Bounded Tilting Estimation

Susanne M. Schennach
Department of Economics
Brown University
Providence, RI, USA
smschenn@brown.edu

Oscar Wahlstrom
AB Hvalfisken
Sweden

August 7, 2024

Abstract

The search for one-step alternatives to the Generalized Method of Moment (GMM) has identified broad classes of potential estimators such as Generalized Empirical Likelihoods (GEL), Empirical Cressie-Read (ECR), Exponentially Tilted Empirical Likelihood (ETEL) and minimum discrepancy (MD) estimators. While Empirical Likelihood (EL) dominates other ECR estimators in terms of higher-order asymptotics, it lacks robustness to model misspecification. ETEL was shown to combine higher-order efficiency and robustness to misspecification, but demands strong moment generating function existence conditions. We show, both theoretically and via simulations, how to achieve the same goal under weaker moment existence conditions, within the class of MD estimators.

Keywords: Generalized Method of Moments, Information Theory, Misspecification, Generalized Empirical Likelihood

1 Introduction

We consider a moment condition model using a d_g-dimensional vector valued non-linear moment function $g(X, \theta)$ where X is a random vector and $\theta \in \Theta$ is a d_θ-dimensional vector of parameters of interest. The moment condition vector pins down the true parameter value θ_0 from the moments of the distribution of the random vector X through the following equation:

$$\mathbb{E}[g(X, \theta)] = 0. \tag{1}$$

When d_g is larger than d_θ, this problem may not have a solution in general. We say that the moment condition model is well specified if a θ_0 satisfying (1) exists, and that the moment condition model is misspecified if such a θ_0 does not exist. The methodology of estimating θ_0 in a well specified model has seen change throughout the last few decades. While the GMM estimator (Hansen, 1982) enjoyed broad popularity, thanks to its desirable first-order asymptotic properties, alternative estimators have been proposed to improve finite-sample

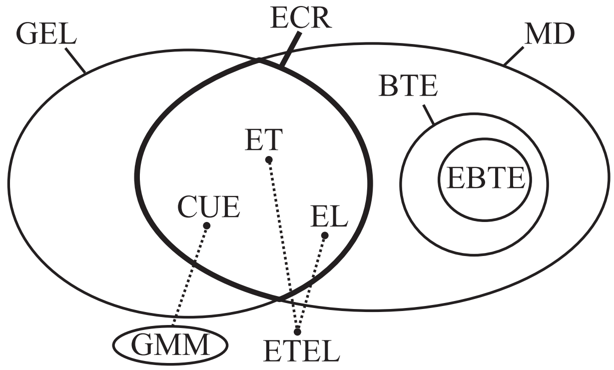

Figure 1: Relationships between existing classes of estimators of moment condition models. Some estimators that are neither GEL nor MD nevertheless have close conceptual ties to members of those classes, as indicated by dotted lines.

performance. These include Empirical Likelihood (EL), Exponential Tilting (ET), and Continuous Updating GMM (CUE) (see, e.g., Owen, 1990, Qin and Lawless, 1994, Imbens et al., 1998, Kitamura and Stutzer, 1997). These estimators are members of a subclass of the Minimum Divergence Estimators or Minimum Discrepancy Estimators (MD) (as detailed in Corcoran, 1998), which this paper will now set its focus on.

The MD estimators are first order equivalent to two-step efficient GMM under regularity conditions, implying that they reach the semi-parametric efficiency bound. They are also one-step estimators, so that reliance on an arbitrary first step estimator is unnecessary. A persistent interest of the recent literature has then been to determine which estimators in the class are most desirable. A particular focus has been set on the Empirical Cressie-Read Estimators (ECR), of which the aforementioned EL, ET, and CUE are members. ECR estimators comprise MD estimators whose dual optimization problem can be cast as a Generalized Empirical Likelihood (GEL) (Newey and Smith, 2004). Figure 1 depicts the relationships between the various classes of estimators. Since ECR estimators all share the same desirable properties mentioned above, the exploration of how to further narrow them down has been funneled into two main directions.

First, there is the question of higher order efficiency. This was explored by Newey and Smith (2004), who showed that bias-corrected EL is second order efficient in the class of Generalized Empirical Likelihood (GEL) estimators. Later, Ragusa (2011) showed that bias-corrected EL is also second order efficient in the class of MD estimators. These analyses suggest that EL is the most desirable estimator from this perspective. The second point of comparison is how the estimators stand up to model misspecification. Imbens et al. (1998) argued that EL could have poor behavior in a misspecified model due to a divergence in its influence function, something which was later formalized by Schennach (2007), who showed that EL is not $\sqrt{n}$ consistent in a misspecified model when the moment function g is unbounded. Schennach observed that ET is robust to misspecification, suggesting that it is the most desirable ECR estimator from this perspective. Schennach also suggested that by using a combination estimator called Exponentially Tilted Empirical Likelihood (ETEL), one can retain the higher order properties of EL, while also being robust to misspecification like ET.

However, in order for ET and ETEL to be robust to misspecification, the Data Generating Process (DGP) of X must satisfy strong assumptions, namely that the moment generating function of the moment function $g\left(X,\theta\right)$ exists. Schennach (2007) showed that this condition was sufficient, while Sueishi (2013) observed that this condition is necessary for the asymptotic reweighting problem associated with these estimators to even have a solution, which is needed to achieve robustness to misspecification. Sufficient conditions which guarantee existence of solutions within the broader ECR class were given by Chaudhuri et al. (2023). They showed that for any choice of discrepancy function, at least some assumptions must be made on the existence of moments of the transformed random variable $Y_\theta = g(X,\theta)$ for all fixed θ. This indicates that, within the ECR subclass, robustness to misspecification guarantees are limited and thus suggests that one should move beyond the ECR subclass into the full class of MD estimators.

The purpose of this paper is to identify a class of estimators within the MD class, which we call Bounded Tilting Estimators, or BTE, that is robust to misspecification without requiring strong moment existence assumptions. Ragusa (2011) suggested that, in analogy with the observation of Imbens et al. (1998) for EL, MD estimators with good properties under misspecification can be obtained by avoiding discrepancies that lead to a divergence in the corresponding estimator's influence function. We formalize this idea by using the techniques of Chaudhuri et al. (2023), extended to general MD estimators, to derive conditions on the discrepancy function that secure the existence of solutions under global misspecification. This result, combined with a derivation of the asymptotic properties of MD estimators under global misspecification, implies that the estimator is robust to misspecification. It is further shown that robustness to misspecification is achievable, while also retaining the higher order efficiency guaranteed by results from Ragusa (2011), thus yielding the class of Efficient BTE, or EBTE.

The rest of this paper is organized as follows. Section 2 reviews the Minimum Divergence Estimators, their first and higher order efficiency, and results about misspecification. Section 3 defines the bounded tilting estimators, proves their higher order efficiency and their behavior under misspecification. Section 4 gives an example of an easily implementable tilting function and shows its performance in simulations, while Section 5 concludes.

2 Minimum Divergence Estimation

2.1 The Estimator

For any value of value $\theta \in \Theta$ one can reweight the sample points $\{x_i\}_{i=1}^n$ so that the moment condition (1) is satisfied. A MD estimator is given by the value θ that minimizes the amount of reweighting, as quantified using a discrepancy function:

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \min_{\{w_i\}} \frac{1}{n}\sum_{i=1}^n q(nw_i) \tag{2}$$

$$\text{subject to } \sum_{i=1}^n w_i g(x_i,\theta) = 0 \text{ and } \sum_{i=1}^n w_i = 1.$$

The discrepancy function q is assumed to satisfy the following:

Assumption 1 *The discrepancy function q is twice continuously differentiable, strictly convex and, without loss of generality, normalized as $q(1) = 0$, $\dot{q}(1) = 0$ and $\ddot{q}(1) = 1$, where dots denote derivatives.*

While the solution of this optimization problem follows well-known techniques (e.g. Newey and Smith, 2004, Kitamura, 2007), we provide here a few key results that are helpful for our presentation. Introducing Lagrange multiplier λ and μ we can write the Lagrangian:

$$L(\theta, w, \lambda, \mu) = \frac{1}{n}\sum_{i=1}^{n} q(nw_i) - \lambda' \sum_{i=1}^{n} w_i g(x_i, \theta) - \mu(\sum_{i=1}^{n} w_i - 1), \tag{3}$$

whose first-order condition, with respect to w_i is

$$\dot{q}(nw_i) - \lambda' g(x_i, \theta) - \mu = 0 \quad \forall i = 1, ..., n.$$

One can then obtain a closed form for w_i in terms of the Lagrange multipliers:

$$w_i = \frac{1}{n}\dot{q}^{-1}(\mu + \lambda' g(x_i, \theta)).$$

The mapping $\dot{q}^{-1}$, known as the tilting function, can then be used to write an equivalent just-identified GMM estimator:

$$\sum_{i=1}^{n} \rho(x_i, \hat{\theta}, \hat{\lambda}, \hat{\mu}) = 0 \tag{4}$$

where, letting $G(x, \theta) = \partial g(x, \theta)/\partial\theta'$,

$$\rho(x, \theta, \lambda, \mu) = \begin{bmatrix} \dot{q}^{-1}(\mu + \lambda' g(x, \theta)) G'(x, \theta)\lambda \\ \dot{q}^{-1}(\mu + \lambda' g(x, \theta)) g(x, \theta) \\ \dot{q}^{-1}(\mu + \lambda' g(x, \theta)) - 1 \end{bmatrix}. \tag{5}$$

These equations are analogous to the well-known first-order conditions of GEL estimators, except that an additional entry (the bottom one) in this vector is needed to ensure that the weights w_i sum up to one, a constraint that is no longer automatically satisfied when q lies outside of the ECR family.

As shown in Ragusa (2011), the dual problem to (2) also admits the following saddle-point formulation, paralleling known results for ECR (e.g. Owen, 1990, Qin and Lawless, 1994, Imbens et al., 1998, Kitamura and Stutzer, 1997, Kitamura, 2007):

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \min_{(\mu, \lambda) \in \Lambda_n(\theta)} \frac{1}{n}\sum_{i=1}^{n} q^*(\mu + \lambda' g(x_i, \theta)) - \mu, \tag{6}$$

with

$$\Lambda_n(\theta) = \{(\mu, \lambda') : \mu + \lambda' g(x_i, \theta) \in \text{Dom}(q^*), \forall i = 1, ..., n\}$$

and where q^* is the convex conjugate of q:

$$q^*(v) := \sup_{u \in \text{Dom}(q)} [uv - q(u)]. \tag{7}$$

where Dom$\,(q)$ denotes the domain of function q (and we similarly define Range$\,(q)$ to be the range of function q). This result can be verified by noting that the first-order conditions for (6) coincide with (5). Assumption 1 implies that the function q^* is twice continuously differentiable, strictly convex and satisfies $q^*(0) = 0$, $\dot{q}^*(0) = 1$ and $\ddot{q}^*(0) = 1$.

2.2 First and Higher order Asymptotics

We now briefly summarize the results from Ragusa (2011) on the asymptotics of the MD estimators that are relevant for our paper.

Assumption 2 *The sample $X_1, \ldots, X_n$ is i.i.d.*

Assumption 3 *(a) $\theta_0 \in \Theta$ is the unique solution to $\mathbb{E}[g(X,\theta)] = 0$; (b) Θ is compact; (c) $g(.,\theta)$ is continuous in θ at all $\theta \in \Theta$ w.p.1; (d) $\mathbb{E}[\sup_\theta \|g(X,\theta)\|^2] < \infty$; (e) $\Omega = \mathbb{E}[g(X,\theta_0)g(X,\theta_0)']$ is non-singular*

Assumption 4 *(a) $\theta \in int(\Theta)$; (b) $g(x,\theta)$ is continuously differentiable in a neighborhood $\mathcal{N}$ of θ_0; (c) $\mathbb{E}[\sup_{\theta\in\mathcal{N}} \|G(X,\theta)\|] < \infty$; (d) $\text{Rank}(G) = \dim(\theta)$, where $G = \mathbb{E}[G(X,\theta_0)]$*

Assumption 5 *There is $b(x)$ with $\mathbb{E}[\|b(X)\|^6] < \infty$ such that for $0 \le j \le 4$ and all x, $\frac{\partial^j g(x,\theta)}{\partial\theta^j}$ exists on a neighborhood $\mathcal{N}$ of θ_0, $\sup_{\theta\in\mathcal{N}} \left\|\partial^j g(x,\theta)/\partial\theta^j\right\| \le b(x)$, and for each $\theta \in \mathcal{N}$, $\left\|\partial^4 g(x,\theta)/\partial\theta^4 - \partial^4 g(x,\theta_0)/\partial\theta^4\right\| \le b(x)\|\theta-\theta_0\|$. Also, q^* is four times continuously differentiable with Lipschitz fourth derivative in a neighborhood of zero.*

Theorem 1 *(Adapted from Ragusa (2011), Theorems 5, 6 and Corollary 1). Let Assumptions 1 and 2 hold.*
(i) Under Assumption 3, $\hat{\theta} \to_p \theta_0$, $\hat{\mu} = o_p(n^{-\frac{1}{2}})$, and $\hat{\lambda} = O_p(n^{-\frac{1}{2}})$.
(ii) Under Assumptions 3 and 4, $\hat{\theta}$ is first-order efficient, i.e., $\sqrt{n}\left(\hat{\theta} - \theta_0\right) \to_d \mathcal{N}\left(0, (G'\Omega^{-1}G)^{-1}\right)$.
(iii) Under Assumptions 3, 4 and 5, bias-corrected $\hat{\theta}$ is second-order efficient iff $q_3^* = 2$, where $q_k^* \equiv \partial^k q^*(0)/\partial x^k$.

Part (i) establishes consistency, while part (ii) shows that the MD estimators have the same first-order asymptotic properties as the GEL estimators. Part (iii) establishes an analogue of the higher-efficiency results of Newey and Smith (2004) for the more general class of MD estimators. The latter two results do not follow from known results on GEL estimators, since MD estimators involve an additional nuisance parameter $\hat{\mu}$. It is noteworthy that the value of q_3^* (under the normalization of Assumption 1) is the only condition which matters for the determination of second order efficiency and hence this property will hold as long as $q^*(0) = 0$, $\dot{q}^*(0) = 1$, $\ddot{q}^*(0) = 1$ and $q_3^*(0) = 2$.

Remark 1 *Newey and Smith (2004) had required $q_4^* = 6$ for their result, but Ragusa (2011) shows that this is not necessary. For an MD estimator with $q_3^*(0) = 2$, the difference between its third order term and that of EL is uncorrelated with the first order term so that the higher order variances are the same, independently of the value of q_4^*. A similar phenomenon occurs in the case of the ETEL estimator (Schennach, 2007).*

2.3 Misspecification, Pseudo-True Values, and Robustness

Let us now formalize the notion of robustness to misspecification. Should there not exist any $\theta \in \Theta$ for which $\mathbb{E}[g(X, \theta)] = 0$, the model is said to be misspecified. In this case, we define the pseudo-true value θ_* as the θ that is the "closest" to satisfying the moment constraints in the population, where closeness is assessed using the previously introduced discrepancy function q by solving a population version of the problem (2):

$$\theta_* = \arg\min_{\theta \in \Theta} \mathbb{E}[q(W_\theta)], \tag{8}$$

where

$$W_\theta = \arg\min_{W} \mathbb{E}[q(W)] \tag{9}$$
$$\text{subject to } \mathbb{E}[Wg(X, \theta)] = 0 \text{ and } \mathbb{E}[W] = 1$$

for all $\theta \in \Theta$. Here, the W are functions mapping a value of X onto a real number, and represent the population version of the sample point weights (hence, in particular, we have $W(X_i) = w_i$ for a sample point X_i with weight w_i). The dependence of W on X is suppressed in the notation, for simplicity. Following Chaudhuri et al. (2023), we refer to W as a *change-of-measure* random variable because multiplying the W effectively amounts to changing the probability measure of the population. Note that the pseudo-true value in general depends on the choice of discrepancy function q.

Definition 1 *An estimator is said to be* robust to misspecification *if θ_* exists and $\left\|\hat{\theta} - \theta_*\right\| = O_p(n^{-1/2})$.*

As pointed out by Chaudhuri et al. (2023), the existence of a θ_* is far from granted and it is thus essential to identify the conditions needed for its existence. First, the definition of θ_* requires W_θ to be defined for all $\theta \in \Theta$. Based on the work of Csiszár (1995), conditions for the existence of W_θ for the Cressie-Read family of discrepancies were given by Chaudhuri et al. (2023). Below, we shall use a similar approach, while exploiting the fact that Csiszar's work is, in fact, applicable to the entire class of MD estimators. The aim of this work is to provide a stronger robustness guarantee than the ETEL estimator (Schennach, 2007). ETEL still requires a bounded moment generating function, while we seek here to avoid any moment existence assumptions (beyond those needed to define the moment conditions and the asymptotic variance of efficient estimators).

The existence of a change-of-measure random variable W_θ demands conditions both on q and on the random variable $Y_\theta \equiv g(X, \theta)$ for every fixed $\theta \in \Theta$. The conditions necessary depend on what assumptions one is willing to make on Y_θ, since there is a trade-off in the assumptions made on Y_θ and how stringent one must be with q. We distinguish the cases, depending on whether Y_θ is bounded or not. For this purpose, let $\|Y_\theta\|$ be the Euclidian norm of Y_θ and denote by $Y_{j,\theta}$ the j-th component of the vector Y_θ.

Assumption 6 *The random variable Y_θ is absolutely continuous with respect to some σ-finite measure for every fixed $\theta \in \Theta$.*

Theorem 2 *(Csiszár (1995) Theorem 3 ii)) If $\|Y_\theta\|$ is bounded for every θ, then, under Assumptions 1 and 6, W_θ exists for every $\theta \in \Theta$.*

To move beyond the boundedness assumption, we need to add two more conditions.

Assumption 7 $\text{Dom}(\dot{q}) = (0, a)$ *with* $a \in (1, \infty]$ *and* $\lim_{x\to a} \dot{q}(x) = \infty$.

Assumption 8 $\mathbb{E}[q^*(\alpha|Y_{j,\theta}|)] < \infty$ *for all* $\alpha > 0$ *and* $j = 1, .., d_g$ *for every* $\theta \in \Theta$*, where* q^* *is the convex conjugate of* q *(see Equation (7)).*

Theorem 3 *(Csiszár (1995) Theorem 3 iii)) Under Assumptions 1, 6, 7, and 8, W_θ exists for every $\theta \in \Theta$.*

We now explain the need for these conditions. Assumption 7 can be understood by analogy to the first order conditions which come from (9). The solution W_θ must satisfy the following first order condition almost surely:

$$\dot{q}(W_\theta) = \mu^* + \lambda^* Y_\theta,$$

where μ^* and λ^* are the values of the Lagrange multipliers that are pinned down from setting the partial derivatives of the Lagrangian to zero under mild differentiability conditions.

If Y_θ is unbounded, then the right-hand side of this equation is also unbounded and can diverge to $+\infty$. Since $\dot{q}$ is strictly increasing (as q is strictly convex), if we want the left-hand side to also have the ability to diverge to $+\infty$, we must allow $\dot{q}(W_\theta)$ to diverge to $+\infty$ as W_θ approaches the upper bound of the domain of $\dot{q}$.

Remark 2 *It is tempting to use an analogous explanation to suggest that one should also impose the condition* $\lim_{x\to 0} \dot{q}(x) = -\infty$*, as Ragusa (2011) suggests. It turns out however that this is not necessary for the existence of a solution to the problem but rather to ensure that the solution W_θ is strictly positive (almost surely), or equivalently, that the sample point weights are positive. To understand why this is the case consider what happens when Y_θ diverges to $-\infty$ for a given θ. Since $\dot{q}$ is strictly increasing, this must be coupled with W_θ moving towards zero. However, when it reaches zero, the constraint associated with λ^* ($\mathbb{E}[WY_\theta)] = 0$) is automatically fulfilled so the Lagrange multiplier becomes zero. This then means that as long as* $\lim_{x\to 0} \dot{q}(x) = \mu^*$*, the first-order condition holds.*

On a technical level, Assumption 8 comes from Csiszar's clever reliance on Orlicz spaces. Intuitively, this can be motivated as follows. For the minimizer W_θ in (9) to exist we must guarantee that it satisfies $\mathbb{E}[W_\theta Y_\theta] = 0$, while the only fact we know is that $\mathbb{E}[q(W)] < \infty$ for all W under consideration in the optimization problem. We then need to ensure that any sequence of W_i which converges to W_θ satisfies $\mathbb{E}[W_i Y_\theta] \to \mathbb{E}[W_\theta Y_\theta]$, which implies what we need. If the sequence W_i converges in L^1, this is guaranteed, and Assumption 8 specifically ensures this convergence.

We finish this section with the fact that the existence theorems of the change-of-measure random variable W_θ, coupled with a continuity argument, can then guarantee existence of a pseudo-true value:

Corollary 1 *Let $\mathcal{Z} = \cup_\theta \mathcal{Z}_\theta$ be the union of $\mathcal{Z}_\theta$, the spaces of absolutely continuous random variables with respect to the same σ-finite measure as Y_θ. Assume that either (i) Assumptions 1 and 6 hold for Y_θ bounded or (ii) Assumptions 1, 6, 7, and 8 hold for Y_θ unbounded.*
If the correspondence $C : \Theta \rightrightarrows \mathcal{Z}$ with $C(\theta) = \{W \in \mathcal{Z} : \mathbb{E}[WY_\theta] = 0\, and\, \mathbb{E}[W] = 1\}$ is continuous and Θ is compact, θ_* exists.

Proof. Under the conditions, we have that W_θ exists for every θ. The correspondence being continuous (i.e., both upper and lower hemicontinuous) makes W_θ continuous by Berge's Maximum Theorem (Berge, 1963, Chapter 6). The optimization problem which defines θ_* is then simply minimizing W_θ over Θ, where W_θ is continuous and Θ is compact — which guarantees the existence of θ_* by the Extreme Value Theorem. ■

3 Beyond ECR: Bounded Tilting

3.1 Existence of pseudo-true value

Now that we are equipped with the necessary tools, we can identify estimators that are both robust to misspecification and higher-order efficient. We begin by explaining the need to look beyond ECR estimators. The discrepancy associated with ECR estimators has the form:

$$q(x) = \begin{cases} \frac{x^{\gamma+1}-1}{\gamma(\gamma+1)}, & \forall \gamma \neq 0, -1 \\ -\log(x), & \gamma = -1 \\ x\log(x), & \gamma = 0 \end{cases}$$

for some $\gamma \in \mathbb{R}$, where the limiting cases $\gamma = 0, -1$ correspond to EL and ET, respectively. When applying Csiszar's conditions to the ECR estimators, Chaudhuri et al. (2023) show that the only estimators in the class which can have a pseudo-true value when Y_θ is unbounded are the ones with $\gamma \geq 0$. Condition 8 in the ECR setting is equivalent to $\mathbb{E}[|Y_{j,\theta}|^{\frac{\gamma+1}{\gamma}}] < \infty$ for all j when $\gamma > 0$ and $\mathbb{E}[\exp(t'Y_{j,\theta})] < \infty$ for all j and $t \in \mathbb{R}$ when $\gamma = 0$. When $\gamma < 0$, Assumption 7 does not hold and these authors show that no solution can exist to the population optimization problem. Hence, for unbounded random variables, this argument rules out the only member of the ECR class which is higher order efficient (EL, with $\gamma = -1$). Hence, we need to consider a broader class of estimators, such as MD estimators, to simultaneously achieve both goals.

We maintain the very weak Assumptions 1 and 6 throughout and seek a choice of q satisfying Assumption 7 that will weaken the additional constraints on Y_θ imposed by Assumption 8 as much as possible. Assumption 8 is the only one involving Y_θ and it can only be guaranteed to hold for unbounded Y_θ if q^* is bounded. Constructing a bounded q^* is facilitated by the following lemma:

Lemma 1 *For q twice continuously differentiable and strictly convex, we have:*

$$q^*(v) = \int_0^v \dot{q}^{-1}(x)dx \tag{10}$$

for all $v \in \text{Range}(\dot{q})$, assuming the normalization $q^*(0) = 0$.
Further, if both $\text{Dom}(q)$ and $\dot{q}^{-1}$ are bounded, then q^* is also bounded.

Proof. A twice continuously differentiable q means that $\dot{q}$ is continuous, while a strictly convex q implies that $\dot{q}$ is strictly increasing, which implies that $\dot{q}$ is invertible. We then notice that $q^*(v) = v\dot{q}^{-1}(v) - q(\dot{q}^{-1}(v))$ and take its derivative:

$$\dot{q}^*(v) = \dot{q}^{-1}(v) + v\frac{\partial}{\partial v}\dot{q}^{-1}(v) - \frac{\partial}{\partial v}\dot{q}^{-1}(v)\dot{q}((\dot{q}^{-1}(v)).$$

Cancelling the last two terms yields the equality: $\dot{q}^*(v) = \dot{q}^{-1}(v)$. Then, integrating with the initial condition $q^*(0) = 0$ establishes (10). Next, for an invertible $\dot{q}$, we have $\text{Range}(\dot{q}^{-1}) = \text{Dom}(q)$, and if $\text{Dom}(q)$ is bounded then $\text{Range}(\dot{q}^{-1})$ is bounded. Integrating a bounded function over a bounded domain yields a bounded integral. Hence, q^* is bounded. ■

Lemma 1 explicitly links the convex conjugate q^* with the tilting function $\dot{q}^{-1}$, thus allowing us to find simple conditions on q that secure suitable properties of q^*. This brings us to the definition of our Bounded Tilting Estimators, which are nested within the MD class but strictly outside of the ECR class (see Figure 1):

Definition 2 *A Bounded Tilting Estimator (BTE) is a MD estimator which uses a bounded tilting discrepancy function q which satisfies:*

a) Assumption 1
b) $\text{Dom}(\dot{q}) = (0, a)$ *with* $a \in (1, \infty)$
c) $\lim_{x \to a} \dot{q}(x) = \infty$

The three conditions ensure, respectively, that a) the tilting function $\dot{q}^{-1}$ is well defined, b) is bounded, and c) allows the first order conditions to hold. Conditions b) and c) jointly imply Assumption 7. Finally, Assumption 8 follows from Lemma 1, since the expectation of a bounded function is also bounded. Hence, by Theorem 3, W_θ exists for $\theta \in \Theta$, which is the key condition for the existence of θ_* in Corollary 1.

3.2 Root n consistency under misspecification

Having provided conditions for the existence of the pseudo-true value θ_*, we now address the second requirement of robustness to misspecification (in Definition 1), namely, $\left\|\hat{\theta} - \theta_*\right\| = O_p(n^{-1/2})$. This is accomplished by deriving the asymptotic distribution of BTE under global misspecification. To simplify the notation, we define $\phi = (\theta, \lambda, \mu)$ and, accordingly, $\rho(x, \phi) \equiv \rho(x, \theta, \lambda, \mu)$, where the latter is defined in Equation (5). The estimator $\hat{\phi}$ is the solution to Equation (4).

Assumption 9 *There exists a set* $\Phi \equiv \Theta \times \Lambda \times \mathcal{U}$ *(where* $\Theta \subset \mathbb{R}^{d_\theta}$, $\Lambda \subset \mathbb{R}^{d_g}$ *and* $\mathcal{U} \subset \mathbb{R}$, *are compact), such that* $E[\rho(X, \phi)] = 0$ *has a unique solution* $\phi = \phi_0$ *in the interior of* Φ.

Assumption 10 $g(x, \theta)$ *is twice continuously differentiable for* $\theta \in \Theta$.

Assumption 11 $E\left[\sup_{\phi\in\Phi} \|g(X, \phi)\|\right] < \infty, E\left[\sup_{\phi\in\Phi} \|G(X, \phi)\|\right] < \infty$ *and* $E\left[\sup_{\phi\in\Phi} \left\|\partial G(X, \phi)/\partial \phi_j\right\|\right] < \infty$ *for* $j = 1, \ldots, d_\theta + d_g + 1$.

Assumption 12 $E[\rho(X, \phi_0)\rho'(X, \phi_0)]$ *is finite and nonsingular.*

Theorem 4 *Under Assumptions 1, 2, 9, 10, 11, 12 and if $\dot{q}^{-1}$ is bounded (i.e. for a Bounded Tilting Estimator), then*

$$\sqrt{n}\left(\hat{\phi}-\phi_0\right) \xrightarrow{d} \mathcal{N}\left(0,(R^{-1}\Sigma R^{-1}\right)$$

where $R = E\left[\partial\rho\left(X,\phi\right)/\partial\phi'\right]$ and $\Sigma = E\left[\rho\left(X,\phi\right)\rho'\left(X,\phi\right)\right]$.

Proof. This result is established by verifying the conditions of Lemma 2.4 and Theorems 2.6 and 3.2 in Newey and McFadden (1994) in the special case of a just-identified GMM estimator (in which the weighting matrix is the identity, without loss of generality). The requisite conditions on the vector of moment functions $\rho\left(X,\theta\right)$ can be trivially expressed in terms of conditions on $g\left(X,\theta\right)$ and the properties of q via Equation (5). ■

Remark 3 *The asymptotic variance of θ can be extracted from the appropriate submatrix of $R^{-1}\Sigma R^{-1}$. Note that using the partitioned inverse formula for R^{-1} does not lead to a simpler expression under misspecification.*

Remark 4 *Once can obtain more explicit expressions for the matrices R and Σ:*

$$\begin{aligned}
\frac{\partial\rho\left(x,\phi\right)}{\partial\phi'} &= \begin{bmatrix} QG'\lambda\lambda'G+\omega\Gamma & QG'\lambda g'+\omega G' & QG'\lambda \\ Qg\lambda'G+\omega G & Qgg' & Qg \\ Q\lambda'G & Qg' & Q \end{bmatrix} \\
\rho\left(x,\phi\right)\rho'\left(x,\phi\right) &= \begin{bmatrix} \omega^2G'\lambda\lambda'G & \omega^2G'\lambda g' & \omega G'\lambda\left(\omega-1\right) \\ \omega^2g\lambda'G & \omega^2gg' & \omega g\left(\omega-1\right) \\ \left(\omega-1\right)\omega\lambda'G & \left(\omega-1\right)\omega g' & \left(\omega-1\right)^2 \end{bmatrix},
\end{aligned}$$

where, for conciseness, we have omitted the (x,θ)-dependence and defined $Q = 1/\ddot{q}(\dot{q}^{-1}\left(\mu+\lambda'g(x,\theta)\right))$, $\omega = \dot{q}^{-1}(\mu+\lambda'g(x,\theta))$, $\Gamma = \partial\left(G'(x,\theta)\lambda\right)/\partial\theta'$. However, implementers may find it more practical to compute the derivative of $\rho\left(x,\theta\right)$ directly via finite differences.

3.3 Higher-order efficiency under correct specification

Since all Bounded Tilting Estimators are equally well-equipped to deal with misspecification, we will now narrow down the scope to those which are also higher order efficient in the absence of misspecification. The additional requirements for the higher-order efficiency result of Theorem 1 are as follows. First, Assumption 5 places constraints on the moment functions and data generating process which we cannot control through the choice of q. In contrast, the remaining conditions can indeed be met through the careful choice of q: the differentiability requirements on q^* imposed by Assumption 5 and the key condition, $q_3^* = 2$, which is implied by $\frac{\partial^2}{\partial x^2}\dot{q}^{-1}(v)|_{v=0} = 2$, by Lemma 1. These observations lead us to define another set of estimators nested within the BTE class (see Figure 1):

Definition 3 *A Higher Order Efficient Bounded Tilting Estimator (EBTE) is a Bounded Tilting Estimator with bounded tilting discrepancy function q that also satisfies:*

a) q is four times continuously differentiable with Lipschitz fourth derivative in a neighborhood of zero

b) $\frac{\partial^2}{\partial x^2}\dot{q}^{-1}(v)|_{v=0} = 2$

Hence, through the choice of q, one can construct an estimator that is robust to misspecification under very weak moment existence conditions and that can achieve higher order efficiency.

4 Suggested EBTE and Simulations

4.1 Suggested EBTE

We now give an example of a tilting function $\dot{q}^{-1}$ which gives a member in this EBTE subclass. The tilting function must be strictly increasing and bounded. To find such a candidate, it makes sense to look at CDFs over the entire real line that satisfy the necessary Lipschitz differentiability criterion. However, the candidate must also satisfy the necessary derivative conditions $q^*(0) = 0$, $\frac{d}{dx}q^*(0) = 1$, $\frac{d^2}{dx^2}q^*(0) = 1$, and $\frac{d^3}{dx^3}q^*(0) = 2$, which translates to $\dot{q}^{-1}(0) = 1$, $\frac{d}{dx}\dot{q}^{-1}(0) = 1$ and $\frac{d^2}{dx^2}\dot{q}^{-1}(0) = 2$ by Lemma 1. The condition $q^*(0) = 0$ is inconsequential because we simply normalize $\int_0^x \dot{q}^{-1}(v)dv|_{x=0} = 0$. This leaves us with 3 equations which must hold simultaneously. A simple family of CDFs which has a sufficient number of adjustable parameters is a combination of the Cauchy and the Logistic CDFs:

$$\dot{q}^{-1}(x) = \frac{2}{\pi}\arctan(\frac{x-a}{b}) + 1 + \frac{2}{1+e^{-\frac{x-c}{d}}},$$

with 4 adjustable parameters a, b, c, d. Our derivative conditions only place 3 constraints, so we arbitrarily set $c = \ln(2)d$, as this choice leads to a simpler closed form expression for the remaining parameters: $a = \sqrt{3}b$, $b = \frac{2\sqrt{6+5\sqrt{3}\pi}-3}{10\pi}$ and $d = \frac{2(2\sqrt{3}\pi+\sqrt{6+5\sqrt{3}\pi})}{9(\sqrt{3}\pi-2)}$. We will call the associated estimator the Cauchy-Logistic EBTE.

4.2 Monte Carlo Simulations

We will now explore how an estimator, based on our suggested tilting function, behaves in simulations. We consider two setups. The first setup explores the higher order efficiency of the proposed estimator, and is based on the design of Hall and Horowitz (1996). We use 13 i.i.d. sequences of independent random variables $\{\{x_{i,j}\}_{i=1}^n\}_{j=1}^{13}$ where:

$$(x_{i,1}, x_{i,2}) \sim \mathcal{N}([0,0], 0.16I), \quad x_{i,3} \sim t_5, \quad x_{i,j} \sim \chi_1^2 \;\; j = 4, ..., 13.$$

We define $h(\theta, x, y) = \exp(-0.72 - \theta(x+y) + 3y) - 1$ and implement four designs:

$$1) \quad g(X,\theta) = [h(\theta, X_1, X_2),\; X_2\, h(\theta, X_1, X_2)]'$$

$$2) \quad g(X,\theta) = [h(\theta, X_1, X_2),\; X_2\, h(\theta, X_1, X_2),\; X_3\, h(\theta, X_1, X_2),\; X_4\, h(\theta, X_1, X_2)]'$$

$$3) \quad g(X,\theta) = [h(\theta, X_1, X_2),\; X_2\, h(\theta, X_1, X_2),\; X_4\, h(\theta, X_1, X_2),\, ..., X_7\, h(\theta, X_1, X_2)]'$$

$$4) \quad g(X,\theta) = [h(\theta, X_1, X_2),\; X_2\, h(\theta, X_1, X_2),\; X_4\, h(\theta, X_1, X_2),\, ..., X_{13}\, h(\theta, X_1, X_2)]'.$$

In all cases, the parameter's true value is $\theta_0 = 3$. Design 1 is symmetric, whereas Designs 2-4 impose skewness and an increase in kurtosis. Design 2 is as suggested by Ragusa (2011)

Table 1: Bias, Variance, Median, and Inter-Quartile Range (IQR) of $\hat{\theta} - \theta_0$ for EL, ET, and Cauchy-Logistic EBTE for $n = 100$ and 10000 replications

	Design 1: n=100		
Estimator	EL	ET	Cauchy-Logistic
Bias	0.056	0.072	0.063
Variance	0.083	0.088	0.086
Median	0.033	0.048	0.039
IQR	0.382	0.388	0.385

Table 2: Bias, Variance, Median, and Inter-Quartile Range (IQR) of $\hat{\theta} - \theta_0$ for EL, ET, and Cauchy-Logistic EBTE for $n = 100$ and 10000 replications

	Design 2: n=100		
Estimator	EL	ET	Cauchy-Logistic
Bias	0.113	0.181	0.149
Variance	0.093	0.118	0.107
Median	0.087	0.144	0.115
IQR	0.340	0.430	0.418

and Designs 3 and 4 are adapted from Schennach (2007). We perform 10000 replications for each design at $n = 100$ and $n = 400$. We discard the few samples where the estimators failed to converge.[1] The results are shown in the tables below, together with the same results for EL and ET for comparison:

In Tables 1-8, we can see that our Cauchy-Logistic EBTE performs more in line with the EL estimator than the ET, with the difference between the estimator performances decreasing significantly with sample size, as expected, due to higher-order efficiency. As we move from Design 1 and 2 to Designs 3 and 4 we see all three estimators performing worse as a result of the increased skewness and kurtosis in the $n = 100$ designs. This effect is much less noticeable in the $n = 400$ designs but still there nevertheless. Designs 1 and 2 were based on Ragusa (2011) and the results are very similar to his for both the EL and the ET estimators, while at the same time the Cauchy-Logistic EBTE performs approximately as

[1]This purely numerical artifact was detected by checking if providing three different starting points to the optimization routine gave different optima.

Table 3: Bias, Variance, Median, and Inter-Quartile Range (IQR) of $\hat{\theta} - \theta_0$ for EL, ET, and Cauchy-Logistic EBTE for $n = 100$ and 10000 replications

	Design 3: n=100		
Estimator	EL	ET	Cauchy-Logistic
Bias	0.162	0.267	0.226
Variance	0.103	0.147	0.131
Median	0.137	0.222	0.188
IQR	0.409	0.462	0.442

Table 4: Bias, Variance, Median, and Inter-Quartile Range (IQR) of $\hat{\theta} - \theta_0$ for EL, ET, and Cauchy-Logistic EBTE for $n = 100$ and 10000 replications

	Design 4: n=100		
Estimator	EL	ET	Cauchy-Logistic
Bias	0.284	0.450	0.399
Variance	0.127	0.213	0.182
Median	0.249	0.382	0.343
IQR	0.458	0.572	0.540

Table 5: Bias, Variance, Median, and Inter-Quartile Range (IQR) of $\hat{\theta} - \theta_0$ for EL, ET, and Cauchy-Logistic EBTE for $n = 400$ and 10000 replications

	Design 1: n=400		
Estimator	EL	ET	Cauchy-Logistic
Bias	0.019	0.022	0.019
Variance	0.020	0.020	0.020
Median	0.014	0.018	0.015
IQR	0.192	0.192	0.191

Table 6: Bias, Variance, Median, and Inter-Quartile Range (IQR) of $\hat{\theta} - \theta_0$ for EL, ET, and Cauchy-Logistic EBTE for $n = 400$ and 10000 replications

	Design 2: n=400		
Estimator	EL	ET	Cauchy-Logistic
Bias	0.031	0.053	0.039
Variance	0.020	0.021	0.021
Median	0.024	0.045	0.032
IQR	0.193	0.196	0.195

Table 7: Bias, Variance, Median, and Inter-Quartile Range (IQR) of $\hat{\theta} - \theta_0$ for EL, ET, and Cauchy-Logistic EBTE for $n = 400$ and 10000 replications

	Design 3: n=400		
Estimator	EL	ET	Cauchy-Logistic
Bias	0.046	0.085	0.064
Variance	0.021	0.023	0.022
Median	0.043	0.080	0.059
IQR	0.190	0.201	0.196

Table 8: Bias, Variance, Median, and Inter-Quartile Range (IQR) of $\hat{\theta} - \theta_0$ for EL, ET, and Cauchy-Logistic EBTE for $n = 400$ and 10000 replications

	Design 4: n=400		
Estimator	EL	ET	Cauchy-Logistic
Bias	0.084	0.156	0.123
Variance	0.021	0.026	0.024
Median	0.079	0.148	0.115
IQR	0.195	0.216	0.206

Table 9: Standard deviations of EL, ET, and Cauchy-Logistic EBTE estimators for Models C, M_1 and M_2 defined in the text with $n = 1000$

n=1000				
Estimator	EL	ET	ETEL	Cauchy-Logistic
Model C	0.032	0.032	0.032	0.032
Model M_1	0.054	0.031	0.038	0.028
Model M_2	0.058	0.046	0.053	0.039

well as Ragusa's Quartic Tilting estimator.

Our second simulation setup is specified to test how the estimators handle misspecification. This setup was also used by both Ragusa and Schennach with moment function:

$$g(x, \theta) = [x - \theta,\, (x - \theta)^2 - 1]'.$$

Our random variable x_i follows either a correctly specified model (we call this model C) or one of two misspecified models (we call these models M_1 and M_2):

$$C : X_i \sim \mathcal{N}(0, 1)$$
$$M_1 : X_i \sim \mathcal{N}(0, 0.64)$$
$$M_2 : X_i \sim (0.8)\,\frac{1}{2} t_4,$$

where $(1/2)\,t_4$ is a student t distribution with 4 degrees of freedom, suitably scaled to have unit variance. To see how our Cauchy-Logistic estimator performs, we compare it to EL, ET and ETEL. Our simulations consider 10000 replications with sample size $n = 1000$, and 2000 replications with sample size $n = 5000$. The results are reported in Tables 9 and 10 respectively, where the standard deviations of the estimators are shown. We do not report means or medians, because different estimators have different pseudo-true values under misspecification, thus making it difficult to meaningfully compare such statistics.

Our Cauchy-Logistic EBTE performs as well as the other estimators in the well specified model, as was also highlighted in the previous simulation setup. As expected, our EBTE outperforms prior estimators in terms of standard deviation for all misspecified models. A more specific analysis of robustness to misspecification can be carried out by comparing the rate at which the standard deviation of $\hat{\theta}$ (denoted $\sigma_{\hat{\theta}}$) decays with sample size. We see that, for both M_1 and M_2, EL exhibits little decay of $\sigma_{\hat{\theta}}$ with sample size. In contrast, for both

Table 10: Standard deviations of EL, ET, and Cauchy-Logistic EBTE estimators for Models C, M_1 and M_2 defined in the text with $n = 5000$

n=5000				
Estimator	EL	ET	ETEL	Cauchy-Logistic
Model C	0.014	0.014	0.014	0.014
Model M_1	0.052	0.014	0.019	0.012
Model M_2	0.038	0.026	0.032	0.018

ET and ETEL, $\sigma_{\hat{\theta}}$ decreases with sample size by roughly the expected $\sqrt{5}$ factor between $n = 1000$ and $n = 5000$ for Model M_1 but not for Model M_2. This reflects the fact that the moment generating function of X_i exists in Model M_1, while it does not in Model M_2. Furthermore, in the case of the EBTE estimator, $\sigma_{\hat{\theta}}$ decreases by roughly the expected $\sqrt{5}$ factor for both Models M_1 and M_2, which illustrates its robustness to misspecification under weak moment existence conditions.

5 Conclusion

Since the introduction of Hansen's GMM, researchers have sought ways to improve upon the technique through the introduction of ECR estimators. Unfortunately, the search for the "best member" of this subclass of estimators has been disappointing, as there are no members which can satisfy the two requirements of robustness to misspecification and higher order efficiency simultaneously. Prior suggestions to look beyond ECR came with the requirement of strong assumptions on the DGP and moment condition function g.

Our approach borrows insights from Ragusa (2011) and Chaudhuri et al. (2023) and focuses on a subclass of MD estimators which have a bounded tilting function (BTE) and thus eliminate the need for strong assumptions to guarantee a solution to the asymptotic tilting problem. By further narrowing our scope to members which satisfy a higher-order derivative condition, we produce the EBTE class which is also higher-order efficient.

We observe that bounded tilting functions can be simply constructed from a combination of scaled CDFs. In particular, we propose an example, which we call the Cauchy-Logistic EBTE, and confirm its desirable properties under both correct specification and misspecification through simulations.

While the EBTE class completes the search for the "best estimators" according to the criteria considered here, it does not yet single out one best estimator. We hope, however, that our results may pave the way for future research that aims to further narrow down the class to yield additional optimality properties.

Acknowledgements

This work was supported by the US National Science Foundation under grants SES-1950969 and SES-2150003.

Disclosure statement

The authors report there are no competing interests to declare.

References

Berge, C. (1963). *Topological Spaces*. Macmillan, New York.

Chaudhuri, S., Renault, E., and Wahlstrom, O. (2023). Identification of beliefs in the presence of disaster risk and misspecification. In Chang, Y., Lee, S., and Miller, J., editors, *Essays in Honor of Joon Y. Park: Econometric Methodology in Empirical Applications*, volume 5B of *Advances in Econometrics*, pages 261–290. Emerald Publishing Limited, Bingley.

Corcoran, S. A. (1998). Bartlett adjustment of empirical discrepancy statistics. *Biometrika*, 85:967–972.

Csiszár, I. (1995). Generalized projections for non-negative functions. *Acta Mathematica Hungarica*, 68:161–186.

Hall, P. and Horowitz, J. L. (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica*, 64:891–916.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50:1029–1054.

Imbens, G. W., Spady, R. H., and Johnson, P. (1998). Information theoretic approaches to inference in moment condition models. *Econometrica*, 66:333–357.

Kitamura, Y. (2007). Empirical likelihood methods in econometrics: Theory and practice. In Blundell, R., Newey, W. K., and Persson, T., editors, *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, volume III, chapter 7. Cambridge University Press.

Kitamura, Y. and Stutzer, M. (1997). An information-theoretic alternative to generalized method of moment estimation. *Econometrica*, 65:861–874.

Newey, W. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In Engel, R. F. and McFadden, D. L., editors, *Handbook of Econometrics*, volume IV. Elsevier Science.

Newey, W. K. and Smith, R. J. (2004). Higher order properties of gmm and generalized empirical likelihood estimators. *Econometrica*, 72:219–255.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Annals of Statistics*, 18:90–120.

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

Ragusa, G. (2011). Minimum divergence, generalized empirical likelihoods, and higher order expansions. *Econometric Reviews*, 30:406–456.

Schennach, S. M. (2007). Point estimation with exponentially tilted empirical likelihood. *The Annals of Statistics*, 35:634–672.

Sueishi, N. (2013). Identification problem of the exponential tilting estimator under misspecification. *Economics Letters*, 118:509–511.