Efficient GMM Estimation Based on Distribution Function: Property and Application

Guoning Yang
Indiana University
April 2, 2008

Abstract

This paper introduces an Efficient Generalized Method of Moment estimator based on distribution function (DF-based EGMM) and studies its asymptotic distribution as well as finite-sample performance. The estimator is consistent and asymptotically Normal. It also has merit of simplicity; the optimal weighting matrix is symmetric and tridiagonal, and is also free of data and parameters under estimation. The method is then applied to estimating the consumer search model, which provides an interesting case because Maximum Likelihood Estimation fails. Monte Carlo study shows that DF-based EGMM performs very well with search model and is more efficient than Bayesian estimation based on theoretical quantiles with no sampling error. At last, I extend the estimator to estimating conditional distribution functions and study its finite-sample performance in a linear model.

1 Introduction

In the literature of game theory and industrial organization, mixed strategy equilibriums are usually defined by cumulative distribution function. For instance, in the search model first introduced by Stigler (1961) and its following variants firms of identical products randomly draw prices from the equilibrium distribution function to price discriminate informed and uninformed customers, and in Janeba (2000) government may randomize its tax from certain distribution function. In addition, equilibriums in the literature of auction typically stem from the heterogeneity in bidder’s valuation, which enters into the optimal bidding strategy in form of distribution function and the optimal strategy is usually a bijective map from bidder’s valuation to his bid. Therefore, distribution functions play important role in modeling various economics
issues. Unique to such models is the strong conclusion about the distribution function of the variables of interest in equilibrium, which provides rich information for estimation. However, in empirical works of such models parameters are usually estimated via likelihood function. As a result, knowledge of distribution function obtained from theoretical framework is not fully exploited and applied econometricians have to derive likelihood function first, which may be very challenging and even fail in certain circumstances.

This paper introduces an Efficient Generalized Method of Moment estimator based on distribution function (DF-based EGMM) and studies its asymptotic properties as well as finite-sample performance. Under regularity conditions, the estimator is consistent and has Normal distribution asymptotically. Not only does this estimator provide an alternative way to estimate parameters of interest by matching theoretical and empirical distribution functions, it is also very efficient in finite samples. For search model where Maximum Likelihood Estimation (MLE) fails, the new estimator stands strong and is more efficient than Bayesian estimation based on theoretical quantiles without any sampling error. And the applicability of this method can be extended to conditional distributions, where the estimator remains consistent and asymptotically Normal although the object function is discontinuous. The estimator also has the merit of simplicity; the optimal weighting matrix is symmetric and tridiagonal, and is also free of data and parameters under estimation. It implies that it is not necessary to estimate weighting matrix from data, and the total number of terms in optimization is $2n + 1$ rather than $n^2$; products of moment conditions more than one rank away do not enter into the optimization because the weights on them are zero.

The remainder of the article is organized as follows. In section 2 I briefly review the Generalized Method of Moment (GMM) and Efficient Generalized Method of Moment (EGMM). Section 3 introduces the DF-based EGMM and derives its asymptotic properties. Monte Carlo studies of the finite-sample performance of the DF-based EGMM in estimating standard Normal distribution and search model are constructed in section 4. In section 4 I extend the application of DF-based EGMM to conditional distribution and study its asymptotic properties and finite-sample performance.
2 Brief Review of Efficient Generalized Method of Moments

First introduced in Hansen (1982), GMM makes use of moment conditions to estimate parameters of interest. Suppose data set \( \{ z_1, z_2, ..., z_n \} \) is a random sample of size \( n \) from a population whose distribution is indexed by parameter vector \( \theta_0 \in \Theta \subset \mathbb{R}^K \), and \( \theta_0 \) satisfies moment conditions

\[
E[g(z_i, \theta_0)] = 0 \quad \forall \ i = 1, 2, ..., n
\]

where function \( g(z_i, \theta) \in \mathbb{R}^L \). Given an \( L \times L \) symmetric, positive semi-definite weighting matrix \( W \), GMM estimator is

\[
\hat{\theta}_{\text{GMM}} = \arg \min_{\theta \in \Theta} \left[ \sum_{i=1}^{n} g(z_i, \theta) \right]' W \left[ \sum_{i=1}^{n} g(z_i, \theta) \right].
\]

**Theorem 1** (Wooldridge 14.1) Assume that (a) \( \Theta \) is compact; (b) for each \( \theta \in \Theta \), \( g(\cdot, \theta) \) is Borel measurable on \( Z \), the collection of possible values of \( z \); (c) for each \( z \in Z \), \( g(z, \cdot) \) is continuous on \( \Theta \); (d) \( |g^j(z, \theta)| \leq b(z) \) for all \( \theta \in \Theta \) and \( j = 1, 2, \ldots, L \), where \( g^j(z, \theta) \) is the \( j \)-th element of \( g(z, \theta) \) and \( b(\cdot) \) is a nonnegative function on \( Z \) such that \( E[b(z)] < \infty \); (e) \( W \xrightarrow{p} W_0 \), an \( L \times L \) positive definite matrix; and (f) \( \theta_0 \) is the unique solution to equation \( E[g(z_i, \theta_0)] = 0 \). Then a random vector \( \hat{\theta}_{\text{GMM}} \) exists and \( \hat{\theta}_{\text{GMM}} \xrightarrow{p} \theta_0 \).\(^1\)

**Theorem 2** (Wooldridge 14.2) In addition to the assumptions in Theorem 1, assume that (a) \( \theta_0 \) is in the interior of \( \Theta \); (b) \( g(z, \cdot) \) is continuously differentiable on the interior of \( \Theta \) for all \( z \in Z \); (c) each element of \( g(z, \theta_0) \) has finite second moment; (d) each element of \( \nabla_\theta g(z, \theta) \) is bounded in absolute value by a function \( b(z) \), where \( E[b(z)] < \infty \); and (e) \( E[\nabla_\theta g(z_i, \theta_0)] \) has rank \( K \). Then

\[
\sqrt{n} \left( \hat{\theta}_{\text{GMM}} - \theta_0 \right) \xrightarrow{d} N \left( 0, A_0^{-1}B_0A_0^{-1} \right)
\]

where

\[
A_0 = E[\nabla_\theta g(z_i, \theta_0)]' W_0 E[\nabla_\theta g(z_i, \theta_0)]
\]

\[
B_0 = E[\nabla_\theta g(z_i, \theta_0)]' W_0 E\left[ g(z_i, \theta_0) g(z_i, \theta_0)' \right] W_0 E[\nabla_\theta g(z_i, \theta_0)]
\]

\(^1\)For detailed proof, refer to Chapter 14 in Wooldridge (2002).
Choosing different weighting matrix gives out different GMM estimators. However, among all the possibilities there exists one that produces the GMM estimator with the smallest asymptotic variance and this optimal weighting matrix $W_{opt}$ has the property that

$$W_{opt} \xrightarrow{p} E \left[ g \left( z_i, \theta_0 \right) g \left( z_i, \theta_0 \right)' \right]^{-1}$$

Then the Efficient Generalized Method of Moments (EGMM) estimator is

$$\hat{\theta}_{EGMM} = \arg \min_{\theta \in \Theta} \left[ \sum_{i=1}^{n} g \left( z_i, \theta \right)' W_{opt} \sum_{i=1}^{n} g \left( z_i, \theta \right) \right]$$

Newey and McFadden (1994), hereafter referred to as NM, shows that under regularity conditions, $\hat{\theta}_{EGMM} \xrightarrow{p} \theta_0$ and

$$\sqrt{n} \left( \hat{\theta}_{EGMM} - \theta_0 \right) \xrightarrow{d} N \left( 0, \left[ E \left[ \nabla_{\theta} g \left( z_i, \theta_0 \right) \right]' E \left[ g \left( z_i, \theta_0 \right) g \left( z_i, \theta_0 \right)' \right]^{-1} E \left[ \nabla_{\theta} g \left( z_i, \theta_0 \right) \right] \right]^{-1} \right)$$

Since it can be shown that

$$A_0^{-1} B_0 A_0^{-1} - \left[ E \left[ \nabla_{\theta} g \left( z_i, \theta_0 \right) \right]' E \left[ g \left( z_i, \theta_0 \right) g \left( z_i, \theta_0 \right)' \right]^{-1} E \left[ \nabla_{\theta} g \left( z_i, \theta_0 \right) \right] \right]^{-1}$$

is positive semi-definite for any $E \left[ g \left( z_i, \theta_0 \right) g \left( z_i, \theta_0 \right)' \right]$ and $W_0$, the EGMM estimator is more efficient than any other GMM estimator based on the same moment conditions.

### 3 EGMM Estimation Based on Distribution Function

Suppose $\bar{x}' = (x_1, x_2, ..., x_n)$ are realizations of i.i.d. variables $(X_1, X_2, ..., X_n)$ with continuous distribution function $F(X, \theta_0)$. Without loss of generality, assume $(x_1, x_2, ..., x_n)$ has been sorted increasingly. It is straightforward that $\forall \ i = 1, 2, ..., n$

$$E \left[ \frac{1 \left( X_i < x \right) + 1 \left( X_i \leq x \right)}{2} - F \left( x, \theta_0 \right) \right] = 0^2$$

Therefore, there are $n$ moment conditions based on observed data $\bar{x}$

$$E \left[ g_i \left( \bar{x}, \theta_0 \right) \right] = 0$$

where

$$g_i \left( \bar{x}, \theta_0 \right) = \left( \begin{array}{c}
\frac{1 \left( X_i < x_1 \right) + 1 \left( X_i \leq x_1 \right)}{2} - F \left( x_1, \theta_0 \right) \\
\frac{1 \left( X_i < x_2 \right) + 1 \left( X_i \leq x_2 \right)}{2} - F \left( x_2, \theta_0 \right) \\
\frac{1 \left( X_i < x_n \right) + 1 \left( X_i \leq x_n \right)}{2} - F \left( x_n, \theta_0 \right)
\end{array} \right)$$
Then the EGMM estimator based on these conditions is
\[
\hat{\theta}_{\text{EGMM}} = \arg \min_{\theta \in \Theta} \left( n^{-1} \sum_{i=1}^{n} g_i (\bar{x}, \theta) \right) W \left( n^{-1} \sum_{i=1}^{n} g_i (\bar{x}, \theta) \right)
\]
where \( W \) is a consistent estimate of \( E \left[ g_i (\bar{x}, \theta_0) g_i (\bar{x}, \theta_0)' \right] \). Define the empirical distribution function (EDF)
\[
\hat{F}(x) = n^{-1} \sum_{i=1}^{n} \frac{1(X_i < x) + 1(X_i \leq x)}{2}
\]
Then the DF-based EGMM estimator can be written as
\[
\hat{\theta}_{\text{EGMM}} = \arg \min_{\theta \in \Theta} \left[ \hat{F}(\bar{x}) - F(\bar{x}, \theta) \right]' W \left[ \hat{F}(\bar{x}) - F(\bar{x}, \theta) \right]
\]
where
\[
W = \begin{pmatrix}
\hat{F}(x_1) [1 - \hat{F}(x_1)] & \hat{F}(x_1) [1 - \hat{F}(x_2)] & \cdots & \hat{F}(x_1) [1 - \hat{F}(x_n)] \\
\hat{F}(x_2) [1 - \hat{F}(x_1)] & \hat{F}(x_2) [1 - \hat{F}(x_2)] & \cdots & \hat{F}(x_2) [1 - \hat{F}(x_n)] \\
\vdots & \vdots & \ddots & \vdots \\
\hat{F}(x_n) [1 - \hat{F}(x_1)] & \hat{F}(x_n) [1 - \hat{F}(x_2)] & \cdots & \hat{F}(x_n) [1 - \hat{F}(x_n)]
\end{pmatrix}^{-1}
\]
As Carrasco and Florens (2000) shows, this \( \hat{\theta}_{\text{EGMM}} \) is minimizing the norm of a continuum of moment conditions in the reproducing kernel Hilbert space associated with the covariance and asymptotically has the same efficiency as MLE. Since in application, \( \hat{\theta}_{\text{EGMM}} \) is based on a finite number of moment conditions, I derive the asymptotic properties given fixed \( x' \).

**Theorem 3** Under the regularity conditions as in Theorem 1, the DF-based EGMM estimator exists and converges in probability to \( \theta_0 \). And under the regularity conditions as in Theorem 2, the DF-based EGMM estimator
\[
\sqrt{n} \left( \hat{\theta}_{\text{EGMM}} - \theta_0 \right) \xrightarrow{d} N \left( 0, \left[ \nabla_\theta F(\bar{x}, \theta_0)' E \left[ g_i (\bar{x}, \theta_0) g_i (\bar{x}, \theta_0)' \right]^{-1} \nabla_\theta F(\bar{x}, \theta_0) \right]^{-1} \right)
\]

**Proof.** Under the regularity conditions, the Uniform Weak Law of Large Numbers implies
\[
\max_{\theta \in \Theta} \left| \hat{F}(x) - F(x, \theta) - E \left[ \frac{1(X_i < x) + 1(X_i \leq x)}{2} - F(x, \theta) \right] \right| \xrightarrow{P} 0
\]
Since \( W \xrightarrow{P} E \left[ g_i (\bar{x}, \theta_0) g_i (\bar{x}, \theta_0)' \right]^{-1} \) as Claim 6 in Appendix shows, one has

\[\text{Claim 6 in Appendix}\]
\[
\left[ \hat{F}(\bar{x}) - F(\bar{x}, \theta) \right]' W \left[ \hat{F}(\bar{x}) - F(\bar{x}, \theta) \right]
\]
converges uniformly in probability to
\[
E \left[ g_i(\bar{x}, \theta) \right]' E \left[ g_i(\bar{x}, \theta_0) g_i(\bar{x}, \theta_0)' \right]^{-1} E \left[ g_i(\bar{x}, \theta) \right]
\]
which is uniquely minimized by \( \theta_0 \) because \( E \left[ g_i(\bar{x}, \theta_0) g_i(\bar{x}, \theta_0)' \right]^{-1} \) is positive definite. Therefore, by Theorem 2.1 in NM
\[
\hat{\theta}_{EGMM} \xrightarrow{p} \theta_0
\]
Under the regularity conditions, the first-order condition of the minimization problem is
\[
\nabla_{\theta} F\left(\bar{x}, \hat{\theta}_{EGMM}\right)' W \left[ \hat{F}(\bar{x}) - F\left(\bar{x}, \hat{\theta}_{EGMM}\right) \right] = 0
\]
Taylor expansion suggests
\[
\nabla_{\theta} F\left(\bar{x}, \hat{\theta}_{EGMM}\right)' W \sqrt{n} \left[ \hat{F}(\bar{x}) - F\left(\bar{x}, \hat{\theta}_{EGMM}\right) \right] = \nabla_{\theta} F\left(\bar{x}, \theta_0\right)' W \sqrt{n} \left[ \hat{F}(\bar{x}) - F\left(\bar{x}, \theta_0\right) \right]
\]
\[
+ \nabla_{\theta}^2 F\left(\bar{x}, \theta_0\right)' W \sqrt{n} \left[ \hat{F}(\bar{x}) - F\left(\bar{x}, \theta_0\right) \right] \left( \hat{\theta}_{EGMM} - \theta_0 \right) + o_p(1)
\]
By the Central Limit Theorem,
\[
\sqrt{n} \left[ \hat{F}(\bar{x}) - F\left(\bar{x}, \theta_0\right) \right] = O_p(1)
\]
Therefore,
\[
\nabla_{\theta} F\left(\bar{x}, \hat{\theta}_{EGMM}\right)' W \sqrt{n} \left[ \hat{F}(\bar{x}) - F\left(\bar{x}, \hat{\theta}_{EGMM}\right) \right] = \nabla_{\theta} F\left(\bar{x}, \theta_0\right)' W \sqrt{n} \left[ \hat{F}(\bar{x}) - F\left(\bar{x}, \theta_0\right) \right] - \nabla_{\theta} F\left(\bar{x}, \theta_0\right)' W \nabla_{\theta} F\left(\bar{x}, \theta_0\right) \sqrt{n} \left( \hat{\theta}_{EGMM} - \theta_0 \right) + o_p(1)
\]
As \( n \to \infty \), the first-order condition suggests
\[
\nabla_{\theta} F\left(\bar{x}, \theta_0\right)' E \left[ g_i\left(\bar{x}, \theta_0\right) g_i\left(\bar{x}, \theta_0\right)' \right]^{-1} \sqrt{n} \left[ \hat{F}(\bar{x}) - F\left(\bar{x}, \theta_0\right) \right] - C_0 \sqrt{n} \left( \hat{\theta}_{EGMM} - \theta_0 \right) = 0
\]
where
\[
C_0 = \nabla_{\theta} F\left(\bar{x}, \theta_0\right)' E \left[ g_i\left(\bar{x}, \theta_0\right) g_i\left(\bar{x}, \theta_0\right)' \right]^{-1} \nabla_{\theta} F\left(\bar{x}, \theta_0\right)
\]
Since $C_0$ is positive definite under the regularity conditions, one has

$$\sqrt{n} \left( \hat{\theta}_{EGMM} - \theta_0 \right) = C_0^{-1} \nabla_\theta F(\bar{x}, \theta_0)' E \left[ g_i(\bar{x}, \theta_0) g_i(\bar{x}, \theta_0)' \right]^{-1} \sqrt{n} \left[ \hat{F}(\bar{x}) - F(\bar{x}, \theta_0) \right]$$

Because, by the Central Limit Theorem,

$$\sqrt{n} \left[ \hat{F}(\bar{x}) - F(\bar{x}, \theta_0) \right] \xrightarrow{d} N \left( 0, E \left[ g_i(\bar{x}, \theta_0) g_i(\bar{x}, \theta_0)' \right]^{-1} \nabla_\theta F(\bar{x}, \theta_0) \right)^{-1}$$

it follows

$$\sqrt{n} \left( \hat{\theta}_{EGMM} - \theta_0 \right) \xrightarrow{d} N \left( 0, \left[ \nabla_\theta F(\bar{x}, \theta_0)' E \left[ g_i(\bar{x}, \theta_0) g_i(\bar{x}, \theta_0)' \right]^{-1} \nabla_\theta F(\bar{x}, \theta_0) \right]^{-1} \right)$$

by the Continuos Mapping Theorem.

One of the most attractive features of this DF-based EGMM method is its merit of simplicity. First, derivation of density function, which may be very complicated, is not indispensable to estimating theories based on distribution function. Second, the weighting matrix $W$ is free of data and parameters under estimation. To see this, note that

$$\hat{F}(\bar{x})' = \left( \frac{1}{2n}, \frac{3}{2n}, \frac{5}{2n}, ..., \frac{2n-1}{2n} \right)$$

which is a function of only $n$. Therefore, it is not necessary to estimate weighting matrix from data. Third, as Claim 8 in Appendix shows, when the sample size is $n$ the optimal weighting matrix

$$W = n \begin{pmatrix}
3 & -1 & 0 & . & 0 & 0 & 0 \\
-1 & 2 & -1 & . & 0 & 0 & 0 \\
0 & -1 & 2 & . & 0 & 0 & 0 \\
. & . & . & . & . & . & . \\
0 & 0 & 0 & . & 2 & -1 & 0 \\
0 & 0 & 0 & . & -1 & 2 & -1 \\
0 & 0 & 0 & . & 0 & -1 & 3
\end{pmatrix}$$

which is symmetric and tridiagonal. It suggests that there are $2n - 1$ terms rather than the $n^2$ terms in the optimization problem; and products of moment conditions more than one rank away do not enter into the optimization because the weights on them are zero.
4 Monte Carlo Simulation

To investigate the finite-sample performance of the DF-based EGMM estimator, I construct two Monte Carlo experiments and briefly discuss the results in this section. The first experiment simulates the estimation of mean and standard deviation of standard Normal distribution. The results are compared with sample means and sample standard deviations from the same simulated samples. The second set of simulations apply DF-based EGMM estimation to non-sequential search model originated by Stigler (1961). Then its efficiency is compared with Bayesian estimation based on theoretical quantiles without sampling error.

4.1 Estimate Normal Distribution

The DF-based EGMM estimation is first applied to estimating center and dispersion of standard Normal distribution. Table 1 presents the results based on 2,000 simulations together with the sample means and sample standard deviations of the simulated samples. For each sample size, the first row in the table is the average of the estimates, the second row (in parenthesis) is the standard deviation, and the third row is the mean squared error (MSE).

In general, DF-based EGMM performs well. It tends to have smaller bias than sample mean and sample standard deviation, suggesting it is also unbiased because sample mean and sample standard deviation are unbiased. Its efficiency improves as sample size increases, but at a rate appears slower than that of sample mean and sample standard deviation. Sample mean and sample standard deviation almost always have smaller standard deviation and MSE. However, when $n = 5$, DF-based EGMM estimator of population mean has smaller standard deviation and MSE, suggesting the estimator tends to be more efficient than sample mean when the sample size is very small.

4.2 Estimate Search Model

In the literature of search model, sellers of homogeneous products randomize their prices because some consumers have imperfect information about sellers’ prices. In this section, I estimate the search model established in Varian (1980) using DF-based EGMM. Varian (1980) models an oligopoly of $n$ sellers of homogeneous products. The sellers have a common constant marginal
Table 1: Estimation of Normal Distribution

<table>
<thead>
<tr>
<th>$\mu = 0$</th>
<th>$\sigma = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{EGMM}$</td>
<td>$\text{Sample Mean}$</td>
</tr>
<tr>
<td>$n = 5$</td>
<td>0.00357650</td>
</tr>
<tr>
<td></td>
<td>(0.4384695)</td>
</tr>
<tr>
<td></td>
<td>0.19217216</td>
</tr>
<tr>
<td>$n = 10$</td>
<td>0.00480013</td>
</tr>
<tr>
<td></td>
<td>(0.3497322)</td>
</tr>
<tr>
<td></td>
<td>0.12227451</td>
</tr>
<tr>
<td>$n = 25$</td>
<td>0.00128415</td>
</tr>
<tr>
<td></td>
<td>(0.2481335)</td>
</tr>
<tr>
<td></td>
<td>0.06154112</td>
</tr>
<tr>
<td>$n = 50$</td>
<td>-0.00559540</td>
</tr>
<tr>
<td></td>
<td>(0.1825208)</td>
</tr>
<tr>
<td></td>
<td>0.03332850</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>-0.00120300</td>
</tr>
<tr>
<td></td>
<td>(0.1310588)</td>
</tr>
<tr>
<td></td>
<td>0.01716915</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>-0.00069831</td>
</tr>
<tr>
<td></td>
<td>(0.0933522)</td>
</tr>
<tr>
<td></td>
<td>0.00871077</td>
</tr>
</tbody>
</table>
cost $c$. Number of consumers is normalized to one. All consumers have unit demand up to a common reservation price $R$ and zero demand when price is higher than $R$. $\mu$ of the consumers have low search cost while the others have high search cost. Therefore, $\mu$ consumers seek and buy at the best price of all the sellers while the other consumers randomly buy from sellers without searching.

It can be shown that there is no pure strategy equilibrium for seller pricing in this search model. Therefore, let cumulative distribution function $F(p)$ be the mixed pricing strategy in symmetric equilibrium. Seller’s task is to balance its profit from the two types of consumers. When a seller charges reservation price $R$, its expected profit is from those consumers who do not search and happen to patronize the seller randomly. That is,

$$\pi(R) = \frac{1 - \mu}{n} (R - c)$$

And when it charges price $p$, its expected profit is

$$\pi(p) = \frac{1 - \mu}{n} (p - c) + \mu (p - c) (1 - F(p))^{n-1}$$

where $\mu (p - c) (1 - F(p))^{n-1}$ is the expected profit from those consumers that search all the prices in the market. In the mixed strategy equilibrium the expected profit is constant, that is,

$$\pi(R) = \pi(p)$$

Then it follows

$$F(p) = \begin{cases} 
0 & \text{if } p \in (0, \mu) \\
1 - \left(\frac{n-1}{\mu} \frac{R-p}{p-c}\right)^{\frac{1}{n-1}} & \text{if } p \in [\mu, R] \\
1 & \text{if } p \in (R, \infty)
\end{cases}$$

where

$$\mu = \frac{n-1}{\mu} \frac{R + c}{n+1}$$

In this simple search model, parameter $\mu$ cannot be estimated using MLE. To see this, first derive the density function

$$f(p) = \frac{1}{n-1} \left[\frac{n-1}{\mu} \frac{R-p}{p-c}\right]^{\frac{1}{n-1}} \left[\frac{n}{\mu} \frac{R-c}{(p-c)^2}\right]$$
Clearly, \( \frac{\partial f(p)}{\partial \mu} < 0 \) over interval \([0, 1]\). Therefore, MLE fails in estimating this search model.

Instead, DF-based EGMM estimator is a straightforward and convenient option because the theoretical solution to the search model is characterized by distribution function\(^4\). To the best of my knowledge, the estimate of \( \mu \) never exceeds 0.4 in the empirical literature. Therefore, I simulate \( \mu = 0.1, 0.2, 0.3, \) and 0.4. In all the simulations, I use \( n = 5, c = 70, \) and \( R = 100 \).

Standard in the estimation of search model is the pre-estimation of lower and upper bound of the pricing interval. These estimates are super-consistent and thus often used as nonstochastic in the following estimations to reduce the dimension of parameters under estimation. However, Table 2 summarizes the observed maximum prices \( \hat{R} \) and minimum prices \( \hat{p} \) in the simulated samples and suggests although using \( \hat{R} \) as \( R \) loses little efficiency, the efficiency of using minimum prices observed is very doubtful when the sample size is small. Therefore, I only use \( \hat{R} \) as nonstochastic in the following DF-based EGMM estimation\(^5\).

Let \( \tilde{p} = (p_1, p_2, ..., p_n) \) be the increasingly sorted observed prices in a sample with size \( n \). Theoretical distribution function at prices \( \tilde{p} \) is

\[
F(\tilde{p}, \mu) = \begin{cases} 
1 - \left( \frac{\mu^{-1} \hat{R} - p_1}{n \hat{p}_1 - c} \right)^{\frac{1}{n-1}} \\
1 - \left( \frac{\mu^{-1} \hat{R} - p_2}{n \hat{p}_2 - c} \right)^{\frac{1}{n-1}} \\
\vdots \\
1 - \left( \frac{\mu^{-1} \hat{R} - p_n}{n \hat{p}_n - c} \right)^{\frac{1}{n-1}} 
\end{cases}
\]

Then the DF-based EGMM estimator of \( \mu \) is

\[
\hat{\mu}_{EGMM} = \arg \min_{\mu} \left[ \tilde{F}(\tilde{p}) - F(\tilde{p}, \mu) \right]' W \left[ \tilde{F}(\tilde{p}) - F(\tilde{p}, \mu) \right]
\]

In the practical operation, optimization does not have to be carried out over the whole interval \([0, 1]\). Instead, \( \hat{\mu}_{EGMM} \) is searched over \( \left[ \left( \frac{n(\hat{p} - c)}{R - \hat{p}} \right) + 1 \right]^{-1}, 1 \). This comes from the fact that

\(^4\)Baye and Morgan (2004) estimates the value of \( \varepsilon \) in the \( \varepsilon \)-equilibrium of the search model by matching empirical distribution function with theoretical distribution function. However, the paper uses non-linear least square rather than EGMM estimator, thus fails to account for the interaction among the moment conditions.

\(^5\)Note that although this introduces extra variability into the estimation of \( \mu \), using \( R \) as given is not practical in application.
\begin{table}
\centering
\begin{tabular}{cccccc}
\toprule
 & \multicolumn{2}{c}{\(p = 89.28571\)} & \multicolumn{2}{c}{\(p = 76.92308\)} & \\
 & \(R = 100\) & \(R = 100\) & \(R = 100\) & \(R = 100\) & \\
\hline
\(n = 10\) & 91.6459556 & 99.9838079 & 79.3865849 & 99.8970553 & \\
 & (1.9727922) & (0.0636366) & (2.6459764) & (0.4982521) & \\
 & 9.46072240 & 0.00430977 & 13.0665466 & 0.25872867 & \\
\hline
\(n = 25\) & 90.335646147 & 99.999229675 & 77.8417089 & 99.9953316 & \\
 & 0.9919182 & 0.0052788 & (0.9628947) & (0.0316263) & \\
 & 2.0857756 & 0.00002845 & 1.77058164 & 0.00102151 & \\
\hline
\(n = 50\) & 89.7930997 & 99.9999400 & 77.3807790 & 99.9996593 & \\
 & (0.5059597) & (0.0003749) & (0.4647694) & (0.0020179) & \\
 & 0.51331145 & 0.00000014 & 0.42539099 & 0.00000019 & \\
\hline
\(n = 100\) & 89.5460923 & 99.9999956 & 77.1392175 & 99.9999674 & \\
 & (0.2445552) & (0.0000346) & (0.2248316) & (0.0002804) & \\
 & 0.12757630 & 0.00000000 & 0.09723938 & 0.00000008 & \\
\hline
\(n = 200\) & 89.4212937 & 99.9999997 & 77.0330391 & 99.9999983 & \\
 & (0.1356613) & (0.0000024) & (0.1139167) & (0.0000135) & \\
 & 0.03677775 & 0.00000000 & 0.02506153 & 0.00000000 & \\
\bottomrule
\end{tabular}
\caption{Super-consistency of Bounds of Pricing Interval}
\end{table}
Table 3: Estimation of Search Model

<table>
<thead>
<tr>
<th></th>
<th>$\mu = 0.1$</th>
<th>$\mu = 0.2$</th>
<th>$\mu = 0.3$</th>
<th>$\mu = 0.4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 10$</td>
<td>0.10273556</td>
<td>0.20425937</td>
<td>0.30015732</td>
<td>0.39843910</td>
</tr>
<tr>
<td></td>
<td>(0.0295777)</td>
<td>(0.0532041)</td>
<td>(0.0718805)</td>
<td>(0.0842246)</td>
</tr>
<tr>
<td></td>
<td>0.00088189</td>
<td>0.00284740</td>
<td>0.00516425</td>
<td>0.00709267</td>
</tr>
<tr>
<td>$n = 25$</td>
<td>0.10039783</td>
<td>0.20147377</td>
<td>0.30046922</td>
<td>0.39961060</td>
</tr>
<tr>
<td></td>
<td>(0.0131849)</td>
<td>(0.0232665)</td>
<td>(0.0315006)</td>
<td>(0.0364608)</td>
</tr>
<tr>
<td></td>
<td>0.00017391</td>
<td>0.00054323</td>
<td>0.00099201</td>
<td>0.00132888</td>
</tr>
<tr>
<td>$n = 50$</td>
<td>0.10051236</td>
<td>0.19991484</td>
<td>0.29990719</td>
<td>0.39934474</td>
</tr>
<tr>
<td></td>
<td>(0.0066551)</td>
<td>(0.0127082)</td>
<td>(0.0170839)</td>
<td>(0.0190835)</td>
</tr>
<tr>
<td></td>
<td>0.00004453</td>
<td>0.00016142</td>
<td>0.00029172</td>
<td>0.00036434</td>
</tr>
<tr>
<td>$n = 100$</td>
<td>0.10018908</td>
<td>0.20015410</td>
<td>0.30012315</td>
<td>0.40004806</td>
</tr>
<tr>
<td></td>
<td>(0.0032214)</td>
<td>(0.0061391)</td>
<td>(0.0082839)</td>
<td>(0.0096729)</td>
</tr>
<tr>
<td></td>
<td>0.00001041</td>
<td>0.00003769</td>
<td>0.00006860</td>
<td>0.00009352</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>0.10002588</td>
<td>0.20000132</td>
<td>0.30016054</td>
<td>0.39991390</td>
</tr>
<tr>
<td></td>
<td>(0.0017712)</td>
<td>(0.0031550)</td>
<td>(0.0040987)</td>
<td>(0.0050148)</td>
</tr>
<tr>
<td></td>
<td>0.00000314</td>
<td>0.00000995</td>
<td>0.00001682</td>
<td>0.00002514</td>
</tr>
</tbody>
</table>

$$
\pi (R) = \pi (p)
$$

implies

$$
\mu^{-1} = \frac{n (p - c)}{R - \tilde{p}} + 1
\leq \frac{n (\hat{p} - c)}{R - \hat{p}} + 1
$$

Table 3 presents the results based on 2,000 replications for each different value of $\mu$. The DF-based EGMM estimator performs very well in estimating $\mu$. Even when the sample size is as small as 10, the biases are less than 3% of the true parameters. As sample size increases, the standard deviation of the DF-based EGMM estimator tends to decrease proportionally rather than by factor $\sqrt{n}$.

For comparison purposes, I report the Bayesian estimation of the search model. The data
Table 4: Bayesian Estimation of Search Model

<table>
<thead>
<tr>
<th>n</th>
<th>μ = 0.1</th>
<th>μ = 0.2</th>
<th>μ = 0.3</th>
<th>μ = 0.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.1421</td>
<td>0.2518</td>
<td>0.3517</td>
<td>0.4462</td>
</tr>
<tr>
<td></td>
<td>(0.074500)</td>
<td>(0.087780)</td>
<td>(0.092920)</td>
<td>(0.08619)</td>
</tr>
<tr>
<td></td>
<td>0.00747720</td>
<td>0.00979680</td>
<td>0.01088321</td>
<td>0.00949992</td>
</tr>
<tr>
<td>25</td>
<td>0.1108</td>
<td>0.2163</td>
<td>0.3192</td>
<td>0.4204</td>
</tr>
<tr>
<td></td>
<td>(0.020620)</td>
<td>(0.032280)</td>
<td>(0.036730)</td>
<td>(0.03979)</td>
</tr>
<tr>
<td></td>
<td>0.00053501</td>
<td>0.00125693</td>
<td>0.00185784</td>
<td>0.00215085</td>
</tr>
<tr>
<td>50</td>
<td>0.1045</td>
<td>0.2072</td>
<td>0.3091</td>
<td>0.4103</td>
</tr>
<tr>
<td></td>
<td>(0.008818)</td>
<td>(0.013770)</td>
<td>(0.017470)</td>
<td>(0.02014)</td>
</tr>
<tr>
<td></td>
<td>0.00009422</td>
<td>0.00025701</td>
<td>0.00039549</td>
<td>0.00047105</td>
</tr>
<tr>
<td>100</td>
<td>0.1020</td>
<td>0.2032</td>
<td>0.3044</td>
<td>0.4049</td>
</tr>
<tr>
<td></td>
<td>(0.003887)</td>
<td>(0.006451)</td>
<td>(0.008752)</td>
<td>(0.009507)</td>
</tr>
<tr>
<td></td>
<td>0.00002021</td>
<td>0.00005301</td>
<td>0.00009093</td>
<td>0.00010596</td>
</tr>
<tr>
<td>200</td>
<td>0.1009</td>
<td>0.2017</td>
<td>0.3022</td>
<td>0.4024</td>
</tr>
<tr>
<td></td>
<td>(0.001858)</td>
<td>(0.003323)</td>
<td>(0.004318)</td>
<td>(0.004743)</td>
</tr>
<tr>
<td></td>
<td>0.00000450</td>
<td>0.00001536</td>
<td>0.00002266</td>
<td>0.00002792</td>
</tr>
</tbody>
</table>

used in Bayesian estimation are the theoretical quantiles of the distribution function, $F^{-1}(\frac{1}{2n})$, $F^{-1}\left(\frac{3}{2n}\right)$, $F^{-1}\left(\frac{5}{2n}\right)$, ..., $F^{-1}\left(\frac{2n-1}{2n}\right)$. Therefore, there is no sampling error with these quantiles. Furthermore, instead of $\hat{R}$ in the sample, $R$ is used as given. I use the prior of uniform distribution over interval $\left[\left(\frac{n(\bar{x} - c)}{\bar{R} - \bar{c}} + 1\right)^{-1}, 1\right]$. Two chains with initial values of 0.5 and the true parameter individually are used to monitor convergence. The Markov Chain Monte Carlo (MCMC) are iterated for 10,000 times with the initial 6,000 iterations as burn-in period$^6$. This MCMC design appears to be adequate because the two chains converge very well and there are no autocorrelations.

I summarize the results from the Bayesian estimation in Table 3. Clearly, DF-based EGMM estimation outclasses Bayesian estimation in terms of bias, standard deviation, and MSE for all the parameter values under simulation. In addition, the bias of Bayesian estimation is remarkably large when the sample size is very small.

---

$^6$All the numbers of iteration mentioned are after I "thin" to every 5th iteration to address autocorrelations.
5 Extension to Conditional Distribution Functions

In this section, I consider the DF-based EGMM estimation using conditional distribution function. Suppose \( \{ \varepsilon_i : i = 1, 2, ..., n \} \) is an independent, identically distributed sequence whose distribution conditional on variable vector \( z_i \) is \( F(r(z_i, \theta_0)) \). For example, many econometrics models assume i.i.d. additive error terms with known distribution, that is,

\[
y_i = m(x_i, \theta_0) + \varepsilon_i
\]

Then the distribution function of \( \varepsilon_i \) conditional on \( (x_i, y_i) \) is \( F(y_i - m(x_i, \theta_0)) \).\(^7\)

Intuitively, I construct a DF-based EGMM estimator

\[
\hat{\theta}_{EGMM} = \arg \min_{\theta \in \Theta} \left[ \bar{F}(\bar{\varepsilon}|z, \theta) - F(\bar{\varepsilon}|z, \theta) \right]' W \left[ \bar{F}(\bar{\varepsilon}|z, \theta) - F(\bar{\varepsilon}|z, \theta) \right]
\]

where \( \bar{\varepsilon} \) is the sorted vector of

\[
\varepsilon_i = r(z_i, \theta)
\]

and

\[
\bar{F}(\varepsilon_i|z, \theta) = n^{-1} \sum_{j=1}^{n} 1(\varepsilon_j < \varepsilon_i|z, \theta) + 1(\varepsilon_j \leq \varepsilon_i|z, \theta)
\]

The challenge of estimating conditional distribution is that the moment conditions are not smooth; \( \bar{F}(\varepsilon_i|z, \theta) \) is not continuous in \( \theta \). However, following the lines in NM, I show that the DF-based EGMM estimator using conditional distribution is consistent and asymptotically Normal under certain conditions.

**Theorem 4** If (a) \( W_0 = E[g(z, \theta_0)g(z, \theta_0)']^{-1} \) is positive semi-definite and \( W_0E[g(z, \theta)] = 0 \) only if \( \theta = \theta_0 \), where vector \( g(z, \theta) = 1(\varepsilon \leq \bar{\varepsilon}|z, \theta) + 1(\varepsilon < \bar{\varepsilon}|z, \theta) - F(\bar{\varepsilon}|z, \theta) \) (b) \( \theta_0 \in \Theta \), which is compact; and (c) \( g(z, \theta) \) is continuous at each \( \theta \in \Theta \) with probability one, then \( \hat{\theta}_{EGMM} \rightarrow \theta_0 \).

**Proof.** The result follows by Theorem 2.1 in NM. Therefore, I check if all the conditions in Theorem 2.1 hold.

By Lemma 2.3 in NM, condition (a) implies \( E[g(z, \theta)]' W_0 E[g(z, \theta)] \) has a unique minimum at \( \theta = \theta_0 \). Therefore, condition 2.1(i) holds.

\(^7\)Note that this is mostly for demonstration convenience; the method is not limited to additive error terms.

15
Condition 2.1(ii) holds by (b).

Note that for all \( \theta \in \Theta, \|g(z, \theta)\| \leq \sqrt{n}, \) where \( \|\cdot\| \) is the Euclidean norm. Then by condition (c) and Lemma 2.4 in NM, \( E[g(z, \theta)] \) is continuous and

\[
\sup_{\theta \in \Theta} \|\hat{g}(\theta) - \bar{g}(\theta)\| \xrightarrow{p} 0
\]

where

\[
\hat{g}(\theta) = \left[ \bar{F}(\bar{z}|z, \theta) - F(\bar{z}|z, \theta) \right]
\]

\[
\bar{g}(\theta) = E[g(z, \theta)]
\]

Thus \( E[g(z, \theta)]' W_0 E[g(z, \theta)] \) is continuous. So condition 2.1(iii) holds.

Because \( \Theta \) is compact, \( E[g(z, \theta)] \) is bounded on \( \Theta. \) Then by triangle and Cauchy-Schwartz inequalities, one has

\[
|\hat{g}(\theta)' W \bar{g}(\theta) - \bar{g}(\theta)' W_0 \bar{g}(\theta)|
\]

\[
\leq |\hat{g}(\theta) - \bar{g}(\theta)' W (\hat{g}(\theta) - \bar{g}(\theta))| + |\bar{g}(\theta)' (W + W_0) (\hat{g}(\theta) - \bar{g}(\theta))| + |\bar{g}(\theta)' (W - W_0) \bar{g}(\theta)|
\]

\[
\leq \|\hat{g}(\theta) - \bar{g}(\theta)\|^2 \|W\| + 2 \|\bar{g}(\theta)\| \|\hat{g}(\theta) - \bar{g}(\theta)\| \|W\| + \|\bar{g}(\theta)\|^2 \|W - W_0\|
\]

Since it has been shown

\[
\sup_{\theta \in \Theta} \|\hat{g}(\theta) - \bar{g}(\theta)\| \xrightarrow{p} 0
\]

and

\( W \xrightarrow{p} W_0 \)

it follows

\[
\sup_{\theta \in \Theta} |\hat{g}(\theta)' W \bar{g}(\theta) - \bar{g}(\theta)' W_0 \bar{g}(\theta)| \xrightarrow{p} 0
\]

and thus condition 2.1(iv) holds.  

**Theorem 5** If in addition to the conditions in Theorem 4, (a) \( E[g(z, \theta)] \) is differentiable at \( \theta_0 \) with derivative \( \nabla_0 \) such that \( \nabla_0 W_0 \nabla_0 \) is nonsingular; (b) \( \theta_0 \) is an interior point of \( \Theta; \) and (c) for any \( \delta_n \to 0, \)

\[
\sup_{\|\theta - \theta_0\| \leq \delta_n} \frac{\sqrt{n} \left\| \left[ \bar{F}(\bar{z}|z, \theta) - F(\bar{z}|z, \theta) \right] - \left[ \bar{F}(\bar{z}|z, \theta_0) - F(\bar{z}|z, \theta_0) \right] - E[g(z, \theta_0)] \right\|}{1 + \sqrt{n} \|\theta - \theta_0\|} \xrightarrow{p} 0
\]
Proof. The proof is a straightforward application of Theorem 7.2 in NM. By the definition of $\hat{\theta}_{EGMM}$,

$$\sqrt{n} \left( \hat{\theta}_{EGMM} - \theta_0 \right) \xrightarrow{d} N \left( 0, \left[ \nabla'_0 E \left[ g(z, \theta_0) g(z, \theta_0)' \right]^{\top} \nabla_0 \right]^{-1} \right)$$

So, condition (iv) in Theorem 7.2 holds.

Therefore, all the conditions of Theorem 7.2 in NM hold and the asymptotic variance is

$$\left( \nabla'_0 W_0 \nabla_0 \right)^{-1} \nabla'_0 E \left[ g(z, \theta_0) g(z, \theta_0)' \right] W_0 \nabla_0 \left( \nabla'_0 W_0 \nabla_0 \right)^{-1} = \left[ \nabla'_0 E \left[ g(z, \theta_0) g(z, \theta_0)' \right]^{\top} \nabla_0 \right]^{-1} \nabla_0$$

By Monte Carlo simulations, I test the efficiency of this method in a simple linear model

$$y = \beta_0 + \beta_1 x + \varepsilon$$

where $\varepsilon \sim N(0, \sigma)$. Then one has

$$F(\varepsilon_i | x, \theta_0) = \Phi \left( \frac{y_i - \beta_0 - \beta_1 x_i}{\sigma} \right)$$

where $\Phi$ is the distribution function of standard Normal distribution.

Two sets of parameters are studied. The Monte Carlo results are reported in Table 4. Estimates of $\beta_0$ tend to be biased upward while those of $\beta_1$ and $\sigma$ tend to be biased downward.
<table>
<thead>
<tr>
<th></th>
<th>$n = 10$</th>
<th>$n = 25$</th>
<th>$n = 50$</th>
<th>$n = 100$</th>
<th>$n = 200$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta_0 = 10$</td>
<td>$\beta_1 = -3$</td>
<td>$\sigma = 2$</td>
<td>$\beta_0 = 10$</td>
<td>$\beta_1 = 3$</td>
</tr>
<tr>
<td>$\beta_0 = 10$</td>
<td>10.1120793</td>
<td>-3.0122399</td>
<td>1.77986254</td>
<td>9.98195553</td>
<td>3.00141589</td>
</tr>
<tr>
<td>(3.1780015)</td>
<td>(0.3484298)</td>
<td>(0.5462396)</td>
<td>(2.7869425)</td>
<td>(0.2583328)</td>
<td>(0.5381361)</td>
</tr>
<tr>
<td>$\sigma = 2$</td>
<td>10.1072053</td>
<td>0.12149243</td>
<td>0.34668898</td>
<td>7.76349044</td>
<td>0.06670445</td>
</tr>
<tr>
<td>$\beta_1 = -3$</td>
<td>10.1075101</td>
<td>-3.0104619</td>
<td>1.88043028</td>
<td>10.1104905</td>
<td>2.99088165</td>
</tr>
<tr>
<td>(2.0712846)</td>
<td>(0.1870162)</td>
<td>(0.3668931)</td>
<td>(2.0269485)</td>
<td>(0.1771105)</td>
<td>(0.3628629)</td>
</tr>
<tr>
<td>$\sigma = 2$</td>
<td>4.29963302</td>
<td>0.03506703</td>
<td>0.14884013</td>
<td>4.11867395</td>
<td>0.03143560</td>
</tr>
<tr>
<td>$\beta_0 = 10$</td>
<td>10.1163472</td>
<td>-3.0115399</td>
<td>1.92662151</td>
<td>10.1179097</td>
<td>2.98871621</td>
</tr>
<tr>
<td>(1.2391693)</td>
<td>(0.1192017)</td>
<td>(0.2553046)</td>
<td>(1.2882808)</td>
<td>(0.1194916)</td>
<td>(0.2510372)</td>
</tr>
<tr>
<td>$\sigma = 2$</td>
<td>1.54830934</td>
<td>0.01433512</td>
<td>0.07053223</td>
<td>1.67274033</td>
<td>0.01439843</td>
</tr>
<tr>
<td>$\beta_1 = -3$</td>
<td>10.1485156</td>
<td>-3.0144755</td>
<td>1.95375094</td>
<td>10.1023474</td>
<td>2.98970475</td>
</tr>
<tr>
<td>(0.8588790)</td>
<td>(0.0810323)</td>
<td>(0.1772211)</td>
<td>(0.8466991)</td>
<td>(0.0832129)</td>
<td>(0.1763741)</td>
</tr>
<tr>
<td>$\sigma = 2$</td>
<td>0.75936116</td>
<td>0.00677250</td>
<td>0.03353060</td>
<td>0.72701592</td>
<td>0.00702692</td>
</tr>
<tr>
<td>$\beta_0 = 10$</td>
<td>10.1346142</td>
<td>-3.0133102</td>
<td>1.96789116</td>
<td>10.1115405</td>
<td>2.9835027</td>
</tr>
<tr>
<td>(0.5833799)</td>
<td>(0.0586362)</td>
<td>(0.1281015)</td>
<td>(0.6289025)</td>
<td>(0.0615652)</td>
<td>(0.1232583)</td>
</tr>
<tr>
<td>$\sigma = 2$</td>
<td>0.35828297</td>
<td>0.00361365</td>
<td>0.01743278</td>
<td>0.40776190</td>
<td>0.00392410</td>
</tr>
</tbody>
</table>
6 Conclusion

In this paper, I introduce an efficient Generalized Method of Moment estimator which matches theoretical and empirical distribution functions. I show that the estimator is consistent and asymptotically Normal under regularity conditions, and is more efficient than any other GMM estimator based on the distribution function. Not only does this estimator provide an alternative way to estimate parameters of interest, it is also very efficient in finite samples. In search model where MLE fails, the new estimator stands strong and is more efficient than Bayesian estimation based on theoretical quantiles without any sampling error. I also extend the method to conditional distribution functions. The estimator remains consistent and asymptotically Normal under certain conditions although the object function is discontinuous. The estimator also has the merit of simplicity; the optimal weighting matrix is symmetric and tridiagonal, and is also free of data and parameters under estimation. It implies that it is not necessary to estimate weighting matrix from data, and products of moment conditions more than one rank away do not enter into the optimization.
7 Appendix

This Appendix provides proof of the following three claims.

Claim 6 \( W \xrightarrow{p} E \left[ g_i(\bar{x}, \theta_0) g_i(\bar{x}, \theta_0)' \right]^{-1} \).

Proof. Let \( g^j_i(x, \theta_0) \) be the \( j \)-th element in vector \( g_i(\bar{x}, \theta_0) \). That is,

\[
  g^j_i(\bar{x}, \theta_0) = \frac{1(X_i < x_j) + 1(X_i \leq x_j)}{2} - F(x_j, \theta_0)
\]

Then

\[
  E \left[ g^j_i(\bar{x}, \theta_0) g^j_i(\bar{x}, \theta_0) \right] = E \left[ \frac{1(X_i < x_j) + 1(X_i \leq x_j)}{2} \right]^2 - [F(x_j, \theta_0)]^2
\]

and for \( j < k \),

\[
  E \left[ g^j_i(\bar{x}, \theta_0) g^k_i(\bar{x}, \theta_0) \right] = E \left[ \frac{1(X_i < x_j) + 1(X_i \leq x_j)}{2} \left( \frac{1(X_i < x_k) + 1(X_i \leq x_k)}{2} \right) \right] - F(x_j, \theta_0) F(x_k, \theta_0)
\]

Therefore,

\[
  E \left[ g_i(\bar{x}, \theta_0) g_i(\bar{x}, \theta_0)' \right] = \begin{pmatrix}
    F(x_1, \theta_0) [1 - F(x_1, \theta_0)] & F(x_1, \theta_0) [1 - F(x_2, \theta_0)] & \cdots & F(x_1, \theta_0) [1 - F(x_n, \theta_0)] \\
    F(x_1, \theta_0) [1 - F(x_2, \theta_0)] & F(x_2, \theta_0) [1 - F(x_2, \theta_0)] & \cdots & F(x_2, \theta_0) [1 - F(x_n, \theta_0)] \\
    \vdots & \vdots & \ddots & \vdots \\
    F(x_1, \theta_0) [1 - F(x_n, \theta_0)] & F(x_2, \theta_0) [1 - F(x_n, \theta_0)] & \cdots & F(x_n, \theta_0) [1 - F(x_n, \theta_0)]
  \end{pmatrix}
\]

Since \( \hat{F}(x) \) is a consistent estimator of \( F(x, \theta_0) \),

\[
  W \xrightarrow{p} E \left[ g_i(\bar{x}, \theta_0) g_i(\bar{x}, \theta_0)' \right]^{-1}
\]
Claim 7 (1) \[ |g^j_i(x, \theta)| \leq b(x) \text{ for all } \theta \in \Theta \text{ and } j = 1, 2, \ldots, n, \text{ where } b(\cdot) \text{ is a nonnegative function on } X, \text{ the collection of possible values of } x, \text{ such that } E[b(x)] < \infty; \]

(2) For any \( n \), \( W \) is positive definite;

(3) Each element of \( g_i(x, \theta_0) \) has finite second moment.

Proof. Note that without given \( F(x, \theta) \), the above are the only conditions that can be checked.

(1) First, for all \( j = 1, 2, \ldots, n \), by the definition of distribution function

\[ |F(x, \theta)| \leq 1 \]

Then it follows

\[
\begin{align*}
|g^j_i(x, \theta)| &\leq \frac{1(X_i < x_j) + 1(X_i \leq x_j)}{2} + |F(x_j, \theta)| \\
&\leq 2
\end{align*}
\]

Therefore, let \( b(x) = 2 \forall x \in X \). Clearly, \( b(\cdot) \) is a nonnegative function on \( X \) such that \( E[b(x)] < \infty \) and \( |g^j_i(x, \theta)| \leq b(x) \) for all \( \theta \in \Theta \) and \( j = 1, 2, \ldots, n \).

(2) For any non-zero vector

\[ a' = (a_1 \ a_2 \ a_3 \ \ldots \ a_n) \]

with weighting matrix \( W \) given in Claim 8, it can be shown that

\[
a'Wa = a_1^2 + a_n^2 + 2 \sum_{i=1}^{n} a_i^2 - 2 \sum_{i=1}^{n-1} a_ia_{i+1}
\]

\[
= 2a_1^2 + 2a_n^2 + \sum_{i=1}^{n-1} (a_i^2 - 2a_ia_{i+1} + a_{i+1}^2)
\]

\[
= 2a_1^2 + 2a_n^2 + \sum_{i=1}^{n-1} (a_i - a_{i+1})^2
\]

which is clearly positive unless vector \( a \) is zero. Therefore, \( W \) is a positive definite matrix.

(3) Since

\[
\frac{1(X_i < x_j) + 1(X_i \leq x_j)}{2} \in [0, 1]
\]
and

\[ F(x_j, \theta) \in [0, 1] \]

The second moment of \( g^j_i (x, \theta) \)

\[
E \left[ g^j_i (\bar{x}, \theta) g^j_i (\bar{x}, \theta) \right] = E \left[ \left( \frac{1(X_i < x_j) + 1(X_i \leq x_j)}{2} \right)^2 \right] - 2E \left[ \frac{1(X_i < x_j) + 1(X_i \leq x_j)}{2} \right] F(x_j, \theta) + [F(x_j, \theta)]^2
\]

is clearly finite. ★

**Claim 8** When the sample size is \( n \), the optimal weighting matrix

\[
W = n \begin{pmatrix}
3 & -1 & 0 & . & 0 & 0 & 0 \\
-1 & 2 & -1 & . & 0 & 0 & 0 \\
0 & -1 & 2 & . & 0 & 0 & 0 \\
. & . & . & . & . & . & . \\
0 & 0 & 0 & . & 2 & -1 & 0 \\
0 & 0 & 0 & . & -1 & 2 & -1 \\
0 & 0 & 0 & . & 0 & -1 & 3 \\
\end{pmatrix}
\]

**Proof.** Let

\[
V = \begin{pmatrix}
\hat{F}(x_1) \left[ 1 - \hat{F}(x_1) \right] & \hat{F}(x_1) \left[ 1 - \hat{F}(x_2) \right] & . & \hat{F}(x_1) \left[ 1 - \hat{F}(x_n) \right] \\
\hat{F}(x_2) \left[ 1 - \hat{F}(x_1) \right] & \hat{F}(x_2) \left[ 1 - \hat{F}(x_2) \right] & . & \hat{F}(x_2) \left[ 1 - \hat{F}(x_n) \right] \\
. & . & . & . \\
\hat{F}(x_n) \left[ 1 - \hat{F}(x_1) \right] & \hat{F}(x_n) \left[ 1 - \hat{F}(x_2) \right] & . & \hat{F}(x_n) \left[ 1 - \hat{F}(x_n) \right] \\
\end{pmatrix}
\]

\( r_i \) be the \( i \)-th row of matrix \( V \), and \( c_{j} \) be the \( j \)-th column of matrix \( W \). Then the \( j \)-th element in \( r_i \).

\[
r_{ij} = \begin{cases}
\frac{2j-1}{2n} - \frac{2n-(2i-1)}{2n} & \text{if } j < i \\
\frac{2i-1}{2n} - \frac{2n-(2j-1)}{2n} & \text{if } j = i \\
\frac{2j-1}{2n} - \frac{2n-(2j-1)}{2n} & \text{if } j > i
\end{cases}
\]

For \( j = 1 \), the \( i \)-th element in \( c_1 \)

\[
c_{i1} = \begin{cases}
3n & \text{if } i = 1 \\
-n & \text{if } i = 2 \\
0 & \text{otherwise}
\end{cases}
\]

For \( j = n \), the \( i \)-th element in \( c_n \)

\[
c_{in} = \begin{cases}
-n & \text{if } i = n-1 \\
3n & \text{if } i = n \\
0 & \text{otherwise}
\end{cases}
\]
For $j \neq 1$ or $n$, the $i$-th element in $c_{j}$

$$c_{ij} = \begin{cases} 
-n & \text{if } i = j - 1 \\
2n & \text{if } i = j \\
-n & \text{if } i = j + 1 \\
0 & \text{otherwise}
\end{cases}$$

Therefore,

$$r_{1} \times c_{1} = \frac{1}{2n} \frac{2n - 1}{2n} \frac{1}{3n} - \frac{1}{2n} \frac{2n - 3}{2n} n = 1$$

For $i \neq 1$,

$$r_{i} \times c_{1} = \frac{1}{2n} \frac{2n - (2i - 1)}{2n} \frac{3n}{3n} - \frac{3}{2n} \frac{2n - (2i - 1)}{2n} n = 0$$

And

$$r_{n} \times c_{1} = - \frac{2n - 3}{2n} \frac{1}{2n} + \frac{2n - 1}{2n} \frac{1}{2n} \frac{3n}{2n} n = 1$$

For $i \neq n$,

$$r_{i} \times c_{n} = - \frac{2i - 1}{2n} \frac{3}{2n} n + \frac{2i - 1}{2n} \frac{1}{2n} \frac{3n}{2n} n = 0$$

For $j \neq 1$ or $n$, and $i > j$,

$$r_{i} \times c_{j} = - \frac{2 (j - 1)}{2n} \frac{1}{2n} \frac{12n - (2i - 1)}{2n} n + \frac{2j - 1}{2n} \frac{12n - (2i - 1)}{2n} \frac{2n}{2n} - \frac{2 (j + 1)}{2n} \frac{12n - (2i - 1)}{2n} n = 0$$
For $j \neq 1$ or $n$, and $i < j$,

$$r_i \times c_j = \frac{-2i-1 2n-[2(j-1)-1]}{2n} n + \frac{-2i-1 2n-(2j-1)}{2n} 2n - \frac{-2i-1 2n-[2(j+1)-1]}{2n} n = 0$$

For $j \neq 1$ or $n$, and $i = j$,

$$r_i \times c_j = \frac{-2(i-1)-1 2n-(2i-1)}{2n} n + \frac{-2i-1 2n-(2i-1)}{2n} 2n - \frac{-2i-1 2n-[2(i+1)-1]}{2n} n = 1$$

Therefore,

$$VW = I$$

where $I$ is an $n \times n$ identity matrix. That is, $W$ is the inverse matrix of $V$. ■
References


