General Doubly Robust Identification and Estimation *

Arthur Lewbel, Jin-Young Choi, and Zhuzhu Zhou

original Nov. 2016, revised August 2018
THIS VERSION IS PRELIMINARY AND INCOMPLETE

Abstract

Consider two different parametric models. Suppose one model is correctly specified, but we don’t know which one (or both could be right). Both models include a common vector of parameters, in addition to other parameters that are separate to each. An estimator for the common parameter vector is called Doubly Robust (DR) if the estimator is consistent no matter which model is correct. We provide a general technique for constructing DR estimators, which we call General Doubly Robust (GDR) estimation. Our GDR estimator is a simple extension of the Generalized Method of Moments, with analogous root-n asymptotics. We illustrate the GDR with a variety of models, including average treatment effect estimation. Our empirical application is an instrumental variables model where either one of two candidate instruments might be invalid.

JEL codes: C51, C36, C31, Keywords: Doubly Robust Estimation, Generalized Method of Moments, Instrumental Variables, Average Treatment Effects, Parametric Models

1 Introduction

Consider two different parametric models, which we will call $G$ and $H$. One of these models is correctly specified, but we don’t know which one (or both could be right). Both models include

*Corresponding Author: Arthur Lewbel, Department of Economics, Maloney 315, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA, 02467, USA. (617)-552-3678, lewbel@bc.edu, https://www2.bc.edu/arthur-lewbel/
the same parameter vector $\alpha$. An estimator $\hat{\alpha}$ is called Doubly Robust (DR) if $\hat{\alpha}$ is consistent no matter which model is correct.

We provide a general technique for constructing doubly robust (DR) estimators, which we call General Doubly Robust (GDR) estimation. The technique can be immediately extended to triply robust and general multiply robust models. Our GDR takes the form of a straightforward extension of Hansen’s (1982) Generalized Method of Moments (GMM) estimator, and we show it has similar associated root-n asymptotics.

The term double robustness was coined by Robins, Rotnitzky, and van der Laan (2000), but is based on Scharfstein, Rotnitzky, and Robins (1999) and the augmented inverse probability weighting average treatment effect estimator introduced by Robins, Rotnitzky, and Zhao (1994). In their application $\alpha$ is a population Average Treatment Effect (ATE). To summarize their application, suppose we have data consisting of $n$ observations of a random vector $Z$. Let $\tilde{G}(Z, \beta)$ be a proposed functional form for the expectation of an outcome given a binary treatment indicator and a vector of other observed covariates. Let $G$ denote the model for $\alpha$ based on $\tilde{G}$, that is, the expectation of the difference between $\tilde{G}$ in the treatment group and the control group. Let $\tilde{H}(Z, \gamma)$ be a proposed functional form for the propensity score, that is, the probability of being given treatment as a function of covariates. Then $H$ is the model for ATE $\alpha$ based on $\tilde{H}$, i.e., expected propensity score weighted outcomes. A DR estimator $\hat{\alpha}$ is then an estimator for the ATE $\alpha$ that is consistent if either $\tilde{G}$ or $\tilde{H}$ is (or both are) correctly specified. See, e.g., Słoczyński and Wooldridge (2018), Wooldridge (2007), Bang and Robins (2005), Rose and van der Laan (2014), Funk, Westreich, Wiesen, Stürmer, Brookhart, and Davidian (2011), Robins, Rotnitzky, and van der Laan (2000), and Scharfstein, Rotnitzky, and Robins (1999).

In this treatment effect example, as in most DR applications, one could consistently estimate $\alpha$ based on a nonparametric estimator of either the conditional outcome or the propensity score. That is, the functional forms of either $\tilde{G}$ or $\tilde{H}$ could be replaced with nonparametric estimators of these functions, which could then be substituted into the models $G$ or $H$ to consistently
estimate $\alpha$. The alternative approach used in DR estimation is to parameterize both $\tilde{G}$ and $\tilde{H}$. DR methods avoid the complications associated with nonparametric estimation, but still provide some insurance against misspecification, since only one of the two functions $\tilde{G}$ or $\tilde{H}$ needs to be correctly specified, and the user doesn’t need to know which one is correct. Our GDR estimator has these same benefits. Unlike nonparametric estimators, GDR requires no smoothing functions, tuning parameters, regularization, or penalty functions, and does not raise rates of convergence issues. And unlike standard parametric models, GDR provides two chances instead of just one to correctly specify a functional form.

An alternative approach to modeling if one thought that either $\tilde{G}$ or $\tilde{H}$ was correctly specified would be to engage in some form of model selection, which would then entail pretesting and the associated complications for inference. Another alternative would be model averaging, which is generally not consistent unless both $\tilde{G}$ and $\tilde{H}$ happened to be correctly specified. Like DR, our GDR avoids these issues.

The main drawback of existing DR estimators is that they are not generic, in that for each problem one needs to design a specific DR estimator, which can then only be used for that one specific application. Specifically, existing DR applications require that one find some clever way of expressing $\alpha$ as the mean of functions of $\tilde{G}(Z, \beta)$ and $\tilde{H}(Z, \gamma)$ that happens to possess the DR property. In the ATE example, this expression is given by equation (6), which has the tricky DR property of equaling the true $\alpha$ if either $\tilde{G}$ or $\tilde{H}$ is correctly specified. No general method exists for finding or constructing such equations, and only a very few example of such models are known in the literature. In contrast, we provide a general method of constructing estimators that have the DR property.

Existing DR applications express the parameter $\alpha$ as a function of $\tilde{G}(Z, \beta)$, $\tilde{H}(Z, \gamma)$, and $Z$, where $\tilde{G}$ and $\tilde{H}$ are conditional mean functions. We further generalize by assuming that the true value of $\alpha$ satisfies either $E[G(Z, \alpha, \beta)] = 0$ or $E[H(Z, \alpha, \gamma)] = 0$ for some known vector valued functions $G$ and $H$. Our GDR estimator then consistently estimates $\alpha$, despite not knowing which
of these two sets of equalities actually holds, for any functions $G$ and $H$ that satisfy some regularity and identification conditions.

Unlike existing DR estimators, we do not need to find some clever, model specific way to combine these moments. All that is needed to apply our estimator is to know the functions $G$ and $H$. For example, for estimation of the average treatment effect $\alpha$, the function $G$ is just embodies the standard expression for $\alpha$ as the difference in expected outcomes between the treatment and control groups, while the function $H$ corresponds to just the standard expression of $\alpha$ as the mean of propensity score weighted outcomes.

Note that we do not claim that our GDR estimator is superior to existing DR estimators in applications where DR estimators are known to exist. Rather, our primary contribution is providing a general method for constructing estimators that possess the double robustness property. Also, our GDR estimator has an extremely simple numerical form, and an ordinary root N consistent, asymptotically normal limiting distribution. The GDR estimator just consists of minimizing an objective function that equals the product of two GMM objective functions.

In the next section we describe our GDR estimator. Section 3 then gives four examples of potential applications. In section 4 we prove consistency of the GDR estimator and provide limiting distribution theory. Later sections provide an empirical application, and discuss extensions, including to triply and other multiply robust estimators.

## 2 The GDR Estimator

In this section we describe the GDR estimator (proof of consistency and limiting distribution theory is provided later). Let $Z$ be a vector of observed random variables, let $\alpha$, $\beta$ and $\gamma$ be vectors of parameters, and assume $G$ and $H$ are known functions. Assume a sample consisting of $n$ iid observations $z_i$ of the vector $Z$. The goal is root-n consistent estimation of $\alpha$.

Let $g_0(\alpha, \beta) \equiv E\{G(Z, \alpha, \beta)\}$, $h_0(\alpha, \gamma) \equiv E\{H(Z, \alpha, \gamma)\}$, $\theta_0 \equiv \{\alpha_0, \beta_0, \gamma_0\}$, and $\theta \equiv \{\alpha, \beta, \gamma\}$.
**Assumption A1:** For a compact set $\Theta$, $\theta_0 \in \Theta$.

**Assumption A2:** Either 1) $g_0(\alpha_0, \beta_0) = 0$, or 2) $h_0(\alpha_0, \gamma_0) = 0$, or both hold.

Assumption A2 says that either the $G$ model is true or the $H$ model is true (or both are true), for some unknown true coefficient values $\alpha_0$, $\beta_0$, and $\gamma_0$. This is a defining feature DR estimators, and hence of our GDR estimator.

**Assumption A3:** For any $\{\alpha, \beta, \gamma\} \in \Theta$, if $g_0(\alpha, \beta) = 0$ then $\{\alpha, \beta\} = \{\alpha_0, \beta_0\}$, and if $h_0(\alpha, \gamma) = 0$ then $\{\alpha, \gamma\} = \{\alpha_0, \gamma_0\}$.

Assumptions A2 and A3 are identification assumptions. They imply that if $G$ is the true model, then the true values of the coefficients $\{\alpha_0, \beta_0\}$ are identified by $g_0(\alpha_0, \beta_0) = 0$, and if $H$ is the true model, then the true values of the coefficients $\{\alpha_0, \gamma_0\}$ are identified by $h_0(\alpha_0, \gamma_0) = 0$. Assumption A3 rules out the existence of alternative pseudo-true values satisfying the ‘wrong’ moments, e.g., this assumption rules out having both $g_0(\alpha_0, \beta_0) = 0$ and $g_0(\alpha_1, \beta_1) = 0$ for some $\alpha_1 \neq \alpha_0$.

Note that Assumption A3 is a potentially strong restriction, and is not required by some existing DR estimators. As our examples later will illustrate, satisfying this assumption generally requires that parameters be over identified, which in turn typically means that the vector $G$ contains more elements than the set $\{\alpha, \beta\}$, and that the vector $H$ contains more elements than the set $\{\alpha, \gamma\}$. Otherwise, as in method of moments estimation, $g_0(\alpha, \beta) = 0$ and $h_0(\alpha, \gamma) = 0$ each have as many equations as unknowns, and so typically a pseudo-true solution will exist for whichever one is misspecified (if one is), thereby violating Assumption A3.

To define our proposed estimator, we define the following functions

$$
\tilde{g}(\alpha, \beta) \equiv \frac{1}{n} \sum_{i=1}^{n} G(Z_i, \alpha, \beta), \quad \tilde{h}(\alpha, \gamma) \equiv \frac{1}{n} \sum_{i=1}^{n} H(Z_i, \alpha, \gamma),
$$

$$
\hat{Q}^g(\alpha, \beta) \equiv \tilde{g}(\alpha, \beta)' \hat{\Omega}_g \tilde{g}(\alpha, \beta), \quad \hat{Q}^h(\alpha, \gamma) \equiv \tilde{h}(\alpha, \gamma)' \hat{\Omega}_h \tilde{h}(\alpha, \gamma),
$$
where $\hat{\Omega}_g$ and $\hat{\Omega}_h$ are positive definite matrices. Note that $\hat{Q}^g(\alpha, \beta)$ is the standard Hansen (1982) and Hansen and Singleton (1982) Generalized Method of Moments (GMM) objective function that would be used to estimate $\alpha$ and $\beta$ if $G$ were correctly specified. Similarly, $\hat{Q}^h(\alpha, \beta)$ is the GMM objective function that would be used to estimate $\alpha$ and $\gamma$ if $H$ were correctly specified. Our proposed GDR estimator is just:

$$\{\hat{\alpha}, \hat{\beta}, \hat{\gamma}\} = \arg \min_{\{\alpha, \beta, \gamma\} \in \Theta} \hat{Q}(\alpha, \beta, \gamma)$$

(1)

where $\hat{Q}(\alpha, \beta, \gamma) \equiv \hat{Q}^g(\alpha, \beta)\hat{Q}^h(\alpha, \gamma)$.

So the GDR objective function is nothing more than the GMM objective function based on the moments $g_0(\alpha, \beta) = 0$ times the GMM objective function based on the moments $h_0(\alpha, \gamma) = 0$.

Regarding the weighting matrix $\hat{\Omega}_g$ and $\hat{\Omega}_h$, note that usual GMM weight matrices are not generally optimal in our GDR. We will discuss it at the end of section 5 after discussing the asymptotic distribution of GDR estimator.

### 3 GDR Examples

Before proceeding to show consistency and deriving the limiting distribution of the GDR estimator, we consider four example applications. The first two examples show how GDR could be used in place of existing DR applications. The second two examples are new applications for which no existing DR estimator were known.

#### 3.1 Average Treatment Effect

Harking back to the earliest DR estimators like Robins, Rotnitzky, and van der Laan (2000), Scharfstein, Rotnitzky, and Robins (1999), and Robins, Rotnitzky, and Zhao (1994), here we describe the construction of DR estimates of average treatment effects, as in, e.g., Bang and Robins (2005), Funk, Westreich, Wiesen, Stürmer, Brookhart, and Davidian (2011), Rose and van der Laan
We then show how this model could alternatively be estimated using our GDR construction.

The assumption in this application is that either the conditional mean of the outcome or the propensity score of treatment is correctly parametrically specified. Let \( Z = \{Y, T, X\} \) where \( Y \) is an outcome, \( T \) is a binary treatment indicator, and \( X \) is a \( J \) vector of other covariates (including a constant). The average treatment effect we wish to estimate is

\[
\alpha = E\{E(Y|T = 1, X) - E(Y|T = 0, X)\}. \tag{2}
\]

As is well known, an alternative propensity score weighted expression for the same average treatment effect is

\[
\alpha = E \left\{ \frac{YT}{E(T|X)} - \frac{Y(1-T)}{1-E(T|X)} \right\}. \tag{3}
\]

Let \( \tilde{G}(T, X, \beta) \) be the proposed functional form of the conditional mean of the outcome, for some \( K \) vector of parameters \( \beta \). So if \( \tilde{G} \) is correctly specified, then \( \tilde{G}(T, X, \beta) = E(Y|T, X) \). Similarly, let \( \tilde{H}(X, \gamma) \) be the proposed functional form of the propensity score for some \( J \) vector of parameters \( \gamma \), so if \( \tilde{H} \) is correctly specified, then \( \tilde{H}(X, \gamma) = E(T|X) \).

One standard estimator of \( \alpha \), based on equation (2), consists of first estimating \( \beta \) by least squares, minimizing the sample average of \( E[(Y - \tilde{G}(T, X, \beta))^2] \), and then estimating \( \alpha \) as the sample average of \( \tilde{G}(1, X, \beta) - \tilde{G}(0, X, \beta) \). This estimator is equivalent to GMM estimation of \( \alpha \) and \( \beta \), using the vector of moments

\[
E \left[ \begin{array}{c}
\{Y - \tilde{G}(T, X, \beta)\}r_1(T, X) \\
\alpha - \{\tilde{G}(1, X, \beta) - \tilde{G}(0, X, \beta)\}
\end{array} \right] = 0 \tag{4}
\]

for some vector valued function \( r_1(T, X) \). Least squares estimation of \( \beta \) specifically chooses \( r_1(T, X) \) to equal \( \partial \tilde{G}(T, X, \beta) / \partial \beta \), but alternative functions could be used, corresponding to, e.g., weighted least squares estimation, or to the score functions associated with a maximum likelihood based estimator of \( \beta \), given a parameterization for the error terms \( Y - \tilde{G}(T, X, \beta) \). Note that to identify the \( K \) vector \( \beta \), the function \( r_1(T, X) \) needs to be a \( \tilde{K} \) vector for some \( \tilde{K} \geq K \). The
problem with this estimator is that in general $\alpha$ will not be consistently estimated if the functional form of $\tilde{G}(T, X, \beta)$ is not the correct specification of $E(Y|T, X)$.

An alternative common estimator of $\alpha$, based on equation (3), consists of first estimating $\gamma$ by least squares, minimizing the sample average of $E[(T - \bar{H}(X, \gamma))^2]$, and then estimating $\alpha$ as the sample average of $\frac{YT}{\bar{H}(X, \gamma)} - \frac{Y(1-T)}{1-H(X, \gamma)}$. This estimator is equivalent to GMM estimation of $\alpha$ and $\gamma$, using the vector of moments

$$
E \left[ \begin{array}{c} \{T - \bar{H}(X, \gamma)\} r_2(X) \\ \alpha - \left\{ \frac{YT}{\bar{H}(X, \gamma)} - \frac{Y(1-T)}{1-H(X, \gamma)} \right\} \end{array} \right] = 0
$$

(5)

for some $\tilde{J}$ vector valued function $r_2(X)$. As above, least squares estimation of $\gamma$ sets $r_2(X)$ equal to $\partial \bar{H}(X, \gamma) / \partial \gamma$, but as above alternative functions could be chosen for $r_2(X)$. To identify the $J$ vector $\gamma$, the function $r_2(X)$ needs to be a $\tilde{J}$ vector for some $\tilde{J} \geq J$. With this estimator, in general $\alpha$ will not be consistently estimated if the functional form of $\bar{H}(X, \gamma)$ is not the correct specification of $E(T|X)$.

A doubly robust estimator like that of Bang and Robins (2005) and later authors assumes $\alpha$ can be expressed as

$$
\alpha = E \left\{ \frac{YT}{\bar{H}(X, \gamma)} - \frac{Y(1-T)}{1-H(X, \gamma)} + \frac{T - \bar{H}(X, \gamma)}{\bar{H}(X, \gamma)} \tilde{G}(1, X, \beta) - \frac{T - \bar{H}(X, \gamma)}{1-H(X, \gamma)} \tilde{G}(0, X, \beta) \right\}.
$$

(6)

Observe that if $\bar{H}(X, \gamma) = E(T|X)$, then the first two terms in the above expectation equal equation (3) and the second two terms have mean zero. By rearranging terms, equation (6) can be rewritten as

$$
\alpha = E \left[ \tilde{G}(1, X, \beta) - \tilde{G}(0, X, \beta) + \frac{T}{\bar{H}(X, \gamma)} \{Y - \bar{G}(1, X, \beta)\} - \frac{1-T}{1-H(X, \gamma)} \{Y - \bar{G}(0, X, \beta)\} \right].
$$

(7)

Rewriting the equation this way, it can be seen that if $\tilde{G}(T, X, \beta) = E(Y|T, X)$, then the first two terms in equation (7) equal equation (2), and the second two terms have mean zero. This shows that equation (6) or equivalently (7) is doubly robust, in that it equals the average treatment effect $\alpha$ if either $\tilde{G}(T, X, \beta)$ or $\bar{H}(X, \gamma)$ is correctly specified. The GMM estimator associated with this
doubly robust estimator estimates $\alpha$, $\beta$, and $\gamma$, using the moments

$$
E \left[ \begin{array}{c}
\{Y - \tilde{G}(T, X, \beta)\}r_1(T, X) \\
\{T - \tilde{H}(X, \gamma)\}r_2(X) \\
\alpha - \left\{ \frac{YT}{H(X, \gamma)} - \frac{Y(1-T)}{1-H(X, \gamma)} + \frac{T-\tilde{H}(X, \gamma)}{H(X, \gamma)} \tilde{G}(1, X, \beta) - \frac{T-\tilde{H}(X, \gamma)}{1-H(X, \gamma)} \tilde{G}(0, X, \beta) \right\}
\end{array} \right] = 0. \quad (8)
$$

Construction of this doubly robust estimator required finding an expression like equation (6) that is special to the problem at hand. In general, finding such expressions for any particular problem may be difficult or impossible.

In contrast, our proposed GDR estimator does not require any such creativity. All that is required for constructing the GDR for this problem is to know the two alternative standard estimators, based on equations (2) and (3), expressed in GMM form, i.e., equation (4) and equation (5). Just define $G(Z, \alpha, \beta)$ to be the vector of functions given in equation (4) and define $H(Z, \alpha, \gamma)$ to be the vector of functions given in equation (5). That is,

$$
G(Z, \alpha, \beta) = \left[ \begin{array}{c}
\{Y - \tilde{G}(T, X, \beta)\}r_1(T, X) \\
\alpha - \{\tilde{G}(1, X, \beta) - \tilde{G}(0, X, \beta)\}
\end{array} \right] \quad (9)
$$

and

$$
H(Z, \alpha, \gamma) = \left[ \begin{array}{c}
\{T - \tilde{H}(X, \gamma)\}r_2(X) \\
\alpha - \left\{ \frac{YT}{H(X, \gamma)} - \frac{Y(1-T)}{1-H(X, \gamma)} \right\}
\end{array} \right]. \quad (10)
$$

These functions can then be plugged into the expressions in the previous section to obtain our GDR estimator, equation (1), without having to find an expression like equation (6) with its difficult to satisfy properties.

The vector $r_2(X)$ can include any functions of $X$ as long as the corresponding moments $E\{H(Z, \alpha, \gamma)\}$ exist. To satisfy Assumption A3, we will want to choose $r_2(X)$ to include $\tilde{J}$ elements where $\tilde{J}$ is strictly greater than $J$. What we require is that, if the propensity score is incorrectly specified, then there is no $\alpha, \gamma$ (in the set of permitted values) that satisfies the moments $E\{H(Z, \alpha, \gamma)\} = 0$, while, if the propensity score is correctly specified, then the only $\alpha, \gamma$ that satisfies $E\{H(Z, \alpha, \gamma)\} = 0$ is $\alpha_0, \gamma_0$. By the same logic, we will want to choose the $\tilde{K}$ vector
$r_1(T, X)$ to include strictly more than $K$ elements. For efficiency, it could be sensible to let $r_2(X)$ and $r_1(T, X)$ include $\partial \tilde{H}(X, \gamma) / \partial \gamma$ and $\partial \tilde{G}(T, X, \beta) / \partial \beta$, respectively.

### 3.2 An Instrumental Variables Additive Regression Model

Okui, Small, Tan, and Rubins (2012) propose a DR estimator for an instrumental variables (IV) additive regression model. The model is the additive regression

$$Y = M(W, \alpha) + \tilde{G}(X) + U,$$

$$E(Q | X) = \tilde{H}(X),$$

$$E(U | X, Q) = 0,$$

where $Y$ is an observed outcome variable, $W$ is a $S$ vector of observed exogenous covariates, $X$ is a $J$ vector of observed confounders, and $Q$ is a $K \geq J$ vector of observed instruments. Note that this model has features that are unusual for instrumental variables estimation, in particular, the assumption that $E(U | X, Q) = 0$ is stronger than the usual $E(U | Q) = 0$ assumption. The function $M(W, \alpha)$ is assumed to be correctly parameterized, and the goal is estimation of $\alpha$.

Okui, Small, Tan, and Rubins (2012) construct a DR estimator assuming that, in addition to the above, either $\tilde{G}(X) = \tilde{G}(X, \beta)$ is correctly parameterized, or that $\tilde{H}(X) = \tilde{H}(X, \gamma)$ is correctly parameterized. Let $Z = \{Y, W, X, Q\}$, and let $r_1(X)$ and $r_2(X)$ be vectors of functions chosen by the user. Define $G(\alpha, \beta, Z)$ and $H(\alpha, \gamma, Z)$ by

$$G(Z, \alpha, \beta) = \begin{bmatrix} \{Y - M(W, \alpha) - \tilde{G}(X, \beta)\}r_1(X) \\ \{Y - M(W, \alpha) - \tilde{G}(X, \beta)\}Q \end{bmatrix},$$

and

$$H(Z, \alpha, \gamma) = \begin{bmatrix} \{Q - \tilde{H}(X, \gamma)\}r_2(X) \\ \{Y - M(W, \alpha)\}\{Q - \tilde{H}(X, \gamma)\} \end{bmatrix}.$$

Okui, Small, Tan, and Rubins (2012) take $r_1(X) = \partial \tilde{G}(X, \beta) / \partial \beta$ and $r_2(X) = \partial \tilde{H}(X, \gamma) / \partial \gamma$. If $\tilde{G}(X, \beta)$ is correctly specified, then $E\{G(Z, \alpha, \beta)\} = 0$, while if $\tilde{H}(X, \gamma)$ is correctly specified then $E\{H(Z, \alpha, \gamma)\} = 0$. 

10
To get their doubly robust estimator, Okui, Small, Tan, and Rubins (2012) first specify 
\( \tilde{G}(X_i, \beta) \) and \( \tilde{H}(X_i, \gamma) \), then estimate \( \hat{\gamma} \) by the moment:

\[
E(Q_i | X_i) = \tilde{H}(X_i, \gamma)
\]

and then estimate \( \alpha \) and \( \beta \) by minimizing a quadratic form of \( \hat{B}(\alpha, \beta; \hat{\gamma}) \), where

\[
\hat{B}(\alpha, \beta; \hat{\gamma}) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \left\{ Y_i - M(W_i, \alpha) - \tilde{G}(X_i, \beta) \right\} \left\{ Q_i - \hat{H}(X_i, \hat{\gamma}) \right\} \right. \\
\left. \quad \left\{ Y_i - M(W_i, \alpha) - \tilde{G}(X_i, \beta) \right\} r_1(X_i) \right\}
\]

In place of the Okui, Small, Tan, and Rubins (2012) DR construction, we could estimate this model using the GDR estimator, equation (1), with \( G \) and \( H \) given by equations (13) and (14). To satisfy Assumption A3, \( r_1(X) \) needs to include more than \( S + J - K \) elements, and \( r_2(X) \) needs to include more than \( J \) elements. So, e.g., we would want to include at least one more function of \( X \) into \( r_1(X) \) and \( r_2(X) \), in addition to the functions \( \partial \tilde{G}(X, \beta) / \partial \beta \) and \( \partial \tilde{H}(X, \gamma) / \partial \gamma \) used by Okui, Small, Tan, and Rubins (2012).

### 3.3 Preference Parameter Estimates

One of the original applications of GMM estimation was the estimation of marginal utility parameters and of pricing kernels. See, e.g., Hansen and Singleton (1982) or Cochrane (2001). Consider a lifetime utility function of the form

\[
u_\tau = E \left\{ \sum_{t=0}^{T} b^t R_t U(C_t, X_t, \rho) \mid W_\tau \right\}
\]

where \( u_\tau \) is expected discounted lifetime utility in time period \( \tau \), \( b \) is the subjective rate of time preference, \( R_t \) is the time \( t \) gross returns from a traded asset, \( U \) is the single period utility function, \( C_t \) is observable consumption expenditures in time \( t \), \( X_t \) is a vector of other observable covariates that affect utility, \( \rho \) is a vector of utility parameters, and \( W_\tau \) is a vector of variables that are observable in time period \( \tau \). Maximization of this expected utility function under a life time budget constraining yields Euler equations of the form

\[
E \left\{ \left\{ b R_{t+1} \frac{U'(C_{t+1}, X_{t+1}, \rho)}{U'(C_t, X_t, \rho)} - 1 \right\} \mid W_\tau \right\} = 0
\]

(15)
where $U'(C_t, X_t, \rho)$ denotes $\partial U(C_t, X_t, \rho) / \partial C_t$. If the functional form of $U'$ is known, then this equation provides moments that allow $b$ and $\rho$ to be estimated using GMM. But suppose we have two different possible specifications of $U'$, and we do not know which specification is correct. Then our GDR estimator can be immediately applied, replacing the expression in the inner parentheses in equation (15) with $G(Z, \alpha, \beta)$ or $H(Z, \alpha, \gamma)$ to represent the two different specifications. Here $\alpha$ would represent parameters that are same in either specification, including the subjective rate of time preference $b$.

To give a specific example, a standard specification of utility is constant relative risk aversion with habit formation, where utility takes the form

$$U(C_t, X_t, \rho) = \frac{(C_t - M(X_t))^{1-\rho} - 1}{1 - \rho}$$

where $X_t$ is a vector of lagged values of $C_t$, the parameter $\rho$ is coefficient of risk aversion, and the function $M(X_t)$ is the habit function. See, e.g., Campbell and Cochrane (1999) or Chen and Ludvigson (2009). While this general functional form has widespread acceptance and use, there is considerable debate about the correct functional form for $M$, including whether $X_t$ should include the current value of $C_t$ or just lagged values. See, e.g., the debate about whether habits are internal or external as discussed in the above papers. Rather than take a stand on which habit model is correct, we could estimate the model by GDR.

To illustrate, suppose that with internal habits the function $M(X_t)$ would be given by $\tilde{G}(X_t, \beta)$, where $\tilde{G}$ is the internal habits functional form. Similarly, suppose with external habits $M(X_t)$ would be given by $\tilde{H}(X_t, \gamma)$ where $\tilde{H}$ is the external habits specification. Then, based on equation (15), we could define $G(Z, \alpha, \beta)$ and $H(Z, \alpha, \gamma)$ by

$$G(Z, \alpha, \beta) = \left[bR_{t+1} \left\{ \frac{C_{t+1} - \tilde{G}(X_{t+1}, \beta)}{C_t - \tilde{G}(X_t, \beta)} \right\}^{-\rho} - 1 \right] W_t$$

$$H(Z, \alpha, \gamma) = \left[bR_{t+1} \left\{ \frac{C_{t+1} - \tilde{H}(X_{t+1}, \gamma)}{C_t - \tilde{H}(X_t, \gamma)} \right\}^{-\rho} - 1 \right] W_t$$
In this example, we would have \( \alpha = (b, \rho) \), and so would consistently estimate the discount rate \( b \) and the coefficient risk aversion \( \rho \), no matter which habit model is correct. To help satisfy Assumption A3, we would generally want \( W_\tau \) to have more elements than \( (\alpha, \beta) \) and more than \((\alpha, \gamma)\).

### 3.4 Alternative Sets of Instruments

Consider a parametric model

\[
Y = M(W, \alpha) + \epsilon
\]

where \( Y \) is an outcome, \( W \) is vector of observed covariates, \( M \) is a known functional form, \( \alpha \) is a vector of parameters to be estimated, and \( \epsilon \) is an unobserved error term. Let \( R \) and \( Q \) denote two different vectors of observed covariates that are candidate instruments. One may be unsure if either \( R \) or \( Q \) are valid instrument vectors are not, where validity is defined as being uncorrelated with \( \epsilon \).

We may then define model \( G \) by \( E(\epsilon R) = 0 \), so \( G(Z, \alpha) = \{Y - M(W, \alpha)\} R \) and define model \( H \) by \( E(\epsilon Q) = 0 \), so \( H(Z, \alpha) = \{Y - M(W, \alpha)\} Q \). With these definition we can then immediately apply the GDR estimator. In this case both \( \beta \) and \( \gamma \) are empty, but more generally, the variables \( R \) and \( Q \) could themselves be functions of covariates and of parameters \( \beta \) and \( \gamma \), respectively.

To give an example, consider a model based on Lewbel (2012). Suppose \( Y = X' \alpha_x + S \alpha_s + \epsilon \), where \( X \) is a \( K \)-vector of observed exogenous covariates (including a constant term) satisfying \( E(\epsilon X) \), and \( S \) is an endogenous or mismeasured covariate that is correlated with \( \epsilon \). The goal is estimation of the set of coefficients \( \alpha = \{\alpha_x, \alpha_s\} \).

A standard instrumental variables based estimator for this model would consist of finding one or more covariates \( L \) such that \( E(\epsilon L) = 0 \). Then the set of instruments \( R \) would be defined by \( R = \{X, L\} \). The equivalent GMM estimator would be based on the moments \( E\{G(Z, \alpha)\} = \)}
where \( G(Z, \alpha) \) is given by the stacked vectors

\[
G(Z, \alpha) = \begin{cases} 
X(Y - X' \alpha_x - S \alpha_s) \\
L(Y - X' \alpha_x - S \alpha_s)
\end{cases}
\]

A special case of this estimator (corresponding to a specific choice of the GMM weighting matrix) is standard linear two stage least squares estimation. The main difficulty with applying this estimator is that one must find one or more covariates \( L \) to serve as instruments. Defining \( L \) have more than one element results in more moments than parameters, helping to satisfy Assumption A3.

To illustrate, consider Engel curve estimation (see Lewbel 2008 for a short survey, and references therein). Suppose \( Y \) is a consumer’s expenditures on food, \( X \) is a vector of covariates that affect the consumer’s tastes, and \( S \) is the consumer’s total consumption expenditures (i.e., their total budget which must be allocated between food and non-food expenditures). Suppose, as is commonly the case, that \( S \) is observed with some mismeasurement error. Then a possible and commonly employed set of instruments \( L \) consist of functions of the consumer’s income. However, validity of functions of income as instruments for total consumption depends on an assumption of separability between the consumer’s decisions on savings and their food expenditure decision, which may or may not be valid.

An alternative method of obtaining potential instruments is by exploiting functional form related assumptions. Lewbel (2012) shows that, under some conditions (including standard assumptions regarding classical measurement error), one may construct a set of potential instruments using the following procedure: Linearly regress \( S \) on \( X \), and obtain the residuals from that regression. Define a vector of instruments \( P \) to be demeaned \( X \) (excluding the constant) times these residuals. This constructed vector \( P \), along with \( X \), would then the set of instruments used to construct a GMM estimator. This estimator is implemented in the STATA module IVREG2H by Baum and Schaffer (2012).

Let \( X_\gamma \) denote the vector \( X \) with the constant removed. Algebraically, we can write the instruments obtained in this way as \( R = \{X, P\} \) where \( P = (X_\gamma - \gamma_1) (S - X' \gamma_2) \), and where the
vectors $\gamma_1$ and $\gamma_2$ in turn satisfy $E (X_- - \gamma_1) = 0$ and $E \{X (S - X'\gamma_2)\} = 0$. An efficient estimator based on this construction would be standard GMM using the moments $E \{H(Z, \alpha, \gamma)\} = 0$ where $H(Z, \alpha, \gamma)$ is a vector that consists of the stacked vectors

$$H(Z, \alpha, \gamma) = \begin{cases} 
X_- - \gamma_1 \\
X (S - X'\gamma_2) \\
X (Y - X'\alpha_x - S\alpha_s) \\
(X_- - \gamma_1)(S - X'\gamma_2)(Y - X'\alpha_x - S\alpha_s)
\end{cases}.
$$

(17)

The estimator will have more moments than parameters if $X_-$ has more than one element. As shown in Lewbel (2012), one set of conditions under which the instruments $P$ are valid (yielding consistency of this estimator) if the measurement error in $S$ is classical and if a component of $\epsilon$ is homoscedastic. So this estimator does not require finding a covariate from outside the model like income to use an instrument, but still could be inconsistent if the measurement error or homoskedasticity assumptions do not hold.

The moments given by $E \{G(Z, \alpha)\} = 0$ or $E \{H(Z, \alpha, \gamma)\} = 0$ correspond to very different sets of identifying conditions. GDR estimation based on these moments therefore allows for consistent estimation of $\alpha$ if either of these sets of conditions hold.

## 4 The GDR Estimator Asymptotics

Here we show identification and consistency of our GDR estimator.

### 4.1 GDR Identification and Consistency

To show consistency of $\hat{\alpha}$, we apply Theorem 2.1 in Newey and McFadden (1994), which provides a set of standard sufficient conditions for consistency of extremum estimators. For an objective function $Q(\theta)$, their conditions include (a) $Q(\theta)$ is uniquely maximized at $\theta_0$; (b) $\Theta$ is compact; (c) $Q(\theta)$ is continuous; (d) $\dot{Q}(\theta)$ converges uniformly in probability to $Q(\theta)$. The first of
these conditions concerns identification based on the probability limit of the objective function. The following Lemma 1 establishes this identification condition for $\alpha_0$, and for either $\beta_0$ or $\gamma_0$, depending on which moment condition is correctly specified.

Let $Q^g_0(\alpha, \beta) \equiv g_0(\alpha, \beta)'\Omega_g g_0(\alpha, \beta)$ and $Q^h_0(\alpha, \gamma) \equiv h_0(\alpha, \gamma)'\Omega_h h_0(\alpha, \gamma)$ for positive definite matrices $\Omega_g$ and $\Omega_h$. We later discuss efficient choices for $\Omega_g$ and $\Omega_h$, using two-step estimation analogous to two-step GMM.

**Lemma 1**: Suppose Assumptions A1, A2, and A3 hold. Let $\{\alpha, \beta, \gamma\} \in \Theta$ that minimizes

$$Q_0(\alpha, \beta, \gamma) \equiv Q^g_0(\alpha, \beta)Q^h_0(\alpha, \gamma).$$

Then $\alpha_1 = \alpha_0$ and either $\beta_1 = \beta_0$ or $\gamma_1 = \gamma_0$, or both.

**Proof**: If A2-1) holds, then $Q^g_0(\alpha, \beta)$ has a unique minimum at $\{\alpha_0, \beta_0\}$ by Lemma 2.3 of Newey and McFadden (1994) regarding GMM estimator. [their proof is: Let $W_g$ be such that $W'_gW_g = \Omega_g$. If $\{\alpha, \beta\} \neq \{\alpha_0, \beta_0\}$, then $\Omega_g g_0(\alpha, \beta) = W'_gW_g g_0(\alpha, \beta) \neq 0$ implies $W_g g_0(\alpha, \beta) \neq 0$. And hence $Q^g_0(\alpha, \beta) = \{W_g g_0(\alpha, \beta)\}'\{W_g g_0(\alpha, \beta)\} > Q^g_0(\alpha_0, \beta_0) = 0$ for $\{\alpha, \beta\} \neq \{\alpha_0, \beta_0\}$. Q.E.D.] Similarly, if A2-2) holds then $Q^h_0(\alpha, \gamma)$ has a unique minimum at $\{\alpha_0, \gamma_0\}$.

Now let $\{\alpha_1, \beta_1, \gamma_1\}$ be any value that minimizes $Q_0$, and consider three possible cases.

Case 1: suppose A2-1) holds, but A2-2) does not. In this case, if $\{\alpha_1, \beta_1\} \neq \{\alpha_0, \beta_0\}$, then by A3

$$Q^g_0(\alpha_1, \beta_1)Q^h_0(\alpha_1, \gamma_1) < Q^g_0(\alpha_0, \beta_0)Q^h_0(\alpha_0, \gamma_1) = 0,$$

which cannot hold because $Q^g_0$ and $Q^h_0$ must be nonnegative. Therefore in this case $\{\alpha_1, \beta_1\} = \{\alpha_0, \beta_0\}$.

Case 2: suppose A2-2) holds, but A2-1) does not. Analogously, if $\{\alpha_1, \gamma_1\} \neq \{\alpha_0, \gamma_0\}$, then by A3

$$Q^g_0(\alpha_1, \beta_1)Q^h_0(\alpha_1, \gamma_1) < Q^g_0(\alpha_0, \beta_1)Q^h_0(\alpha_0, \gamma_0) = 0,$$

which again cannot hold, so in this case $\{\alpha_1, \gamma_1\} = \{\alpha_0, \gamma_0\}$.
Case 3: Finally, suppose both A2-1) and A2-2) hold. If both \( \{\alpha_1, \beta_1\} \neq \{\alpha_0, \beta_0\} \) and \( \{\alpha_1, \gamma_1\} \neq \{\alpha_0, \gamma_0\} \) then by A3 and by \( Q_0^g \) and \( Q_0^h \) being quadratics, we again get a contradiction, so in this case either \( \{\alpha_1, \beta_1\} = \{\alpha_0, \beta_0\} \) or \( \{\alpha_1, \gamma_1\} = \{\alpha_0, \gamma_0\} \), or both. Q.E.D.

To satisfy the remaining conditions of Theorem 2.1 in Newey and McFadden (1994), continuity and uniform convergence, we make the following additional assumptions.

**Assumption A4**: \( G(Z, \alpha, \beta) \) and \( H(Z, \alpha, \gamma) \) are continuous at each \( \{\alpha, \beta, \gamma\} \in \Theta \) with probability one.

**Assumption A5**: \( E[\sup_{\{\alpha, \beta\} \in \Theta} ||G(Z, \alpha, \beta)||] < \infty \) and \( E[\sup_{\{\alpha, \gamma\} \in \Theta} ||H(Z, \alpha, \gamma)||] < \infty \).

**Theorem 1**: Suppose that \( z_i, i = 1, 2, \ldots \), are iid, \( \tilde{\Omega}_g \to^P \Omega_g, \tilde{\Omega}_h \to^P \Omega_h, \Omega_g \) and \( \Omega_h \) are positive definite, and Assumptions A1 to A5 hold. Then \( \tilde{\alpha} \to^P \alpha_0 \).

**Proof**: The proof proceeds by verifying the four conditions of Theorem 2.1 of in Newey and McFadden (1994). Conditions 1-2 of Theorem 2.1 in Newey and McFadden follow by A1-3 and Lemma 1. By A4-5 and the Uniform Law of Large Numbers (ULLN), one obtains \( \sup_{\{\alpha, \beta\} \in \Theta} ||\tilde{g}(\alpha, \beta) - g_0(\alpha, \beta)|| \to^P 0 \), \( \sup_{\{\alpha, \gamma\} \in \Theta} ||\tilde{h}(\alpha, \gamma) - h_0(\alpha, \gamma)|| \to^P 0 \), and \( g_0(\alpha, \beta) \) and \( h_0(\alpha, \gamma) \) are continuous. Thus, Condition 3 of Theorem 2.1 in Newey and McFadden holds by continuity of \( Q_0^g(\alpha, \beta) \) and \( Q_0^h(\alpha, \beta) \). By the triangle and Cauchy-Schwarz inequality, we have

\[
|\tilde{Q}(\alpha, \beta, \gamma) - Q_0(\alpha, \beta, \gamma)| = |\tilde{Q}_0^g(\alpha, \beta)\tilde{Q}_0^h(\alpha, \gamma) - Q_0^g(\alpha, \beta)Q_0^h(\alpha, \gamma)|
\]

\[
\leq |\tilde{Q}_0^g(\alpha, \beta)||\tilde{Q}_0^h(\alpha, \gamma) - Q_0^h(\alpha, \gamma)|| + |\{\tilde{Q}_0^g(\alpha, \beta) - Q_0^g(\alpha, \beta)\}||Q_0^h(\alpha, \gamma)||
\]

\[
\leq ||\tilde{Q}_0^g(\alpha, \beta) - Q_0^g(\alpha, \beta)||||\tilde{Q}_0^h(\alpha, \gamma) - Q_0^h(\alpha, \gamma)|| + ||Q_0^g(\alpha, \beta)||||\tilde{Q}_0^h(\alpha, \gamma) - Q_0^h(\alpha, \gamma)||
\]

\[
\quad + ||\tilde{Q}_0^g(\alpha, \beta) - Q_0^g(\alpha, \beta)||||Q_0^h(\alpha, \gamma)||
\]

\[
\leq ||\tilde{Q}_0^g(\alpha, \beta) - Q_0^g(\alpha, \beta)||||\tilde{Q}_0^h(\alpha, \gamma) - Q_0^h(\alpha, \gamma)||
\]

\[
\quad + ||g_0(\alpha, \beta)||^2||\Omega_g||||\tilde{Q}_0^h(\alpha, \gamma) - Q_0^h(\alpha, \gamma)||
\]

\[
\quad + ||\tilde{Q}_0^g(\alpha, \beta) - Q_0^g(\alpha, \beta)||||h_0(\alpha, \gamma)||^2||\Omega_h||.
\]
By A1, \( g_0(\alpha, \beta) \) is bounded on \( \Theta \), and following the same argument of Theorem 2.6 of Newey and McFadden, we obtain
\[
|\hat{Q}^g(\alpha, \beta) - Q_0^g(\alpha, \beta)| \leq \|\hat{g}(\alpha, \beta) - g_0(\alpha, \beta)\|^2 \|\hat{\Omega}_g\| + 2\|g_0(\alpha, \beta)\|\|\hat{g}(\alpha, \beta) - g_0(\alpha, \beta)\|\|\hat{\Omega}_g\|
+ \|g_0(\alpha, \beta)\|^2 \|\hat{\Omega}_g - \Omega_g\|,
\]
so that \( \sup_{(\alpha, \beta) \in \Theta} |\hat{Q}^g(\alpha, \beta) - Q_0^g(\alpha, \beta)| \to^P 0 \). And analogously, \( \sup_{(\alpha, \gamma) \in \Theta} |\hat{Q}^h(\alpha, \gamma) - Q_0^h(\alpha, \gamma)| \to^P 0 \) also holds. Thus Condition 4 of Theorem 2.1 in Newey and McFadden holds. Q.E.D.

4.2 Limiting Distribution

To be completed.

5 Simulation Results

To be completed.

6 Empirical Application: Engel Curve Estimation

Here we consider the example discussed in section 3.4. \( Y \) is a consumer’s expenditures on food, \( X \) is a vector of covariates that affect the consumer’s tastes, and \( S \) is the consumer’s total consumption expenditures (i.e., their total budget which must be allocated between food and non-food expenditures). The budget \( S \) is observed with some mismeasurement error. \( L \) consists of two or more functions of the consumer’s income.

To be completed.

7 Conclusions

To be completed.
REFERENCES


of Econometrics, 141(2), 1281-1301.