Empirical Likelihood-Based Constrained Nonparametric Regression with an Application to Option Price and State Price Density Estimation

Guangyi Ma*†

Texas A&M University

This version: January 20, 2011

Abstract

Economic models often imply that the proposed functional relationship between economic variables must satisfy certain shape restrictions. This paper develops a constrained nonparametric regression method to estimate a function and its derivatives subject to such restrictions. We construct a set of constrained local quadratic (CLQ) estimators based on empirical likelihood. Under standard regularity conditions, the proposed CLQ estimators are shown to be weakly consistent and have the same first order asymptotic distribution as the conventional unconstrained estimators. The CLQ estimators are guaranteed to be within the inequality constraints imposed by economic theory, and display similar smoothness as the unconstrained estimators. At a location where the unconstrained estimator for a curve (e.g., the second derivative) violates a restriction, the corresponding CLQ estimator is adjusted towards to the true function. Interestingly, such bias reduction can also be achieved when the binding effect is from a restriction on another curve (e.g., the first derivative). This finite sample gain is achieved through joint estimation of the functions, using the same empirical likelihood weights. We apply this procedure to estimate the day-to-day option pricing function and the corresponding state-price density function with respect to different strike prices.

Key Words**:** Constrained Nonparametric Regression, Empirical Likelihood, Derivative Estimation, Option Pricing, State-Price Density.

JEL Classification Numbers: C13, C14, C58.

---

*Address: Department of Economics, Texas A&M University, 4228 TAMU, College Station, TX 77843, USA, telephone: 1-979-599-8327, e-mail: *gma@econmail.tamu.edu*.

†I am indebted to Qi Li and Ke-Li Xu for their guidance. I am also grateful to the participants of the Econometrics Seminar at the Department of Economics, Texas A&M University. All errors are my own responsibility.

1 Introduction

Nonparametric regression methods are known to be robust to functional form misspecification, hence they are useful when the researcher does not have a theory specifying the exact relationship between economic variables. However, in many cases, economic theory indicates that the functional relationship between two variables X and Y, say, $Y = m(X)$, should be under certain shape restrictions such as monotonicity, convexity, homogeneity, etc. Because the estimation results from nonparametric regression are not guaranteed to satisfy these ex-ante model restrictions, it is desirable to develop a methodology to accommodate such conventional restrictions in nonparametric estimation.

In previous literature, various approaches to nonparametric regression which satisfy monotonic restriction have been developed. See Matzkin (1994) for a comprehensive survey. A popular approach in the existing research literature is the isotonic regression method (e.g., Hansen, Pledger, and Wright, 1973; Dykstra, 1983; Goldman and Rudd, 1992; Rudd, 1995; etc.)[1]. A less desirable feature of the isotonic regression technique is that the estimated function might not be smooth. To produce monotonic yet still smooth estimation results, one can add a kernel-based smoothing step with the isotonic regression (e.g., Mukerjee, 1988; Mammen, 1991). Recently, Aït-Sahalia and Duarte (2003) proposed a similar two-step procedure to estimate option price function nonparametrically. In the first step, they adopt Dykstra's (1983) constrained least square algorithm to trim the data so that the estimates from the succeeding kernel smoothing step are guaranteed to be monotonic and convex. In this constrained least square method, the numerical search is performed iteratively in a subset of the n-dimensional Euclidean space, where n is the sample size. This algorithm is potentially computation-intensive when the sample size is large. Consequently, it is practically useful to combine theory-imposed restrictions with a procedure (e.g., kernel smoothing method) that will yield smooth estimated functions, and reduce computational burden.

[1] Another vast line in the literature is "constrained smoothing splines" (e.g., Yatchew and Bos (1997) etc.). See a recent survey by Henderson and Parmeter (2009) for this and other methods.

In this paper, we explore the possibilities of imposing shape restrictions in the local polynomial regression framework. We construct constrained local quadratic (CLQ) estimators specifically for the functions $m(X)$, $m'(X)$, and $m''(X)$. The proposed estimators can be viewed as reweighted versions of the standard local quadratic (LQ) estimators and the weights are determined via empirical likelihood (EL) maximization. Empirical likelihood (Owen, 1988, 1990, 1991) is a nonparametric likelihood method, in contrast to the widely known parametric likelihood method. See Kitamura (2006) for a comprehensive survey of EL in econometrics. EL can be applied in both parametric and nonparametric models. In parametric estimation, a generalized empirical likelihood estimator is shown to have advantages, in terms of higher order asymptotic properties, over the GMM estimator (Newey and Smith 2004). The idea of parametric estimation via EL is to maximize a nonparametric likelihood ratio $\prod_{i=1}^{n}(np_i)$ between a probability measure $\{p_i : i = 1, \cdots, n\}$ given on the sample points and the empirical distribution $\{1/n, \cdots, 1/n\}$, subject to moment conditions (Qin and Lawless, 1994; Kitamura, Tripathi, and Ahn, 2004). EL can also be used in combination with nonparametric models. For a given nonparametric estimator, confidence intervals via EL has demonstrated advantages over asymptotic normality-based approaches. Refer to Hall and Owen (1993), Chen (1996) for density function estimation; Chen and Qin (2000), Qin and Tsao (2005) for local linear estimators of conditional mean function; Cai (2002) for conditional distribution and regression quantiles; Xu (2009) for local linear estimators in continuous-time diffusion models. The common approach in this literature is to maximize the nonparametric likelihood ratio $\prod_{i=1}^{n}(np_i)$ subject to the EL-weighted estimating equations which can be viewed as counterparts of the moment conditions in parametric settings.

Following this line, we consider the empirical likelihood profile $\{p_i : i = 1, \cdots, n\}$ embedded on a set of local quadratic estimators and we maximize $\prod_{i=1}^{n}(np_i)$ under the desired shape restrictions. If the restrictions are true for the underlying data generating process, then the empirical likelihood profile asymptotically converges to $\{1/n, \cdots, 1/n\}$ as n goes to infinity. Hence our CLQ estimators have the same first order asymptotic distribution as the standard local quadratic estimators. Our procedure offers estimation results that are smooth functions, and reduces the dimensions of numerical optimization from sample size n to the number of restrictions. Moreover, the procedure estimates the function $Y = m(X)$ and its first and second derivative simultaneously, so it is particularly useful when one is interested in estimating the derivatives. When multiple nonparametric functions are jointly estimated, it is common for only some of the restrictions to be violated by the unconstrained estimator. By adjusting those violations to meet the constraints, our EL approach can meanwhile tune other functional estimates at the same location towards the corresponding true value. In addition, the procedure can be applied in more general situations when constraints on the function and its derivatives vary by location, unlike the constant constraints considered in existing methods.

Hall and Huang (2001) proposed an EL-based nonparametric regression approach to estimate a function, subject to monotonicity constraints. Under certain assumptions of the weight functions of the original estimator (kernel or local linear weights, etc.), they show the existence of a set of "location independent" EL weights which guarantee the reweighted estimator to be monotonic. Racine, Parmeter, and Du (2009) extend Hall and Huang's (2001) approach to multivariate and multi-constraint cases. Our study is different from these two papers in several aspects. First, we jointly estimate the regression function and its derivatives and investigate the asymptotic distribution of the EL weighted CLQ estimators, whereas the above two papers focus on the estimation of the regression function itself. Second, our EL weights are "location dependent" so that we can accommodate constraints varying upon the domain of X. Third, we allow constraints on the regression function as well as its derivatives, while Racine, Parmeter, and Du's (2009) theoretical results are not directly applicable to such a case where there are multiple constraints on the first and second derivatives with respect to the same explanatory variable X.

As an application, we use the EL-based CLQ estimator to investigate the nonparametric estimation of daily call option prices C as a function of strike prices X. As implied by finance theory, under the assumption of market completeness and no arbitrage opportunities, the price of a call option $C = C(X)$ must be a decreasing and convex function of the option's strike price X. These shape restrictions can be expressed as $C_{t,\tau}(X) \in \left[\max\left(0, S_t e^{-\delta_{t,\tau}\tau} - X e^{-r_{t,\tau}\tau}\right), S_t e^{-\delta_{t,\tau}\tau}\right]$, $C'_{t,\tau}(X) \in [-e^{-r_{t,\tau}\tau}, 0]$, and $C''_{t,\tau}(X) \in [0, \infty)$, where t is the current time, τ is the time-to-expiration, r is the risk free interest rate, and δ is the dividend yield of the underlying asset with price S_t. We estimate $C_{t,\tau}(X)$, $C'_{t,\tau}(X)$, and $C''_{t,\tau}(X)$ under these constraints and compare the results of our CLQ estimation with the results of standard LQ estimation. In a simulation study, we adopt the same simulation set-up as in Aït-Sahalia and Duarte (2003), and find that our results are comparable with theirs in this extremely small sample setting, whereas our procedure exhibits potential advantages such as less intensive computation when the sample size becomes larger. Last, we apply this method to estimate the S&P 500 index options in a typical trading day in May 2009.

The remainder of this paper is organized as follows. In Section 2 we introduce the definition of EL for the local quadratic estimators, subject to inequality constraints. Then we show the equivalence of two saddlepoint problems from the EL formulation to ease the following asymptotic analysis. Next, in Section 3 we study the asymptotic properties of the EL-based CLQ estimator. In Section 4 we apply the constrained estimation procedure to estimate option price function and the state-price density. Section 5 concludes. The proofs and figures are presented in the Appendix.

2 Empirical Likelihood-Based Constrained Local Quadratic Regression

We consider a random sample $\{(X_i, Y_i) : i = 1, \cdots, n\}$ generated from a bivariate distribution. Let us denote the conditional mean function of Y given X by $m(x) = E(Y|X = x)$ and the conditional variance function by $\sigma^2(x) = Var(Y|X = x)$, then the nonparametric regression model under consideration is $Y = m(X) + \sigma(X)u$, where $E(u|X) = 0$ and $Var(u|X) = 1$. We also denote the marginal density of X by $f(\cdot)$. In this section we develop the empirical likelihood formulation in the context of local quadratic regression model subject to inequality constraints.

2.1 The Local Quadratic Estimator

Because our empirical motivation is to impose theory-motivated constraints on the estimators of the functions $m(\cdot)$, $m'(\cdot)$, and $m''(\cdot)$, we focus on the local quadratic regression model, which provides estimators for the three functions simultaneously. The local quadratic estimators can be derived from the following minimization problem

$$\min_{\beta_j(x):\, j=0,1,2} \sum_{i=1}^{n} \left[Y_i - \beta_0(x) - \beta_1(x)(X_i - x) - \beta_2(x)(X_i - x)^2\right]^2 K\left(\frac{X_i - x}{h}\right). \quad (1)$$

Denote $K_i = K((X_i - x)/h)$, and hereafter we shall slightly abuse the notation and write $(m_0(\cdot), m_1(\cdot), m_2(\cdot))^\intercal = \left(m(\cdot), m^{(1)}(\cdot), m^{(2)}(\cdot)/2\right)^\intercal$, then the local quadratic estimator for $(m_0(x), m_1(x), m_2(x))^\intercal$ can be written as

$$\widehat{\beta}(x) = \left(\widehat{\beta}_0(x), \widehat{\beta}_1(x), \widehat{\beta}_2(x)\right)^\intercal,$$

where for $j = 0, 1, 2$,

$$\widehat{\beta}_j(x) = \frac{1}{h^j} \frac{\frac{1}{n}\sum_{i=1}^{n} W_{ji}(x) Y_i}{\frac{1}{n}\sum_{i=1}^{n} W_{0i}(x)}$$

and

$$W_{0i}(x) = \left[\left(s_2 s_4 - s_3^2\right) - \left(s_1 s_4 - s_2 s_3\right)\left(\frac{X_i - x}{h}\right) - \left(s_2^2 - s_1 s_3\right)\left(\frac{X_i - x}{h}\right)^2\right] K_i,$$

$$W_{1i}(x) = \left[\left(s_2 s_3 - s_1 s_4\right) - \left(s_2^2 - s_0 s_4\right)\left(\frac{X_i - x}{h}\right) - \left(s_0 s_3 - s_1 s_2\right)\left(\frac{X_i - x}{h}\right)^2\right] K_i,$$

$$W_{2i}(x) = \left[\left(s_1 s_3 - s_2^2\right) - \left(s_0 s_3 - s_1 s_2\right)\left(\frac{X_i - x}{h}\right) - \left(s_1^2 - s_0 s_2\right)\left(\frac{X_i - x}{h}\right)^2\right] K_i,$$

$$s_j = \frac{1}{nh}\sum_{i=1}^{n}\left(\frac{X_i - x}{h}\right)^j K_i \text{ for } j = 0, 1, 2, 3, 4.$$

2.2 Empirical Likelihood for the Local Quadratic Estimating Equations

In this subsection we construct empirical likelihood for local quadratic regression model. Let $\{p_1, \cdots, p_n\}$ be a discrete probability distribution on the sample $\{(X_i, Y_i) : i = 1, \cdots, n\}$. That is, $\{p_1, \cdots, p_n\}$ is a set of nonnegative numbers adding to unity. At a location x in the domain of X, the profile empirical likelihood ratio at a set of candidate values $\beta(x) = (\beta_0(x), \beta_1(x), \beta_2(x))^\intercal$ of

$$E\left[\widehat{\beta}(x)\right] = \left(E\left[\widehat{\beta}_0(x)\right], E\left[\widehat{\beta}_1(x)\right], E\left[\widehat{\beta}_2(x)\right]\right)^\intercal$$

is defined as[2]

$$L(\beta) = \max_{\{p_1, \cdots, p_n\}}\left\{\prod_{i=1}^{n} np_i \mid p_i \geqslant 0,\ \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} p_i U_i(\beta) = 0\right\}, \tag{2}$$

where $U_i(\beta) = (U_{0i}(\beta), U_{1i}(\beta), U_{2i}(\beta))^\intercal$ and $U_{ji}(x, \beta(x)) = W_{ji}(x)\left[Y_i - (X_i - x)^j \beta_j(x)\right]$. The three equations

$$\sum_{i=1}^{n} p_i U_i(\beta) = 0 \tag{3}$$

[2]Hereafter, when it is clear we shall omit the explicit dependence of a variable on the location x for brevity of notations.

are labelled as "*estimating equations*" in the empirical likelihood literature. Heuristically, if we take $\{p_1, \cdots, p_n\} = \{1/n, \cdots, 1/n\}$, then the estimating equations become

$$\frac{1}{n}\sum\nolimits_{i=1}^{n} W_{ji}(x)\left[Y_i - (X_i - x)^j \beta_j(x)\right] = 0,$$

which can be viewed as reformulations of the first order conditions of the weighted least square problem (1). From the above equations, we can solve, for $j = 0, 1, 2$,

$$\beta_j(x) = \frac{\frac{1}{n}\sum\nolimits_{i=1}^{n} W_{ji}(x) Y_i}{\frac{1}{n}\sum\nolimits_{i=1}^{n} W_{ji}(x)(X_i - x)^j} = \frac{1}{h^j}\frac{\frac{1}{n}\sum\nolimits_{i=1}^{n} W_{ji}(x) Y_i}{\frac{1}{n}\sum\nolimits_{i=1}^{n} W_{0i}(x)}$$

by recognizing that

$$\frac{1}{nh}\sum\nolimits_{i=1}^{n} W_{0i}(x) = \frac{1}{nh}\sum\nolimits_{i=1}^{n} W_{1i}(x)\left(\frac{X_i - x}{h}\right) = \frac{1}{nh}\sum\nolimits_{i=1}^{n} W_{2i}(x)\left(\frac{X_i - x}{h}\right)^2. \quad (4)$$

That is, the candidate values $\beta_j(x)$ coincide with the local quadratic estimators $\widehat{\beta}_j(x)$. In general, the candidate values $\beta_j(x)$ are not fixed at $\widehat{\beta}_j(x)$, and the corresponding $\{p_1, \cdots, p_n\}$ are different from uniform weights $1/n$. As a digression on notation, we will reserve D_n for the common value in (4). That is, we denote

$$D_n = \frac{1}{nh}\sum\nolimits_{i=1}^{n} W_{ji}(x)\left(\frac{X_i - x}{h}\right)^j = s_0\left(s_2 s_4 - s_3^2\right) - s_1(s_1 s_4 - s_2 s_3) - s_2\left(s_2^2 - s_1 s_3\right).$$

By using the log empirical likelihood ratio, we can modify the EL maximization problem (2) to be

$$l(\beta) = \max_{(p_1, \cdots, p_n)}\left\{\sum\nolimits_{i=1}^{n} \log(np_i) \mid p_i \geqslant 0, \ \sum\nolimits_{i=1}^{n} p_i = 1, \ \sum\nolimits_{i=1}^{n} p_i U_i(\beta) = 0\right\}. \quad (5)$$

Further, by introducing Lagrange multipliers $\lambda(\beta) = (\lambda_0(\beta), \lambda_1(\beta), \lambda_2(\beta))^\intercal$ for the estimating equations (3) respectively, we can form the Lagrangian as

$$\mathcal{L} = \sum_{i=1}^{n} \log\left(np_i\right) - \gamma \left(\sum_{i=1}^{n} p_i - 1\right) - n\lambda\left(\beta\right)^{\intercal} \sum_{i=1}^{n} p_i U_i\left(\beta\right),$$

and solve for

$$p_i\left(\beta\right) = \frac{1}{n\left(1 + \lambda\left(\beta\right)^{\intercal} U_i\left(\beta\right)\right)}.$$

Now the log empirical likelihood ratio $l\left(\beta\right)$ can be expressed as

$$l\left(\beta\right) = \min_{\lambda \in \Lambda} \left[-\sum_{i=1}^{n} \log\left(1 + \lambda\left(\beta\right)^{\intercal} U_i\left(\beta\right)\right)\right] = \max_{\lambda \in \Lambda} \sum_{i=1}^{n} \log\left(1 + \lambda\left(\beta\right)^{\intercal} U_i\left(\beta\right)\right), \tag{6}$$

and

$$\widetilde{\lambda}\left(\beta\right) = \arg\max_{\lambda \in \Lambda} \sum_{i=1}^{n} \log\left(1 + \lambda\left(\beta\right)^{\intercal} U_i\left(\beta\right)\right), \tag{7}$$

where

$$\Lambda = \left\{\lambda\left(\beta\right) \in \mathbb{R}^3 | 1 + \lambda\left(\beta\right)^{\intercal} U_i\left(\beta\right) \geqslant 1/n,\ i = 1, \cdots, n\right\}.$$

The domain Λ of $\lambda\left(\beta\right)$ is derived from $p_i \in [0, 1]$ and it is needed to ensure that the arguments of the logarithm are strictly positive.

2.3 Empirical Likelihood Formulation under Inequality Constraints

In this section we investigate the log empirical likelihood ratio (6) under inequality constraints. In the regression model $Y = m\left(X\right) + \sigma\left(X\right) u$, we consider shape restrictions imposed by economic theory in the form of lower and upper bounds on the function $m\left(\cdot\right)$ and its derivatives. More specifically, let $\underline{b}\left(x\right) = \left(\underline{b}_0\left(x\right), \underline{b}_1\left(x\right), \underline{b}_2\left(x\right)\right)^{\intercal}$ and $\overline{b}\left(x\right) = \left(\overline{b}_0\left(x\right), \overline{b}_1\left(x\right), \overline{b}_2\left(x\right)\right)^{\intercal}$, then the restrictions can be expressed as

$$\underline{b}_0(x) \leqslant m_0(x) \leqslant \bar{b}_0(x),$$
$$\underline{b}_1(x) \leqslant m_1(x) \leqslant \bar{b}_1(x),$$
$$\underline{b}_2(x) \leqslant m_2(x) \leqslant \bar{b}_2(x).$$

For example, in the estimation of option price function and its derivatives, we know that $\underline{b}(X) = \left(\max\left(0, S_t e^{-\delta\tau} - X e^{-r\tau}\right), -e^{-r\tau}, 0\right)^\intercal$, and $\bar{b}(X) = \left(S_t e^{-\delta\tau}, 0, \infty\right)^\intercal$. Our goal is to accommodate these constraints in the nonparametric estimation of $m(\cdot)$ and its derivatives.

Because the log empirical likelihood ratio (6) depends on candidate values $\beta(x) = (\beta_0(x), \beta_1(x), \beta_2(x))^\intercal$, we can stack the above inequality constraints and impose them on the candidate values:

$$\underline{b}(x) \leqslant \beta(x) \leqslant \bar{b}(x). \tag{8}$$

Then (6) is modified as

$$\min_{\underline{b}\leqslant\beta\leqslant\bar{b}} l(\beta) = \min_{\underline{b}\leqslant\beta\leqslant\bar{b}} \max_{\lambda\in\Lambda} G_n(\beta,\lambda), \tag{9}$$

where

$$G_n(\beta,\lambda) = \sum_{i=1}^{n} \log\left(1 + \lambda(\beta)^\intercal U_i(\beta)\right).$$

This can be viewed as a saddlepoint problem, and let its solution be $\left(\widetilde{\beta}, \widetilde{\lambda}\right)$.

Remark 1 *The objective function in (5) corresponds to -1 times the Kullback–Leibler distance between the probability distribution $\{p_1, \cdots, p_n\}$ and the empirical distribution $\{1/n, \cdots, 1/n\}$. Thus maximizing the log empirical likelihood ratio $l(\beta)$ can be interpreted as minimizing the Kullback–Leibler distance between $\{p_1, \cdots, p_n\}$ and $\{1/n, \cdots, 1/n\}$. On the other hand, it is easy to verify that (5) attains its global maximum at $\{1/n, \cdots, 1/n\}$, corresponding to the candidate values $\beta(x)$ being equal to the standard LQ estimators $\widehat{\beta}(x)$. Indeed $\widehat{\beta}(x)$ are the minimizers of $l(\beta)$ without imposing the inequality constraints (8). Thus $\widetilde{\beta}(x)$, as the minimizers of $l(\beta)$ in (9), are designed to minimally adjust the standard LQ estimators such* *that the inequality constraints (8) are satisfied.*

Remark 2 *The empirical likelihood formulated so far is for* $E\left[\widehat{\beta}(x)\right] = m(x) + bias$*, rather than for* $m(x)$*. This point has been observed in previous studies of empirical likelihood-based inference for nonparametric models (Chen and Qin (2000), Qin and Tsao (2005), etc.). To reduce the bias, we use an "undersmoothing" bandwidth condition* $nh^7 \to 0$*, as recommended in the literature. We will discuss this approach further in the asymptotic analysis.*

To facilitate the asymptotic analysis of the CLQ estimator $\left(\widetilde{\beta}, \widetilde{\lambda}\right)$, we need to introduce another saddlepoint problem

$$\min_{\beta} \max_{\lambda \in \Lambda, \nu \in \mathbb{R}_+^6} G_n^*(\beta, \lambda, \nu) \tag{10}$$

where

$$G_n^*(\beta, \lambda, \nu) = G_n(\beta, \lambda) + n\underline{\nu}^\intercal \left(\underline{b} - \beta\right) + n\overline{\nu}^\intercal \left(\beta - \overline{b}\right)$$

and $\nu = \left(\underline{\nu}^\intercal, \overline{\nu}^\intercal\right)^\intercal$ is a set of Lagrangian multipliers for the inequalities

$$\underline{b} - \beta \leqslant 0,$$

$$\beta - \overline{b} \leqslant 0.$$

The following lemma states that the two problems (9) and (10) have the same saddlepoints.

Lemma 1 $\left(\widetilde{\beta}, \widetilde{\lambda}\right)$ *is a saddlepoint of* $G_n(\beta, \lambda)$ *and solve (9) if and only if* $\left(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{\nu}\right)$ *is a saddlepoint of* $G_n^*(\beta, \lambda, \nu)$ *that solves (10), where for* $j = 0, 1, 2$*,*

$$\widetilde{\underline{\nu}}_j\left(x, \widetilde{\beta}\right) = \begin{cases} -\frac{1}{n}\sum_{i=1}^n \frac{\widetilde{\lambda}_j\left(\widetilde{\beta}\right) W_{ji}(x)(X_i - x)^j}{1 + \widetilde{\lambda}\left(\widetilde{\beta}\right)^\intercal U_i\left(\widetilde{\beta}\right)} & if\ \underline{b}_j - \widetilde{\beta}_j = 0, \\ 0 & if\ \underline{b}_j - \widetilde{\beta}_j < 0, \end{cases}$$

$$\widetilde{\overline{\nu}}_j\left(x, \widetilde{\beta}\right) = \begin{cases} \frac{1}{n}\sum_{i=1}^n \frac{\widetilde{\lambda}_j\left(\widetilde{\beta}\right) W_{ji}(x)(X_i - x)^j}{1 + \widetilde{\lambda}\left(\widetilde{\beta}\right)^\intercal U_i\left(\widetilde{\beta}\right)} & if\ \widetilde{\beta}_j - \overline{b}_j = 0, \\ 0 & if\ \widetilde{\beta}_j - \overline{b}_j < 0. \end{cases} \tag{11}$$

Remark 3 *In practical implementation, one can program according to the saddlepoint problem (9). Essentially, at each evaluating location* x*, searching for* $\widetilde{\beta}(x)$ *is performed in* $\left[\underline{b}(x), \overline{b}(x)\right]$*. This can be viewed as an "outer loop". While for each candidate* $\beta(x)$*, searching for* $\lambda(\beta)$ *is done in the "inner loop" via maximizing* $G_n(\beta, \lambda)$*. Plugging* p_i *in the EL weighted estimating equations* $\sum_{i=1}^{n} p_i U_i(\beta) = 0$*, we have*

$$\frac{1}{n}\sum\nolimits_{i=1}^{n} \frac{U_i(\beta)}{1+\lambda(\beta)^{\intercal} U_i(\beta)} = 0,$$

which can also be viewed as the first order conditions for $\lambda(\beta)$ *in (6) divided by* n*. In the "inner loop", given a candidate value of* $\beta(x)$*, we can equivalently solve for* $\lambda(\beta)$ *from these first order conditions.*

Remark 4 *The saddlepoint problem (10) is useful in the following asymptotic analysis of the CLQ estimators* $\widetilde{\beta}(x)$*. Since (9) and (10) are equivalent, we use the same notation* $l(\beta)$ *in the remaining of this paper. That is, we denote*

$$\begin{aligned} l(\beta) &= \max_{\lambda\in\Lambda, \nu\in\mathbb{R}_+^6} G_n^*(\beta, \lambda, \nu) \\ &= \max_{\lambda\in\Lambda, \nu\in\mathbb{R}_+^6} \sum\nolimits_{i=1}^{n} \log\left(1+\lambda(\beta)^{\intercal} U_i(\beta)\right) + n\underline{\nu}^{\intercal}\left(\underline{b}-\beta\right) + n\overline{\nu}^{\intercal}\left(\beta-\overline{b}\right). \end{aligned}$$

3 Asymptotic Analysis of the Constrained Local Quadratic Estimators

In this section we first show that, under proper regularity conditions, $\lambda(m) = (\lambda_0(m), \lambda_1(m), \lambda_2(m))^{\intercal}$ converges to zero as $n \to \infty$. This result is presented in Theorem 1. Then we show in Theorem 2 that the CLQ estimators $\widetilde{\beta}(x)$ and the standard LQ estimators $\widehat{\beta}(x)$ are asymptotically equivalent, that is, $\widetilde{\beta}(x)$ and $\widehat{\beta}(x)$ have the same first-order asymptotic distribution. As a starting point, we need the following assumptions:

Assumption 1 *The kernel function $K(\cdot)$ is a symmetric bounded density function compactly supported on $[-1,1]$.*

Assumption 2 *$f(\cdot)$ and $\sigma(\cdot)$ have continuous derivatives up to the second order in a neighborhood of x, and both $f(x)>0$ and $\sigma(x)>0$. Also $m(\cdot)$ has continuous derivatives up to the third order in a neighborhood of x.*

Assumption 3 *$h\to 0$, $nh\to\infty$, and $nh^7\to 0$ as $n\to\infty$.*

Lemma 2 *Under Assumptions 1, 2, and 3, as $n\to\infty$, we have the asymptotic distribution*

$$\sqrt{nh}\left(\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_i(m)-\frac{h^3}{6}m^{(3)}(x)\,f^3(x)\,B_U\right)\xrightarrow{d}N\left(0,\sigma^2(x)\,f^5(x)\,V_U\right), \qquad (12)$$

where

$$B_U=\begin{pmatrix}0\\ \mu_4^2-\mu_2^2\mu_4\\ 0\end{pmatrix},\; V_U=\begin{pmatrix}\omega_0 & 0 & \omega_2\\ 0 & \omega_3 & 0\\ \omega_2 & 0 & \omega_5\end{pmatrix},$$

$$\omega_0=\mu_2^2\left(\mu_4^2\nu_0-2\mu_2\mu_4\nu_2+\mu_2^2\nu_4\right),$$

$$\omega_2=-\mu_2^2\left(\mu_2\mu_4\nu_0-\left(\mu_4+\mu_2^2\right)\nu_2+\mu_2\nu_4\right),$$

$$\omega_3=\left(\mu_4-\mu_2^2\right)^2\nu_2,$$

$$\omega_5=\mu_2^2\left(\mu_2^2\nu_0-2\mu_2\nu_2+\nu_4\right).$$

Remark 5 *To investigate the asymptotic behavior of the EL-based CLQ estimators, first we need to find the asymptotic distribution of the estimating equations (3). Lemma 2 presents the asymptotic distribution of (3) evaluated at the set of true values $m(x)=(m_0(x),m_1(x),m_2(x))$. This result can be derived from the asymptotic distribution of the LQ estimators $\widehat{\beta}(x)$ because (3) can be viewed as a reformulation of $\widehat{\beta}(x)-m(x)$. Essen-*

tially, from Lemma 2 we have

$$\frac{1}{nh}\sum_{i=1}^{n}U_i(m)=O_p\left((nh)^{-1/2}+h^3\right).$$

Remark 6 *Notice that the leading bias terms of* $\frac{1}{nh}\sum_{i=1}^{n}U_{0i}(m)$ *and* $\frac{1}{nh}\sum_{i=1}^{n}U_{2i}(m)$ *in (12) are actually zeros because of the symmetry of the kernel* $K(\cdot)$*. As suggested by Chen and Qin (2000), we use an "undersmoothing" condition* $nh^7\to 0$ *to reduce the bias of* $\frac{1}{nh}\sum_{i=1}^{n}U_{1i}(m)$*. With this condition, we can still use the optimal bandwidth* $h=O\left(n^{-1/9}\right)$ *for the estimation of the regression function and the second derivative.*

Lemma 3 *Under Assumptions 1, 2, and 3, we have*

$$\frac{1}{nh}\sum_{i=1}^{n}U_i(m)U_i(m)^{\intercal}=\Omega_U+o_p(1),$$

where

$$\Omega_U=f^3(x)\begin{pmatrix}\sigma^2(x)\omega_0 & \sigma^2(x)\omega_1 & \sigma^2(x)\omega_2\\ \sigma^2(x)\omega_1 & \left[\sigma^2(x)+m^2(x)\right]\omega_3 & \left[\sigma^2(x)+m^2(x)\right]\omega_4\\ \sigma^2(x)\omega_2 & \left[\sigma^2(x)+m^2(x)\right]\omega_4 & \left[\sigma^2(x)+m^2(x)\right]\omega_5\end{pmatrix}.$$

Theorem 1 *Assume that* $E|Y_i|^s<\infty$ *for some* $s>2$ *and that Assumptions 1, 2, and 3 hold. Then*

$$\lambda(m)=O_p\left((nh)^{-1/2}+h^3\right)=o_p\left(n^{-3/7}\right),$$

also

$$\lambda(m)=\left[\frac{1}{nh}\sum_{i=1}^{n}U_i(m)U_i(m)^{\intercal}\right]^{-1}\left[\frac{1}{nh}\sum_{i=1}^{n}U_i(m)\right]+o_p\left((nh)^{-1/2}+h^3\right).$$

Remark 7 *From Theorem 1 we have that* $p_i(m)=n^{-1}\left(1+\lambda(m)^{\intercal}U_i(m)\right)^{-1}$ *converges to* $1/n$ *with increasing sample size and proper selected bandwidth. Hence the EL-based CLQ* *estimators*

$$\widetilde{\beta}_j(x) = \frac{1}{h^j} \frac{\sum_{i=1}^{n} p_i W_{ji}(x) Y_i}{\sum_{i=1}^{n} p_i W_{0i}(x)} \quad (j = 0, 1, 2)$$

converge to the unconstrained LQ estimators

$$\widehat{\beta}_j(x) = \frac{1}{h^j} \frac{\frac{1}{n}\sum_{i=1}^{n} W_{ji}(x) Y_i}{\frac{1}{n}\sum_{i=1}^{n} W_{0i}(x)} \quad (j = 0, 1, 2)$$

Lemma 4 *Assume that Assumptions 1, 2, and 3 hold, further assume that $nh^5 \to \infty$ as $n \to \infty$. Then $G_n^*(\beta, \lambda, \nu)$ attains its saddlepoint at $\left(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{\nu}\right)$ where $\widetilde{\beta} = \left(\widetilde{\beta}_0, \widetilde{\beta}_1, \widetilde{\beta}_2\right)$ is such that $\left|\widetilde{\beta}_j(x) - m_j(x)\right| \leqslant h^{2-j}$; $\widetilde{\lambda} = \lambda\left(\widetilde{\beta}\right)$ is given by (7); and $\widetilde{\nu}$ is given by (11). Further, $\left(\widetilde{\beta}, \widetilde{\lambda}\right)$ satisfies*

$$g_{1n}\left(\widetilde{\beta}, \widetilde{\lambda}\right) = 0, \ g_{1n}\left(\widetilde{\beta}, \widetilde{\lambda}\right) = 0,$$

where

$$g_{1n}(\beta, \lambda) = \frac{1}{nh} \sum_{i=1}^{n} \frac{U_i(\beta)}{1 + \lambda(\beta)^{\intercal} U_i(\beta)},$$
$$g_{2n}(\beta, \lambda) = \frac{1}{nh} \sum_{i=1}^{n} \frac{D_i(x) \lambda(\beta)}{1 + \lambda(\beta)^{\intercal} U_i(\beta)},$$

and $D_i(x)$ is a 3×3 matrix such that $D_i(x) = diag\left\{-W_{ji}(x)\left((X_i - x)/h\right)^j\right\}$.

Remark 8 *Lemma 4 shows the existence of a saddlepoint of $G_n^*(\beta, \lambda, \nu)$ in the interior of a (asymptotically shrinking) neighborhood of $m(x)$,*

$$\left\{\beta(x) : |\beta_j(x) - m_j(x)| \leqslant h^{2-j}, \ j = 0, 1, 2\right\}. \tag{13}$$

This is achieved by establishing a lower bound of $l(\beta)$ out of (13) and then we show that this lower bound is of a larger stochastic order than $l(m)$.

Theorem 2 *Suppose that the assumptions of Theorem 1 hold. We also assume that $nh^5 \to$*

∞ *as* $n \to \infty$*. Then for* $j = 0, 1, 2$*, the EL-based CLQ estimator*

$$\widetilde{\beta}_j(x) = \widehat{\beta}_j(x) + o_p\left(\left(nh^{1+2j}\right)^{-1/2} + h^{3-j}\right),$$

where for each j*,* $\widehat{\beta}_j(x)$ *is the corresponding LQ estimator. As* $n \to \infty$*, the asymptotic distribution of* $\widetilde{\beta}(x)$ *is given by*

$$diag\left(\sqrt{nh^{1+2j}}\right)\left(\widetilde{\beta}(x) - m(x) - \frac{h^2}{6}m^{(3)}(x)B_{\widetilde{\beta}}\right) \xrightarrow{d} N\left(0, \frac{\sigma^2(x)}{f(x)}V_{\widetilde{\beta}}\right), \tag{14}$$

where

$$B_{\widetilde{\beta}} = \begin{pmatrix} 0 \\ \mu_4/\mu_2 \\ 0 \end{pmatrix}, \; V_{\widetilde{\beta}} = \frac{V_U}{\mu_2^2\left(\mu_4 - \mu_2^2\right)^2}.$$

Remark 9 *Theorem 2 shows that the EL-based CLQ estimators* $\widetilde{\beta}(x)$ *and the standard LQ estimators* $\widehat{\beta}(x)$ *have the same asymptotic distribution up to the first order. This result is naturally expected because, as the sample size increases,* $\widehat{\beta}(x)$ *converges to the true function values which are in the bounded region* $\left[\underline{b}(x), \overline{b}(x)\right]$*, hence the inequality constraints become unbinding (*$\widetilde{\nu} \xrightarrow{p} 0$*) and* $\widetilde{\beta}(x)$ *and* $\widehat{\beta}(x)$ *are asymptotically first-order equivalent.*

4 Application: Option Pricing Function and State-Price Density Estimation under Shape Restrictions

4.1 Restrictions Imposed by Option Pricing Theory

To show the usefulness of the CLQ estimation procedure proposed in this paper, we estimate the daily option pricing function and the state-price density function by incorporating various shape restrictions. In summary, given market completeness and no arbitrage assumptions, implication from financial market theory suggests that the price of a call option, as a function of its strike price, must be decreasing and convex.

Let us consider an European call option with price C_t at time t, and expiration time T. Denote by $\tau = T - t$ the maturity, and X the strike price. Also denote by $r_{t,\tau}$ the risk free interest rate and $\delta_{t,\tau}$ the dividend yield of the underlying asset with price S_t. Using these notations, we can give the call option price C_t by

$$C\left(X, S_t, \tau, r_{t,\tau}, \delta_{t,\tau}\right) = e^{-r_{t,\tau}\tau} \int_0^{+\infty} \max\left(0, S_T - X\right) f^*\left(S_T | S_t, \tau, r_{t,\tau}, \delta_{t,\tau}\right) dS_T,$$

where $f^*\left(S_T | S_t, \tau, r_{t,\tau}, \delta_{t,\tau}\right)$ is the state-price density (SPD), also called the risk-neutral density (denoted as $f^*\left(S_T\right)$ for brevity in what follows). Asset pricing theory imposes no arbitrage bounds for the price function as

$$\max\left(0, S_t e^{-\delta_{t,\tau}\tau} - X e^{-r_{t,\tau}\tau}\right) \leqslant C_{t,\tau}\left(X\right) \leqslant S_t e^{-\delta_{t,\tau}\tau}, \tag{15}$$

where $C_{t,\tau}\left(X\right)$ is used to denote $C\left(X, S_t, \tau, r_{t,\tau}, \delta_{t,\tau}\right)$ since we focus on the call option price C as a function of X. For the first derivative

$$\frac{\partial C}{\partial X} = -e^{-r_{t,\tau}\tau} \int_X^{+\infty} f^*\left(S_T\right) dS_T,$$

the no-arbitrage assumption requires C to be a decreasing function of X, and the above derivative larger than $-e^{-r_{t,\tau}\tau}$. Thus we have

$$-e^{-r_{t,\tau}\tau} \leqslant C'_{t,\tau}\left(X\right) \leqslant 0 \tag{16}$$

from the positivity and integrability to one of the SPD. The second derivative is

$$C''_{t,\tau}\left(X\right) = e^{-r_{t,\tau}\tau} f^*\left(X\right) \geqslant 0 \tag{17}$$

since the SPD must be positive.

Given the data (X_i, C_i) recorded at time t (typically in one trading day) with the same maturity τ, our objective is to estimate $C_{t,\tau}(X)$, $C'_{t,\tau}(X)$, and $C''_{t,\tau}(X)$ under constraints (15), (16), and (17).

4.2 Monte-Carlo Simulation

To compare the performance of our procedure with existing research, we adopt the simulation setup as that in Aït-Sahalia and Duarte (2003). Specifically, the true call option price function is assumed to be parametric as in the Black-Scholes/Merton model

$$C_{BS}(X, F_{t,\tau}, \tau, r_{t,\tau}, \sigma) = e^{-r_{t,\tau}\tau}\left[F_{t,\tau}\Phi(d_1) - X\Phi(d_2)\right],$$

where $F_{t,\tau} = S_t e^{(r_{t,\tau}-\delta_{t,\tau})\tau}$ is the forward price of the underlying asset at time t and

$$d_1 = \frac{\log(F_{t,\tau}/X)}{\sigma\sqrt{\tau}} + \frac{\sigma\sqrt{\tau}}{2},\ d_2 = \frac{\log(F_{t,\tau}/X)}{\sigma\sqrt{\tau}} - \frac{\sigma\sqrt{\tau}}{2},$$

and $\sigma = \sigma(X/F_{t,\tau}, \tau)$ is the volatility parameter. To generate data for simulation, we calibrate parameter values from real observations of S&P 500 index options on May 13, 1999. The parameter values and domain of strike prices are set as

$$\begin{aligned} S_t &= 1365, & \tau &= 30/252, \\ r_{t,\tau} &= 4.5\%, & X_i &\in [1000, 1700], \\ \delta_{t,\tau} &= 2.5\%, & \sigma_i &= -X_i/140 + 432/35. \end{aligned}$$

In the first simulation, the strike prices X are equally spaced between 1000 and 1700 with a sample size of 25. That is, in each sample there are 25 distinct strike prices and each of them corresponds to one call option price. In the second simulation, we generate 10 call option prices for each distinct strike price, so the sample size is 250.[3] To generate option prices, in the first simulation ($n = 25$), we add uniform noise to the true option price function,

[3] This is similar with the simulation setup in Yatchew and Hardle (2006).

which ranges from 3% of the true price value for deep in the money options ($X = 1000$) to 18% for deep out of the money options ($X = 1700$). We double the noise size in the second simulation ($n = 250$).[4] We use the Epanechnikov kernel in the local qradratic estimation and adopt a rule-of-thumb bandwidth as in Fan and Mancini (2009). In each simulation experiment, we generate and estimate 1000 samples and show the average, 5%, and 95% quantiles as confidence bands in each graph.

For samples with 25 observations in the first simulation, the estimation results are shown in Figure 1. The sample size in this simulation is tiny, so the unconstrained local quadratic estimators, especially the estimators for first and second derivatives, $\widehat{C'}(X)$ and $\widehat{C''}(X)$, perform poorly and violate the constraints frequently. Although difficult to distinguish in the graph, the estimator for the option price, $\widehat{C}(X)$, also violates the lower bound when the strike price is low for deep in the money options. This violation of constraint can be adjusted in our EL-based estimator, while not in Ait-Sahalia and Duarte (2003). Turning to the constrained estimation by our EL-based procedure, we can find that all three estimators, $\widetilde{C}(X)$, $\widetilde{C'}(X)$, and $\widetilde{C''}(X)$, are guaranteed to satisfy the constraints, and the estimators for first and second derivatives have smaller confidence bands in both of the boundary areas of the domain of X. An interesting finding is that, by correcting the violation of constraints in the first derivative estimate, the EL-based procedure also adjusts the second derivative estimate towards to its true function in corresponding boundary areas, although the unconstrained estimate itself, $\widehat{C''}(X)$, may not violate its nonnegative lower bound.[5]

In the second simulation with 250 observations (Figure 2), performance of the unconstrained local quadratic estimators is better than in the previous small sample design in spite of the doubled noise size. With a sample size as large as 250, the unconstrained estimate ($\widehat{C}(X)$) of the option price function and its true value become undistinguishable. But for the estimation of derivatives, the unconstrained estimators ($\widehat{C'}(X)$ and $\widehat{C''}(X)$) still

[4] If we use the same noise design in the second simulation with larger sample size, the unconstrained local quadratic estimates will violate constrains less so the constrained and unconstrained estimation results will be close.

[5] Note that the 5% quantile of $\widetilde{C'}(X)$ corresponds to the 95% quantile of $\widetilde{C''}(X)$ and vice versa.

violate the constraints when the strike price is very low or very high. In comparison, the EL-based constrained estimators ($\widetilde{C'}(X)$, and $\widetilde{C''}(X)$) are strictly within the constraints and have much narrower confidence bands, specially, in the left boundary area.

Last, we compare the integrated mean squared errors (IMSE) from constrained and unconstrained estimation in Figure 3. We focus on the first simulation design with sample size 25. The plots show that the IMSE's are much lower for the constrained estimators in all three functional estimations. Also we find a U-shaped IMSE curve in all three cases, showing that there exists an optimal bandwidth minimizing the IMSE.

4.3 Empirical Analysis

To investigate the empirical performance of our EL-based CLQ estimators, we estimate the option price function and the state price density (a scaled second derivative of the option price function). We consider closing prices of European call options on the S&P 500 index (symbol SPX). The SPX index option is one of the most actively traded options and has been studied extensively in empirical option pricing literature. The data are downloaded from *OptionMetrics*. We collect options on May 18, 2009 for a maturity of 61 days corresponding to the expiration on July 18, 2009. Following Aït-Sahalia and Lo (1998), Fan and Mancini (2009), we use the bid-ask average of closing price as the option price, and we delete less liquid options with implied volatility larger than 70%, or price less than or equal to 0.125. Finally we reach a sample of 81 call option prices with strike prices ranging from 715 to 1150. The closing spot price of the S&P 500 index on that day was 909.71, and the risk free interest rate for the 2-month maturity was 0.51%. The dividend yield is retrieved from the put-call parity. Figure 4 presents the estimation results of this daily cross-sectional option prices data set. From the results we can find that the unconstrained estimate of the first derivative significantly violates the constraints at both the in-the-money and the out-of-the-money areas. In contrast, the constrained estimate of this function is bounded in both areas.

5 Conclusions

We propose an empirical likelihood-based constrained local quadratic regression procedure to accomodate general shape restrictions imposed by economic theory. The resulted estimates can satisfy the constraints on the function and its first and second derivatives. Compared with the traditional "isotonic regression and smoothing" two-step method, the EL-based approach can be less computationally intensive, and can accommodate more general constraints, hence it shows potential to be useful in a wide range of applications. We study the empirical performance of the constrained estimation method in both simulations and real data applications.

A part of the follow-up work is to analyze more extensively the EL-based CLQ estimators, both asymptotically and in finite sample, such as the comparison of mean squared errors between the constrained and unconstrained estimators. Another direction which might enrich the scope of this paper is to develop tests on shape restrictions such as monotonicity and convexity, based on the asymptotic chi-square distribution of the log EL ratio statistic.

References

Aït-Sahalia, Y., Duarte, J., 2003. Nonparametric option pricing under shape restrictions. Journal of Econometrics 116, 9–47.

Aït-Sahalia, Y., Lo, A., 1998. Nonparametric estimation of state-price densities implicit in 2nancial asset prices. Journal of Finance 53, 499–547.

Cai, Z., 2001. Weighted Nadaraya-Watson regression estimation. Statistics and Probability Letters 51, 307–318.

Cai, Z., 2002. Regression quantiles for time series. Econometric Theory 18, 169–192.

Chen, S.X., 1996. Empirical likelihood confidence intervals for nonparametric density estimation. Biometrika 83, 329–341.

Chen, S.X., Qin, Y.-S., 2000. Empirical likelihood confidence interval for a local linear smoother. Biometrika 87, 946–953.

Dykstra, R.L., 1983. An algorithm for restricted least squares. Journal of the American Statistical Association 78, 837–842.

Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and its Applications. Chapman & Hall, London.

Fan, J., Mancini, L., 2009. Option pricing with model-guided nonparametric methods. Journal of the American Statistical Association 104, 1351–1372.

Hall, P., Huang, H., 2001. Nonparametric kernel regression subject to monotonicity constraints. The Annals of Statistics 29, 624–647.

Hall, P., Owen, A.B., 1993. Empirical likelihood confidence bands in density estimation. Journal of Computational and Graphical Statistics 2, 273–289.

Hall, P., Presnell, B. 1999. Intentionally biased bootstrap methods. Journal of the Royal Statistical Society, Series B, 61, 143–158.

Henderson, D., Parmeter C., 2009. Imposing economic constraints on nonparametric regression: Survey, implementation and extensions. In: Li, Q., Racine, J. S. (Eds.), Advances in Econometrics: Nonparametric Methods. Elsevier Science, Vol. 25, 433–469.

Matzkin, R.L., 1994. Restrictions of economic theory in nonparametric methods. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, Vol. 4, North Holland, Amsterdam.

Moon, H.R., Schorfheide, F., 2009. Estimation with overidentifying inequality moment conditions. Journal of Econometrics 153, 136–154.

Kitamura, Y., 2006. Empirical likelihood methods in econometrics: Theory and practice. In: Blundell, R., Torsten, P., Newey, W.K. (Eds.), Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress. Cambridge University Press, Cambridge.

Kitamura, Y., Tripathi G., Ahn H., 2004. Empirical likelihood based inference in conditional moment restriction models, Econometrica 72, 1667–1714.

Li, Q., Racine, J., 2007. Nonparametric Econometrics: Theory and Practice. Princeton, NJ: Princeton University Press.

Newey, W.K., Smith R.J., 2004. Higher order properties of GMM and generalized empirical likelihood estimators, Econometrica 72, 219–256.

Owen, A.B., 1988. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75, 237–249.

Owen, A.B., 1990. Empirical likelihood confidence regions. Annals of Statistics 18, 90–120.

Owen, A.B., 1991. Empirical likelihood for linear models. Annals of Statistics 19, 1725–1747.

Owen, A.B., 2001. Empirical Likelihood. Chapman and Hall, New York.

Pagan, A., Ullah, A., 1999. Nonparametric Econometrics. New York: Cambridge University Press.

Qin, G., Tsao, M., 2005. Empirical likelihood based inference for the derivative of the nonparametric regression function. Bernoulli 11, 715–735.

Qin, J., Lawless, J., 1994. Empirical likelihood and general estimating equations. Annals of Statistics 22, 300–325.

Racine, J.S., Parmeter, C.F., Du, P., 2009. Constrained nonparametric kernel regression: Estimation and inference. Virginia Tech AAEC Working Paper.

Xu, K.-L., 2009. Empirical likelihood based inference for recurrent nonparametric diffusions. Journal of Econometrics 153, 65–82.

Xu, K.-L., 2009, Re-weighted functional estimation of diffusion models. Econometric Theory 26, 541–563.

Yatchew, A.J., Bos, L., 1997. Nonparametric regression and testing in economic models. Journal of Quantitative Economics 13, 81–131.

Yatchew, A., Hardle, W., 2006. Nonparametric state price density estimation using constrained least squares and the bootstrap. Journal of Econometrics 133, 579–599.

Appendix A: Proofs

Proof of Lemma 1

Proof. (i) Let $\left(\widetilde{\beta},\widetilde{\lambda},\widetilde{\nu}\right)$ be a saddlepoint of $G_n^*(\beta,\lambda,\nu)$ solving (10). First we look at the upper bounds $\overline{b}$. Suppose there is $j\in\{0,1,2\}$ such that $\widetilde{\beta}_j>\overline{b}_j$, then there must exist $\overline{\nu}_j'>\widetilde{\overline{\nu}}_j\geqslant 0$ such that $\overline{\nu}_j'\left(\widetilde{\beta}_j-\overline{b}_j\right)>\widetilde{\overline{\nu}}_j\left(\widetilde{\beta}_j-\overline{b}_j\right)$, so $G_n^*\left(\widetilde{\beta},\widetilde{\lambda},\widetilde{\underline{\nu}},\widetilde{\overline{\nu}}_{-j},\overline{\nu}_j'\right)>G_n^*\left(\widetilde{\beta},\widetilde{\lambda},\widetilde{\nu}\right)$, which contradicts with the definition of $\left(\widetilde{\beta},\widetilde{\lambda},\widetilde{\nu}\right)$. Therefore $\widetilde{\beta}_j\leqslant\overline{b}_j$ for all $j=0,1,2$. This implies that $\widetilde{\overline{\nu}}^{\intercal}\left(\widetilde{\beta}-\overline{b}\right)\leqslant 0$ since $\widetilde{\overline{\nu}}\geqslant 0$. Further, if $\widetilde{\beta}_j<\overline{b}_j$, then $\widetilde{\overline{\nu}}_j=0$. Together we have $\widetilde{\overline{\nu}}^{\intercal}\left(\widetilde{\beta}-\overline{b}\right)=0$. Similarly we can show that $\widetilde{\underline{\nu}}^{\intercal}\left(\underline{b}-\widetilde{\beta}\right)=0$. So

$$G_n^*\left(\widetilde{\beta},\widetilde{\lambda},\widetilde{\nu}\right)=G_n\left(\widetilde{\beta},\widetilde{\lambda}\right)\geqslant G_n(\beta,\lambda)$$

for any $\beta\in\left(\underline{b},\overline{b}\right)$ and $\lambda\in\Lambda$. That is, $\left(\widetilde{\beta},\widetilde{\lambda}\right)$ is a saddlepoint of $G_n(\beta,\lambda)$ that solves (9).

(ii) Let $\left(\widetilde{\beta},\widetilde{\lambda}\right)$ be a saddlepoint of $G_n(\beta,\lambda)$ solving (9) and $\widetilde{\nu}$ as defined in the lemma. We want to show that $\left(\widetilde{\beta},\widetilde{\lambda},\widetilde{\nu}\right)$ is a saddlepoint of (10). First, since $\widetilde{\overline{\nu}}^{\intercal}\left(\widetilde{\beta}-\overline{b}\right)=0$ and $\widetilde{\underline{\nu}}^{\intercal}\left(\underline{b}-\widetilde{\beta}\right)=0$ by the definition of $\widetilde{\nu}$, also since $\widetilde{\beta}\in\left[\underline{b},\overline{b}\right]$, we have $\widetilde{\overline{\nu}}^{\intercal}\left(\widetilde{\beta}-\overline{b}\right)\geqslant\overline{\nu}^{\intercal}\left(\widetilde{\beta}-\overline{b}\right)$ and $\widetilde{\underline{\nu}}^{\intercal}\left(\underline{b}-\widetilde{\beta}\right)\geqslant\underline{\nu}^{\intercal}\left(\underline{b}-\widetilde{\beta}\right)$ for any $\nu=\left(\underline{\nu}^{\intercal},\overline{\nu}^{\intercal}\right)^{\intercal}\in\mathbb{R}_+^6$, hence

$$G_n^*\left(\widetilde{\beta},\widetilde{\lambda},\widetilde{\nu}\right)\geqslant G_n^*\left(\widetilde{\beta},\widetilde{\lambda},\nu\right). \tag{18}$$

Second, let $\widetilde{\beta}=\left(\widetilde{\beta}_j,\widetilde{\beta}_{-j}\right)$ such that $\widetilde{\beta}_j=\overline{b}_j$ (or $\widetilde{\beta}_j=\underline{b}_j$) and $\widetilde{\beta}_{-j}\in\left(\underline{b}_{-j},\overline{b}_{-j}\right)$. Then for any $\beta\in\left[\underline{b},\overline{b}\right]$, we make the same partition $\beta=(\beta_j,\beta_{-j})$ and have

$$G_n^*\left(\beta,\widetilde{\lambda},\widetilde{\nu}\right)=G_n^*\left(\beta_j,\beta_{-j},\widetilde{\lambda},\widetilde{\nu}\right)=G_n^*\left(\beta_j,\widetilde{\beta}_{-j},\widetilde{\lambda},\widetilde{\nu}\right),$$

where the second equality holds because by definition the part in $\widetilde{\nu}$ corresponding to β_{-j} are zeros. Further, we have

$$G_n^*\left(\beta_j, \widetilde{\beta}_{-j}, \widetilde{\lambda}, \widetilde{\nu}\right) \geqslant G_n^*\left(\widetilde{\beta}_j, \widetilde{\beta}_{-j}, \widetilde{\lambda}, \widetilde{\nu}\right)$$

since $G_n^*(\beta, \lambda, \nu)$ is globally convex in β_j, and by definition of $\widetilde{\nu}$,

$$\left.\frac{\partial G_n^*(\beta, \lambda, \nu)}{\partial \beta_j}\right|_{\widetilde{\beta}, \widetilde{\lambda}, \widetilde{\nu}} = 0.$$

Together we have

$$G_n^*\left(\beta, \widetilde{\lambda}, \widetilde{\nu}\right) \geqslant G_n^*\left(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{\nu}\right). \tag{19}$$

Finally, (18) and (19) imply that $\left(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{\nu}\right)$ is a saddlepoint of (10). ■

Proof of Lemma 2

For the general local polynomial estimators, the asymptotic conditional bias and variance terms are discussed in Fan and Gijbels (1996), Theorem 3.1. Following their notations, we denote, in the case of local quadratic estimator,

$$S = \begin{pmatrix} \mu_0 & \mu_1 & \mu_2 \\ \mu_1 & \mu_2 & \mu_3 \\ \mu_2 & \mu_3 & \mu_4 \end{pmatrix}, \ S^* = \begin{pmatrix} \nu_0 & \nu_1 & \nu_2 \\ \nu_1 & \nu_2 & \nu_3 \\ \nu_2 & \nu_3 & \nu_4 \end{pmatrix}, \ c_2 = \begin{pmatrix} \mu_3 \\ \mu_4 \\ \mu_5 \end{pmatrix}, \ \widetilde{c}_2 = \begin{pmatrix} \mu_4 \\ \mu_5 \\ \mu_6 \end{pmatrix},$$

where $\mu_j = \int u^j K(u)\, du$, $\nu_j = \int u^j K^2(u)\, du$. Note that $\mu_0 = 1$, and for a symmetric kernel, $\mu_1 = \mu_3 = \mu_5 = \nu_1 = \nu_3 = 0$. Then the asymptotic bias is given by

$$Bias\left(\widehat{\beta}_j(x) \,|\mathbf{X}\right) = e_{j+1}^{\intercal} S^{-1} c_2 \frac{m^{(3)}(x)}{6} h^{3-j} + o_p\left(h^{3-j}\right)$$

for $j = 1$, and

$$Bias\left(\widehat{\beta}_j(x)\,|\mathbf{X}\right) = e_{j+1}^{\intercal} S^{-1} \widetilde{c}_2 \frac{1}{24}\left(m^{(4)}(x) + 4m^{(3)}(x)\frac{f^{(1)}(x)}{f(x)}\right) h^{4-j} + o_p\left(h^{4-j}\right)$$

for $j = 0, 2$. The asymptotic variances are given by

$$Var\left(\widehat{\beta}_j(x)\,|\mathbf{X}\right) = e_{j+1}^{\intercal} S^{-1} S^{*} S^{-1} e_{j+1} \frac{\sigma^2(x)}{f(x)\,nh^{1+2j}} + o_p\left(\frac{1}{nh^{1+2j}}\right)$$

for $j = 0, 1, 2$. It is known that the leading term in the asymptotic bias is of a smaller order for j being even than in the case for j being odd. Explicitly, we have

$$\begin{aligned}
Bias\left(\widehat{\beta}_0(x)\,|\mathbf{X}\right) &= \frac{h^4}{24}\frac{\mu_4^2 - \mu_2\mu_6}{\mu_4 - \mu_2^2}\left(m^{(4)}(x) + 4m^{(3)}(x)\frac{f^{(1)}(x)}{f(x)}\right) + o_p\left(h^4\right), \\
Bias\left(\widehat{\beta}_1(x)\,|\mathbf{X}\right) &= \frac{h^2}{6}\frac{\mu_4}{\mu_2} m^{(3)}(x) + o_p\left(h^2\right), \\
Bias\left(\widehat{\beta}_2(x)\,|\mathbf{X}\right) &= \frac{h^2}{24}\frac{\mu_6 - \mu_2\mu_4}{\mu_4 - \mu_2^2}\left(m^{(4)}(x) + 4m^{(3)}(x)\frac{f^{(1)}(x)}{f(x)}\right) + o_p\left(h^2\right), \\
Var\left(\widehat{\beta}_0(x)\,|\mathbf{X}\right) &= \frac{1}{nh}\frac{\mu_4^2\nu_0 - 2\mu_2\mu_4\nu_2 + \mu_2^2\nu_4}{\left(\mu_4 - \mu_2^2\right)^2}\frac{\sigma^2(x)}{f(x)} + o_p\left(\frac{1}{nh}\right), \\
Var\left(\widehat{\beta}_1(x)\,|\mathbf{X}\right) &= \frac{1}{nh^3}\frac{\nu_2}{\mu_2^2}\frac{\sigma^2(x)}{f(x)} + o_p\left(\frac{1}{nh^3}\right), \\
Var\left(\widehat{\beta}_2(x)\,|\mathbf{X}\right) &= \frac{1}{nh^5}\frac{\mu_2^2\nu_0 - 2\mu_2\nu_2 + \nu_4}{\left(\mu_4 - \mu_2^2\right)^2}\frac{\sigma^2(x)}{f(x)} + o_p\left(\frac{1}{nh^5}\right).
\end{aligned}$$

To derive the asymptotic distribution for the estimating equations, we need to introduce more notations. Let $S^{-1} = T/D$, where

$$T = \begin{pmatrix} t_0 & t_1 & t_2 \\ t_1 & t_3 & t_4 \\ t_2 & t_4 & t_5 \end{pmatrix} = \begin{pmatrix} \mu_2\mu_4 - \mu_3^2 & \mu_2\mu_3 - \mu_1\mu_4 & \mu_1\mu_3 - \mu_2^2 \\ \mu_2\mu_3 - \mu_1\mu_4 & \mu_0\mu_4 - \mu_2^2 & \mu_1\mu_2 - \mu_0\mu_3 \\ \mu_1\mu_3 - \mu_2^2 & \mu_1\mu_2 - \mu_0\mu_3 & \mu_0\mu_2 - \mu_1^2 \end{pmatrix},$$

$$D = \det(S) = \mu_0\left(\mu_2\mu_4 - \mu_3^2\right) - \mu_1\left(\mu_1\mu_4 - \mu_2\mu_3\right) - \mu_2\left(\mu_2^2 - \mu_1\mu_3\right),$$

then

$$S^{-1}S^{*}S^{-1}=\frac{1}{D^{2}}TS^{*}T.$$

Note that we have already denoted

$$\begin{aligned}D_{n}&=\frac{1}{nh}\sum\nolimits_{i=1}^{n}W_{0i}\left(x\right)\\&=s_{0}\left(s_{2}s_{4}-s_{3}^{2}\right)-s_{1}\left(s_{1}s_{4}-s_{2}s_{3}\right)-s_{2}\left(s_{2}^{2}-s_{1}s_{3}\right),\end{aligned}$$

thus we have $D_{n}\xrightarrow{p}f^{3}\left(x\right)D$ because $s_{j}\xrightarrow{p}f\left(x\right)\mu_{j}$ for $j=0,1,2,3,4$.

The three estimating equations evaluated at the true values $\left(m_{0}\left(x\right),m_{1}\left(x\right),m_{2}\left(x\right)\right)^{\intercal}$ are

$$\begin{aligned}\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_{0i}\left(m\right)&=\left(\widehat{\beta}_{0}\left(x\right)-m_{0}\left(x\right)\right)D_{n},\\\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_{1i}\left(m\right)&=\left(\widehat{\beta}_{1}\left(x\right)-m_{1}\left(x\right)\right)hD_{n},\\\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_{2i}\left(m\right)&=\left(\widehat{\beta}_{2}\left(x\right)-m_{2}\left(x\right)\right)h^{2}D_{n},\end{aligned}$$

so we can derive that, by assuming $h\to0$, $nh\to\infty$, $nh^{7}\to0$ as $n\to\infty$,

$$\sqrt{nh}\left(\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_{i}\left(m\right)-\frac{h^{3}}{6}m^{(3)}\left(x\right)f^{3}\left(x\right)Tc_{2}\right)\xrightarrow{d}N\left(0,\sigma^{2}\left(x\right)f^{5}\left(x\right)TS^{*}T\right),$$

where

$$Tc_{2}=\begin{pmatrix}\mu_{3}\left(\mu_{2}\mu_{4}-\mu_{3}^{2}\right)-\mu_{4}\left(\mu_{1}\mu_{4}-\mu_{2}\mu_{3}\right)-\mu_{5}\left(\mu_{2}^{2}-\mu_{1}\mu_{3}\right)\\\mu_{3}\left(\mu_{2}\mu_{3}-\mu_{1}\mu_{4}\right)-\mu_{4}\left(\mu_{2}^{2}-\mu_{0}\mu_{4}\right)-\mu_{5}\left(\mu_{0}\mu_{3}-\mu_{1}\mu_{2}\right)\\\mu_{3}\left(\mu_{1}\mu_{3}-\mu_{2}^{2}\right)-\mu_{4}\left(\mu_{0}\mu_{3}-\mu_{1}\mu_{2}\right)-\mu_{5}\left(\mu_{1}^{2}-\mu_{0}\mu_{2}\right)\end{pmatrix},\ TS^{*}T=\begin{pmatrix}\omega_{0}&\omega_{1}&\omega_{2}\\\omega_{1}&\omega_{3}&\omega_{4}\\\omega_{2}&\omega_{4}&\omega_{5}\end{pmatrix},$$

and

$$\omega_0 = t_0^2\nu_0 + 2t_0t_1\nu_1 + \left(2t_0t_2 + t_1^2\right)\nu_2 + 2t_1t_2\nu_3 + t_2^2\nu_4,$$

$$\omega_1 = t_0t_1\nu_0 + \left(t_0t_3 + t_1^2\right)\nu_1 + \left(t_0t_4 + t_1t_2 + t_1t_3\right)\nu_2 + \left(t_1t_4 + t_2t_3\right)\nu_3 + t_2t_4\nu_4,$$

$$\omega_2 = t_0t_2\nu_0 + \left(t_0t_4 + t_1t_2\right)\nu_1 + \left(t_0t_5 + t_1t_4 + t_2^2\right)\nu_2 + \left(t_2t_4 + t_1t_5\right)\nu_3 + t_2t_5\nu_4,$$

$$\omega_3 = t_1^2\nu_0 + 2t_1t_3\nu_1 + \left(2t_1t_4 + t_3^2\right)\nu_2 + 2t_3t_4\nu_3 + t_4^2\nu_4,$$

$$\omega_4 = t_1t_2\nu_0 + \left(t_1t_4 + t_2t_3\right)\nu_1 + \left(t_1t_5 + t_2t_4 + t_3t_4\right)\nu_2 + \left(t_3t_5 + t_4^2\right)\nu_3 + t_4t_5\nu_4,$$

$$\omega_5 = t_2^2\nu_0 + 2t_2t_4\nu_1 + \left(2t_2t_5 + t_4^2\right)\nu_2 + 2t_4t_5\nu_3 + t_5^2\nu_4.$$

For a symmetric kernel $K(\cdot)$, remind that $\mu_1 = \mu_3 = \mu_5 = \nu_1 = \nu_3 = 0$, so $t_0 = \mu_2\mu_4$, $t_2 = -\mu_2^2$, $t_3 = \mu_4 - \mu_2^2$, $t_5 = \mu_2$, $t_1 = t_4 = 0$, and

$$Tc_2 = \begin{pmatrix} 0 \\ \mu_4^2 - \mu_2^2\mu_4 \\ 0 \end{pmatrix},$$

$$TS^*T = \mu_2^2 \begin{pmatrix} \mu_4^2\nu_0 - 2\mu_2\mu_4\nu_2 + \mu_2^2\nu_4 & 0 & -\mu_2\mu_4\nu_0 + \left(\mu_4 + \mu_2^2\right)\nu_2 - \mu_2\nu_4 \\ 0 & \left(\mu_4/\mu_2 - \mu_2\right)^2\nu_2 & 0 \\ -\mu_2\mu_4\nu_0 + \left(\mu_4 + \mu_2^2\right)\nu_2 - \mu_2\nu_4 & 0 & \mu_2^2\nu_0 - 2\mu_2\nu_2 + \nu_4 \end{pmatrix}.$$

Proof of Lemma 3

Lemma 3 states the stochastic order for the squared sums of $U_i(m)$. The proof is similar as that of Lemma 2 in Qin and Tsao (2005). Using the same notation as in Section 2.2, we let

$K_i = K\left(\left(X_i - x\right)/h\right)$, and

$$\begin{aligned}
W_{0i}(x) &= \left[\left(s_2 s_4 - s_3^2\right) - \left(s_1 s_4 - s_2 s_3\right)\left(\frac{X_i - x}{h}\right) - \left(s_2^2 - s_1 s_3\right)\left(\frac{X_i - x}{h}\right)^2\right] K_i \\
&= \left[T_0 + T_1\left(\frac{X_i - x}{h}\right) + T_2\left(\frac{X_i - x}{h}\right)^2\right] K_i, \\
W_{1i}(x) &= \left[\left(s_2 s_3 - s_1 s_4\right) - \left(s_2^2 - s_0 s_4\right)\left(\frac{X_i - x}{h}\right) - \left(s_0 s_3 - s_1 s_2\right)\left(\frac{X_i - x}{h}\right)^2\right] K_i \\
&= \left[T_1 + T_3\left(\frac{X_i - x}{h}\right) + T_4\left(\frac{X_i - x}{h}\right)^2\right] K_i, \\
W_{2i}(x) &= \left[\left(s_1 s_3 - s_2^2\right) - \left(s_0 s_3 - s_1 s_2\right)\left(\frac{X_i - x}{h}\right) - \left(s_1^2 - s_0 s_2\right)\left(\frac{X_i - x}{h}\right)^2\right] K_i \\
&= \left[T_2 + T_4\left(\frac{X_i - x}{h}\right) + T_5\left(\frac{X_i - x}{h}\right)^2\right] K_i,
\end{aligned}$$

then for $j = 0, 1, 2,$

$$\sum\nolimits_{i=1}^{n} U_{ji}(m) = \sum\nolimits_{i=1}^{n} W_{ji}(x)\left[Y_i - m_j(x)\left(X_i - x\right)^j\right].$$

The conclusion in Lemma 3 can be verified as follows.

Lemma A 1 $\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_{0i}^2(m_0) = \sigma^2(x) f^3(x)\omega_0 + o_p(1).$

Proof. Write

$$\begin{aligned}
\frac{1}{nh}\sum U_{0i}^2(m_0) &= \frac{1}{nh}\sum W_{0i}^2(x)\left[Y_i - m(x)\right]^2 \\
&= \frac{1}{nh}\sum W_{0i}^2(x)\left[Y_i - m(X_i)\right]^2 + \frac{2}{nh}\sum W_{0i}^2(x)\left[Y_i - m(X_i)\right]\left[m(X_i) - m(x)\right] \\
&+ \frac{1}{nh}\sum W_{0i}^2(x)\left[m(X_i) - m(x)\right]^2 \\
&= J_1 + 2J_2 + J_3.
\end{aligned}$$

First,

$$
\begin{aligned}
J_1 &= T_0^2 \frac{1}{nh} \sum K_i^2 \sigma^2 (X_i) u_i^2 + 2T_0 T_1 \frac{1}{nh} \sum \left(\frac{X_i - x}{h} \right) K_i^2 \sigma^2 (X_i) u_i^2 \\
&+ \left(2T_0 T_2 + T_1^2 \right) \frac{1}{nh} \sum \left(\frac{X_i - x}{h} \right)^2 K_i^2 \sigma^2 (X_i) u_i^2 \\
&+ 2T_1 T_2 \frac{1}{nh} \sum \left(\frac{X_i - x}{h} \right)^3 K_i^2 \sigma^2 (X_i) u_i^2 + T_2^2 \frac{1}{nh} \sum \left(\frac{X_i - x}{h} \right)^4 K_i^2 \sigma^2 (X_i) u_i^2,
\end{aligned}
$$

since for $j = 0, 1, 2, 3, 4,$

$$
\begin{aligned}
\frac{1}{nh} \sum \left(\frac{X_i - x}{h} \right)^j K_i^2 \sigma^2 (X_i) u_i^2 &= E \left[\left(\frac{X_1 - x}{h} \right)^j \frac{1}{h} K_1^2 \sigma^2 (X_1) u_1^2 \right] + o_p (1) \\
&= \sigma^2 (x) f (x) \int u^j K^2 (u) \, du + o_p (1) \\
&= \sigma^2 (x) f (x) \nu_j + o_p (1) ,
\end{aligned}
$$

and for $j = 0, 1, 2, 3, 4,$

$$
T_j = f^2 (x) t_j + o_p (1) ,
$$

so

$$
\begin{aligned}
J_1 &= \sigma^2 (x) f^3 (x) \left[t_0^2 \nu_0 + 2t_0 t_1 \nu_1 + \left(2t_0 t_2 + t_1^2 \right) \nu_2 + 2t_1 t_2 \nu_3 + t_2^2 \nu_4 \right] + o_p (1) \\
&= \sigma^2 (x) f^3 (x) \omega_0 + o_p (1) .
\end{aligned}
$$

Second, $J_2 = o_p(1)$ since for $j = 0, 1, 2, 3, 4,$

$$\begin{aligned}
&\frac{1}{nh}\sum\left(\frac{X_i - x}{h}\right)^j K_i^2 \sigma(X_i) u_i (m(X_i) - m(x)) \\
&= E\left[\left(\frac{X_1 - x}{h}\right)^j \frac{1}{h} K_1^2 \sigma(X_1) u_1 (m(X_1) - m(x))\right] + o_p(1) \\
&= E\left[\left(\frac{X_1 - x}{h}\right)^j \frac{1}{h} K_1^2 \sigma(X_1) E(u_1|X_1)(m(X_1) - m(x))\right] + o_p(1) \\
&= o_p(1).
\end{aligned}$$

Third, $J_3 = o_p(1)$ since for $j = 0, 1, 2, 3, 4,$

$$\begin{aligned}
&\frac{1}{nh}\sum\left(\frac{X_i - x}{h}\right)^j K_i^2 [m(X_i) - m(x)]^2 \\
&= \frac{1}{nh}\sum\left(\frac{X_i - x}{h}\right)^j K_i^2 \left[m^{(1)}(x)(X_i - x) + o_p(h)\right]^2 \\
&= \frac{\left(m^{(1)}(x)\right)^2 h}{n}\sum\left(\frac{X_i - x}{h}\right)^{j+2} K_i^2 + o_p(h)\frac{2m^{(1)}(x)}{n}\sum\left(\frac{X_i - x}{h}\right)^{j+1} K_i^2 \\
&+ o_p\left(h^2\right)\frac{1}{nh}\sum\left(\frac{X_i - x}{h}\right)^j K_i^2 \\
&= \left(m^{(1)}(x)\right)^2 h^2 [\nu_{j+2} f(x) + o_p(1)] + o_p(h) 2m^{(1)}(x) h [\nu_{j+1} f(x) + o_p(1)] + o_p\left(h^2\right)[\nu_j f(x) + o_p(1)] \\
&= O_p\left(h^2\right) + o_p\left(h^2\right) = o_p(1).
\end{aligned}$$

where the first equality is because the kernel function is bounded in $[-1, 1]$. ■

Lemma A 2 $\frac{1}{nh}\sum_{i=1}^n U_{1i}^2(m_1) = \left[\sigma^2(x) + m^2(x)\right] f^3(x)\,\omega_3 + o_p(1).$

Proof. Write

$$
\begin{aligned}
&\frac{1}{nh}\sum U_{1i}^2(m_1) \\
&= \frac{1}{nh}\sum W_{1i}^2(x)\left[Y_i - m^{(1)}(x)(X_i - x)\right]^2 \\
&= \frac{1}{nh}\sum W_{1i}^2(x)\left[Y_i - m(X_i)\right]^2 + \frac{2}{nh}\sum W_{1i}^2(x)\left[Y_i - m(X_i)\right]\left[m(X_i) - m^{(1)}(x)(X_i - x)\right] \\
&+ \frac{1}{nh}\sum W_{1i}^2(x)\left[m(X_i) - m^{(1)}(x)(X_i - x)\right]^2 \\
&= J_1 + 2J_2 + J_3,
\end{aligned}
$$

where $J_1 = \sigma^2(x) f^3(x)\omega_3 + o_p(1)$, $J_2 = o_p(1)$ because of similar proof for corresponding parts in Lemma A1. Next,

$$
\begin{aligned}
J_3 &= \frac{m^2(x)}{nh}\sum W_{1i}^2(x) + \frac{2m(x)}{nh}\sum W_{1i}^2(x)\left[m(X_i) - m(x) - m^{(1)}(x)(X_i - x)\right] \\
&+ \frac{1}{nh}\sum W_{1i}^2(x)\left[m(X_i) - m(x) - m^{(1)}(x)(X_i - x)\right]^2 \\
&= m^2(x)J_4 + 2m(x)J_5 + J_6,
\end{aligned}
$$

where $m^2(x)J_4 = m^2(x)f^3(x)\omega_3 + o_p(1)$, since for $j = 0, 1, 2, 3, 4$,

$$
\frac{1}{nh}\sum\left(\frac{X_i - x}{h}\right)^j K_i^2 = E\left[\left(\frac{X_1 - x}{h}\right)^j \frac{1}{h}K_1^2\right] + o_p(1) = f(x)\nu_j + o_p(1),
$$

and for $j = 0, 1, 2, 3, 4$,

$$
T_j = f^2(x)t_j + o_p(1).
$$

Also,

$$
\begin{aligned}
J_5 &= \frac{1}{nh}\sum W_{1i}^2(x)\left[\frac{1}{2}m^{(2)}(x)(X_i-x)^2+o_p\left(h^2\right)\right] \\
&= \frac{m^{(2)}(x)}{2nh}\sum W_{1i}^2(x)(X_i-x)^2+o_p\left(h^2\right)J_4 \\
&= O_p\left(h^2\right)+o_p\left(h^2\right)=o_p(1), \\
J_6 &= \frac{1}{nh}\sum W_{1i}^2(x)\left[\frac{1}{2}m^{(2)}(x)(X_i-x)^2+o_p\left(h^2\right)\right]^2 \\
&= O_p\left(h^4\right)+o_p\left(h^4\right)=o_p(1).
\end{aligned}
$$

■

Lemma A 3 $\frac{1}{nh}\sum_{i=1}^{n}U_{2i}^2(m_2)=\left[\sigma^2(x)+m^2(x)\right]f^3(x)\omega_5+o_p(1)$.

Proof. Write

$$
\begin{aligned}
\frac{1}{nh}\sum U_{2i}^2(m_2) &= \frac{1}{nh}\sum W_{2i}^2(x)\left[Y_i-\frac{1}{2}m^{(2)}(x)(X_i-x)^2\right]^2 \\
&= \frac{1}{nh}\sum W_{2i}^2(x)\left[Y_i-m(X_i)\right]^2 \\
&+ \frac{2}{nh}\sum W_{2i}^2(x)\left[Y_i-m(X_i)\right]\left[m(X_i)-\frac{1}{2}m^{(2)}(x)(X_i-x)^2\right] \\
&+ \frac{1}{nh}\sum W_{2i}^2(x)\left[m(X_i)-\frac{1}{2}m^{(2)}(x)(X_i-x)^2\right]^2 \\
&= J_1+2J_2+J_3,
\end{aligned}
$$

where $J_1=\sigma^2(x)f^3(x)\omega_5+o_p(1)$, $J_2=o_p(1)$ because of similar proof for corresponding parts in Lemma A1. Next, let

$$
\begin{aligned}
A_1 &= m(X_i)-m(x)-m^{(1)}(x)(X_i-x)-\frac{1}{2}m^{(2)}(x)(X_i-x)^2=o_p\left(h^2\right), \\
A_2 &= m(x)+m^{(1)}(x)(X_i-x),
\end{aligned}
$$

then

$$J_3 = \frac{1}{nh}\sum W_{2i}^2(x)\,A_1^2 + \frac{2}{nh}\sum W_{2i}^2(x)\,A_1A_2 + \frac{1}{nh}\sum W_{2i}^2(x)\,A_2^2$$
$$= J_4 + 2J_5 + J_6,$$

where

$$J_4 = \frac{1}{nh}\sum W_{2i}^2(x)\left[o_p\left(h^2\right)\right]^2 = o_p\left(h^4\right),$$
$$J_5 = \frac{m(x)}{nh}\sum W_{2i}^2(x)\left[o_p\left(h^2\right)\right] + \frac{m^{(1)}(x)}{nh}\sum W_{2i}^2(x)\left[o_p\left(h^2\right)(X_i - x)\right] = o_p\left(h^2\right) + o_p\left(h^3\right),$$
$$J_6 = \frac{m^2(x)}{nh}\sum W_{2i}^2(x) + \frac{2m(x)\,m^{(1)}(x)}{nh}\sum W_{2i}^2(x)(X_i - x) + \frac{\left[m^{(1)}(x)\right]^2}{nh}\sum W_{2i}^2(x)(X_i - x)^2$$
$$= J_7 + O_p(h) + O_p\left(h^2\right),$$

and $J_7 = m^2(x)\,f^3(x)\,\omega_5 + o_p(1)$ as J_4 in Lemma A2. So $J_3 = m^2(x)\,f^3(x)\,\omega_5 + o_p(1)$. ■

Lemma A 4 $\frac{1}{nh}\sum_{i=1}^{n} U_{0i}(m_0)\,U_{1i}(m_1) = \sigma^2(x)\,f^3(x)\,\omega_1 + o_p(1)$.

Proof. Write

$$\frac{1}{nh}\sum U_{0i}(m_0)\,U_{1i}(m_1)$$
$$= \frac{1}{nh}\sum W_{0i}(x)\,W_{1i}(x)\,[Y_i - m(x)]\left[Y_i - m^{(1)}(x)(X_i - x)\right]$$
$$= \frac{1}{nh}\sum W_{0i}(x)\,W_{1i}(x)\,[Y_i - m(X_i)]^2$$
$$+ \frac{1}{nh}\sum W_{0i}(x)\,W_{1i}(x)\,[Y_i - m(X_i)]\left[m(X_i) - m^{(1)}(x)(X_i - x)\right]$$
$$+ \frac{1}{nh}\sum W_{0i}(x)\,W_{1i}(x)\,[Y_i - m(X_i)]\,[m(X_i) - m(x)]$$
$$+ \frac{1}{nh}\sum W_{0i}(x)\,W_{1i}(x)\,[m(X_i) - m(x)]\left[m(X_i) - m^{(1)}(x)(X_i - x)\right]$$
$$= J_1 + J_2 + J_3 + J_4,$$

where $J_1 = \sigma^2(x) f^3(x)\omega_1 + o_p(1)$, $J_2 = J_3 = o_p(1)$ because of similar proof for corresponding parts in Lemma A1, and

$$\begin{aligned} J_4 &= \frac{1}{nh}\sum W_{0i}(x) W_{1i}(x) [m(X_i) - m(x)]^2 \\ &+ \frac{1}{nh}\sum W_{0i}(x) W_{1i}(x) [m(X_i) - m(x)] \left[m(x) - m^{(1)}(x)(X_i - x)\right] \\ &= J_{41} + J_{42}, \end{aligned}$$

where $J_{41} = o_p(1)$ as J_3 in Lemma A1, and

$$\begin{aligned} J_{42} &= \frac{1}{nh}\sum W_{0i}(x) W_{1i}(x) \left[m^{(1)}(x)(X_i - x) + o_p(h)\right] \left[m(x) - m^{(1)}(x)(X_i - x)\right] \\ &= \frac{m(x)}{nh}\sum W_{0i}(x) W_{1i}(x) \left[m^{(1)}(x)(X_i - x) + o_p(h)\right] \\ &- \frac{m^{(1)}(x)}{nh}\sum W_{0i}(x) W_{1i}(x) \left[m^{(1)}(x)(X_i - x) + o_p(h)\right](X_i - x) \\ &= O_p(h) + o_p(h) + O_p\left(h^2\right) + o_p\left(h^2\right) = o_p(1). \end{aligned}$$

■

Lemma A 5 $\frac{1}{nh}\sum_{i=1}^{n} U_{0i}(m_0) U_{2i}(m_2) = \sigma^2(x) f^3(x)\omega_2 + o_p(1)$.

Proof. Write

$$
\begin{aligned}
&\frac{1}{nh}\sum U_{0i}\left(m_0\right)U_{2i}\left(m_2\right)\\
&=\frac{1}{nh}\sum W_{0i}\left(x\right)W_{2i}\left(x\right)\left[Y_i-m\left(x\right)\right]\left[Y_i-\frac{1}{2}m^{(2)}\left(x\right)\left(X_i-x\right)^2\right]\\
&=\frac{1}{nh}\sum W_{0i}\left(x\right)W_{2i}\left(x\right)\left[Y_i-m\left(X_i\right)\right]^2\\
&+\frac{1}{nh}\sum W_{0i}\left(x\right)W_{2i}\left(x\right)\left[Y_i-m\left(X_i\right)\right]\left[m\left(X_i\right)-\frac{1}{2}m^{(2)}\left(x\right)\left(X_i-x\right)^2\right]\\
&+\frac{1}{nh}\sum W_{0i}\left(x\right)W_{2i}\left(x\right)\left[Y_i-m\left(X_i\right)\right]\left[m\left(X_i\right)-m\left(x\right)\right]\\
&+\frac{1}{nh}\sum W_{0i}\left(x\right)W_{2i}\left(x\right)\left[m\left(X_i\right)-m\left(x\right)\right]\left[m\left(X_i\right)-\frac{1}{2}m^{(2)}\left(x\right)\left(X_i-x\right)^2\right]\\
&=J_1+J_2+J_3+J_4,
\end{aligned}
$$

where $J_1=\sigma^2\left(x\right)f^3\left(x\right)\omega_2+o_p\left(1\right)$, $J_2=J_3=o_p\left(1\right)$ because of similar proof for corresponding parts in Lemma A1, and

$$
\begin{aligned}
J_4&=\frac{1}{nh}\sum W_{0i}\left(x\right)W_{2i}\left(x\right)\left[m\left(X_i\right)-m\left(x\right)\right]^2\\
&+\frac{1}{nh}\sum W_{0i}\left(x\right)W_{2i}\left(x\right)\left[m\left(X_i\right)-m\left(x\right)\right]\left[m\left(x\right)-\frac{1}{2}m^{(2)}\left(x\right)\left(X_i-x\right)^2\right]\\
&=J_{41}+J_{42},
\end{aligned}
$$

where $J_{41}=o_p\left(1\right)$ as J_3 in Lemma A1, and

$$
\begin{aligned}
J_{42}&=\frac{1}{nh}\sum W_{0i}\left(x\right)W_{2i}\left(x\right)\left[m^{(1)}\left(x\right)\left(X_i-x\right)+o_p\left(h\right)\right]\left[m\left(x\right)-\frac{1}{2}m^{(2)}\left(x\right)\left(X_i-x\right)^2\right]\\
&=\frac{m\left(x\right)}{nh}\sum W_{0i}\left(x\right)W_{2i}\left(x\right)\left[m^{(1)}\left(x\right)\left(X_i-x\right)+o_p\left(h\right)\right]\\
&-\frac{m^{(2)}\left(x\right)}{2nh}\sum W_{0i}\left(x\right)W_{2i}\left(x\right)\left[m^{(1)}\left(x\right)\left(X_i-x\right)+o_p\left(h\right)\right]\left(X_i-x\right)^2\\
&=O_p\left(h\right)+o_p\left(h\right)+O_p\left(h^3\right)+o_p\left(h^3\right)=o_p\left(1\right).
\end{aligned}
$$

■

Lemma A 6 $\frac{1}{nh}\sum_{i=1}^{n} U_{1i}(m_1) U_{2i}(m_2) = [\sigma^2(x) + m^2(x)] f^3(x) \omega_4 + o_p(1)$.

Proof. Write

$$\begin{aligned}
&\frac{1}{nh}\sum U_{1i}(m_1) U_{2i}(m_2) \\
&= \frac{1}{nh}\sum W_{1i}(x) W_{2i}(x) \left[Y_i - m^{(1)}(x)(X_i - x)\right] \left[Y_i - \frac{1}{2} m^{(2)}(x)(X_i - x)^2\right] \\
&= \frac{1}{nh}\sum W_{1i}(x) W_{2i}(x) [Y_i - m(X_i)]^2 \\
&+ \frac{1}{nh}\sum W_{1i}(x) W_{2i}(x) [Y_i - m(X_i)] \left[m(X_i) - \frac{1}{2} m^{(2)}(x)(X_i - x)^2\right] \\
&+ \frac{1}{nh}\sum W_{1i}(x) W_{2i}(x) [Y_i - m(X_i)] \left[m(X_i) - m^{(1)}(x)(X_i - x)\right] \\
&+ \frac{1}{nh}\sum W_{1i}(x) W_{2i}(x) \left[m(X_i) - m^{(1)}(x)(X_i - x)\right] \left[m(X_i) - \frac{1}{2} m^{(2)}(x)(X_i - x)^2\right] \\
&= J_1 + J_2 + J_3 + J_4,
\end{aligned}$$

where $J_1 = \sigma^2(x) f^3(x) \omega_4 + o_p(1)$, $J_2 = J_3 = o_p(1)$ because of similar proof for corresponding parts in Lemma A1, and

$$\begin{aligned}
J_4 &= \frac{1}{nh}\sum W_{1i}(x) W_{2i}(x) \left[m(X_i) - m(x) - m^{(1)}(x)(X_i - x) + m(x)\right] \\
&\left[m(X_i) - m(x) - m^{(1)}(x)(X_i - x) - \frac{1}{2} m^{(2)}(x)(X_i - x)^2 + m(x) + m^{(1)}(x)(X_i - x)\right] \\
&= \frac{1}{nh}\sum W_{1i}(x) W_{2i}(x) [o_p(h) + m(x)] \left[o_p\left(h^2\right) + m(x) + m^{(1)}(x)(X_i - x)\right] \\
&= \frac{m^2(x)}{nh}\sum W_{1i}(x) W_{2i}(x) + \frac{m(x) m^{(1)}(x)}{nh}\sum W_{1i}(x) W_{2i}(x)(X_i - x) + o_p(h) \\
&= J_5 + O_p(h) + o_p(h),
\end{aligned}$$

where $J_5 = m^2(x) f^3(x) \omega_4 + o_p(1)$ as J_4 in Lemma A2. ■

Proof of Theorem 1

Proof. Write $\lambda(m) = \rho\theta$ where $\rho \geqslant 0$ and $\|\theta\| = 1$. Also denote $\overline{U} = \sum_{j=0}^{2} \left[\frac{1}{nh}\sum_{i=1}^{n} U_{ji}(m_j)\right]$ for $j = 0, 1, 2$, and $\overline{\overline{U}} = \frac{1}{nh}\sum_{i=1}^{n} U_i(m) U_i(m)^\intercal$. Note that

$$p_i = \frac{1}{n\left(1 + \rho\theta^\intercal U_i(m)\right)} \in [0, 1],$$

from which we have $1 + \rho\theta^\intercal U_i(m) > 0$. From the three EL weighted estimating equations, $\sum_{i=1}^{n} p_i U_i(m) = 0$, we have

$$\begin{aligned}
0 &= \left\| \frac{1}{nh}\sum\nolimits_{i=1}^{n} \frac{U_i(m)}{1 + \rho\theta^\intercal U_i(m)} \right\| \\
&\geqslant \left| \frac{1}{nh}\sum\nolimits_{i=1}^{n} \frac{\theta^\intercal U_i(m)}{1 + \rho\theta^\intercal U_i(m)} \right| \\
&= \left| \theta^\intercal \left(\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i(m) - \rho\frac{1}{nh}\sum\nolimits_{i=1}^{n} \frac{U_i(m)\left[\theta^\intercal U_i(m)\right]}{1 + \rho\theta^\intercal U_i(m)} \right) \right| \\
&\geqslant \rho\theta^\intercal \left(\frac{1}{nh}\sum\nolimits_{i=1}^{n} \frac{U_i(m) U_i(m)^\intercal}{1 + \rho\theta^\intercal U_i(m)} \right)\theta - \left|\overline{U}\right| \\
&\geqslant \frac{\rho}{1 + \rho Z_n}\theta^\intercal \overline{\overline{U}}\theta - \left|\overline{U}\right|,
\end{aligned}$$

where in the right hand side of the last inequality, $Z_n = \max\limits_{1 \leqslant i \leqslant n} \|U_i(m)\|$ so $Z_n \geqslant \theta^\intercal U_i(m)$ for each i. Therefore

$$\frac{\rho}{1 + \rho Z_n}\theta^\intercal \overline{\overline{U}}\theta \leqslant \left|\overline{U}\right|$$

implies

$$\rho\left(\theta^\intercal \overline{\overline{U}}\theta - Z_n\left|\overline{U}\right|\right) \leqslant \left|\overline{U}\right|.$$

Since (i) by Lemma 2, $\left|\overline{U}\right| = O_p\left((nh)^{-1/2} + h^3\right)$, (ii) by Lemma 3, $\overline{\overline{U}} = \Omega_U + o_p(1)$, (iii) $Z_n = o_p\left(n^{1/s}\right)$ from the assumption of $E\left|Y_i\right|^s < \infty$ for $s > 2$, we have

$$\|\lambda(m)\| = \rho = O_p\left((nh)^{-1/2} + h^3\right).$$

Moreover, by a Taylor expansion of the EL weighted estimating equations at $\lambda = 0$, we have

$$0 = \frac{1}{nh} \sum\nolimits_{i=1}^{n} U_i(m) - \left[\frac{1}{nh} \sum\nolimits_{i=1}^{n} U_i(m) U_i(m)^{\intercal} \right] \lambda(m) + o(\|\lambda(m)\|),$$

hence

$$\lambda(m) = \left[\frac{1}{nh} \sum\nolimits_{i=1}^{n} U_i(m) U_i(m)^{\intercal} \right]^{-1} \left[\frac{1}{nh} \sum\nolimits_{i=1}^{n} U_i(m) \right] + o_p\left((nh)^{-1/2} + h^3 \right).$$

■

Proof of Lemma 4

Proof. Without losing generality, for the saddlepoint $\left(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{\nu}\right)$ of $G_n^*(\beta, \lambda, \nu)$, we only consider the case $\widetilde{\nu} = 0$. That is, the inequality constraints $\underline{b} \leqslant \beta \leqslant \overline{b}$ are not binding in the large sample context. Therefore the "inner" optimization problem

$$\max_{\lambda \in \Lambda, \nu \in \mathbb{R}_+^6} \sum\nolimits_{i=1}^{n} \log\left(1 + \lambda(\beta)^{\intercal} U_i(\beta)\right) + n\underline{\nu}^{\intercal}\left(\underline{b} - \beta\right) + n\overline{\nu}^{\intercal}\left(\beta - \overline{b}\right)$$

is simplified as

$$l(\beta) = \max_{\lambda \in \Lambda} \sum\nolimits_{i=1}^{n} \log\left(1 + \lambda(\beta)^{\intercal} U_i(\beta)\right).$$

We point out that the following proof also holds without this simplification.

Denote $\overline{\beta} = \left(\overline{\beta}_0, \overline{\beta}_1, \overline{\beta}_2,\right)^{\intercal}$, and for $j = 0, 1, 2$, $\overline{\beta}_j = m_j - h^{2-j} u_j$, where $u_j \in \mathbb{R}$ is such that $u = (u_0, u_1, u_2)^{\intercal}$, $\|u\| = 1$. First, following the argument in the proof of Lemma 1 in Qin and Lawless(1994), we establish a lower bound for $l(\beta)$ at $\overline{\beta}$. To do this, notice that:

(i) by Lemma 2,

$$\frac{1}{nh}\sum_{i=1}^{n}U_{ji}\left(\overline{\beta}_j\right) = h^2 u_j\left[\frac{1}{nh}\sum_{i=1}^{n}W_{ji}\left(x\right)\left(\frac{X_i - x}{h}\right)^j\right] + \frac{1}{nh}\sum_{i=1}^{n}U_{ji}\left(m_j\right)$$
$$= h^2 u_j f^3\left(x\right)D + o_p\left(h^2\right) + O_p\left(\left(nh\right)^{-1/2} + h^3\right)$$
$$= h^2 u_j f^3\left(x\right)D + o_p\left(h^2\right),$$

since $D_n = \frac{1}{nh}\sum_{i=1}^{n}W_{ji}\left(x\right)\left(\left(X_i - x\right)/h\right)^j = f^3\left(x\right)D + o_p\left(1\right)$;

(ii) by Lemma 3,

$$\frac{1}{nh}\sum_{i=1}^{n}U_i\left(\overline{\beta}\right)U_i\left(\overline{\beta}\right)^{\intercal} = \frac{1}{nh}\sum_{i=1}^{n}U_i\left(m\right)U_i\left(m\right)' + o_p\left(1\right) = \Omega_U + o_p\left(1\right),$$

where

$$\frac{1}{nh}\sum_{i=1}^{n}U_{0i}^2\left(\overline{\beta}_0\right) = \frac{1}{nh}\sum_{i=1}^{n}U_{0i}^2\left(m_0\right) + O_p\left(h^3\right),$$
$$\frac{1}{nh}\sum_{i=1}^{n}U_{1i}^2\left(\overline{\beta}_1\right) = \frac{1}{nh}\sum_{i=1}^{n}U_{1i}^2\left(m_1\right) + O_p\left(h^2\right),$$
$$\frac{1}{nh}\sum_{i=1}^{n}U_{2i}^2\left(\overline{\beta}_2\right) = \frac{1}{nh}\sum_{i=1}^{n}U_{2i}^2\left(m_2\right) + O_p\left(h^2\right),$$
$$\frac{1}{nh}\sum_{i=1}^{n}U_{0i}\left(\overline{\beta}_0\right)U_{1i}\left(\overline{\beta}_1\right) = \frac{1}{nh}\sum_{i=1}^{n}U_{0i}\left(m_0\right)U_{1i}\left(m_1\right) + O_p\left(h^2\right),$$
$$\frac{1}{nh}\sum_{i=1}^{n}U_{0i}\left(\overline{\beta}_0\right)U_{2i}\left(\overline{\beta}_2\right) = \frac{1}{nh}\sum_{i=1}^{n}U_{0i}\left(m_0\right)U_{2i}\left(m_2\right) + O_p\left(h^2\right),$$
$$\frac{1}{nh}\sum_{i=1}^{n}U_{1i}\left(\overline{\beta}_1\right)U_{2i}\left(\overline{\beta}_2\right) = \frac{1}{nh}\sum_{i=1}^{n}U_{1i}\left(m_1\right)U_{2i}\left(m_2\right) + O_p\left(h^2\right).$$

As in the proof of Theorem 1, from (i) and (ii), we have

$$\lambda\left(\overline{\beta}\right) = \left[\frac{1}{nh}\sum_{i=1}^{n}U_i\left(\overline{\beta}\right)U_i\left(\overline{\beta}\right)^{\intercal}\right]^{-1}\left[\frac{1}{nh}\sum_{i=1}^{n}U_i\left(\overline{\beta}\right)\right] + o_p\left(h^2\right) \tag{20}$$
$$= O_p\left(h^2\right).$$

Therefore by a Taylor expansion at $\lambda = 0$ and by (20),

$$
\begin{aligned}
l\left(\overline{\beta}\right) &= nh\left\{\lambda\left(\overline{\beta}\right)^{\intercal}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_i\left(\overline{\beta}\right)\right]-\frac{1}{2}\lambda\left(\overline{\beta}\right)^{\intercal}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_i\left(\overline{\beta}\right)U_i\left(\overline{\beta}\right)^{\intercal}\right]\lambda\left(\overline{\beta}\right)+o_p\left(\left\|\lambda\left(\overline{\beta}\right)\right\|^2\right)\right\} \\
&= \frac{nh}{2}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_i\left(\overline{\beta}\right)\right]^{\intercal}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_i\left(\overline{\beta}\right)U_i\left(\overline{\beta}\right)^{\intercal}\right]^{-1}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_i\left(\overline{\beta}\right)\right]+o_p\left(nh^5\right) \\
&= \frac{nh}{2}\left[h^2uf^3\left(x\right)D+o_p\left(h^2\right)\right]^{\intercal}\Omega_U^{-1}\left[h^2uf^3\left(x\right)D+o_p\left(h^2\right)\right]+o_p\left(nh^5\right) \\
&\geqslant nh^5\left(c-\epsilon\right),
\end{aligned}
$$

where $c-\epsilon>0$ and c is the smallest eigenvalue of $f^6\left(x\right)D^2\left(u^{\intercal}\Omega_U^{-1}u\right)$.

Similarly,

$$
\begin{aligned}
l\left(m\right) =& \frac{nh}{2}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_i\left(m\right)\right]^{\intercal}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_i\left(m\right)U_i\left(m\right)'\right]^{-1}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_i\left(m\right)\right] \\
&+o_p\left(nh\left(\left(nh\right)^{-1/2}+h^3\right)^2\right) \\
&= \frac{nh}{2}O_p\left(\left(nh\right)^{-1/2}+h^3\right)^{\intercal}\Omega_U^{-1}O_p\left(\left(nh\right)^{-1/2}+h^3\right)+o_p\left(nh^7\right) \\
&= O_p\left(nh^7\right).
\end{aligned}
$$

Since $l\left(\beta\right)$ is continuous in the interior of

$$
\left\{\beta\left(x\right):\left|\beta_j\left(x\right)-m_j\left(x\right)\right|\leqslant h^{2-j},\ j=0,1,2\right\}, \tag{21}
$$

$l\left(\beta\right)$ attains minimum value $\widetilde{\beta}$ in (21). Moreover, we have

$$
\begin{aligned}
\left.\frac{\partial l\left(\beta\right)}{\partial\beta}\right|_{\beta=\widetilde{\beta}} &= \left(\partial\lambda\left(\beta\right)^{\intercal}/\partial\beta\right)\sum\nolimits_{i=1}^{n}\left.\frac{U_i\left(\beta\right)}{1+\lambda\left(\beta\right)^{\intercal}U_i\left(\beta\right)}\right|_{\beta=\widetilde{\beta}} \\
&+\sum\nolimits_{i=1}^{n}\left.\frac{\left(\partial U_i\left(\beta\right)^{\intercal}/\partial\beta\right)\lambda\left(\beta\right)}{1+\lambda\left(\beta\right)^{\intercal}U_i\left(\beta\right)}\right|_{\beta=\widetilde{\beta}} \\
&= 0
\end{aligned} \tag{22}
$$

Note that we already have

$$g_{1n}\left(\widetilde{\beta},\widetilde{\lambda}\right)=\frac{1}{nh}\sum\nolimits_{i=1}^{n}\frac{U_i\left(\beta\right)}{1+\lambda\left(\beta\right)^{\intercal}U_i\left(\beta\right)}\Big|_{\beta=\widetilde{\beta}}=0$$

as discussed in Remark 1. Therefore by (22),

$$\sum\nolimits_{i=1}^{n}\frac{\left(\partial U_i\left(\beta\right)^{\intercal}/\partial\beta\right)\lambda\left(\beta\right)}{1+\lambda\left(\beta\right)^{\intercal}U_i\left(\beta\right)}\Big|_{\beta=\widetilde{\beta}}=0,$$

where $\partial U_i\left(\beta\right)^{\intercal}/\partial\beta = diag\left\{-W_{ji}\left(x\right)\left(X_i-x\right)^j\right\}$. Denote $H_3 = diag\left\{h^j\right\}$, then $D_i\left(x\right) = \left(\partial U_i\left(\beta\right)^{\intercal}/\partial\beta\right)H_3^{-1}$ and

$$g_{2n}\left(\widetilde{\beta},\widetilde{\lambda}\right)=\frac{1}{nh}\sum\nolimits_{i=1}^{n}\frac{D_i\left(x\right)\lambda\left(\beta\right)}{1+\lambda\left(\beta\right)^{\intercal}U_i\left(\beta\right)}\Big|_{\beta=\widetilde{\beta}}=0.$$

■

Proof of Theorem 2

Proof. Taking derivatives of $g_{1n}\left(\beta,\lambda\right)$ and $g_{2n}\left(\beta,\lambda\right)$ and evaluating at $(m,0)$, we have

$$\begin{aligned}
\frac{\partial g_{1n}\left(m,0\right)}{\partial\beta^{\intercal}} &= \frac{1}{nh}\sum\nolimits_{i=1}^{n}\left(\partial U_i\left(\beta\right)^{\intercal}/\partial\beta\right)=\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n}D_i\left(x\right)\right]H_3,\\
\frac{\partial g_{1n}\left(m,0\right)}{\partial\lambda^{\intercal}} &= -\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_i\left(m\right)U_i\left(m\right)^{\intercal},\\
\frac{\partial g_{2n}\left(m,0\right)}{\partial\beta^{\intercal}} &= 0,\\
\frac{\partial g_{2n}\left(m,0\right)}{\partial\lambda^{\intercal}} &= \frac{1}{nh}\sum\nolimits_{i=1}^{n}D_i\left(x\right).
\end{aligned}$$

Note that $\frac{1}{nh}\sum_{i=1}^{n} D_i(x) = -D_n I_3$ since $D_n = \frac{1}{nh}\sum_{i=1}^{n} W_{ji}(x)\left((X_i - x)/h\right)^j$ for $j = 0, 1, 2$. By Taylor expanding $g_{1n}\left(\widetilde{\beta}, \widetilde{\lambda}\right)$ and $g_{2n}\left(\widetilde{\beta}, \widetilde{\lambda}\right)$ at $(m, 0)$, we have

$$
\begin{aligned}
0 &= g_{1n}\left(\widetilde{\beta}, \widetilde{\lambda}\right) \\
&= g_{1n}(m, 0) + \frac{\partial g_{1n}(m, 0)}{\partial \beta^{\intercal}}\left(\widetilde{\beta} - m\right) + \frac{\partial g_{1n}(m, 0)}{\partial \lambda^{\intercal}}\left(\widetilde{\lambda} - 0\right) + o_p(\delta) \\
&= \frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i(m) - D_n H_3\left(\widetilde{\beta} - m\right) - \left[\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i(m)\, U_i(m)^{\intercal}\right]\left(\widetilde{\lambda} - 0\right) + o_p(\delta), \\
0 &= g_{2n}\left(\widetilde{\beta}, \widetilde{\lambda}\right) \\
&= g_{2n}(m, 0) + \frac{\partial g_{2n}(m, 0)}{\partial \beta^{\intercal}}\left(\widetilde{\beta} - m\right) + \frac{\partial g_{2n}(m, 0)}{\partial \lambda^{\intercal}}\left(\widetilde{\lambda} - 0\right) + o_p(\delta) \\
&= 0 + 0\left(\widetilde{\beta} - m\right) - D_n I_3\left(\widetilde{\lambda} - 0\right) + o_p(\delta),
\end{aligned}
$$

where $\delta = \left\|H_3\left(\widetilde{\beta} - m\right)\right\| + \left\|\widetilde{\lambda}\right\|$. Hence we have

$$
\begin{pmatrix} H_3\left(\widetilde{\beta} - m\right) \\ \widetilde{\lambda} \end{pmatrix} = \Omega_g^{-1} \begin{pmatrix} \frac{1}{nh}\sum_{i=1}^{n} U_i(m) + o_p(\delta) \\ o_p(\delta) \end{pmatrix},
$$

where

$$
\Omega_g = \begin{pmatrix} D_n I_3 & \frac{1}{nh}\sum_{i=1}^{n} U_i(m)\, U_i(m)^{\intercal} \\ 0 & D_n I_3 \end{pmatrix} \xrightarrow{p} \begin{pmatrix} f^3(x)\, D I_3 & -\Omega_U \\ 0 & f^3(x)\, D I_3 \end{pmatrix}.
$$

By this and $\frac{1}{nh}\sum_{i=1}^{n} U_i(m) = O_p\left((nh)^{-1/2} + h^3\right)$, we know that $\delta = O_p\left((nh)^{-1/2} + h^3\right)$. For the limit distribution of $\widetilde{\beta}$, we have

$$
H_3\left(\widetilde{\beta} - m\right) = D_n^{-1}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i(m)\right] + o_p\left((nh)^{-1/2} + h^3\right),
$$

that is, for $j = 0, 1, 2,$

$$\widetilde{\beta}_j(x) - m_j(x) = h^{-j} \frac{\frac{1}{nh}\sum_{i=1}^{n} W_{ji}(x)\left[Y_i - m_j(x)(X_i - x)^j\right]}{\frac{1}{nh}\sum_{i=1}^{n} W_{ji}(x)\left(\frac{X_i - x}{h}\right)^j} + h^{-j} o_p\left((nh)^{-1/2} + h^3\right)$$
$$= \widehat{\beta}_j(x) - m_j(x) + o_p\left(\left(nh^{1+2j}\right)^{-1/2} + h^{3-j}\right).$$

Thus

$$\sqrt{nh^{1+2j}}\left(\widetilde{\beta}_j(x) - m_j(x)\right) = \sqrt{nh^{1+2j}}\left(\widehat{\beta}_j(x) - m_j(x)\right) + \sqrt{nh^{1+2j}} o_p\left(\left(nh^{1+2j}\right)^{-1/2} + h^{3-j}\right)$$
$$= \sqrt{nh^{1+2j}}\left(\widehat{\beta}_j(x) - m_j(x)\right) + o_p\left(1 + \sqrt{nh^7}\right).$$

■

Appendix B: Figures

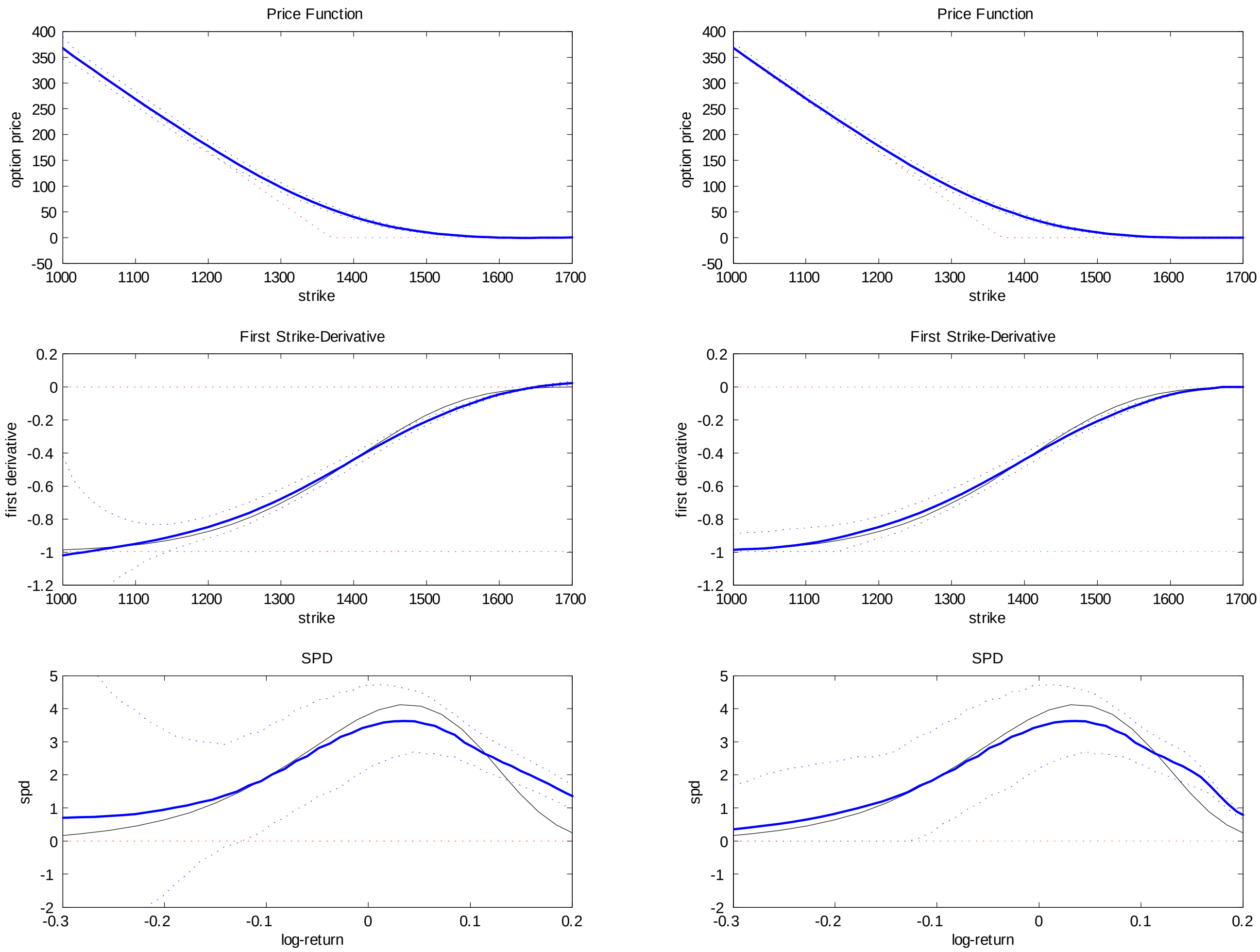

Figure 1: Simulation results for $n = 25$

Left column from top to bottom: Unconstrained estimates $\widehat{C}(X)$, $\widehat{C'}(X)$, and $e^{r_{t,\tau}\tau}\widehat{C''}(X)$. Right column from top to bottom: Constrained estimates $\widetilde{C}(X)$, $\widetilde{C'}(X)$, and $e^{r_{t,\tau}\tau}\widetilde{C''}(X)$. Legend: Solid black line: True function; Solid blue line: Average estimate; Dot blue line: 95% confidence band; Dot red line: Constraints.

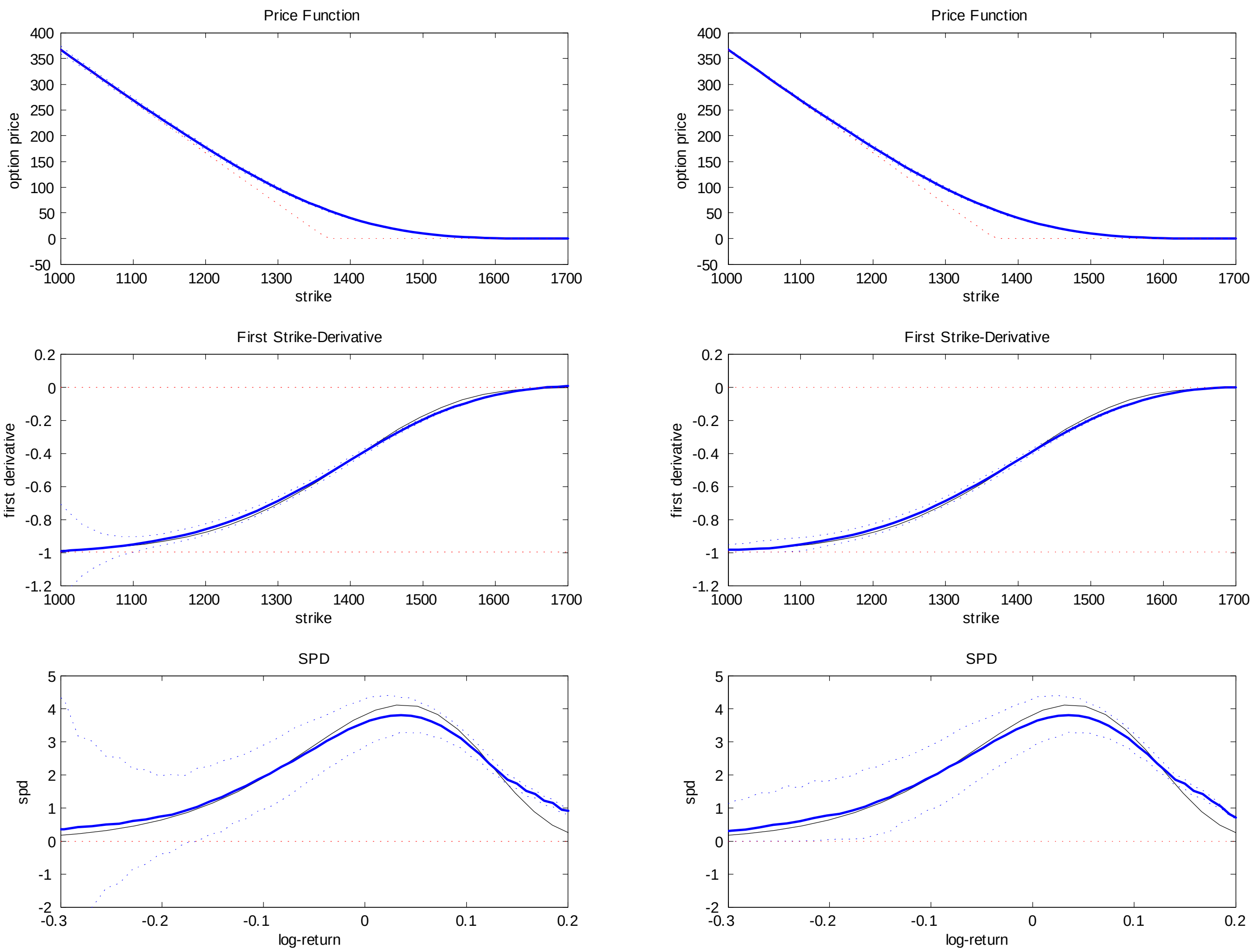

Figure 2: Simulation results for $n = 250$

Left column from top to bottom: Unconstrained estimates $\widehat{C}(X)$, $\widehat{C'}(X)$, and $e^{r_{t,\tau}\tau}\widehat{C''}(X)$. Right column from top to bottom: Constrained estimates $\widetilde{C}(X)$, $\widetilde{C'}(X)$, and $e^{r_{t,\tau}\tau}\widetilde{C''}(X)$. Legend: Solid black line: True function; Solid blue line: Average estimate; Dot blue line: 95% confidence band; Dot red line: Constraints.

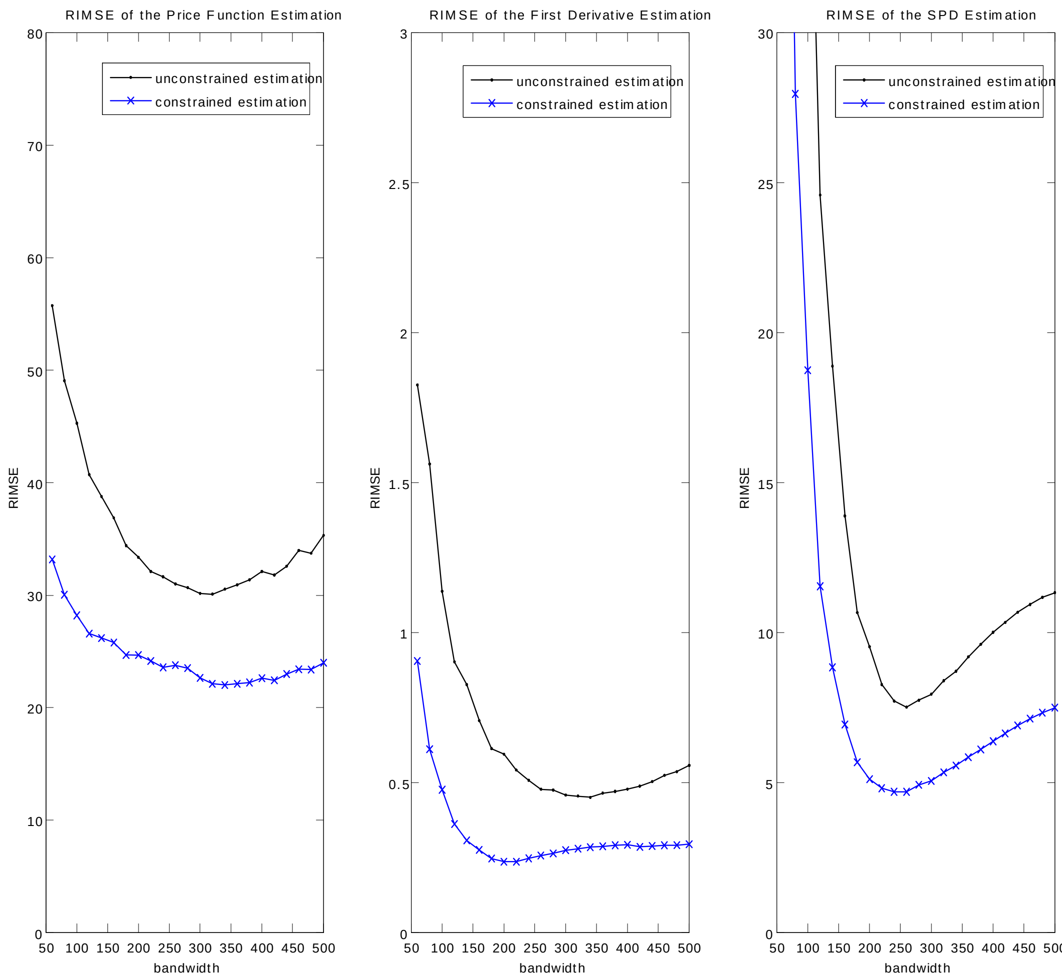

Figure 3: Root of integrated mean squared errors for different bandwidths

Left: RIMSE for the option pricing function estimation, unconstrained and constrained; Middle: RIMSE for the first derivative estimation, unconstrained and constrained; Right: RIMSE for the SPD estimation, unconstrained and constrained. Legend: Black line with dot: Unconstrained estimates; Blue line with cross: Constrained estimates.

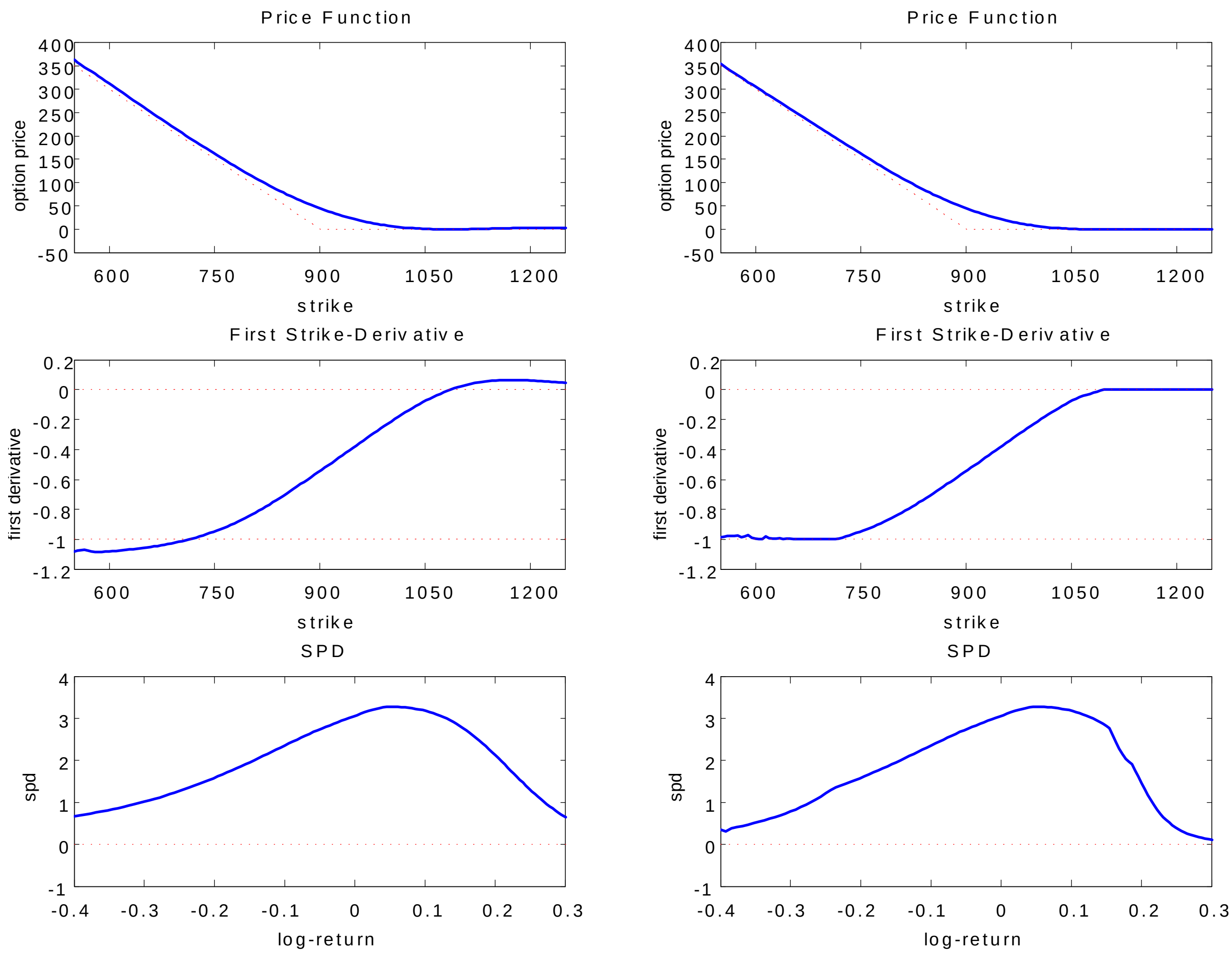

Figure 4: Estimation results of S&P 500 options, July expiration on May 18, 2009

Left column from top to bottom: Unconstrained estimates $\widehat{C}\,(X)$, $\widehat{C'}\,(X)$, and $e^{r_{t,\tau}\tau}\widehat{C''}\,(X)$. Right column from top to bottom: Constrained estimates $\widetilde{C}\,(X)$, $\widetilde{C'}\,(X)$, and $e^{r_{t,\tau}\tau}\widetilde{C''}\,(X)$. Legend: Solid blue line: Estimate; Dot red line: Constraints