a positively biased risk aversion estimator, which is used to partially offset the effect of probabilistic risk aversion. Under either form of probability weighting, the PLE outperforms the misspecified MLE for all considered bandwidths and trimming levels. The upward bias in the MLE can lift its MSE to five times that of the PLE, while without probability weighting, the PLE has less than double the MSE of the MLE for reasonably large bandwidths and trimming values. The PLE performs particularly well under the inverse-S shape weighting function of Tversky and Kahneman (1992): its MSE values are the lowest under this DGP, and even lower than those of the MLE under Z_0. In contrast, under the nearly globally concave Prelec (1998) parametrization, the weighting function and the utility function have a similar effect on the pricing kernel, which complicates their separate identification. This leads to a higher PL bias and variance under Z_P than under Z_{TK}. The integrated mean squared error (IMSE) of the nonparametric probability weighting function estimator, estimated as the inverse of (3.5), is fairly stable across tuning parameters and positively correlated with the MSE. This is not surprising as the nonparametric estimator directly depends on the parametric risk aversion estimator.

Our asymptotic theory establishes the consistency of the PLE assuming a trimming level $v^* > 0$. Indeed, without trimming, the simulated PLE under Z_0 displays a clear downward bias. This can be explained by the downward boundary bias of the kernel density estimator, which incentivizes a lower γ to increase values of $U_{t+1}(\gamma)$ that are close to zero. Trimming effectively counters this bias, as the trimmed likelihood only evaluates the kernel density away from the boundaries.[14] Still, under Z_{TK} and Z_P, the PLE actually performs best without trimming. Both forms of probability weighting imply that $g(v)$ vanishes at the boundaries, which removes the leading

[13] The factor half is motivated by the faster convergence of the kernel CDF estimator. With fourth-order kernels, the optimal bandwidth rates for the kernel PDF and CDF estimator are $O(T^{-1/9})$ and $O(T^{-1/6})$, respectively. For $T = 300$ and equal constants, the CDF bandwidth should scale by $T^{1/9-1/6} = 0.43$.

[14] Alternatively, a simple first-order boundary bias correction can be applied to the kernel density estimator. We found that this correction reduces the bias under Z_0, but also substantially increases the estimation variance for all three DGPs. As this results in higher MSE levels, we do not consider the boundary correction further.

Table 1: Profile likelihood performance.

	$v^* = 0$				$v^* = 0.001$				$v^* = 0.01$				$v^* = 0.05$			
h	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE
0.15	-0.24	1.22	1.55	0.88	0.01	1.28	1.65	0.96	0.20	1.47	2.21	1.20	0.74	1.40	2.50	1.41
0.20	-0.09	0.97	0.95	0.55	0.14	1.02	1.06	0.62	0.39	1.18	1.55	0.84	0.86	1.24	2.28	1.23
0.25	0.09	0.85	0.74	0.41	0.31	0.89	0.88	0.51	0.63	1.03	1.46	0.77	1.13	1.16	2.63	1.35
0.30	0.30	0.82	0.76	0.40	0.52	0.84	0.97	0.52	0.90	0.96	1.73	0.88	1.48	1.07	3.34	1.65
MLE	1.92	0.76	4.26	1.88	1.92	0.76	4.26	1.88	1.90	0.77	4.21	1.85	1.81	0.80	3.93	1.71

(a) Z_{TK}

	$v^* = 0$				$v^* = 0.001$				$v^* = 0.01$				$v^* = 0.05$			
h	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE
0.15	-0.13	1.47	2.18	0.72	0.52	1.54	2.63	0.85	0.85	1.76	3.82	1.13	1.13	1.45	3.38	1.18
0.20	0.02	1.33	1.76	0.54	0.61	1.37	2.24	0.68	1.05	1.55	3.50	0.99	1.23	1.47	3.68	1.13
0.25	0.14	1.22	1.51	0.42	0.68	1.25	2.02	0.57	1.18	1.40	3.36	0.91	1.39	1.54	4.29	1.20
0.30	0.26	1.19	1.48	0.37	0.77	1.20	2.05	0.54	1.31	1.35	3.52	0.93	1.62	1.55	5.02	1.34
MLE	2.04	1.12	5.41	1.26	2.05	1.11	5.44	1.26	2.06	1.12	5.52	1.27	2.08	1.15	5.68	1.28

(b) Z_P

	$v^* = 0$				$v^* = 0.001$				$v^* = 0.01$				$v^* = 0.05$			
h	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE
0.15	-1.91	1.54	6.02	1.52	-0.97	1.60	3.52	1.00	-0.44	1.83	3.53	0.99	0.68	1.41	2.44	0.89
0.20	-1.90	1.40	5.58	1.36	-1.04	1.45	3.20	0.87	-0.32	1.64	2.79	0.77	0.41	1.48	2.35	0.77
0.25	-1.89	1.29	5.22	1.22	-1.09	1.32	2.94	0.76	-0.27	1.48	2.25	0.62	0.34	1.57	2.59	0.74
0.30	-1.85	1.24	4.99	1.13	-1.10	1.27	2.82	0.69	-0.21	1.41	2.04	0.56	0.47	1.63	2.86	0.75
MLE	0.07	1.18	1.40	0.28	0.08	1.18	1.39	0.28	0.07	1.19	1.42	0.28	0.06	1.22	1.48	0.26

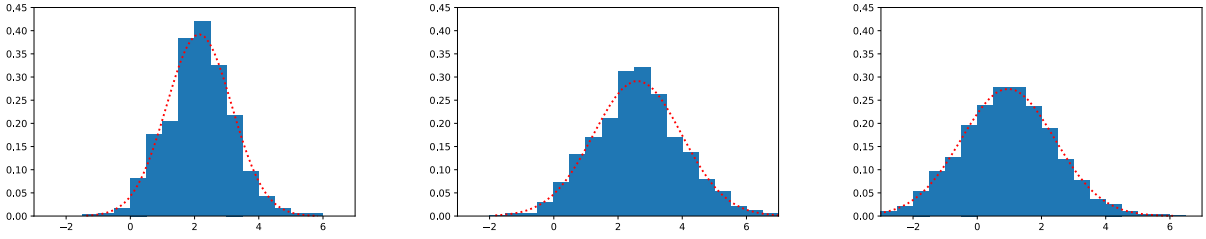
(c) Z_0

Note: Subtables display the bias, standard deviation, and mean squared error (MSE) of the profile likelihood estimator $\hat{\gamma}$ over $N = 1000$ replications of samples of length $T = 300$ for various levels of trimming v^*, with bandwidth $h^* = h/2$ for the CDF estimator. Each subpanel represents a different true probability weighting function. Columns labeled IMSE display the integrated mean squared error of the nonparametric estimator $\widehat{Z}$, multiplied by 1,000. The bottom rows represent the maximum likelihood estimator of the expected utility model, which fixes the weighting function as the identity map.

boundary bias in the kernel density estimator. Meanwhile, trimming slightly increases the variance of the PLE as it does not use all information in the data. Overall, since not trimming is more harmful under Z_0 than it is beneficial under Z_{TK} and Z_P, a small amount of trimming appears prudent when it is unknown whether probability weighting is present.

Figure 2 displays histograms of the PLE $\hat{\gamma}$ at $h = 0.2$ and $v^* = 0.001$ for the three models considered. The histograms confirm that the PLE is approximately Gaussian for all three models at the realistic sample size of $T = 300$. The choice of tuning mainly affects the location and scale of the sampling distributions. The simulated nonparametric estimates of $\widehat{Z}$ are displayed in Figure 3 for the same bandwidth and trimming level. When either form of probability weighting is present, the estimated functions follow the shape of the true weighting function, and their pointwise means almost completely overlap with the true functions. In the expected utility model, the PL estimate is biased under the chosen tuning, which carries over to the nonparametric probability weighting function estimator. Nonetheless, the simulated pointwise quantiles uniformly contain the true weighting function.

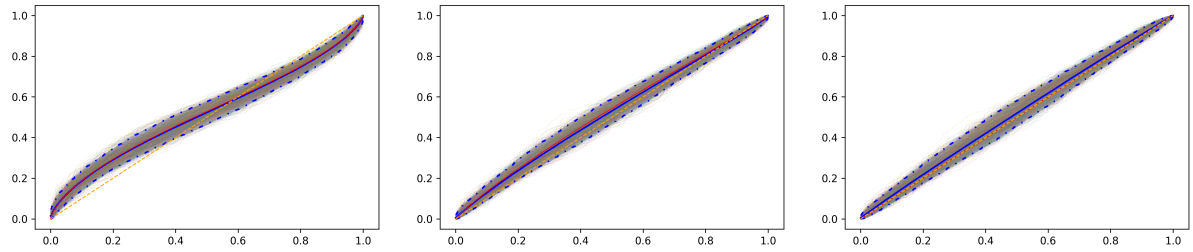
Figure 2: Simulation estimates of the utility parameter.



(a) Z_{TK} (b) Z_P (c) Z_0

Note: The subpanels display histograms of the Monte Carlo estimates $\hat{\gamma}$ for $h = 0.2$ and $v^* = 0.001$, for three different true probability weighting functions and $\gamma_0 = 2$. The dotted red curve is a normal density with mean and standard deviation equal to that of $\hat{\gamma}$ over the $N = 1000$ replications.

Figure 3: Simulation estimates of the probability weighting function.



(a) Z_{TK} (b) Z_P (c) Z_0

Note: The subpanels display the nonparametric estimates $\widehat{Z}$ of the probability weighting function, for three different true probability weighting functions. Each of the $N = 1000$ lines represents a single simulation. The mean (blue, solid), lower and upper 2.5% percentiles (blue, dash-dotted), and the true PW function (red, solid), are also displayed, along with the 45-degree line (orange, dashed).

4.2.1 Bootstrap confidence intervals

To compute confidence intervals for the PL estimation error $\hat{\gamma} - \gamma_0$, we consider the bootstrap described in Section 3.9. We consider confidence intervals based on the bootstrap quantiles as well as one exploiting the asymptotic normal distribution using the bootstrap standard deviation. Table 2 reports the associated coverage rates based on $K = 500$ bootstrap repetitions for each of the $N = 1000$ simulations. Both types of the bootstrapped intervals achieve coverage rates that are close to the nominal rates of 90% and 95%, with some slight under-coverage that can be explained by the parameter estimation biases. The similar performance of the quantile- and standard deviation-based intervals confirms the approximate normality of the PLE.

4.2.2 Testing for the presence of nonlinear probability weighting

If the probability weighting function is truly linear, then the PITs $U_{t+1}(\gamma_0)$ have a standard uniform distribution. Moreover, the MLE is correctly specified and has good finite-sample properties (see Table 1). Therefore, in this subsection, we consider testing the null of linear probability weighting, leveraging the method of Bai (2003). This test uses a Kolmogorov-Smirnov test for the uniformity of the PITs, while using a martingale transformation to correct for the estimation error of the unknown parameter. The correction relies on the score function,

Table 2: Bootstrap results.

		Z_{TK}		Z_P		Z_0	
	CI (%)	90	95	90	95	90	95
Quantile-based	Coverage (%)	88.3	92.7	88.8	93.7	85.7	92.4
	Width	3.23	3.84	4.39	5.22	5.59	6.65
St.d.-based	Coverage (%)	89.8	93.9	89.6	94.8	89.0	94.9
	Width	3.25	3.87	4.41	5.26	5.62	6.70

Note: This table displays the result of the nonparametric bootstrap, see Section 3.9. The columns are based on different DGPs, in the form of a different true probability weighting function. The bandwidth is set at $h = 0.2$, while the trimming level is set at $v^* = 0.01$ for Z_0 and $v^* = 0$ for Z_{TK} and Z_P. We consider $K = 500$ bootstrap replications for each of the $N = 1000$ simulations.

which is straightforward to compute under the expected utility model.[15]

As we consider the correct utility function in the Monte Carlo simulations, the only asymptotic cause of rejection is the presence of nonlinear probability weighting. The rejection rates of the Bai (2003) test are displayed in Table 3 below, based on the MLE using various levels of trimming. The empirical size of the test corresponds to the rejection rates under Z_0. They are close to, and generally below, their nominal value for all choices of α and v^*. Meanwhile, the power of the test against the inverse-S shaped alternative Z_{TK} is excellent, but less so for the nearly globally concave alternative Z_P, which is harder to distinguish from the marginal utility function, as discussed above. For a confidence level of $\alpha = 0.05$, the power of the test is slightly increasing with v^*, whereas its size is slightly decreasing.

Table 3: Testing the probability weighting function.

		Z_{TK}			Z_P			Z_0		
	α (%)	10	5	1	10	5	1	10	5	1
Rejection rate (%)	$v^* = 0$	100	100	100	23.9	14.6	4.1	9.8	5.4	0.8
	$v^* = 0.001$	100	100	100	24.1	14.9	4.1	9.9	5.2	0.8
	$v^* = 0.01$	100	100	100	26.0	15.5	4.2	9.5	4.4	0.7
	$v^* = 0.05$	100	100	99.9	24.5	15.6	3.9	9.0	4.1	0.7

Note: This table displays the rejection rate of the Bai (2003) test at level α. The first two sets of three columns consider the power under two different alternatives, the last set considers a DGP where the null is true and display the size of the test. The rejection rates are based on $N = 1000$ simulations. Critical values are constructed by simulation.

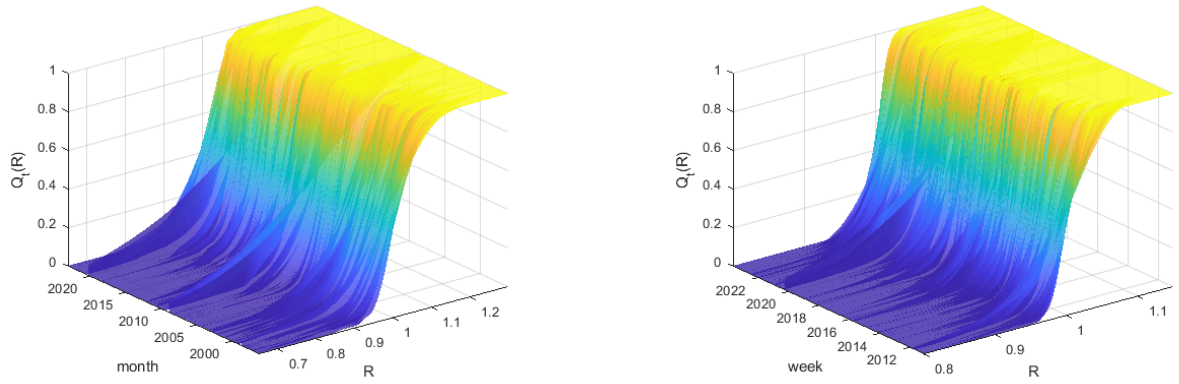
5 Empirical Results

In this section, we illustrate our econometric method using European-style option prices on the S&P 500 index (SPX) obtained from OptionMetrics. We consider both monthly option data from January 1996 to January 2023 and weekly data from January 2011 to August 2023.

[15] Estimating the correction term requires reducing the domain in the Kolmogorov-Smirnov statistic from $[0, 1]$ to $[0, 1 - \varepsilon]$ for some $\varepsilon > 0$. Though Bai (2003) suggests taking ε small, we find that this leads to near-singularity of a matrix that needs to be inverted to compute the correction term. However, the simple choice of $\varepsilon = \frac{1}{2}$ results in nearly nominal size, whereas lower values lead to poor size, and higher values to low power. The resulting test focuses on the left half of the distribution, so we apply it to both the primal and dual probability weighting functions separately.

In particular, we consider monthly option contracts expiring on the third Friday of each month, and record their mid-quote prices on the last trading day with at least $\tau = 29$ days to maturity. The shorter weekly dataset contains the latest recorded mid prices for options expiring the next Friday with at least $\tau = 7$ days to maturity. The relevance of weekly options is discussed in detail in, e.g., Andersen et al. (2017). We remove options that violate static arbitrage bounds, have negative bid-ask spreads, or whose implied volatility is unavailable. We then apply the constrained least squares method of Aït-Sahalia and Duarte (2003) to ensure each period's cross section of option prices is monotone and convex in the strike price. Finally, the risk-neutral distributions Q_t of the futures return are estimated from the resulting option price cross-sections using the local cubic kernel estimator of Dalderop (2020), with a plug-in pilot bandwidth based on fitting the single-factor stochastic volatility jump-diffusion model in Bates (1996).[16] Figure 4 shows that our estimation method produces valid and smooth distributions throughout the sample. We complete the data set with futures returns computed from the SPX forward prices for matching expiry dates in OptionMetrics.

Figure 4: Nonparametrically estimated risk-neutral cumulative distributions for S&P 500 index returns.



(a) Monthly risk-neutral CDFs.

(b) Weekly risk-neutral CDFs.

Note: These figures display nonparametric kernel estimators of the risk-neutral CDFs estimated from monthly and weekly option prices.

With the sample $\{R_{t+1}, Q_t\}_{t=1}^{T}$ of returns and risk-neutral CDFs at hand, we start the empirical analysis by estimating an expected utility model. In particular, we define the maximum likelihood (ML) estimator by maximizing (3.12) as a function of the utility parameter γ, while fixing the probability weighting function as the identity map. We use a trimming level $v^* = 0.05$ for the ML estimator, for which the Bai (2003) test has the highest power and lowest size in simulations for confidence level $\alpha = 0.05$. The normalization constants are computed using the integration-by-parts formulae (3.8) and (3.9). The resulting parameter estimates $\hat{\gamma}^{\mathrm{ML}}$ for both frequencies are displayed in the bottom row of Table 4. We test for the null of expected utility against the general alternative of non-expected utility, by computing the test statistic of Bai (2003) for the estimated utility-adjusted PITs $U_{t+1}(\hat{\gamma}^{\mathrm{ML}})$. For the monthly data this results in

[16] Since nonparametric methods become unstable in the tails of the density due to sparse trading of deep OTM options, we 'paste' the tails of the Bates (1996) model to match at the lower and upper moneyness thresholds that leave 10 observed option prices in either tail. On average, this leaves about 1% of the risk-neutral probability mass in either tail, with the highest fractions in the early years of the sample.

p-values of 0.031 and 0.089 for the primal (left half) and dual (right half), respectively. For the weekly data the corresponding p-values are 0.170 and 0.581. This can be interpreted as significant evidence of non-expected utility in the monthly data, observable in the left-tail of the distribution. We further interpret these test results after discussing the estimated probability weighting functions in Figure 5.

Next, we estimate the rank-dependent utility model using our profile likelihood estimation procedure. Thus, we maximize the likelihood (3.11) as a function of the utility parameter γ by profiling out the probability weighting function based on the kernel density estimator (3.6) of the utility-adjusted PITs computed using (3.7). We report our estimates in the top row of Table 4. The probability weighting function is then estimated based on the PITs at the profile likelihood maximizer $\hat{\gamma}^{\mathrm{PL}}$, see Figure 5.

The bootstrap confidence intervals of the PL estimates contain zero at both frequencies, so that we cannot reject linear utility. This contrasts with the significantly positive risk aversion parameter estimate in the expected utility model. The lower risk aversion level is partially picked up by the probability weighting function, which leads to the estimated weighting functions in Figure 5 that are visibly distinct from the 45-degree line. The magnitude of the respective divergences from linearity are in line with the p-values found in the Bai (2003) test. For monthly data, the probability weighting function exhibits the typical inverse-S shape, although with larger concave than convex regions. The probability weighting function based on weekly data is closer to being linear, but appears to be globally concave. The latter shape may contribute to the higher p-values of the Bai (2003) test for weekly data, according to Table 3.

Table 4: Estimates for γ.

	Monthly	Weekly
PL	0.59 (-1.24, 2.43)	2.87 (-0.26, 6.00)
ML	2.34 (0.62, 4.06)	4.39 (1.49, 7.28)

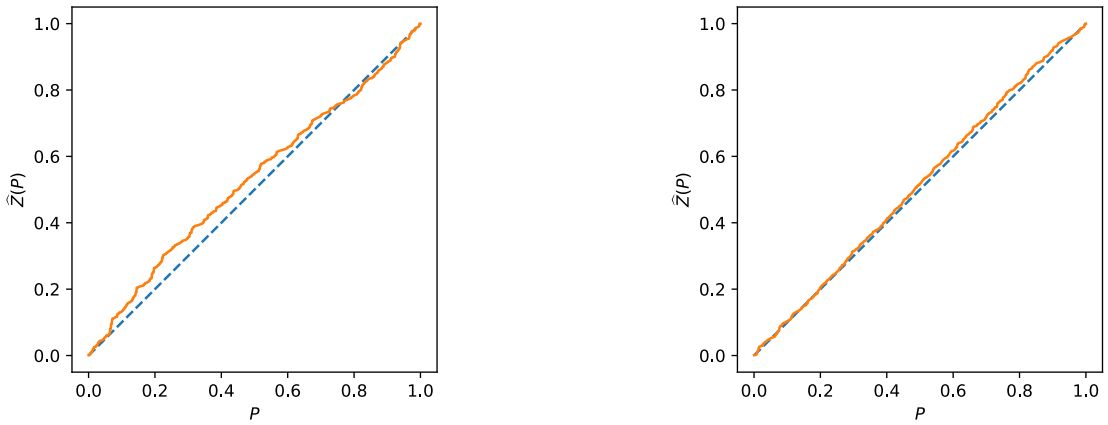
Note: The top row displays the profile likelihood estimates $\hat{\gamma}$, using a fourth-order Gaussian kernel, bandwidth $h = 0.2$, and trimming level $v^* = 0.001$. The bottom row displays the maximum likelihood estimates, which correspond to linear weighting, and uses a trimming level of $v^* = 0.05$. The associated 95% bootstrap confidence intervals, based on a bootstrap estimate of the standard deviation and the Gaussian limit, are displayed in brackets.

The deviations from the 45-degrees line in Figure 5 suggest pronounced probability weighting particularly at the monthly frequency. However, these estimates are subject to the assumption of CRRA utility over return outcomes. An alternative explanation for the findings is that the marginal utility function is misspecified. For robustness, we therefore consider the wider class of marginal utility specifications of the exponential-polynomial form

$$u'(r;\gamma) = \exp\left(-\sum_{l=1}^{L} \gamma_l (\log r)^l\right),$$

which nests power utility as the special case $L = 1$. Rosenberg and Engle (2002) also fit flexible pricing kernel specifications of this form.

Figure 5: Nonparametric estimates of the probability weighting function under CRRA.



(a) Using monthly data.

(b) Using weekly data.

Note: Plots display the estimates of the probability weighting function as the inverse of the empirical CDF (3.5) at $\hat{\gamma}^{\mathrm{PL}}$. The dashed, blue curve is the 45-degrees line.

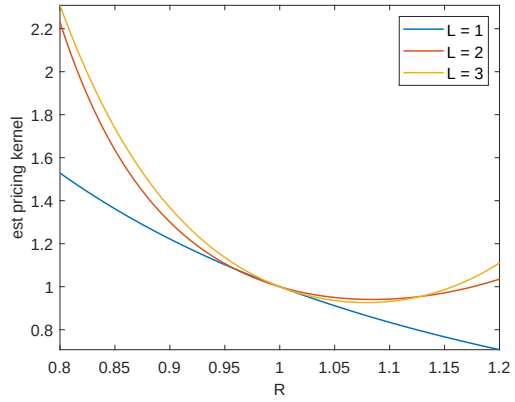
Figure 6 shows the estimated marginal utility functions with and without probability weighting, for varying order L. The parameter space is chosen such that marginal utilities at returns $\pm 50\%$ are within a range of reasonable CRRA utility values. Without probability weighting, estimated marginal utility displays clear non-monotonic U-shapes for monthly data when $L \geq 2$; for $L = 1$, marginal utility is monotone by construction. The increasing marginal utility for positive returns is in line with the literature on the 'pricing kernel puzzle' (see Cuesdeanu and Jackwerth (2018) for a survey). For weekly data, the higher order marginal utility estimates are not strictly monotonically decreasing either, as they flatten for positive returns, but to a much lesser extent than for monthly data. However, when probability weighting is allowed for, the marginal utility estimates decrease monotonically on the range $(0.8, 1.2)$ for all orders of L and both frequencies considered. Thus, the presence of probability weighting allows replacing puzzling U-shaped marginal utilities by more plausible monotonically decreasing functions that closely resemble CRRA utility. Moreover, probability weighting substantially reduces the amount of curvature in the marginal utility functions required to match option prices and returns.

Figure 7 shows the resulting estimates of the probability weighting function as the inverse of the empirical CDF (3.5). The weighting functions are nearly indistinguishable for different orders L of the exponential-polynomial utility model. Using monthly data yields particularly pronounced inverse-S shapes, while using weekly data yields weighting functions that are closer to linear. The plots also reveal a slight asymmetry, with the nonlinear weighting being stronger for the left than the right tail, implying the probabilities of large negative returns are most overweighted.

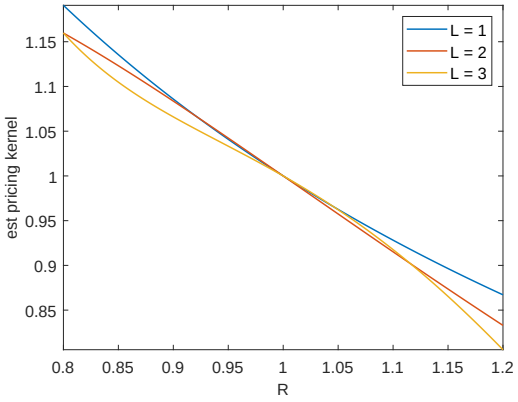
6 Conclusion

The intertwined nature of attitudes toward wealth and toward probabilities challenges their joint identification based on financial market data. Using the time series of risk-neutral distributions implicit in option contracts and the realized returns, this paper develops a semiparametric profile likelihood estimator of the rank-dependent utility model that jointly identifies the probability

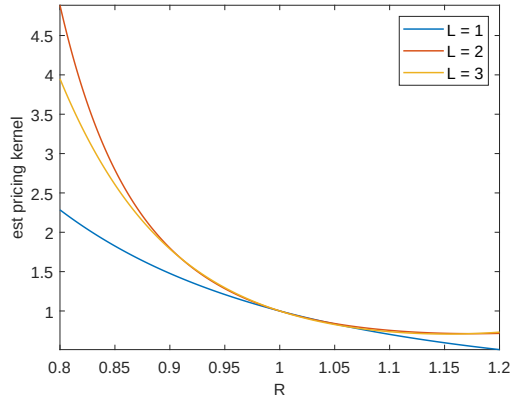
Figure 6: Estimated exponential-polynomial marginal utility functions for varying order L.



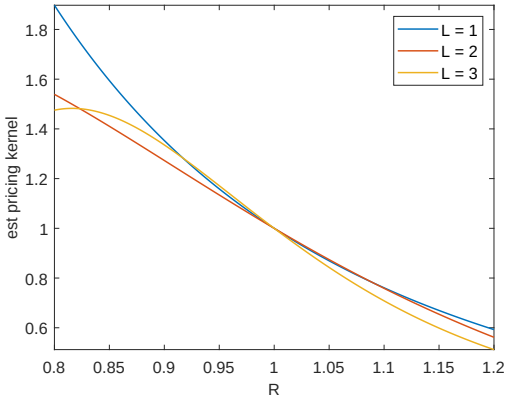
(a) Without probability weighting, monthly data.



(b) With probability weighting, monthly data.



(c) Without probability weighting, weekly data.

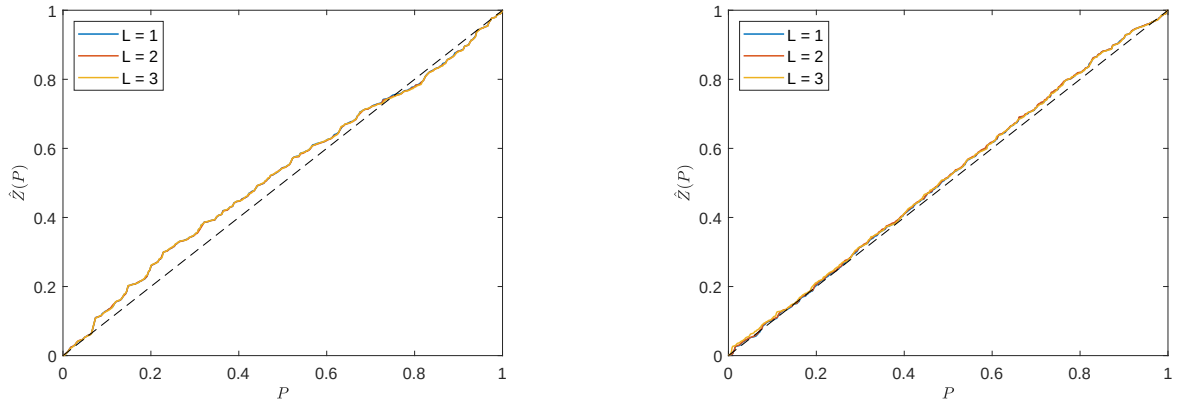


(d) With probability weighting, weekly data.

Note: Plots display (profile) maximum likelihood-estimated exponential-polynomial marginal utility functions for order $L = 1, 2, 3$, based on the local linear risk-neutral CDF estimator, and a Gaussian fourth-order kernel with bandwidth $h = 0.25$ for monthly and $h = 0.20$ for weekly data and trimming constant $v^* = 0.001$ for the probability weighting density estimator. For $L \geq 2$, trimming is based on the parameter set $\Theta_L = \{\theta \in \mathbb{R}^L : R_l^{\gamma_u} \leq u_L'(R_l;\theta) \leq R_l^{-\gamma_l}$ and $R_u^{\gamma_l} \leq u_L'(R_u;\theta) \leq R_u^{-\gamma_u}$ for $(R_l, R_u) = (0.5, 1.5), (\gamma_l, \gamma_u) = (-5, 10)\}$.

weighting and utility functions, without restrictive assumptions on the physical return dynamics. We establish the asymptotic properties of our estimation procedure. Monte Carlo simulations demonstrate the favorable performance of our approach in finite samples. Our empirical analysis of two large samples of monthly and weekly S&P 500 index option prices and returns unveils the importance of probability weighting. Probability weighting is particularly pronounced at the monthly horizon, where the weighting function implicit in option prices is found to be inverse S-shaped. These findings appear robust to the parametric specification of the utility function. Our results, and the nonlinearities in probabilities of investors' risk evaluation that they entail, contribute to our understanding of risk preferences, and have practical implications for option pricing, hedging, and risk management.

Figure 7: Nonparametric estimates of the probability weighting function under exponential-polynomial utility.



(a) Using monthly data.

(b) Using weekly data.

Note: Plots display the estimates of the probability weighting function as the inverse of the empirical CDF (3.5) at the profile ML estimates $\hat{\gamma}^{\text{PL}}$ for varying order L of exponential-polynomial utility models from Figure 6. The weighting functions are nearly indistinguishable for different orders L. The dashed, blue curve is the 45-degrees line.

Appendix

A Proofs

The following lemma is used in several of the proofs that follow.

Lemma D. *Suppose the following conditions hold for some non-negative integer i and multi-index j:*

(i) $u'(r;\gamma)$ *is positive and* (i,j)*-times (continuously) differentiable on* $\mathbb{R}_{++} \times \bar{\Theta}$*, with* $\bar{\Theta} \subseteq \Theta$*;*

(ii) $Z'(P)$ *is positive and* $i + |j|$ *times (continuously) differentiable on* $(0,1)$*;*

(iii) $f_t(r)$ *is positive and* $\max\{i + |j| - 1, 0\}$ *times (continuously) differentiable a.s. on* $\mathbb{R}_{++}$*;*

(iv) $c_t(\gamma)$ *is finite and* j *times (continuously) differentiable a.s. on* $\bar{\Theta}$*.*

Then $g(v;\gamma)$ *is positive and (continuously) differentiable up to order* (i,j) *on* $(0,1) \times \bar{\Theta}$*.*

Proof. The function $U_t(r;\gamma)$ is differentiable w.r.t. r a.s., with derivative

$$U_t'(r;\gamma) = \frac{\partial}{\partial r}\left[c_t(\gamma)\int_0^r \frac{q_t(s)}{u'(s;\gamma)}\,\mathrm{d}s\right] = c_t(\gamma)\frac{q_t(r)}{u'(r;\gamma)} = \frac{c_t(\gamma)}{c_t(\gamma_0)}\frac{u'(r;\gamma_0)}{u'(r;\gamma)}Z'(F_t(r))f_t(r). \tag{A.1}$$

This derivative is $\mathbb{P}$-a.s. positive since all components on the RHS of (A.1) are. Moreover, it is (continuously) differentiable w.r.t. (v,γ) up to order (i,j). Therefore, $r \mapsto U_t(r;\gamma)$ has a strictly monotonic inverse function $v \mapsto U_t^{-1}(v;\gamma)$, whose derivative $\frac{\partial}{\partial v}U_t^{-1}(v;\gamma) = 1/U_t'(U_t^{-1}(v;\gamma);\gamma) > 0$ is (i,j) times (continuously) differentiable.

Using the law of iterated expectations, the CDF of the utility-adjusted PITs equals

$$G_\gamma(v) = \mathbb{P}(U_{t+1}(\gamma) \le v) = \mathbb{P}(R_{t+1} \le U_t^{-1}(v;\gamma)) = \mathbb{E}(F_t(U_t^{-1}(v;\gamma))). \tag{A.2}$$

Its density $g_\gamma(v) = \frac{\partial}{\partial v}G_\gamma(v)$ equals

$$g_\gamma(v) = \mathbb{E}(f_t(U_t^{-1}(v;\gamma))\frac{\partial}{\partial v}U_t^{-1}(v;\gamma)) = \mathbb{E}\left(\frac{c_t(\gamma_0)}{c_t(\gamma)}\frac{u'(U_t^{-1}(v;\gamma);\gamma)}{u'(U_t^{-1}(v;\gamma);\gamma_0)}\frac{1}{Z'(F_t(U_t^{-1}(v;\gamma)))}\right), \tag{A.3}$$

inheriting positivity and order of (continuous) differentiability from c_t, u', Z', F_t, and $\frac{\partial}{\partial v}U_t^{-1}$. □

Proof of Lemma E. Let $\tilde{g}_\gamma$ be a family of densities on $(0,1)$ indexed by γ. Jensen's inequality yields that

$$\ell(\gamma,\tilde{g}_\gamma) - \ell(\gamma,g_\gamma) = \mathbb{E}\left(\log\frac{\tilde{g}_\gamma(U_{t+1}(\gamma))}{g_\gamma(U_{t+1}(\gamma))}\right) \le \log\mathbb{E}\left(\frac{\tilde{g}_\gamma(U_{t+1}(\gamma))}{g_\gamma(U_{t+1}(\gamma))}\right) = \log\int_0^1 \frac{\tilde{g}_\gamma(u)}{g_\gamma(u)}g_\gamma(u)\,\mathrm{d}u = 0,$$

with equality if and only if $\tilde{g}_\gamma(U_{t+1}(\gamma)) = g_\gamma(U_{t+1}(\gamma))$ *a.s.* This requires that the densities are identical on any set in the support of $U_{t+1}(\gamma)$, and equal zero outside this support in order to

integrate to one. Hence, $\tilde{g}_\gamma$ and g_γ are identical everywhere, so that g_γ is the unique maximizing function. □

Proof of Lemma I. Denote $\mathbb{E}_t(\cdot) := \mathbb{E}(\cdot|\mathcal{F}_t)$. For any γ, the law of iterated expectations and Jensen's inequality yield

$$\begin{aligned}\ell(\gamma) - \ell(\gamma_0) = \mathbb{E}\left(\log \frac{f_t(R_{t+1};\gamma,g_\gamma)}{f_t(R_{t+1};\gamma_0,g_{\gamma_0})}\right) &= \mathbb{E}\left(\mathbb{E}_t\left(\log \frac{f_t(R_{t+1};\gamma,g_\gamma)}{f_t(R_{t+1};\gamma_0,g_{\gamma_0})}\right)\right) \\ &\leq \mathbb{E}\left(\log \mathbb{E}_t\left(\frac{f_t(R_{t+1};\gamma,g_\gamma)}{f_t(R_{t+1};\gamma_0,g_{\gamma_0})}\right)\right) = 0,\end{aligned}$$

where the inequality holds with equality if and only if (3.11) holds.

By Lemma D, the density $g_\gamma(v)$ is positive so that $G_\gamma(u)$ is strictly monotonic for all $\gamma \in \Theta$. Equation (3.11) is therefore equivalent to any of the following statements:

$$\begin{aligned} f_t(r;\gamma,g_\gamma) = f_t(r;\gamma_0,g_{\gamma_0}) \text{ a.s. } \forall r &\Leftrightarrow F_t(r;\gamma,g_\gamma) = F_t(r;\gamma_0,g_{\gamma_0}) \text{ a.s. } \forall r \\ &\Leftrightarrow F_t(F_t^{-1}(v);\gamma,g_\gamma) = v \text{ a.s. } \forall v \in (0,1) \\ &\Leftrightarrow G_\gamma\left(U_t(F_t^{-1}(v);\gamma)\right) = v \text{ a.s. } \forall v \in (0,1) \\ &\Leftrightarrow U_t(F_t^{-1}(v);\gamma) = G_\gamma^{-1}(v) \text{ a.s. } \forall v \in (0,1).\end{aligned}$$

The LHS of the final equation describes the conditional quantile function of $U_{t+1}(\gamma)$. For the final statement to hold, it should be constant over time for any u. However, the quantile density function equals

$$\frac{\partial}{\partial v}U_t(F_t^{-1}(v);\gamma) = c_t(\gamma)\frac{\partial}{\partial v}\int_0^{F_t^{-1}(v)} \frac{q_t(r)}{u'(r;\gamma)}\,\mathrm{d}r = \frac{c_t(\gamma)}{c_t(\gamma_0)}\frac{u'(F_t^{-1}(v);\gamma_0)}{u'(F_t^{-1}(v);\gamma)}Z'(v),$$

using (2.4) in the second equation, and its log-derivative equals

$$\frac{\partial}{\partial v}\log\left(\frac{\partial}{\partial v}U_t(F_t^{-1}(v);\gamma)\right) = \frac{1}{f_t(F_t^{-1}(v))}\left(\frac{u''(F_t^{-1}(v);\gamma_0)}{u'(F_t^{-1}(v);\gamma_0)} - \frac{u''(F_t^{-1}(v);\gamma)}{u'(F_t^{-1}(v);\gamma)}\right) + \frac{Z''(v)}{Z'(v)}. \quad \text{(A.4)}$$

The RHS of (A.4) can only be constant over time for a given $v \in (0,1)$ if (i) $\frac{u''(F_t^{-1}(v);\gamma_0)}{u'(F_t^{-1}(v);\gamma_0)} = \frac{u''(F_t^{-1}(v);\gamma)}{u'(F_t^{-1}(v);\gamma)}$ a.s. or if (ii) $f_t(F_t^{-1}(v)) = a(v)\left(\frac{u''(F_t^{-1}(v);\gamma_0)}{u'(F_t^{-1}(v);\gamma_0)} - \frac{u''(F_t^{-1}(v);\gamma)}{u'(F_t^{-1}(v);\gamma)}\right)$ a.s. for some non-zero function $a(v)$.

Let $\mathcal{V}(\bar{r})$ be the support of $F_t(\bar{r})$. Then case (i) cannot hold for any $v \in \mathcal{V}(\bar{r})$ since I*(ii)* implies that $ARA(r;\gamma) \neq ARA(r;\gamma_0)$ in a neighborhood of $\bar{r}$. Case (ii) implies for any $(v_l, v_h) \in \mathcal{V}(\bar{r})$ with $v_l < v_h$ that

$$\log \frac{\frac{\partial}{\partial v}U_t(F_t^{-1}(v_h);\gamma)}{\frac{\partial}{\partial v}U_t(F_t^{-1}(v_l);\gamma)} = \int_{v_l}^{v_h}\left(\frac{1}{a(v)} + \frac{Z''(v)}{Z'(v)}\right)\mathrm{d}v. \quad \text{(A.5)}$$

The constant RHS term implies a one-to-one relation between $F_t^{-1}(v_l)$ and $F_t^{-1}(v_h)$, which is ruled out by I*(iii)*. Therefore, for some $v \in \mathcal{V}(\bar{r})$ neither (i) nor (ii) holds, so that (A.4) is non-

deterministic. As a result, none of the statements equivalent to (3.11) hold, and $\ell(\gamma) < \ell(\gamma_0)$ for $\gamma \neq \gamma_0$. □

Proof of Lemma I.* The censored log-likelihood population criterion is given by

$$\ell^*(\gamma) := \mathbb{E}\left(\log\left(G_\gamma(U_t^l(\gamma))\right)^{1_{t+1}^l}\left(c_t(\gamma)/u'(R_{t+1};\gamma)g_\gamma\left(U_{t+1}(\gamma)\right)\right)^{1_{t+1}^m}\left(1-G_\gamma(U_t^u(\gamma))\right)^{1_{t+1}^u}\right).$$

It holds that

$$\begin{aligned}
&\ell^*(\gamma) - \ell^*(\gamma_0) \\
&\leq \mathbb{E}\left(\log \mathbb{E}_t\left(\left(\frac{F_t(R_{t,l}^*;\gamma)}{F_t(R_{t,l}^*;\gamma_0)}\right)^{1_{t+1}^l}\left(\frac{f_t(R_{t+1};\gamma,g_\gamma)}{f_t(R_{t+1};\gamma_0,g_{\gamma_0})}\right)^{1_{t+1}^m}\left(\frac{1-F_t(R_{t,u}^*;\gamma)}{1-F_t(R_{t,u}^*;\gamma_0)}\right)^{1_{t+1}^u}\right)\right) \\
&= \mathbb{E}\left(\log\left(F_t(R_{t,l}^*;\gamma) + \int_{R_{t,l}^*}^{R_{t,u}^*} f_t(r;\gamma,g_\gamma)\,\mathrm{d}r + (1-F_t(R_{t,u}^*;\gamma))\right)\right) = 0,
\end{aligned}$$

where the Jensen's inequality holds with equality if and only if

$$F_t(R_{t,i}^*;\gamma) = F_t(R_{t,i}^*;\gamma_0) \text{ for } i \in \{l,u\} \text{ and } 1_{t+1}^m\left(f_t(R_{t+1};\gamma,g_\gamma) - f_t(R_{t+1};\gamma_0,g_{\gamma_0})\right) = 0 \quad \text{a.s.} \tag{A.6}$$

By the equivalence statements in the proof of Lemma I, this requires that $F_t(F_t^{-1}(u);\gamma,g_\gamma) = u$ for all $u \in (F_t(R_{t,l}^*), F_t(R_{t,u}^*))$ a.s. Since $(R_{t,l}^*, R_{t,u}^*)$ contains $\bar{r}$ with positive probability, case (ii) in the proof of Lemma I then implies that (A.5) holds for any pair $(v_l, v_h) \in \text{supp}(F_t(\bar{r}) \mid R_{t,l}^* \leq \bar{r} \leq R_{t,u}^*)$. However, the resulting one-to-one relation between $F_t^{-1}(v_l)$ and $F_t^{-1}(v_h)$ cannot hold for the pair (v_1, v_2) in the statement of Lemma I*. As a result, (A.6) cannot hold, and $\ell^*(\gamma) < \ell^*(\gamma_0)$ for $\gamma \neq \gamma_0$. □

Proof of Proposition C. If (i) for every neighborhood Θ_0 of γ_0, $\max_{\gamma\in\Theta\setminus\Theta_0}\ell^*(\gamma) < \ell^*(\gamma_0)$ and (ii) $\sup_{\gamma\in\Theta}\left|\ell_T^*(\gamma) - \ell^*(\gamma)\right| \xrightarrow{p} 0$, then $\hat{\gamma} \xrightarrow{p} \gamma_0$ (e.g., Andrews, 1994, Lemma A-1). The identification condition (i) follows from Lemma I* given Assumption I. To establish the uniform convergence condition (ii), write

$$\sup_{\gamma\in\Theta}\left|\ell_T^*(\gamma) - \ell^*(\gamma)\right| \leq \sup_{\gamma\in\Theta}\left|\ell_T^*(\gamma) - \ell_T^*(\gamma,g_\gamma)\right| + \sup_{\gamma\in\Theta}\left|\ell_T^*(\gamma,g_\gamma) - \ell^*(\gamma)\right|. \tag{A.7}$$

The first term in (A.7) is bounded by

$$\begin{aligned}
&|\ell_T^*(\gamma) - \ell_T^*(\gamma,g_\gamma)| \\
&\quad\leq \frac{1}{T}\sum_{t=1}^T 1_{t+1}^l\left|\log\frac{\widehat{G}_\gamma(U_t^l(\gamma))}{G_\gamma(U_t^l(\gamma))}\right| + 1_{t+1}^m\left|\log\frac{\widehat{g}_\gamma(U_{t+1}(\gamma))}{g_\gamma(U_{t+1}(\gamma))}\right| + 1_{t+1}^u\left|\log\frac{1-\widehat{G}_\gamma(U_t^u(\gamma))}{1-G_\gamma(U_t^u(\gamma))}\right|.
\end{aligned}$$

Since g_γ is positive and continuous by Lemma D, the constant $c_g := \min_{\gamma\in\Theta}\min_{v\in[v^*,1-v^*]} g_\gamma(v)$ is positive. Using $\left|\log\frac{x}{y}\right| \leq \frac{1}{x\wedge y}|y-x|$ for any $x, y > 0$, uniform convergence of $\widehat{g}$ therefore

implies that for any constant $0 < \delta < 1$, with probability approaching (w.p.a.) 1

$$\sup_{\gamma\in\Theta} \frac{1}{T}\sum_{t=1}^{T} 1_{t+1}^{m} \left|\log \frac{\widehat{g}_\gamma(U_{t+1}(\gamma))}{g_\gamma(U_{t+1}(\gamma))}\right| \leq \frac{1}{\delta c_g} \sup_{\gamma\in\Theta} \|\widehat{g}_\gamma - g_\gamma\|_{\infty,v^*} \xrightarrow{p} 0.$$

Similarly, consider the positive constant $c_G^l := \min_{\gamma\in\Theta} G_\gamma(v^*)$. Then w.p.a. 1

$$\sup_{\gamma\in\Theta} \frac{1}{T}\sum_{t=1}^{T} 1_{t+1}^{l} \left|\log \frac{\widehat{G}_\gamma(U_t^l(\gamma))}{G_\gamma(U_t^l(\gamma))}\right| \leq \frac{1}{\delta c_G^l} \sup_{\gamma\in\Theta} \left\|\widehat{G}_\gamma - G_\gamma\right\|_{\infty,v^*} \xrightarrow{p} 0,$$

where the final step follows from the uniform convergence of $\widehat{g}$ and the uniform in γ consistency of $\widehat{G}$ at v^*. The same result holds for the third term for the positive constant $c_G^u := 1 - \max_{\gamma\in\Theta} G_\gamma(1-v^*)$.

For the second term in (A.7), pointwise convergence $\ell_T^*(\gamma, g_\gamma) \xrightarrow{p} \ell^*(\gamma)$ follows from the ergodic theorem (Davidson, 1994, Theorem 13.12), as the summands of $\ell_T^*(\gamma, g_\gamma)$ are measurable in the stationary and ergodic sequence $\{R_{t+1}, X_t\}$ and integrable for all $\gamma \in \Theta$. Moreover, the summands are continuous in γ a.s. by Lemma D and dominated by an integrable function, since

$$\begin{aligned} &1_{t+1}^{m} \left|\log c_t(\gamma) - \log u'(R_{t+1};\gamma) + \log g_\gamma(U_{t+1}(\gamma))\right| \\ &\quad \leq \sup_{\gamma\in\Theta} \left(|\log c_t(\gamma)| + \left|\log u'(R_{t+1};\gamma)\right| + 1_{t+1}^{m} \left|\log g_\gamma(U_{t+1}(\gamma))\right|\right). \end{aligned}$$

The first two of these terms are integrable by assumption, while continuity of g_γ implies that the third term is bounded on $[v^*, 1-v^*]$, uniformly over γ. Since Θ is compact, $\ell_T^*(\gamma)$ thus satisfies a uniform weak law of large numbers, e.g. Andrews (1992, Theorem 4). □

Proof of Lemma A. The remainder term equals

$$\begin{aligned} \xi_T &:= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \left(s(R_{t+1}, q_t, \gamma_0, \widehat{g}) - s(R_{t+1}, q_t, \gamma_0, g_0)\right) \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \Big[1_{t+1}^{m} \left(S(R_{t+1}, q_t, \widehat{g}) - S(R_{t+1}, q_t, g_0)\right) \\ &\quad + 1_{t+1}^{l} \left(S(R_{t,l}^*, q_t, \widehat{G}) - S(R_{t,l}^*, q_t, G_0)\right) + 1_{t+1}^{u} \left(S(R_{t,u}^*, q_t, 1-\widehat{G}) - S(R_{t,u}^*, q_t, 1-G_0)\right)\Big]. \end{aligned}$$

First, we show that for all g with $\|g_{\gamma_0} - g_{0,\gamma_0}\|_{\infty,v^*}$ small enough and (r,q) such that $R_l^*(q,v^*) \leq r \leq R_u^*(q,v^*)$,

$$\|S(r,q,g) - S(r,q,g_0) - D_0^m(r,q,g-g_0)\| \leq b(r,q)\|g-g_0\|^2, \tag{A.8}$$

where $\mathbb{E}\left(1_{t+1}^{m} b_j(R_{t+1}, q_t)\right) < \infty$ for $j = 1, \ldots, k$. Here $D_0^m(r,q,g) := D_S(r,q,g_0)[g]$, with D_S

the pathwise derivative of S in g, defined for any direction $\bar{g}$ as

$$\begin{aligned} D_S(r,q,g)[\bar{g}] &:= \lim_{\tau\to 0}\frac{1}{\tau}\left(S(r,q,g+\tau\bar{g})-S(r,q,g)\right) \\ &= \frac{\partial}{\partial\tau}\frac{\dot{g}_{\gamma_0}(U_0)+\tau\dot{\bar{g}}_{\gamma_0}(U_0)+\left(g'_{\gamma_0}(U_0)+\tau\bar{g}'_{\gamma_0}(U_0)\right)\dot{U}_0}{g_{\gamma_0}(U_0)+\tau\bar{g}_{\gamma_0}(U_0)}\Big|_{\tau=0} \\ &= \frac{1}{g_{\gamma_0}(U)}\dot{\bar{g}}_{\gamma_0}(U)+\frac{\dot{U}}{g_{\gamma_0}(U)}\bar{g}'_{\gamma_0}(U)-\frac{\dot{g}_{\gamma_0}(U)+g'_{\gamma_0}(U)\dot{U}}{g_{\gamma_0}(U)^2}\bar{g}_{\gamma_0}(U), \end{aligned}$$

which is linear in $\bar{g}$.

Using the relation $\frac{\tilde{a}}{\tilde{b}}-\frac{a}{b}=\frac{1}{b}\left(1-\frac{1}{\tilde{b}}(\tilde{b}-b)\right)\left(\tilde{a}-a-\frac{a}{b}(\tilde{b}-b)\right)$, the remainder of the linearization of $\frac{\tilde{a}}{\tilde{b}}$ around $\frac{a}{b}$ takes the form $\frac{1}{b\tilde{b}}\left(\tilde{b}-b\right)\left(\tilde{a}-a-\frac{a}{b}(\tilde{b}-b)\right)$. Therefore,

$$\begin{aligned} &\|S(r,q,g)-S(r,q,g_0)-D_0(r,q,g-g_0)\| \\ &= \left\|\frac{g_{\gamma_0}(U_0)-g_{0,\gamma_0}(U_0)}{g_{\gamma_0}(U_0)g_{0,\gamma_0}(U_0)}\left(\dot{g}_{\gamma_0}(U_0)-\dot{g}_{0,\gamma_0}(U_0)+(g'_{\gamma_0}(U_0)-g'_{0,\gamma_0}(U_0))\dot{U}_0-S(R,q,g_0)\left(g_{\gamma_0}(U_0)-g_{0,\gamma_0}(U_0)\right)\right)\right\| \\ &\le \frac{2+\|S(R,q,g_0)\|+\|\dot{U}_0\|}{g_{\gamma_0}(U_0)g_{0,\gamma_0}(U_0)}\left(\left(g_{\gamma_0}(U_0)-g_{0,\gamma_0}(U_0)\right)^2+\|\dot{g}_{\gamma_0}(U_0)-\dot{g}_{0,\gamma_0}(U_0)\|^2+\left(g'_{\gamma_0}(U_0)-g'_{0,\gamma_0}(U_0)\right)^2\right). \end{aligned}$$

As a result, (A.8) holds with $b(r,q):=\frac{2+\|S(r,q,g_0)\|+\|\dot{U}_0\|}{\delta c_g g_{0,\gamma_0}(U_0)}\le\frac{2+\|\dot{g}_{0,\gamma_0}(U_0)+g'_{0,\gamma_0}(U_0)\dot{U}_0\|/c_g+\|\dot{U}_0\|}{\delta c_g^2}$, with δ and c_g as in the proof of Proposition C. The conditions in A *(i)* imply that g_{γ_0}, g'_{γ_0}, and $\dot{g}_{\gamma_0}$ are continuous and g_{γ_0} is positive on $(0,1)$ by Lemma D. Therefore $\dot{g}_{0,\gamma_0}$ and g'_{0,γ_0} are bounded on $[v^*,1-v^*]$, so that the integrability of $b(r,q)$ follows from that of $\dot{U}_0$.

Similar bounds on the scores of the trimming terms can be obtained by replacing g by G in the above linearization. In particular, for all G with $\|G-G_0\|$ small enough,

$$\|S(R_l^*(q,v^*),q,G)-S(R_l^*(q,v^*),q,G_0)-D_0^l(q,G-G_0)\|\le B^l(q)\|G-G_0\|^2, \tag{A.9}$$

where $D_0^l(q,\bar{G}):=D_S(R_l^*(q,v^*),q,G_0)[\bar{G}]$, $B^l(q):=\frac{1}{\delta c_G^l}\left(2+\|\dot{G}_{0,\gamma_0}(U_0^l)\|+|1+g_{0,\gamma_0}(U_0^l)|\|\dot{U}_0^l\|\right)$, with $U_0^i:=U^i(\gamma_0)$ where $U^i(\gamma):=U(R_i^*(q,v^*);q,\gamma)$ and $\dot{U}_0^i$ defined analogously for $i\in\{l,u\}$. The bound for the right tail follows symmetrically by replacing G with $1-G$ and superscript l with u. Since U_0^l and U_0^u are in $[v^*,1-v^*]$ a.s., the integrability of $B^l(q_t)$ and $B^u(q_t)$ follows from that of $\dot{U}_0$.

Define the linear score approximator

$$D_0(R_{t+1},q_t,\bar{g})=1_{t+1}^m D_0^m(R_{t+1},q_t,\bar{g})+1_{t+1}^l D_0^l(q_t,\bar{G})+1_{t+1}^u D_0^u(q_t,\bar{G}).$$

Then, for $j=1,\dots,k$ w.p.a. 1

$$\begin{aligned} \|\xi_{Tj}\| \le{}& \left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T D_{0,j}(R_{t+1},q_t,\widehat{g}-g_0)\right\|+\left|\frac{1}{T}\sum_{t=1}^T 1_{t+1}^m b_j(R_{t+1},q_t)\right|\sqrt{T}\|\widehat{g}-g_0\|^2 \\ &+\left|\frac{1}{T}\sum_{t=1}^T\left(1_{t+1}^l B_j^l(q_t)+1_{t+1}^u B_j^u(q_t)\right)\right|\sqrt{T}\|\widehat{G}-G_0\|^2. \end{aligned}$$

By Assumption A*(ii)*, $\|\widehat{g} - g_0\|$ and $\|\widehat{G} - G_0\| \leq \|\widehat{g}_{\gamma_0} - g_{0,\gamma_0}\|_{\infty,v^*} + |\widehat{G}_{\gamma_0}(v^*) - G_{0,\gamma_0}(v^*)| + \|\dot{\widehat{G}}_{\gamma_0}(v^*) - \dot{G}_{0,\gamma_0}(v^*)\|$ are $o_p(T^{-\frac{1}{4}})$. By Assumption C*(iv)* and the continuity of $u'(\cdot;\gamma_0)$ and g_{γ_0}, $q_t(r)$ is a measurable function of (R,S). The ergodic theorem and Assumption A*(iii)* therefore imply that $\xi_T = o_p(1)$. □

Proof of Proposition SE. First, we conjecture and verify that $\int D_0(r,q,\widehat{g}-g_0)\,\mathrm{d}F(r,q) = 0$, where F is the joint physical distribution of (R_{t+1}, q_t). The score approximator is equal to

$$D_0(r,q,\bar{g}) = \begin{cases} \frac{\dot{\bar{G}}_{\gamma_0}(U_0^l)}{G_{0,\gamma_0}(U_0^l)} + \frac{\partial}{\partial\gamma}\frac{\bar{G}_{\gamma_0}(U^l(\gamma))}{G_{0,\gamma}(U^l(\gamma))}\Big|_{\gamma=\gamma_0} & \text{for } r < R_l^*(q,v^*); \\ \frac{\dot{\bar{g}}_{\gamma_0}(U_0)}{g_{0,\gamma_0}(U_0)} + \frac{\partial}{\partial\gamma}\frac{\bar{g}_{\gamma_0}(U(\gamma))}{g_{0,\gamma}(U(\gamma))}\Big|_{\gamma=\gamma_0} & \text{for } R_l^*(q,v^*) \leq r \leq R_u^*(q,v^*); \\ \frac{-\dot{\bar{G}}_{\gamma_0}(U_0^u)}{1-G_{0,\gamma_0}(U_0^u)} + \frac{\partial}{\partial\gamma}\frac{1-\bar{G}_{\gamma_0}(U^u(\gamma))}{1-G_{0,\gamma}(U^u(\gamma))}\Big|_{\gamma=\gamma_0} & \text{for } r > R_u^*(q,v^*), \end{cases}$$

where $U(\gamma) := U(r;q,\gamma)$. For $\bar{g} = \widehat{g} - g_0$, integrating the lower, middle, and upper parts with respect to the probability distribution cancels out the terms $\frac{1}{G_0}$, $\frac{1}{g_0}$ and $\frac{1}{1-G_0}$, respectively, to yield

$$\begin{aligned} \int D_0(r,q,\widehat{g}-g_0)\,\mathrm{d}F(r,q) &= \frac{1}{T}\sum_{t=1}^{T}\dot{U}_{0,t+1}\int\left(K_h'(U_{0,t+1}-v) - \dot{g}_0(v)\right)\mathrm{d}v \\ &\quad + \frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\gamma}\int\left(K_h(U_{t+1}(\gamma)-v) - g_0(v;\gamma)\right)\mathrm{d}v\Big|_{\gamma=\gamma_0} \\ &= \frac{1}{Th}\sum_{t=1}^{T}\dot{U}_{0,t+1}\int K'(z)\,\mathrm{d}z - \frac{\partial}{\partial\gamma}\int g_0(v;\gamma)\,\mathrm{d}v\Big|_{\gamma=\gamma_0} - 0 = 0, \end{aligned}$$

using that K and g_0 integrate to one.

Since D_0 is linear in $\bar{g}$, decompose $D_0(R_{t+1},q_t,\widehat{g}-g_0) = D_0(R_{t+1},q_t,\widehat{g}-\bar{g}_h) + D_0(R_{t+1},q_t,\bar{g}_h-g_0)$, where $\bar{g}_h(v;\gamma) := \mathbb{E}\left(\widehat{g}(v;\gamma)\right)$. The second component is a small bias term whose scaled and centered time average

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T}D_0(R_{t+1},q_t,\bar{g}_h-g_0) - \int D_0(r,q,\bar{g}_h-g_0)\,\mathrm{d}F(r,q)\right) = o_p(1)$$

by Chebyshev's inequality, as $\mathbb{V}\mathrm{ar}\left(D_0(R_{t+1},q_t,\bar{g}_h-g_0)\right) = O\left(\|\bar{g}_h-g_0\|^2\right) = O(h^2)$ for second or higher order kernels. For the first component, define $K_{h,s+1}(u,\gamma) := K_h(U_{s+1}(\gamma) - u)$, $d_T(R_{t+1},q_t,R_{s+1},q_s) := D_0(R_{t+1},q_t,K_{h,s+1})$, and the related marginals $d_{T1}(R_{t+1},q_t) := \int d_T(R_{t+1},q_t,r,q)\,\mathrm{d}F(r,q)$ and $d_{T2,s}(R_{s+1},q_s) := \int d_T(r,q,R_{s+1},q_s)\,\mathrm{d}F(r,q)$. Define also the shorthand $d_{T,ts} := d_T(R_{t+1},q_t,R_{s+1},q_s)$, $d_{T1,t} := d_{T1}(R_{t+1},q_t)$, and $d_{T2,s} := d_{T2}(R_{s+1},q_s)$ for

$t, s = 1, \ldots, T$. This allows expressing

$$\begin{aligned}
&\sqrt{T}\left(\frac{1}{T}\sum_{t=1}^{T} D_0(R_{t+1}, q_t, \widehat{g} - \bar{g}_h) - \int D_0(r, q, \widehat{g} - \bar{g}_h)\,\mathrm{d}F(r, q)\right) \\
&\quad = \sqrt{T}\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} d_{T,ts} - \frac{1}{T}\sum_{t=1}^{T} d_{T1,t} - \frac{1}{T}\sum_{t=1}^{T} d_{T2,t} + \mathbb{E}\left(d_{T1,t}\right)\right) \\
&\quad = \sqrt{T}\left(\frac{1}{T^2}\sum_{t=1}^{T} \tilde{d}_{T,t} + \frac{T-1}{T} U_T\right)
\end{aligned}$$

in terms of the second-order, centered U-statistic with symmetric kernel $u_T(R_{t+1}, q_t, R_{s+1}, q_s)$, or $u_{T,ts}$ for short:

$$U_T = \frac{1}{T(T-1)}\sum_{t=1}^{T}\sum_{s<t}\left(u_{T,ts} - u_{T1,t} - u_{T1,s} + \mathbb{E}\left(u_{T1,t}\right)\right),$$

where $u_{T,ts} = d_{T,ts} + d_{T,st}$, $u_{T1,t} = d_{T1,t} + d_{T2,t}$, and $\tilde{d}_{T,t} = d_{T,tt} - u_{T1,t} + \mathbb{E}\left(d_{T1,t}\right)$.

First, we analyze these terms for $R^*_{t,l} \leq R_{t+1} \leq R^*_{t,u}$, in which case

$$\begin{aligned}
d_T(R_{t+1}, q_t, R_{s+1}, q_s) &= \frac{1}{g_{\gamma_0}(U_{0,t+1})}\left(\dot{U}_{0,s+1} - \dot{U}_{0,t+1}\right) K_h'(U_{0,s+1} - U_{0,t+1}) \\
&\quad - \frac{\dot{g}_{\gamma_0}(U_{0,t+1}) + g'_{\gamma_0}(U_{0,t+1})\dot{U}_{0,t+1}}{g_{\gamma_0}(U_{0,t+1})^2} K_h(U_{0,s+1} - U_{0,t+1}) \\
&= \frac{1}{g_{\gamma_0}(U_{0,t+1})}\frac{\partial}{\partial\gamma} K_h\left(U_{t+1}(\gamma) - U_{s+1}(\gamma)\right)\Big|_{\gamma=\gamma_0} \\
&\quad - \frac{\frac{\partial}{\partial\gamma} g_\gamma(U_{t+1}(\gamma))|_{\gamma=\gamma_0}}{g_{\gamma_0}(U_{0,t+1})^2} K_h(U_{0,s+1} - U_{0,t+1}) \\
&= \frac{\partial}{\partial\gamma}\frac{K_h\left(U_{t+1}(\gamma) - U_{s+1}(\gamma)\right)}{g_\gamma(U_{t+1}(\gamma))}\Big|_{\gamma=\gamma_0}.
\end{aligned}$$

This allows computing the following for $1^m_{t+1} = 1$:

$$\begin{aligned}
d_T(R_{t+1}, q_t, R_{t+1}, q_t) &= -\frac{\left(\dot{g}_{\gamma_0}(U_{0,t+1}) + g'_{\gamma_0}(U_{0,t+1})\dot{U}_{0,t+1}\right)}{g_{\gamma_0}(U_{0,t+1})^2} h^{-1} K(0) = O_p\left(h^{-1}\right), \\
d_{T1}(R_{t+1}, q_t) &= \int \frac{\partial}{\partial\gamma}\frac{K_h\left(U_{t+1}(\gamma) - U(\gamma)\right)}{g_\gamma(U_{t+1}(\gamma))}\Big|_{\gamma=\gamma_0}\mathrm{d}F(r, q) \\
&= \frac{\partial}{\partial\gamma}\frac{\int K(z) g_\gamma(U_{t+1}(\gamma) + hz)\,\mathrm{d}z}{g_\gamma(U_{t+1}(\gamma))}\Big|_{\gamma=\gamma_0} = O_p\left(h\right), \\
d_{T2}(R_{t+1}, q_t) &= \int \frac{\partial}{\partial\gamma}\frac{K_h\left(U_{t+1}(\gamma) - U(\gamma)\right)}{g_\gamma(U(\gamma))}\Big|_{\gamma=\gamma_0}\mathrm{d}F(r, q) \\
&= \frac{\partial}{\partial\gamma}\int K_h\left(U_{t+1}(\gamma) - u\right)\mathrm{d}u\Big|_{\gamma=\gamma_0} = 0.
\end{aligned}$$

Therefore $\mathbb{E}\left(h\|1^m_{t+1}\tilde{d}_{T,t}\|\right) < \infty$, so by the ergodic theorem $\frac{1}{T\sqrt{T}}\sum_{t=1}^{T} 1^m_{t+1}\tilde{d}_{T,t} = O_p\left(\frac{1}{\sqrt{T}h}\right)$.

Focusing on the U-statistic, suppose that for some $\delta > 0$, $M < \infty$, positive series a_T, and all T

$$\iint \|u_T(R_{t+1}, q_t, R_{s+1}, q_s)\|^{2+\delta} \, \mathrm{d}F(R_{t+1}, q_t) \, \mathrm{d}F(R_{s+1}, q_s) \leq M a_T, \tag{A.10}$$

and for all integers $t_1 < t_2$

$$\mathbb{E}\left(\|u_T(R_{t_2+1}, q_{t_2}, R_{t_1+1}, q_{t_1})\|^{2+\delta} \right) \leq M a_T. \tag{A.11}$$

By Assumptions A*(i)* and C*(iv)*, q_t is a measurable function of X_t so that measurable functions of (R_{t+1}, q_t) are mixing at the rate in Assumption SE*(i)*. Yoshihara (1976, Lemma 2) then implies that (A.10) and (A.11) are sufficient for $\mathbb{E}\left(\|U_T\|^2\right) = O\left(T^{-1-\eta} a_T^{\frac{2}{2+\delta}}\right)$ with $\eta = 2\frac{\delta - \delta'}{\delta'(2+\delta)} > 0$. Writing $u_{T,ts} = \sum_{i,j \in \{l,m,u\}} 1_{t+1}^i 1_{s+1}^j u_{T,ts}$, we can establish (A.10) and (A.11) for the middle region ($i = j = m$), the tails ($i = j \neq m$), and the cross-terms ($i \neq j$) separately.

To establish (A.10) on the middle region, use $\left|\frac{1}{g_{\gamma_0}(U_{0,t+1})} - \frac{1}{g_{\gamma_0}(U_{0,s+1})}\right| \leq C \left|U_{0,t+1} - U_{0,s+1}\right|$ to write

$$\begin{aligned} 1_{t+1}^m 1_{s+1}^m \|u_T(R_{t+1}, q_t, R_{s+1}, q_s)\| \leq C & \left\| (U_{0,t+1} - U_{0,s+1}) \left(\dot{U}_{0,s+1} - \dot{U}_{0,t+1} \right) K_h'\left(U_{0,s+1} - U_{0,t+1}\right) \right\| \\ & + \left(C_0 + C_1 \left(\|\dot{U}_{0,s+1}\| + \|\dot{U}_{0,t+1}\| \right) \right) \left| K_h\left(U_{0,s+1} - U_{0,t+1}\right) \right|, \end{aligned} \tag{A.12}$$

where C, C_0 and C_1 are some large, positive constants. For the first term in (A.12), note

$$\begin{aligned} & \iint 1_{t+1}^m 1_{s+1}^m \left\| (U_{0,t+1} - U_{0,s+1}) \left(\dot{U}_{0,s+1} - \dot{U}_{0,t+1} \right) K_h'\left(U_{0,s+1} - U_{0,t+1}\right) \right\|^{2+\delta} \mathrm{d}F(R_{t+1}, q_t) \, \mathrm{d}F(R_{s+1}, q_s) \\ & \leq 2C \int 1_{s+1}^m \int_{v^*}^{1-v^*} \left\| (U_{0,t+1} - U_{0,s+1}) \, \dot{U}_{0,s+1} K_h'\left(U_{0,s+1} - U_{0,t+1}\right) \right\|^{2+\delta} \mathrm{d}G_{\gamma_0}(U_{0,t+1}) \, \mathrm{d}F(R_{s+1}, q_s) \\ & \leq \frac{C}{h^{1+\delta}} \int \left| z K'(z) \right|^{2+\delta} \mathrm{d}z \max_{v^* \leq v \leq 1-v^*} \mathbb{E}\left(\|\dot{U}_0\|^{2+\delta} \mid U_0 = v \right) g_{\gamma_0}(v) = O\left(\frac{1}{h^{1+\delta}} \right), \end{aligned}$$

using Minkowski's inequality and stationarity in the first step, and continuity of the functions inside the maximum in the second. To establish (A.11) in the middle region, for the first term in (A.12) we derive for any $s > t$ that

$$\begin{aligned} & \mathbb{E}\left(1_t^m 1_s^m \left\| (U_{0,t} - U_{0,s}) \left(\dot{U}_{0,s} - \dot{U}_{0,t} \right) K_h'\left(U_{0,s} - U_{0,t}\right) \right\|^{2+\delta} \right) \\ & \leq \int_{v^*}^{1-v^*} \int_{v^*}^{1-v^*} \left| (U_{0,t} - U_{0,s}) \, K_h'\left(U_{0,s} - U_{0,t}\right) \right|^{2+\delta} \mathbb{E}\left(\|\dot{U}_{0,s} - \dot{U}_{0,t}\|^{2+\delta} \mid U_{0,t}, U_{0,s} \right) \mathrm{d}G_{\gamma_0}(U_{0,t}) \, \mathrm{d}G_{\gamma_0}(U_{0,s}) \\ & \leq \frac{C}{h^{1+\delta}} \int \left| z K'(z) \right|^{2+\delta} \mathrm{d}z \max_{v^* \leq v \leq 1-v^*} \mathbb{E}\left(\|\dot{U}_{0,s} - \dot{U}_{0,t}\|^{2+\delta} \mid U_{0,s} = U_{0,t} = v \right) + O\left(h^{-\delta} \right), \end{aligned}$$

which is $O\left(\frac{1}{h^{1+\delta}}\right)$ since the conditional moment inside the maximum is bounded on $[v^*, 1 - v^*]$ by SE*(ii)*. The integrals in (A.10) and (A.11) for the second term in (A.12) are also $O\left(\frac{1}{h^{1+\delta}}\right)$ by the moment condition on $\dot{U}_0$.

Meanwhile, when $1^l_{t+1} = 1$, with $\dot{U}^i_{0,t} := \dot{U}^i_t(\gamma_0)$ where $\dot{U}^i_t(\gamma) := \dot{U}_t(R^*_{t,i};\gamma)$ for $i \in \{l,u\}$,

$$\begin{aligned} d_T(R_{t+1},q_t,R_{s+1},q_s) &= \frac{1}{G_{0,\gamma_0}(U^l_{0,t})}\left(\dot{U}^l_{0,t} - \dot{U}_{0,s+1}\right) K_h\left(U^l_{0,t} - U_{0,s+1}\right) \\ &\quad - \frac{\dot{G}_{0,\gamma_0}(U^l_{0,t}) + g_{0,\gamma_0}(U^l_{0,t})\dot{U}^l_{0,t}}{G_{0,\gamma_0}(U^l_{0,t})^2} F_K\left(\frac{U^l_{0,t} - U_{0,s+1}}{h}\right) \\ &= \frac{\partial}{\partial\gamma}\frac{F_K\left(\frac{1}{h}\left(U^l_t(\gamma) - U_{s+1}(\gamma)\right)\right)}{G_{0,\gamma}(U^l_t(\gamma))}\Big|_{\gamma=\gamma_0}. \end{aligned}$$

Similar steps as for the middle region allow computing that $\mathbb{E}\left\|1^l_{t+1}d_{T,tt}\right\| = O(1)$, $\mathbb{E}\left\|1^l_{t+1}d_{T1,t}\right\| = O(h)$, and $\mathbb{E}\left\|1^l_{t+1}d_{T2,t}\right\| = O(1)$. The first of these is now $O(1)$ rather than $O(h^{-1})$ due to integrating the kernel function, while the second is now $O(h)$ rather than $O(h^2)$ as F_K is asymmetric. The term $d_{T2}(R_{t+1},q_t)$ is bounded since $U^l_0(q) \in [v^*, 1-v^*]$ for all q. Combining implies $\mathbb{E}\left(\|1^l_{t+1}\tilde{d}_{T,t}\|\right) < \infty$, so by the ergodic theorem $\frac{1}{T\sqrt{T}}\sum_{t=1}^T 1^l_{t+1}\tilde{d}_T(R_{t+1},q_t) = O_p\left(\frac{1}{\sqrt{T}}\right)$. Furthermore, $U^l_{0,t} \in [v^*, 1-v^*]$ a.s. implies

$$\begin{aligned} 1^l_{t+1}\|d_T(R_{t+1},q_t,R_{s+1},q_s)\| &\le C\left\|\dot{U}^l_{0,t} - \dot{U}_{0,s+1}\right\|\left|K_h\left(U^l_{0,t} - U_{0,s+1}\right)\right| \\ &\quad + \left(C_0 + C_1\|\dot{U}^l_{0,t}\|\right)\left|F_K\left(\frac{U^l_{0,t} - U_{0,s+1}}{h}\right)\right|. \end{aligned} \tag{A.13}$$

The first term in (A.13) is similar to the second term in (A.12). Using that

$$\mathbb{E}(\|\dot{U}^l_{0,t}\|^{2+\delta}) = \mathbb{E}\left(\mathbb{E}(\|\dot{U}_{0,t+1}\|^{2+\delta} \mid U_0 = U^l_{0,t})\right) < \infty,$$

its integrals in (A.10) and (A.11) are $O\left(\frac{1}{h^{1+\delta}}\right)$. The integrals over the second term in (A.13) are $O(1)$.

For cross-terms with one observation in the middle and the other in the lower tail, write

$$1^m_{t+1}1^l_{s+1}u_{T,ts} + 1^l_{t+1}1^m_{s+1}u_{u,ts} = (1^m_{t+1}1^l_{s+1}d_{ts,T} + 1^l_{t+1}1^m_{s+1}d_{st,T}) + (1^m_{t+1}1^l_{s+1}d_{st,T} + 1^l_{t+1}1^m_{s+1}d_{ts,T}).$$

The first bracketed term is ordered to satisfy the middle region bound (A.12), while the terms in the second bracket satisfy the tail bound (A.13). Therefore, their integrals in (A.10) and (A.11) are also $O\left(\frac{1}{h^{1+\delta}}\right)$. By symmetry, the same order is obtained when one or both observations are in the upper tail, i.e. when $1^u_{t+1} = 1$.

We conclude that (A.10) and (A.11) hold with $a_T = \frac{1}{h^{1+\delta}}$. As a result, we find $\sqrt{T}U_T = O_p\left(\left(T^{\frac{\delta-\delta'}{\delta'}}h^{1+\delta}\right)^{\frac{-1}{2+\delta}}\right)$, which is $o_p(1)$ under the bandwidth conditions in SE*(iv)*. □

Proof of Proposition N. This follows by combining Proposition C, Lemma A, Assumption N, and the mean-value expansion in (3.14). We show that $\frac{1}{\sqrt{T}}\sum_{t=1}^T s(R_{t+1},q_t,\gamma_0,g_0) \xrightarrow{d} N(0,V)$.

The score of each observation equals

$$
\begin{aligned}
&s(R_{t+1}, q_t, \gamma, g) \\
&= 1^m_{t+1} \frac{\partial}{\partial \gamma} \log f_t(R_{t+1}; \gamma, g_\gamma) + 1^l_{t+1} \frac{\partial}{\partial \gamma} \log G_\gamma(U^l_t(\gamma)) + 1^u_{t+1} \frac{\partial}{\partial \gamma} \log(1 - G_\gamma(U^u_t(\gamma))).
\end{aligned}
$$

Under specification (3.3), $\mathbb{E}_t s(R_{t+1}, q_t, \gamma_0, g_0) = 0$ for all t. Therefore, applying a central limit theorem for martingale difference sequences to the scores at the true parameter yields the asymptotic normality. Here the covariance matrix V exists due to the moment conditions in N*(i)* and the boundedness of $\frac{\dot{G}_{\gamma_0}(v)}{G_{\gamma_0}(v)}$ on the trimmed domain. Furthermore, the uniform convergence of the Jacobian to the family of matrices $M(\gamma) := \partial \mathbb{E}^{\mathbb{P}} \left(s(R_{t+1}, q_t, \gamma, g_0) \right) / \partial \gamma$ follows from the triangle inequality

$$
\begin{aligned}
&\sup_{\gamma \in \mathcal{N}_0} \left\| \frac{\partial^2}{\partial \gamma \partial \gamma^\top} \ell^*_T(\gamma) - M(\gamma) \right\| \\
&\leq \sup_{\gamma \in \mathcal{N}_0} \left\| \frac{\partial^2}{\partial \gamma \partial \gamma^\top} (\ell^*_T(\gamma) - \ell^*_T(\gamma, g_{0,\gamma})) \right\| + \sup_{\gamma \in \mathcal{N}_0} \left\| \frac{\partial^2}{\partial \gamma \partial \gamma^\top} \ell^*_T(\gamma, g_{0,\gamma}) - M(\gamma) \right\|.
\end{aligned} \tag{A.14}
$$

Define the norm $\|g\|_{\mathcal{N}_0, v^*, l} = \max_{i+|j| \leq l} \sup_{\gamma \in \mathcal{N}_0} \| \frac{\partial^{i+|j|}}{\partial v^i \partial \gamma^j} g \|_{\infty, v^*}$ for non-negative integers (i, l) and multi-index j with sum $|j|$. To control the first term in (A.14), we show there is some integrable $\tilde{b}$ such that $\sup_{\gamma \in \mathcal{N}_0} \| \frac{\partial}{\partial \gamma} s(R_{t+1}, q_t, \gamma, g) - \frac{\partial}{\partial \gamma} s(R_{t+1}, q_t, \gamma, g_0) \| \leq \tilde{b}(R_{t+1}, q_t) \|g - g_0\|_{\mathcal{N}_0, v^*, 2}$ for small enough $\|g - g_0\|_{\mathcal{N}_0, v^*, 2}$. The partial derivative of the score, using the twice continuous differentiability of $g_\gamma(v)$ by Lemma D and Assumption N*(ii)*, equals

$$
\begin{aligned}
\frac{\partial}{\partial \gamma} s(R_{t+1}, q_t, \gamma, g) = 1^m_{t+1} \left(\frac{\partial^2}{\partial \gamma \partial \gamma^\top} \log c_t(\gamma) - \frac{\partial^2}{\partial \gamma \partial \gamma^\top} \log u'(R_{t+1}; \gamma) + \frac{\partial^2}{\partial \gamma \partial \gamma^\top} \log g_\gamma(U_{t+1}(\gamma)) \right) \\
+ 1^l_{t+1} \frac{\partial^2}{\partial \gamma \partial \gamma^\top} \log G_\gamma(U^l_t(\gamma)) + 1^u_{t+1} \frac{\partial^2}{\partial \gamma \partial \gamma^\top} \log(1 - G_\gamma(U^u_t(\gamma))).
\end{aligned}
$$

For $R^*_l(q, v^*) \leq r \leq R^*_u(q, v^*)$, we have $\left\| \frac{\partial}{\partial \gamma} s(r, q, \gamma, g) - \frac{\partial}{\partial \gamma} s(r, q, \gamma, g_0) \right\| \leq (*) + (**)$, where

$$
\begin{aligned}
(*) &= \left\| \frac{\frac{\partial^2}{\partial \gamma \partial \gamma^\top} g_\gamma\left(U(\gamma)\right)}{g_\gamma\left(U(\gamma)\right)} - \frac{\frac{\partial^2}{\partial \gamma \partial \gamma^\top} g_{0,\gamma}\left(U(\gamma)\right)}{g_{0,\gamma}\left(U(\gamma)\right)} \right\| \\
&\leq \frac{1}{g_\gamma(U(\gamma))} \left\| \frac{\partial^2}{\partial \gamma \partial \gamma^\top} \left(g_\gamma\left(U(\gamma)\right) - g_{0,\gamma}\left(U(\gamma)\right) \right) \right\| \\
&\quad + \frac{\| \frac{\partial^2}{\partial \gamma \partial \gamma^\top} g_{0,\gamma}\left(U(\gamma)\right) \|}{g_\gamma(U(\gamma)) g_{0,\gamma}(U(\gamma))} \left| g_\gamma\left(U(\gamma)\right) - g_{0,\gamma}\left(U(\gamma)\right) \right| \\
&\leq \frac{1 + 2\|\dot{U}(\gamma)\| + \|\ddot{U}(\gamma)\|}{g_\gamma(U(\gamma))} \|g - g_0\|_{\mathcal{N}, v^*, 2} + \frac{\| \frac{\partial^2}{\partial \gamma \partial \gamma^\top} g_{0,\gamma}\left(U(\gamma)\right) \|}{g_\gamma(U(\gamma)) g_{0,\gamma}(U(\gamma))} \|g - g_0\|_{\mathcal{N}, v^*, 0},
\end{aligned}
$$

and

$$
\begin{aligned}
(**) &= \left\| \frac{\frac{\partial}{\partial\gamma} g_\gamma(U(\gamma)) \frac{\partial}{\partial\gamma} g_\gamma(U(\gamma))^\top}{g_\gamma(U(\gamma))^2} - \frac{\frac{\partial}{\partial\gamma} g_{0,\gamma}(U(\gamma)) \frac{\partial}{\partial\gamma} g_{0,\gamma}(U(\gamma))^\top}{g_{0,\gamma}(U(\gamma))^2} \right\| \\
&\leq \left\| \frac{\frac{\partial}{\partial\gamma} g_\gamma\left(U(\gamma)\right)}{g_\gamma\left(U(\gamma)\right)} - \frac{\frac{\partial}{\partial\gamma} g_{0,\gamma}\left(U(\gamma)\right)}{g_{0,\gamma}\left(U(\gamma)\right)} \right\| \cdot \left\| \frac{\frac{\partial}{\partial\gamma} g_\gamma\left(U(\gamma)\right)}{g_\gamma\left(U(\gamma)\right)} + \frac{\frac{\partial}{\partial\gamma} g_{0,\gamma}\left(U(\gamma)\right)}{g_{0,\gamma}\left(U(\gamma)\right)} \right\|^\top \\
&\leq \left(\frac{1 + \|\dot{U}(\gamma)\| + \frac{1}{g_\gamma(U(\gamma))} \|\dot{g}_\gamma(U(\gamma)) + g'_\gamma(U(\gamma))\dot{U}(\gamma)\|}{g_\gamma(U(\gamma))} \|g - g_0\|_{\mathcal{N}_0, v^*, 1} + o_P\left(\|g - g_0\|_{\mathcal{N}_0, v^*, 1}\right) \right) \\
&\quad \cdot 2 \sup_{\gamma \in \mathcal{N}_0} \frac{\|\dot{g}_\gamma(U(\gamma)) + g'_\gamma(U(\gamma))\dot{U}(\gamma)\|}{g_\gamma(U(\gamma))},
\end{aligned}
$$

using the same asymptotic linearization as in the proof of Lemma A. The first two derivatives of g in both arguments are continuous and therefore bounded on the trimmed domain, so that integrability of $\tilde{b}$ follows from that of $\dot{U}$ and $\ddot{U}$.

When $r < R_l^*(q, v^*)$ or $r > R_u^*(q, v^*)$, $\left\| \frac{\partial}{\partial\gamma} s(r, q, \gamma, g) - \frac{\partial}{\partial\gamma} s(r, q, \gamma, g_0) \right\|$ is bound using the same expansion in the order of $\|G - G_0\|_{\mathcal{N}_0, v^*, 2}$, which vanishes when $\|g - g_0\|_{\mathcal{N}_0, v^*, 2}$ and $\sup_{\gamma \in \mathcal{N}_0} |\frac{\partial^{|j|}}{\partial\gamma^j}(G_\gamma(v^*) - G_{0,\gamma}(v^*))|$ vanish for all multi-indices j with $|j| \leq 2$.

The second term in (A.14) vanishes by dominated convergence. In particular, the continuous differentiability assumptions ensure that $\frac{\partial}{\partial\gamma} s(R_{t+1}, q_t, \gamma, g_0)$ is continuous in γ at γ_0 a.s., while the uniform integrability assumptions assure that $\mathbb{E}\left(\sup_{\gamma \in \mathcal{N}_{\gamma_0}} \|\frac{\partial}{\partial\gamma} s(R_{t+1}, q_t, \gamma, g_0)\|\right) < \infty$. □

B Uniform Convergence of the Kernel Estimator

This subsection verifies the uniform convergence conditions imposed in Assumptions C, A, and N for the kernel estimators of g and G given by (3.6) and (3.13), respectively. Throughout, we make the following assumptions on the dynamic properties of (R, q) and the kernel K, where $\mu_k(K) = \int z^k K(z)\,\mathrm{d}z$:

Assumption K.

(i) *$f_t(r) = f(r|X_t)$ is a positive, continuous function of the stationary process (R_{t+1}, X_t) on $\mathbb{R}_{++} \times \mathcal{X}$;*

(ii) *(R_{t+1}, X_t) is strongly mixing with mixing coefficients $\alpha(j) \leq Aj^{-\beta}$ for some finite, positive A and β;*

(iii) *$Z'(P)$, $u'(r;\gamma)$, and $c_t(\gamma)$ are positive and continuous on $(0,1)$, $\mathbb{R}_{++} \times \Theta$, and Θ a.s., respectively;*

(iv) *The kernel K is Lipschitz continuous, has compact support, and $\mu_0(K) = 1$.*

Consider the family of time series processes $(U_{t+1}(\gamma))_{t=1}^T$ indexed by $\gamma \in \Theta \subset \mathbb{R}^k$. Let $g(u;\gamma)$ and $g_j(u, v;\gamma)$ denote the densities of $U_t(\gamma)$ and $(U_t(\gamma), U_{t+j}(\gamma))$, which do not depend on (t, T) due to the stationarity assumption. Assumptions K*(i)*-K*(iii)* imply that $g(u;\gamma)$ is positive and

continuous by Lemma D. Furthermore, they imply that $q_t(r)$ is a measurable function of the mixing process X_t, and is therefore itself strongly mixing with polynomially decaying coefficients. Moreover, this property carries over to measurable functions of q_t and R_{t+1}, such as $U_{t+1}(\gamma)$.

First, we establish sufficient conditions for $\sup_{\gamma\in\Theta} \|\widehat{g}_\gamma - g_\gamma\|_{\infty,v^*} \xrightarrow{p} 0$ by adapting a result from Kristensen (2009).

Lemma K-1. *Suppose Assumption K and the following conditions hold:*

(i) $Z'(P)$, $u'(r;\gamma)$, *and* $c_t(\gamma)$ *are continuously differentiable on* $(0,1)$, $\mathbb{R}_{++}\times\Theta$, *and* Θ *a.s., respectively;*

(ii) $\mathbb{E}\left(\|\dot{U}_t(\gamma)\|^{2+\delta}|U_t(\gamma)=u\right)g(u;\gamma)$ *is continuous on* $(0,1)\times\Theta$ *for some* $\delta>0$ *such that* $\beta > \frac{1+(1+\delta)(2+k)}{\delta}$;

(iii) For some $M\geq 0$, $\sup_{|t-s|\geq M}\mathbb{E}\left(|\dot{U}_t(\gamma)\dot{U}_s(\gamma)^\top|\middle|U_t(\gamma)=u,U_s(\gamma)=v\right)g_{t-s}(u,v;\gamma)$ *is continuous on* $(0,1)^2\times\Theta$;

(iv) The bandwidth satisfies $h\to 0$ *and* $\log T/(T^\theta h)\to 0$ *with*

$$\theta = \frac{\beta-2-k-(1+\beta)/(1+\delta)}{\beta+2-k-(1+\beta)/(1+\delta)}.$$

Then $\sup_{\gamma\in\Theta}\|\widehat{g}_\gamma - g_\gamma\|_{\infty,v^*}\xrightarrow{p}0$.

Proof. Assumptions *(i)*-*(iii)* are sufficient for A.1-A.4 and A.6.1 in Kristensen (2009, Theorem 1(i)) with $X_{t,T}(\gamma)=U_t(\gamma)$, $c_T=1$, and $d_T=d$ for some large $d>0$. In particular, *(i)* implies that g' is continuous and the derivative $\dot{U}_t(\gamma)$ exists. Condition *(ii)* implies that $\mathbb{E}(1^m_{t+1}\|\dot{U}_{t+1}(\gamma)\|^{2+\delta})<\infty$ by the law of iterated expectations. Continuity of the function in *(iii)* guarantees its boundedness on $[v^*,1-v^*]^2$. The theorem yields that $\sup_{\gamma\in\Theta}\|\widehat{g}_\gamma-\mathbb{E}\widehat{g}_\gamma\|_{\infty,v^*} = O_p(\sqrt{\log T/(Th)}) = o_p(1)$ under bandwidth condition *(iv)*. Furthermore, $\sup_{\gamma\in\Theta}\|\mathbb{E}\widehat{g}_\gamma - g_\gamma\|_{\infty,v^*} = O(h)$ by a Taylor expansion and the continuity of g'. The conclusion follows from the triangle inequality. □

Next we verify that $\sup_{\gamma\in\Theta}|\widehat{G}_\gamma(v^*)-G_\gamma(v^*)|\xrightarrow{p}0$ for the kernel CDF estimator (3.13). Its uniform convergence would follow directly from Lemma K-1, provided Assumptions *(i)* and *(iii)* hold on the closed interval $[0,1]$. However, this rules out asymptotes in g_γ, g'_γ and $\dot{G}_\gamma$ when v goes to 0 or 1 for any γ. Instead, the following Lemma requires some additional continuity and moment conditions:

Lemma K-1*. *The following conditions, in addition to those in Lemma K-1, imply that* $\sup_{\gamma\in\Theta}|\widehat{G}_\gamma(v^*)-G_\gamma(v^*)|\xrightarrow{p}0$ *for the kernel CDF estimator:*

(i) $u'(r;\gamma)$ *and* $c_t(\gamma)$ *are twice continuously differentiable in* γ *on* $\mathbb{R}_{++}\times\Theta$ *and* Θ *a.s., respectively;*

(ii) For some $\delta>0$ *and all* $\gamma\in\Theta$, $\mathbb{E}(\|\ddot{U}_t(\gamma)\|^{2+\delta})<\infty$ *and* $\mathbb{E}(\|\dot{U}_t(\gamma)\|^{4+2\delta})<\infty$;

(iii) The following quantities are finite and continuous in γ on Θ:

$$\tilde{B}_1^*(\gamma) = \mathbb{E}\left(\|W_t(\gamma)\| \mid U_t(\gamma) = v^*\right)$$
$$\tilde{B}_2^*(\gamma) = \sup_{|t-s|\geq M} \mathbb{E}\left(\|W_t(\gamma)W_s(\gamma)^\top\| \middle| U_t(\gamma) = v^*, U_s(\gamma) = v^*\right) g_{t-s}(v^*, v^*; \gamma),$$

for $W_t(\gamma) \in \{\frac{\partial^{1+|j|}}{\partial\gamma^{1+j}}U_t(\gamma), \dot{U}_t(\gamma)\frac{\partial^{|j|}}{\partial\gamma^j}U_t(\gamma)\}$ for multi-indices $|j| \leq 1$ and some $M > 0$.

Proof. The derivative of $\widehat{G}_\gamma(v^*)$ w.r.t. γ is the kernel-weighted average

$$\dot{\widehat{G}}_\gamma(v^*) = -\frac{1}{T}\sum_{t=1}^{T} \dot{U}_{t+1}(\gamma)K_h(v^* - U_{t+1}(\gamma)).$$

Conditions *(i)*-*(iii)* imply those required in Kristensen (2009, Thm. 1(i)) for the uniform convergence $\sup_{\gamma\in\Theta}\|\dot{\widehat{G}}_\gamma(v^*) - \mathbb{E}\dot{\widehat{G}}_\gamma(v^*)\| \xrightarrow{p} 0$. In particular, *(i)* ensures that $U_t(\gamma)$ is a.s. twice differentiable. Meanwhile,

$$\sup_{\gamma\in\Theta}\|\mathbb{E}\dot{\widehat{G}}_\gamma(v^*)\| = \sup_{\gamma\in\Theta}\left\|\mathbb{E}\left(\dot{U}_{t+1}(\gamma)K_h(v^* - U_{t+1}(\gamma))\right)\right\| = \sup_{\gamma\in\Theta}\|\dot{G}_\gamma(v^*)\| + O(h),$$

where the last equation uses that $\dot{g}(v;\gamma)$ is continuous in a neighborhood of v^* for all $\gamma \in \Theta$ under K-1*(i)*. The mean-value theorem thus establishes stochastic equicontinuity of $\widehat{G}(v^*;\gamma)$, whose pointwise convergence follows from the boundedness of F_K and the ergodic theorem. Therefore $\sup_{\gamma\in\Theta}|\widehat{G}_\gamma(v^*) - \mathbb{E}\widehat{G}_\gamma(v^*)| \xrightarrow{p} 0$ by a uniform law of large numbers such as (Andrews, 1992, Thm. 3). A Taylor expansion implies that $\sup_{\gamma\in\Theta}|\mathbb{E}\widehat{G}_\gamma(v^*) - G_\gamma(v^*)| = O(h)$. The conclusion follows from the triangle inequality. □

Next, we provide primitive conditions for the kernel estimator to satisfy the uniform convergence rate assumptions A*(ii)*. The rate conditions on the density estimator and its first derivatives follow from existing results for kernel estimators, such as Andrews (1995). In particular, the following Lemma provides sufficient conditions.

Lemma K-2. *Suppose Assumption K and the following conditions hold, for some $\omega \geq 4$:*

(i) $Z'(P)$ is ω times continuously differentiable on $(0,1)$, $u'(r;\gamma_0)$ and $\dot{u}'(r;\gamma_0)$ are ω and $\omega-1$ times continuously differentiable on $\mathbb{R}_{++}$, $f_t(r)$ is $\omega-1$ times continuously differentiable a.s. on $\mathbb{R}_{++}$, and $c_t(\gamma)$ is continuously differentiable a.s. at γ_0;

(ii) $\mathbb{E}\|\dot{U}_0\|^{2+\delta} < \infty$ for some $\delta > 0$ such that $\beta > \frac{2+\delta}{\delta}$;

(iii) $\mu_k(K) = 0$ for $k = 1, \ldots, \omega-2$;

(iv) $h = O(T^{-\psi})$ and $h^{-1} = O(T^{\psi})$ for some $\psi > 0$ that satisfies $\frac{1}{4(\omega-1)} < \psi < \frac{1}{8}$.

Then $\sqrt{T}\|\widehat{g} - g_0\|^2 \xrightarrow{p} 0$.

Proof. Condition *(i)* implies that g_{γ_0} and $\dot{g}_{\gamma_0}$ are continuously differentiable on $(0,1)$ up to orders $\omega \geq 4$ and $\omega - 1$, respectively, by Lemma D. Condition *(ii)* establishes that $(\dot{U}_0, U_0)$ are

strongly mixing with coefficients such that $\sum_{l=1}^{\infty} \alpha(l)^{\frac{\delta}{2+\delta}} < \infty$. The result follows from Andrews (1995, Lemma A-1) by verifying its assumptions NP1-NP5 with $(Y_t, X_t) = (1, U_0)$ with $\lambda = 0, 1$, and $(Y_t, X_t) = (\dot{U}_0, U_0)$ with $\lambda = 1$, to establish the uniform convergence of $\widehat{g}_{\gamma_0}$, $\widehat{g}'_{\gamma_0}$, and $\widehat{\dot{g}}_{\gamma_0}$, respectively, on $[v^*, 1 - v^*]$ for any $v^* > 0$. □

Lemma K-2*. *If the conditions in Lemma K-2 hold, and* $\mathbb{E}\left(\|\dot{U}_0\|^2 \mid U_0 = v\right)$ *exists and is continuously differentiable at* v^**, then* $\widehat{G}_{\gamma_0}(v^*) - G_{\gamma_0}(v^*) = o_p\left(T^{-1/4}\right)$ *and* $\widehat{\dot{G}}_{\gamma_0}(v^*) - \dot{G}_{\gamma_0}(v^*) = o_p\left(T^{-1/4}\right)$ *for the kernel CDF estimator.*

Proof. For the result on $\widehat{G}_{\gamma_0}(v^*)$, suppose that $h \leq v^*$, which happens for T large enough. Using partial integration, the expectation of the kernel CDF estimator can then be written as

$$\begin{aligned}
\mathbb{E}F_K\left(\frac{v^* - U_{0,t+1}}{h}\right) &= \int_0^1 F_K\left(\frac{v^* - v}{h}\right) g_{\gamma_0}(v)\,\mathrm{d}v \\
&= F_K(z)\, G_{\gamma_0}(v^* - hz)\Big|_{(v^*-1)/h}^{v^*/h} + \int_{-1}^{1} K(z)\, G_{\gamma_0}(v^* - hz)\,\mathrm{d}z \\
&= G_{\gamma_0}(v^*) + O(h^{\omega-1}),
\end{aligned}$$

using the higher order kernel Assumption K-2*(iii)*, and the $\omega+1$ times continuous differentiability of G_{γ_0} at v^*. Meanwhile, since the summands $F_K\left(\frac{v^*-U_{0,t+1}}{h}\right)$ inherit the i.i.d. property of $U_{0,t+1}$, the variance term equals $\mathbb{V}\mathrm{ar}\,\widehat{G}_{\gamma_0}(v^*) = \frac{1}{T}\mathbb{V}\mathrm{ar}\left(F_K\left(\frac{v^*-U_{0,t+1}}{h}\right)\right) \leq \frac{C}{T}$ since F_K is bounded. We conclude that $\widehat{G}_{\gamma_0}(v^*) - G_{\gamma_0}(v^*) = O(h^{\omega-1}) + O_p\left(T^{-1/2}\right) = o_p\left(T^{-1/4}\right)$ under the bandwidth condition K-2*(iv)*.

For the second result, differentiating G_γ w.r.t. γ yields the relation

$$\dot{G}_{\gamma_0}(v) = -\mathbb{E}(\dot{U}_0 \mid U_0 = v) g_{\gamma_0}(v).$$

With $\widehat{\dot{G}}_{\gamma_0}(v^*) = -\frac{1}{T}\sum_{t=1}^{T} \dot{U}_{0,t+1} K_h(v^* - U_{0,t+1})$, we can decompose $\widehat{\dot{G}}_{\gamma_0}(v^*) - \dot{G}_{\gamma_0}(v^*) = B_T + V_T + D_T$ into bias, variance, and density estimation terms, where the bias term

$$B_T = -\frac{1}{T}\sum_{t=1}^{T}\left(\mathbb{E}(\dot{U}_{0,t+1} \mid U_{0,t+1}) - \mathbb{E}(\dot{U}_{0,t+1} \mid U_{0,t+1} = v^*)\right) K_h(v^* - U_{0,t+1})$$

has i.i.d. summands with $\mathbb{E}B_T = O(h^{\omega-1})$ and $\mathbb{V}\mathrm{ar}\,B_T = O\left(\frac{1}{Th}\right)$, the variance term equals

$$V_T = -\frac{1}{T}\sum_{t=1}^{T}\left(\dot{U}_{0,t+1} - \mathbb{E}(\dot{U}_{0,t+1} \mid U_{0,t+1})\right) K_h(v^* - U_{0,t+1}),$$

and the density estimation term, which is $o_p\left(T^{-1/4}\right)$ by Lemma K-2, equals

$$D_T = -\mathbb{E}(\dot{U}_{0,t+1} \mid U_{0,t+1} = v^*)\left(\hat{g}_{\gamma_0}(v^*) - g_{\gamma_0}(v^*)\right).$$

The variance term has mean $\mathbb{E}V_T = 0$ and long-run covariance matrix $\mathbb{E}V_T V_T^\top = V_{1T} + V_{2T}$,

where the contemporaneous covariance is

$$V_{1T} := \frac{1}{T}\mathbb{E}\left(\dot{U}^c_{0,t+1}\dot{U}^{c\top}_{0,t+1}K_h^2\left(v^* - U_{0,t+1}\right)\right)$$
$$= \frac{g_{\gamma_0}(v^*)}{Th}\mathbb{E}(\dot{U}_{0,t+1}\dot{U}^{\top}_{0,t+1} \mid U_{0,t+1} = v^*)\int K^2(z)dz + o\left(\frac{1}{Th}\right),$$

with $\dot{U}^c_{0,t+1} := \dot{U}_{0,t+1} - \mathbb{E}(\dot{U}_{0,t+1} \mid U_{0,t+1})$, and the between-period covariances sum to

$$V_{2T} := \frac{1}{T^2}\sum_{l=1}^{T}\left(1 - \frac{l}{T}\right)\mathbb{E}\left(\dot{U}^c_{0,1}\dot{U}^{c\top}_{0,1+l}K_h\left(v^* - U_{0,1}\right)K_h\left(v^* - U_{0,1+l}\right)\right)$$
$$\leq C\frac{1}{T^2h^2}\left(\mathbb{E}||\dot{U}_0||^{2+\delta}\right)^{\frac{2}{2+\delta}}\sum_{l=1}^{T}\left(1 - \frac{l}{T}\right)\alpha(l)^{\frac{\delta}{2+\delta}},$$

which is $O\left(\frac{1}{T^2h^2}\right)$ by K-2*(ii)*. We conclude that $\widehat{\dot{G}}_{\gamma_0}(v^*) - \dot{G}_{\gamma_0}(v^*) = O_p\left(h^{\omega-1} + \frac{1}{\sqrt{Th}}\right) + o_p\left(T^{-1/4}\right) = o_p\left(T^{-1/4}\right)$ by K-2*(iv)*. □

Lemma K-3. *Suppose Assumption K and the following conditions hold for some neighborhood $\mathcal{N}_0$ of γ_0:*

(i) For some $\delta > 0$, all $\gamma \in \mathcal{N}_0$, and multi-indices j,

$$\mathbb{E}\left(\left|\frac{\partial^{|j|}}{\partial\gamma^j}U_t(\gamma)\right|^{2+\delta}\right) < \infty \text{ for } |j| \leq 3,$$
$$\mathbb{E}\left(\left|\frac{\partial^{|j|}}{\partial\gamma^j}U_t(\gamma)\right|^{2+\delta}\|\dot{U}_t(\gamma)\|^{2+\delta}\right) < \infty \text{ for } |j| \leq 2;$$

(ii) The mixing exponent β satisfies $\beta > \frac{1+(1+\delta)(2+k)}{\delta}$;

(iii) The following quantities are continuous on $(0,1)\times\mathcal{N}_0$:

$$\tilde{B}_1(\gamma, v) = \mathbb{E}\left(\|W_t(\gamma)\| \mid U_t(\gamma) = v\right),$$
$$\tilde{B}_2(\gamma, u, v) = \sup_{|t-s|\geq M}\mathbb{E}\left(\|W_t(\gamma)W_s(\gamma)^\top\|\middle|U_t(\gamma) = u, U_s(\gamma) = v\right)g_{t-s}(u, v; \gamma),$$

for $W_t(\gamma) \in \{\frac{\partial^{1+|j|}}{\partial\gamma^{1+j}}U_t(\gamma), \dot{U}_t(\gamma)\frac{\partial^{|j|}}{\partial\gamma^j}U_t(\gamma)\}$ for multi-indices $|j| \leq 2$ and some $M > 0$;

(iv) K is three times differentiable;

(v) $\frac{\log T}{Th^5} \to 0$ and $\frac{\log T}{T^\theta h} \to 0$ with $\theta = \frac{\beta-2-k-(1+\beta)/(1+\delta)}{\beta+2-k-(1+\beta)/(1+\delta)}$;

(vi) $Z'(P)$, $u'(r;\gamma)$, and $c_t(\gamma)$ are three times, and $f_t(r)$ two times, continuously differentiable on $(0,1)$, $\mathbb{R}_{++}\times\mathcal{N}_0$, $\mathcal{N}_0$ a.s., and $\mathbb{R}_{++}$ a.s., respectively.

Then $\sup_{\gamma\in\mathcal{N}_0}\|\frac{\partial^{i+|j|}}{\partial v^i\partial\gamma^j}(\widehat{g}_\gamma - g_\gamma)\|_{\infty,v^*} \xrightarrow{p} 0$ for any non-negative integer i and multi-index j with $i + |j| \leq 2$.

Proof. Condition *(vi)* implies that $g(v;\gamma)$ is three times continuously differentiable on $(0,1)\times\mathcal{N}_0$ by Lemma D, and that $U_t(\gamma)$ is a.s. three times differentiable on $\mathcal{N}_0$. The result follows from Kristensen (2009, Thm. 1(i)) by verifying its assumptions A.1-A.4 and A.6.1 with $(Y_t(\gamma), X_t(\gamma)) = (\frac{\partial^{|j|}}{\partial\gamma^j}U_t(\gamma), U_t(\gamma))$ and $K_i(z) = \frac{\partial^i}{\partial^i z}K(z)$ for each (i,j) with $i+|j|\leq 2$, while setting $c_T = 1$ and $d_T = \gamma_0 + \epsilon$ for some $\epsilon > 0$. It establishes that

$$\sup_{\gamma\in\mathcal{N}_0}\left\|\frac{\partial^{i+|j|}}{\partial v^i\partial\gamma^j}(\widehat{g}_\gamma - \mathbb{E}\widehat{g}_\gamma)\right\|_{\infty,v^*} = O_p\left(\sqrt{\frac{\log T}{Th^{1+2(i+|j|)}}}\right),$$

which is $o_p(1)$ by Assumption *(v)*. The conclusion then follows from $\sup_{\gamma\in\mathcal{N}_0}\|\frac{\partial^{i+|j|}}{\partial v^i\partial\gamma^j}(\mathbb{E}\widehat{g}_\gamma - g_\gamma)\|_{\infty,v^*} = O(h^{\omega-1-i})$ by a Taylor expansion of order $\omega-1$, and the triangle inequality. □

Lemma K-3*. *If the conditions in K-3 hold, then* $\sup_{\gamma\in\mathcal{N}_0}\left|\frac{\partial^{|j|}}{\partial\gamma^j}(\widehat{G}_\gamma(v^*) - G_\gamma(v^*))\right| \xrightarrow{p} 0$ *for* $|j|\leq 2$.

Proof. For $|j| = 0,1$ the result follows analogous to the proof of Lemma K-1*, whose conditions are satisfied by those in K-3 with $\mathcal{N}_0$ instead of Θ. For $|j| = 2$, write $\widehat{\ddot{G}}_\gamma(v^*) = \frac{1}{T}\sum_{t=1}^T \dot{U}_{t+1}(\gamma)\dot{U}_{t+1}^\top(\gamma)K_h'\left(v^* - U_{t+1}(\gamma)\right) - \frac{1}{T}\sum_{t=1}^T \ddot{U}_{t+1}(\gamma)K_h\left(v^* - U_{t+1}(\gamma)\right)$. Both components are kernel averages whose uniform convergence follows from Kristensen (2009, Thm. 1(i)) with $(Y_t(\gamma), X_t(\gamma)) = (\dot{U}_t(\gamma)\dot{U}_t^\top(\gamma), U_t(\gamma))$ and $(Y_t(\gamma), X_t(\gamma)) = (\ddot{U}_t(\gamma), U_t(\gamma))$ with kernels K' and K, respectively. The conclusion then follows from $\sup_{\gamma\in\mathcal{N}_0}|\frac{\partial^{|j|}}{\partial\gamma^j}(\mathbb{E}\widehat{G}_\gamma(v^*) - G_\gamma(v^*))| = O(h)$ by K-3*(vi)* and the triangle inequality. □

C Additional Simulation Results

C.1 Weekly observations

This appendix considers the same DGP as Section 4, but changes the observation scheme to match weekly data instead of monthly. Historic weekly data does not extend as far back as monthly data, so we consider a sample of 12 years, or $T = 624$. This is not a *ceteris paribus* sample size increase, as the shortened observation span affects the shape of the distributions.

Table 5: Profile likelihood performance

	$v^*=0$				$v^*=0.001$				$v^*=0.01$				$v^*=0.05$			
h	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE
0.15	-0.05	1.76	3.09	0.41	0.38	1.82	3.47	0.48	0.74	1.97	4.44	0.60	1.00	2.09	5.40	0.72
0.20	0.24	1.44	2.14	0.31	0.60	1.48	2.55	0.38	0.98	1.62	3.58	0.50	1.22	1.88	5.01	0.67
0.25	0.55	1.27	1.92	0.29	0.86	1.30	2.42	0.36	1.26	1.44	3.65	0.51	1.55	1.71	5.32	0.71
0.30	0.89	1.20	2.22	0.33	1.17	1.21	2.84	0.41	1.58	1.35	4.33	0.58	1.96	1.55	6.24	0.81
MLE	3.29	1.09	12.04	1.34	3.30	1.09	12.08	1.34	3.30	1.10	12.07	1.34	3.22	1.14	11.69	1.29

(a) Z_{TK}

	$v^*=0$				$v^*=0.001$				$v^*=0.01$				$v^*=0.05$			
h	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE
0.15	0.55	2.24	5.32	0.39	1.52	2.24	7.32	0.52	1.98	2.44	9.84	0.66	1.89	2.53	9.98	0.71
0.20	0.91	2.02	4.93	0.34	1.77	2.01	7.18	0.49	2.24	2.18	9.79	0.63	2.18	2.47	10.89	0.73
0.25	1.19	1.85	4.83	0.32	1.97	1.83	7.22	0.47	2.44	2.00	9.95	0.63	2.41	2.36	11.41	0.74
0.30	1.43	1.78	5.22	0.33	2.15	1.75	7.72	0.49	2.62	1.93	10.58	0.65	2.64	2.28	12.13	0.78
MLE	3.92	1.64	18.03	1.00	3.94	1.61	18.08	1.00	4.00	1.62	18.64	1.03	4.09	1.67	19.48	1.06

(b) Z_P

	$v^*=0$				$v^*=0.001$				$v^*=0.01$				$v^*=0.05$			
h	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE
0.15	-2.82	2.32	13.38	0.71	-1.49	2.33	7.68	0.46	-0.68	2.58	7.13	0.45	0.05	2.46	6.07	0.44
0.20	-2.76	2.11	12.07	0.62	-1.58	2.11	6.96	0.40	-0.70	2.31	5.85	0.36	-0.16	2.50	6.26	0.41
0.25	-2.71	1.91	10.98	0.55	-1.63	1.90	6.26	0.35	-0.72	2.11	4.97	0.31	-0.13	2.43	5.93	0.38
0.30	-2.64	1.82	10.30	0.51	-1.63	1.80	5.92	0.32	-0.71	2.03	4.63	0.28	0.05	2.39	5.73	0.36
MLE	0.01	1.66	2.77	0.14	0.04	1.65	2.73	0.14	0.06	1.68	2.83	0.14	0.04	1.72	2.96	0.13

(c) Z_0

Note: Subtables display the bias, standard deviation, and mean squared error (MSE) of the profile likelihood estimator $\hat{\gamma}$ over $N = 1000$ replications of samples of length $T = 300$ for various levels of trimming v^*, with bandwidth $h^* = h/2$ for the CDF estimator. Each subpanel represents a different true probability weighting function. Columns labeled IMSE display the integrated mean squared error of the nonparametric estimator $\widehat{Z}$, multiplied by 1,000. The bottom rows represent the maximum likelihood estimator of the expected utility model, which fixes the weighting function as the identity map.

Table 6: Testing the probability weighting function

		Z_{TK}			Z_P			Z_0		
	α (%)	10	5	1	10	5	1	10	5	1
Rejection rate (%)	$v^*=0$	100	100	100	49.4	37.9	18.3	9.9	5.6	1.3
	$v^*=0.001$	100	100	100	49.9	38.3	17.9	9.8	5.4	1.2
	$v^*=0.01$	100	100	100	51.9	39.6	19.5	8.8	4.8	1.3
	$v^*=0.05$	100	100	100	48.5	37.2	17.6	8.3	4.0	1.1

Note: This table displays the rejection rate of the Bai (2003) test at level α. The first two sets of three columns consider the power under two different alternatives, the last set considers a DGP where the null is true and display the size of the test. The rejection rates are based on $N = 1000$ simulations. Critical values are constructed by simulation.

C.2 Different CRRA parameter: Higher risk aversion

This appendix considers the same DGP as Section 4, but increases the CRRA parameter to $\gamma_0 = 10$ from its baseline value of $\gamma_0 = 2$. This is perhaps not an empirically realistic case; we study it primarily to investigate the robustness of our estimator to relatively extreme values of γ_0. The interval for the parameter under trimming is set to $\Theta = [0, 15]$, to maintain the same width relative to the baseline case.

Table 7: Profile likelihood performance

	$v^*=0$				$v^*=0.001$				$v^*=0.01$				$v^*=0.05$			
h	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE
0.15	-0.27	1.30	1.76	1.02	-0.18	1.27	1.64	0.95	0.01	1.45	2.10	1.15	-0.36	1.38	2.03	1.22
0.20	-0.16	1.23	1.54	0.94	-0.03	1.01	1.02	0.60	0.11	1.14	1.32	0.74	-0.02	1.23	1.51	0.90
0.25	0.03	1.12	1.25	0.76	0.14	0.88	0.80	0.45	0.24	1.00	1.07	0.59	0.27	1.14	1.37	0.80
0.30	0.28	0.95	0.98	0.54	0.36	0.84	0.83	0.44	0.42	0.94	1.05	0.57	0.49	1.03	1.31	0.76
MLE	1.92	0.76	4.26	1.88	1.92	0.76	4.26	1.88	1.93	0.76	4.30	1.89	2.00	0.80	4.65	2.02

(a) Z_{TK}

	$v^*=0$				$v^*=0.001$				$v^*=0.01$				$v^*=0.05$			
h	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE
0.15	-0.13	1.47	2.18	0.72	0.05	1.50	2.25	0.78	0.08	1.73	2.98	0.97	-0.87	1.38	2.66	1.08
0.20	0.02	1.33	1.76	0.54	0.17	1.34	1.82	0.60	0.13	1.49	2.25	0.72	-0.32	1.39	2.04	0.80
0.25	0.14	1.22	1.51	0.42	0.28	1.22	1.57	0.48	0.18	1.36	1.88	0.58	0.05	1.44	2.09	0.74
0.30	0.26	1.19	1.48	0.37	0.39	1.18	1.55	0.43	0.25	1.31	1.79	0.53	0.26	1.44	2.15	0.72
MLE	2.04	1.12	5.41	1.26	2.05	1.11	5.42	1.26	2.08	1.11	5.55	1.29	2.14	1.15	5.93	1.34

(b) Z_P

	$v^*=0$				$v^*=0.001$				$v^*=0.01$				$v^*=0.05$			
h	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE
0.15	-1.91	1.54	6.02	1.52	-1.54	1.59	4.91	1.31	-1.29	1.84	5.04	1.35	-1.33	1.37	3.64	1.20
0.20	-1.90	1.40	5.58	1.36	-1.57	1.44	4.53	1.17	-1.33	1.61	4.38	1.15	-1.08	1.45	3.26	1.00
0.25	-1.89	1.29	5.22	1.22	-1.58	1.31	4.22	1.04	-1.36	1.46	3.97	1.02	-0.89	1.55	3.19	0.91
0.30	-1.85	1.24	4.99	1.13	-1.57	1.26	4.05	0.96	-1.35	1.40	3.77	0.95	-0.75	1.59	3.09	0.86
MLE	0.07	1.18	1.40	0.28	0.08	1.18	1.39	0.28	0.08	1.18	1.40	0.28	0.06	1.21	1.47	0.27

(c) Z_0

Note: Subtables display the bias, standard deviation, and mean squared error (MSE) of the profile likelihood estimator $\hat{\gamma}$ over $N = 1000$ replications of samples of length $T = 300$ for various levels of trimming v^*, with bandwidth $h^* = h/2$ for the CDF estimator. Each subpanel represents a different true probability weighting function. Columns labeled IMSE display the integrated mean squared error of the nonparametric estimator $\widehat{Z}$, multiplied by 1,000. The bottom rows represent the maximum likelihood estimator of the expected utility model, which fixes the weighting function as the identity map.

Table 8: Testing the probability weighting function

		Z_{TK}			Z_P			Z_0		
	α (%)	10	5	1	10	5	1	10	5	1
Rejection rate (%)	$v^*=0$	100	100	100	23.9	14.6	4.1	9.9	5.4	0.8
	$v^*=0.001$	100	100	100	24.2	15.1	4.1	9.8	5.1	0.7
	$v^*=0.01$	100	100	100	25.8	15.8	4.3	9.2	4.2	0.8
	$v^*=0.05$	100	100	99.8	23.1	13.8	3.7	8.2	4.2	0.7

Note: This table displays the rejection rate of the Bai (2003) test at level α. The first two sets of three columns consider the power under two different alternatives, the last set considers a DGP where the null is true and display the size of the test. The rejection rates are based on $N = 1000$ simulations. Critical values are constructed by simulation.

C.3 Different $\mathbb{P}$-dynamics

This appendix considers a different DGP as Section 4, using instead the following $\mathbb{P}$-dynamics

$$\begin{aligned} \mathrm{d}\log F_t &= \left(-\tfrac{1}{2}V_t - \mu_J\lambda_t\right)\mathrm{d}t + \sqrt{V_t}\,\mathrm{d}W_{1,t} + J_t\,\mathrm{d}N_t, \\ \mathrm{d}V_t &= 8(0.015 - V_t)\,\mathrm{d}t + 0.3\sqrt{V_t}\,\mathrm{d}W_{2,t} + J_t^V \mathbb{1}_{\{J_t<0\}}\,\mathrm{d}N_t, \end{aligned} \tag{C.1}$$

where $W_{1,t}$ and $W_{2,t}$ are correlated with coefficient $\rho = -0.6$, and $\lambda_t = 30V_t$. Jump parameters, and utility and probability weighting functions are the same as in Section 4. This model does not satisfy Assumption I, as there is only one state variable. Nonetheless, the utility parameter does seem to be identified.

Table 9: Profile likelihood performance

	$v^* = 0$				$v^* = 0.001$				$v^* = 0.01$				$v^* = 0.05$			
h	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE
0.15	-0.15	1.64	2.70	0.83	0.12	1.69	2.86	0.89	0.30	1.87	3.58	1.09	0.66	1.85	3.87	1.25
0.20	0.04	1.30	1.70	0.53	0.27	1.34	1.86	0.58	0.51	1.48	2.46	0.76	0.80	1.68	3.46	1.09
0.25	0.26	1.13	1.35	0.42	0.47	1.16	1.57	0.49	0.76	1.29	2.25	0.69	1.11	1.54	3.61	1.11
0.30	0.50	1.07	1.39	0.42	0.71	1.09	1.69	0.52	1.03	1.21	2.53	0.76	1.50	1.37	4.15	1.25
MLE	2.27	0.94	6.01	1.58	2.27	0.94	6.03	1.58	2.26	0.94	6.00	1.57	2.20	0.99	5.80	1.50

(a) Z_{TK}

	$v^* = 0$				$v^* = 0.001$				$v^* = 0.01$				$v^* = 0.05$			
h	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE
0.15	0.20	2.00	4.02	0.68	0.79	1.99	4.58	0.79	1.24	2.27	6.67	1.08	1.24	2.00	5.53	1.08
0.20	0.39	1.81	3.41	0.54	0.92	1.79	4.06	0.66	1.40	1.99	5.92	0.94	1.38	2.08	6.21	1.08
0.25	0.54	1.64	2.98	0.44	1.03	1.62	3.70	0.57	1.52	1.80	5.54	0.86	1.56	2.09	6.80	1.12
0.30	0.69	1.58	2.96	0.42	1.15	1.56	3.75	0.56	1.65	1.72	5.68	0.86	1.80	2.04	7.39	1.19
MLE	2.57	1.42	8.62	1.20	2.59	1.42	8.70	1.21	2.61	1.43	8.84	1.22	2.66	1.48	9.25	1.25

(b) Z_P

	$v^* = 0$				$v^* = 0.001$				$v^* = 0.01$				$v^* = 0.05$			
h	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE	Bias	St.d.	MSE	IMSE
0.15	-2.14	2.12	9.06	1.23	-1.21	2.11	5.91	0.87	-0.49	2.42	6.11	0.94	0.29	1.97	3.98	0.82
0.20	-2.14	1.93	8.31	1.09	-1.29	1.93	5.40	0.76	-0.48	2.14	4.79	0.73	-0.03	2.12	4.51	0.77
0.25	-2.12	1.75	7.58	0.97	-1.34	1.75	4.88	0.66	-0.48	1.92	3.92	0.60	-0.06	2.20	4.83	0.76
0.30	-2.08	1.67	7.13	0.89	-1.35	1.68	4.64	0.60	-0.45	1.83	3.56	0.54	0.12	2.21	4.91	0.76
MLE	-0.01	1.55	2.39	0.28	0.01	1.54	2.38	0.28	0.02	1.56	2.44	0.27	0.01	1.60	2.56	0.27

(c) Z_0

Note: Subtables display the bias, standard deviation, and mean squared error (MSE) of the profile likelihood estimator $\hat{\gamma}$ over $N = 1000$ replications of samples of length $T = 300$ for various levels of trimming v^*, with bandwidth $h^* = h/2$ for the CDF estimator. Each subpanel represents a different true probability weighting function. Columns labeled IMSE display the integrated mean squared error of the nonparametric estimator $\widehat{Z}$, multiplied by 1,000. The bottom rows represent the maximum likelihood estimator of the expected utility model, which fixes the weighting function as the identity map.

Table 10: Testing the probability weighting function

		Z_{TK}			Z_P			Z_0		
	α (%)	10	5	1	10	5	1	10	5	1
Rejection rate (%)	$v^* = 0$	100	100	100	27.4	19.6	6.3	9.9	4.9	0.6
	$v^* = 0.001$	100	100	100	27.5	19.8	6.3	9.6	4.8	0.6
	$v^* = 0.01$	100	100	100	28.4	20.3	6.9	9.0	4.3	0.4
	$v^* = 0.05$	100	100	100	28.2	20.1	6.7	9.1	4.1	0.7

Note: This table displays the rejection rate of the Bai (2003) test at level α. The first two sets of three columns consider the power under two different alternatives, the last set considers a DGP where the null is true and display the size of the test. The rejection rates are based on $N = 1000$ simulations. Critical values are constructed by simulation.

References

Abdellaoui, M. (2000): "Parameter-free elicitation of utility and probability weighting functions," *Management Science*, 46, 1497–1512.

Abel, A. B. (1990): "Asset prices under habit formation and catching up with the Joneses," *American Economic Review*, 80, 38–42.

Ai, C. (1997): "A semiparametric maximum likelihood estimator," *Econometrica*, 65, 933–963.

Ai, H. (2005): "Smooth nonexpected utility without state independence," *Working paper, Federal Reserve Bank of Minneapolis*.

Aït-Sahalia, Y. and J. Duarte (2003): "Nonparametric option pricing under shape restrictions," *Journal of Econometrics*, 116, 9–47.

Aït-Sahalia, Y. and A. W. Lo (2000): "Nonparametric risk management and implied risk aversion," *Journal of Econometrics*, 94, 9–51.

Allais, M. (1953): "Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école Américaine," *Econometrica*, 21, 503–546.

Andersen, T. G., N. Fusari, and V. Todorov (2015): "The risk premia embedded in index options," *Journal of Financial Economics*, 117, 558–584.

——— (2017): "Short-term market risks implied by weekly options," *The Journal of Finance*, 72, 1335–1386.

Andrews, D. W. (1992): "Generic uniform convergence," *Econometric Theory*, 8, 241–257.

——— (1994): "Asymptotics for semiparametric econometric models via stochastic equicontinuity," *Econometrica*, 62, 43–72.

——— (1995): "Nonparametric kernel estimation for semiparametric models," *Econometric Theory*, 11, 560–586.

Baele, L., J. Driessen, S. Ebert, J. M. Londono, and O. G. Spalt (2019): "Cumulative prospect theory, option returns, and the variance premium," *The Review of Financial Studies*, 32, 3667–3723.

Bai, J. (2003): "Testing parametric conditional distributions of dynamic models," *Review of Economics and Statistics*, 85, 531–549.

Barberis, N. C. (2013): "Thirty years of prospect theory in economics: A review and assessment," *Journal of Economic Perspectives*, 27, 173–196.

Barseghyan, L., F. Molinari, T. O'Donoghue, and J. C. Teitelbaum (2013): "The nature of risk preferences: Evidence from insurance choices," *American Economic Review*, 103, 2499–2529.

Bates, D. S. (1996): "Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options," *The Review of Financial Studies*, 9, 69–107.

Bliss, R. R. and N. Panigirtzoglou (2004): "Option-implied risk aversion estimates," *The Journal of Finance*, 59, 407–446.

Boswijk, H. P., R. J. A. Laeven, and E. Vladimirov (2024): "Estimating option pricing

models using a characteristic function-based linear state space representation," *Journal of Econometrics*, 244, 105864.

Bruggen, P. v., R. J. A. Laeven, and G. van de Kuilen (2024): "Higher-order risk attitudes for non-expected utility," Working paper, Center for Economic Research, centER Discussion Paper Nr. 2024-019, https://research.tilburguniversity.edu/files/101140130/2024-019.pdf.

Bruhin, A., H. Fehr-Duda, and T. Epper (2010): "Risk and rationality: Uncovering heterogeneity in probability distortion," *Econometrica*, 78, 1375–1412.

Chabi-Yo, F., R. Garcia, and E. Renault (2008): "State dependence can explain the risk aversion puzzle," *The Review of Financial Studies*, 21, 973–1011.

Chen, T.-Y., Y.-L. Lin, and L. Y. Tzeng (2024): "Estimating probability weighting functions through option pricing bounds," *The Review of Asset Pricing Studies*, raae008.

Chen, X. and Y. Fan (2006): "Estimation of copula-based semiparametric time series models," *Journal of Econometrics*, 130, 307–335.

Chew, S., E. Karni, and Z. Safra (1987): "Risk aversion in the theory of expected utility with rank dependent probabilities," *Journal of Economic Theory*, 42, 370–381.

Cicchetti, C. J. and J. A. Dubin (1994): "A microeconometric analysis of risk aversion and the decision to self-insure," *Journal of Political Economy*, 102, 169–186.

Cuesdeanu, H. and J. C. Jackwerth (2018): "The pricing kernel puzzle: Survey and outlook," *Annals of Finance*, 14, 289–329.

Dalderop, J. (2020): "Nonparametric filtering of conditional state-price densities," *Journal of Econometrics*, 214, 295–325.

Dalderop, J. and O. B. Linton (2024): "Estimating a conditional density ratio model for asset returns and option demand," *Available at SSRN 4779268*.

Davidson, J. (1994): *Stochastic Limit Theory: An Introduction for Econometricians*, OUP Oxford.

Dierkes, M. (2013): "Probability weighting and asset prices," *Available at SSRN 2253817*.

Duffie, D., J. Pan, and K. Singleton (2000): "Transform analysis and asset pricing for affine jump-diffusions," *Econometrica*, 68, 1343–1376.

Eeckhoudt, L. R., C. Gollier, and H. Schlesinger (2005): *Economic and Financial Decisions under Risk*, Princeton: Princeton University Press.

Eeckhoudt, L. R. and R. J. A. Laeven (2022): "Dual moments and risk attitudes," *Operations Research*, 70, 1330–1341.

Efron, B. and R. J. Tibshirani (1993): *An Introduction to the Bootstrap*, Chapman & Hall, New York.

Gonzalez, R. and G. Wu (1999): "On the shape of the probability weighting function," *Cognitive Psychology*, 38, 129–166.

Gonçalves, S. and L. Kilian (2004): "Bootstrapping autoregressions with conditional het-

eroskedasticity of unknown form," *Journal of Econometrics*, 123, 89–120.

Hafner, C. M., H. Herwartz, and S. Wang (2024): "Statistical identification of independent shocks with kernel-based maximum likelihood estimation and an application to the global crude oil market," *Journal of Business & Economic Statistics*, 1–16.

Harrison, G. W. and J. T. Swarthout (2023): "Cumulative prospect theory in the laboratory: A reconsideration," in *Models of Risk Preferences: Descriptive and Normative Challenges*, ed. by G. W. Harrison and D. Ross, Leeds: Emerald Group Publishing Limited, vol. 22, 107–192.

Jullien, B. and B. Salanié (2000): "Estimating preferences under risk: The case of racetrack bettors," *Journal of Political Economy*, 108, 503–530.

Kliger, D. and O. Levy (2009): "Theories of choice under risk: Insights from financial markets," *Journal of Economic Behavior & Organization*, 71, 330–346.

Kristensen, D. (2009): "Uniform convergence rates of kernel estimators with heterogeneous dependent data," *Econometric Theory*, 25, 1433–1445.

Linn, M., S. Shive, and T. Shumway (2018): "Pricing kernel monotonicity and conditional information," *The Review of Financial Studies*, 31, 491–531.

Linton, O., S. Sperlich, and I. Van Keilegom (2008): "Estimation of a semiparametric transformation model," *The Annals of Statistics*, 36, 686–718.

Liu, X., M. B. Shackleton, S. J. Taylor, and X. Xu (2007): "Closed-form transformations from risk-neutral to real-world distributions," *Journal of Banking & Finance*, 31, 1501–1520.

Newey, W. K. (1994): "The asymptotic variance of semiparametric estimators," *Econometrica*, 62, 1349–1382.

Polkovnichenko, V. and F. Zhao (2013): "Probability weighting functions implied in options prices," *Journal of Financial Economics*, 107, 580–609.

Prelec, D. (1998): "The probability weighting function," *Econometrica*, 66, 497–527.

Qiu, J. and E.-M. Steiger (2011): "Understanding the two components of risk attitudes: An experimental analysis," *Management Science*, 57, 193–199.

Quiggin, J. (1982): "A theory of anticipated utility," *Journal of Economic Behavior & Organization*, 3, 323–343.

Roëll, A. (1987): "Risk aversion in Quiggin and Yaari's rank-order model of choice under uncertainty," *The Economic Journal*, 97, 143–159.

Rosenberg, J. V. and R. F. Engle (2002): "Empirical pricing kernels," *Journal of Financial Economics*, 64, 341–372.

Schreindorfer, D. and T. Sichert (2023): "Volatility and the pricing kernel," *Swedish House of Finance Research Paper*.

Snowberg, E. and J. Wolfers (2010): "Explaining the favorite–long shot bias: Is it risk-love or misperceptions?" *Journal of Political Economy*, 118, 723–746.

Song, Z. and D. Xiu (2016): "A tale of two option markets: Pricing kernels and volatility risk," *Journal of Econometrics*, 190, 176–196.

Tversky, A. and D. Kahneman (1992): "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, 5, 297–323.

Van de Kuilen, G. and P. P. Wakker (2011): "The midweight method to measure attitudes toward risk and ambiguity," *Management Science*, 57, 582–598.

Yaari, M. (1987): "The dual theory of choice under risk," *Econometrica*, 55, 95–115.

Yoshihara, K.-i. (1976): "Limiting behavior of U-statistics for stationary, absolutely regular processes," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 35, 237–252.