The sign of a mean regression: characterisation, estimation and applications*

Spyros Skouras

Athens University of Economics and Business

spyros@santafe.edu

Abstract

There exist a number of important applications in which interest centers on the sign rather than the overall shape of a mean regression. These include certain decision-making problems as well as the problems of calibration and econometric equation inversion. We show that the sign of a mean regression may be viewed as a generalisation of a quantile regression for the sign of the regressand. Using this, we provide necessary and sufficient conditions for a parametric model to have the same sign as a mean regression and use this to construct three related estimators that require only a very weak specification condition for the estimated model's sign to be consistent. Further properties of one of the estimators are studied in a simulation.

Key Words: Regression sign, regression zeros, robust consistent estimation, quantile regression, decision rule estimation, calibration, econometric equation inversion.

JEL: C13, C14, C25, C44, G11.

* Acknowledgements to be included.

1 Introduction

Consider the random variables Y and X such that:

$$\begin{aligned} Y &= \mu_Y(X) + U, \\ E(U|x) &= 0, \end{aligned} \tag{1}$$

where $\mu_Y : \mathcal{X} \to \mathbb{R}$ is an unknown and for now unrestricted mean regression, U is an unobserved scalar disturbance term and $E(U|x)$ is the expectation function of U conditional on realisations x of X. The 'sign of the mean regression μ_Y' is an indicator function $I_{\mu_Y}(X) \equiv I(\mu_Y(X))$ where $I(z) = 1$ if $z > 0$ and $I(z) = 0$ otherwise.[1] The problem we deal with in this paper is the estimation of a model the sign of which converges to I_{μ_Y} even when little is known about the behaviour of $\mu_Y(X)$. By redefining Y as $Y - a$ for a constant a, a solution to this problem also solves the problem of estimating a model that provides an indicator of whether $\mu_Y(X)$ exceeds the threshold a.

Although this seems to have gone unnoticed, there are a number of applications in which the sign of a mean regression is the object of interest. In particular, many decision problems with discrete choice sets or discrete solutions can be solved if the sign of a mean regression is known. We provide a simple example of a job search decision (an optimal stopping problem) in which the choice set is discrete (accept or reject a job at a given wage) and the optimal decision is a function of the sign of a mean regression. We also discuss the decision problem of a risk neutral investor as an example of a problem with a discrete solution, and in a companion paper (Skouras 1998) find that the methods herein proposed are very successful in practice. Finally, we show that the classic problems of *calibration* (Eisenhart 1939) and *econometric equation inversion* (Hendry and Ericsson 1991) can be seen as problems requiring the estimation of a mean regression's zeros and discuss how such estimates can be obtained from estimates of signs.

The most straightforward approach to estimating the sign of a mean regression is to view the sign as a particular function of the parameters of a model for the mean regression. Estimation of the sign can thereby be viewed as a special case of the problem of estimating a function of model parameters rather than the parameters themselves. It is well known that functions of good estimates for parameters can lead to poor estimates for those functions of the parameters, particularly in finite samples.[2] In response to this problem with 'naive' estimators, a number of (primarily Bayesian) techniques for estimating arbitrary functions of model parameters have been developed which however are quite complicated to implement (Zellner 1978, Park and Zellner 1979).

[1] This definition of the sign of μ_Y distinguishes only between positive and non-positive values of μ_Y. In Section 5 we will also consider situations where the further distinction between negative and zero values of μ_Y must be made.

[2] For example, one might consider using the reciprocal of a maximum likelihood estimate for the mean of a random variable as an estimate of the reciprocal of that mean. However, such an estimate typically does not have finite moments and has infinite risk relative to a number of standard loss functions.

The insight on which this paper builds is that because the sign is a *many-to-one* function, models that are misspecified may nevertheless have a sign that is correctly specified. We derive necessary and sufficient conditions for a model's sign to coincide with that of a mean regression and find that they are rather weak. Somewhat suprisingly, these conditions characterise the sign of a mean regression as a generalised form of quantile regression for a binary variable[3] (see e.g. Manski and Thompson 1989) and we explore this relationship as much as possible. The conditions are stated as moment extremum conditions with sample analogs leading to two estimators that generalise the maximum score estimator for binary median regression models of Manski (1975) and another that generalises the smoothed maximum score estimator of Horowitz (1992). The advantage of these estimators relative to naive/Bayesian 'likelihood-based' counterparts is that they are far more robust, in that they are consistent under far weaker conditions on model specification. A disadvantage is that, like the maximum score estimator, their further properties remain analytically intractable.

Estimation of the sign is closely related to estimation of the zeros of a mean regression - a problem that has previously been addressed by Härdle and Nixdorf (1987) and Tsybakov (1988). These papers build on the observation that nonparametric estimation of a mean regression function involves a great deal of learning that is redundant when only its zeros are of interest. Exploiting this fact, nonparametric estimators are proposed that are computationally cheaper because they only learn the mean regression's relevant attribute. Our estimators provide a semiparametric counterpart to this approach which will be preferable in the usual situations; for example, our estimators do not require assuming continuity and smoothness of μ_Y or low dimensionality of X.

In the next section we introduce and discuss the assumption that the semiparametric model for μ_Y is correctly specified for its sign (but not necessarily for μ_Y). We characterise the correct parametrisation in terms of necessary and sufficient extremum conditions it must satisfy and show that this characterisation provides an interpretation of the sign of a mean regression as a generalised form of binary quantile regression. In Section 3 sample analogs of these conditions are shown to lead to consistent estimates for the sign of the mean regression. In Section 4 we discuss further properties of our estimators and reasons for which they may be better than standard estimators even when the model is correctly specified for the conditional mean. In Section 5 we motivate our focus on the sign of the mean regression by discussing a number of significant applications in which this sign is the object of interest. Section 6 concludes and discusses useful directions for future research.

2 Models for the sign of a mean regression

2.1 Model specification assumptions

Our analysis will be restricted to the situation where a potentially misspecified parametric model $\mathcal{G}$ for $F_{Y|X}$ is postulated, taking the form:

$$\mathcal{G} = \left\{ G_{Y|[X,c]}, c \in B \subseteq \mathbb{R}^k \right\}, \tag{2}$$

[3] In this case the sign of Y.

where $G_{Y|[X,c]}$ is a conditional distribution function for Y given X and c is a parameter in B. The models are assumed to have conditional means that exist, given by:

$$m(X,c) = \int Y \cdot dG_{Y|[X,c]}. \tag{3}$$

We will denote the sign of a model $I(m(X,c))$ as $I(x,c)$. Our analysis will be restricted to pairs $\left(\mathcal{G}, F_{Y|X}\right)$ such that $m : \mathcal{X} \times B \to \mathbb{R}^1$ is correctly specified for the sign of the conditional mean on B - that is, for some $b \in B$, $I(x,b) = I_{\mu_Y}(x)$ for almost all x.[4] We formalise our specification requirement in the following assumption.

Assumption 1 (Correct specification of model sign) *Let $m : \mathcal{X} \times B \to \mathbb{R}^1$ be a model such that for some $b \in B$, $m(x,b)$ has the same sign as $\mu_Y(x)$ almost everywhere, i.e.:*

$$m(x,b)\, w(x) = \mu_Y(x) \text{ a.e.}, \tag{4}$$

for some (not necessarily known) strictly positive $w : \mathcal{X} \to \mathbb{R}^1_{++}$.

An example illustrating the weak restriction this imposes on the relation of m and μ_Y is given in Figure 1 below.

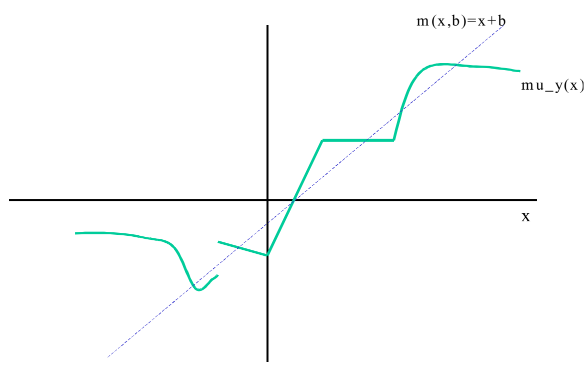

Figure 1. A simple model m with the same sign as a complex mean regression μ_Y.

Correct specification of the model for the conditional mean is much stronger than Assumption 1 in that it requires (4) to hold for $w(x) = 1$ rather than for some $w : \mathcal{X} \to \mathbb{R}^1_{++}$. It is easy to imagine a situation in which we cannot be confident that a parsimonious model is correctly specified for the conditional mean yet is for the sign of the conditional mean. For example, if x is a scalar and μ_Y is an unknown, possibly nonlinear and discontinuous function then a model as simple as $m(x,c) = c + x$ will satisfy Assumption 1 if μ_Y satisfies a single-crossing condition at zero (e.g. Manski and Thompson 1989) with $\lim_{x\to-\infty} \mu_Y(x) < 0$. Given our weak correct specification requirement and the assumptions that follow, models

[4] This is a requirement of correct specification for a very particular conditional attribute of a model. See White (1994) for definitions of correct specification in other more standard conditional attributes, such as the mean or median.

in $\mathcal{G}$ will not need to specify aspects of the conditional distribution of Y on X beyond the conditional mean.

2.2 Necessary and sufficient extremum conditions for correct specification of the sign of a mean regression

The estimators we will propose are based on sample analogs of extremum conditions satisfied by a parameter b for which (4) holds (i.e. the sign of $m(*, b)$ coincides with the sign of μ_Y). The first condition is that:

$$b \in \arg\max_{c \in B} \int Y \cdot I(X, c)\, dF. \tag{5}$$

where F is the true joint disrtibution of (Y, X). This condition is an immediate consequence of the following Lemma:

Lemma 1 *Let $h : \mathcal{X} \to \mathbb{R}_{++}$ be any strictly positive (bounded, integrable, measurable) function.*

Suppose Assumption 1 is satisfied.

Then $I(x, b) = I_{\mu_Y}(x)$ almost everywhere if and only if:

$$b \in \arg\max_{c \in B} \int h(X) \cdot Y \cdot I(X, c) dF. \tag{6}$$

Proof. $\Rightarrow$

$$\begin{aligned} \int h(X) \cdot Y \cdot I(X, c) dF &= \int h(X) \cdot \mu_Y(X) \cdot I(X, c) dF \\ &\leq \int h(X) \cdot \mu_Y(X) \cdot I(X, c) \cdot I_{\mu_Y}(X) dF \\ &\leq \int h(X) \cdot \mu_Y(X) \cdot I_{\mu_Y}(X) dF. \end{aligned} \tag{7}$$

By A1 there is a $c' \in B$ such that $I(x, c') = I_{\mu_Y}(x)$. Clearly in this case, the above relations hold with equality. Therefore if b is a maximising value, the above relations must also hold with equality and hence:

$$\int h(X) \cdot \mu_Y(X) \cdot I(X, b) dF = \int h(X) \cdot \mu_Y(X) \cdot I_{\mu_Y}(X) dF.$$

This implies that $I(X, b) = I_{\mu_Y}(X)$ almost surely.

$\Leftarrow$

We have already shown in (7) that:

$$\int h(X) \cdot Y \cdot I(X, c) dF \leq \int h(X) \cdot \mu_Y(X) \cdot I_{\mu_Y}(X) dF$$

Hence b is a maximising value of the LHS if $I(X, b) = I_{\mu_Y}(X)$. $\blacksquare$

The next Lemma, provides another useful condition satisfied by b.

Lemma 2 *Let $A : \mathcal{X} \to [0,1]$ be a mapping such that:*

$$A(x) \equiv \frac{E(|Y|\,|I(Y)=0,x)}{E(|Y|\,|I(Y)=0,x)+E(|Y|\,|I(Y)=1,x)},\;\; x \in \mathcal{X} \tag{8}$$

Suppose Assumption 1 is satisfied.

$I(x,b) = I_{\mu_Y}(x)$ *almost everywhere if and only if:*

$$b \in \arg\max_{c \in B} \int [I(Y) - A(X)] \cdot I(X,c)dF \tag{9}$$

Proof. Notice that for any function $h : \mathcal{X} \to \mathbb{R}$:

$$\int h(X) \cdot Y \cdot I(X,c)dF = \int h(X) \cdot \mu_Y(X) \cdot I(X,c)dF. \tag{10}$$

It is easy to show that:

$$\mu_Y(x) = \Pr(I(Y)=1|x)\,[E(|Y|\,|I(Y)=0,x)+E(|Y|\,|I(Y)=1,x)] - E(|Y|\,|I(Y)=0,x) \tag{11}$$

Define the strictly positive function $h : \mathcal{X} \to \mathbb{R}_{++}$ by:

$$h(x) \equiv \frac{1}{E(|Y|\,|I(Y)=0,x)+E(|Y|\,|I(Y)=1,x)} \tag{12}$$

Using (10),(11) and (12).

$$\int h(X) \cdot Y \cdot I(X,c)dF = \int [\Pr(I(Y)=1|X) - A(X)] \cdot I(X,c)dF$$

Lemma 1 together with the law of iterated expectations can now be applied to obtain the desired result. ■

If the 'nuisance function' A is constant ($A(x) = a$ for all x) then $I(x,b)$ will be the a'th quantile of I_Y and therefore the sign of a mean regression is given by the a'th quantile of the distribution of the sign of Y conditional on x. Quantile regression of binary variables is discussed in Manski and Thompson (1989). In particular, for $a = \frac{1}{2}$ the sign of a mean regression coincides with the median regression for the sign of Y. More generally, it is straightforward to verify that $I(x,b)$ is the $A(x)$'th quantile of I_Y conditional on x (see e.g. Skouras (1998)) and in this sense the sign of a mean regression $I(x,b)$ may be viewed as a generalisation of a quantile regression for the sign of Y. This is because $A(x)$ is a measure of the magnitude of Y when negative relative to the magnitude of Y when positive, constructed so that if the probability that Y is positive exceeds $A(x)$ then the mean of Y will be positive. The sign of the mean of Y therefore coincides with the $A(x)$'th quantile of the sign of Y.

3 Estimators for the sign of a mean regression

In this section, we propose three moment extremum estimators for semiparametric models of a mean regression. We show that under reasonable conditions these estimators lead to models that asymptotically have the same sign as the mean of the true conditional distribution even when the model is substantially misspecified.

3.1 Assumptions used in proving estimator consistency

We now introduce and discuss assumptions (drawing on Manski (1988)) to be used in proving estimator consistency.

Assumption 2 (Compactness) *The parameter space $B \subseteq \mathbb{R}^k$ is compact.*

Assumption 3 (SLLN) *The sample $\{y_n, x_n\}_{n=1}^{N}$ satisfies a strong law of large numbers (a variety of which can be found in White (1984)).*

Assumption 4 (Identifiability) *There exists a unique $b \in B$ such that:*

$$\int Y \cdot I\left(X, b\right) dF = \max_{c \in B} \int Y \cdot I(X, c) dF.$$

Whether this identifiability assumption holds will depend on the interaction of $\{m, F, B\}$ and must be ensured on a case-by-case basis by appropriate specification of m and B given our priors regarding the behaviour of F.

Assumption 5 (Boundary) *The following boundary condition is satisfied:*

$$\lim_{\alpha \to 0} \sup_{c \in B} \int |Y| \cdot I\left(\alpha - |m\left(X, c\right)|\right) dF = 0.$$

This ensures that the probability that $m(X, c)$ is close to zero for all c is small in an appropriate sense. It serves to ensure continuity of $\int Y \cdot I(X, c) dF$.

Assumption 6 (Equicontinuity) *There exists a mapping $v : \mathcal{X} \to \mathbb{R}^1_{++}$ such that $v\left(x\right) m(x, c)$ is equicontinuous on B, i.e. $\forall\ \alpha > 0$,*

$$\exists\ \gamma_\alpha : |a - c| < \gamma_\alpha \Rightarrow \sup_{x \in \mathcal{X}} |v\left(x\right) m(x, a) - v\left(x\right) m(x, c)| < \alpha, \quad (a, c) \in B \times B.$$

Sufficient conditions for equicontinuity of a function are provided by Manski (1988, Lemma 7, pp. 109-110) reproduced for reference purposes in Appendix B. The restrictions it imposes on the form models can take is evidently very weak. Its role is to introduce appropriate smoothness in $\int Y \cdot I(X, c) dF$ that does not depend on the behaviour of F.

It is relevant to note that if $m\left(*, c\right)$ is linear (for example because we restrict our attention to best linear prediction) A4-6 become immediately satisfied under regularity conditions given in Appendix B and in the proposition that follows, but only parameter scale can be identified so B must not include c and c' such that $c = ac'$ (where a is a positive scalar). When $m\left(*, c\right)$

is non-linear it is difficult to derive conditions on F ensuring A4-6 will hold. We note that when the parameter set B is discrete (see e.g. the application in Skouras (2001)) consistency only requires A3.

Sufficient conditions for identifiability are given by the following proposition, which extends a result in Manski (1985).

Proposition 1 *Sufficient conditions for the Identifiability Assumption are:*

1. *There exists* $\rho : \mathcal{X} \times B \to \mathbb{R}^1_{++}$ *and* $v : \mathcal{X} \to \mathbb{R}^k$ *such that:*
$$\rho\left(x, c\right) m(x, c) = v(x)'c, \qquad (x, c) \in (\mathcal{X}, B)\,,$$
with $\rho\left(x, b\right) = w\left(x\right)$ *of Assumption 1,*
2. *the support of* $F_{v(X)}$ *is not contained in any proper linear subspace of* $\mathbb{R}^k$*,*
3. $b_l \neq 0$ *for some* l *and* $\forall v\left(X_{-l}\right) \equiv \left(v\left(X_1\right), v\left(X_2\right), ..., v\left(X_{l-1}\right), v\left(X_{l+1}\right), ..., v\left(X_k\right)\right)$ *the distribution of* $v\left(X_l\right)|v\left(X_{-l}\right)$ *has everywhere positive Lebesgue density,*

Proof. Consider the difference

$$\begin{aligned} S\left(c\right) &\equiv \int Y \cdot I\left(X, b\right) dF - \int Y \cdot I\left(X, c\right) dF \\ &= \int \mu_Y\left(X\right)\left[I\left(X, b\right) - I\left(X, c\right)\right] dF_X \end{aligned}$$

Define

$$\begin{aligned} \mathcal{X}_c^+ &\equiv \left\{x \in \mathbb{R}^k : I\left[X, c\right] \neq I\left[X, b\right], I\left[X, b\right] = 1\right\}, \\ \mathcal{X}_c^- &\equiv \left\{x \in \mathbb{R}^k : I\left[X, c\right] \neq I\left[X, b\right], I\left[X, b\right] = 0\right\}. \end{aligned}$$

So

$$S\left(c\right) = \int_{\mathcal{X}_c^+} \mu_Y\left(X\right) dF_X - \int_{\mathcal{X}_c^-} \mu_Y\left(X\right) dF_X.$$

By assumption 1,

$$S\left(c\right) = \int_{\mathcal{X}_c^+ \cup \mathcal{X}_c^-} \left|\mu_Y\left(X\right)\right| dF_X \geq 0,$$

with the inequality being strict if $\int_{\mathcal{X}_c^+ \cup \mathcal{X}_c^-} dF_X > 0$.

The conditions we have assumed imply by the proof of Manski (1985, Lemma 2, p. 317) that for all $c \neq b$,

$$\begin{gathered} \int_{\mathcal{X}_c} dF_X > 0, \\ \mathcal{X}_c \equiv \left\{x \in \mathbb{R}^k : I\left[v(x)'c\right] \neq I\left[v(x)'b\right]\right\} \\ = \left\{x \in \mathbb{R}^k : I\left[x, c\right] \neq I\left[x, b\right]\right\} \\ = \mathcal{X}_c^+ \cup \mathcal{X}_c^-. \end{gathered}$$

Therefore the inequality will be strict and therefore $S(c)$ will be positive which gives the desired result. ■

3.2 A step function M-estimator

Consider the estimator b_N given by the sample analog of (5):

$$b_N \in \arg\max_{c \in B} \int Y \cdot I(X, c) \, dF_N, \tag{13}$$

where F_N is the empirical c.d.f. of $\{y_n, x_n\}_{n=1}^N$. This estimator is structurally similar to Manski's (1975) maximum score estimator in which Y is replaced by $I(Y)$ and $I(X, c)$ represents the sign of a median regression for Y (it may or may not be that the mean and median regression coincide). Our next proposition proves that b_N is consistent for the parameter b for which the model's sign is the same as the mean regression's sign. This is achieved by applying a theorem of Manski (1988) that extends standard conditions under which M-estimators are consistent to the case where the objective functions being optimised are step functions of the model parameter.

Proposition 2 *If Assumptions 1-6 hold, then*

$$\Pr\left(\lim_{N \to \infty} |b_N - b| = 0\right) = 1, \tag{14}$$

where $I(x, b) = I_{\mu_Y}(x)$ *almost everywhere.*

Proof. Theorem 3', Chapter 7 of Manski (1988) applies to b_N and ensures

$$\Pr\left(\lim_{N \to \infty} |b_N - b| = 0\right) = 1,$$

with

$$b = \arg\max_{c \in B} \int Y \cdot I(X, c) \, dF.$$

Lemma 1 in the Appendix ensures that the above is a necessary and sufficient condition for

$$I(x, b) = I_{\mu_Y}(x) \text{ a.e.}$$

■

3.3 A quasi generalised step function M-estimator

Another estimator that highlights some similarities between maximum score estimation and estimation of the sign of a mean regression is b_N^a, defined as:

$$b_N^a \in \arg\max_{c \in B} \int [I(Y) - A_N(X)] \cdot I(X, c) dF_N. \tag{15}$$

where A_N is an estimated model that converges uniformly to A given by (8).

Gourieroux and Monfort (1995) refer to estimators optimising objective functions that contain consistent estimates of certain nuisance parameters as 'quasi generalised M estimators' (Definition 8.2, p. 214). We will refer to b_N^a as a 'quasi generalised step function M-estimator' modifying their terminology slightly since in this case the estimator involves a nuisance function rather than a nuisance parameter. We establish the estimator's consistency in the following proposition.

Proposition 3 *If*

$$\Pr\left(\lim_{N\to\infty}\sup_{x\in\mathcal{X}}|A_N(x)-A(x)|=0\right)=1.$$

and Assumptions 1-6 hold, then

$$\Pr\left(\lim_{N\to\infty}|b_N^a-b|=0\right)=1, \tag{16}$$

where $I(x,b)=I_{\mu_Y}(x)$ almost everywhere.

Proof. Lemmata 3, 4 and 5 of Appendix A imply that:

$$\Pr\left(\lim_{N\to\infty}|b_N^a-b|=0\right)=1,$$

where

$$b=\arg\max_{c\in B}\int[I(Y)-A(X)]\cdot I(x,c)dF.$$

Lemma 2 ensures that the above is a necessary and sufficient condition for

$$I(x,b)=I_{\mu_Y}(x)\ \text{ a.e.}$$

■

The quasi generalised step function M-estimator may be viewed as a generalisation of the estimators for binary quantile regression in Manski and Thompson (1989) which correspond to the case that A is a known constant. One situation in which this special case may arise is when Y represents financial returns R_{t+1} in period $t+1$ and X are lagged returns. This is because empirical evidence suggests it is acceptable to assume that $|R_{t+1}|$ is independent of the sign of returns $I_{R_{t+1}}$ (Granger and Ding 1994a, Henriksson and Merton 1981) in which case $A(x)=\frac{1}{2}$ and for $A_N(x)=A(x)=\frac{1}{2}$ the first assumption in the above proposition is satisifed trivially and our estimator coincides with the maximum score estimator. Correspondingly, the sign of the mean regression coincides with a binary median regression.

Remark 1 *In the case $A(x)=a$ for all x, there exist statistical techniques (Zheng 1998) which allow us to judge whether m is a correctly specified model for the a'th quantile of $I(Y)$ and hence evaluate whether Assumption 1 holds. Given that many optimal decision rules of interest can be expressed as functions of the sign of a mean regression, we would thus also obtain specification tests for models of optimal decisions. It may be possible to extend these* *results to the case where $A(x)$ varies with x to provide a general specification test for models of the sign of a mean regression.*

Financial time series also provide instances in which the first assumption of the proposition may be reasonable even when $A(x)$ is non-constant because there is a strong relationship between future absolute returns and past returns (Taylor 1986, Schwert 1989, Granger and Ding 1994b, Mills 1996, Fornari and Mele 1994). This indicates that it may be feasible to obtain a good model for

$$E\left(|R_{t+1}|\,|I_{R_{t+1}}, R_t, R_{t-1}, ...R_{t-T}\right),$$

and hence also $A(R_t, R_{t-1}, ...R_{t-T})$. To be precise, what we would need is a model converging to $A(R_t, R_{t-1}, ...R_{t-T})$ uniformly on $(R_t, R_{t-1}, ...R_{t-T})$.[5] While this may be too much to expect, it seems reasonable that with a good model for $A(x)$ this estimator might be an interesting competitor to the step function M-estimator of the previous section. Of course this remains to be verified in the context of a specific application.

3.4 A smoothed estimator

Horowitz (1992) provides a smoothed maximum score estimator that he shows has various desirable tractable properties that the maximum score estimator does not. Analogously, we smooth our simple step function M-estimator to obtain a third consistent estimator for the sign of a mean regression. In future research we aim to derive further asymptotic properties of our smoothed estimator. The estimator b_N^s is defined as:

$$b_N^s \equiv \arg\max_{c\in B} \int Y \cdot K\left[\frac{m(X,c)}{\zeta_N}\right] dF_N, \tag{17}$$

where $\{\zeta_N\}$ is a sequence s.t. $\zeta_N > 0$ and $\lim_{N\to\infty}\zeta_N = 0$ and $K : \mathbb{R}\to\mathbb{R}$ is a continuous function satisfying $|K(v)| < \infty$, $\lim_{v\to-\infty} K(v) = 0$, $\lim_{v\to\infty} K(v) = 1$. The following Proposition shows consistency of the smoothed estimator.

Proposition 4 *If Assumptions 1-6 hold, then*

$$\Pr\left(\lim_{N\to\infty} |b_N^s - b| = 0\right) = 1, \tag{18}$$

where $I(x,b) = I_{\mu_Y}(x)$ almost everywhere.

[5] We conjecture that if this is the case, the quasi-generalised estimator will be more efficient than the simple step function M-estimator. This is because we would expect use of any additional available information regarding the structure of F to improve our estimators and A captures some of this structure.

Proof. Let

$$
\begin{aligned}
f(c) &\equiv \int Y \cdot I(X,c)dF, \\
f_N(c) &\equiv \int Y \cdot I(X,c)dF_N, \\
S_N(c) &\equiv \int Y \cdot K\left[\frac{m(X,c)}{\zeta_N}\right] dF_N.
\end{aligned}
$$

Notice that for any $\alpha > 0:$

$$
\begin{aligned}
|S_N(c) - f_N(c)| \leq \int \left| Y \cdot \left[K\left(\frac{m(X,c)}{\zeta_N}\right) - I(X,c)\right]\right| \cdot I\left(|m(X,c)| - \alpha\right) dF_N \\
+ \int \left| Y \cdot \left[K\left(\frac{m(X,c)}{\zeta_N}\right) - I(X,c)\right]\right| \cdot I\left[\alpha - |m(X,c)|\right] dF_N.
\end{aligned}
$$

Since $\Pr(m(X,c) = 0) = 0$ and by our assumptions, K is bounded with

$$
\lim_{N\to\infty} K\left(\frac{m(x,c)}{\zeta_N}\right) = \left\{ \begin{array}{c} 1 \text{ if } m(x,c) > 0 \\ 0 \text{ if } m(x,c) < 0 \end{array} \right\},
$$

it follows that the first term on the RHS of the previous expression converges to zero uniformly on c as $N \to \infty$. The second term is smaller than:

$$
k \int |Y| \cdot I\left(\alpha - |m(X,c)|\right) dF_N
$$

uniformly over c (where k is some positive constant).

By equicontinuity, there exists a finite δ_α such that for all a, c:

$$
|a - c| < \delta_\alpha \Rightarrow |m(x,c) - m(x,a)| < \alpha.
$$

Fix $a > 0$. Then for $c \in B$ such that $|a - c| < \delta_\alpha$

$$
\begin{aligned}
|m(x,c))| &< \alpha \\
&\Rightarrow |m(x,c)| + |m(x,a) - m(x,c)| < 2\alpha \\
&\Rightarrow |m(x,a)| < 2\alpha.
\end{aligned}
$$

Hence:

$$
\begin{aligned}
|a - c| &< \delta_\alpha \\
&\Rightarrow \int |Y| \cdot I\left(\alpha - |m(X,c)|\right) dF_N \leq \int |Y| \cdot I\left(2\alpha - |m(X,a)|\right) dF_N.
\end{aligned}
$$

Since B is a compact set, we may define the finite set B_α such that for all c in B there is an a in B_α such that $|a-c|<\delta_\alpha$. Then there exists an a in B_α such that for all c in B:

$$\int |Y| \cdot I\left(\alpha - |m(X,c)|\right) dF_N \leq \int |Y| \cdot I\left(2\alpha - |m(X,a)|\right) dF_N,$$

so for all c in B :

$$\int |Y| \cdot I\left(\alpha - |m(X,c)|\right) dF_N \leq \max_{a\in B_\alpha} \int |Y| \cdot I\left(2\alpha - |m(X,a)|\right) dF_N$$

By the Strong Law of Large Numbers, as $N \to \infty$,

$$\max_{a\in B_\alpha} \int |Y| \cdot I\left(2\alpha - |m(X,a)|\right) dF_N \to \max_{a\in B_\alpha} \int |Y| \cdot I\left(2\alpha - |m(X,a)|\right) dF \quad \text{a.s.}$$

Hence for all c in B, for $N > N_\alpha$

$$\left|S_N\left(c\right) - f_N\left(c\right)\right| \leq k \max_{a\in B_\alpha} \int |Y| \cdot I\left(2\alpha - |m(X,a)|\right) dF,$$

and therefore for $N > N_\alpha$

$$\begin{aligned} \sup_{c\in B} \left|S_N\left(c\right) - f_N\left(c\right)\right| &\leq k \max_{a\in B_\alpha} \int |Y| \cdot I\left(2\alpha - |m(X,a)|\right) dF \\ &\leq k \sup_{a\in B} \int |Y| \cdot I\left(2\alpha - |m(X,a)|\right) dF. \end{aligned}$$

Since this holds for all α, for N large enough:

$$\sup_{c\in B} \left|S_N\left(c\right) - f_N\left(c\right)\right| \leq k \lim_{\alpha\to 0} \sup_{a\in B} \int |Y| \cdot I\left(2\alpha - |m(X,a)|\right) dF,$$

which by the Boundary Condition implies

$$\Pr\left(\lim_{N\to\infty} \sup_{c\in B} \left|S_N\left(c\right) - f_N\left(c\right)\right| \leq 0\right) = 1.$$

Using also the fact that by Manski's (1988) Lemma 6, p.106:

$$\Pr\left(\lim_{N\to\infty} \sup_{c\in B} \left|f\left(c\right) - f_N\left(c\right)\right| \leq 0\right) = 1,$$

it follows that

$$\Pr\left(\lim_{N\to\infty} \sup_{c\in B} \left|S_N\left(c\right) - f\left(c\right)\right| \leq 0\right) = 1.$$

Using this in Lemma 5, we obtain that

$$\Pr\left(\lim_{N\to\infty} \left|b_N^s - b\right| = 0\right) = 1.$$

Lemma 1 now ensures that the above is a necessary and sufficient condition for

$$I(x, b) = I_{\mu_Y}(x) \text{ a.e.}$$

■

4 Further estimator properties

It would be very convenient to have analytical results that would allow us to estimate at least the asymptotic distribution of the estimators. However, this type of result is extremely difficult to obtain and is not even available for exhaustively studied special cases such as the maximum score estimator. We will therefore conduct a simulation study of estimator properties (focusing on a single estimator to save space) as well as a comparison with standard estimators. We choose to focus on the step function M-estimator because it is consistent under the weakest set of assumptions.

4.1 Step function M-estimator

Kim and Pollard (1990) show that a broad class of estimators optimising step functions converge at cube-root rate to an analytically intractable distribution. We have not been able to show that our step function M-estimator belongs in this class but it seems reasonable to conjecture that this is the case given that its only unusual feature is that it optimises a step function - which according to Kim and Pollard (1990, p.194) "is the main distinguishing feature of estimation problems that exhibit cube-root asymptotics". The following simulation provides corroborating evidence for this conjecture.

SIMULATION 4.1

Consider the following DGP:

$$\begin{aligned} R_{t+1} &= b_0 + b_1 R_t + U, && (19)\\ U &\sim N\left(0, \sigma^2\right) \text{ iid}, && (20)\\ R_0 &\sim N\left\{\frac{b_0}{1-b_1}, \sigma^2\left[1-(b_1)^2\right]^{-1}\right\}, && (21)\\ b_0 &= 0.00015;\ b_1 = 0.0330;\ \sigma = 0.0108. && (22) \end{aligned}$$

The parameters of this DGP were determined using OLS to estimate an AR(1) model on a series of returns drawn from the empirical distribution of IBM daily closing prices[6] from 1st January 1990 through to 6th November 1997 (2049 observations).

Suppose the parametric form of the DGP is known, but that the values of the parameters are not. As discussed, our estimator can only provide a consistent estimate of the scale of the parameters of a linear model. That is, in order for identifiability to be ensured by Proposition

[6] Obtained from DATASTREAM on the last day in the dataset.

4, the model to be estimated takes the form:[7]

$$m(R_t, c) = c + R_t.$$

It is worth noting that the model has less parameters than would be required if we were interested in aspects of the conditional mean beyond its sign. By Proposition 1, the step function M-estimator b_N will converge to the parameter b such that $b + R_t$ has the same sign as $b_0 + b_1 R_t$; it is easy to see that this will be the case for $b = \frac{b_0}{b1}$.

The estimator satisfies:[8]

$$b_N \in \arg\max_c \int R_{t+1} \cdot I\left[c + R_t\right] dF_N. \tag{23}$$

We took $T = 10^4$ draws from the c.d.f. F_{b_N} of b_N to obtain the simulated distribution $\widehat{F}_{b_N}$ and repeated for various sample sizes. In particular, we set: $N = \left\{100 \cdot 2^{l-1}\right\}_{l=1}^{7}$. The size of T was chosen so that the standard deviation $\widehat{\sigma}_{b_N}$ of $\widehat{F}_{b_N}$ did not change 'much' between $\frac{T}{2}$ and T for any N. The distributions $\widehat{F}_{b_N}$ were then used to investigate various properties of F_{b_N} that are of interest.

Asymptotic consistency: According to Proposition 1, b_N is consistent so it must converge to $\frac{.00015}{.0330} = .0045455$ almost surely as N becomes large. This is confirmed in Table 1, which illustrates the convergence of $\widehat{F}_{b_N}$ to b.

N	100	200	400	800	1600	3200	6400	∞
$10^3 \cdot \widehat{\mu}_{b_N}$	1.2797	1.3502	2.4048	2.9877	3.9049	4.099	4.4858	4.5455
$10^2 \cdot \widehat{\sigma}_{b_N}$	1.3603	1.3378	1.2694	1.1699	1.0162	0.86258	0.69860	-

Table 1 Each column corresponds to a sample size N for which we compute the mean $\widehat{\mu}_{b_N}$ and standard deviation $\widehat{\sigma}_{b_N}$ of the simulated distribution of b_N. The column corresponding to ∞ gives the true value b.

Finite sample bias: Table 1 also indicates that b_N is biased downwards meaning that forecasts tend to be positive more often than optimal forecasts. This occurs because this series has a positive mean and therefore the impact of a forecast that is wrongly positive on the objective function in (23) is smaller (on average) than the effect of a forecast that is wrongly negative. In finite samples the estimates contain a negative bias to insure against estimation uncertainty.

Rate of convergence: The most similar estimator to the step function M-estimator for which there exists an analytical characterisation is the maximum score estimator of Manski (1975) for which Kim and Pollard (1990) show convergence occurs at cube-root. We therefore

[7] Strictly speaking, we should also estimate the model $m(X, c) = c - R$ unless we know the sign of b_1, but (for simplicity) we will assume here (and whenever we estimate linear models) that this is indeed known. Often either theory or related empirical evidence suggest this sign.

[8] To fix b_N at a particular point, we took (both here and later) the value closest to zero satisfying (23).

conjecture that b_N converges to an asymptotic distribution D at cube-root rate, i.e. that:

$$N^{\frac{1}{3}}\left(b_N - \frac{b_0}{b_1}\right) \overset{a}{\sim} D. \tag{24}$$

If our conjecture is true, under standard regularity conditions, an implication is that:

$$\lim_{N\to\infty} N^{\frac{1}{3}}\sigma_{b_N} = \sigma_D, \tag{25}$$

where σ_{b_N} is the standard deviation of F_{b_N} and σ_D is the standard deviation of D.

Figure 2 is a plot of $(N_k)^{\frac{1}{3}}\,\widehat{\sigma}_{b_N}$ for each N_k which supports (25) and provides an estimate $\widehat{\sigma}_D \simeq 0.13$ for σ_D. In this figure σ_D is the asymptote to which $N^{\frac{1}{3}}\widehat{\sigma}_{b_N}$ tends. If convergence was not at cube-root rate, the asymptote would not exist.

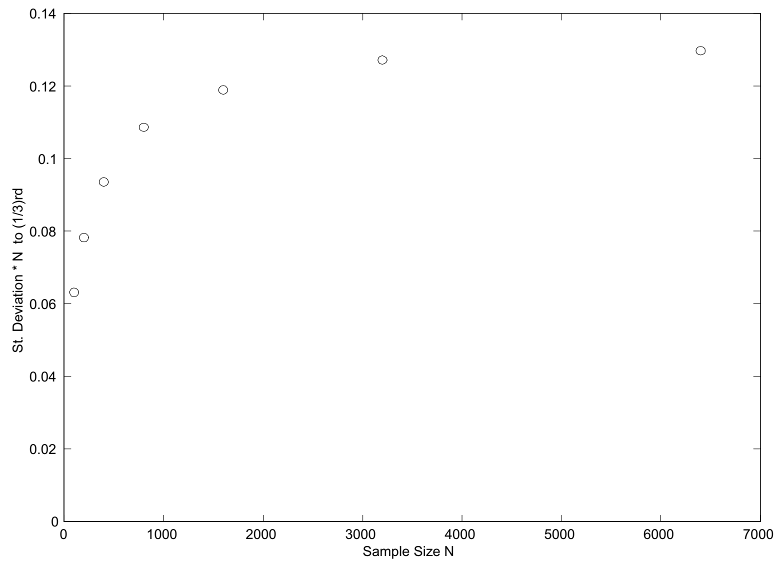

Figure 2. This figure provides supportive evidence for the conjecture that convergence occurs at cube-root rate.

Asymptotic distribution: That D is a non-normal distribution is clearly illustrated by the following QQplot of a normal distribution and the distribution of $N^{\frac{1}{3}}\widehat{F}_{b_N}$ for $N = 6400$.

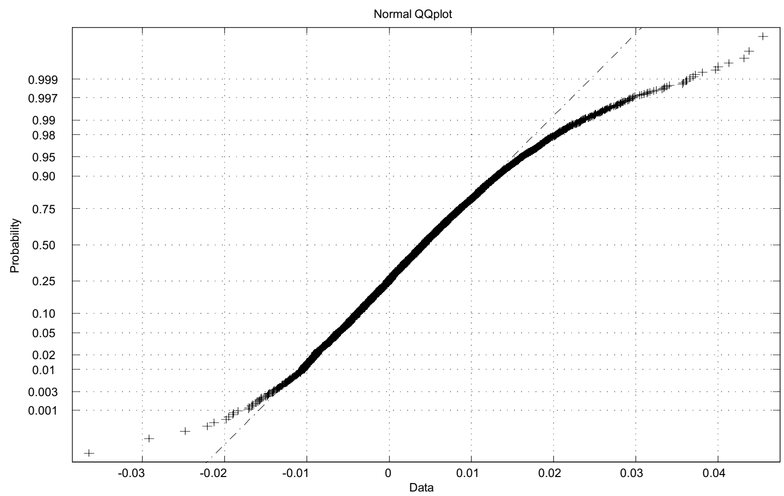

Figure 3. This figure plots the quantiles of the step estimator's asymptotic distribution as obtained from 10,000 simulations each with 6,400 observations. The axes are set so that quantiles of any normal distribution lie on a staight line.

4.2 Comparison to likelihood-based estimators

Asymptotic comparison: Under suitable regularity conditions, quasi- maximum likelihood estimators as well as Bayesian estimators will asymptotically converge to

$$\arg\max_{c\in B}\int\log\left(\frac{d}{dY}G_{Y|[X,c]}\right)dF, \tag{26}$$

i.e. the parameter minimising the Kullback-Leibler information discrepancy of the model.

Evidently, this condition does not guarantee correct specification of the sign as in (5) unless the model is correctly specified, which implies Assumption 1 holds for $w(x)=1$ and may require correct specification of model moments beyond the mean. Confidence in this stronger version of Assumption 1 requires a much greater degree of prior information about the behaviour of μ_Y that is not always available. Hence the estimators previously introduced will in general be much more robust than 'standard' estimators.

How bad misspecification may be will depend on the particular form of G and F and it therefore does not seem useful to state any general results. However, it should be clear that under Assumption 1, misspecification of likelihood based models can be arbitrarily bad even in simple cases. For example, let F be given by:

$$\begin{aligned} Y &= \log(X)+U, \\ U &\sim N(0,\sigma), \\ X &\sim Un\left[\frac{1}{2},\frac{3}{2}+k\right] \end{aligned}$$

where N is the normal distribution, $Un\left[\frac{1}{2}, \frac{3}{2}+k\right]$ is the uniform disribution in the interval $\left[\frac{1}{2}, \frac{3}{2}+k\right]$ and U and X are assumed to be independent. Assume that the model $G_{Y|[X,c]}$ is linear:

$$Y = c_0 + c_1 X$$

with correct specification of the distributions of U and X.

It is straightforward to verify that in this example, the parameters satisfying (26) are those of the least squares line going through $\log(X)$ in the interval $\left[\frac{1}{2}, \frac{3}{2}+k\right]$. Therefore, as k increases, likelihood based models estimate the sign arbitrarily badly. This is in contrast to our estimators that will converge to parameters that estimate the sign exactly.

Finite sample comparison: Even in the restrictive cases where we believe that we have an accurate model for the conditional mean, deriving estimates from its sign based on a model for the conditional mean estimated using standard methods is not always advisable. This is because the sign of the estimated model's conditional mean may be badly estimated, even when the conditional mean is estimated well. Indeed, it is quite a general observation that functions of 'good' estimates may be badly behaved due to the way estimation uncertainty interacts with the function of interest; a variety of estimators have been developed specifically to get around this problem (see e.g. Zellner (1978), Park and Zellner (1979)). Because our estimators directly penalise uncertainty in estimates of the sign of the conditional mean, their performance is not subject to this effect.

To illustrate this effect, Figure 4 plots the cumulative distribution function of parameter estimators for the AR(1) process of simulation 4.1, obtained by generating parameter estimates for 10,000 time series generated from this process. Similar cdf's were obtained for $b_0 \in [-0.05, 0.05]$, $b_1 \in [-0.03, 0.03]$ and sample sizes $N = 100, 400, 800$, but here we report results only for $b = (0.0002, 0.02)$, $\sigma = 0.01$ and $N = 400$. The estimators being compared are the maximum likelihood and step function M-estimator of a model $m(R_t, c) = c_0 + c_1 R_t$ (with the restriction $c_1 = \{-1, 1\}$ in the second case to ensure identification). Evidently, the dispersion of the maximum likelihood estimator is far larger, even though for this process it can be shown to be an asymptotically efficient estimator. Note that the probability that $\widehat{c}_1$ is smaller than zero for each estimator was .3604 and .3676 respectively.

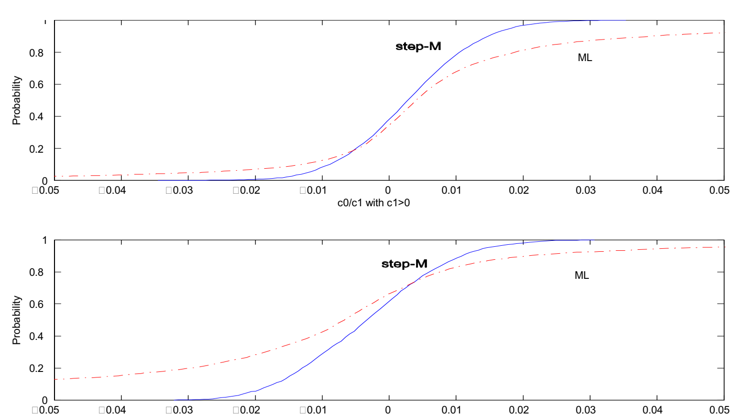

Figure 4: Simulated approximations to $F_{\frac{\widehat{c}_0}{\widehat{c}_1}} | [\widehat{c}_1 \leq 0]$ and $F_{\frac{\widehat{c}_0}{\widehat{c}_1}} | [\widehat{c}_1 > 0]$ for ML and step-M estimates.

Empirical comparison: In Skouras (1998), the step function M-estimator is applied to estimate a model as above and is compared to a quasi-maximum likelihood estimator. The application considered there is a special case of the decision problem of section 5.2 and the data were returns on the Dow Jones Industrial Average from 01/01/1896 to 31/12/1996 (27567 observations). A rolling window of 1,000 observations (approximately four years of data) was updated daily and used to estimate the sign of a mean regression by quasi-maximum likelihood and using the step function M-estimator respectively. The out of sample behaviour in terms of the objective function in (5) is substantially better and this leads the proposed estimator to be substantially superior for decision-makers.

5 Applications

Brandt (1999) discusses the importance of estimating solutions to decision problems and proposes a nonparametric method for estimating optimal decisions that can be described in terms of Euler equations. Many interesting decision problems that do not have solutions of this form can instead be described by the sign of a mean regression. The decision-maker can therefore use the methods discussed in the previous section to estimate optimal decision rules when, as is usually the case, these are unknown. In the first two applications presented we describe situations in which a decision-maker's optimal decision rule is a simple function of a threshold crossing indicator for μ_Y, i.e. the sign of a mean regression $\mu_{Y,a}(x) \equiv \mu_Y(x) - a$ for some mean regression μ_Y and a known constant a. The third and fourth applications are situations in which the modeller is interested primarily in the *zeros* of $\mu_{Y,a}$, i.e. the values of x such that $\mu_{Y,a}(x) = 0$. These zeros are also the x's such that the sign of neither $\mu_{Y,a}$ nor $-\mu_{Y,a}$ is positive. These applications provide a broad range of circumstances in which the sign of a mean regression is of interest and can be consistently estimated using the methods previously discussed (but note that now $\mu_{Y,a}$ rather than μ_Y must satisfy assumption 1).

5.1 Estimating solutions to binary decision dynamic programming problems

A broad class of binary decision dynamic programming problems of agents at time τ can be written in the general form:

$$\max_{d_t \in \{0,1\}} E\left[\sum_{t=\tau}^{T} \beta^{t-\tau}\left[\left(R_{t,1} - R_{t,0}\right) d_t + R_{t,0}\right] | x_\tau\right], \tag{27}$$

where T is possibly infinite, $0 < \beta < 1$ is a discount rate, $E(*|x_\tau)$ is the expectation of a random variable conditional on $X_\tau = x_\tau$ (information at time τ) and $R_{t,j}$ is the (random at $t-1$) reward to action j obtained by an individual choosing action j at time t.

Many such problems have the property that $\{d_t = 1\}$ is an optimal action at time t if and only if:

$$E\left(Y_{t+1}|x_t\right) \geq a_t, \tag{28}$$

where Y_{t+1} is the realisation of the agent's value function if $d_t = 1$ is chosen at time t. In particular, this is the form of solutions to optimal stopping problems - that is problems in which if the optimal decision is $d_t = 1$ (stop) at $t = k$, then $d_t = 1$ is also optimal for $t > k$ (see e.g. Eckstein and Wolpin (1989) for a review of the substantial literature on problems of this form).[9]

Suppose now that $E(Y_{t+1}|x_t)$ is unknown to an agent solving this problem who must estimate his optimal decision using a sample of past data $\{y_n, x_n\}_{n=1}^{N}$. Then as is evident from (28) it suffices that at time t he estimates the sign of the mean regression $\mu_{Y_{t+1},a}(x_t) = E(Y_{t+1}|x_t) - a_t$. This mean regression must be re-estimated at each t until the first instance in which it is positive.

Example 1 ***A canonical two period[10] job-search problem***

Let $R_{t,0} = b$ be a 'benefits' payment, i.e. the reward from remaining unemployed at time $t = 1, 2$ and $R_{1,1} = Y_1$ be the reward from being employed in the first period at a known wage Y_1. If a job is accepted in the first period, no new job offers are received in the second (employment leaves no time to search for new jobs) but the worker can keep the job he has. However, if a job is not accepted, a new wage offer Y_2 is received in the second period. The reward from being employed in the second period is then:

$$R_{2,1} = \left\{ \begin{array}{c} Y_1 \text{ if } d_1 = 1 \\ Y_2 \text{ if } d_1 = 0 \end{array} \right\},$$

where $d_t = 1$ indicates a job offer is accepted in period t and $d_t = 0$ indicates it is not. Note that at $t = 1$, Y_2 is a random variable.

The agent's optimisation problem at $t = 1$ is a special case of (27) which may be more simply written as:

$$\max_{d^1, d^2 \in \{0,1\}} d_1 (Y_1 - b) + b + \beta \left[d_2 \left(d_1 \cdot (Y_1 - E(Y_2|x)) + E(Y_2|x) - b \right) + b \right]$$

where β is the agent's discount factor and x is a vector containing public information at $t = 1$ predicting the distribution of wages Y_2 (such as unemployment rates or other business cycle variables).

In the non-trivial case where $Y_1 > b$, the optimal decision rule (obtained by backward induction) is:

$$\begin{aligned} d_2 &= 1 \text{ iff } \left\{ \left[Y_1 > b \text{ and } d^1 = 1 \right] \text{ or } \left[Y_2 > b \text{ and } d^1 = 0 \right] \right\} \\ d_1 &= 1 \text{ iff } \left\{ \begin{array}{c} Y_1 + \frac{1}{\beta}(Y_1 - (\beta + 1) b) > 0 \\ \text{or } E(Y_2|x) - Y_1 - \frac{1}{\beta}(Y_1 - b) < 0 \end{array} \right\} \end{aligned}$$

*which can be solved without knowledge of $E(Y_2|x)$ if the **sign** of $\mu_Y = E(Y_2|x) - Y_1 - \frac{1}{\beta}(Y_1 - b)$ is known. Estimation of the optimal decision rule can be achieved by estimation of the sign of μ_Y.*

[9] A very different dynamic programming problem with a solution of the same form is the 'cost-loss ratio' problem extensively analysed by meteorologists - see for example Katz and Murphy (1990).

[10] Granger and Pesaran (2000) argue that even single-period binary decision problems arise frequently and present an analysis of such problems (e.g. to grit or not to grit a road that may or may not become icy). Here we discuss a slightly more general problem but solutions to either could be solved by estimating the sign of a regression mapping.

5.2 Estimating solutions to continuous decision problems with corner solutions

The following example illustrates that even when decision spaces are convex optimal decision rules may be corner solutions that can be represented as the sign of an appropriately chosen mean regression.

Example 2 ***Risk Neutral bond-portfolio investment decision with transaction costs***
Consider a Risk Neutral Investor's single-period problem of allocating wealth between a 'riskless' asset such as a bond and a single risky asset (or portfolio). This can be described as:

$$\max_{a_t \in [-s,l]} E\left\{W_t\left[(1+R_{t+1})\,a_t + \left(1+R^f_{t+1}\right)(1-a_t)\right]|x_t\right\} - \kappa W_t\left|a_t - a'_t\right| \tag{29}$$

where x_t is a vector variable used to forecast the returns R_{t+1} to the portfolio from t to $t+1$, R^f_{t+1} are the returns to the bond, W_t is wealth at time t, κ are proportional transaction costs, a'_t is the proportion of W_t held in the asset before a decision is made at t and a_t is the proportion of wealth chosen for investment in the risky asset. The investor can borrow at the riskless interest rate up to $(l-1)*100\%$ of his wealth ($l \geq 1$) and take a short position in the asset of a size up to $s*100\%$ of his wealth ($s \geq 0$).[11]

Any solution to (29) also solves:

$$\max_{a_t \in [-s,l]} E\left(R^e_{t+1}|X_t\right) a_t - \kappa\left|a_t - a'_t\right| \tag{30}$$

where $R^e_{t+1} \equiv R_{t+1} - R^f_{t+1}$ are 'excess returns'.

Let $\mu_{Y,1}(x) \equiv E\left(R^e_{t+1}|X_t = x\right) - \kappa$ and $\mu_{Y,2}(x) \equiv E\left(R^e_{t+1}|X_t = x\right) + \kappa$. A solution to (30) can be expressed as:

$$a_t(x) = \left\{\begin{array}{c} -s \text{ if } 0 \geq \mu_{Y,2}(x) \\ a'_t \text{ if } \mu_{Y,2}(x) > 0 \geq \mu_{Y,1}(x) \\ l \text{ if } \mu_{Y,1}(x) > 0 \end{array}\right\}. \tag{31}$$

Estimation of optimal decisions a_t can be achieved by estimation of the sign of the mean regression $\mu_{Y,1}$ and $\mu_{Y,2}$.

5.3 Estimating the zeros of a mean regression

The set of $\mathcal{Z}$eros of μ_Y is:

$$\mathcal{Z} \equiv \{x : \mu_Y(x) = 0, x \in \mathcal{X}\} \tag{32}$$

which is identical to

$$\left\{x : I_{\mu_Y}(x) = -I(-\mu_Y(x)), x \in \mathcal{X}\right\}. \tag{33}$$

An estimate for the set $\mathcal{Z}$ can be obtained as a function of estimates for the sign of the

[11] The 'Risk Neutral Forecasting' decision problem of Skouras (1998) is a special case of this with $\kappa = 0$ and $R^f = 0$.

mean regression μ_Y and $-\mu_Y$:

$$\widehat{\mathcal{Z}} \equiv \left\{x : I\left(x, b_N\right) = I'\left(x, b'_N\right), x \in \mathcal{X}\right\}, \tag{34}$$

where $I(*, b_N)$ is an estimate for the sign of μ_Y and $I'(*, b'_N)$ is an estimate for the sign of $-\mu_Y$. Note that if Assumption 1 applies to (m, μ_Y) then it will also apply to $(m, -\mu_Y)$ and we can expect that if b_N and b'_N are obtained using the proposed estimators, $\widehat{\mathcal{Z}}$ will be a consistent estimate of $\mathcal{Z}$. We now discuss various interesting applications in which the set $\mathcal{Z}$ is of interest.

5.3.1 Calibration

Suppose we have $J \geq 1$ observations of Y drawn from $F_Y|x^*$ (x^* unknown). The calibration problem is the task of estimating x^* given $\{y_j\}_{j=1}^J$ and a sample $\{y_n, x_n\}_{n=1}^N$ of draws from F. Observe that x^* is in the set:

$$\mathcal{Z} \equiv \{x : \mu_Y(x) - \mu_Y(x^*) = 0, x \in \mathcal{X}\}$$

and if this set can be assumed to be single-valued, the calibration problem (estimating x^*) reduces to the problem of estimating $\mathcal{Z}$.

The ‘classical’ approach to calibration (Eisenhart 1939) draws on the fact that the set:

$$\widehat{\mathcal{Z}}' = \left\{x : m\left(x, \widehat{c}\right) - \frac{\Sigma_{j=1}^J y_j}{J} = 0, x \in \mathcal{X}\right\}$$

should provide a good estimate for $\mathcal{Z}$ under weak conditions on $F_Y|x^*$ when $m'(*, \widehat{c})$ is a good model for μ_Y, with estimated parameter $\widehat{c}$. However, standard parametric estimates $\widehat{c}$ have been shown to lead to highly unstable estimates $\widehat{\mathcal{Z}}'$ (Kruthckoff 1967) because of the nonlinearity of the mapping from $\widehat{c}$ to $\widehat{\mathcal{Z}}$. This is a special case of a situation discussed in the introduction - functions of ‘good’ parameter estimates do not necessarily provide good estimates of parameter functions.

To see that our estimators can be applied to estimation of $\mathcal{Z}$, let $\mu_Y(x) = \mu'_Y(x) - \frac{\Sigma_{j=1}^J y_j}{J}$ and $m(*, c) = m'(*, c) - \frac{\Sigma_{j=1}^J y_j}{J}$ be a model for $\mu_Y(*)$. Then a parameter estimate b_N obtained using one of our estimators may lead to consistent estimation of $\mathcal{Z}$ even when $m'(*, c)$ is substantially misspecified.

5.3.2 Econometric equation inversion

Let X_i denote the i'th element of X and X_{-i} denote $(X_1, ..., X_{i-1}, X_{i+1}, ..., X_I)$. When $Y = \mu_Y(X) + U$ - as in (1) - it is sometimes desirable to estimate the value x_i^* such that for given $\left(y^*, x_{-i}^*\right)$ it is the case that $y^* = \mu_Y\left(x_i^*, x_{-i}^*\right)$. This is referred to as the problem of ‘econometric equation inversion’, since it is the inverse of $\mu_Y(x_i, x_{-i})$ with respect to x_i that needs to be estimated at the point $\left(y^*, x_{-i}^*\right)$. To see that econometric equation inversion is a special case of the problem of estimating the zeros of a mean regression define $\mu_{x_{-i},y^*}(x_i) \equiv \mu_Y(x_i, x_{-i}) - y^*$. Then the zeros of $\mu_{x_{-i},y^*}(x_i)$ are the object of interest and can be estimated using the proposed methods. The following two examples illustrate contexts in which econometric equation inversion has received widespread attention and in which standard estimators have been shown to be problematic.

Macroeconometric models Substantial research effort has been devoted to analysing the implications of econometric equation inversion in the context of macroeconomic models - see for example Hendry and Ericsson (1991), Ericsson, Hendry, and Mizon (1998) and the references therein. The standard setting in which this arises is when (1) is assumed to represent a money demand equation with Y being money demand and X including prices, interest rates etc. Based on these assumptions, it is often required to obtain the level of prices (or interest rates) x_i that are consistent with a given amount of money demand y^* and interest rates x_{-i} (or prices).

The most common way of estimating this x_i is to invert $m(x_i, x_{-i}, c)$ with respect to x_i so to obtain a relationship of the form $x_i = f(y^*, x_{-i}, c)$ and use f as if it were a correctly specified model for $E(x_i | [y^*, x_{-i}])$. Under this assumption, c can be estimated using standard methods but the assumption only holds under very special and unrealistic conditions such as joint normality of (Y, X). Estimation following this course may therefore lead to inconsistent and unreliable estimates for x_i in practice.

It seems preferable to estimate the zeros of $\mu_{x_{-i},y^*}(x_i) = \mu'_Y(x_i, x_{-i}) - y^*$ and treat them as the price level for which the expected money demand is y^* when interest rates are x_i.

Decisions with Targets Consider the following classic problem in the control of passively observed systems which arises routinely in economic models. A decision maker has some control over the distribution of X_i which he uses to ensure that the expected value of Y is equal to a 'target level' y^*, given any realisation of X_{-i}. It is typically assumed that $\mu'_Y(x)$ is not affected by any changes in the distribution of X_i which the decision maker effects. In this case, if μ_Y is unknown, the decision maker will be interested in estimating the zeros of $\mu_{x_{-i},y^*}(x_i)$.

6 Conclusions and future directions

We have presented a study of the sign of a mean regression suggesting it is an important attribute of econometric relationships that in many cases deserves special treatment. On the basis of extremum conditions characterising the sign of a mean regression, we have proposed three estimators for semiparametric models of mean regressions that will provide consistent estimates for this sign. We have given a number of examples of important modelling situations in which this is the case and have argued that in such contexts the proposed estimators may have substantial advantages over standard estimators.

The reasons for this are twofold: Firstly, our estimators will lead to models that are consistent for the sign of the mean regression even when the model is highly misspecified while this will not be the case for standard estimators. The conditions required for consistency are weak and can accommodate, for example, arbitrary heteroskedasticity of errors and a non-functional relationship between regressor and regressand. Secondly, if models are estimated using standard procedures, model sign may be subject to extreme estimation uncertainty because it is typically a complicated non-linear function of the estimated parameters. In contrast, the proposed estimators penalise uncertainty about model sign and therefore account to some extent for estimation uncertainty.

Comparison of our estimators on the basis of analytical results is difficult because most of their properties seem to be intractable. The properties of one of the estimators are investigated using simulations and, despite irregularities, it performs well relative to standard methods as is also confirmed in the empirical application of Skouras (1998). The properties of the smoothed estimator may be tractable and we intend to investigate this in future research. This is particularly appealing since such results might make it possible to conduct inference about properties of the sign of a mean regression and therefore also about the form of optimal decisions that can be expressed as functions of this sign. Of course, it will also be useful to consider how other stages of the modelling process beyond estimation might be modified to improve modelling of the sign of a mean regression.

7 Appendix

7.1 A. Lemmata used in proofs of propositions

Lemma 3 *Let*

$$\begin{aligned} h(x,c) &\equiv (I(Y) - A(x)) \cdot I(x,c), \\ h_N(x,c) &\equiv (I(Y) - A_N(x)) \cdot I(x,c), \\ v(x) &\equiv I(Y) - A(x), \\ v_N(x) &\equiv I(Y) - A_N(x). \end{aligned}$$

If $A_N(x)$ converges uniformly to $A(x)$ almost surely, the SLLN holds, and B_α is a finite cardinality subset of B then

$$\Pr\left(\lim_{N\to\infty}\max_{c\in B_\alpha}\left|\int h_N(X,c)dF_N - \int h(X,c)dF\right| = 0\right) = 1 \tag{35}$$

and

$$\Pr\left(\lim_{N\to\infty}\max_{c\in B_\alpha}\left|\int v_N(X)dF_N - \int v(X)dF\right| = 0\right) = 1. \tag{36}$$

Proof. As the proofs of (35) and (36) are identical, we prove only (35).

Since $h_N(X,c)$ is a linear function of $A_N(X)$, the assumption directly implies that for all $\varepsilon > 0$ there exists a N_ε such that for $N > N_\varepsilon$, uniformly on $(\mathcal{X}, B)$:

$$|h_N(X,c) - h(X,c)| < \varepsilon \text{ a.s.}$$

So:

$$\int |h_N(X,c) - h(X,c)|\, dF_N < \int \varepsilon dF_N = \varepsilon$$

Since $\int |h_N(X,c) - h(X,c)|\, dF_N \geq \left|\int h_N(X,c) - h(X,c)dF_N\right|$ this implies:

$$\left|\int h_N(X,c) - h(X,c)dF_N\right| < \varepsilon$$

$$\Rightarrow \Pr\left(\lim_{N\to\infty}\left|\int h_N(X,c) - h(X,c)dF_N\right| = 0\right) = 1 \tag{37}$$

The Strong Law of Large Numbers ensures that for any c:

$$\lim_{N\to\infty}\left|\int h(X,c)dF_N - \int h(X,c)dF\right| = 0 \text{ a.s.} \tag{38}$$

So combining (37) and (38), for any c:

$$\lim_{N\to\infty}\left|\int h_N(X,c)dF_N - \int h(X,c)dF_N\right| + \left|\int h(X,c)dF_N - \int h(X,c)dF\right| = 0 \text{ a.s.} \tag{39}$$

Now notice that since $|a-b|+|b-c|\geq|a-c|$,

$$\left|\int h_N(X,c)dF_N-\int h(X,c)dF\right|\leq\left|\int h_N(X,c)dF_N-\int h(X,c)dF_N\right|$$
$$+\left|\int h(X,c)dF_N-\int h(X,c)dF\right|$$

Using this fact and (39), we obtain that for any $c\in B_\alpha$:

$$\lim_{N\to\infty}\left|\int h_N(X,c)dF_N-\int h(X,c)dF\right|=0\text{ a.s.}$$

Since this holds for any $c\in B$, an immediate consequence is (35) ■

Lemma 4 *Let*

$$\begin{aligned}h(x,c)&\equiv v(x)\cdot I(x,c),\\ h_N(x,c)&\equiv v_N(x)\cdot I(x,c),\\ v(x)&\equiv I(Y)-A(x),\\ v_N(x)&\equiv I(Y)-A_N(x).\end{aligned}$$

If $A_N(x)$ converges uniformly to $A(x)$ almost surely and A2-6 are satisfied, then

$$\Pr\left(\lim_{N\to\infty}\max_{c\in B}\left|\int h_N(X,c)dF_N-\int h(X,c)dF\right|=0\right)=1.$$

Proof. We follow the logic of Manski (1988), Lemmata 5 and 6, pp 104-108:

$$\left|\int h_N(X,a)-h_N(X,c)dF_N\right|$$
$$=\left|\int v_N(X)[I(X,a)-I(X,c)]dF_N\right|$$
$$\leq\int|v_N(X)[I(X,a)-I(X,c)]|\,dF_N$$

So

$$\left|\int h_N(X,a)-h_N(X,c)dF_N\right|\leq\int_{\mathcal{X}(a,c)}|v_N(X)|\,dF_N \tag{40}$$

where

$$\mathcal{X}(a,c)\equiv\{x\in\mathcal{X}:m(x,a)\leq0\leq m(x,c)\cup m(x,a)\geq0\geq m(x,c)\}.$$

For $\alpha>0$ and $c\in B$, by the equicontinuity assumption (A6) which for notational simplicity (but without loss of generality) we assume holds for $w(x)=1$, it follows that there exists a δ_α such that for all $(a,x)\in(B,\mathcal{X})$

$$|a-c|<\delta_\alpha\Rightarrow\left\{\begin{array}{c}m(x,c)>\alpha\Rightarrow m(x,a)>0\\ m(x,c)<-\alpha\Rightarrow m(x,a)<0\end{array}\right\}$$

Hence there exists a δ_α such that for all $(a,x) \in (B, \mathcal{X})$

$$|a-c| < \delta_\alpha \Rightarrow \mathcal{X}(a,c) \subset \mathcal{X}_{c\alpha} \equiv \{x \in \mathcal{X} : -\alpha < m(x,c) < \alpha\}. \tag{41}$$

Using (40) and (41):

$$|a-c| < \delta_\alpha \Rightarrow \left| \int h_N(X,a) - h_N(X,c) dF_N \right| \leq \int_{Xc\alpha} |v_N(X)| \, dF_N$$

By identical reasoning, this condition holds if we replace $h_N(x,c)$ with $h(x,c)$, $v_N(x)$ with $v(x)$ and F_N with F.

Hence,

$$\begin{aligned} |a-c| < \delta_\alpha \Rightarrow \left| \int h(X,a) - h(X,c) dF \right| + \left| \int h_N(X,a) - h_N(X,c) dF_N \right| \\ \leq \int_{\mathcal{X}_{c\alpha}} |v(X)| \, dF + \int_{\mathcal{X}_{c\alpha}} |v_N(X)| \, dF_N \end{aligned} \tag{42}$$

Now notice that:

$$\begin{aligned} &\left| \int h_N(X,a) dF_N - \int h(X,a) dF \right| \\ &\qquad = \left| \begin{array}{c} \int (h_N(X,a) - h_N(X,c)) \, dF_N - \int (h(X,a) - h(X,c)) \, dF + \\ + \int h_N(X,c) dF_N - \int h(X,c) dF \end{array} \right| \end{aligned}$$

$$\begin{aligned} \Rightarrow &\left| \int h_N(X,a) dF_N - \int h(X,a) dF \right| \\ &\leq \left| \int h_N(X,a) - h_N(X,c) dF^N \right| + \left| \int h(X,a) - h(X,c) dF \right| \\ &\qquad + \left| \int h_N(X,c) dF^N - \int h(X,c) dF \right| \end{aligned}$$

Hence combining the above relationship with (42):

$$\begin{aligned} |a-c| &< \delta_\alpha \Rightarrow \\ & \left| \int h_N(X,a) dF_N - \int h(X,a) dF \right| \\ &\leq \int_{Xc\alpha} |v(X)| \, dF + \int_{Xc\alpha} |v_N(X)| \, dF_N + \left| \int h_N(X,c) dF_N - \int h(X,c) dF \right| \end{aligned}$$

Now Compactness of B implies that there exists a $B_\alpha \subset B$ such that $card(B_\alpha) < \infty$ and for any $a \in B$ there is a $c \in B_\alpha$ such that $|a - c| < \delta_\alpha$. Therefore for all $a \in B$,

$$\left| \int h_N(X,a)\, dF - \int h(X,a)\, dF \right| \leq \max_{c \in B_\alpha} \int_{\mathcal{X}_{c\alpha}} |v(X)|\, dF + \max_{c \in B_\alpha} \int_{\mathcal{X}_{c\alpha}} |v_N(X)|\, dF_N$$
$$+ \max_{c \in B_\alpha} \left| \int h_N(X,c) dF_N - \int h(X,c) dF \right|$$

By Lemma 1, for all α, $\eta > 0$, there exists a $N_{\alpha\eta} < \infty$ such that for all $N > N_{\alpha\eta}$

$$\begin{aligned} \sup_{a \in B} \left| \int h_N(X,a) dF_N - \int h(X,a) dF \right| &\leq 2[\max_{c \in B_\alpha} \int_{Xc\alpha} |v(X)|\, dF + \eta \\ &\leq 2[\sup_{c \in B} \int_{Xc\alpha} |v(X)|\, dF + \eta \end{aligned}$$

Now the Boundary condition can be written as:

$$\lim_{\alpha \to 0} \sup_{c \in B} \int_{c\alpha} |Y|\, dF = 0$$

implying that as $(\alpha, \eta) \to 0$ the required result is obtained. ■

Lemma 5 *If A2 holds and*

$$\int [I(Y) - A_N(x)] \cdot I(x,c) dF_N$$

converges uniformly almost surely as N tends to infinity to

$$\int [I(Y) - A(x)] \cdot I(x,c) dF,$$

then:

$$\Pr\left(\lim_{N \to \infty} \left| b_N^\alpha - b^j \right| = 0 \right) = 1.$$

Proof. See for example Amemiya (1985) Theorem 4.1.1 as modified by footnote 1. ■

7.2 B: Sufficient conditions for assumptions required for asymptotic consistency

Sufficient conditions for Equicontinuity

By Manski (1988), Lemma 7, pp. 109-110, for some $\rho : \mathcal{X} \times B \to \mathbb{R}^1_{++}$, *at least one* of (1), (2) or (3) hold:

1. $\mathcal{X} \times B$ is a compact metric space and $\rho(*,*)\, m(*,*)$ is continuous on it.
2. $\rho(*,*)\, m(*,*)$ is bounded on $\mathcal{X} \times C$ and $m(x,*)$ is convex on C for all $x \in \mathcal{X}$, where $B \subset C \subset \mathbb{R}^k$ and C is an open convex set.

3. $\rho\,(x,c)\,m(x,c) = v(x)'c,\ (x,c) \in (\mathcal{X}, B)\,,$ for some $v : \mathcal{X} \to \mathbb{R}^k$.

Sufficient conditions for the Boundary condition

According to Manski (1988, Lemma 8, pp. 110-111) the following three conditions are sufficient:

1. For some $\rho : \mathcal{X} \times B \to \mathbb{R}^1_{++}$, there exists a $v : \mathcal{X} \to Z \subset \mathbb{R}^k$ such that $\rho\,(x)\,m(x,c) = v(x)'c,\ (x,c) \in (\mathcal{X}, B)\,,$
2. $\forall\ (c,\omega) \in B \times V$, where V is the range space of $|Y|$, the probability measure $F_{v(x)'c}|Y$ is absolutely continuous w.r.t. the Lebesgue measure μ and also $\exists\ \lambda < \infty$ s.t. $\frac{dF_{v(x)'c}|Y}{d\mu} < \lambda$,
3. $\int |Y|\, dF_Y$ exists

References

Amemiya, T. (1985): *Advanced Econometrics*. Basil Blackwell.

Brandt, M. W. (1999): "Estimating Portfolio and Consumption Choice: A Conditional Euler Equations Approach," *Journal of Finance*, 54(5), 1609–45.

Eckstein, Z., and K. I. Wolpin (1989): "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models," *Journal of Human Resources*, 24(4), 562–98.

Eisenhart, C. (1939): "The Interpretation of Certain Regression Methods and their Use in Biological and Industrial Research," *Annals of Mathematical Statistics*, 10.

Ericsson, N. R., D. F. Hendry, and G. E. Mizon (1998): "Exogeneity, Cointegration and Economic Policy Analysis," *Journal of Business and Economic Statistics*, 16(4), 370–387.

Fornari, F., and A. Mele (1994): "A Stochastic Variance Model for Absolute Returns," *Economics Letters*, 46, 211–214.

Gourieroux, C., and A. Monfort (1995): *Statistics and Econometric Models*, vol. 1. Cambridge University Press.

Granger, C. W., and Z. Ding (1994a): "Modeling Volatility Persistence of Speculative Returns: A New Approach," *Journal of Econometrics*, 73, 185–215.

——— (1994b): "Stylized Facts on the Temporal and Distributional Properties of Daily Data from Speculative Markets," working paper 93-38, University of California at San Diego.

Granger, C. W., and M. H. Pesaran (2000): "A Decision Theoretic Approach to Forecast Evaluation," in *Statistics and Finance: An Interface*, ed. by W. S. Chan, W. K. Li, and H. Tong, working paper 15, pp. 261–78. Imperial College Press, London.

Härdle, W., and R. Nixdorf (1987): "Nonparametric Sequential Estimation of Zeros and Extrema of Regression Functions," *IEEE Transactions on information systems*, IT-33(3), 367–372.

Hendry, D. F., and N. R. Ericsson (1991): "Modeling the Demand for Narrow Money in the United Kingdom and the United States," *European Economic Review*, 35, 833–886.

Henriksson, R. D., and R. C. Merton (1981): "On Market Timing and Investment Performance II: Statistical Procedures for Evaluating Forecasting Skills," *Journal of Business*, 54, 513–533.

Horowitz, J. L. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60(3), 506–531.

Katz, R. W., and A. H. Murphy (1990): "Quality/Value Relationships for Imperfect Weather Forecasts in a Prototype Multistage Decision-Making Model," *Journal of Forecasting*, 9, 75–86.

Kim, J., and D. Pollard (1990): "Cube Root Asymptotics," *Annals of Statistics*, 18, 191–219.

Kruthckoff, R. (1967): "Classical and Inverse Regression Methods of Calibration," *Technometrics*, pp. 425–439.

Manski, C. (1975): "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, (3), 205–228.

——— (1985): "Semiparametric Analysis of Discrete Response," *Journal of Econometrics*, pp. 313–333.

Manski, C. F. (1988): *Analog Methods in Econometrics*, vol. 39 of *Monographs on Statistics and Applied Probability*. Chapman and Hall.

Manski, C. F., and S. T. Thompson (1989): "Estimation of Best Predictors of Binary Response," *Journal of Econometrics*, 40(1), 97–123.

Mills, T. C. (1996): "Non-Linear Forecasting of Financial Time-Series: An Overview and some New Models," *Journal of Forecasting*, 15, 123–135.

Park, S., and A. Zellner (1979): "Minimum Expected Loss (MELO) Estimators for Functions of Parameters and Structural Coefficients of Econometric Models," *Journal of the American Statistical Association*, 74(365), 185–93.

Schwert, W. G. (1989): "Why Does Stock Market Volatility Change over Time," *Journal of Finance*, XLIV(5).

Skouras, S. (1998): "Risk Neutral Forecasting," Working paper 98-40, European University Institute, Revised 2001.

——— (2001): "Financial Returns and Efficiency as Seen by an Artificial Technical Analyst," *Journal of Economic Dynamics and Control*, 25(1-2), 213–244.

Taylor, S. J. (1986): *Modelling Financial Time Series*. John Wiley and Sons.

Tsybakov, A. (1988): "Passive Stochastic Approximation," Working paper A-207, University of Bonn.

White, H. (1984): *Asymptotic Theory for Econometricians*. Academic Press.

——— (1994): *Estimation, Inference and Specification Analysis*, Econometric Society Monographs, 22. Cambridge University Press.

Zellner, A. (1978): "Estimation of Functions of Population Means and Regression Coefficients Including Structural Coefficients: A Minimum Expected Loss (MELO) Approach," *Journal of Econometrics*, 8, 127–58.

Zheng, X. (1998): "A Consistent Nonparametric Test of Parametric Regression Models under Conditional Quantile Restrictions," *Econometric Theory*, 14, 123–38.