Over-Identified Doubly Robust Identification and Estimation[*]

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

Original 2018, Revised January 2022

Abstract

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 . An estimator is called Doubly Robust (DR) if is consistent no matter which model is correct. 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 is a popula in the ATE application). However, we do not advise using our ODR for applications where DR methods construct weights to yield the DR consistency property and for relative efficiency.

Analogous to $_g$, let $_h$ denote the estimator of based on the moments $E[H(Z; _0; _0)] = 0$, so $_h$ and $_h$ minimize a quadratic GMM objective function $Q^h(;)$, and are asymptotically e¢ cient if model H is true and model G is not true. Finally, let $_f$ (f)Tj lly, let

If $\boldsymbol{G}$ is correctly specified, so $\boldsymbol{g}_0(\ _0; \ _0) = 0$, then there does not exist any $\{\ ;\ \} \in$ with $\{\ ;\ \} \neq \{\ _0$

model $\boldsymbol{H}$. In our applications, we likewise use the standard e¢ cient two step GMM method for estimating the matrices $\hat{}_g$ and $\hat{}_h$.

De…ne $Q_0^g(\ ;\ )$ and $Q_0^h(\ ;\ )$ by

$$Q_0^g(\ ;\ ) \quad g_0(\ ;\ )^0 \ {}_g g_0(\ ;\ ) \quad \text{and} \quad Q_0^h(\ ;\ ) \quad h_0(\ ;\ )^0 \ {}_h h_0(\ ;\ )$$

for given positive de…nite matrices $_g$ and $_h$, where $\hat{}_g \; ! \;^p \; _g$ and $\hat{}_h \; ! \;^p \; _h$.

Assumption A4: Assume there exists f $_g(\ _g); \ _g(\ _g)$g 2 such that $Q_0^g(\ _g(\ _g); \ _g(\ _g)) < Q_0^g(\ ;\ )$ for all f $;$ g 2 nf $_g(\ _g); \ _g(\ _g)$g and there exists f $_h(\ _h); \ _h(\ _h)$g 2 such that $Q_0^h(\ _h(\ _h); \ _h(\ _h)) < Q_0^h(\ ;\ )$ for all f $;$ g 2 nf $_h(\ _h); \ _h(\ _h)$g.

Assumption A4 says that, for each of the models $\boldsymbol{G}$ and $\boldsymbol{H}$, there exists a unique value of the parameters that minimizes the limiting value of the GMM objective function. Given Assumptions A2 and A3, Assumption A4 will automatically be satis…ed for model $\boldsymbol{G}$ when $\boldsymbol{G}$ is correctly speci…ed, with f $_g(\ _g); \ _g(\ _g)$g = f $_0; \ _0$g, and similarly for f $_h(\ _h); \ _h(\ _h)$g when $\boldsymbol{H}$ is correctly speci…ed, by Lemma 2.3 of Newey and McFadden (1994). That is, for correctly speci…ed models, the minimizing value is the true value.

The dependence of , , and on the weighting matrices $_g$ and $_h$ in Assumption A4 re‡ects the fact that, when model $\boldsymbol{G}$ or $\boldsymbol{H}$ is incorrectly speci…ed, the parameter values that minimize the GMM criterion functions $Q_0^g(\ ;\ )$ and $Q_0^h(\ ;\ )$ may depend on the choice of weighting matrices $_g$ and $_h$. To save notation, we will omit this dependence when $_g$ and $_h$ are the standard e¢ cient two step GMM weighting matrices. We have similarly dropped the dependence of $Q_0^g(\ ;\ )$ and $Q_0^h(\ ;\ )$ on $_g$ and $_h$ to save notation.

Together with our other Assumptions, Assumption A4 implies that GMM estimators of $\boldsymbol{G}$ or $\boldsymbol{H}$ will also converge to some (pseudo-true) values when they are misspeci…ed. Consider, e.g., applying the standard then be f $_g($ $_g);$ $_g($ $_g)$g, based on this construction of while if H is correctly speci…ed and and G is not, then

$$\hat{W}_g \; ! \; {}^p \; \frac{c_g^0 \quad {}_g c_g = k_g}{c_g^0 \quad {}_g c_g = k_g + 0} = 1:$$

Before getting to our ODR estimator given by equation (1), consider the simpler estimator e de…ned by

$$= \hat{W}_g \quad {}_h + \quad 1 \quad \hat{W}_g \quad {}_g: \qquad (4)$$

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

The intuition behind is straightforward (the asymptotic statements in this paragraph are proved formally in the next section). Suppose model H $\hat{Q}$ b_h b_g

Although the SODR has the desired DR property, it also 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 . Second, when both G and H are correct, may be ineﬃcient, relative to a GMM estimator that eﬃciently combines the moments from both models.

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

$$\{ \; {}_f; \; {}_f; \; {}_f \} = \arg \min_{\{ \; ; \; ; \; \} 2}$$

as shown earlier has the same limiting value as either $_g$ or $_h$, depending on which is correctly speci…ed.

The estimator therefore, like , has the desired DR property. We show later that avoids the asymptotic issues has when both G and H are correctly speci…ed, and that generally performs better than in …nite samples. This is why consider di¤erent choices of in our applications. Overall, we found that the exponential where $U^0(C_t; X_t; \ )$ denotes $@U(C_t; X_t; \ )=@C_t$. If the functional form of U^0 is known, then this equation provides moments that allow b and to be estimated using GMM. But suppose we have two di¤erent possible speci…cations of U^0, and we do not know which speci…cation is correct. Then our ODR estimator can be immediately applied, replacing the expression in the inner parentheses in equation (7) with $G(Z; \ ; \ )$ or $H(Z; \ ; \ )$ to represent the two di¤erent speci…cations. Here would represent parameters that are the same in either speci…cation, including the subjective rate of time preference b.

To give a speci…c example, a standard speci…cation of utility is constant relative risk aversion with habit formation, where utility takes the form

$$U(C_t; X_t; \ ) = \frac{[C_t \quad M(X_t)]^1 \qquad 1}{1}$$

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

To illustrate, suppose that with internal habits the function $M(X_t)$ would be given by $G(X_t; \ )$, where G is the internal habits functional form. Similarly, suppose with external habits $M(X_t)$ would be given by $H(X_t; \ )$ where H is the external habits speci…cation. Then, based on equation (7), we could de…ne $G(Z; \ ; \ )$ and $H(Z; \ ; \ )$ by

$$G(Z; \ ; \ ) = \quad bR_{t+1} \frac{C_{t+1} \quad G(X_{t+1}; \ )}{}$$

@bR)g798234 T Td(b)27. /TT1 10.909 Tf 21.207 125.878 Td6 (Tj /TT3 10.909 Tf 9.954 0 Td (()Tj /TT1 10.909

1Tj /T /TT2 7.97 Tf 6.516 -2.715 Td (900)Tj /TT4 7.97 Tf ( )Tj ET q 1 97 9 Tf 16.801 Td 8611.6615 0

$(\,;\,)$. This would generally be the case, because the potential information set of consumers at time t is large relative the the number of parameters in the model.

3.2 Alternative Sets of Instruments

Consider a parametric model

$$Y = M(W; \;) +$$

where Y is an outcome, W is a vector of observed covariates, M is a known functional form, is a vector of parameters to be estimated, and is an unobserved error term. The errors may be correlated with W, so to estimate the model we wish to …nd instruments that are uncorrelated with . Let R and Q denote

0 where $\boldsymbol{G}(\boldsymbol{Z};\ )$ is given by the stacked vectors

$$\boldsymbol{G}(\boldsymbol{Z};\ ) = \begin{array}{c} \boldsymbol{X}\ \ \boldsymbol{Y}\ \ \boldsymbol{X}'\ {}_{x}\ \ \boldsymbol{S}\ {}_{s} \\ \boldsymbol{L}\ \ \boldsymbol{Y}\ \ \boldsymbol{X}'\ {}_{x}\ \ \boldsymbol{S}\ {}_{s} \end{array} . \tag{8}$$

The main difficulty with applying this two stage least squares or GMM estimator is that one must find one or more covariates $\boldsymbol{L}$ to serve as instruments.

Lewbel (2012) proposes an alternative estimator that, rather than requiring that one find instruments $\boldsymbol{L}$, instead constructs instruments based on assumptions regarding heteroscedasticity. This estimator consists of first linearly regressing $\boldsymbol{S}$ on $\boldsymbol{X}$, and obtaining the residuals from that regression. Then a vector of instruments $\boldsymbol{P}$ is constructed by setting $\boldsymbol{P}$ equal to demeaned $\boldsymbol{X}$ (excluding the constant) times these residuals. This constructed vector $\boldsymbol{P}$ is then used instead of $\boldsymbol{L}$ above as instruments.[8] As shown in Lewbel (2012), one set of conditions under which the vector $\boldsymbol{P}$ can be a valid set of instruments is when the endogeneity in $\boldsymbol{S}$ is due to classical measurement error in $\boldsymbol{S}$.

Let $\boldsymbol{X_c}$ denote the vector $\boldsymbol{X}$ with the constant removed. Algebraically, we can write the instruments obtained in this way as $\boldsymbol{R} = \{\boldsymbol{X};\boldsymbol{P}\}$ where $\boldsymbol{P} = (\boldsymbol{X_c}\quad {}_1)(\boldsymbol{S}$

=X

$\boldsymbol{X}$ is a vector of covariates that affect the consumer's tastes, and $\boldsymbol{S}$ 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 $\boldsymbol{S}$ is observed with some measurement error. To deal with this budget measurement error, a commonly employed set of instruments $\boldsymbol{L}$ consists 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, and this behavioral assumption may or may not be valid. It is therefore useful to consider the alternative set of potential instruments $\boldsymbol{P}$ defined above. Use of $\boldsymbol{P}$ does not require finding covariates from outside the model, like income, to use as instruments, but does require that certain measurement error assumptions hold. Our later empirical application applies ODR to this application, thereby obtaining consistent estimates of if either $\boldsymbol{L}$ or $\boldsymbol{P}$ are valid instruments.

4 The ODR Estimator Asymptotics

In this section we show consistency of our ODR estimator , and then derive its limiting distribution, which is root n consistent and asymptotically normal. We make the following additional assumptions. What these assumptions mostly do is ensure that GMM estimates of models $\boldsymbol{G}$, $\boldsymbol{H}$, and $\boldsymbol{F}$ are each asymptotically normal around the true values when correctly specified, and are suitably bounded in probability around the pseudo-true values when misspecified. We do not require asymptotic normality under misspecification.

Assumption A5: $\boldsymbol{G}(\boldsymbol{Z};\ ;\ )$, $\boldsymbol{H}(\boldsymbol{Z};\ ;\ )$ and $\boldsymbol{F}(\boldsymbol{Z};\ ;\ ;\ )$ are continuous at f ; g 2 , f ; g 2 , and f ; ; g 2 respectively, with probability one.

Assumptionband35.9(p)-26(b)[(g)]729 Tf 38( )pb ;

f $_{h};$ $_{h}$g; and ^{f} f $_{f};$ $_{f};$ $_{f}$g. If the models $\boldsymbol{G}$ and $\boldsymbol{H}$ are correctly speci…ed, $^{g}_{0}$ = ^{g}, $^{h}_{0}$ = ^{h}, and $^{f}_{0}$ = ^{f}.

Assumption A7: With probability one, $\boldsymbol{G}(\boldsymbol{Z};\ ;\ )$, $\boldsymbol{H}(\boldsymbol{Z};\ ;\ )$, and $\boldsymbol{F}(\boldsymbol{Z};\ ;\ ;\ )$ are twice continuously di¤erentiable in a neighborhood @^{g} of ^{g}, @^{h} of ^{h}, and @^{f} of ^{f}, respectively.

Assumption A8: $\boldsymbol{H}_{g}(\ ^{g}_{0})$ r $\boldsymbol{g}_{0}(\ ^{g}_{0})$ $_{g}$r

4.1 ODR Consistency

Lemma 1 : Suppose Assumptions A1 to A15 hold. Then, for any with $0 <\quad < 1$, ^

Case 1) Suppose both $g_0(\ _0, \ _0) = 0$ and $h_0(\ _0, \ _0) = 0$. Then $\{\ _g, \ _g\} \xrightarrow{p} \{\ _0, \ _0\}$, $\{\ _h, \ _h\} \xrightarrow{p} \{\ _0, \ _0\}$, and $\{\ _f, \ _f, \ _f\} \xrightarrow{p} \{\ _0, \ _0, \ _0\}$, so $\hat{Q}^g \xrightarrow{p} 0, \hat{Q}^h \xrightarrow{p} 0$, and $\hat{Q}^f \xrightarrow{p} 0$. By Lemma 1, $\hat{W}_f$ and $\hat{W}_f \hat{W}_g$

The first part of Theorem 2 states that the ODR estimator is root n consistent and asymptotically normal, while the second part gives a consistent estimator for the limiting variance of . The proof of Theorem 2 is given in the Supplemental Appendix. The basic structure of the proof follows Newey and McFadden (1994) for multistep parametric estimators.

Note that while consistency only requires $0 < \; < 1$, Theorem 2 assumes $> 1/2$ to ensure $\sqrt{n}$-consistency of . This condition is only required for the case where $_g \neq \, _h$.

The estimator of V given in equation (10) does not require knowing which of the models G or H is correct. Nevertheless, as shown in the Supplemental Appendix, V will either equal a matrix V^g or V^h or V^f, depending on whether models G, H, or both are correctly specified.

A fact that complicates the derivation of Theorem 2 is that ^h_i does not consistently estimate the influence function of $_h$ if model H is not correctly specified. Similarly, ^g_i is not consistent if model G is misspecified, and ^f_i is not consistent if either G or H is misspecified. However, it turns out that to estimate the limiting variance of , we do not need to consistently estimate the influence function of any incorrectly specified GMM. For example, in the limiting variance formula for , the function ^h_i is multiplied by $\hat{W}$

) Tj /TT5 7.97 Tf 10.303 1.793 Td ... Tf -3.281 -2.758 Td (W) Tj /TT5 7.97 Tf 10

weights $\hat{W}_g$ and $\hat{W}_f$). It is therefore numerically desirable in …nite samples to have these matrices be estimators, the numerator of the weight on model $\boldsymbol{H}$ depends on the criterion for model $\boldsymbol{G}$ (i.e., on $\boldsymbol{Q^g}$) designed to put all weight on model $\boldsymbol{H}$ when model $\boldsymbol{G}$ is wrong but $\boldsymbol{H}$ is correct, and vice versa.

A di¤erence between $^{\boldsymbol{MG}}$ and SODR (witp SODRv26.996(76onT2 .996(76tia1 10.909 4f 14.161 0 Tdd (2713)Tj that vary by correlations $_{Rj}$ and $_{Qj}$. The first design takes $_{Rj} = \ _{Qj} = 0$, which makes both models right (both sets of instruments are valid). The second takes $_{R1} = \ _{R2} = 0$, $_{Q1} = 0.4$, and $_{Q2} = 0.6$, which makes model G 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). The third takes $_{R1} = 0.4$, $_{R2} = 0.6$ and $_{Q1} = \ _{Q2} = 0$, which makes model H right and model G wrong.

For the tuning function discussed in sections 2.3 and 4.4, we consider two different choices, $_1\ n\hat{Q} = \exp\ n\hat{Q}\quad 1$ and $_2(n\hat{Q}) = (n\hat{Q})^2$ so the weighting functions $\hat{W}_g$ and $\hat{W}_f$ are

$$_1 : \hat{W}_g = \frac{\exp\{n\hat{Q}^g(\ _g;\ _g)\}\quad 1}{\exp\{n\hat{Q}^g(\ _g;\ _g)\} + \exp\{n\hat{Q}^h(\ _h;\ _h)\}\quad 2},\ \hat{W}_f = 1\quad \frac{1}{\exp\{n\quad \hat{Q}^f(\ _f;\ _f;\ _f)\}}; \tag{12}$$

$$_2 : \hat{W}_g = \frac{\{n\hat{Q}^g(\ _g;\ _g)\}^2}{\{n\hat{Q}^g(\ _g;\ _g)\}^2 + \{n\hat{Q}^h(\ _h;\ _h)\}^2},\ \hat{W}_f = 1\quad \frac{1}{\{n\quad \hat{Q}^f(\ _f;\ _f;\ _f)\}^2 + 1}. \tag{13}$$

For the tuning parameter , we use $= 1\quad p$, where p is the p-value of the Wald statistic as discussed in section 2.3.

We report eight estimates of $_1$ and $_0$ 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

We report skewness (Skew) and kurtosis (Kurt) of these t-statistics across simulations, and the frequency (Freq) that these t-statistics are less than 2 in magnitude, corresponding to the frequency with which a 2 estimated standard error con…dence interval contains the true parameter value. Also, to check the accuracy of the standard error estimates, we report the average of the estimated standard errors (SE), and standard deviation of the estimated standard errors (SD_{SE}), across the simulations. The last …ve summary statistics are not reported for $SODR$, because we do not consider its limiting distribution due to the random probability limit of $\hat{W}_g$.

When both sets of instruments are valid, ODR estimates are almost as precise as GMM_f, and when either set of instruments is invalid, ODR estimates are more precise than inconsistent GMM estimators. The $SODR$ estimates are found to be less e¢ cient than ODR when both G and H models are valid (as expected), but when one model is invalid, $SODR$ is similar to ODR. In this application, the cost in e¢ ciency of choosing the simp7true9 T.ations, OD (aslargJ -45-33in-45-33aba [illegible]

Table 1. Simulation Results of $_1$ (

Table 2. Simulation Results of $_{0}$ ($\boldsymbol{n}$ = 100)

	Bias	Mde	RMSE	MAE	SD	Skew	Kurt	Freq	SE	$\boldsymbol{SD}_{\mathrm{SE}}$
Both correct										
$\boldsymbol{GMM}_{\mathrm{g}}$	-0.0038	-0.0048	0.0112	0.0687	0.1058	0.0005	3.1738	0.9415	0.1009	0.0089
$\boldsymbol{GMM}_{\mathrm{h}}$	-0.0024	-0.0090	0.0134	0.0757	0.1157	-0.0131	2.9788	0.9490	0.1115	0.0182
$\boldsymbol{GMM}_{\mathrm{f}}$	-0.0046	-0.0073	0.0113	0.0688	0.1063	0.0212	3.1124	0.9350	0.0981	0.0085
$\boldsymbol{MG}$	-0.0022	-0.0063	0.0115	0.0697	0.1071	0.0291	3.0642	0.9440	0.1022	0.0110
$\boldsymbol{ODR}_{1}$	-0.0039	-0.0062	0.0113	0.0686	0.1063	0.0583	3.0524	0.9370	0.0989	0.0092
$\boldsymbol{ODR}_{2}$	0.0001	-0.0017	0.0105	0.0687	0.1025	-0.0532	3.1744	0.9525	0.0990	0.0088
$\boldsymbol{SODR}_{1}$	-0.0016	-0.0067	0.0120	0.0703	0.1097					
$\boldsymbol{SODR}_{2}$	0.0014	0.0023	0.0108	0.0707	0.1041					
G correct										
$\boldsymbol{GMM}_{\mathrm{g}}$	-0.0038	-0.0060	0.0112	0.0683	0.1060	-0.0390	3.1287	0.9395	0.1009	0.0108
$\boldsymbol{GMM}_{\mathrm{h}}$	-0.2005	-0.1977	0.0554	0.1977	0.1234	0.1485	3.0509	0.5750	0.1103	0.0179
$\boldsymbol{GMM}_{\mathrm{f}}$	-0.0744	-0.0737	0.0219	0.0999	0.1280	-0.0354	3.1266	0.7540	0.0867	0.0074
$\boldsymbol{MG}$	-0.0401	-0.0396	0.0140	0.0774	0.1115	-0.1154	3.1855	0.8885	0.0954	0.0109
$\boldsymbol{ODR}_{1}$	-0.0258	-0.0198	0.0147	0.0722	0.1186	-0.2332	3.2476	0.9010	0.0996	0.0120
$\boldsymbol{ODR}_{2}$	-0.0245	-0.0198	0.0136	0.0744	0.1139	-0.2004	3.0110	0.9065	0.0995	0.0114
$\boldsymbol{SODR}_{1}$	-0.0258	-0.0198	0.0147	0.0722	0.1186					
$\boldsymbol{SODR}_{2}$	-0.0240	-0.0194	0.0136	0.0745	0.1142					
H correct										
$\boldsymbol{GMM}_{\mathrm{g}}$	-0.1151	-0.1166	0.0230	0.1198	0.0989	0.0139	2.8983	0.6735	0.0808	0.0069
$\boldsymbol{GMM}_{\mathrm{h}}$	-0.0028	-0.0088	0.0133	0.0722	0.1153	-0.2405	2.9748	0.9530	0.1123	0.0344
$\boldsymbol{GMM}_{\mathrm{f}}$	-0.0963	-0.0966	0.0203	0.1039	0.1050	-0.0085	2.9169	0.7095	0.0791	0.0068
$\boldsymbol{MG}$	-0.0035	-0.0094	0.0133	0.0720	0.1151	-0.2389	2.9660	0.9515	0.1120	0.0343
$\boldsymbol{ODR}_{1}$	-0.0051	-0.0105	0.0131	0.0725	0.1146	-0.2535	2.9609	0.9475	0.1109	0.0320
$\boldsymbol{ODR}_{2}$	-0.0084	-0.0187	0.0135	0.0753	0.1159	-0.1964	3.0290	0.9380	0.1095	0.0287
$\boldsymbol{SODR}_{1}$	-0.0029	-0.0089	0.0133	0.0722	0.1153					
$\boldsymbol{SODR}_{2}$	-0.0038	-0.0144	0.0138	0.0760	0.1176					

Table 4. Simulation Results of $_0$ ($\boldsymbol{n}$ = 500)

	Bias	Mde	RMSE	MAE	SD	Skew	Kurt	Freq	SE	$\boldsymbol{SD}_{\mathrm{SE}}$
Both correct										
$\boldsymbol{GMM}_g$	-0.0010	-0.0002	0.0021	0.0315	0.0458	-0.1391	2.9732	0.9565	0.0459	0.0018
$\boldsymbol{GMM}_h$	-0.0008	0.0005	0.0024	0.0328	0.0492	-0.1701	3.0631	0.9500	0.0491	0.0030
$\boldsymbol{GMM}_f$	-0.0011	0.0000	0.0021	0.0311	0.0458	-0.1335	2.9799	0.9550	0.0454	0.0017
$\boldsymbol{MG}$	-0.0007	0.0004	0.0021	0.0311	0.0463	-0.1527	3.0340	0.9570	0.0462	0.0021
$\boldsymbol{ODR}_1$	-0.0010	0.0000	0.0021	0.0310	0.0459	-0.1327	3.0063	0.9540	0.0455	0.0018
$\boldsymbol{ODR}_2$	0.0009	-0.0005	0.0022	0.0315	0.0471	0.0061	2.9664	0.9445	0.0455	0.0019
$\boldsymbol{SODR}_1$	-0.0005	0.0003	0.0022	0.0321	0.0468					
$\boldsymbol{SODR}_2$	0.0010	0.0003	0.0023	0.0334	0.0483					
G correct										
$\boldsymbol{GMM}_g$	-0.0010	-0.0003	0.0021	0.0314	0.0458	-0.1566	2.9735	0.9570	0.0459	0.0021
$\boldsymbol{GMM}_h$	-0.2000	-0.2000	0.0428	0.2000	0.0529	0.0663	3.1573	0.0225	0.0495	0.0033
$\boldsymbol{GMM}_f$	-0.0732	-0.0731	0.0084	0.0739	0.0554	0.0501	2.9813	0.5400	0.0402	0.0014
$\boldsymbol{MG}$	-0.0012	-0.0004	0.0021	0.0314	0.0458	-0.1553	2.9705	0.9570	0.0458	0.0021
$\boldsymbol{ODR}_1$	-0.0010	-0.0003	0.0021	0.0314	0.0458	-0.1563	2.9744	0.9570	0.0459	0.0021
$\boldsymbol{ODR}_2$	-0.0020	-0.0011	0.0021	0.0315	0.0459	-0.1685	2.9918	0.9550	0.0457	0.0021
$\boldsymbol{SODR}_1$	-0.0010	-0.0003	0.0021	0.0314	0.0458					
$\boldsymbol{SODR}_2$	-0.0020	-0.0011	0.0021	0.0315	0.0459					
H correct										
$\boldsymbol{GMM}_g$	-0.1122	-0.1121	0.0146	0.1121	0.0448	-0.0037	3.0575	0.1945	0.0367	0.0013
$\boldsymbol{GMM}_h$	-0.0007	-0.0007	0.0024	0.0329	0.0494	-0.2688	3.0914	0.9480	0.0492	0.0048
$\boldsymbol{GMM}_f$	-0.0938	-0.0948	0.0111	0.0948	0.0481	-0.0661	2.9792	0.3445	0.0366	0.0013
$\boldsymbol{MG}$	-0.0007	-0.0007	0.0024	0.0329	0.0494	-0.2688	3.0914	0.9480	0.0492	0.0048
$\boldsymbol{ODR}_1$	-0.0007	-0.0007	0.0024	0.0329	0.0494	-0.2688	3.0914	0.9480	0.0492	0.0048
$\boldsymbol{ODR}_2$	-0.0011	-0.0038	0.0025	0.0340	0.0500	-0.1804	2.9318	0.9555	0.0491	0.0049
$\boldsymbol{SODR}_1$	-0.0007	-0.0007	0.0024	0.0329	0.0494					
$\boldsymbol{SODR}_2$	-0.0011	-0.0037	0.0025	0.0340	0.0500					

GMM. This suggests a modest advantage of the exponential tuning function $_1$.

One should expect correctly specified GMM estimators to be more efficient than ODR, and that is indeed the case. But in many of the simulations, the loss in efficiency from using ODR is very low. In particular, when model G is invalid, so only the weaker instruments are valid, the precision of ODR is almost identical to that of the efficient GMM_h. So, using our ODR, there is little loss in efficiency from not knowing which specification is correct. In summary, we conclude that our proposed ODR works well, even at low sample sizes.

6 Empirical Application: Engel Curve Estimation

Here we empirically estimate the Engel curve example discussed in section 3.1. Y is the food budget share, S is log real total consumption expenditures, and X is a vector of other covariates that serve as controls[11]. 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. The vector L consists of two candidate external instrumental 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 3.2 above are valid. Model which are heteroscedasticity based constructed instruments, GMM_f is the GMM estimator that uses both sets of instruments, and $SODR$ and ODR are our new estimators given in equations (4) and (1) with the tuning functions $_1$ and $_2$.

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 di¤erence in strength can be seen in the standard errors of $\hat{}_s$, 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.

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

Table 5. Engel Curve Estimates [13]

	GMM_{g0}	GMM_g	GMM_h	GMM_f	$SODR_1$	ODR_1	$SODR_2$	ODR_2
$\hat{}_s$	-0.0859 (0.0198)	-0.0840 (0.0197)	-0.0521 (0.0546)	-0.0862 (0.0177)	-0.0812	-0.0862 (0.0192)	-0.0831	-0.0862 (0.0192)
$\hat{}_0$	0.336 (0.0122)	0.335 (0.0120)	0.317 (0.0328)	0.337 (0.0109)	0.333	0.337 (0.0118)	0.335	0.337 (0.0118)
^{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				
$\hat{W}_g; \hat{W}_f; p$					0.09, 0.004, 0.86		0.03, 0.000, 0.86	

[13] Table 5 notes: We report coe¢ cient estimates with associated standard errors in parentheses, except SODR. Also reported is ^{2}, the Hansen (1982) test statistics for overidenti…ed GMM, along with their degrees of freedom $d.f.$ and p-values. $\hat{Q}$ is the normalized minimand of the GMM estimators. The last row reports weights $\hat{W}_g$, $\hat{W}_f$, and gives p, which is the p-value of the Wald statistic testing the null hypothesis that $b_g = b_h$. This p is used to construct $= 1 \quad p$ in $\hat{W}_f$ in equation (5), as explained in section 2.3.

The estimated weight $\hat{w}_g$ is 0.09 with the tuning function $_1$ and 0.03 with $_2$, so $SODR$ puts over ten times as much weight on model G as on model H. However, in ODR the weight on model F, $1 \ \hat{w}_f$, is 0.996 with $_1$ and is one to three decimal places with $_2$. The very small difference in $\hat{w}_f$ between $_1$ and $_2$ is why both of the ODR estimates appear the same in Table 5 (they actually differ in the fourth significant digit: -0.08617 vs. -0.08619 for $\hat{\ }_s$).

The very high weight on model F strongly suggests that both models are likely to be correctly specified. This therefore implies that the difference between GMM_g and GMM_h is likely due to imprecision of GMM_h rather than misspecification of the constructed instruments in model H. Further evidence that both are all $s \geq 0$, and the third either converges to a constant or diverges depending on s (and sometimes γ) as discussed below.[14]

First suppose model G is locally misspecified with $s > 1/2$. Then $nQ^g(\hat{\theta}_g; \hat{\Omega}_g) \to^d \chi^2_{k_g}(0)$, which is the same limit as when G is correctly specified, and similarly for H. As a result, in this case the SODR and ODR estimators have the same $\sqrt{n}$ consistent, asymptotically normal limiting distribution as they have when G is correctly specified, and similarly for H. Note this means that instead of requiring that either G or H (or both) be correctly specified, it is sufficient to assume that either G or H (or both) are locally misspecified with $s > 1/2$, noting that correct specification is the special case of $s = 1$.

If model G is locally misspecified with $s < 1/2$, then $nQ^g(\hat{\theta}_g; \hat{\Omega}_g)$ diverges, and the SODR has the same $\sqrt{n}$ consistent, asymptotically normal limiting distribution as when G is globally misspecified. The ODR will also have the same limiting distribution as when G is globally misspecified, as long as the tuning parameter γ has $\gamma > s + 0.5$. This then guarantees that model G will asymptotically have zero weight. Since these cases are equivalent asymptotically to G being globally misspecified, we need to assume that H is either correctly specified, or locally misspecified with its $s > 1/2$. This generalizes our original theorems that simply assumed either G or H is correctly specified.

Finally, suppose model G is locally misspecified with $s = 1/2$. Then nQ^g converges to a noncentral chi-squared distribution. Specifically, $nQ^g(\hat{\theta}_g; \hat{\Omega}_g) \to^d \chi^2_{k_g}(\omega_g' \Omega_g \omega_g)$, where the object in parentheses is the noncentrality parameter and the formula for ω_g is given in the Supplemental Appendix. In this case the GMM estimator of model G is consistent but not $\sqrt{n}$ consistent, as established in, e.g., Newey and McFadden (1994). Here nQ^g is still bounded in probability, and if H is correctly specified (or locally misspecified with its $s > 1/2$), then nQ^f is also bounded in probability. Thus, ODR will asymptotically put weight on model F, which then is consistent but may not be $\sqrt{n}$ consistent. As a result, in this knife edge case, ODR will be consistent, but not $\sqrt{n}$ consistent, since $\sqrt{n}(\hat{}$

will have the same limiting distribution as e¢ cient GMM with both G and H correctly speci…ed. If just G is locally misspeci…ed with $s > 1=2$ (again including as a special case having G be correctly speci…ed by $s = 1$), and H is either misspeci…ed or locally misspeci…ed with $s < 1=2$, then (assuming $> s + 0{:}5$) ODR will have the same limiting distribution as e¢ cient GMM based just on model G (and vice versa, exchanging the roles of G and H). Equivalently we can say that our earlier Theorem 2 still holds, replacing "correctly speci…ed model" with "locally misspeci…ed model having any $s > 1=2$, including $s = 1$" and replacing "incorrectly speci…ed model" with "locally misspeci…ed model having any $s < 1=2$, including $s = 0$."

We conclude this section with some additional Monte Carlo results (reported in Tables 6 and 7 in the Supplemental Appendix), which we …nd support these conclusions. We use the same designs and estimators as in section 5 but with a drift parameter s for the locally misspeci…ed cases. Since ODR performed better with the tuning function $_1$ in section 5, to save space we only report ODR_1, along with GMM_g, GMM_h, and GMM_f. In these tables, model H is either globally misspeci…ed, or locally misspeci…ed with s equal to $0{:}25$, $0{:}50$, or $0{:}75$. In Tables 6-1 and 6-2 model G is correctly speci…ed, while in Tables 7-1 and 7-2, G is locally misspeci…ed with $s = 0{:}75$.

The …nite heandav-3339 Td (tl)]TJ /TT.97 -247 -24.388 Td [7 -247 -3n estimation of would then be the weighted average

$$= \frac{\hat{Q}^g(\ _g;\ _g)\hat{Q}^h(\ _h;\ _h)\ _l + \hat{Q}^l(\ _l;\ _l)\hat{Q}^h(\ _h;\ _h)\ _g + \hat{Q}^l(\ _l;\ _l)\hat{Q}^g(\ _g;\ _g)\ _h}{\hat{Q}^g(\ _g;\ _g)\hat{Q}^h(\ _h;\ _h) + \hat{Q}^l(\ _l;\ _l)\hat{Q}^h(\ _h;\ _h) + \hat{Q}^l(\ _l;\ _l)\hat{Q}^g(\ _g;\ _g)} \tag{14}$$

$$= \frac{\frac{_l}{\hat{Q}^l(\ _l;\ _l)} + \frac{_g}{\hat{Q}^g(\ _g;\ _g)} + \frac{_h}{\hat{Q}^h(\ _h;\ _h)}}{\frac{1}{\hat{Q}^l(\ _l;\ _l)} + \frac{1}{\hat{Q}^g(\ _g;\ _g)} + \frac{1}{\hat{Q}^h(\ _h;\ _h)}}: \tag{15}$$

In equation (14), the weight on $_l$ is proportional to the product of objective functions for the other models, $\hat{Q}^g\hat{Q}^h$, and similarly for the weights on $_g$ and $_h$.

The above estimator is a simple extension of our $SODR$ estimator because the $SODR$ can be rewriten as

$$= \frac{\frac{_g}{\hat{Q}^g(\ _g;\ _g)} + \frac{_h}{\hat{Q}^h(\ _h;\ _h)}}{\frac{1}{\hat{Q}^g(\ _g;\ _g)} + \frac{1}{\hat{Q}^h(\ _h;\ _h)}}:$$

The logic of is the same as for the $SODR$ estimator. For example, if model G is right and models L and H are wrong, then only $_g$ will get a nonzero weight asymptotically. Now suppose two but6getbuly del8 10.9

can suffer from well known finite sample biases when models have many more moments than parameters, and particularly when some moments might be weak. In such cases, it may be desirable to let models G and H equal just a subset of the available moments for each. Existing moment selection methods such as Andrews and Lu (2001), Caner (2009), or Liao (2013) might be used prior to applying ODR, though this then introduces pretest bias that ODR is intended to avoid. A potential subject for future work could be more formally modifying ODR to deal with333(dea(A)-341(p)-p)-28.004(os(used)-]TJ 0 -24.ectisused)-002/999(ac24

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

Average Treatment E¤ects", 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.

Supplemental Appendix: Over-Identi…ed Doubly Robust Identi…cation and Estimation

by Arthur Lewbel, Jin-Young Choi, and Zhuzhu Zhou

Original 2018, Revised November 2021

This Supplemental Appendix consists of …ve parts. The …rst is a proof of Lemma 1 and of Theorem 2, which give the asypmtotic properties of ODR described in Section 4.1-2. The second part is a proof of Lemma 2 and of Theorems 3 and 4, which provide asymptotic properties of ODR where k_g is the degrees of freedom of the chi-squared statistic that $n\hat{Q}^g$ converges to if the G model is true. That is the integer k_g is the number of moments in G (k_g) minus the number of elements in and (k_g) which is positive as discussed earlier. For notational simplicity, let $\hat{Q}^g$, $\hat{Q}^h$, and $\hat{Q}^f$ denote $\hat{Q}^g(\hat{}_g; \hat{}_g)$; $\hat{Q}^h(\hat{}_h; \hat{}_h)$; and $\hat{Q}^f(\hat{}_f; \hat{}_f; \hat{}_f)$; respectively. The population version of $\hat{Q}$ corresponding to each model is

$$Q_0^g \quad \frac{c_g^0 \quad {}_g c_g}{k_g}; \quad Q_0^h \quad \frac{c_h^0 \quad {}_h c_h}{k_h}; \quad Q_0^f \quad \frac{c_f^0 \quad {}_f c_f}{k_f}:$$

If the model is correctly speci…ed, then $Q_0^j = 0$, $j = g; h; f$.

Our proposed ODR estimator is a weighted average of b_g, b_h, and b_f, taking the form

$$b = \hat{W}_f \hat{W}_g b_h + \hat{W}_f \quad 1 \quad \hat{W}_g \quad b_g + (1 \quad \hat{W}_f) b_f \tag{1}$$

Proof of Lemma 1.

To obtain the probability limits of $\hat{W}_g$ and $\hat{W}_f$, …rst we consider without loss of generality the probability limit of $\hat{Q}^g$ when model G is correctly speci…ed, and when it's misspeci…ed. The asymptotics for $\hat{Q}^h$ and $\hat{Q}^f$ are obtained following the same logic. After these derivations, we then obtain the probability limits of $\hat{W}_g$ and $\hat{W}_f$ based on $\hat{Q}^g$, $\hat{Q}^h$ and $\hat{Q}^f$. First we have

$$n\hat{Q}^g = f \hat{}_g^{1=2} \sqrt{n}\bar{g}(\hat{}^g)g^0 f \hat{}_g^{1=2} \sqrt{n}\bar{g}(\hat{}^g)g \frac{1}{k_g}: \tag{2}$$

From the …rst order condition for $\hat{}^g$ minimizing $\hat{Q}^g(\ )$, we have

$$\sqrt{n} r \quad \bar{g}(\hat{}^g) \quad \hat{}_g \bar{g}(\hat{}^g) = 0:$$

Taylor-expanding the last term $\bar{g}(\hat{}^g)$ around ^{g} gives

$$\begin{aligned} 0 &= \sqrt{n} r \quad \bar{g}(\hat{}^g) \quad \hat{}_g f \bar{g}(\ ^g) + r \quad {}_0\bar{g}(\ ^{-g})(\hat{}^g \quad ^g) g \\ &= r \quad \bar{g}(\hat{}^g) \quad \hat{}_g \sqrt{n} \bar{g}(\ ^g) + r \quad \bar{g}(\hat{}^g) \quad \hat{}_g r \quad {}_0\bar{g}(\ ^{-g}) \sqrt{n}(\hat{}^g \quad ^g) \end{aligned}$$

where $^{-g}$ is a mean value between ^{g} and $\hat{}^g\hat{}$

where $\bar{\theta}^g$ is a mean value between θ^g_0 and $\hat{\theta}^g$. Plug equation (3) with θ^g replaced by θ^g_0 into this equation to get

$$
\begin{aligned}
\hat{\Omega}_g^{-1/2}\sqrt{n}\,\bar{g}(\hat{\theta}^g) &= \hat{\Omega}_g^{-1/2}\sqrt{n}\,\bar{g}(\theta^g_0) - \hat{\Omega}_g^{-1/2}\nabla_{\theta'}\bar{g}(\bar{\theta}^g)(\hat{H}^g)^{-1}\nabla_{\theta}\bar{g}(\hat{\theta}^g)'\hat{\Omega}_g^{-1}\sqrt{n}\,\bar{g}(\theta^g_0) \\
&= \{I_{\tilde{k}_g} - \hat{\Omega}_g^{-1/2}\nabla_{\theta'}\bar{g}(\bar{\theta}^g)(\hat{H}^g)^{-1}\nabla_{\theta}\bar{g}(\hat{\theta}^g)'\hat{\Omega}_g^{-1/2}\}\,\hat{\Omega}_g^{-1/2}\sqrt{n}\,\bar{g}(\theta^g_0) = \hat{\Lambda}_g\hat{\Omega}_g^{-1/2}\sqrt{n}\,\bar{g}(\theta^g_0); \qquad (4) \\
&\text{where } \hat{\Lambda}_g \equiv I_{\tilde{k}_g} - \hat{\Omega}_g^{-1/2}\nabla_{\theta'}\bar{g}(\bar{\theta}^g)(\hat{H}^g)^{-1}\nabla_{\theta}\bar{g}(\hat{\theta}^g)'\hat{\Omega}_g^{-1/2}
\end{aligned}
$$

and $I_{\tilde{k}_g}$ is the $\tilde{k}_g \times \tilde{k}_g$ identity matrix and $\tilde{k}_g$ is the number of moments in the model G.

By Assumption A10 and the Lindberg-Levy CLT, $\sqrt{n}\,\bar{g}(\theta^g_0) \xrightarrow{d} N$

Under Assumption A13, $\sqrt{n}\{\nabla \hat{g}(\;^{g}) - \nabla g_0(\;^{g})\}$ is asymptotically normal with mean zero by the Lindberg-Levy CLT, so it is bounded in probability. And $\sqrt{n}(\hat{\;}_g - \;_g)$ is also bounded in probability by Assumption A12. Under Assumption A7, A10, A11, A14, and the consistency of $\hat{\;}$

and $\hat{W}_f \hat{W}_g \to^p 0$.

Case 3). Suppose now $g_0(\ _0;\ _0) \neq 0$ but $h_0(\ _0;\ _0) = 0$. Then $\{\hat{}_g; \hat{}_g\} \to^p \{\ _g;\ _g\}$, $\{\hat{}_h; \hat{}_h\} \to^p \{\ _0;\ _0\}$, and $\{\hat{}_f; \hat{}_f; \hat{}_f\} \to^p \{\ _f;\ _f;\ _f\}$. So $\hat{Q}^g \to^p Q_0^g = c_g^{0}\ _g c_g / k_g > 0$, $\hat{Q}^h \to^p Q_0^h = c_h^{0}\ _h c_h / k_h = 0$, and $\hat{Q}^f \to^p Q_0^f = c_f^{0}\ _f c_f / k_f > 0$. Following the same argument as in Case 2), $\hat{W}_g \to^p 1$ and $\hat{W}_f \to^p 1$. In short, the probability limits of $\hat{W}_f$ and $\hat{W}_g \hat{W}_f$ are categorized as follows:

Case 1) Both G and H are correctly specified $\Rightarrow$ $\hat{W}_f \to^p 0$ and $\hat{W}_f \hat{W}_g \to^p 0$;
Case 2) G is correctly specified, but H is not $\Rightarrow$ $\hat{W}_f \to^p 1$ and $\hat{W}_f \hat{W}_g \to^p 0$;
Case 3) H is correctly specified, but G is not $\Rightarrow$ $\hat{W}_f \to^p 1$ and $\hat{W}_f \hat{W}_{fg} \to^p 1$:

Q.E.D.

Proof of Theorem 2 .
Recall equation (1) and rewrite it as

$$\hat{} = \ _0 + \hat{W}_f \hat{W}_g (\hat{}_h \quad _0) + \hat{W}_f \quad 1 \quad \hat{W}_g \quad (\hat{}_g \quad _0) + (1 \quad \hat{W}_f)(\hat{}_f \quad _0):$$

From this, we have

$$\begin{aligned} \sqrt{n}(\hat{} \quad _0) &= \hat{W}_f \hat{W}_g \sqrt{n}(\hat{}_h \quad _0) + \hat{W}_f \quad 1 \quad \hat{W}_g \quad \sqrt{n}(\hat{}_g \quad _0) + (1 \quad \hat{W}_f)\sqrt{n}(\hat{}_f \quad _0) \\ &= \hat{W}_f \hat{W}_g \sqrt{n}(\hat{}_h \quad _h) + \hat{W}_f \quad 1 \quad \hat{W}_g \quad \sqrt{n}(\hat{}_g \quad _g) + (1 \quad \hat{W}_f)\sqrt{n}(\hat{}_f \quad _f) \qquad (9) \end{aligned}$$

30.9(9531d)1001)620144245r5)4379721.7Tj3 0T[(1 116.a)8100Tr2.4596Td)9(y9.i59r0(h)Tj 0TT1 11.955 f 0

d (0)Tj 2d (0)Tj 2

Case 2). Suppose G is correct, but H is not (h 0 h 6= 0). In this case, F is also misspeci…ed (f 0 f 6= 0). (9) can be rewritten as

p n(^ 0) = Ŵ f fn^ W021 08d [())-278(=)]TJ22(+711 3.2.383d (^)Tj /TT1 11.955 Tf -3.412 -3.022 Td (W

Appendix II: Proof of Lemma 2 and Theorems 3 and 4

Let the model G be "locally misspeci…ed" when the parameter in the data generating process takes the form $^{g} = {}^{g}_{0} + {}_{g} n^{s}$ for a constant ${}_{g}$ and $s > 0$, while ${}^{g}_{0}$ satis…es $E f G(Z; {}^{g}_{0}) g = 0$ due to Assumption A3${}_{g}$Ahemn

Following the same steps as in Case ii) of Lemma 1, we can rewrite the last term other than $(\hat{H}^{g})^{1}$ in (15) as

$$r \quad \hat{g}(\hat{\ }^{g})\,\hat{\ }_{g}{}^{p}$$

Case 4). Suppose that model G is correct, but H is locally misspeci…ed with $^h = {}^h_0 + {}_h n^{s}$. In this case, F is also locally misspeci…ed with $^f = {}^f_0 + {}_f n^{s}$ for some $_f$.

Case 4-1). If $s = 1=2$, as shown in Case iii), $n\hat{Q}^h \; !_d \; {}^2_{k_h}(!\,{}^0_h \; {}_h!\,{}_h)=k_h$ and $n\hat{Q}^f \; !_d \; {}^2_{k_f}(!\,{}^0_f \; {}_f!\,{}_f)=k_f$ as $n \; ! \; 1$. Thus $\hat{W}_g = n\hat{Q}^g(\hat{}_g;\hat{}_g)=fn\hat{Q}^g(\hat{}_g;\hat{}_g) + n\hat{Q}^h(\hat{}_h;\hat{}_h)g$ converges to a distribution on $(0;1)$. For $\hat{W}_f$, we have

$$\hat{W}_f = 1 \quad \frac{1}{n\;\hat{Q}^f + 1} = 1 \quad \frac{1}{n^{\;1}\, n\hat{Q}^f + 1} \; !^{\,p} 0;$$

because $n\hat{Q}^f$ is bounded in probability, and $n^{\;1} \; !^{\,p} 0$. Thus, $\hat{W}_g\hat{W}_f \; !^{\,p} 0$.

Case 4-2). If $s > 1=2$, $n\hat{Q}^h \; !_d \; {}^2_{k_h}=k_h$, and $n\hat{Q}^f \; !_d \; {}^2_{k_f}=k_f$. Therefore, it is asymptotically the same as Case 1) of Lemma 1.

Case 4-3). If $s < 1=2$, $n\hat{Q}^h$ and $n\hat{Q}^f$ are $O_p(n^{2(1=2\;s)})$, as each is a squared version of a term analogous to (18). In this case, whereas $\hat{W}_g \; !^{\,p} 0$, convergence of $\hat{W}_f$ depends on the relationship between and s. Because $n\;\hat{Q}^f = O(n^{\;1})O_p(n^{2(1=2\;s)}) = O_p(n^{\;2s})$, when $> 2s$, $n\;\hat{Q}^f$ diverges to result in $\hat{W}_f \; !^{\,p} 1$ and $\hat{W}_g\hat{W}_f \; !^{\,p} 0$. When $< 2s$, $n\;\hat{Q}^f \; !^{\,p} 0$, and consequently $\hat{W}_f \; !^{\,p} 0$ and $\hat{W}_f\hat{W}_g \; !^{\,p} 0$. When $= 2s$, however, (18) shows that $n\;\hat{Q}^f \; !^{\,p} \; !\,{}^0_f \; {}_f!\,{}_f$ because only the last term of (18) matters, so that $\hat{W}_f \; !^{\,p} W_f \quad 1 \quad (!\,{}^0_f \; {}_f!\,{}_f + 1)^{\;1}$ and $\hat{W}_g\hat{W}_f \; !^{\,p} 0$.

Case 5). Suppose that model G is locally misspeci…ed with $^g = {}^g_0 + {}_g n^{s}$, but model H is correct. Then essentially the same arguments as in Case 4) apply.

Case 5-1). If $s = 1=2$, then $n\hat{Q}^g \; !_d \; {}^2_{k_g}(!\,{}^0_g \; {}_g!\,{}_g)=k_g$ and $n\hat{Q}^f \; !_d \; {}^2_{k_f}(!\,{}^0_f \; {}_f!\,{}_f)=k_f$. Thus, $\hat{W}_f \; !^{\,p} 0$ and $\hat{W}_g\hat{W}_f \; !^{\,p} 0$.

Case 5-2). If $s > 1=2$, then $n\hat{Q}^g \; !_d \; {}^2_{k_g}=k_g$s37nnd

Case 4). Suppose that G is correct, but H is the locally misspeci…ed with $^h = {}^h_0 + {}_h n^{s}$. By Theorem 9.1 of in Newey and McFadden (1994), still $\{\hat{}_g; \hat{}_g\} \xrightarrow{p} \{{}_0; {}_0\}$, $\{\hat{}_h; \hat{}_h\} \xrightarrow{p} \{{}_0; {}_0\}$ and $\{\hat{}_f; \hat{}_f; \hat{}_f\} \xrightarrow{p} \{{}_0; {}_0; {}_0\}$. By Lemma 2, if s 1=2, then $\hat{W}_f \xrightarrow{p} 0$ and $\hat{W}_g\hat{W}_f \xrightarrow{p} 0$, and the consistency of $\hat{}$ in (1) follows from consistency of $\hat{}_f$. If s < 1=2, the probability limits of $\hat{W}_f$ and $\hat{W}_g\hat{W}_f$ depend on the relationship between and s. If s < 1=2 and < 2s, the limits are the

same0e 0e068 Td (0e068 Td (0e068 11.955 .955 Tf 15j /Tg65511.955 Tf 22.97 0 TdTj 441.955 Tf (=7[)-28

By Assumption A12 and A13, $\sqrt{n}\{\nabla \hat{h}(\theta^h) - \nabla h_0(\theta^h)\}$ and $\sqrt{n}(\hat{\theta}_h - \theta_h)$ are bounded in probability. Given $\hat{\theta}_h \to^p \theta_h$, the last two terms in b_h converge to zero because $\delta_h n^{-s} \to 0$ as $n \to \infty$ for $s > 0$. Therefore,

$$\sqrt{n}(\hat{\theta}_h - \theta_h) = -A_h^{-1} \nabla \hat{h}(\hat{\theta}^h)' \Omega_h \sqrt{n}\{\hat{h}(\theta^h) - \delta_h n^{-s}\} + o_p(1).$$

By Assumption A7, A9, A10, A11, and the consistency of $\hat{\theta}_h$ for θ_0, $\sqrt{n}\{\hat{h}(\theta^h) - \delta_h n^{-s}\} \to^d N(0; \Sigma_h)$ where $\Sigma_h = Var[H(Z; \theta_h; \eta_h)]$, $\nabla \hat{h}(\bar{\theta}^h) \to^p \nabla h_0(\theta_0^h)$, $\nabla \hat{h}(\hat{\theta}^h) \to^p \nabla h_0(\theta_0^h)$, and $\hat{H}^h \to^p H^h$ which is non-singular by Assumption A8. Thus, by the continuous mapping theorem, we get

$$\sqrt{n}(\hat{\theta}_h - \theta_h) \to^d N(0; V^h);$$

where V^h is the same asymptotic variance as in Case 3) of Theorem 2 as if model H were correct. Analogously, the same argument holds for $\sqrt{n}(\hat{\theta}_f - \theta_f)$, so that we have $\sqrt{n}(\hat{\theta}_f - \theta_f) \to^d N(0; V^f)$. Hence, all of $\sqrt{n}(\hat{\theta}_g - \theta_0)$, $\sqrt{n}(\hat{\theta}_h - \theta_h)$ and $\sqrt{n}(\hat{\theta}_f - \theta_f)$ in the first line of (19) are asymptotically normal with mean zero and variance being that of the corresponding GMM estimator under correct specification.

Recall (19):

$$\begin{aligned}\sqrt{n}(\hat{\theta} - \theta_0) = {} & \hat{W}_f \hat{W}_g \sqrt{n}(\hat{\theta}_h - \theta_h) + \hat{W}_f (1 - \hat{W}_g) \sqrt{n}(\hat{\theta}_g - \theta_0) + (1 - \hat{W}_f) \sqrt{n}(\hat{\theta}_f - \theta_f) \\ & + \hat{W}_f \hat{W}_g \delta_h n^{1/2 - s} + (1 - \hat{W}_f) \delta_f n^{1/2 - s}.\end{aligned}$$

Recalling (14) and its "squared version", we have

$$n\hat{Q}^h = O_p(n^{2(1/2-s)}) \text{ and } n\hat{Q}^f = O_p(n^{2(1/2-s)}) \Longrightarrow \hat{Q}^f = n^{-1} n\hat{Q}^f = O_p(n^{-1+2(1/2-s)}) = O_p(n^{-2s}).$$

Consequently, for the last two terms in (19), we get

$$\begin{aligned}\hat{W}_f \hat{W}_g \delta_h n^{1/2-s} + (1 - \hat{W}_f) \delta_f n^{1/2-s} &= \left(1 - \frac{1}{n\hat{Q}^f + 1}\right) \left(\frac{n\hat{Q}^g}{n\hat{Q}^g + n\hat{Q}^h} \delta_h n^{1/2-s}\right) + \frac{1}{n\hat{Q}^f + 1} \delta_f n^{1/2-s} \\ &= \left(1 - \frac{1}{O_p(n^{-2s}) + 1}\right) \frac{O_p(1) O(n^{1/2-s})}{O_p(1) + O_p(n^{2(1/2-s)})} + \frac{1}{O_p(n^{-2s}) + 1} O(n\end{aligned}$$

Case 4-2). If $s > 1/2$, $\hat{W}_f \xrightarrow{p} 0$ and $(1 - \hat{W}_f)\,{}_f\, n^{1/2 - s} \xrightarrow{p} 0$ as $n \to \infty$

Under Assumption A7 and A9, the following …rst-order conditions hold:

$$FD^f = \frac{@Q^f(\hat{}^f)}{@} = r \quad f\!b(\hat{}^f)\,\hat{}_f\, f\!b(\hat{}^f) = 0; \qquad FD^f = \frac{@Q^f(\hat{}^f)}{@} = r \quad f\!b(\hat{}^f)\,\hat{}_f\, f\!b(\hat{}^f) = 0;$$

$$FD^f = \frac{@Q^f(\hat{}^f)}{@} = r \quad f\!b(\hat{}^f)\,\hat{}_f\, f\!b(\hat{}^f) = 0:$$

Expend $f\!b$ around the unique minimizer $^f \quad f \quad _f;\ _f;\ _f\, g$ to get

$$f\!b(\hat{}^f) = f\!b(\ ^f) + r \quad _0 f\!b(\bar{}^f)(\hat{}_f \quad _f) + r \quad _0 f\!b(\bar{}^f)(\hat{} \quad _f) + r \quad _0 f\!b(\bar{}^f)(\hat{} \quad _f);$$

where $\bar{}^f$ is the mean value to apply the mean value theorem. Substitute these into each FD^f to get

$$FD^f = r \quad f\!b(\hat{}^f)\,\hat{}_f\, f f\!b(\ ^f) + r \quad _0 f\!b(\bar{}^f)(\hat{}_f \quad _f) + r \quad _0 f\!b(\bar{}^f)(\hat{}_f \quad _f) + r \quad _0 f\!b(\bar{}^f)(\hat{} \quad _f)g;$$

$$FD^f = r \quad f\!b(\hat{}^f)\,\hat{}_f\, f f\!b(\ ^f) + r \quad _0 f\!b(\bar{}^f)(\hat{}_f \quad _f) + r \quad _0 f\!b(\bar{}^f)(\hat{}_f \quad _f) + r \quad _0 f\!b(\bar{}^f)(\hat{} \quad _f)g:$$

$$FD^f = r \quad f\!b(\hat{}^f)\,\hat{}_f\, f f\!b(\ ^f) + r \quad _0 f\!b(\bar{}^f)(\hat{}_f \quad _f) + r \quad _0 f\!b(\bar{}^f)(\hat{}_f \quad _f) + r \quad _0 f\!b(\bar{}^f)(\hat{} \quad _f)g;$$

$FD^f = fFD^f; FD^f; FD^f g = b^f + H^f(\hat{}^f \quad ^f)$, and from these, $\sqrt{n}(\hat{}^f \quad ^f) = H^{f\,1}\sqrt{n}\,b^f$;

$$b^f \begin{bmatrix} r & f\!b(\hat{}^f)\,\hat{}_f\, f\!b(\ ^f) \\ r & f\!b(\hat{}^f)\,\hat{}_f\, f\!b(\ ^f) \\ r & f\!b(\hat{}^f)\,\hat{}_f\, f\!b(\ ^f) \end{bmatrix}; H^f \begin{bmatrix} r\ f\!b(\hat{}^f)\,\hat{}_f\, r\ _0 f\!b(\bar{}^f) & r\ f\!b(\hat{}^f)\,\hat{}_f\, r\ _0 f\!b(\bar{}^f) & r\ f\!b(\hat{}^f)\,\hat{}_f\, r\ _0 f\!b(\bar{}^f) \\ r\ f\!b(\hat{}^f)\,\hat{}_f\, r\ _0 f\!b(\bar{}^f) & r\ f\!b(\hat{}^f)\,\hat{}_f\, r\ _0 f\!b(\bar{}^f) & r\ f\!b(\hat{}^f)\,\hat{}_f\, r\ _0 f\!b(\bar{}^f) \\ r\ f\!b(\hat{}^f)\,\hat{}_f\, r\ _0 f\!b(\bar{}^f) & r\ f\!b(\hat{}^f)\,\hat{}_f\, r\ _0 f\!b(\bar{}^f) & r\ f\!b(\hat{}^f)\,\hat{}_f\, r\ _0 f\!b(\bar{}^f) \end{bmatrix}:$$

In this expression for $\sqrt{n}(\hat{}^f \quad ^f)$, examine the part for $\sqrt{n}(\hat{}_f \quad _f)$, i.e., the …rst $k \quad 1$ components:

$$\sqrt{n}(\hat{}_f \quad _f) = A_f^{b\,1}\, r \quad f\!b(\hat{}^f)\,\hat{}_f \sqrt{n} f\!b(\ ^f); \quad \text{where } A_f^b \quad r \quad f\!b(\hat{}^f)\,\hat{}_f\, r \quad _0 f\!b(\bar{}^f); \quad \hat{}_f \quad \hat{}_f^{1=2}\,\hat{}_f\,\hat{}_f^{1=2}; \tag{22}$$

$$\hat{}_f \quad I_{k_f} \quad \hat{}_f^{1=2} r \quad f\!b(\bar{}^f) f r \quad f\!b(\hat{}^f)\,\hat{}_f\, r \quad _0 f\!b(\bar{}^f) g^{1}\, r \quad f\!b(\hat{}^f)\,\hat{}_f^{1=2}$$
$$\hat{}_f^{1=2} r \quad f\!b(\bar{}^f) f r \quad f\!b(\hat{}^f)\,\hat{}_f\, r \quad _0 f\!b(\bar{}^f) g^{1}\, r \quad f\!b(\hat{}^f)\,\hat{}_f^{1=2}:$$

Then, we have

$$\sqrt{n}(\hat{}_f \quad _f) = \frac{1}{\sqrt{n}} \sum_i b_i^f; \qquad b_i^f \quad A_f^{b\,1}\, r \quad f\!b(\hat{}^f)\,\hat{}_f\, F(Z_i;\ ^f); \tag{23}$$

and b_i^f is the in‡uence function of167.97 Tfuence iifib/TT3 11.955 T.45T1 . 1119. 11.955 Tf -0.riban‡cccd

TTd a(ff)4.98 7124.3 S. 0 TE 95 2 9 0 4

Expend l_g around the unique minimizer ^{g} f $_g$; $_g$g to get

$$l_g(\hat{}^g) = l_g(\ ^g) + \mathrm{f}r\ _0 l_g(^{-g})\mathrm{g}(\hat{}_g\ \ _g) + \mathrm{f}r\ _0 l_g(^{-g})\mathrm{g}(\hat{}\ \ _g)$$

where $^{-g}$ is the value for the mean value theorem. Substitute these into each FD^g to get

$$FD^g = r\ \ l_g(\hat{}^g)\hat{}_g[l_g(\ ^g) + r\ _0 l_g(^{-g})(\hat{}_g\ \ _g) + r\ _0 l_g(^{-g})(\hat{}_g\ \ _g)];$$
$$FD^g = r\ \ l_g(\hat{}^g)\hat{}_g[l_g(\ ^g) + r\ _0 l_g(^{-g})(\hat{}_g\ \ _g) + r\ _0 l_g(^{-g})(\hat{}_g\ \ _g)];$$
$$FD^g = \mathrm{f}FD$$

an outcome, T is a binary treatment indicator, and X is a J vector of other covariates (including

Observe that if $\widetilde{H}(X;\ ) = E(T|X)$, then the first two terms in the above expectation equal equation (27) and the second two terms have mean zero. By rearranging terms, equation (30) can be rewritten as

$$= E\left[\widetilde{G}(1;X;\ )\quad \widetilde{G}(0;X;\ ) + \frac{T}{\widetilde{H}(X;\ )}\{Y \quad \widetilde{G}(1;X;\ )\} \quad \frac{1\quad T}{1\quad \widetilde{H}(X;\ )}\{Y\quad \widetilde{G}(0;X;\ )\}\right]. \tag{31}$$

Rewriting the equation this way, it can be seen that if $\widetilde{G}(T;X;\ ) = E(Y|T;X)$, then the first two terms in equation (31) equal equation (26), and the second two terms have mean zero. This shows that equation (30) or equivalently (31) is doubly robust, in that it equals the average treatment effect if either $\widetilde{G}(T;X;\ )$ or $\widetilde{H}(X;\ )$ is correctly specified. The GMM estimator associated with this doubly robust estimator estimates , , and , using the moments

$$E\left[\begin{matrix} \{Y \quad \widetilde{G}(T;X;\ )\}r_1(T;X) \\ \{T \quad \widetilde{H}(X;\ )\}r_2(X) \\ \frac{YT}{\widetilde{H}} \end{matrix}\right.$$

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

$$Y = M(W; \ ) + \tilde{G}(X) + U; \tag{35}$$

$$E(Q \mid X) = \tilde{H}(X);$$
$$E(U \mid X; Q) = 0; \tag{36}$$

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 $K \quad S$ 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 $M(W; \ )$ is assumed to be correctly parameterized, and the goal is estimation of

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

$$G(Z; \ ; \ ) = \begin{bmatrix} \{Y \quad M(W; \ ) \quad \tilde{G}(X; \ )\} r_1(X) \\ \{Y \quad M(W; \ ) \quad \tilde{G}(X; \ )\} Q \end{bmatrix} \tag{37}$$

and

$$H(Z; \ ; \ ) = \begin{bmatrix} \{Q \quad \tilde{H}(X; \ )\} r_2(X) \\ \{Y \quad M(W; \ )\}\{Q \quad \tilde{H}(X; \ )\} \end{bmatrix}. \tag{38}$$

Okui, Small, Tan, and Robins (2012) take $r_1(X) = @\tilde{G}(X; \ )/@$ and $r_2(X) = @\tilde{H}(X; \ )/@$. If $\tilde{G}(X; \ )$ is correctly specified, then $E\{G(Z; \ ; \ )\} = 0$, while if $\tilde{H} = @$ and [illegible]

Table 6-1. Model G is Correctly Speci…ed and Model H is Misspeci…ed (n = 500)

$_1$	Bias	Mde	RMSE	MAE	SD	Skew	Kurt	Freq	SE	SD_{SE}
s=0.25										
GMM_g	0.0002	0.0006	0.0001	0.0075	0.0111	0.2310	3.1966	0.9465	0.0108	0.0011
GMM_h	0.2374	0.2367	0.0566	0.2367	0.0157	0.1558	3.1392	0.0000	0.0139	0.0016
GMM_f	0.1094	0.1094	0.0121	0.1094	0.0112	0.0817	3.0557	0.0000	0.0068	0.0005
ODR_1	0.0002	0.0006	0.0001	0.0075	0.0111	0.2311	3.1963	0.9460	0.0108	0.0011
s=0.5										
GMM_g	0.0002	0.0006	0.0001	0.0075	0.0110	0.1255	3.081			

Table 7-1. Model G is Misspeci…ed with s = 0:75 and Model H is Misspeci…ed (n = 500)

$_1$	Bias	Mde	RMSE	MAE	SD	Skew	Kurt	Freq	SE	SD