General Doubly Robust Identification and Estimation *

Arthur Lewbel, Jin-Young Choi, and Zhuzhu Zhou
Boston College, Xiamen University, and Boston College

Original 2018, Revised December 2019

Abstract

Consider two parametric models. At least one is correctly specified, but we don't know which. Both models include a common vector of parameters. An estimator for this common parameter vector is called Doubly Robust (DR) if it's consistent no matter which model is correct. We provide a general technique for constructing DR estimators. Our General Doubly Robust (GDR) technique is a simple extension of the Generalized Method of Moments. We illustrate our GDR with a variety of models, including average treatment effect estimation. Our empirical application is instrumental variables estimation, where either one of two instrument vectors might be invalid.

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

*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/

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 the same parameter vector α. An estimator $\widehat{\alpha}$ is called *Doubly Robust* (DR) if $\widehat{\alpha}$ 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. Our GDR takes the form of a weighted average of Hansen's (1982) Generalized Method of Moments (GMM) based estimates of α, and 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 α 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 $\widetilde{G}(Z,\beta)$ 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 α based on $\widetilde{G}$, that is, the expectation of the difference between $\widetilde{G}$ in the treatment group and the control group. Let $\widetilde{H}(Z,\gamma)$ 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 the ATE α based on $\widetilde{H}$, i.e., the expected difference between propensity score weighted outcomes. A DR estimator $\widehat{\alpha}$ is then an estimator for the ATE α that is consistent if either $\widetilde{G}$ or $\widetilde{H}$ is (or both are) correctly specified. See, e.g., Scharfstein, Rotnitzky, and Robins (1999), Rose and van der Laan (2000), Bang and Robins (2005), Wooldridge (2007), Funk, Westreich, Wiesen, Stürmer, Brookhart, and Davidian (2011), Robins, Rotnitzky, and van der Laan (2014), Słoczyński and Wooldridge (2018), and Chernozhukov, Escanciano, Ichimura, Newey, and Robins (2018).

In this treatment effect example, one could consistently estimate α based on a nonparametric

estimator of either the conditional outcome or the propensity score. That is, as in most DR applications, the functional forms of either $\widetilde{G}$ or $\widetilde{H}$ could be replaced with nonparametric estimators of these functions, which would then be substituted into the models G or H to consistently estimate α. But there are a number of potential problems associated with nonparametric estimation, e.g., it can be impractical at moderate sample sizes due to the curse of dimensionality, and it can be sensitive to the choice of regularizations like kernel and bandwidth choice.

The alternative to nonparametric estimation provided by DR estimators is to parameterize both $\widetilde{G}$ and $\widetilde{H}$. DR methods avoid the complications associated with nonparametric estimation, but still provide some insurance against misspecification, since only one of the two models G or 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, regularization, or penalty functions, and converges at the parametric root N rate. 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 G or H was correctly specified would be to engage in some form of model selection. Model selection has some disadvantages relative to doubly robust methods, e.g., one needs to correct limiting distributions for pretest bias, and tests for which model is superior can be inconclusive. In the context of GMM based models, selection methods like Andrews and Lu (2001), Caner (2009), and Liao (2013) use test-based methods or shrinkage penalties to select moments that are most likely to be valid.

Another alternative would be model averaging, which is generally not consistent unless both G and H happen to be correctly specified. Like DR, our GDR avoids these issues. However, our GDR estimator does take the form of a weighted average of GMM estimates of α, and so closely resembles GMM model averaging. A number of model averaging estimators exist for GMM and related models. Kuersteiner and Okui (2010) apply Hansen's (2007) model averaging criterion for instruments in linear instrumental variables models. Averaging across instruments or moments

in GMM models is also considered by Martins and Gabriel (2014), Sueishi (2013), and DiTraglia (2016). Unlike these papers, we do not use typical model averaging criteria like mean squared error or Bayes weights or information criteria to choose weights. Instead, we construct weights to yield the DR consistency property.

The main drawback of existing DR estimators is that they are not generic, meaning that for each problem, one needs to design a specific DR estimator, which can then only be used only for that one specific application. Existing DR applications require that one find some clever ways of expressing α as the mean of functions of both $\widetilde{G}(Z,\beta)$ and $\widetilde{H}(Z,\gamma)$ that happens to possess the DR property. In the ATE example, this expression is given by equation (6) below. No general method exists for finding or constructing such equations, and only a few examples of such models are known in the literature.

Perhaps the closest thing to a general method is Chernozhukov, Escanciano, Ichimura, Newey, and Robins (2018). These authors derive a set of locally robust estimators, provide a characterization result showing when these estimators will also be doubly robust and thereby provide some new examples of constructing doubly robust estimators.[1] In contrast, our GDR provides a simple general method of constructing estimators for a very wide class of models that have the DR property.

Existing DR applications express the parameter α as a function of $\widetilde{G}(Z,\beta)$, $\widetilde{H}(Z,\gamma)$, and Z, where $\widetilde{G}$ and $\widetilde{H}$ are conditional mean functions. We further generalize by assuming that the true value of α 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 α, 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 a clever, model specific way to combine

[1] Chernozhukov, Escanciano, Ichimura, Newey, and Robins (2018) also show that their DR estimators possess some additional useful asymptotic properties that the GDR estimators we construct may not possess. Ideally, some different terminology would distinguish between estimators that just have the DR property (including ours and theirs) vs. estimators that have the additional properties, including local robustness, that they document.

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 α, the vector valued function G just consists of moments implied by the standard expression for α as the difference in expected outcomes between treated and control groups, while the function H consists of moments implied by the standard expression of α as the mean of propensity score weighted outcomes.

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 DR property in applications where no DR estimator is known. Also, our GDR estimator has an extremely simple numerical form, and an ordinary root N consistent, asymptotically normal limiting distribution.

Consider three different possible estimators for the vector α, called $\widehat{\alpha}_g$, $\widehat{\alpha}_h$, and $\widehat{\alpha}_f$. The estimator $\widehat{\alpha}_g$ is a GMM estimator of α that is asymptotically efficient if just the model G is correctly specified, i.e., if the moments $E\left[G\left(Z,\alpha,\beta\right)\right]$ do equal zero at the true α_0 (and β_0). Specifically, $\widehat{\alpha}_g$ (along with $\widehat{\beta}_g$) minimizes the Hansen (1982) two-step quadratic GMM objective function, which we will call $\widetilde{Q}^g(\alpha,\beta)$. This $\widehat{\alpha}_g$ will generally be inconsistent if G is not correctly specified. If model G is correctly specified, then $n\widetilde{Q}^g(\widehat{\alpha}_g,\widehat{\beta}_g)$ will be asymptotically chi-squared. But more importantly for us, $\widetilde{Q}^g(\widehat{\alpha}_g,\widehat{\beta}_g)$ itself will converge to zero in probability if G is correctly specified, and not converge to zero otherwise. We use this property to construct our GDR estimator.

Analogous to $\widehat{\alpha}_g$, let $\widehat{\alpha}_h$ denote the estimator of α based on the moments $E\left[H\left(Z,\alpha,\gamma\right)\right]=0$, so $\widehat{\alpha}_h$ and $\widehat{\gamma}_h$ minimize a quadratic GMM objective function $\widetilde{Q}^h(\alpha,\gamma)$. Finally, let $\widehat{\alpha}_f$ be the GMM estimator of α based on both sets of moments $E\left[G\left(Z,\alpha,\beta\right)\right]=0$ and $E\left[H\left(Z,\alpha,\gamma\right)\right]=0$. This $\widehat{\alpha}_f$ along with $\widehat{\beta}_f$ and $\widehat{\gamma}_f$ minimizes a GMM objective function $\widetilde{Q}^f(\alpha,\beta,\gamma)$, and is asymptotically efficient (generally more efficient than either $\widetilde{Q}^g$ or $\widetilde{Q}^h$) if both models G and H are correctly specified, but will otherwise generally be inconsistent.

Our proposed GDR estimator is a weighted average of $\widehat{\alpha}_g$, $\widehat{\alpha}_h$, and $\widehat{\alpha}_f$, taking the form

$$\widehat{\alpha}=\hat{W}_f\hat{W}_g\widehat{\alpha}_h+\hat{W}_f\left(1-\hat{W}_g\right)\widehat{\alpha}_g+(1-\hat{W}_f)\widehat{\alpha}_f \tag{1}$$

where the weight $\hat{W}_f$ is a function of the GMM objective function $\widetilde{Q}^f$, and the weight $\hat{W}_g$ is function of $\widetilde{Q}^g$ and $\widetilde{Q}^h$. The expressions for these weights are given later in equations (21) and (19). These weights are not based on the typical model averaging criteria. Instead, they are constructed to make $\widehat{\alpha}$ have the same probability limit as either $\widehat{\alpha}_g$, $\widehat{\alpha}_h$, or $\widehat{\alpha}_f$, depending on whether model G, model H, or both are correctly specified. We construct weights with this property by exploiting the feature noted above that $\widetilde{Q}^g$ itself will converge to zero if G is correctly specified, and similarly for $\widetilde{Q}^h$ and $\widetilde{Q}^f$.

The conditions that will be required to apply our GDR are mostly the same as those needed for GMM estimation. However, one key additional requirement we will need is that models G and H each be overidentified, so the number of elements of the vector G is greater than the number of elements of α and β, and the number of elements of the vector H is greater than the number of elements of α and γ. More formally, Assumption A3 (given later) 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 $\alpha_1 \neq \alpha_0$, and similarly for model H. Ruling out existence of such wrong solutions generally requires overidentification.

In the next section, we give four examples of potential applications of our GDR estimator. In section 3 we describe our GDR estimator in more detail. In section 4 we show consistency and provide limiting distribution theory for our GDR. Later sections provide Monte Carlo simulations, an empirical application, and discuss extensions.

2 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.

2.1 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 (2014), Lunceford and Davidian (2004), Słoczyński and Wooldridge (2018) and Wooldridge (2007). We then show how this model could alternatively be estimated using our GDR construction. Note that other DR estimators of treatment effects also exist, e.g., Lee and Lee (2018).

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

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

and β, using the vector of moments

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

for some vector valued function $r_1\left(T,X\right)$. Least squares estimation of β specifically chooses $r_1\left(T,X\right)$ to equal $\partial\widetilde{G}\left(T,X,\beta\right)/\partial\beta$, 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 β, given a parameterization for the error terms $Y-\widetilde{G}\left(T,X,\beta\right)$. Note that to identify the K vector β, the function $r_1\left(T,X\right)$ needs to be a $\widetilde{K}$ vector for some $\widetilde{K}\geq K$. The problem with this estimator is that in general α will not be consistently estimated if the functional form of $\widetilde{G}\left(T,X,\beta\right)$ is not the correct specification of $E(Y|T,X)$.

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

$$E\left[\begin{array}{c}\{T-\widetilde{H}\left(X,\gamma\right)\}r_2\left(X\right)\\ \alpha-\left\{\frac{YT}{\widetilde{H}(X,\gamma)}-\frac{Y(1-T)}{1-\widetilde{H}(X,\gamma)}\right\}\end{array}\right]=0 \tag{5}$$

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

A doubly robust estimator like that of Bang and Robins (2005) and other authors assumes α can be expressed as

$$\alpha=E\left\{\frac{YT}{\widetilde{H}\left(X,\gamma\right)}-\frac{Y(1-T)}{1-\widetilde{H}\left(X,\gamma\right)}+\frac{T-\widetilde{H}\left(X,\gamma\right)}{\widetilde{H}\left(X,\gamma\right)}\widetilde{G}\left(1,X,\beta\right)-\frac{T-\widetilde{H}\left(X,\gamma\right)}{1-\widetilde{H}\left(X,\gamma\right)}\widetilde{G}\left(0,X,\beta\right)\right\}. \tag{6}$$

Observe that if $\widetilde{H}\left(X,\gamma\right)=E(T|X)$, 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

$$\alpha = E\left[\widetilde{G}\left(1,X,\beta\right) - \widetilde{G}\left(0,X,\beta\right) + \frac{T}{\widetilde{H}\left(X,\gamma\right)}\{Y - \widetilde{G}\left(1,X,\beta\right)\} - \frac{1-T}{1-\widetilde{H}\left(X,\gamma\right)}\{Y - \widetilde{G}\left(0,X,\beta\right)\}\right]. \tag{7}$$

Rewriting the equation this way, it can be seen that if $\widetilde{G}\left(T,X,\beta\right) = E(Y|T,X)$, 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 α if either $\widetilde{G}\left(T,X,\beta\right)$ or $\widetilde{H}\left(X,\gamma\right)$ is correctly specified. The GMM estimator associated with this doubly robust estimator estimates α, β, and γ, using the moments

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

Construction of this doubly robust estimator required finding equation (6) which is special to the problem at hand and possesses the DR property. 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 our 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,

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

and

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

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 the required overidentification (discussed earlier, and formally given later in Assumption A3), we will want to choose $r_2(X)$ to include $\widetilde{J}$ elements where $\widetilde{J}$ is strictly greater than J. What we require is that, if the propensity score is incorrectly specified, then there is no α, γ (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 α, γ that satisfies $E\{H(Z,\alpha,\gamma)\} = 0$ is α_0, γ_0. By the same logic, we will want to choose the $\widetilde{K}$ 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 $\partial\widetilde{H}(X,\gamma)/\partial\gamma$ and $\partial\widetilde{G}(T,X,\beta)/\partial\beta$, respectively.

2.2 An Instrumental Variables Additive Regression Model

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

$$Y = M(W,\alpha) + \widetilde{G}(X) + U, \tag{11}$$

$$E(Q \mid X) = \widetilde{H}(X),$$

$$E(U \mid X, Q) = 0, \tag{12}$$

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 $K \geq S$ vector of observed instruments. Note that this model has features that are unusual for instrumental variables estimation, in particular, the assumption that $E(U \mid X, Q) = 0$ is stronger than the usual $E(U \mid Q) = 0$ assumption. The function $M(W,\alpha)$ is assumed to be correctly parameterized, and the goal is estimation of α.

Okui, Small, Tan, and Robins (2012) construct a DR estimator assuming that, in addition to the above, either $\widetilde{G}(X) = \widetilde{G}(X,\beta)$ is correctly parameterized, or that $\widetilde{H}(X) = \widetilde{H}(X,\gamma)$ 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

$$G(Z, \alpha, \beta) = \begin{bmatrix} \{Y - M(W, \alpha) - \widetilde{G}(X, \beta)\} r_1(X) \\ \{Y - M(W, \alpha) - \widetilde{G}(X, \beta)\} Q \end{bmatrix} \tag{13}$$

and

$$H(Z, \alpha, \gamma) = \begin{bmatrix} \{Q - \widetilde{H}(X, \gamma)\} r_2(X) \\ \{Y - M(W, \alpha)\}\{Q - \widetilde{H}(X, \gamma)\} \end{bmatrix}. \tag{14}$$

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

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

$$E(Q|X_i) = \widetilde{H}(X_i, \gamma)$$

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

$$\hat{B}(\alpha, \beta; \hat{\gamma}) = \frac{1}{n} \sum_{i=1}^{n} \begin{bmatrix} \{Y_i - M(W_i, \alpha) - \widetilde{G}(X_i, \beta)\}\{Q_i - \widetilde{H}(X_i, \hat{\gamma})\} \\ \{Y_i - M(W_i, \alpha) - \widetilde{G}(X_i, \beta)\} r_1(X_i) \end{bmatrix}.$$

In place of the Okui, Small, Tan, and Robins (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 the required overidentification (Assumption A3), $r_1(X)$ and $r_2(X)$ need 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 $\partial \widetilde{G}(X, \beta)/\partial\beta$ and $\partial \widetilde{H}(X, \gamma)/\partial\gamma$ used by Okui, Small, Tan, and Robins (2012).

2.3 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

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

where u_τ is expected discounted lifetime utility in time period τ, 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, ρ is a vector of utility parameters, and W_τ is a vector of variables that are observable in time period τ. Maximization of this expected utility function under a lifetime budget constraining yields Euler equations of the form

$$E\left[\left\{bR_{t+1}\frac{U'\left(C_{t+1}, X_{t+1}, \rho\right)}{U'\left(C_t, X_t, \rho\right)} - 1\right\} \mid W_\tau\right] = 0 \tag{15}$$

where $U'\left(C_t, X_t, \rho\right)$ denotes $\partial U\left(C_t, X_t, \rho\right)/\partial C_t$. If the functional form of U' is known, then this equation provides moments that allow b and ρ 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 α would represent parameters that are the 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

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

where X_t is a vector of lagged values of C_t, the parameter ρ is the coefficient of risk aversion, and the function $M\left(X_t\right)$ 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 $\widetilde{G}(X_t, \beta)$, where $\widetilde{G}$ is the internal habits functional form. Similarly, suppose with external habits $M(X_t)$ would be given by $\widetilde{H}(X_t, \gamma)$ where $\widetilde{H}$ is the external habits specification. Then, based on equation (15), we could define $G(Z, \alpha, \beta)$ and $H(Z, \alpha, \gamma)$ by

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

and

$$H(Z, \alpha, \gamma) = \left[bR_{t+1} \frac{\left\{ C_{t+1} - \widetilde{H}(X_{t+1}, \gamma) \right\}^{-\rho}}{\left\{ C_t - \widetilde{H}(X_t, \gamma) \right\}^{-\rho}} - 1 \right] W_\tau.$$

In this example, we would have $\alpha = (b, \rho)$, and so would consistently estimate the discount rate b and the coefficient risk aversion ρ, no matter which habit model is correct. To help satisfy the required overidentification (Assumption A3), we would want W_τ to have more elements than (α, β) and more than (α, γ).

2.4 Alternative Sets of Instruments

Consider a parametric model

$$Y = M(W, \alpha) + \epsilon$$

where Y is an outcome, W is a vector of observed covariates, M is a known functional form, α is a vector of parameters to be estimated, and ϵ 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 or not, where validity is defined as being uncorrelated with ϵ.

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 β and γ are empty, but more generally, the variables R and Q could themselves be functions of covariates and of parameters β and γ, respectively.

A simple example that we consider in our Monte Carlo analysis is where $M(W, \alpha) = \alpha' W$, so the G model consists of the moments $E\left[(Y - \alpha' W) R\right] = 0$ and the H model is the moments $E\left[(Y - \alpha' W) Q\right] = 0$. The earlier discussed overidentification condition (later given formally as Assumption A3) is generally satisfied when Q and R each have more elements than W. Note this simple example has no β or γ parameters.

Next consider a richer example, which that includes some parameters other than α. This example, which we later empirically apply, is based on a model of 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 $E(\epsilon X) = 0$, and S is an endogenous or mismeasured covariate that is correlated with ϵ. 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 $E\{G(Z, \alpha)\} = 0$ where $G(Z, \alpha)$ is given by the stacked vectors

$$G(Z, \alpha) = \left\{ \begin{array}{c} X\left(Y - X'\alpha_x - S\alpha_s\right) \\ L\left(Y - X'\alpha_x - S\alpha_s\right) \end{array} \right\}. \tag{16}$$

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 overidentification Assumption A3.

A slightly more complicated example, which we later use for our empirical application, involves 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 measurement 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 in a food Engel curve assumes separability between the consumer's decisions on savings and their within period 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, then comprises 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_c denote the vector X with the constant removed. Algebraically, we can write the instruments obtained in this way as $R = \{X, P\}$ where $P = (X_c - \gamma_1)(S - X'\gamma_2)$, and where the vectors γ_1 and γ_2 in turn satisfy $E(X_c - \gamma_1) = 0$ 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

$$H(Z,\alpha,\gamma) = \left\{ \begin{array}{c} X_c - \gamma_1 \\ X(S - X'\gamma_2) \\ X\left(Y - X'\alpha_x - S\alpha_s\right) \\ (X_c - \gamma_1)(S - X'\gamma_2)\left(Y - X'\alpha_x - S\alpha_s\right) \end{array} \right\}. \tag{17}$$

This estimator will have more moments than parameters if X_c 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) is if the measurement error in S is classical and if a component of

ϵ 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 required measurement error assumptions do not hold.

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

3 The GDR Estimator

In this section we describe the GDR estimator in detail. Proof of its consistency and limiting distribution theory is then provided in the next section. Let Z be a vector of observed random variables, let α, β and γ be vectors of parameters, and assume G and H are known functions. Assume a sample consisting of n independent, identically distributed (iid) observations z_i of the vector Z.[2] The goal is root-n consistent, asymptotically normal estimation of α.

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 compact sets Θ_α, Θ_β, and Θ_γ, $\alpha_0 \in \Theta_\alpha$, $\beta_0 \in \Theta_\beta$, and $\gamma_0 \in \Theta_\gamma$. Let $\Theta = \Theta_\alpha \times \Theta_\beta \times \Theta_\gamma$.

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 α_0, β_0, and γ_0. This is a defining feature of DR estimators, and hence of our GDR estimator.

Assumption A3: The vector $G(Z,\alpha,\beta)$ has more elements than α,β, and vector $H(Z,\alpha,\gamma)$

[2] We assue iid data just for convenience. Our GDR is a straightforward generalization of GMM, so our GDR can generally be applied with almost any data generating process for which GMM estimation is root n consistent and asymptotically normal.

has more elements than α, γ. 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. Satisfying this assumption implies that each model's parameters are over identified. The first part of Assumption A3 is typically necessary to satisfy the second part. E.g., if G contained the same number of elements as the set $\{\alpha, \beta\}$, then the equation $g_0(\alpha, \beta) = 0$ would have as many equations as unknowns, and so typically a pseudo-true solution α_1, β_1 would exist satisfying $g_0(\alpha_1, \beta_1) = 0$ even if G were misspecified.

Define the following functions:

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

$$\widetilde{Q}^g(\alpha, \beta) \equiv \widehat{g}(\alpha, \beta)' \hat{\Omega}_g \widehat{g}(\alpha, \beta), \qquad \widetilde{Q}^h(\alpha, \gamma) \equiv \widehat{h}(\alpha, \gamma)' \hat{\Omega}_h \widehat{h}(\alpha, \gamma),$$

where $\hat{\Omega}_g$ and $\hat{\Omega}_h$ are positive definite matrices. As we can see in the above definition, $\widetilde{Q}^g(\alpha, \beta)$ is the standard Hansen (1982) and Hansen and Singleton (1982) Generalized Method of Moments (GMM) objective function, which the GMM estimator minimizes to estimate α and β, assuming G were correctly specified. Similarly, minimizing $\widetilde{Q}^h(\alpha, \gamma)$ is the standard GMM estimator for model H. Define $\widehat{\alpha}_g$, $\widehat{\beta}_g$, $\widehat{\alpha}_h$, and $\widehat{\gamma}_h$ by

$$\{\widehat{\alpha}_g, \widehat{\beta}_g\} = \arg \min_{\{\alpha, \beta\} \in \Theta_\alpha \times \Theta_\beta} \widetilde{Q}^g(\alpha, \beta) \quad \text{and} \quad \{\widehat{\alpha}_h, \widehat{\gamma}_h\} = \arg \min_{\{\alpha, \gamma\} \in \Theta_\alpha \times \Theta_\gamma} \widetilde{Q}^h(\alpha, \gamma). \tag{18}$$

So $\{\widehat{\alpha}_g, \widehat{\beta}_g\}$ is the standard GMM estimate of model G, and $\{\widehat{\alpha}_h, \widehat{\gamma}_h\}$ is the standard GMM estimate of model H.

Let $\widetilde{Q}_0^g(\alpha,\beta) \equiv g_0(\alpha,\beta)'\Omega_g g_0(\alpha,\beta)$ and $\widetilde{Q}_0^h(\alpha,\gamma) \equiv h_0(\alpha,\gamma)'\Omega_h h_0(\alpha,\gamma)$ for positive definite matrices Ω_g and Ω_h, where $\hat{\Omega}_g \to_p \Omega_g$ and $\hat{\Omega}_h \to_p \Omega_h$.

Assumption A4: Assume there exists a unique $\{\alpha_g,\beta_g\} \in \Theta_\alpha \times \Theta_\beta$ that minimizes $\widetilde{Q}_0^g(\alpha,\beta)$, and there exists a unique $\{\alpha_h,\gamma_h\} \in \Theta_\alpha \times \Theta_\gamma$ that minimizes $\widetilde{Q}_0^h(\alpha,\gamma)$.

Given assumptions A1 to A4, if the minimized value $\widetilde{Q}_0^g(\alpha_g,\beta_g) = 0$, then G is correctly specified, and $\{\alpha_g,\beta_g\}$ will equal $\{\alpha_0,\beta_0\}$. Otherwise, if the minimized value $Q_0^g(\alpha_g,\beta_g) > 0$, then G is not a correct model, and in this case we can think of $\{\alpha_g,\beta_g\}$ as unique values that are pseudo-true, in the sense that they are the values that GMM estimation of model G will converge to if model G is wrong. Assumption A4 requires that these pseudo-true values are unique. The same holds with $\{\alpha_h,\gamma_h\}$ in model H. This should not be a strong restriction, since non-uniqueness would generally imply that the separate models are not identified.

Let $c_g \equiv g_0(\alpha_g,\beta_g)$. Under minimal, standard regularity conditions (see details in the next section), we have $\widetilde{Q}^g(\widehat{\alpha}_g,\widehat{\beta}_g) \to_p c_g'\Omega_g c_g$. If G is correctly specified, then $\alpha_g = \alpha_0$ and $\beta_g = \beta_0$, which makes $c_g = 0$, so $c_g'\Omega_g c_g = 0$. What is important for our GDR estimator is that the probability limit of $\widetilde{Q}^g(\widehat{\alpha}_g,\widehat{\beta}_g)$ is zero if G is correctly specified, and positive otherwise.

Having G correctly specified also means (again with minimal regularity), that $n^{1/2}\widehat{g}(\widehat{\alpha}_g,\widehat{\beta}_g)\Omega_g^{1/2} \to_d N\left(0, I_{k_g}\right)$ so $n\widetilde{Q}^g(\widehat{\alpha}_g,\widehat{\beta}_g) \to_d \chi^2_{k_g}$. However, if G is incorrectly specified, then $c_g \neq 0$, so $c_g'\Omega_g c_g > 0$ and $n\widetilde{Q}^g(\widehat{\alpha}_g,\widehat{\beta}_g)$ diverges. Analogous statements similarly hold for model H.

Let $\hat{Q}^g(\alpha,\beta) \equiv \widetilde{Q}^g(\alpha,\beta)/k_g$ and $\hat{Q}^h(\alpha,\gamma) \equiv \widetilde{Q}^h(\alpha,\gamma)/k_h$, where the integer k_g is the degrees of freedom of the chi-squared statistic that $n\widetilde{Q}^g$ converges to if the G model is true. This is the number of moments in G minus the number of elements in α and β, which is positive as discussed earlier. Similarly, k_h is the degrees of freedom of the chi-squared statistic that $n\widetilde{Q}^h$ equals if the H model is true. This scaling by k_g and k_h is not necessary for our estimator, but improves its finite sample performance (see below for details).

Define $\hat{W}_g$ by

$$\hat{W}_g \equiv \frac{\hat{Q}^g(\widehat{\alpha}_g, \widehat{\beta}_g)}{\hat{Q}^g(\widehat{\alpha}_g, \widehat{\beta}_g) + \hat{Q}^h(\widehat{\alpha}_h, \widehat{\gamma}_h)}. \tag{19}$$

From the above derivations, we have that, if G is correctly specified and H is not,

$$\hat{W}_g \to_p \frac{0}{0 + c_h^{'}\Omega_h c_h / k_h} = 0,$$

while if H is correctly specified and and G is not,

$$\hat{W}_g \to_p \frac{c_g^{'}\Omega_g c_g / k_g}{c_g^{'}\Omega_g c_g / k_g + 0} = 1.$$

Before getting to our GDR estimator given by equation (1), consider the simpler estimator $\widetilde{\alpha}$ defined by

$$\widetilde{\alpha} = \hat{W}_g \widehat{\alpha}_h + \left(1 - \hat{W}_g\right) \widehat{\alpha}_g. \tag{20}$$

So $\widetilde{\alpha}$ is simply a weighted average of the GMM estimates $\widehat{\alpha}_g$ and $\widehat{\alpha}_h$, where the weights are proportional to $\hat{Q}^g$ and $\hat{Q}^h$. We will call $\widetilde{\alpha}$ the SGDR (simpler GDR) estimator.

The intuition behind $\widetilde{\alpha}$ is straightforward (the asymptotic statements in this paragraph are proved formally in the next section). Suppose model H is wrong and model G is right, so $E\left[H(Z, \alpha, \gamma)\right] \neq 0$ for any α and γ, and $E\left[G(Z, \alpha_0, \beta_0)\right] = 0$. Then $\hat{Q}^g(\widehat{\alpha}_g, \widehat{\beta}_g)$ goes in probability to zero while the limiting value of $\hat{Q}^h(\widehat{\alpha}_h, \widehat{\gamma}_h)$ is nonzero, so $\hat{W}_g$, the weight on $\widehat{\alpha}_h$ in equation (20) will go to zero, and $\left(1 - \hat{W}_g\right)$, the weight on $\widehat{\alpha}_g$, will go to one. As a result, $\widetilde{\alpha}$ will have the same probability limit as $\widehat{\alpha}_g$, and since model G is right, this probability limit will be α_0. The same logic applies if model H is right and G is wrong, switching the roles of g and h, and the roles of β and γ. Finally, if both models are right, then $\widetilde{\alpha}$ is just a weighted average of consistent estimators of α_0, and so is consistent no matter what values the weights take on. We therefore obtain the double robustness property that, whichever model is right, $\widetilde{\alpha} \to_p \alpha_0$.[3]

[3]Notice that when both G and H are correctly specified, $\hat{W}_g$ converges to a ratio of correlated chi-squared distributions, not to a constant. Nevertheless, $\widetilde{\alpha}$ is still consistent because $\widetilde{\alpha} = \widehat{\alpha}_g + (\widehat{\alpha}_h - \widehat{\alpha}_g)\hat{W}_g$, and when both are correctly specified, $\widehat{\alpha}_g \to_p \alpha_0$ and $\widehat{\alpha}_h - \widehat{\alpha}_g \to_p 0$.

We could have defined the weight $\hat{W}_g$ without scaling each GMM objective function by its degrees of freedom. Asymptotically, the estimator would still be doubly robust. The reason we scale is because, even when a model is correctly specified, in finite samples the greater is the degrees of freedom of a model, the larger its GMM objective function is likely to be. Asymptotically, the mean of $n\widetilde{Q}^g$ converges to k_g when g is correctly specified, and similarly for h. So, by scaling, when both models are correctly specified, both $n\hat{Q}^g$ and $n\hat{Q}^h$ will asymptotically have mean one. Otherwise, if we didn't scale, whichever model has more moments will tend to have a larger GMM objective function, which would then undesirably penalize that model in finite samples.

Since $\widetilde{\alpha}$ has the desired DR property, we could use $\widetilde{\alpha}$ itself as a GDR estimator. However, $\widetilde{\alpha}$ has two drawbacks. First, when G and H are both correct, the ratio $\hat{W}_g$ converges to a random variable rather than a constant, which complicates the limiting distribution of $\widetilde{\alpha}$. Second, when both G and H are correct, $\widetilde{\alpha}$ may be very inefficient, relative to a GMM estimator that efficiently combines the moments from both models.

To address both of these issues, consider a third model F, defined as the union of moments of the models G and H. Specifically, let $F(Z, \alpha, \beta, \gamma)$ be the vector valued function consisting of the union of elements of $G(Z, \alpha, \beta)$ and $H(Z, \alpha, \gamma)$. Then, letting $\widehat{f}(\alpha, \beta, \gamma) \equiv \frac{1}{n}\sum_{i=1}^{n} F(Z_i, \alpha, \beta, \gamma)$, we can define a third GMM estimator

$$\{\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f\} = \arg \min_{\{\alpha,\beta,\gamma\} \in \Theta_\alpha \times \Theta_\beta \times \Theta_\gamma} \widetilde{Q}^f(\alpha, \beta, \gamma)$$

where $\widetilde{Q}^f(\alpha, \beta, \gamma) \equiv \widehat{f}(\alpha, \beta, \gamma)' \hat{\Omega}_f \widehat{f}(\alpha, \beta, \gamma)$. This is GMM assuming both specifications are correct, and so uses all the moments from both. If models G and H are correctly specified, then $\widehat{\alpha}_f$ is generally more asymptotically efficient than $\widehat{\alpha}_g$, $\widehat{\alpha}_h$, and $\widetilde{\alpha}$. Let $c_f \equiv f_0(\alpha_f, \beta_f, \gamma_f) \equiv E\{F(Z, \alpha_f, \beta_f, \gamma_f)\}$. As with the other GMM estimators, we will get that $\widetilde{Q}^f(\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f) \to_p c_f' \Omega_f c_f$, which equals zero if both models G and H are correctly specified, and is positive otherwise.

We again scale by the degrees of freedom (number of moments in F minus number of elements of α, β, and γ), denoted k_f, defining $\hat{Q}^f(\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f) \equiv \widetilde{Q}^f(\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f)/k_f$. We then define the weight

$\hat{W}_f$ by

$$\hat{W}_f \equiv 1 - \frac{1}{n^\tau \hat{Q}^f(\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f) + 1} \tag{21}$$

for some τ having $0 < \tau < 1$. Later we discuss selection of the tuning parameter τ, but note for now that the only property τ will need to satisfy is lying between zero and one. Our GDR estimator, given by equation (1), can be equivalently written as

$$\widehat{\alpha} = \hat{W}_f \widetilde{\alpha} + \left(1 - \hat{W}_f\right) \widehat{\alpha}_f. \tag{22}$$

The intuition now is, if both G and H are correctly specified, then $\hat{Q}^f(\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f) \to_p 0$ and $n\hat{Q}^f(\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f)$ converges in distribution to a chi-squared statistic (divided by its degrees of freedom), which means that $n^\tau \hat{Q}^f(\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f)$ for $0 < \tau < 1$ converges in probability to zero. Alternatively, if either G or H is incorrectly specified, then $\hat{Q}^f(\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f)$ converges in probability to a positive value, so $n^\tau \hat{Q}^f(\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f)$ diverges to infinity. Therefore, if both G and H are correctly specified then $\hat{W}_f \to_p 0$ and so $\widehat{\alpha}$ has the same limiting value as $\widehat{\alpha}_f$, while if either G or H is incorrectly specified, then $\widehat{\alpha}$ has the same limiting value as $\widetilde{\alpha}$, which as shown earlier has the same limiting value as either $\widehat{\alpha}_g$ or $\widehat{\alpha}_h$, depending on which is correctly specified.

The estimator $\widehat{\alpha}$ therefore, like $\widetilde{\alpha}$, has the desired DR property. We show later that $\widehat{\alpha}$ avoids the asymptotic issues $\widetilde{\alpha}$ has when both G and H are correctly specified, and that $\widehat{\alpha}$ generally performs better than $\widetilde{\alpha}$ in finite samples. This is why $\widehat{\alpha}$ is our preferred GDR estimator. However $\widehat{\alpha}$ has the disadvantages of being a little more complicated to estimate (since it requires estimating the third model F), and it requires selection of a tuning parameter τ. In our monte carlo simulations, we find that a good choice is $\tau = 1 - p$ where p is the p-value of the Wald statistic testing the null hypothesis that $\widehat{\alpha}_g = \widehat{\alpha}_h$. This works well because it puts more weight on $\widehat{\alpha}_f$ when both G and H are likely to be correctly specified, which is when $\alpha_g = \alpha_h$.

One final remark concerns the weighting matrices $\hat{\Omega}_g$, $\hat{\Omega}_h$, and $\hat{\Omega}_f$. The usual weighting matrices obtained in two step GMM by construction yield asymptotic efficiency of the separate GMM estimators $\widehat{\alpha}_g$, $\widehat{\alpha}_h$, and $\widehat{\alpha}_f$. However, it is possible that these are not the most efficient choices for

our GDR estimator $\widehat{\alpha}$. Nevertheless, there are advantages to just using the standard GMM weight matrices, which we discuss in the next section

4 The GDR Estimator Asymptotics

In this section we show consistency of our GDR estimator $\widehat{\alpha}$, and then derive its limiting distribution, showing it is root n consistent and asymptotically normal.

4.1 GDR Consistency

To show consistency of $\widehat{\alpha}$, we apply Theorem 2.1 in Newey and McFadden (1994), which provides a set of sufficient conditions for identification and consistency of extremum estimators. To satisfy their continuity and uniform convergence conditions, we make the following two additional assumptions.

Assumption A5: $G(Z, \alpha, \beta)$, $H(Z, \alpha, \gamma)$ and $F(Z, \alpha, \beta, \gamma)$ are continuous at $\{\alpha, \beta\} \in \Theta_\alpha \times \Theta_\beta$, $\{\alpha, \gamma\} \in \Theta_\alpha \times \Theta_\gamma$, and $\{\alpha, \beta, \gamma\} \in \Theta_\alpha \times \Theta_\beta \times \Theta_\gamma$ respectively, with probability one.

Assumption A6: $E[\sup_{\{\alpha,\beta\}\in\Theta_\alpha\times\Theta_\beta} ||G(Z, \alpha, \beta)||] < \infty$, $E[\sup_{\{\alpha,\gamma\}\in\Theta_\alpha\times\Theta_\gamma} ||H(Z, \alpha, \gamma)||] < \infty$, and $E[\sup_{\{\alpha,\beta,\gamma\}\in\Theta_\alpha\times\Theta_\beta\times\Theta_\gamma} ||F(Z, \alpha, \beta, \gamma)||] < \infty$.

Together, Assumptions A1, A2, A3, A5, and A6, are standard conditions that suffice for consistency of the GMM estimates $\{\widehat{\alpha}_g, \widehat{\beta}_g\}$ in equation (18) if model G is correctly specified, and similarly yield consistency of $\{\widehat{\alpha}_h, \widehat{\gamma}_h\}$ if model H is correctly specified. If both models are correctly specified, the conditions also suffice for consistency of the GMM estimates $\{\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f\}$.

Lemma 1 : Suppose $\hat{\Omega}_g \rightarrow_p \Omega_g$, $\hat{\Omega}_h \rightarrow_p \Omega_h$, $\hat{\Omega}_f \rightarrow_p \Omega_f$, Ω_g, Ω_h and Ω_f are positive definite, and Assumptions A1 to A6 hold. Then, for any τ having $0 < \tau < 1$, $\hat{W}_f$ given by equation (21) has a finite probability limit. Specifically, if either G or H is misspecified, then $\hat{W}_f \rightarrow_p 1$, and if both G and H are correctly specified, then $\hat{W}_f \rightarrow_p 0$. Also, for $\hat{W}_g$ given by equation (19), $\hat{W}_g\hat{W}_f$

has a finite probability limit. Specifically, if G is correctly specified but H is misspecified, then $\hat{W}_g\hat{W}_f \to_p 0$; if H is correctly specified but G is misspecified, then $\hat{W}_g\hat{W}_f \to_p 1$; and if both if G and H are correctly specified, then $\hat{W}_g\hat{W}_f \to_p 0$.

Lemma 1 is proved in the Appendix, but the intuition is as follows. When either G or H is misspecified, under the regularity conditions $\hat{Q}^f \to_p c_f'\Omega_f c_f/k_f > 0$ so $n^\tau\hat{Q}^f \to_p \infty$ and $\hat{W}_f$ goes to one. If G is correct but H is false, then $\hat{Q}^g$ converges in probability to zero while the limiting value of $\hat{Q}^h$ is nonzero. Thus, $\hat{W}_g$ will go to zero and so $\hat{W}_g\hat{W}_f$ will also go to zero. If H is right but G is wrong, following the same logic but switching the roles of g and h, $\hat{W}_g$ will go to one and so $\hat{W}_g\hat{W}_f$ will also go to one. When both G and H are correctly specified, so F is correctly specified, then $\hat{Q}^f \to_p c_f'\Omega_f c_f/k_f = 0$ so $n^\tau\hat{Q}^f \to_p 0$ and therefore $\hat{W}_f \to_p 0$. And because both $n\hat{Q}^g$ and $n\hat{Q}^h$ converge to chi-squared distributions, $\hat{W}_g$ converges to a ratio of possibly dependent chi-squareds, which is bounded in probability, so in this case $\hat{W}_g\hat{W}_f \to_p 0$.

Under Assumptions A1 to A6 along with Lemma 1, the following theorem shows consistency of the GDR estimates $\widehat{\alpha}$ in equation (1).

Theorem 1 : Suppose that z_i, $i = 1, 2, ...$, are iid, $\hat{\Omega}_g \to_p \Omega_g$, $\hat{\Omega}_h \to_p \Omega_h$, $\hat{\Omega}_f \to_p \Omega_f$, Ω_g, Ω_h and Ω_f are positive definite, and Assumptions A1 to A6 hold. Then for $\widehat{\alpha}$ given by equation (1), $\widehat{\alpha} \to_p \alpha_0$.

We discuss choice of $\hat{\Omega}_g$, $\hat{\Omega}_h$, and $\hat{\Omega}_f$ later, but note for now that these can be standard GMM weight matrix estimates.

Proof of Theorem 1: By A1, A2, A3, A5, and A6, the conditions of Theorem 2.1 of in Newey and McFadden (1994) (uniqueness, compactness, continuity, and uniform convergence) hold for GMM based on model G, model H, or both when these moments are correctly specified. Therefore when either A2-1), A2-2), or both hold, the corresponding GMM estimators are consistent.

Now consider $\widehat{\alpha}$. For simplicity, let $\hat{Q}^g \equiv \hat{Q}^g(\widehat{\alpha}_g, \widehat{\beta}_g)$, $\hat{Q}^h \equiv \hat{Q}^h(\widehat{\alpha}_h, \widehat{\gamma}_h)$, $\hat{Q}^f \equiv \hat{Q}^f(\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f)$, $Q_0^g \equiv c_g'\Omega_g c_g/k_g$, $Q_0^h \equiv c_h'\Omega_h c_h/k_h$, and $Q_0^f \equiv c_f'\Omega_f c_f/k_f$ unless we specify otherwise. We show

consistency of $\widehat{\alpha}$ given by equation (1),

$$\widehat{\alpha} = \hat{W}_f \hat{W}_g \widehat{\alpha}_h + \hat{W}_f \left(1 - \hat{W}_g\right) \widehat{\alpha}_g + (1 - \hat{W}_f)\widehat{\alpha}_f,$$

by considering the three possible cases of just G being correctly specified, just H being correctly specified, or both correctly specified.

Case 1: Suppose in Assumption A2 that $g_0(\alpha_0, \beta_0) = 0$ and $h_0(\alpha_0, \gamma_0) \neq 0$. Then $\{\widehat{\alpha}_g, \widehat{\beta}_g\} \rightarrow_p \{\alpha_0, \beta_0\}$, $\{\widehat{\alpha}_h, \widehat{\gamma}_h\} \rightarrow_p \{\alpha_h, \gamma_h\}$, and $\{\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f\} \rightarrow_p \{\alpha_f, \beta_f, \gamma_f\}$. By Lemma 1, $\hat{W}_g$ converges to zero and $\hat{W}_f$ converges to one in probability. The consistency of $\widehat{\alpha}$ in (1) therefore follows from consistency of $\widehat{\alpha}_g$.

Case 2: Suppose in Assumption A2 that $g_0(\alpha_0, \beta_0) \neq 0$ and $h_0(\alpha_0, \gamma_0) = 0$. Then $\{\widehat{\alpha}_g, \widehat{\beta}_g\} \rightarrow_p \{\alpha_g, \beta_g\}$, $\{\widehat{\alpha}_h, \widehat{\gamma}_h\} \rightarrow_p \{\alpha_0, \gamma_0\}$, and $\{\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f\} \rightarrow_p \{\alpha_f, \beta_f, \gamma_f\}$. By Lemma 1, $\hat{W}_g$ converges to one and $\hat{W}_f$ also converges to one in probability. The consistency of $\widehat{\alpha}$ in (1) therefore follows from consistency of $\widehat{\alpha}_h$.

Case 3: Suppose in Assumption A2 that both $g_0(\alpha_0, \beta_0) = 0$ and $h_0(\alpha_0, \gamma_0) = 0$. Then $\{\widehat{\alpha}_g, \widehat{\beta}_g\} \rightarrow_p \{\alpha_0, \beta_0\}$, $\{\widehat{\alpha}_h, \widehat{\gamma}_h\} \rightarrow_p \{\alpha_0, \gamma_0\}$, and $\{\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f\} \rightarrow_p \{\alpha_0, \beta_0, \gamma_0\}$, so $\hat{Q}^g \rightarrow_p 0$, $\hat{Q}^h \rightarrow_p 0$, and $\hat{Q}^f \rightarrow_p 0$. By Lemma 1, $\hat{W}_f$ and $\hat{W}_f \hat{W}_g$ converge to zero and the consistency of $\widehat{\alpha}$ in (1) therefore follows from consistency of $\widehat{\alpha}_f$. Q.E.D.

4.2 Limiting Distribution

In this section we provide the asymptotic distribution of $\widehat{\alpha}$, and a simple consistent estimator of its limiting variance. Let $\theta_0^g \equiv \{\alpha_0, \beta_0\}$, $\theta_0^h \equiv \{\alpha_0, \gamma_0\}$, $\theta_0^f \equiv \{\alpha_0, \beta_0, \gamma_0\}$, $\theta^g \equiv \{\alpha_g, \beta_g\}$, $\theta^h \equiv \{\alpha_h, \gamma_h\}$, and $\theta^f \equiv \{\alpha_f, \beta_f, \gamma_f\}$. Let $\nabla_\theta g_0(\theta^g)$, $\nabla_\theta h_0(\theta^h)$, and $\nabla_\theta f_0(\theta^f)$ denote $E\{\nabla_{\theta^g} G(Z, \alpha_g, \beta_g)\}$, $E\{\nabla_{\theta^h} H(Z, \alpha_h, \gamma_h)\}$, and $E\{\nabla_{\theta^f} F(Z, \alpha_f, \beta_f, \gamma_f)\}$, respectively.

Assumption A7: With probabilities approaching one, $G(Z, \alpha, \beta)$, $H(Z, \alpha, \gamma)$, and $F(Z, \alpha, \beta, \gamma)$ are twice continuously differentiable in neighborhoods $\aleph^g$ of θ^g, $\aleph^h$ of θ^h, and $\aleph^f$ of θ^f, respectively.

Assumption A8: $\{\nabla_\theta g_0(\theta_0^g)\}'\Omega_g\{\nabla_\theta g_0(\theta_0^g)\}$, $\{\nabla_\theta h_0(\theta_0^h)\}'\Omega_h\{\nabla_\theta h_0(\theta_0^h)\}$, and $\{\nabla_\theta f_0(\theta_0^f)\}'\Omega_f\{\nabla_\theta f_0(\theta_0^f)\}$ are non-singular.

Assumption A7 are regularity conditions used to show asymptotic normality of standard GMM. Assumptions A8 is roughly analogous to ruling out perfect collinearity in linear regression.

Assumption A9: $\{\alpha_g, \beta_g\}$, $\{\alpha_h, \gamma_h\}$, and $\{\alpha_f, \beta_f, \gamma_f\}$ lie in the interior of $\Theta_\alpha \times \Theta_\beta$, $\Theta_\alpha \times \Theta_\gamma$, and $\Theta_\alpha \times \Theta_f \times \Theta_\gamma$.

Assumption A10: $E[||G(Z, \alpha, \beta)||^2] < \infty$, $E[||H(Z, \alpha, \gamma)||^2] < \infty$, and $E[||F(Z, \alpha, \beta, \gamma)||^2] < \infty$.

Assumption A11: $E[\sup_{\{\alpha,\beta\}\in\aleph^g} ||\nabla_{\theta^g} G(Z, \alpha, \beta)||] < \infty$, $E[\sup_{\{\alpha,\gamma\}\in\aleph^h} ||\nabla_{\theta^h} H(Z, \alpha, \gamma)||] < \infty$, and $E[\sup_{\{\alpha,\beta,\gamma\}\in\aleph^f} ||\nabla_{\theta^f} F(Z, \alpha, \beta, \gamma)||] < \infty$.

Assumptions A9, A10 and A11 are standard regularity conditions. Let $\widehat{\eta}_i^g$, $\widehat{\eta}_i^h$ and $\widehat{\eta}_i^f$ be consistent estimators of the GMM influence functions associated with the standard GMM estimators $\widehat{\alpha}_g$, $\widehat{\alpha}_h$ and $\widehat{\alpha}_f$. These influence function formulas, based on Hansen (1982), are given in the Appendix.

Theorem 2: Suppose $\hat{\Omega}_g \to_p \Omega_g$, $\hat{\Omega}_h \to_p \Omega_h$, $\hat{\Omega}_f \to_p \Omega_f$, Ω_g, Ω_h and Ω_f are positive definite, and assumptions A1 to A12 hold. Then there exists a matrix $\widetilde{V}$ such that

$$\sqrt{n}(\widehat{\alpha} - \alpha_0) \to^d N(0, \widetilde{V}),$$

and

$$\frac{1}{N}\sum_i \widehat{\eta}_i \widehat{\eta}_i' \to_p \widetilde{V} \tag{23}$$

$$\text{where } \widehat{\eta}_i \equiv \hat{W}_f \hat{W}_g \widehat{\eta}_i^h + \hat{W}_f(1 - \hat{W}_g)\widehat{\eta}_i^g + (1 - \hat{W}_f)\widehat{\eta}_i^f.$$

The first part of Theorem 2 states that the GDR estimator $\widehat{\alpha}$ is root N consistent and asymptotically normal, while the second part gives a consistent estimator for the limiting variance of $\widehat{\alpha}$.

The proof of Theorem 2 is given in the Appendix. The basic structure of the proof follows Newey and McFadden (1994) for multistep parametric estimators.

In the Appendix we define variance matrices $\widetilde{V}^g$, $\widetilde{V}^h$, and $\widetilde{V}^f$. The matrix $\widetilde{V}$ equals $\widetilde{V}^g$ if model G is correctly specified and model H is not, while $\widetilde{V}$ equals $\widetilde{V}^h$ if H is correctly specified and G is not, and $\widetilde{V}$ equals $\widetilde{V}^f$ when both G and H are correctly specified. Importantly, the consistent estimator of $\widetilde{V}$ given in equation (23) does not require knowing which of the models G or H is correct.

A complication in the derivation of Theorem 2 is that, if model H is wrong, then we cannot consistently estimate the influence function η_i^h for model H. However, in the limiting variance formula for $\widehat{\alpha}$, the function η_i^h is multiplied by $\hat{W}_f\hat{W}_g$, so if model H is wrong then $\hat{W}_f\hat{W}_g$ goes to zero. We therefore only need an estimate for η_i^h that is consistent when model H is right, and that estimate is the standard GMM influence function $\widehat{\eta}_i^h$. A similar analysis applies to the influence function $\widehat{\eta}_i^g$ for model G when model G is wrong.

4.3 Efficiency and Numerical Issues

For asymptotic efficiency of α, we could consider estimates of the weighting matrices $\hat{\Omega}_g$, $\hat{\Omega}_h$, and $\hat{\Omega}_f$ that minimize the variance given by equation (23). However, the standard two step GMM estimates of $\hat{\Omega}_g$, $\hat{\Omega}_h$, and $\hat{\Omega}_f$ (leading to asymptotically efficient GMM estimation) should also be close to efficient for $\widehat{\alpha}$. This is because the GDR objective function is asymptotically dominated by the GMM objective function of the correct model when either G or H is correct, and dominated by the GMM objective function of model F when both models are correct.

Also, the scaling or units of moments affect the relative magnitudes of $\hat{Q}^g$, $\hat{Q}^h$, and $\hat{Q}_f$ (and hence the estimated weights $\hat{W}_g$ and $\hat{W}_f$). It is therefore numerically desirable in finite samples to have these matrices be comparable in magnitude. The standard two step GMM estimates of $\hat{\Omega}_g$, $\hat{\Omega}_h$, and $\hat{\Omega}_f$ help make $\hat{Q}^g$, $\hat{Q}^h$, and $\hat{Q}_f$ comparable. Specifically, $\hat{Q}^g$ will then asymptotically have a mean of one when model G is right, and the same is true for $\hat{Q}^h$ when H is right, and for $\hat{Q}^f$ when

both G and H are right (this is also the role of scaling each by the degrees of freedom k_g, k_h, and k_f, respectively). We therefore find it desirable to use the standard GMM estimates of $\hat{\Omega}_g$ and $\hat{\Omega}_h$, even if that potentially sacrifices a small amount of asymptotic efficiency.

5 Simulation Results

Here we do some Monte Carlo analyses to investigate small sample properties of our estimator. Our design is two competing sets of instruments as in section 3.4. For each simulation, we draw $n = 100$ or $n = 500$ independent, identically distributed observations of the random vector $(Y, W, R_1, R_2, Q_1, Q_2)$. We generate data from the model

$$Y = \alpha_0 + \alpha_1 W + \epsilon.$$

The goal is estimation of $\alpha = (\alpha_0, \alpha_1) = (1, 1)$. The regressor W is endogenous (correlated with ϵ), so estimation is by instrumental variables. Model G assumes $E(\epsilon) = E(\epsilon R_1) = E(\epsilon R_2) = 0$, meaning that $R = (1, R_1, R_2)'$ is a vector of valid instruments for instrumental variables estimation. Model H assumes $E(\epsilon) = E(\epsilon Q_1) = E(\epsilon Q_2) = 0$, making $Q = (1, Q_1, Q_2)'$ be a vector of valid instruments. Here $Z = (Y, W, R, Q)$, $G(Z, \alpha) = (Y - \alpha_0 - \alpha_1 W) R$, and $H(Z, \alpha) = (Y - \alpha_0 - \alpha_1 W) Q$. In this application there is no β or γ.

We let $W = 1 + 4R_1 + R_2 + 2Q_1 + Q_2 + \epsilon$. Having the 4 and 2 in this equation means that the G model has stronger instruments (i.e., instruments more highly correlated with the endogenous regressor W) than the H model, and that R_1 and Q_1 are stronger instruments than R_2 and Q_2. A general property of instrumental variables estimation is that, the stronger the instruments (when valid) the more precise are the resulting estimates.

We let R_1, R_2, Q_1, Q_2, and ϵ be standard normals, with $corr(R_j, \epsilon) = \rho_R$, $corr(Q_j, \epsilon) = \rho_Q$, and all the other correlations among these normals are zero. The correlations ρ_R and ρ_Q are each set to either 0.4 or zero.

We consider three different simulation designs. The first takes $\rho_R = \rho_Q = 0$, which makes both

models be right (both sets of instruments are valid). The second takes $\rho_R = 0$ and $\rho_Q = 0.4$, which makes model G be right (i.e., R are valid instruments so G is correctly specified) and model H be wrong (i.e., Q are not valid instruments, because they correlate with the model error ϵ). The third takes $\rho_R = 0.4$ and $\rho_Q = 0$, which makes model H be right and model G wrong.

We report five estimates of α_1 and α_0 for each simulation. First is GMM based on the model G moments, denoted by GMM_g (which is only consistent if model G is right), second is GMM based on the H moments, denoted by GMM_h (which is only consistent if model H is right), third is GMM based on both sets of moments, denoted by GMM_f (which is consistent, and more efficient than either the first or second set of estimates, only if both models are right), fourth is our simpler GDR estimator in equation (20), denoted by $SGDR$ (which is consistent if either set of moments is valid), and fifth is our GDR estimator in equation (1), denoted by GDR (which is also consistent if either set of moments is valid, and should be more efficient than $SGDR$ when both sets of moments are valid).

Tables 1 and 2 present simulation results of 1000 repetitions for each of the five estimators. The reported summary statistics are, respectively, the bias (Bias), median error (MdE), root mean-squared error (RMSE), median absolute error (MAE), and standard deviation (SD). Also to verify the finite sample accuracy of our limiting distribution, the average (across the simulations) of the estimated standard errors (SE) and their standard deviation (SD_{SE}) are reported.

As expected, when correctly specified GMM_g estimates are more precise (by all measures) than GMM_h estimates, and when both sets of instruments are valid, GDR estimates are almost as precise as GMM_f. The GDR estimates are more precise than all of the inconsistent estimators, with one exception: in the smaller sample size when the H model is invalid, GMM_f is more accurate than GDR. In this case the GMM_f weighting matrix put relatively little weight on the moments associated with the invalid weak instruments. But by the larger sample size, GDR is always superior to estimators containing any invalid instruments. As expected, the $SGDR$ estimates are less precise than GDR when both G and H models are valid. However, when one model is invalid, $SGDR$ is

similar to GDR. So the cost in efficiency of choosing the simpler $SGDR$ is small.[4]

Table 1. Simulation Results of α_1

	$n = 100$							$n = 500$						
	Bias	MdE	RMSE	MAE	SD	SE	SD_{SE}	Bias	MdE	RMSE	MAE	SD	SE	SD_{SE}
Both Correct														
GMM_g	0.0006	0.0000	0.0007	0.0179	0.0257	0.0240	0.0040	-0.0005	-0.0006	0.0001	0.0074	0.0109	0.0109	0.0007
GMM_h	-0.0029	0.0010	0.0022	0.0294	0.0471	0.0458	0.0128	-0.0011	-0.0010	0.0004	0.0133	0.0204	0.0202	0.0022
GMM_f	0.0009	0.0014	0.0005	0.0167	0.0233	0.0205	0.0030	-0.0005	-0.0004	0.0001	0.0066	0.0096	0.0095	0.0006
SGDR	-0.0025	-0.0010	0.0011	0.0202	0.0337			-0.0010	-0.0014	0.0002	0.0082	0.0143		
GDR	0.0003	0.0008	0.0006	0.0169	0.0249	0.0211	0.0036	-0.0006	-0.0005	0.0001	0.0068	0.0104	0.0097	0.0010
G Correct														
GMM_g	-0.0000	0.0010	0.0006	0.0167	0.0254	0.0238	0.0045	-0.0005	-0.0000	0.0001	0.0073	0.0110	0.0108	0.0009
GMM_h	0.1766	0.1743	0.0326	0.1743	0.0376	0.0355	0.0097	0.1778	0.1776	0.0319	0.01776	0.0169	0.0159	0.0018
GMM_f	0.0583	0.0591	0.0038	0.0591	0.0211	0.0173	0.0024	0.0577	0.0577	0.0034	0.0577	0.0096	0.0081	0.0005
SGDR	0.0647	0.0512	0.0083	0.0515	0.0637			0.0274	0.0150	0.0022	0.0163	0.0383		
GDR	0.0636	0.0527	0.0071	0.0529	0.0552	0.0205	0.0058	0.0285	0.0165	0.0022	0.0177	0.0370	0.0094	0.0015
H Correct														
GMM_g	0.1007	0.1017	0.0105	0.1017	0.0193	0.0171	0.0025	0.1011	0.1009	0.0103	0.1009	0.0080	0.0077	0.0005
GMM_h	-0.0012	0.0044	0.0025	0.0303	0.0499	0.0471	0.0241	-0.0000	0.0013	0.0004	0.0136	0.0204	0.0201	0.0031
GMM_f	0.0827	0.0835	0.0072	0.0835	0.0178	0.0152	0.0021	0.0828	0.0830	0.0069	0.0830	0.0076	0.0070	0.0004
SGDR	0.0138	0.0197	0.0026	0.0350	0.0493			0.0033	0.0049	0.0004	0.0143	0.0204		
GDR	0.0349	0.0399	0.0028	0.0439	0.0399	0.0289	0.0146	0.0076	0.0092	0.0005	0.0151	0.0199	0.0184	0.0031

[4]However, $SGDR$ incurs the additional cost of possibly not having a simple limit normal distribution when both G and H are correctly specified.

Table 2. Simulation Results of α_0

	$n = 100$							$n = 500$						
	Bias	MdE	RMSE	MAE	SD	SE	SD_{SE}	Bias	MdE	RMSE	MAE	SD	SE	SD_{SE}
Both Correct														
GMM_g	0.0016	0.0012	0.0111	0.0731	0.1052	0.1015	0.0092	0.0005	-0.0006	0.0021	0.0311	0.0460	0.0460	0.0017
GMM_h	0.0052	0.0067	0.0135	0.0771	0.1160	0.1119	0.0175	0.0012	0.0010	0.0025	0.0316	0.0498	0.0494	0.0031
GMM_f	0.0011	0.0039	0.0114	0.0742	0.1067	0.0985	0.0087	0.0004	-0.0010	0.0021	0.0308	0.0459	0.0455	0.0016
SGDR	0.0051	0.0077	0.0119	0.0747	0.1091			0.0009	0.0001	0.0022	0.0306	0.0471		
GDR	0.0021	0.0064	0.0114	0.0742	0.1068	0.0989	0.0090	0.0005	-0.0016	0.0021	0.0304	0.0461	0.0456	0.0017
G Correct														
GMM_g	-0.0002	-0.0015	0.0111	0.0734	0.1054	0.1013	0.0099	0.0015	0.0020	0.0021	0.0312	0.0463	0.0459	0.0020
GMM_h	-0.1740	-0.1709	0.0431	0.1709	0.1131	0.1076	0.0153	-0.1770	-0.1779	0.0340	0.1779	0.0513	0.0483	0.0029
GMM_f	-0.0565	-0.0594	0.0146	0.0839	0.1070	0.0892	0.0072	-0.0568	-0.0581	0.0056	0.0584	0.0487	0.0412	0.0014
SGDR	-0.0632	-0.0669	0.0173	0.0920	0.1153			-0.0266	-0.0218	0.0042	0.0393	0.0594		
GDR	-0.0620	-0.0641	0.0160	0.0903	0.1104	0.0946	0.0093	-0.0277	-0.0233	0.0042	0.0394	0.0584	0.0439	0.0024
H Correct														
GMM_g	-0.1002	-0.0999	0.0184	0.1020	0.0914	0.0831	0.0070	-0.1011	-0.1029	0.0118	0.1029	0.0397	0.0377	0.0013
GMM_h	0.0037	-0.0064	0.0139	0.0762	0.1180	0.1137	0.0332	0.0004	-0.0001	0.0023	0.0315	0.0479	0.0492	0.0045
GMM_f	-0.0819	-0.0813	0.0155	0.0909	0.0936	0.0818	0.0069	-0.0826	-0.0835	0.0086	0.0835	0.0415	0.0377	0.0013
SGDR	-0.0116	-0.0214	0.0130	0.0755	0.1134			-0.0029	-0.0034	0.0023	0.0314	0.0476		
GDR	-0.0334	-0.0399	0.0120	0.0762	0.1044	0.0955	0.0187	-0.0073	-0.0074	0.0023	0.0321	0.0470	0.0476	0.0042

One should expect correctly specified GMM estimators to be more efficient than GDR, and that is indeed the case. But in many of the simulations, the loss in efficiency from using GDR is very low. In particular, when model G is invalid, so only the weaker instruments are valid, the precision of GDR is almost identical to that of the efficient GMM_h. So, using our GDR, there is little loss

in efficiency from not knowing which specification is correct.

In conclusion, the proposed GDR is shown to work well (even at low sample sizes) regardless of which moments are misspecified, and GDR improves as the sample size increases.

6 Empirical Application: Engel Curve Estimation

Here we empirically estimate the Engel curve example discussed in section 3.4. Y is the food budget share, S is log real total consumption expenditures, and X is a vector of additional covariates that affect the consumer's tastes (e.g. age, spouse's age, and several indicators). The goal is estimation of the coefficient of S in a regression of Y on S and X. Total consumption S is observed with measurement error, so instrumental variables estimation is used to correct for the resulting endogeneity (i.e., the correlation between S and the regression error). The vector L consists of two candidate external instrument variables, real total income and real total income squared. Model G assumes these external instruments are valid. Model H instead assumes that constructed instruments based on heteroscedasticity as described by Lewbel (2012) and summarized in section 2.4 above are valid.

The data consist of 854 households collected from the UK Family Expenditure Survey 1980-1982 as studied by Banks, Blundell, and Lewbel (1997), Lewbel (2012), and Baum and Schaffer (2012). The sample means are $\overline{Y} = 0.285$ and $\overline{S} = 0.599$, and the standard deviations are $SD(Y) = 0.106$ and $SD(S) = 0.410$.

The parameter of interest is the coefficient of log real total expenditure α_s. Table 3 summarizes estimates of α_s and of the constant term α_0. GMM_{g0} is the estimate reported in Lewbel (2012) and Baum and Schaffer (2012), which used log total income as a single instrument and so is not over identified. GMM_g is the GMM estimator using the moments in equation (16), which makes use of the external instruments L, which are income and income squared. GMM_h is the GMM estimator that uses the moments in equation (17), which are heteroscedasticity based constructed instruments, as described in section 2.4, based on Lewbel (2012). GMM_f is the GMM estimator

that uses both sets of instruments, and $SGDR$ and GDR are our new estimators given in equations (20) and (1), respectively.[5]

Table 3. Engel Curve Estimates

	GMM_{g0}	GMM_g	GMM_h	GMM_f	$SGDR$	GDR
$\hat{\alpha}_s$	−0.0859 (0.0197)	−0.0839 (0.0194)	−0.0521 (0.0550)	−0.0862 (0.0174)	−0.0795	−0.0862 (0.0174)
$\hat{\alpha}_0$	0.336 (0.0121)	0.335 (0.0120)	0.317 (0.0330)	0.337 (0.0107)	0.333	0.336 (0.0107)
χ^2		0.191	12.91	15.94		
$d.f.$		1	11	13		
P-value		0.662	0.299	0.252		
$\hat{Q}$		0.0002	0.0014	0.0014		
$\hat{W}_g, \hat{W}_f, p$				0.140,	0.003,	0.89

[6]

The estimated results show that the external instruments of model G are much stronger than the constructed instruments of model H. This is not surprising since the constructed instruments are based on higher moments of the data. This difference in strength can be seen in the standard errors of $\hat{\alpha}_s$, which are much lower in model G than in model H, and also in model GMM_f which gives estimates much closer to GMM_g than GMM_h.

[5] As Table 3 shows, the estimates of GMM_{g0} and GMM_g are very similar. There's a similar small difference between GMM_f and the models based on both sets of moments reported in Lewbel (2012) and Baum and Schaffer (2012), again because GMM_f uses the two external instruments income and income squared, instead of just using income.

[6] Note: For simplicity, we implemented the estimator using demeaned regressors, and then recovered the reported intercepts based on the estimated coefficients and variable means. We report coefficient estimates with associated standard errors in parentheses, except SGDR. Also reported are the Hansen (1982) test statistics for overidentified GMM, along with their degrees of freedom and p-values. $\hat{Q}$ is the normalized minimand of the GMM estimators. p denotes the p-value of the Wald statistic testing the null hypothesis that $\widehat{\alpha}_g = \widehat{\alpha}_h$. This p is used to set up $\tau = 1 - p$ of $\hat{W}_f$ in equation (21), as explained in section 3.

The point estimates of GMM_g and GMM_h are substantially different, which could be due to having one of these sets of instruments be invalid. However, this difference could also just be due to imprecision, particularly of GMM_h. This illustrates the usefulness of our GDR, which does not require resolving which set of instruments is valid, or if both are valid.

Based on the reported values of the objective functions $\hat{Q}^g$, $\hat{Q}^h$, and $\hat{Q}^f$, the estimated weight $\hat{W}_g$ is 0.140, so $SGDR$ puts over six times as much weight on model G as on model H (0.860 vs. 0.140). However, $\hat{W}_f$ is 0.003, making the GDR estimate very much closer to GMM_f than to either GMM_g or GMM_h. These weights suggest that both models are likely to be correctly specified, and so the difference between GMM_g and GMM_h is likely due to the imprecision of GMM_h rather than misspecification. Further evidence that both are correctly specified is given by the chi-squared statistics in Table 3, which test validity of the moments comprising each of the GMM estimates. This situation, where both models appear to be correctly specified, is when we would expect GDR to perform better than $SGDR$.

Lewbel (2012) observes that a virtue of the constructed instruments is that they are valid under very different conditions than those required for validity of the external instruments, and suggests that they therefore are useful for testing overidentification. Our proposed GDR estimator makes further use of these instruments, by delivering estimates that are consistent if either (or both) sets of instruments are valid.

7 Extensions: Multiple Robustness and Alternative GDR's

It is possible to construct triply and higher multiply robust estimators. Suppose we have a third model, called model L, with GMM objective function $\hat{Q}^l(\alpha, \delta)$. The GMM estimator of model L is $\{\widehat{\alpha}_l, \widehat{\delta}_l\} = \arg\min_{\{\alpha,\delta\}\in\Theta_\alpha\times\Theta_\delta} \hat{Q}^l(\alpha, \delta)$. A possible formula for triply robust estimation of α would then be the weighted average

$$\widetilde{\alpha} = \frac{\hat{Q}^g(\widehat{\alpha}_g, \widehat{\beta}_g)\hat{Q}^h(\widehat{\alpha}_h, \widehat{\gamma}_h)\widehat{\alpha}_l + \hat{Q}^l(\widehat{\alpha}_l, \widehat{\delta}_l)\hat{Q}^h(\widehat{\alpha}_h, \widehat{\gamma}_h)\widehat{\alpha}_g + \hat{Q}^l(\widehat{\alpha}_l, \widehat{\delta}_l)\hat{Q}^g(\widehat{\alpha}_g, \widehat{\beta}_g)\widehat{\alpha}_h}{\hat{Q}^g(\widehat{\alpha}_g, \widehat{\beta}_g)\hat{Q}^h(\widehat{\alpha}_h, \widehat{\gamma}_h) + \hat{Q}^l(\widehat{\alpha}_l, \widehat{\delta}_l)\hat{Q}^h(\widehat{\alpha}_h, \widehat{\gamma}_h) + \hat{Q}^l(\widehat{\alpha}_l, \widehat{\delta}_l)\hat{Q}^g(\widehat{\alpha}_g, \widehat{\beta}_g)}. \tag{24}$$

In equation (24), the weight on $\widehat{\alpha}_l$ is proportional to the product of objective functions for the other models, $\hat{Q}^g\hat{Q}^h$, and similarly for the weights on $\widehat{\alpha}_g$ and $\widehat{\alpha}_h$.

The logic of this estimator is the same as for our $SGDR$ estimator, that is, $\widetilde{\alpha}$ as given by equation (20). For example, if model G is right and models L and H are wrong, then only $\widehat{\alpha}_g$ will get a nonzero weight asymptotically. Now suppose two but not all three models are right, e.g., suppose models G and H are right and L is wrong. Then all the weights in both the numerator and denominator of equation (24) go to zero. However, in this case we can divide the numerator and denominator by $\hat{Q}^g(\widehat{\alpha}_g, \widehat{\beta}_g)$. Both $\hat{Q}^g(\widehat{\alpha}_g, \widehat{\beta}_g)$ and $\hat{Q}^h(\widehat{\alpha}_h, \widehat{\gamma}_h)$ converge to zero, but if $n\hat{Q}^g(\widehat{\alpha}_g, \widehat{\beta}_g)/n\hat{Q}^h(\widehat{\alpha}_h, \widehat{\gamma}_h)$ is finite and nonzero, then the limiting weights on $\widehat{\alpha}_g$ and $\widehat{\alpha}_h$ will be nonzero while the limiting weight on $\widehat{\alpha}_l$ will be zero, as desired.

As with $SGDR$, the limiting distribution of the triply robust estimator $\widetilde{\alpha}$ in equation (24) is complicated by the potential limiting randomness of ratios like $n\hat{Q}^g(\widehat{\alpha}_g, \widehat{\beta}_g)/n\hat{Q}^h(\widehat{\alpha}_h, \widehat{\gamma}_h)$ in the weights. In the doubly robust case, we avoided this problem using the additional weight W_f for when both models are correctly specified. An analogous construction for triply robust estimation will be more difficult, since we would also need to consider the cases where any pair of models is correct, and when all three are correct.

It is possible to construct alternative GDR's based on the idea of our proposed GDR. For example, one could replace $\hat{Q}^g$ and $\hat{Q}^h$ in equation (19) with $\zeta\left(\hat{Q}^g\right)$ and $\zeta\left(\hat{Q}^h\right)$ for any strictly monotonically increasing function ζ having $\zeta(0) = 0$. Another alternative estimator would be

$$\widetilde{\alpha}^a = \arg\min_{\alpha\in\Theta_\alpha} \hat{Q}^g(\alpha, \widehat{\beta}_g)\hat{Q}^h(\alpha, \widehat{\gamma}_h). \tag{25}$$

Consistency of the alternative $\widetilde{\alpha}^a$ follows a similar logic to the SGDR. For example, if model G is right and H is not, then minimizing $\hat{Q}^g(\alpha, \widehat{\beta}_g)\hat{Q}^h(\alpha, \widehat{\gamma}_h)$ will be asymptotically equivalent to minimizing $\hat{Q}^g(\alpha, \widehat{\beta}_g)$ because $\hat{Q}^g$ will go to zero while $\hat{Q}^h$ cannot. A disadvantage of this alternative estimator is that it's more numerically complicated than SGDR, because it entails an additional numerical search for $\widehat{\alpha}$ instead of just taking a weighted average of $\widehat{\alpha}_g$ and $\widehat{\alpha}_h$. It will also have similar limiting distribution issues as $SGDR$. However, one advantage of $\widetilde{\alpha}^a$ is that it more readily

extends to triple and higher multiply robust cases. For example, the alternative general triply robust estimator would just be

$$\widehat{\alpha}^a = \arg\min_{\alpha\in\Theta_\alpha} \hat{Q}^l(\alpha,\widehat{\delta}_l)\hat{Q}^g(\alpha,\widehat{\beta}_g)\hat{Q}^h(\alpha,\widehat{\gamma}_h). \tag{26}$$

8 Conclusions

In this paper, we provided a general technique for constructing doubly robust estimators. Our General Doubly Robust (GDR) technique is a simple extension of the Generalized Method of Moments. It takes the form of a weighted average of Hansen's (1982) Generalized Method of Moments (GMM) based estimates, and has similar associated root-n asymptotics. We illustrated our GDR with a variety of potential applications, including average treatment effect estimation. The proposed estimator appears to work well in a small Monte Carlo study and in an empirical application to instrumental variables estimation, where either one of two sets of instrument vectors might be invalid.

9 References

Andrews, D.W.K. and Lu, B. (2001): "Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models", *Journal of Econometrics*, 101(1), 123-164.

Banks, J., Blundell, R., and Lewbel, A. (1997): "Quadratic Engel Curves and Consumer Demand", *Review of Economics and Statistics*, 79, 527-539.

Bang, H., and Robins, J. (2005): "Doubly Robust Estimation in Missing Data and Causal Inference Models", *Biometrics*, 61(4), 962-973.

Baum, C., and Schaffer, M. (2012): "IVREG2H: Stata Module to Perform Instrumental Variables Estimation Using Heteroskedasticity-based Instruments", Statistical Software Components S457555, Boston College Department of Economics, revised 18 Feb 2018.

Campbell, J., and Cochrane, J. (1999): "By Force of Habit: A Consumption?Based Explanation of Aggregate Stock Market Behavior", *Journal of Political Economy*, 107(2), 205-251.

Caner, M. (2009): "Lasso-type GMM Estimator," *Econometric Theory*, 25(1) 270-290.

Chen, X., and Ludvigson, S. (2009): "Land of Addicts? an Empirical Investigation of Habit-based Asset Pricing Models", *Journal of Applied Econometrics*, 24(7), 1057-1093.

Chernozhukov, V., Escanciano, J.C., Ichimura, H., Newey, W. and Robins, J. (2018): "Locally Robust Semiparametric Estimation," Unpublished Manuscript.

Cochrane, J. (2001): "Long-Term Debt and Optimal Policy in the Fiscal Theory of the Price Level", *Econometrica*, 69(1), 69-116.

DiTraglia, F. (2016): "Using Invalid Instruments on Purpose: Focused Moment Selection and Averaging for GMM", *Journal of Econometrics*, 195(2), 187-208.

Funk, M., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M., and Davidian, M. (2011): "Doubly Robust Estimation of Causal Effects", *American Journal of Epidemiology*, 173(7), 761-7.

Hansen, B. (2007): "Least Squares Model Averaging", *Econometrica*, 75, 1175-1189.

Hansen, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, 50(4), 1029-1054.

Hansen, L., and Singleton, K. (1982): "Generalized Instrumental Variables Estimation of Non-linear Rational Expectations Models", *Econometrica*, 50(5), 1269-1286.

Kuersteiner, G. and Okui, R. (2010): "Constructing Optimal Instruments by First-Stage Prediction Averaging.", *Econometrica*, 78(2), 697-718.

Lee, M.J., and Lee, S. (2019): "Double Robustness Without Weighting", *Statistics and Probability Letters*, 146, 175-180.

Lewbel, A. (2008): "Engel curves", entry for The New Palgrave Dictionary of Economics, 2nd Edition, MacMillan Press.

Lewbel, A. (2012): "Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models", *Journal of Business and Economic Statistics*, 30(1), 67-80.

Liao, Z. (2013): "Adaptive GMM Shrinkage Estimation With Consistent Moment Selection", *Econometric Theory*, 29(5), 857-904.

Lunceford, J.K., and Davidian, M. (2004): "Stratification and Weighting via the Propensity Score in Estimation of Causal Treatment Effects: a Comparative Study", *Statistics in Medicine*, 23(19), 2937–2960.

Martins, L.F., and Gabriel, V.J. (2014): "Linear Instrumental Variables Model Averaging Estimation", *Computational Statistics and Data Analysis*, 71, 709-724.

Newey, W. and McFadden, D. (1994): "Chapter 36 Large Sample Estimation and Hypothesis Testing", in Handbook of Econometrics, 4, 2111-2245.

Okui, R., Small, D., Tan, Z., and Robins, J. (2012): "Doubly Robust Instrumental Variable Regression", *Statistica Sinica*, 22(1), 173-205.

Robins, J., Rotnitzky, A., and Van Der Laan, M. (2000): "On Profile Likelihood: Comment", *Journal of the American Statistical Association*, 95(450), 477-482.

Robins, J., Rotnitzky, A., and Zhao, L. (1994): "Estimation of Regression Coefficients When Some Regressors are not Always Observed", *Journal of the American Statistical Association*, 89(427), 846-866.

Rose, S., and Van der Laan, M. (2014): "A Double Robust Approach to Causal Effects in Case-Control Studies", *American Journal of Epidemiology*, 179(6), 663-669.

Scharfstein, D., Rotnitzky, A., and Robins, J. (1999): "Adjusting for Nonignorable Drop-Out Using Semiparametric Nonresponse Models", *Journal of the American Statistical Association*, 94(448), 1096-1120.

Słoczyński, T., and Wooldridge, J. (2018): "A General Double Robustness Result for Estimating Average Treatment Effects", *Econometric Theory*, 34(01), 112-133.

Sueishi, M. (2013): "Generalized Empirical Likelihood-Based Focused Information Criterion and Model Averaging", *Econometrics*, 1(2), 141-156.

Wooldridge, J. (2007): "Inverse Probability Weighted Estimation for General Missing Data

Problems", *Journal of Econometrics*, 141(2), 1281-1301.

Appendix

Proof of Lemma 1: By A1, A2, A3, A5, and A6, the conditions of Theorem 2.1 of in Newey and McFadden (1994) (uniqueness, compactness, continuity, and uniform convergence) hold for GMM_g, GMM_h and GMM_f. Therefore when either A2-1) or A2-2), or both hold, the corresponding GMM estimators are consistent.

For simplicity, let $\hat{Q}^g \equiv \hat{Q}^g(\widehat{\alpha}_g, \widehat{\beta}_g)$, $\hat{Q}^h \equiv \hat{Q}^h(\widehat{\alpha}_h, \widehat{\gamma}_h)$, $\hat{Q}^f \equiv \hat{Q}^f(\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f)$, $Q_0^g \equiv c_g^{'}\Omega_g c_g/k_g$, $Q_0^h \equiv c_h^{'}\Omega_h c_h/k_h$, and $Q_0^f \equiv c_f^{'}\Omega_f c_f/k_f$ unless we specify otherwise.

Case 1: Suppose in Assumption A2 that $g_0(\alpha_0, \beta_0) = 0$ and $h_0(\alpha_0, \gamma_0) \neq 0$. Then $\{\widehat{\alpha}_g, \widehat{\beta}_g\} \to_p \{\alpha_0, \beta_0\}$, $\{\widehat{\alpha}_h, \widehat{\gamma}_h\} \to_p \{\alpha_h, \gamma_h\}$, and $\{\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f\} \to_p \{\alpha_f, \beta_f, \gamma_f\}$. By the continuous mapping theorem and uniform convergence of $\hat{Q}^g$ and $\hat{Q}^h$, $\hat{Q}^g \to_p Q_0^g = c_g^{'}\Omega_g c_g/k_g = 0$, $\hat{Q}^h \to_p Q_0^h = c_h^{'}\Omega_h c_h/k_h > 0$, and $\hat{Q}^f \to_p Q_0^f = c_f^{'}\Omega_f c_f/k_f > 0$. So, $\hat{W}_g = \frac{\hat{Q}^g}{\hat{Q}^g + \hat{Q}^h}$ converges to zero, and for $\hat{W}_f$ we observe

$$\hat{W}_f = 1 - \frac{1}{n^\tau \hat{Q}^f + 1} = \frac{\hat{Q}^f}{\hat{Q}^f + 1/n^\tau} \to_p 1$$

because $\hat{Q}^f \to_p Q_0^f > 0$ and $1/n^\tau \to_p 0$ for $\tau > 0$. Thus, $\hat{W}_g \hat{W}_f$ converges to zero in probability.

Case 2: Suppose in Assumption A2 that $g_0(\alpha_0, \beta_0) \neq 0$ and $h_0(\alpha_0, \gamma_0) = 0$. Then $\{\widehat{\alpha}_g, \widehat{\beta}_g\} \to_p \{\alpha_g, \beta_g\}$, $\{\widehat{\alpha}_h, \widehat{\gamma}_h\} \to_p \{\alpha_0, \gamma_0\}$, and $\{\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f\} \to_p \{\alpha_f, \beta_f, \gamma_f\}$. So $\hat{Q}^g \to_p Q_0^g = c_g^{'}\Omega_g c_g/k_g > 0$, $\hat{Q}^h \to_p Q_0^h = c_h^{'}\Omega_h c_h/k_h = 0$, and $\hat{Q}^f \to_p Q_0^f = c_f^{'}\Omega_f c_f/k_f > 0$. Following the same argument in case 1), $\hat{W}_g \to_p 1$ and $\hat{W}_f \to_p 1$ in probability.

Case 3: Suppose in Assumption A2 that both $g_0(\alpha_0, \beta_0) = 0$ and $h_0(\alpha_0, \gamma_0) = 0$. Then, $f_0(\alpha_0, \beta_0, \gamma_0) = 0$ because f_0 is a union of the correct moments. By A1, A2, A3, A5, and A6, $\{\widehat{\alpha}_g, \widehat{\beta}_g\} \to_p \{\alpha_0, \beta_0\}$, $\{\widehat{\alpha}_h, \widehat{\gamma}_h\} \to_p \{\alpha_0, \gamma_0\}$, and $\{\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f\} \to_p \{\alpha_0, \beta_0, \gamma_0\}$, so $\hat{Q}^g \to_p 0$, $\hat{Q}^h \to_p 0$,

and $\hat{Q}^f \to_p 0$. For $n^\tau \hat{Q}^f$, we observe

$$\begin{aligned} n^\tau \hat{Q}^f &= n^\tau \left\{ \frac{1}{N} \sum_{i=1}^{n} F(\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f) \right\} \hat{\Omega}_g \left\{ \frac{1}{N} \sum_{i=1}^{n} F(\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f) \right\} \frac{1}{k_f} \\ &= \frac{1}{n^{1-\tau}} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} F(\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f) \right\}' \hat{\Omega}_g \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} F(\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f) \right\} \frac{1}{k_f} \\ &= \frac{1}{n^{1-\tau}} \hat{\chi}^2_{k_f} / k_f, \end{aligned}$$

where $\hat{\chi}^2_{k_f} \equiv \{1/\sqrt{n} \sum_{i=1}^{n} F(\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f)\}' \hat{\Omega}_g \{1/\sqrt{n} \sum_{i=1}^{n} F(\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f)\}$. Because $\hat{\chi}^2_{k_f} \to^d \chi^2_{k_f}$, by Slutsky's theorem, $n^\tau \hat{Q}^f \to_p 0$ for $\tau < 1$. And therefore

$$\hat{W}_f = 1 - \frac{1}{n^\tau \hat{Q}^f + 1} \to_p 0.$$

For $\hat{W}_g$, because $n\hat{Q}^g(\widehat{\alpha}_g, \widehat{\beta}_g) \to_d \chi^2_{k_g}/k_g$ and $n\hat{Q}^h(\widehat{\alpha}_h, \widehat{\gamma}_h) \to_d \chi^2_{k_h}/k_h$, thus $\hat{W}_g = \frac{n\hat{Q}^g(\widehat{\alpha}_g, \widehat{\beta}_g)}{n\hat{Q}^g(\widehat{\alpha}_g, \widehat{\beta}_g) + n\hat{Q}^h(\widehat{\alpha}_h, \widehat{\gamma}_h)}$ will converge to a ratio of possibly dependent chi-squared random values that, with probability one, lies between zero and one, so long as H includes at least one moment that is excluded from G. But this exclusion must hold, because if it didn't then H would be nested in G, and so Assumption A2 would be violated.

We do not establish what distribution $\hat{W}_g$ converges to.[7] All that is relevant here is that, by construction, the limiting random value of $\hat{W}_g$ is bounded in probability, ensuring that $\hat{W}_f \hat{W}_g \to_p 0$.

Proof of Theorem 2:

Recall equation (1) and rewrite as follows

$$\begin{aligned} \widehat{\alpha} &= \hat{W}_f \hat{W}_g \widehat{\alpha}_h + \hat{W}_f \left(1 - \hat{W}_g\right) \widehat{\alpha}_g + (1 - \hat{W}_f) \widehat{\alpha}_f, \\ &= \alpha_0 + \hat{W}_f \hat{W}_g (\widehat{\alpha}_h - \alpha_0) + \hat{W}_f \left(1 - \hat{W}_g\right) (\widehat{\alpha}_g - \alpha_0) + (1 - \hat{W}_f)(\widehat{\alpha}_f - \alpha_0), \end{aligned}$$

$$\Rightarrow \sqrt{N}(\widehat{\alpha} - \alpha_0) = \hat{W}_f \hat{W}_g \sqrt{N}(\widehat{\alpha}_h - \alpha_0) + \hat{W}_f \left(1 - \hat{W}_g\right) \sqrt{N}(\widehat{\alpha}_g - \alpha_0) + (1 - \hat{W}_f) \sqrt{N}(\widehat{\alpha}_f - \alpha_0). \quad (27)$$

[7] If $\hat{Q}^g$ and $\hat{Q}^f$ happen to be independent, then $\hat{W}_g$ would be a ratio of independent Chi-squareds and so converges to a beta distribution with shape parameters $k_g/2$ and $k_k/2$. But there is no reason to impose that these distributions be independent.

Now we will show the asymptotic normality of $\widehat{\alpha}$ and the form of $\widetilde{V}$ depending on which model is correctly specified.

Case 1) Suppose G is correctly specified, but H is not. By Theorem 1, $(\widehat{\alpha}_g, \widehat{\beta}_g) \to_p (\alpha_0, \beta_0)$ and $\hat{Q}^g(\widehat{\alpha}_g, \widehat{\beta}_g) \to_p c_g' \Omega_g c_g / k_g = 0$, while $p\lim(\widehat{\alpha}_g, \widehat{\gamma}_h)$ is not (α_0, γ_0) but (α_h, γ_h) (the pseudo-true value by A4) so that we have $\hat{Q}^h(\widehat{\alpha}_h, \widehat{\gamma}_h) \to_p c_h' \Omega_h c_h / k_h \neq 0$. By following the same argument as given in Theorem 3.4 of Newey and McFadden, under A7, A9, A10 and A11, the central limit theorem yields $\frac{1}{\sqrt{N}} \sum_i G(Z_i, \alpha_0, \beta_0) \to^d N(0, \Sigma_g)$ where $\Sigma_g = E\{G(Z, \alpha_0, \beta_0) G(Z, \alpha_0, \beta_0)'\}$. Along with $\widehat{g}(\widehat{\alpha}, \widehat{\beta}_g) \to_p g_0(\theta_0^g) = 0$ and $\nabla_\alpha \widehat{g}(\widehat{\alpha}, \widehat{\beta}_g) \to_p \nabla_\alpha g_0(\theta_0^g)$, we can establish asymptotic normality of $\sqrt{N}(\widehat{\alpha}_g - \alpha_0)$. Therefore, by the asymptotic normality of $\widehat{\alpha}_g$, Lemma 1 on $\hat{W}_f \to_p 1$ and $\hat{W}_g \hat{W}_f \to_p 0$, and the continuous mapping theorem,

$$\sqrt{N}(\widehat{\alpha} - \alpha_0) \to^d N(0, \widetilde{V}^g),$$

and by A8

$$\frac{1}{N} \sum_i \widehat{\eta}_i^g \widehat{\eta}_i^{g\prime} \to_p \widetilde{V}^g \equiv E(\eta^g \eta^{g\prime}) \quad \text{where} \quad \frac{1}{\sqrt{N}} \sum_i \widehat{\eta}_i^g = \sqrt{N}(\widehat{\alpha}_g - \alpha_0).$$

Here $\widehat{\eta}_i^g$ is the influence function of the first-stage estimate $\widehat{\alpha}_g$ (the expression for $\widehat{\eta}_i^g$ is given below), making $\widetilde{V}^g = \widetilde{V}$ be the asymptotic variance of $\widehat{\alpha}_g$.

Case 2) Suppose H is correctly specified, but G is not. Then the same argument as Case 1 applies, replacing $\hat{W}_g$ with $1 - \hat{W}_g$, and switching the roles of β and γ, and switching the roles of g and h.

Case 3) Suppose G and H are both correctly specified. By following the same argument as given in Theorem 3.4 of Newey and McFadden, under A7, A9, A10, A11 and the consistency of $\{\widehat{\alpha}_f, \widehat{\beta}_f, \widehat{\gamma}_f\}$, the asymptotic normality of $\sqrt{N}(\widehat{\alpha}_f - \alpha_0)$ is established. Therefore, by Lemma 1 on $\hat{W}_f \to_p 0$ and $\hat{W}_g \hat{W}_f \to_p 0$, and the asymptotic normality of $\sqrt{N}(\widehat{\alpha}_f - \alpha_0)$,

$$\sqrt{N}(\widehat{\alpha} - \alpha_0) \to^d N(0, \widetilde{V}^f),$$

and by A8

$$\frac{1}{N}\sum_i \widehat{\eta}_i^f \widehat{\eta}_i^{f\prime} \to_p \widetilde{V}^f \equiv E(\eta^f \eta^{f\prime}) \text{ where } \frac{1}{\sqrt{N}}\sum_i \widehat{\eta}_i^f = \sqrt{N}(\widehat{\alpha}_f - \alpha_0).$$

Here $\widehat{\eta}_i^f$ is the influence function of the GMM estimate $\widehat{\alpha}_f$, making $\widetilde{V}^f = \widetilde{V}$ be the asymptotic variance of $\widehat{\alpha}_f$. Q.E.D.

Derivation of the form of $\widehat{\eta}_i^g$ and $\widehat{\eta}_i^h$:

To find the influence functions $\widehat{\eta}_i^g$, let $\widehat{\theta}^g$ denote the first-stage estimator

$$\widehat{\theta}^g \equiv (\widehat{\alpha}_g, \widehat{\beta}_g) = \arg \min_{\{\alpha,\beta\}\in\Theta_\alpha\times\Theta_\beta} \widetilde{Q}^g(\alpha,\beta) = \widehat{g}(\alpha,\beta)\hat{\Omega}_g\widehat{g}(\alpha,\beta).$$

By A7, A10-12 with probability approaching one the following first-order conditions in the first-stage for $\widehat{\theta}^g$ are satisfied

$$FD_\alpha^g = \frac{\partial\widetilde{Q}^g(\widehat{\theta}^g)}{\partial\alpha} = \{\nabla_\alpha\widehat{g}(\widehat{\theta}^g)\}'\hat{\Omega}_g\widehat{g}(\widehat{\theta}^g) = 0,$$
$$FD_\beta^g = \frac{\partial\widetilde{Q}^g(\widehat{\theta}^g)}{\partial\beta} = \{\nabla_\beta\widehat{g}(\widehat{\theta}^g)\}'\hat{\Omega}_g\widehat{g}(\widehat{\theta}^g) = 0.$$

Expending $\widehat{g}$ around the unique minimizer $\theta^g \equiv \{\alpha_g, \beta_g\}$ to get

$$\widehat{g}(\widehat{\theta}^g) = \widehat{g}(\theta^g) + \{\nabla_\alpha\widehat{g}(\overline{\theta}^g)\}(\widehat{\alpha}_g - \alpha_g) + \{\nabla_\beta\widehat{g}(\overline{\theta}^g)\}(\widehat{\beta} - \beta_g)$$

where $\overline{\theta}^g$ is a value from the mean value theorem. Substitute these into each FD^g to get

$$FD_\alpha^g = \{\nabla_\alpha\widehat{g}(\widehat{\theta}^g)\}'\hat{\Omega}_g[\widehat{g}(\theta^g) + \{\nabla_\alpha\widehat{g}(\overline{\theta}^g)\}(\widehat{\alpha}_g - \alpha_g) + \{\nabla_\beta\widehat{g}(\overline{\theta}^g)\}(\widehat{\beta}_g - \beta_g)]$$
$$FD_\beta^g = \{\nabla_\beta\widehat{g}(\widehat{\theta}^g)\}'\hat{\Omega}_g[\widehat{g}(\theta^g) + \{\nabla_\alpha\widehat{g}(\overline{\theta}^g)\}(\widehat{\alpha}_g - \alpha_g) + \{\nabla_\beta\widehat{g}(\overline{\theta}^g)\}(\widehat{\beta}_g - \beta_g)]$$
$$FD^g = \{FD_\alpha^g, FD_\beta^g\} = \widehat{I}^g + \widehat{H}^g(\widehat{\theta}^g - \theta^g)$$
$$\Longrightarrow \sqrt{N}(\widehat{\theta}^g - \theta^g) = \widehat{H}^{g-1}\sqrt{N}\widehat{I}^g$$

where $\widehat{I}^g \equiv \begin{bmatrix}\{\nabla_\alpha\widehat{g}(\widehat{\theta}^g)\}'\hat{\Omega}_g\widehat{g}(\theta^g)\\ \{\nabla_\beta\widehat{g}(\widehat{\theta}^g)\}'\hat{\Omega}_g\widehat{g}(\theta^g)\end{bmatrix}$,

$$\widehat{H}^g \equiv \begin{bmatrix} \{\nabla_\alpha\widehat{g}(\widehat{\theta}^g)\}'\hat{\Omega}_g\{\nabla_\alpha\widehat{g}(\overline{\theta}^g)\} & \{\nabla_\alpha\widehat{g}(\widehat{\theta}^g)\}'\hat{\Omega}_g\{\nabla_\beta\widehat{g}(\overline{\theta}^g)\} \\ \{\nabla_\beta\widehat{g}(\widehat{\theta}^g)\}'\hat{\Omega}_g\{\nabla_\alpha\widehat{g}(\overline{\theta}^g)\} & \{\nabla_\beta\widehat{g}(\widehat{\theta}^g)\}'\hat{\Omega}_g\{\nabla_\beta\widehat{g}(\overline{\theta}^g)\} \end{bmatrix}.$$

In this expression for$\sqrt{n}(\widehat{\theta}^g-\theta^g)$, examine the part for $\sqrt{n}(\widehat{\alpha}_g-\alpha_g)$, i.e., the first $k_\alpha\times 1$ components:

$$\sqrt{n}(\widehat{\alpha}_g-\alpha_g)=\widehat{A}_g^{-1}\hat{\Gamma}_g\sqrt{n}\widehat{g}(\theta^g)$$

where

$$\begin{aligned}\widehat{A}_g\equiv&\{\nabla_\alpha\widehat{g}(\widehat{\theta}^g)\}'\hat{\Omega}_g\{\nabla_\alpha\widehat{g}(\overline{\theta}^g)\}\\&+\{\nabla_\alpha\widehat{g}(\widehat{\theta}^g)\}'\hat{\Omega}_g\{\nabla_\beta\widehat{g}(\overline{\theta}^g)\}[\{\nabla_\beta\widehat{g}(\widehat{\theta}^g)\}'\hat{\Omega}_g\{\nabla_\beta\widehat{g}_n(\overline{\theta}^g)\}]^{-1}\{\nabla_\beta\widehat{g}(\widehat{\theta}^g)\}'\hat{\Omega}_g\{\nabla_\alpha\widehat{g}(\overline{\theta}^g)\}\end{aligned}$$

and

$$\hat{\Gamma}_g\equiv[\{\nabla_\alpha\widehat{g}(\widehat{\theta}^g)\}'\hat{\Omega}_g-\{\nabla_\alpha\widehat{g}(\widehat{\theta}^g)\}'\hat{\Omega}_g\{\nabla_\beta\widehat{g}(\overline{\theta}^g)\}[\{\nabla_\beta\widehat{g}(\widehat{\theta}^g)\}'\hat{\Omega}_g\{\nabla_\beta\widehat{g}(\overline{\theta}^g)\}]^{-1}\{\nabla_\beta\widehat{g}(\widehat{\theta}^g)\}'\hat{\Omega}_g].$$

Then, we have

$$\sqrt{N}(\widehat{\alpha}_g-\alpha_g)=\frac{1}{\sqrt{N}}\sum_i\widehat{\eta}_i^g,$$

where

$$\widehat{\eta}_i^g\equiv\widehat{A}_g^{-1}\hat{\Gamma}_gG(Z_i,\theta^g),$$

and $\widehat{\eta}_i^g$ is the influence function of the first-stage estimate $\widehat{\alpha}_g$. If G is correct, θ^g is replaced by θ_0^g.

Analogously for the influence functions $\widehat{\eta}_i^h$, let $\widehat{\theta}^h$ denote the second-stage estimator

$$\widehat{\theta}^h\equiv(\widehat{\alpha}_h,\widehat{\gamma}_h)=\arg\min\widetilde{Q}^h(\alpha,\gamma)=h(\alpha,\gamma)\hat{\Omega}_hh(\alpha,\gamma).$$

By A7-A9 with probability approaching one the following first-order conditions in the second-stage for $\widehat{\theta}^h$ are satisfied

$$\begin{aligned}FD_\alpha^h=\frac{\partial\widetilde{Q}^h(\widehat{\theta}^h)}{\partial\alpha}=\{\nabla_\alpha\widehat{h}(\widehat{\theta}^g)\}'\hat{\Omega}_h\widehat{h}(\widehat{\theta}^g)=0,\\FD_\gamma^h=\frac{\partial\widetilde{Q}^h(\widehat{\theta}^h)}{\partial\gamma}=\{\nabla_\gamma\widehat{h}(\widehat{\theta}^h)\}'\hat{\Omega}_h\widehat{h}(\widehat{\theta}^h)=0.\end{aligned}$$

Expending $\widehat{h}$ around the unique minimizer $\theta^h \equiv \{\alpha_h, \gamma_h\}$ and substitute it into each FD^h to get

$$
\begin{aligned}
FD^h_\alpha &= \{\nabla_\alpha \widehat{h}(\widehat{\theta}^g)\}' \hat{\Omega}_h [\widehat{h}(\theta^h) + \{\nabla_\alpha \widehat{h}(\overline{\theta}^h)\}(\widehat{\alpha}_h - \alpha_h) + \{\nabla_\gamma \widehat{h}(\overline{\theta}^h)\}(\widehat{\gamma} - \gamma_h)] \\
FD^h_\gamma &= \{\nabla_\gamma \widehat{h}(\widehat{\theta}^h)\}' \hat{\Omega}_h [\widehat{h}(\theta^h) + \{\nabla_\alpha \widehat{h}(\overline{\theta}^h)\}(\widehat{\alpha}_h - \alpha_h) + \{\nabla_\gamma \widehat{h}(\overline{\theta}^h)\}(\widehat{\gamma} - \gamma_h)] \\
FD^h &= \{FD^h_\alpha, FD^h_\gamma\} = I^h_n + \widehat{H}^h_n(\widehat{\theta}^h - \theta^h) \\
&\Longrightarrow \sqrt{N}(\widehat{\theta}^h - \theta^h) = \widehat{H}^{h-1}_n \sqrt{N} \widehat{I}^h \\
\text{where } \widehat{I}^h &\equiv \begin{bmatrix} \{\nabla_\alpha \widehat{h}(\widehat{\theta}^h)\}' \hat{\Omega}_h \widehat{h}(\theta^h) \\ \{\nabla_\gamma \widehat{h}(\widehat{\theta}^h)\}' \hat{\Omega}_h \widehat{h}(\theta^h) \end{bmatrix}, \\
\widehat{H}^h &\equiv \begin{bmatrix} \{\nabla_\alpha \widehat{h}(\widehat{\theta}^h)\}' \hat{\Omega}_h \{\nabla_\alpha \widehat{h}(\overline{\theta}^h)\} & \{\nabla_\alpha \widehat{h}(\widehat{\theta}^h)\}' \hat{\Omega}_h \{\nabla_\gamma \widehat{h}(\overline{\theta}^h)\} \\ \{\nabla_\gamma \widehat{h}(\widehat{\theta}^h)\}' \hat{\Omega}_h \{\nabla_\alpha \widehat{h}(\overline{\theta}^h)\} & \{\nabla_\gamma \widehat{h}(\widehat{\theta}^h)\}' \hat{\Omega}_h \{\nabla_\gamma \widehat{h}(\overline{\theta}^h)\} \end{bmatrix}.
\end{aligned}
$$

In this expression for$\sqrt{n}(\widehat{\theta}^h - \theta^h)$, examine the part for $\sqrt{n}(\widehat{\alpha}_h - \alpha_h)$, i.e., the first $k_\alpha \times 1$ components:

$$\sqrt{n}(\widehat{\alpha}_h - \alpha_h) = \widehat{A}^{-1}_h \hat{\Gamma}_h \sqrt{n}\widehat{h}(\theta^h)$$

where

$$
\begin{aligned}
\widehat{A}^{-1}_h \equiv\ & \{\nabla_\alpha \widehat{h}(\widehat{\theta}^h)\}' \hat{\Omega}_h \{\nabla_\alpha \widehat{h}(\overline{\theta}^h)\} \\
& + \{\nabla_\alpha \widehat{h}(\widehat{\theta}^h)\}' \hat{\Omega}_h \{\nabla_\gamma \widehat{h}(\overline{\theta}^h)\} [\{\nabla_\gamma \widehat{h}(\widehat{\theta}^h)\}' \hat{\Omega}_h \{\nabla_\gamma \widehat{h}(\overline{\theta}^h)\}]^{-1} \{\nabla_\gamma \widehat{h}(\widehat{\theta}^h)\}' \hat{\Omega}_h \{\nabla_\alpha \widehat{h}(\overline{\theta}^h)\}
\end{aligned}
$$

and

$$\hat{\Gamma}_h \equiv [\{\nabla_\alpha \widehat{h}(\widehat{\theta}^h)\}' \hat{\Omega}_h - \{\nabla_\alpha \widehat{h}(\widehat{\theta}^h)\}' \hat{\Omega}_h \{\nabla_\gamma \widehat{h}(\overline{\theta}^h)\} [\{\nabla_\gamma \widehat{h}(\widehat{\theta}^h)\}' \hat{\Omega}_h \{\nabla_\gamma \widehat{h}(\overline{\theta}^h)\}]^{-1} \{\nabla_\gamma \widehat{h}(\widehat{\theta}^h)\}' \hat{\Omega}_h].$$

Then, we have

$$\sqrt{n}(\widehat{\alpha}_h - \alpha_h) = \frac{1}{\sqrt{N}} \sum_i \widehat{\eta}^h_i,$$

where

$$\widehat{\eta}^h_i \equiv \widehat{A}^{-1}_h \hat{\Gamma}_h H(Z_i, \theta^h),$$

and $\widehat{\eta}^h_i$ is the influence function of the first-stage estimate $\widehat{\alpha}_h$. If H is correct, θ^h is replaced by θ^h_0.

Analogously, we can obtain the influence function $\widehat{\eta}^f_i$ switching the roles of g or h to f.