THE CENTRE FOR MARKET AND PUBLIC ORGANISATION

Estimating Structural Mean Models with Multiple Instrumental Variables using the Generalised Method of Moments

Paul S Clarke, Tom M Palmer
and Frank Windmeijer

August 2011

Working Paper No. 11/266

Centre for Market and Public Organisation
Bristol Institute of Public Affairs
University of Bristol
2 Priory Road
Bristol BS8 1TX
http://www.bristol.ac.uk/cmpo/

Tel: (0117) 33 10952
Fax: (0117) 33 10705
E-mail: cmpo-admin@bristol.ac.uk

The Centre for Market and Public Organisation (CMPO) is a leading research centre, combining expertise in economics, geography and law. Our objective is to study the intersection between the public and private sectors of the economy, and in particular to understand the right way to organise and deliver public services. The Centre aims to develop research, contribute to the public debate and inform policy-making.

CMPO, now an ESRC Research Centre was established in 1998 with two large grants from The Leverhulme Trust. In 2004 we were awarded ESRC Research Centre status, and CMPO now combines core funding from both the ESRC and the Trust.

ISSN 1473-625X

CMPO Working Paper Series No. 11/266

Estimating Structural Mean Models with Multiple Instrumental Variables using the Generalised Method of Moments

Paul S. Clarke[1], Tom M. Palmer[2]
and Frank Windmeijer[3]

[1]*CMPO, University of Bristol*
[2]*MRC CAiTE Centre, School of Social & Community Medicine, University of Bristol*
[3] *CMPO and Department of Economics, University of Bristol & CEMMAP/IFS*

August 2011

Abstract
Instrumental variables analysis using genetic markers as instruments is now a widely used technique in epidemiology and biostatistics. As single markers tend to explain only a small proportion of phenotypical variation, there is increasing interest in using multiple genetic markers to obtain more precise estimates of causal parameters. Structural mean models (SMMs) are semi-parametric models that use instrumental variables to identify causal parameters, but there has been little work on using these models with multiple instruments, particularly for multiplicative and logistic SMMs. In this paper, we show how additive, multiplicative and logistic SMMs with multiple discrete instrumental variables can be estimated efficiently using the generalised method of moments (GMM) estimator, how the Hansen *J*-test can be used to test for model mis-specification, and how standard GMM software routines can be used to fit SMMs. We further show that multiplicative SMMs, like the additive SMM, identify a weighted average of local causal effects if selection is monotonic. We use these methods to reanalyse a study of the relationship between adiposity and hypertension using SMMs with two genetic markers as instruments for adiposity. We find strong effects of adiposity on hypertension, but no evidence of unobserved confounding.

Keywords Structural Mean Models, Multiple Instrumental Variables, Generalised Method of Moments, Mendelian Randomisation, Local Average Treatment Effects.
JEL Classification C13, C14, C26

Electronic version www.bristol.ac.uk/cmpo/publications/papers/2011/wp266.pdf

Acknowledgements
This work was funded by UK Economic & Social Research Council grant RES-060-23-0011, UK Medical Research Council grants G0601625 and G0600705, and European Research Council grant 269874 - DEVHEALTH. The authors would like to thank Borge Nordestgaard for access to the Copenhagen General Population Study data. We also thank George Davey Smith, Nicholas Timpson, Vanessa Didelez, Roger Harbord, Nuala Sheehan and conference participants in London and Hannheim for helpful comments.

Address for correspondence
CMPO, Bristol Institute of Public Affairs
University of Bristol
2 Priory Road
Bristol BS8 1TX
frank.windmeijer@bristol.ac.uk
www.bristol.ac.uk/cmpo/

1 Introduction

Additive and multiplicative structural mean models (SMMs) and G-estimaton were introduced by Robins (1989, 1994) for estimating the causal effects of treatment regimes on outcomes from encouragement designs, namely, randomised controlled trials (RCTs) affected by non-compliance. Additive SMMs are parameterised in terms of average treatment effects, and multiplicative SMMs in terms of causal risk ratios; the G-estimators for these models are consistent, asymptotically normal and can be semi-parametrically efficient. Vansteelandt and Goetghebeur (2003) subsequently developed a class of estimators for generalized SMMs and, in particular, for estimating causal odds ratios using the 'double logistic' SMM; see also Robins and Rotnitzky (2004), Goetghebeur and Vansteelandt (2005), and van der Laan et al. (2007).

The application of SMMs is not limited to encouragement designs, however, and extends to the analysis of observational studies using instrumental variables; see e.g. Hernán and Robins (2006). Instrumental variables analysis involves estimating the causal relationship between an outcome and a temporally antecedent predictor variable using an instrumental variable that is associated with the outcome *only* through its association with the predictor. Instrumental variables analyis has historically been the domain of econometrics, but is now frequently used within epidemiology and biostatistics. In particular, genetic markers were proposed as instruments for modifiable risk factors by Davey Smith and Ebrahim (2003). Epidemiological studies using genetic markers are known as Mendelian randomisation studies after the assumption that each individual's genotype is randomly assigned at conception, which implies that the genetic marker is an instrumental variable if it at least partly explains variation in the risk factor. In practice, genetic markers explain only a small proportion of phenotypic variation, and so large sample sizes are required to obtain any reasonable precision. The number of genome-wide association studies has increased as the costs of genotyping have decreased, which has led

to the identification of multiple genetic variants for the same risk factor. An important attraction of using multiple genetic variants as instrumental variables is that more precise causal estimates can potentially be obtained.

In this paper, we propose a framework for the estimation and testing of SMMs using multiple instrumental variables. Techniques for multiple instruments in linear instrumental variables analysis are already in use; see e.g. Palmer et al. (2011). However, our framework extends to non-linear semi-parametric models which are suitable for the study of binary and discrete outcomes, and the estimation of causal risk ratios and causal odds ratios. We use this framework to reanalyse data from the study of the relationship between hypertension and adiposity by Timpson et al. (2009). In the original study, two genetic markers were used as instruments for adiposity and analysed using linear instrumental variables models. We reanalyse this study by focusing on hypertension as a binary outcome, and estimating causal effects of adiposity using multiplicative and logistic SMMs.

The framework we propose does not come from extending the existing estimating equations for SMMs. Instead, we show how the basic model assumptions for SMMs lead straightforwardly to a generalised method of moments (GMM) estimator; see Hansen (1982). The theory and application of GMM are already standard within econometrics, where Chamberlain (1987) established results on the asymptotic efficiency of GMM estimators, and the Hansen J-test is a widely used test of instrument validity in applications involving multiple instruments. Moreover, routines for parameter estimation using GMM are already implemented in standard software packages like Stata and R; see Chaussé (2010). These routines can be used to obtain asymptotically correct inferences for all the SMMs considered here, and so make this important technique straightforwardly accessible to applied researchers.

We also consider the interpretation of additive and multiplicative SMMs with multiple

instruments when a key SMM assumption fails, namely, that of no effect modification by the instrumental variables (NEM). In such circumstances, an additive SMM with one binary instrument identifies a 'local' average treatment effect (LATE) - also known as a 'complier' average causal effect (CACE) - provided that selection is monotonic, and multiplicative SMMs identify local causal risk ratios; see e.g. Clarke and Windmeijer (2010). When there are multiple instruments, Imbens and Angrist (1994) show that a GMM estimator for the additive SMM identifies a weighted average of LATEs. We extend their analysis to multiplicative SMMs to show that a GMM estimator identifies weighted averages of local risk ratios.

The remainder of the paper is organised as follows. In Section 2 we review the potential outcomes framework and the additive, multiplicative and logistic SMMs for a single binary instrument. In Section 3, we discuss the GMM estimation procedure and rework the SMM moment conditions to fit into the GMM framework. Section 4 discusses the estimation of SMMs using GMM when there are multiple instruments. Section 5 presents some Monte Carlo results for the multiplicative and logistic SMMs. In Section 6 we derive the multiple instruments results for the local risk ratio. Section 7 applies the estimation procedures to the adiposity and hypertension data of Timpson et al. (2009). Finally, in Section 8 we make some concluding remarks. The Appendix provides Stata and R code for the estimation of the three SMMs by GMM.

2 Structural Mean Models

2.1 The basic set-up

To introduce SMMs, we follow the exposition in Hernán and Robins (2006) and focus on SMMs for a randomised controlled trial where Z_i, X_i and Y_i are *i.i.d.* dichotomous random variables for individual subjects $i = 1, \ldots, n$ drawn from the target population. For individual i, let Z_i to be a binary indicator of treatment assignment following random-

ization, X_i the selected treatment, and Y_i the study outcome. For notational simplicity the subject index is sometimes suppressed for the random variables.

The potential outcomes can now be defined in the usual way. The potential treatments X_0 and X_1 are the treatments selected by the individual following assignment to treatment $z = 0, 1$, respectively. Similarly, the potential study outcome Y_{xz} is that obtained if the individual is assigned to treatment z but given treatment x. Using potential outcomes notation, we can now state five key conditions that must be satisfied for causal inference: (i) the 'stable unit treatment value assumption' that each individual's potential treatments and potential study outcomes are mutually independent of those for any other individual; (ii) the 'consistency assumption' $X = X_Z$ and $Y = Y_{XZ}$ that links the observed realisations to the potential outcomes; (iii) the 'causal relationship' assumption that $E(X_z)$ and $E(Y_{xz})$ are non-trivial functions of z and (x, z), respectively; (iv) the 'exclusion restriction' $Y_{xz} = Y_x$; and (v) the 'independence assumption' that Z is independent of (X_0, X_1, Y_0, Y_1). Alternative statements of these key conditions are given by Robins and Rotnitzky (2004) and Tan (2010).

2.2 SMM Identification

For the basic set-up defined above, the generalised SMM of Vansteelandt and Goetghebeur (2003) is

$$h\{E(Y|X,Z)\} - h\{E(Y_0|X,Z)\} = (\psi_0 + \psi_1 Z)\, X, \tag{1}$$

where Y_0 is often referred to as the exposure-free potential outcome, and h is the link function that determines the interpretation of the target causal parameters ψ_0 and $\psi_0 + \psi_1$. For example, the identity link leads to the additive SMM $E(Y|X,Z) - E(Y_0|X,Z) = (\psi_0 + \psi_1 Z)\, X$, where $\psi_0 = E(Y_1 - Y_0|X = 1, Z = 0)$ and $\psi_0 + \psi_1 = E(Y_1 - Y_0|X = Z = 1)$ are both average treatment effects; the log link leads to the multiplicative SMM $E(Y|X,Z)/E(Y_0|X,Z) = \exp\{(\psi_0 + \psi_1 Z)\, X\}$, where $\exp(\psi_0) = E(Y_1|X = 1, Z =$

$0)/E(Y_0|X = 1, Z = 0)$ and $\exp(\psi_0 + \psi_1) = E(Y_1|X = Z = 1)/E(Y_0|X = Z = 1)$ are causal risk ratios. These models are saturated, or non-parametric, because each has one parameter for each value of Z.

In both cases, the SMM parameters are identified by exploiting the conditional mean independence (CMI), or randomisation, assumption

$$E(Y_0|Z = 0) = E(Y_0|Z = 1) = E(Y_0), \tag{2}$$

which follows automatically from the independence condition defined above. Under the additive SMM, $E(Y_0|Z) = E\{Y - (\psi_0 + \psi_1 Z)\, X|Z\}$ and its G-estimator is based on the moment condition

$$E\{Y - (\psi_0 + \psi_1)\, X|Z = 1\} = E(Y - \psi_0 X|Z = 0), \tag{3}$$

which follows from CMI. However, it is clear from (3) that further assumptions are required to identify ψ_0 and ψ_1 because there are two unknowns but only one moment condition; the multiplicative SMM is similarly non-identified without further assumptions.

Hernán and Robins (2006) highlight the identification assumption that there is no effect modification by Z (NEM), which constrains $\psi_1 = 0$ so that the two conditional causal effects in the model are equal. Under NEM, the identified parameter of the additive SMM is $\psi_0 = E(Y_1 - Y_0|X = 1)$, that is, the average treatment effect among the treated; for the multiplicative SMM it is $\exp(\psi_0) = E(Y_1|X = 1)/E(Y_0|X = 1)$, that is, the causal risk ratio among the treated.

The logistic SMM is given by

$$\text{logit}\{E(Y|X, Z)\} - \text{logit}\{E(Y_0|X, Z)\} = (\psi_0 + \psi_1 Z)\, X,$$

where $\text{logit}(p) = \log\{p/(1-p)\}$ and the parameters $\exp(\psi_0)$ and $\exp(\psi_0 + \psi_1)$ are causal odds ratios for the $(X, Z) = (1, 0)$ and $(1, 1)$ groups, respectively. Under NEM, $E(Y_0|Z) = E[\text{expit}\{\text{logit}(E(Y|X, Z)) - \psi_0 X\}|Z]$, where $\text{expit}(a) = \exp(a)/\{1 + \exp(a)\}$.

G-estimation cannot be used for the logistic SMM because the moment conditions following from CMI depend on $E(Y|X,Z)$; see e.g. Robins (1999). The estimating equations for the logistic SMM must be adjusted for estimates of an 'association model' for $E(Y|X,Z)$. Vansteelandt and Goetghebeur (2003) proposed the double-logistic SMM based on a logistic association model; in this example, $E(Y|X,Z) = \text{expit}(\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ)$ so that

$$E(Y_0|Z) = E[\text{expit}\{\beta_0 + (\beta_1 - \psi_0)X + \beta_2 Z + \beta_3 XZ\}|Z].$$

In general, a saturated association model like this one implies no further identifying assumptions because the estimator is effectively non-parametric; see Vandsteelandt et al. (2011). However, a non-saturated association model implies further semi-parametric assumptions that will lead to bias if this model is mis-specified; see Robins and Rotnitzky (2004) and Vansteelandt et al. (2011).

Standard approaches to estimating the SMMs discussed above are based on estimating equations for the moment condition

$$E[\{Z - E(Z)\}E(Y_0|Z)] = 0,$$

which holds under CMI. Asymptotic inference is based on theory for semi-parametric models; see e.g. Tsiatis (2006).

3 The Generalised Method of Moments

The three SMMs above are all just-identified in the sense that each has one parameter and one moment condition under CMI (conveniently taking $\boldsymbol{\beta}$ to be known for the double-logistic SMM). The solutions to these moment conditions for the just-identified models are unique; for example, the solution to (3) under NEM ($\psi_1 = 0$) gives

$$\psi_0 = \frac{E(Y|Z=1) - E(Y|Z=0)}{E(X|Z=1) - E(X|Z=0)}, \tag{4}$$

namely, the classical instrumental variable estimator; see e.g. Hernán and Robins (2006).

More generally, there will be fewer SMM parameters than moment conditions and the model is said to be 'over-identified'. There is no unique solution to the CMI moment conditions for over-identified models, but it is still possible to construct an estimator that is consistent and efficient. Hansen (1982) proposed the generalised method of moments (GMM) estimator for 'models' of the form $E\{\mathbf{g}(\boldsymbol{\delta})\} = \mathbf{0}$, where $\mathbf{g}(\boldsymbol{\delta})$ is a random vector and a function of parameter $\boldsymbol{\delta}$, and $\mathbf{0}$ is an appropriately dimensioned column vector of zeros. The GMM estimator can be written as

$$\widehat{\boldsymbol{\delta}} = \arg\min_{\boldsymbol{\delta}} \left\{ n^{-1} \textstyle\sum_{i=1}^{n} \mathbf{g}_i'(\boldsymbol{\delta}) \right\} W_n^{-1} \left\{ n^{-1} \textstyle\sum_{i=1}^{n} \mathbf{g}_i(\boldsymbol{\delta}) \right\}, \tag{5}$$

where $\mathbf{g}_i(\boldsymbol{\delta})$ is the random vector for subject i, $\mathbf{g}_i'(\boldsymbol{\delta})$ is the matrix transpose of $\mathbf{g}_i(\boldsymbol{\delta})$, and the weight matrix W_n determines the efficiency of the estimator.

Instrumental variables models take the form $\mathbf{g}(\boldsymbol{\delta}) = v(\boldsymbol{\delta})\mathbf{S}$, where $v(\boldsymbol{\delta})$ is a 'residual' depending on Y, X and $\boldsymbol{\delta}$, and $\mathbf{S}$ is a random vector of instrumental variables. The moment conditions for the three SMMs introduced above can be written in this form: the CMI moment conditions can be expressed as $E(Y_0|Z=z) - \alpha_0 = 0$ for $z = 0, 1$, where $\alpha_0 = E(Y_0)$ is treated as a parameter and results in the additional moment condition $E(Y_0) - \alpha_0 = 0$. It follows that one of $E(Y_0|Z=z) - \alpha_0 = 0$ is redundant because Z is discrete and $E\{E(Y_0|Z)\} = \alpha_0$ by definition. However, using the additional $E(Y_0) - \alpha_0 = 0$ moment condition allows the system of moment conditions to be expressed in a convenient form. For example, under the additive SMM it follows that

$$\begin{bmatrix} E(Y - \psi_0 X) - \alpha_0 \\ E(Y - \psi_0 X | Z = 1) - \alpha_0 \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow E \begin{bmatrix} Y - \psi_0 X - \alpha_0 \\ (Y - \psi_0 X - \alpha_0) Z \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{6}$$

that is, $E\{\mathbf{g}(\psi_0, \alpha_0)\} = \mathbf{0}$, where $\mathbf{g}(\psi_0, \alpha_0) = (Y - \psi_0 X - \alpha_0)\mathbf{S}$, and $\mathbf{S} = (1, Z)'$. Similarly, for the multiplicative SMM it follows that

$$E \begin{bmatrix} Y \exp(-\psi_0 X) - \alpha_0 \\ \{Y \exp(-\psi_0 X) - \alpha_0\} Z \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{7}$$

and for the double-logistic SMM,

$$E\left[\begin{array}{c}\text{expit}(\beta_0+\beta_1X+\beta_2Z+\beta_3XZ-\psi_0X)-\alpha_0\\ \{\text{expit}(\beta_0+\beta_1X+\beta_2Z+\beta_3XZ-\psi_0X)-\alpha_0\}Z\end{array}\right]=\binom{0}{0}, \tag{8}$$

where in both cases $\alpha_0 = E(Y_0)$.

The estimators for these three models are trivial special cases of GMM because each is just-identified, but it is clear that moment conditions (6-7) are of the form $E\{v(\boldsymbol{\delta})\,\mathbf{S}\} = \mathbf{0}$, where $\mathbf{0}$ is an appropriately dimensioned vector of zeros. This generalises to over-identified models through different choices of $\mathbf{S}$ and specifications of $v(\boldsymbol{\delta})$. It is also clear that the moment condition for the double-logistic SMM has the more complicated form $E\{\mathbf{g}(\boldsymbol{\delta};\boldsymbol{\beta})\} = \mathbf{0}$, where $\boldsymbol{\beta}$ is the vector of association model parameters. In practice, this complicates variance estimation because $\boldsymbol{\beta}$ must be estimated, which we discuss in Section 4.2.

4 Multiple Instruments

Mendelian randomisation studies justify the use of genetic markers as instrumental variables by arguing that a) the random allocation of genes from parents to offspring mimics a randomised experiment, and b) there is an established relationship between the marker and some modifiable risk factor of interest; see e.g. Davey Smith and Ebrahim (2003) and Lawlor et al. (2008).

The genetic variant typically has three forms: homozygous for the common allele; and heterozygous and homozygous for the rare allele. If we code these 0, 1, and 2, respectively, then the resulting instrument Z is multivalued. In fact, this is a simple multiple instruments example because the three-level variable can be coded using two dichotomous variables; for example, $Z_1 = I(Z = 1)$ and $Z_2 = I(Z = 2)$, where I is the indicator function.

The additive SMM for multiple instruments in this case can be written as

$$E(Y|X,Z_1,Z_2) - E(Y_0|X,Z_1,Z_2) = (\psi_0+\psi_1Z_1+\psi_2Z_2)\,X$$

where NEM corresponds to constraining $\psi_1 = \psi_2 = 0$ and CMI yields the moment conditions

$$\left\{ \begin{array}{c} E(Y - \psi_0 X - \alpha_0) \\ E(Y - \psi_0 X - \alpha_0 | Z_1 = 1) \\ E(Y - \psi_0 X - \alpha_0 | Z_2 = 1) \end{array} \right\} = \left(\begin{array}{c} 0 \\ 0 \\ 0 \end{array} \right),$$

where $\alpha_0 = E(Y_0)$ as before. The unconditional moment condition is

$$E\{(Y - \psi_0 X - \alpha_0)\mathbf{S}\} = \mathbf{0},$$

where $\mathbf{S} = (1, Z_1, Z_2)'$ is a random vector representing the multiple instruments. In fact, this model is linear and so the parameters can be consistently estimated using standard Two-Stage Least Squares (2SLS). The 2SLS estimator can be obtained as the OLS estimator from the regression of Y on $\widehat{X}$, where $\widehat{X}$ is the prediction from the first-stage regression of X on $\mathbf{S}$. The 2SLS estimator is a special case of a 'one-step' GMM estimator with $W_n = n^{-1} \sum_i \mathbf{S}_i \mathbf{S}_i'$ (see next section), and is commonly used for linear instrumental variables analysis with multiple instruments; see Palmer et al. (2011) for its use with Medelian randomisation studies.

4.1 Multiplicative SMM

The saturated multiplicative SMM for the two instruments is

$$E(Y|X, Z_1, Z_2)/E(Y_0|X, Z_1, Z_2) = \exp\{(\psi_0 + \psi_1 Z_1 + \psi_2 Z_2)\, X\},$$

where NEM here corresponds to $\psi_1 = \psi_2 = 0$. Using the same vector of instrumental variables $\mathbf{S}$, there are three over-identified systems of multiplicative SMM moment conditions:

$$E\left[\left\{\frac{Y}{\exp(X\psi_0)} - \alpha_0\right\}\mathbf{S}\right] = \mathbf{0}, \tag{9}$$

$$E\left\{\frac{Y - \exp(\alpha_0^* + X\psi_0)}{\exp(X\psi_0)}\mathbf{S}\right\} = \mathbf{0}, \tag{10}$$

$$E\left\{\frac{Y - \exp(\alpha_0^* + X\psi_0)}{\exp(\alpha_0^* + X\psi_0)}\mathbf{S}\right\} = \mathbf{0}, \tag{11}$$

where $\alpha_0^* = \log(\alpha_0)$ and (11) is obtained by dividing (10) by $\exp(\alpha_0^*) \neq 0$. The last of these expressions equal the moment conditions for exponential mean models proposed by Mullahy (1997). For example, consider a GMM estimator based on moment conditions (9); the GMM estimator for $\boldsymbol{\delta} = (\alpha_0, \psi_0)'$ is the solution to (5) with $\mathbf{g}(\boldsymbol{\delta}) = \{Y \exp(-X\psi_0) - \alpha_0\}\mathbf{S}$.

There are two choices of weight matrix W_n to consider. A 'one-step' GMM estimator $\widehat{\boldsymbol{\delta}}_1$ is obtained by choosing an initial weight matrix such as $W_n = n^{-1}\sum_i \mathbf{S}_i\mathbf{S}_i'$. The 'two-step' GMM estimator $\widehat{\boldsymbol{\delta}}_2$ is obtained using

$$W_n\left(\widehat{\boldsymbol{\delta}}_1\right) = n^{-1}\sum_{i=1}^n \mathbf{g}_i\left(\widehat{\boldsymbol{\delta}}_1\right)\mathbf{g}_i'\left(\widehat{\boldsymbol{\delta}}_1\right),$$

that is, the weighting matrix W_n for the two-step GMM is estimated using the one-step GMM estimator $\widehat{\boldsymbol{\delta}}_1$. We will refer throughout to the one-step and two-step GMM estimators as those obtained using $W_n = n^{-1}\sum_i \mathbf{S}_i\mathbf{S}_i'$ as the initial weight matrix.

Under standard regularity conditions, the limiting distributions of the one-step and two-step GMM estimators are

$$\begin{aligned} n^{1/2}\left(\widehat{\boldsymbol{\delta}}_1 - \boldsymbol{\delta}_0\right) &\xrightarrow{d} N\left\{\mathbf{0}, \left(C_0'WC_0\right)^{-1}C_0W\Omega_0WC_0\left(C_0'WC_0\right)^{-1}\right\} \\ n^{1/2}\left(\widehat{\boldsymbol{\delta}}_2 - \boldsymbol{\delta}_0\right) &\xrightarrow{d} N\left\{\mathbf{0}, \left(C_0'\Omega_0C_0\right)^{-1}\right\}, \end{aligned}$$

respectively, where $\boldsymbol{\delta}_0$ is the true parameter value, $\xrightarrow{d}$ indicates convergence in distribution, N indicates a normally distributed random vector,

$$C_0 = E\left\{\frac{\partial \mathbf{g}(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'}\right\}, \Omega_0 = E\left\{\mathbf{g}(\boldsymbol{\delta}_0)\mathbf{g}'(\boldsymbol{\delta}_0)\right\},$$

and $W = E(\mathbf{S}_i\mathbf{S}_i')$ is the probability limit of the one-step GMM estimator's weight matrix.

Chamberlain (1987) shows that the two-step GMM estimator is asymptotically efficient because the instruments are mutually exclusive indicators that follow a multinomial distribution. The GMM estimators based on (10) and (11) have the same asymptotic distribution and efficiency, but will differ in finite samples for over-identified models.

A useful property of two-step GMM for over-identified models is that it admits the use of the Hansen J-test for the validity of the moment conditions; see Hansen (1982). The test statistic, and its limiting distribution under the null that the moment conditions are valid, are given by

$$J\left(\widehat{\boldsymbol{\delta}}_2\right) = n\left\{n^{-1}\sum_{i=1}^{n}\mathbf{g}_i'\left(\widehat{\boldsymbol{\delta}}_2\right)\right\}W_n^{-1}\left(\widehat{\boldsymbol{\delta}}_1\right)\left\{n^{-1}\sum_{i=1}^{n}\mathbf{g}_i\left(\widehat{\boldsymbol{\delta}}_2\right)\right\} \xrightarrow{d} \chi_q^2,$$

where χ_q^2 indicates a chi-square random variable with q degrees of freedom, and q is the number of parameters by which the model is over-identified (e.g. $q = 1$ in this illustration).

4.2 Logistic SMM

Under NEM, the logistic SMM for the two instruments is

$$\text{logit}\{E(Y|X, Z_1, Z_2)\} - \text{logit}\{E(Y_0|X, Z_1, Z_2)\} = \psi_0 X,$$

and the saturated association model for $E(Y|X, Z_1, Z_2)$ is specified as

$$\text{expit}\{m_{\boldsymbol{\beta}}(X, Z_1, Z_2)\} = \text{expit}(\beta_0 + \beta_1 X + \beta_2 Z_1 + \beta_3 Z_2 + \beta_4 X Z_1 + \beta_5 X Z_2). \tag{12}$$

Denoting $\widehat{\boldsymbol{\beta}}$ as the maximum likelihood estimator for the parameters of the association model, it follows that

$$E\{\mathbf{g}(\boldsymbol{\delta}; \widehat{\boldsymbol{\beta}})\} = E[\{q(\psi_0; \widehat{\boldsymbol{\beta}}) - \alpha_0\}\mathbf{S}] = \mathbf{0} \tag{13}$$

where $\boldsymbol{\delta} = (\psi_0, \alpha_0)'$, $q(\psi_0; \boldsymbol{\beta}) = \text{expit}\{m_{\boldsymbol{\beta}}(X, Z_1, Z_2) - X\psi_0\}$ and $\mathbf{S} = (1, Z_1, Z_2)'$. Point estimation is carried out exactly as before, but standard error estimates obtained by fixing $\widehat{\boldsymbol{\beta}}$ and plugging it into the asymptotic covariance matrices presented above will be biased because the first stage estimation of $\boldsymbol{\beta}$ is ignored. However, theory for 'two-stage' GMM estimators (2SGMM) has been developed by Gouriéroux et al. (1996). The 2SGMM $\widehat{\boldsymbol{\delta}}_{1,\boldsymbol{\beta}}$ is the solution to (5) and its asymptotic distribution is

$$n^{1/2}\left(\widehat{\boldsymbol{\delta}}_{1,\boldsymbol{\beta}} - \boldsymbol{\delta}_0\right) \xrightarrow{d} N\left\{\mathbf{0}, \left(C_0'WC_0\right)^{-1}C_0W\Omega_0^*WC_0\left(C_0'WC_0\right)^{-1}\right\},$$

where C_0 and W are both defined as above, and Ω_0^* is the asymptotic variance of the limiting normal distribution of

$$n^{-1/2}\sum_{i=1}^{n}\mathbf{g}_i\left(\boldsymbol{\delta}_0;\boldsymbol{\beta}_0\right)+E\left\{\frac{\partial\mathbf{g}\left(\boldsymbol{\delta}_0;\boldsymbol{\beta}_0\right)}{\partial\boldsymbol{\beta}'}\right\}n^{1/2}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right),$$

which has the consistent estimator

$$n\widehat{\Omega}^*=\sum_{i=1}^{n}\widehat{\mathbf{g}}_i\widehat{\mathbf{g}}_i'+\widehat{G}_{\boldsymbol{\beta}}'\widehat{V(\boldsymbol{\beta})}\widehat{G}_{\boldsymbol{\beta}}+\widehat{G}_{\boldsymbol{\beta}}'\widehat{V(\boldsymbol{\beta})}\left(\sum_{i=1}^{n}Q_i\mathbf{R}_i\widehat{\mathbf{g}}_i'\right)+\left(\sum_{i=1}^{n}Q_i\widehat{\mathbf{g}}_i\mathbf{R}_i'\right)\widehat{V(\boldsymbol{\beta})}\widehat{G}_{\boldsymbol{\beta}},$$

with $\widehat{\mathbf{g}}_i=\mathbf{g}_i(\widehat{\boldsymbol{\delta}}_{1,\boldsymbol{\beta}};\widehat{\boldsymbol{\beta}})$, $\widehat{G}_{\boldsymbol{\beta}}=\sum_i\partial\mathbf{g}_i'(\widehat{\boldsymbol{\delta}}_{1,\boldsymbol{\beta}};\widehat{\boldsymbol{\beta}})/\partial\boldsymbol{\beta}$, $\widehat{V(\boldsymbol{\beta})}=\left(\sum_i\widehat{p}_i\left(1-\widehat{p}_i\right)\mathbf{R}_i\mathbf{R}_i'\right)^{-1}$, $\mathbf{R}_i=(1,X_i,Z_{1i},Z_{2i},X_iZ_{1i},X_iZ_{2i})'$, $\widehat{p}_i=\text{expit}\{m_{\widehat{\boldsymbol{\beta}}}(X_i,Z_{i1},Z_{i2})\}$ and $Q_i=Y_i-\widehat{p}_i$. $\widehat{\Omega}^*$ is also the weight matrix for the asymptotically efficient two-step 2SGMM estimator, and so the limiting distribution of the Hansen J-test statistic (with $W_n=\widehat{\Omega}^*$) is also valid.

Vansteelandt and Goetghebeur (2003) developed estimating equations for the double-logistic SMM by expanding its system of estimating equations to include those for the association model. In the same spirit, a 'joint' GMM estimator can be obtained by applying the GMM estimator to

$$\mathbf{g}\left(\boldsymbol{\delta};\boldsymbol{\beta}\right)=\begin{pmatrix}\left[Y-\text{expit}\left\{m_{\boldsymbol{\beta}}\left(X,Z_1,Z_2\right)\right\}\right]\mathbf{R}\\ \left[\text{expit}\{m_{\boldsymbol{\beta}}\left(X,Z_1,Z_2\right)-\psi_0X-\alpha_0\right]\mathbf{S}\end{pmatrix},\tag{14}$$

where $\mathbf{R}$ is defined above and $\boldsymbol{\delta}=(\alpha_0,\psi_0)'$. Gouriéroux et al. (1996) show that the asymptotic distributions of the 2SGMM and the joint GMM estimators are the same. An important advantage of using the joint moments (14) is that standard GMM software can be used to make asymptotically correct inferences about the target parameter ψ_0. Further details on how the *gmm* command in Stata and the *gmm*() function in R can be used to implement these estimators are given in the Appendix.

4.3 A note on combining multiple intruments

The one-step GMM estimator combines multiple instruments in the following way. Note that the GMM estimator is the solution to the first derivative of the objective function

in (5) evaluated at zero. For the multiplicative SMM based on (9), this gives

$$\begin{aligned} & \left\{ n^{-1} \sum_{i=1}^{n} \frac{\partial \mathbf{g}_i'(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}} \right\} W_n^{-1} \left\{ n^{-1} \sum_{i=1}^{n} \mathbf{g}_i(\boldsymbol{\delta}) \right\} \\ = & \left\{ n^{-1} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ Y_i X_i \exp(-X_i \psi_0) \end{pmatrix} \mathbf{S}_i' \right\} W_n^{-1} \left\{ n^{-1} \sum_{i=1}^{n} \mathbf{g}_i(\boldsymbol{\delta}) \right\} = \mathbf{0}, \end{aligned}$$

where $\mathbf{g}_i(\boldsymbol{\delta}) = \{Y \exp(-X\psi_0) - \alpha_0\}\mathbf{S}$. This system can be expressed as

$$D'S(S'S)^{-1}S'\mathbf{v} = \mathbf{0},$$

where $D = \{\mathbf{d}_i'\}$ and $S = \{\mathbf{S}_i'\}$ are matrices obtained by stacking the vectors $\mathbf{d}_i' = (1, Y_i X_i \exp(-X_i\psi_0))$ and $\mathbf{S}_i'$, respectively, and $\mathbf{v} = \{v_i\}$ is a column vector with elements given by $v_i = Y_i \exp(-X_i\psi_0) - \alpha_0$. It is thus apparent that the GMM estimator combines the instruments in the projection $S(S'S)^{-1}S'D$, that is, the multiple instruments for each individual are replaced by the linear projection of $\mathbf{d}_i$ onto the space spanned by S; alternatively put, the combined instrumental variable can be thought of as the prediction from a linear regression of $\mathbf{d}_i$ on the instruments $\mathbf{S}_i$.

For the binary variables case considered here, we have that

$$Y_i X_i \exp(-X_i\psi_0) = Y_i X_i \exp(-\psi_0) \tag{15}$$

and therefore the one-step GMM combination of instruments is equivalent to the simple projection of YX onto the space spanned by S.

It also follows that an equivalent one-step GMM estimate of ψ_0 is obtained by specifying moment conditions based on $\mathbf{g}_i(\psi_0) = Y_i \exp(-X_i\psi_0)\widetilde{\mathbf{S}}_i$, where $\widetilde{\mathbf{S}}_i = \left(Z_{i1} - \overline{Z}_1, Z_{i2} - \overline{Z}_2\right)'$ and $\overline{Z}_j$ is the sample average of Z_j $(j = 1, 2)$. This is the transformation generally used for semi-parametric estimation of SMMs; see e.g. Vansteelandt and Goetgebheur (2003).

The logistic SMM estimator also has the form of a linear projection of $\mathbf{d}_i$ onto the space spanned by S, but here $\mathbf{d}_i' = \left(1, q_i(\psi_0; \widehat{\boldsymbol{\beta}})\left\{1 - q_i(\psi_0; \widehat{\boldsymbol{\beta}})\right\} X_i\right)$. In fact, for both

the multiplicative and logistic SMMs, these are the combinations of multiple instruments as proposed by Bowden and Vansteelandt (2011).

In the simple set-up involving only binary variables, the one-step GMM estimator for the multiplicative SMM can be expressed as a linear 2SLS estimator. Following Angrist (2001), note that $\exp(-\psi_0 X) = (1-X) + X\exp(-\psi_0)$ and therefore

$$Y\exp(-\psi_0 X) - \alpha_0 = Y(1-X) + YX\exp(-\psi_0) - \alpha_0.$$

Hence, the moment conditions can be expressed as the linear (in $\exp(-\psi_0)$) moments

$$E[\{Y(1-X) + YX\exp(-\psi_0) - \alpha_0\}]\mathbf{S} = \mathbf{0}, \tag{16}$$

from which we see that the one-step GMM estimator for $\exp(-\psi_0)$ using moment condition (9) is identical to the 2SLS estimator from regressing $Y(X-1)$ on $\widehat{YX}$, where $\widehat{YX}$ are the predictions from the linear regression of YX on S.

Multiplying (16) by the risk ratio $\exp(\psi_0)$, we obtain

$$E[\{YX + Y(1-X)\exp(\psi_0) - \gamma_0\}\mathbf{S}] = \mathbf{0} \tag{17}$$

where $\gamma_0 = \alpha_0\exp(\psi_0)$. In this case, the same estimator as the one-step GMM estimator for $\exp(\psi_0)$ is obtained from a linear instrumental variable estimator where $Y(X-1)$ is instrumented by $\widehat{YX}$. We will use this result later in Section 6 when deriving results for Local Risk Ratios.

5 Monte Carlo Studies

5.1 Multiplicative SMM

We now present two Monte Carlo simulation studies to demonstrate the properties of GMM estimators with multiple instruments. First, we consider the multiplicative SMM by generating data from population model M_1, which satisfies the multiplicative SMM

under both the NEM and CMI restrictions. Population model M_1 is defined so that

$$E(Y|X, Z_1, Z_2) = \exp\{\beta_0 + (\beta_1 + \psi_0) X + \beta_2 Z_1 + \beta_3 Z_2 + \beta_4 X Z_1 + \beta_5 X Z_2\},$$

where $\psi_0 = 0.6$ is the treatment effect. To define the distribution of the observed data, we further define Z to follow the marginal distribution given by $P(Z = 1) = 0.3$ and $P(Z = 2) = 0.2$, and $P(X = 1|Z = z) = p_{10} + 0.15 \times z$ for $z = 0, 1, 2$. To define the joint distribution of the observed and potential outcomes, we set the expected treatment-free outcome in the population to be $\alpha_0 = E(Y_0) = 0.19$, which leads to $\alpha_0^* = \log E(Y_0) = -1.6607$ in moment conditions (10) and (11), and $E(Y) = 0.25$, $\beta_1 = 0.15$, $\beta_4 = -0.6$ and $\beta_5 = 0.6$. The other parameter values are then numerically found in order for CMI and NEM to hold: $\beta_0 = -1.766$, $\beta_2 = -0.1307$, $\beta_3 = 0.0827$ and $p_{10} = 0.2321$.

Table 1. Monte Carlo estimation results for multiplicative SMM

	Single instrument	Multipe Instruments	
Instruments **S**	$1, Z$	$1, Z_1, Z_2$	
Moment conditions	(10) or (11)	(10)	(11)
One-Step GMM			
α_0^*	-1.6597	-1.6618	-1.6588
	(.0844)	(.0570)	(.0570)
	[.0844]	[.0566]	[.0566]
ψ_0	0.6153	0.6122	0.6053
	(.2177)	(.1376)	(.1371)
	[.2163]	[.1360]	[.1355]
Two-Step GMM			
α_0^*		-1.6619	-1.6588
		(.0570)	(.0570)
		[.0565]	[.0565]
ψ_0		0.6116	0.6045
		(.1373)	(.1369)
		[.1358]	[.1352]
Hansen J		.9895	.9885
rej freq, 5%		.0499	.0490

Notes: Sample size 10,000; means based on 10,000 Monte Carlo replications; std. eror in brackets; means of estimated standard errors in square brackets; data drawn from population model M_1 as described in Section 5.1; $\alpha_0^* = -1.6607$ and $\psi_0 = 0.6$.

Table 1 presents some estimation results for $10,000$ samples of size $10,000$ drawn from population model M_1. Three different versions of the GMM estimator are applied: the first column of Table 1 contains the results of the just-identified model using the multivalued instrument $Z \in \{0,1,2\}$ as a single instrument so that $\mathbf{S} = (1, Z)'$; in the second and third columns, we present the one- and two-step GMM estimates for moment conditions (10) and (11) respectivley, using multiple instruments so that $\mathbf{S} = (1, Z_1, Z_2)'$.

All of the estimators display a small positive bias for $\psi_0 = 0.6$, and the mean estimated standard errors are very close to the true standard errors. Among the two estimators using multiple instruments, this bias is slightly larger for the estimator based on moment condition (10). There is here a neglible gain in precision from using the two-step GMM estimator as compared to the one-step estimator. However, there is a substantial gain in efficiency from using two instrumental variables rather than one, with the standard error decreasing from 0.22 for the just-identified model to 0.14 for the two-step GMM estimators. This is because the GMM projection (15) in this case is not linear in Z, even though the conditional probablities $P(X = 1|Z)$ are. More specifically, the coefficient on Z_2 in the regression of YX on $(1, Z_1, Z_2)$ from (15) is actually smaller than that of Z_1. Under this particular population model (but not generally) the relationship between the coefficients is roughly linear: the average coefficient on Z_1 is equal to 0.1067 and for Z_2 it equals 0.0557. Hence, a single instrument that takes the value 1 if $Z = 2$ and 2 if $Z = 1$ leads to a just-identified estimator which is likely to be almost as efficient as the over-identified GMM estimators. Further simulations show that this is indeed the case, with the just-identified estimator for ψ_0 just described having an average of 0.6077 and a standard error of 0.1375, which are both virtually identical to those of the over-identified GMM estimators.

We repeated the analysis above for a similar design to M_1 but with the instrument Z taking the six values $0, 1, ..., 5$; full details of this design are available from the authors.

The GMM estimators are again well behaved. Using moment conditions (11), the mean based on 10,000 Monte Carlo estimates using the two-step GMM estimator is 0.5966 with a standard error 0.0801; the mean estimated standard error equals 0.0806. The rejection frequency of the J-test is 5.1% at the 5% level.

Returning to the design with Z taking the values $0, 1, 2$, we modify population model M_1 so as to study how the multiplicative GMM performs when Z does not satisfy the key conditions of an instrumental variable. We do this by keeping all M_1 parameters the same but making the "instrument" Z_1 invalid. This is done by specifying

$$E(Y|X, Z_1, Z_2) = \exp\{\beta_0 + (\beta_1 + \psi_0) X + (\beta_2 + \phi) Z_1 + \beta_3 Z_2 + \beta_4 X Z_1 + \beta_5 X Z_2\},$$

with $\phi = 0.15$. The GMM estimators are now severely biased upwards. The mean based on 10,000 Monte Carlo estimates of the two-step GMM estimator using moments (11) is equal to 1.1191, with a standard error of 0.1681. The mean (variance) of Hansen's J-test is equal to 3.56 (3.70) with a rejection frequency at the 5% level of 34%. If instead we change the coefficient on Z_2 to $\beta_3 + 0.15$, we get a much smaller bias, with the mean (std. error) of the estimator equal to 0.6452 (0.1370), but the rejection frequency of the J-test is now much larger, namely, 93% at the 5% level.

5.2 Logistic SMM

To consider the performance of the GMM estimators for the logistic SMM, we generate data from population M_2 satisfying the logistic SMM model and its corresponding NEM and CMI identification restrictions. More specifically, the data are generated from

$$E(Y|X, Z_1, Z_2) = \text{expit}\{\beta_0 + (\beta_1 + \psi_0) X + \beta_2 Z_1 + \beta_3 Z_2 + \beta_4 X Z_1 + \beta_5 X Z_2\},$$

where the treatment effect is again $\psi_0 = 0.6$. Similarly to model M_1, we set $P(Z = 1) = 0.3$, $P(Z = 2) = 0.2$, $P(X = 1|Z = z) = p_{10} + 0.15 \times z$, $E(Y_0) = 0.19$, $E(Y) = 0.25$, $\beta_1 = 0.15$, $\beta_4 = -0.6$ and $\beta_5 = 0.6$. The other parameters are such that CMI and NEM hold: $\beta_0 = -1.518$, $\beta_2 = 0.3183$, $\beta_3 = -0.5202$, and $p_{10} = 0.4404$.

Table 2. Monte Carlo estimation results for logistic SMM

	Single instrument	Multiple instruments	
Instruments $\mathbf{S}$	$1, Z$	$1, Z_1, Z_2$	$1, Z_1, Z_2$
Moment conditions	Joint/2SGMM	2SGMM	Joint-GMM
One-Step GMM			
α_0	0.1912	0.1905	0.1907
	(.0168)	(.0153)	(.0153)
	[.0167]	[.0152]	[.0152]
ψ_0	0.5970	0.6033	0.6001
	(.1905)	(.1729)	(.1731)
	[.1899]	[.1722]	[.1721]
Two-Step GMM			
α_0		0.1904	0.1911
		(.0153)	(.0154)
		[.0152]	[.0152]
ψ_0		0.6038	0.5957
		(.1729)	(.1735)
		[.1722]	[.1722]
Hansen J		0.9882	0.9827
rej-freq 5%		0.0503	0.0495

Notes: Sample size 10,000; means based on 10,000 Monte Carlo replications; std. [error] in brackets; means of estimated standard errors in square brackets; data drawn from population model M_2 as described in Section 5.2; α_0= 0.19 and ψ_0= 0.6.

Table 2 contains estimation results for $10,000$ samples of size $10,000$ drawn from population model M_2. Three different versions of the GMM estimator for the logistic SMM are applied: the first column of Table 1 contains the results of the just-identified model using multivalued Z as a single instrument; in the second column, we present the one- and two-step GMM estimates for the 2SGMM using multiple instruments; and the third column contains the corresponding results for the joint-GMM estimator based on (14). Both the 2SGMM and joint-GMM estimators use saturated logistic models for $\boldsymbol{\beta}$ as in (12)

All of the estimators are virtually unbiased and the means of the estimated standard errors are close to Monte Carlo standard errors. There is an efficiency gain from using the instruments separately: the standard error in the just identified case is 0.1905, compared

to 0.1729 for the 2SGMM estimator. The performances of the 2SGMM estimator and the GMM estimator using the joint moment conditions are virtually identical. The Hansen J-tests are well behaved in both cases. There is no efficiency gain from using the two-step GMM estimators as compared to the one-step estimators in this design.

As with the multiplicative SMM, we also find that the estimators behave well for instruments with 6 or even 11 values, although we find that the 2SGMM estimator has a small upward bias in the designs we considered. For example, for an instrument with values $0, 1, 2, ..., 10$, we get means (std. error) of the two-step GMM estimates of 0.6323 (0.1073) for 2SGMM and 0.5999 (0.1066) for the joint moments GMM estimator. Details of this design are available from the authors.

Finally, we return to the design with Z taking the values $0, 1, 2$, and modify population model M_2 so as to study how these estimators perform when Z is not a valid instrumental variable. We keep all parameters the same but make the "instrument" Z_2 invalid, by changeing the parameter of Z_2 to $\beta_3 + \tau$ with $\tau = 0.25$. The GMM estimators are now severely biased upwards. The mean of 10,000 Monte Carlo estimates of the two-step GMM estimator using the joint moments (14) is equal to 1.2805, with a standard error of 0.1511. However, in this case the mean (variance) of Hansen's J-test is equal to 1.26 (3.09), with a rejection frequency at the 5% level of only 8.5%. In contrast, if we instead change the parameter of Z_1 to $\beta_2 + \tau$ with $\tau = 0.1$, the estimator has a much smaller bias, with a mean of 0.5527 and standard error of 0.1660, but the J-test has much more power in this case as it rejects 49.4% of the time at the 5% level.

6 Local Average Treatment Effects

The parameters of the SMMs we have considered thus far are all identified by the assumption of no effect modification by the instruments (NEM). For the case where we have two instruments Z_1 and Z_2, recall that the NEM assumption for the identification

of the conditional causal relative risk is that

$$\frac{E(Y|X,Z_1,Z_2)}{E(Y_0|X,Z_1,Z_2)} = \exp\left(\psi_0 X\right),$$

i.e., the instruments Z_1 and Z_2 do not modify the causal effect of X on the risk. In this section, we consider how the failure of NEM impacts on GMM estimators for additive and multiplicative SMMs with multiple instruments.

Clarke and Windmeijer (2010) review identification results concerning the additive and multiplicative SMMs in the simple case of a single binary instrument where both X and Y are also binary. If the NEM assumption fails then a causal effect is identified if selection is 'monotonic'. In this simple case, where Z is randomised treatment assignment and X is the selected treatment, selection is monotonic if

$$P(X_1 - X_0 \geq 0) = 1,$$

that is, subjects cannot defy their treatment assignments in every potential scenario, so that $\{X_1 = 0, X_0 = 1\}$ has zero probability. Under monotonicity, the additive SMM estimator (4) identifies the 'local average treatment effect' (LATE), and the multiplicative SMM identifies the 'local risk ratio' (LRR), where

$$\text{LATE} = E(Y_1 - Y_0|X_1 > X_0); \qquad \text{LRR} = \frac{E(Y_1|X_1 > X_0)}{E(Y_0|X_1 > X_0)}.$$

LATE is the average treatment effect for the subgroup of subjects who actually and counterfactually accept the treatments to which they have been assigned, that is, $X_1 = 1$ and $X_0 = 0$; for this reason, these subjects are also known as 'compliers' and LATE is also known as the 'complier average causal effect' (CACE). The logistic SMM does not estimate a local causal effect when NEM fails, but for binary outcomes the local odds ratio can be estimated by taking the ratio of LRR estimates obtained by fitting multiplicative SMMs to binary Y and $1 - Y$.

If we have two instruments, then these instruments could in principle define two different local causal effects, provided that the two instruments can be combined into a

single multivalued instrument. We consider using the single K-valued instrument $Z \in \{0, 1, 2, ..., K-1\}$ for binary X. In this scenario, monotonic selection does not have the convenient 'no defiers' interpretation; instead, selection is monotonic if $z > \widetilde{z}$ implies that $X_z \geq X_{\widetilde{z}}$ with probability 1, for any two values $z \neq \widetilde{z}$ of the instrument. From this, we can define the analogue of (4) for $z > \widetilde{z}$ as

$$\beta_{z,\widetilde{z}} = \frac{E(Y|Z=z) - E(Y|Z=\widetilde{z})}{E(X|Z=z) - E(X|Z=\widetilde{z})},$$

where $\beta_{z,\widetilde{z}} = E(Y_1 - Y_0|X_z > X_{\widetilde{z}}) \equiv \text{LATE}_{z,\widetilde{z}}$ under monotonicity.

The 2SLS estimator for the additive SMM is obtained as the OLS estimator from the regression of Y on $\widehat{X}$, where $\widehat{X}$ is the prediction from the first-stage regression of X on $\mathbf{S} = \{1, Z_1, ..., Z_{K-1}\}'$ and $Z_k = I(Z = k)$. Let monotonicity hold and the values of Z be ordered such that $E(X|Z=k) > E(X|Z=k-1)$. Imbens and Angrist (1994) show that the 2SLS estimator is consistent for

$$\beta_z = \sum_{k=1}^{K-1} \mu_k \beta_{k,k-1}$$

where

$$\mu_k = \{E(X|Z=k) - E(X|Z=k-1)\} \frac{\sum_{l=k}^{K-1} \{E(X|Z=l) - E(X)\}\pi_l}{\sum_{l=0}^{K-1} E(X|Z=l)\{E(X|Z=l) - E(X)\}\pi_l},$$

and $\pi_l = P(Z=l)$ such that $0 \leq \mu_k \leq 1$ and $\sum_{l=1}^{K-1} \mu_k = 1$; see also Angrist and Imbens (1995) and Angrist and Pischke (2009). In other words, when NEM fails but selection is monotonic, the 2SLS estimator is not consistent for $E(Y_1 - Y_0|X=1)$, but for a weighted sum of local average treatment effects.

Alternatively, if we define

$$\beta_{k,0} = \frac{E(Y|Z=k) - E(Y|Z=0)}{E(X|Z=k) - E(X|Z=0)},$$

then, following the proof given by Angrist and Imbens (1995), it is easily established that

$$\beta_z = \sum_{k=1}^{K-1} \lambda_k \beta_{k,0},$$

where

$$\lambda_k = \{E(X|Z=k) - E(X|Z=0)\}\frac{\{E(X|Z=k) - E(X)\}\pi_k}{\sum_{l=0}^{K-1} E(X|Z=l)\{E(X|Z=l) - E(X)\}\pi_l},$$

such that $\sum_{l=1}^{K-1} \lambda_k = 1$. However, in this case, β_z is only a weighted average of the $\beta_{k,0}$ (i.e., $0 \leq \lambda_k \leq 1$) if $E(X|Z=1) > E(X)$.

We now extend this result to the multiplicative SMM and give an analogous result for local risk ratios. In Section 4.3 we established that the one-step GMM estimator for $\exp(-\psi_0)$ using moment condition (9) was equivalent to a linear 2SLS estimator because

$$Y\exp(-X\psi_0) - \alpha_0 = Y(1-X) + YX\exp(-\psi_0) - \alpha_0.$$

We can therefore straightforwardly generalise the above results of Imbens and Angrist (1994) for the additive SMM to the multiplicative SMM for the inverse local risk ratio. As above, let

$$e_{k,k-1}^{-\beta} = \frac{E\{Y(X-1)|Z=k\} - E\{Y(X-1)|Z=k-1\}}{E(YX|Z=k) - E(YX|Z=k-1)},$$

where

$$e_{k,k-1}^{-\beta} = \frac{E(Y_0|X_k > X_{k-1})}{E(Y_1|X_k > X_{k-1})} \equiv \text{ILRR}_{k,k-1},$$

is the *inverse* local risk ratio under monotonicity; see Angrist (2001). We then get equivalent results to the above for the linear SMM, namely,

$$e_z^{-\beta} = \sum_{k=1}^{K-1} \mu_k e_{k,k-1}^{-\beta},$$

where

$$\mu_k = \{E(YX|Z=k) - E(YX|Z=k-1)\}\frac{\sum_{l=k}^{K-1}\{E(YX|Z=l) - E(YX)\}\pi_l}{\sum_{l=0}^{K-1} E(YX|Z=l)\{E(YX|Z=l) - E(YX)\}\pi_l},$$

which is a weighted average if $E(YX|Z=k) > E(YX|Z=k-1)$.

For the local risk ratio, we use the results from Section 4.3 that the one-step GMM estimator for $\exp(\psi_0)$ can be obtained from a linear IV estimator in the additive SMM

with YX as the 'outcome' and $Y(X-1)$ as the 'treatment', but with instruments a constant and $E(YX|S)$. Let

$$e^{\beta}_{k,k-1} = \frac{E(YX|Z=k) - E(YX|Z=k-1)}{E\{Y(X-1)|Z=k\} - E\{Y(X-1)|Z=k-1\}},$$

where $e^{\beta}_{k,k-1} = E(Y_1|X_k > X_{k-1})/E(Y_0|X_k > X_{k-1}) \equiv \text{LRR}_{k,k-1}$ under monotonicity. It follows that

$$e^{\beta}_{z} = \sum_{k=1}^{K-1} \tau_k e^{\beta}_{k,k-1},$$

where

$$\begin{aligned}\tau_k &= \{E(Y(X-1)|Z=k) - E(Y(X-1)|Z=k-1)\} \\ &\quad \times \frac{\sum_{l=k}^{K-1}\{E(YX|Z=l) - E(YX)\}\pi_l}{\sum_{l=0}^{K-1} E\{Y(X-1)|Z=l\}\{E(YX|Z=l) - E(YX)\}\pi_l},\end{aligned}$$

is a weighted average of local risk ratios if $E(YX|Z=k) > E(YX|Z=k-1)$ and $E\{Y(X-1)|Z=k\} > E\{Y(X-1)|Z=k-1\}$.

As an example, consider an instrument that takes the values $Z = \{0,1,2,3\}$, with Y and X generated form a bivariate normal distribution as

$$\begin{aligned} X &= I(c_0 + c_1 Z_1 + c_2 Z_2 + c_3 Z_3 - V > 0), \\ Y &= I(b_0 + b_1 X - U > 0), \\ \begin{pmatrix} U \\ V \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right), \end{aligned}$$

with, as before, $Z_k = I(Z=k)$. Setting $\pi_l = P(Z=l) = 0.25$ for all l; the c_l parameters are such that $P(X=1|Z=l) = 0.1 + 0.1 \times l$; $b_0 = \Phi^{-1}(0.4)$; $b_1 = 0.5$ and $\rho = 0.8$. The local risk ratios in this population are $\text{LRR}_{1,0} = 1.1585$, $\text{LRR}_{2,1} = 1.3227$ and $\text{LRR}_{3,2} = 1.5303$; the population τ-weights are

$$\tau_1 = 0.3725,\ \tau_2 = 0.3991,\ \tau_3 = 0.2285.$$

Clarke and Windmeijer (2010) show that the NEM assumption does not hold under this design, and so the one-step GMM estimator based on moment conditions (9) identifies

the weighted average $\tau_1 \,\mathrm{LRR}_{1,0} + \tau_2 \,\mathrm{LRR}_{2,1} + \tau_3 \,\mathrm{LRR}_{3,2} = 1.3090$. Table 3 presents some estimation results confirming this, for a sample of size 40,000 and for 10,000 Monte Carlo replications. Using the two-step GMM results, the Hansen J-test rejects the null 47% of the time at the 5% level, therefore clearly having power to reject this violation of the NEM assumption.

Table 3. Risk ratio estimation results

	$e^{\beta}_{1,0}$	$e^{\beta}_{2,1}$	$e^{\beta}_{3,2}$	e^{β}_{z}	τ_1	τ_2	τ_3
mean	1.1644	1.3304	1.5415	1.3113	0.3726	0.3995	0.2279
st. dev.	0.0946	0.1213	0.1601	0.0377	0.0268	0.0321	0.0216

Notes: Estimation results from 10,000 MC replications. Sample size 40,000.

7 The Effect of Adiposity on Hypertension

Timpson et al. (2009) used multiple genetic instruments to estimate the causal effect of adiposity on hypertension from the Copenhagen General Population Study; full details of the variable definitions and selection criteria are given in that paper. We apply the procedures described above to reanalyse these data using additive, multiplicative and logistic SMMs, using the same genetic markers as instruments for adiposity. Furthermore, our sample includes additional individuals who have been recruited into the study since the previous study was published; the total number of individuals in our analyses is 55,523.

The binary outcome variable is an indicator of whether an individual has hypertension, which is defined as a systolic blood pressure of >140 mmHg, diastolic blood pressure of > 90 mmHg, or the taking of antihypertensive drugs. The intermediate adiposity phenotype is being overweight, defined as having a BMI>25. The two Single Nucleotide Polymorphisms (SNPs) that were used as instruments by Timpson et al. (2009) and that have been consistently shown to relate to BMI and adiposity are the *FTO* (rs9939609) and *MC4R* (rs17782313) loci; see Frayling et al. (2007) and Loos et al. (2008). Lawlor

et al. (2008) provide further details on the use of genes as instruments in Mendelian Randomisation studies.

FTO is specified as having three categories: no risk alleles (homozygous TT), one risk allele (heterozygous AT) and two risk alleles (homozygous AA). Due to the nature of the association between *MC4R* and adiposity (a dominant genetic model), *MC4R* is specified as having two categories: no risk alleles (TT) versus one or two risk alleles (CT or CC). Combining the two instruments together results in an instrument with 6 different values, but we found that two pairs of combinations of alleles gave the same predicted value of being overweight, also for the projection in the multiplicative SMM, and we therefore condensed the number of values of the instrument to four. The combinations for the four values are given in Table 4. Table 5 gives the frequency distributions for the hypertension (Y) and overweight (X) variables.

Table 4. Combinations of instruments

FTO	*MC4R*	Z	Freq
0	0	0	0.20
0	1	1	0.15
1	0	1	0.27
1	1	2	0.21
2	0	2	0.09
2	1	3	0.07

Table 5. Frequency distributions for the hypertension (Y) and overweight (X) variables

	All		$Z=0$		$Z=1$		$Z=2$		$Z=3$	
	X		X		X		X		X	
Y	0	1	0	1	0	1	0	1	0	1
0	0.18	0.12	0.19	0.12	0.19	0.12	0.17	0.13	0.16	0.13
1	0.25	0.44	0.27	0.42	0.26	0.43	0.23	0.46	0.23	0.48

The estimation results for the linear, multiplicative and logistic SMM estimators are presented in Table 6. The instrument set for the GMM estimators is $\mathbf{S} = (1, Z_1, Z_2, Z_3)'$. For the linear SMM, the 2SLS and two-step GMM estimates are virtually identical to

the OLS estimate. As the F-statistic in the regression of overweight on **S** is equal to 113, this is not due to a weak instrument problem and therefore indicates that there is no unobserved confounding bias. The estimate of the risk difference is quite large and equal to 0.20. The J-test does not reject the null of the validity of the model assumptions, including the NEM assumption. We find similar results for the multiplicative and logistic SMMs. There is no indication of biases due to unobserved confounding, as the GMM estimates are virtually identical to the Gamma and the logistic regression estimates respectively, and all estimates indicate that being overweight leads to hypertension. The Gamma estimate for the risk ratio is equal to 1.3464 (95% CI, 1.3300-1.3630), whereas the logistic regression odds ratio is equal to 2.5823 (95% CI, 2.4885-2.6797). We present and compare the multiplicative SMM results to that of the Gamma with log link here, because moment conditions (9)-(11) when using X as an instrument for itself are equivalent to the first-order condition of the Gamma with log link GLM.

Table 6. SMM estimation results of the effect of being overweight on hypertension

Additive	OLS	2SLS	GMM2	J-test
ψ_0	0.2009 (0.0039)	0.2091 (0.0819)	0.2094 (0.0819)	0.2965
Multiplicative	Gamma	GMM1	GMM2	J-test
ψ_0	0.2974 (0.0063)	0.3090 (0.1192)	0.3104 (0.1192)	0.3071
Logistic	Logistic regression	GMM1	GMM2	J-test
ψ_0	0.9487 (0.0189)	1.0409 (0.4220)	1.0528 (0.4217)	0.2924

Notes: Sample size 55,523. Gamma regression uses log link;
Multiplicative SMM uses moments (9);
logistic SMM uses joint moments (14); Instruments, $S = \{1, Z_1, Z_2, Z_3\}$;
Standard errors in brackets; p-values are reported for the J-test.

Although the J-test results do not indicate that the NEM assumptions are not valid, we present in Table 7 the local risk ratio estimation results as described in Section 6.

The most precisely estimated risk ratio is $e^{\beta}_{2,1} = LRR_{2,1}$ which gets the largest weight, $\tau_2 = 0.81$.

Table 7. Local Risk Ratio estimation results

	$e^{\beta}_{1,0}$	$e^{\beta}_{2,1}$	$e^{\beta}_{3,2}$	e^{β}_{z}	τ_1	τ_2	τ_3
Coeff	2.2065	1.1086	2.6935	1.3621	0.1037	0.8082	0.0881
95% CI	0.548-8.884	0.791-1.553	0.588-12.336	1.078-1.720			
Sample Size	34,896	40,552	20,627	55,523			

7.1 Continuous Exposure

Following Vansteelandt and Goetghebeur (2003), we can use the same GMM format to estimate the logistic SMM with a continuous exposure X. With a continuous exposure, parametric modelling assumptions have to be made in order to identify causal parameters. As in Vansteelandt and Goetghebeur (2003) and Vansteelandt et al. (2011), we impose that the exposure effect is linear in the exposure on the odds ratio scale and independent of the instrumental variable:

$$\frac{\text{odds}\,(Y=1|X,Z)}{\text{odds}(Y_0=1|X,Z)} = \exp\left(\xi_0 X\right),$$

where $\text{odds}\,(Y=1|X,Z) = P(Y=1|X,Z)/P(Y=0|X,Z)$. Further, we specify the association model as

$$\begin{aligned}\text{logit}\left\{P(Y=1|X,Z)\right\} &= \text{logit}\left\{m_\beta(X,Z_1,Z_2,Z_3)\right\} \\ &= \beta_0+\beta_1 X+\beta_2 Z_1+\beta_3 Z_2+\beta_4 Z_3+\beta_5 XZ_1+\beta_6 XZ_2+\beta_7 XZ_3,\end{aligned}$$

and estimate the parameters using the joint moment conditions as in (14).

For the continuous exposure we use $\left(BMI-\overline{BMI}\right)$, $10\left(\ln BMI-\overline{\ln BMI}\right)$ and $10\left(\ln RELBMI\right)$, where $\ln BMI$ is the natural logarithm of BMI, and $\ln RELBMI$ are the residuals of the regression of $\ln BMI$ on sex, age, age squared, ln(height) and

an age-sex interaction, as used in Timpson et al. (2009) to represent relative BMI. We subtract the mean from BMI and $\ln BMI$ to ensure that zero exposure is part of the data range. We further multiply the $\ln BMI$ and $\ln RELBMI$ by a factor 10 so that the estimated odds ratio is for an increase in exposure of aproximately 10%.

Table 8 presents the two-step estimation results for the three exposure measures. Again, we find a strong positive effect of adiposity on hypertension, with the effects of $\ln BMI$ and $\ln RELBMI$ virtually identical.

Table 8. Estimation results for double-logistic SMM with continuous exposure

Exposure	BMI	$\ln BMI$	$\ln RELBMI$
ξ_0	0.1122	0.3035	0.2879
	(0.0384)	(0.1069)	(0.1016)
J-test	0.4714	0.4828	0.5004

Notes: Sample size 55,523. Two-step GMM estimtates, using joint moments (14). Instruments, $\mathbf{S}=\{1, Z_1, Z_2, Z_3\}$. BMI and $\ln BMI$ taken in deviation from the mean. $\ln BMI$ and $\ln RELBMI$ multiplied by a factor 10. Standard errors in brackets; p-values are reported for the J-test.

8 Conclusions

We have shown how the moment conditions that identify additive, multiplicative and logistic SMMs can be formulated such that the causal parameters can be estimated by a standard GMM estimator of the type widely used in econometrics. The key to this formulation is simply to treat $E(Y_0)$ as a parameter to be estimated directly, from which estimators using multivalued and multiple instrumental variables can be straightforwardly derived. For discrete instrumental variables, these estimators are consistent and fully efficient without having to centre the instruments, as is commonly done using other estimating equation-based approaches such as G-estimation. Another major advantage is that standard GMM routines are available in statistical software packages.

We give example Stata and R syntax in the Appendix below, for easy use by applied researchers. These estimation routines provide correct asymptotic inference, even for the logistic SMM when the two sets of model parameters are estimated jointly.

We have also found in some Monte Carlo analyses that the Hansen J-test has power to detect violations of the CMI and NEM assumptions. Moreover, if the NEM assumption fails and selection is monotonic, then we have shown that the one-step GMM estimator for the multiplicative SMM is consistent for a weighted average of the instrument specific local risk ratios.

Although we have concentrated on relatively simple SMMs in this paper, the class of GMM estimators we propose enables efficient estimation of a more general class of SMMs where the treatment X is a random variable with a finite countable or compact support, and pre-exposure covariates $\mathbf{C}$ are available. The GMM estimator can thus fit generalised SMMs of the form

$$h\{E(Y|X,Z,\mathbf{C})\} - h\{E(Y_0|X,Z,\mathbf{C})\} = \eta_{\boldsymbol{\psi}}(X,Z,\mathbf{C}),$$

where $\boldsymbol{\psi}$ is the finite-dimensional SMM parameter, $\eta_{\boldsymbol{\psi}}(X,Z,\mathbf{C})$ is subject to $\eta_{\boldsymbol{\psi}}(0,Z,\mathbf{C}) = 0$; NEM for these models corresponds to the assumption that $\eta_{\boldsymbol{\psi}}(X,Z,\mathbf{C}) = \eta_{\boldsymbol{\psi}}(X,\mathbf{C})$. Introducing further variables necessitates semi-parametric modelling assumptions to avoid the curse of dimensionality. For example, introducing continuous X and $\mathbf{C}$ we may choose the SMM $\eta_{\boldsymbol{\psi}}(X,Z,\mathbf{C}) = \psi_0 + X\psi_1 + X^2\psi_2 + I(X \neq 0)\mathbf{C}'\boldsymbol{\psi}_3$. For this model to hold, it must be assumed that the covariates each have linear effects, and the exposure X has a quadratic effect, on the scale of the link function h, and that the quadratic exposure effect is the same given $\mathbf{C}$. For double-logistic SMMs, this also necessitates semi-parametric assumptions for the association model.

Tan (2010) notes that NEM is not crucial for identification in these more complex scenarios, provided that alternative plausible semi-parametric assumptions are available that identify the SMM parameters; see also Vansteelandt and Goetghebeur (2005). Tan

(2010) also proposes alternative GMM estimators for generalised SMMs designed explicitly to address the problems posed by mis-specification of semi-parametric modelling assumptions. Rather than using the standard GMM formulation from econometrics, he constructs estimating equations that are doubly robust and applies classical results from Hansen (1982). These GMM estimators are doubly robust in the sense of remaining consistent if one but not both of the following user-specified models are mis-specified: a) the instrument propensity score $P(Z|\mathbf{C})$; and b) both the treatment propensity score $P(X|Z,\mathbf{C})$ and association model $m_{\boldsymbol{\beta}}(X,Z,\mathbf{C})$. The double robustness property is attractive, but these estimators are not available in standard software, and further work is required to explore fully, rather than locally, efficient choices of weights for the estimating equations.

Acknowledgements

This work was funded by UK Economic & Social Research Council grant RES-060-23-0011, UK Medical Research Council grants G0601625 and G0600705, and European Research Council grant 269874 - DEVHEALTH. The authors would like to thank to thank Borge Nordestgaard for access to the Copenhagen General Population Study data. We also thank George Davey Smith, Nicholas Timpson, Vanessa Didelez, Roger Harbord, Nuala Sheehan and conference participants in London and Mannheim for helpful comments.

References

[1] Angrist, J.D. (2001), Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice, Journal of Business and Economic Statistics 19, 2-16.

[2] Angrist, J.D. and Imbens, G.W. (1995), Two-Stage Least Squares estimation of average causal effects in models with variable treatment intensity, Journal of the

American Statistical Association 90, 431-442.

[3] Angrist, J.D., Imbens, G.W. and Rubin, D.B. (1996), Identification of causal effects using instrumental variables, Journal of the American Statistical Association 91, 444-455.

[4] Bowden, J. and Vansteelandt, S. (2011), Mendelian randomisation analysis of case-control data using Structural Mean Models, Statistics in Medicine 30, 678-694.

[5] Chamberlain, G. (1987), Asymptotic efficiency in estimation with conditional moment conditions, Journal of Econometrics 34, 305-334.

[6] Chaussé, P. (2010), Computing Generalized Method of Moments and Generalized Empirical Likelihood with R, Journal of Statistical Software 34, 1-35.

[7] Clarke, P.S. and Windmeijer, F. (2010), Identification of causal effects on binary outcomes using Structural Mean Models, Biostatistics 11, 756-770.

[8] Davey Smith, G., and Ebrahim S. (2003), 'Mendelian Randomization': Can genetic epidemiology contribute to understanding environmental determinants of disease?, International Journal of Epidemiology 32, 1-22.

[9] Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., McCarthy, M.I., et al. (2007), A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity, Science 316, 889-94.

[10] Goetghebeur, E. and Vansteelandt, S. (2005), Structural mean models for compliance analysis in randomized clinical trials and the impact on measures of exposure, Statistical Methods in Medical Research 14, 397-415.

[11] Gouriéroux, C., Monfort, A. and Renault, E. (1996), Two-stage generalized moment method with applications to regressions with heteroscedasticity of unknown from, Journal of Statistical Planning and Inference 50, 37-63.

[12] Hansen, L.P. (1982), Large sample properties of generalized method of moments estimators, Econometrica 50, 1029-1054.

[13] Hernán, M.A. and Robins, J.M. (2006), Instruments for causal inference: an epidemiologist's dream?, Epidemiology 17, 360-372.

[14] Imbens, G.W. and Angrist, J. (1994), Identification and estimation of local average treatment effects, Econometrica 62, 467-476.

[15] Lawlor, D.A., Harbord, R.M., Sterne, J.A., Timpson, N. and Davey Smith, G. (2008). Mendelian Randomization: Using genes as instruments for making causal inferences in epidemiology, Statistics in Medicine 27, 1133-63.

[16] Loos, R.J.F., Lindgren, C.M., Li, S., Wheeler, E., Barosso, I. et al. (2008). Common variants near MC4R are associated with fat mass, weight and risk of obesity, Nature Genetics 40, 768-75.

[17] Mullahy, J. (1997), Instrumental variable estimation of Poisson regression models: application to models of cigarette smoking behavior, Review of Economics and Statistics 79, 586-593.

[18] Palmer, T.M., Lawlor, D.A., Harbord, R.M., Sheehan, N.A., Tobias, J.H., Timpson, N.J., Davey Smith, G. and Sterne, J.A.C. (2011), Using multiple genetic variants as instrumental variables for modifiable risk factors, Statistical Methods in Medical Research (Advance Access: January 7, 2011, DOI: 10.1177/0962280210394459).

[19] Robins, J.M. (1989), The analysis of randomised and non-randomised AIDS treatment trials using a new approach to causal inference in longitudinal studies,

in: Sechrest, L., Freeman, H. and Mulley, A. (eds.), Health Service Research Methodology: A Focus on AIDS, 113-159, Washington, DC: US Public Health Service, National Center for Health Services Research.

[20] Robins, J.M. (1994), Correcting for non-compliance in randomized trials using structural nested mean models, Communications in Statistics - Theory and Methods 23, 2379-2412.

[21] Robins, J.M. (1999), Marginal structural models versus structural nested models as tools for causal inference, in: Halloran, E. and Berry, D. (eds.), Statistical Methods in Epidemiology: The Environment and Clinical Trials, 95-134, New York: Springer.

[22] Robins, J.M. and Rotnitzky, A. (2004). Estimation of treatment effects in randomised trials with non-compliance and a dichotomous outcome using structural mean models, Biometrika 91, 763-783.

[23] Tan, Z. (2010), Marginal and nested structural models using instrumental variables, *Journal of the American Statistical Association* 105, 157-169.

[24] Timpson, N.J., Harbord, R., Davey Smith, G., Zacho, J., Tybjærg-Hansen, A. and Nordestgaard, B.G. (2009), Does greater adiposity increase blood pressure and hypertension risk? Mendelian randomisation using the *FTO/MC4R* genotype, Hypertension 54, 84-90.

[25] Tsiatis, A.A. (2006), *Semiparametric Theory and Missing Data*, Springer: New York.

[26] van der Laan, M.J., Hubbard, A. and Jewell, N.P. (2007). Estimation of treatment effects in randomized trials with non-compliance and a dichotomous outcome. Journal of the Royal Statistical Society, Series B 69, 463-482.

[27] Vansteelandt, S. and Goetghebeur, E. (2003), Causal inference with generalized structural mean models, Journal of the Royal Statistical Society, Series B 65, 817-835.

[28] Vansteelandt, S. and Goetghebeur, E. (2005), Sense and sensitivity when correcting for observed exposures in randomized controlled trials, Statistics in Medicine 24, 191-210.

[29] Vansteelandt, S., Bowden, J. Babanezhad, M. and Goetghebeur, E. (2011), On instrumental variables estimation of causal odds ratios, Statistical Science (in press).

Appendix: Stata and R syntax

In this section we present example Stata (version 11) and R (version 2.13.1) syntax to fit SMMs using generalised method of moments routines. Our example code uses the notation of Y the outcome, X the exposure, and two instrumental variables, Z_1, Z_2, in addition to the constant vector of 1s. Both syntaxes easily generalise to more instruments, and allow different association models in the double logistic SMM.

In both Stata and R it is possible to specify analytic first derivatives, which we find greatly reduces the time for the models to fit. Also both syntaxes allow the inclusion of covariates. We have not included these extra syntaxes here but they are available on request.

Stata syntax

The Stata syntax uses the `gmm` command; and `{ey0}` denotes $E(Y_0)$ the mean exposure free potential outcome. After fitting each SMM using two-step estimation we perform the Hansen over-identification test using the `estat overid` post-estimation command. The `gmm` command automatically includes a vector of 1s as instruments to allow estimation of the constant ($E(Y_0)$) term, hence we just need to list `z1` and `z2` in the `instruments()` option.

Additive SMM

Here `{psi}` denotes the causal effect (which is a risk difference for a binary outcome).

```
gmm (y - {ey0} - x*{psi}), instruments(z1 z2)
estat overid
```

This is equivalent to Stata's built in `ivregress` command:

```
ivregress gmm y (x = z1 z2)
estat overid
```

Multiplicative SMM

Here `{psi}` denotes the log causal risk ratio, and hence we display the exponentiated estimate using the `lincom` command with its `eform` option after fitting the model.

```
gmm (y*exp(-1*x*{psi}) - {ey0}), instruments(z1 z2)
lincom [psi]_cons, eform // causal risk ratio
estat overid
```

We also give the Stata syntax for the alternative Multiplicative SMM moments. Here `{logey0}` denotes $\log\{E(Y_0)\}$ and so we additionally display the exponentiated form of this parameter after fitting the model.

```
gmm (y*exp(-x*{psi} - {logey0}) - 1), instruments(z1 z2)
lincom [psi]_cons, eform // causal risk ratio
lincom [logey0]_cons, eform // E[Y(0)]
estat overid
```

Logistic SMM

Here `{psi}` denotes the log causal odds ratio. In the joint estimation we use the `gmm` command's linear predictor substitution syntax (we denote the linear predictor for the association model by `{xb:}`). We collect the association and causal model parameter estimates in a matrix called `from`, we then use these estimates as initial values in the joint estimation. Also in the joint estimation we specify the `winitial(unadjusted, independent)` option so that the moments are assumed to independent in the first step of estimation. Note in Stata: $\text{invlogit}(x) = \text{expit}(x) = e^x/(1+e^x)$.

```
* generate interactions
gen xz1 = x*z1
gen xz2 = x*z2

* association model
logit y x z1 z2 xz1 xz2
matrix from = e(b)
predict xblog, xb

* causal model with incorrect SEs
gmm (invlogit(xblog - x*{psi}) - {ey0}), instruments(z1 z2)
matrix from = (from,e(b))

* joint estimation of association and causal models
gmm (y - invlogit({xb:x z1 z2 xz1 xz2} + {b0})) ///
  (invlogit({xb:} + {b0} - x*{psi}) - {ey0}), ///
  instruments(1:x z1 z2 xz1 xz2) ///
  instruments(2:z1 z2) ///
  winitial(unadjusted, independent) from(from)
lincom [psi]_cons, eform // causal odds ratio
estat overid
```

R syntax

The R syntax uses the `gmm()` function in the GMM package (Chaussé, 2010), which we first load using `library(gmm)`. After fitting each SMM using two-step estimation we perform the Hansen over-identification test using the `specTest()` function. The R code

assumes our data is in a matrix called `data` whose columns contain the values of the variables Y, X, Z_1, and Z_2 in this order with column names `"y"`, `"x"`, `"z1"`, `"z2"`.

In this code we have specified the `vcov="iid"` option which assumes the moment conditions are independent. We find specifying this option is necessary for the models to converge on reasonably sized datasets. We also find that changing the optimization algorithm used in the estimation through the `method` option can reduce the time it takes the models to fit (we find the BFGS and L-BFGS-B methods are the fastest).

Additive SMM

Firstly we fit the Additive SMM using the `gmm()` function's formula syntax for linear models.

```
asmm <- gmm(data[,"y"] ~ data[,"x"], x=data[,c("z1","z2")], vcov="iid")
print(summary(asmm))
print(cbind(coef(asmm),confint(asmm))) # estimates and 95% CI
print(specTest(asmm))
```

We can also pass the moment conditions to `gmm()` using its function syntax. In order to do this we first define a function `asmmMoments()` which returns the ASMM moments. This function must have two arguments; the first of which `theta` denotes the vector of parameters to be estimated, where `theta[1]` is $E(Y_0)$ and `theta[2]` is the causal risk difference. The second argument `x` is the data matrix, the user must avoid confusion here with the single variable `X`. In the `gmm()` function the `t0` option specifies the initial values of the parameter estimates. After we have fitted the model with the call to `gmm()` we print out the model summary, then the estimates and their 95% CIs, and finally the over-identification test using `specTest()`.

```
asmmMoments <- function(theta,x){
  # extract variables from x
  Y <- x[,"y"]
  X <- x[,"x"]
  Z1 <- x[,"z1"]
  Z2 <- x[,"z2"]
  # moments
  m1 <- (Y - theta[1] - theta[2]*X)
  m2 <- (Y - theta[1] - theta[2]*X)*Z1
  m3 <- (Y - theta[1] - theta[2]*X)*Z2
  return(cbind(m1,m2,m3))
}

asmm2 <- gmm(asmmMoments, x=data, t0=c(0,0), vcov="iid")
print(summary(asmm2))
print(cbind(coef(asmm2),confint(asmm2))) # estimates and 95% CI
print(specTest(asmm2))
```

Multiplicative SMM

We again use the `gmm()` function syntax to fit the Multiplicative SMM. Firstly we define the function `msmmMoments()` to return the moments. After fitting the model we print the model summary. Here `theta[2]` is the log causal risk ratio, and so we print the exponentiated form of this parameter.

```
msmmMoments <- function(theta,x){
  # extract variables from x
  Y <- x[,"y"]
  X <- x[,"x"]
  Z1 <- x[,"z1"]
  Z2 <- x[,"z2"]
  # moments
  m1 <- (Y*exp(- X*theta[2]) - theta[1])
  m2 <- (Y*exp(- X*theta[2]) - theta[1])*Z1
  m3 <- (Y*exp(- X*theta[2]) - theta[1])*Z2
  return(cbind(m1,m2,m3))
}

msmm <- gmm(msmmMoments, x=data, t0=c(0,0), vcov="iid")
print(summary(msmm))
print(exp(cbind(coef(msmm), confint(msmm))[2,])) # causal risk ratio
print(cbind(coef(msmm), confint(msmm))[1,]) # E[Y(0)]
print(specTest(msmm))
```

We can also fit the alternative MSMM moments in the same way. Here `theta[1]` denotes $\log\{E(Y_0)\}$ and so we print out the exponentiated form of both estimates.

```
msmmAltMoments <- function(theta,x){
  # extract variables from x
  Y <- x[,"y"]
  X <- x[,"x"]
  Z1 <- x[,"z1"]
  Z2 <- x[,"z2"]
  # moments
  m1 <- (Y*exp(-theta[1] - X*theta[2]) - 1)
  m2 <- (Y*exp(-theta[1] - X*theta[2]) - 1)*Z1
  m3 <- (Y*exp(-theta[1] - X*theta[2]) - 1)*Z2
  return(cbind(m1,m2,m3))
}

msmm2 <- gmm(msmmAltMoments, x=data, t0=c(0,0), vcov="iid")
print(exp(cbind(coef(msmm2), confint(msmm2)))) # exponentiate estimates & 95% CI
print(specTest(msmm2))
```

Logistic SMM

In estimation of the logistic SMM, especially with the joint moments, it is important to check that convergence has been reached, either by inspecting the model summary

or checking that the model `algoInfo$convergence` attribute is equal to 0. If convergence has not been reached a higher iteration limit (say 5000) can be specified in `gmm()` through the option `control=list(maxit=5000)`. Note in R qlogis$(p) = \log(p/(1-p))$ and plogis$(x) = \text{expit}(x) = e^x/(1+e^x)$.

First we fit the association model using the `glm()` function to fit the logistic regression. Again we collect the parameter estimates and predicted values. We then fit the causal model using the function `cmMoments()` to return its moment conditions. In this function `theta[1]` denotes $E(Y_0)$ and `theta[2]` denotes the log causal odds ratio.

In the joint estimation the function `lsmmMoments()` returns the moment conditions. In this function `theta[1:6]` are the coefficients in the association model, `theta[7]` denotes $E(Y_0)$ and `theta[8]` denotes the log causal odds ratio.

```
# association model
am <- glm(y ~ x + z1 + z2 + x*z1 + x*z2, as.data.frame(data), fam=binomial)
print(summary(am))
amfit <- coef(am)
xblog <- qlogis(fitted.values(am))

# causal model with incorrect SEs
cmMoments <- function(theta,x){
  # extract variables from x
  X <- x[,"x"]
  Z1 <- x[,"z1"]
  Z2 <- x[,"z2"]
  # moments
  c1 <- (plogis(xblog - theta[2]*X) - theta[1])
  c2 <- (plogis(xblog - theta[2]*X) - theta[1])*Z1
  c3 <- (plogis(xblog - theta[2]*X) - theta[1])*Z2
  return(cbind(c1,c2,c3))
}

cm <- gmm(cmMoments, x=data, t0=c(0,0), vcov="iid")
cmfit <- coef(cm)

lsmmMoments <- function(theta,x){
  # extract variables from x
  Y <- x[,"y"]
  X <- x[,"x"]
  Z1 <- x[,"z1"]
  Z2 <- x[,"z2"]
  XZ1 <- X*Z1
  XZ2 <- X*Z2
  # association model moments
  xb <- theta[1] + theta[2]*X + theta[3]*Z1 + theta[4]*Z2 + theta[5]*XZ1 + theta[6]*XZ2
  a1 <- (Y - plogis(xb))
  a2 <- (Y - plogis(xb))*X
  a3 <- (Y - plogis(xb))*Z1
  a4 <- (Y - plogis(xb))*Z2
```

```
  a5 <- (Y - plogis(xb))*XZ1
  a6 <- (Y - plogis(xb))*XZ2
  # causal model moments
  c1 <- (plogis(xb - theta[8]*X) - theta[7])
  c2 <- (plogis(xb - theta[8]*X) - theta[7])*Z1
  c3 <- (plogis(xb - theta[8]*X) - theta[7])*Z2
  return(cbind(a1,a2,a3,a4,a5,a6,c1,c2,c3))
}

lsmm <- gmm(lsmmMoments, x=data, t0=c(amfit,cmfit), vcov="iid")
print(summary(lsmm))
print(cbind(coef(lsmm), confint(lsmm))[8]) # E[Y(0)]
print(exp(cbind(coef(lsmm), confint(lsmm))[-7,])) # exponentiate other estimates
print(specTest(lsmm))
```