Center for Economic Institutions

Working Paper Series

No. 2010-9

"Testing a Parametric Function Against a Nonparametric Alternative in IV and GMM Settings"

Tue Gørgens and
Allan W¨urtz

January 2011

Center for Economic Institutions

Working Paper Series

Institute of Economic Research
Hitotsubashi University
2-1 Naka, Kunitachi, Tokyo, 186-8603 JAPAN
http://cei.ier.hit-u.ac.jp/English/index.html
Tel:+81-42-580-8405/Fax:+81-42-580-8333

Testing a parametric function against a nonparametric alternative in IV and GMM settings

Tue Gørgens* Allan Würtz†

December 2010

Abstract: This paper develops a specification test for functional form for models identified by moment restrictions, including IV and GMM settings. The general framework is one where the moment restrictions are specified as functions of data, a finite-dimensional parameter vector, and a nonparametric real function (an infinite-dimensional parameter vector). The null hypothesis is that the real function is parametric. The test is relatively easy to implement and its asymptotic distribution is known. The test performs well in simulation experiments.

Keywords: Generalized method of moments, specification test, nonparametric alternative, LM statistic, generalized arc-sine distribution.

JEL classification codes: C12, C14, C52.

Acknowledgment: We thank Joel Horowitz for sharing his computer code, Bo Honoré and Chris Skeels for comments on early drafts, anonymous referees for constructive criticism, and CREATES and the Danish Social Science Research Council for financial support. Part of this paper was written while Tue Gørgens was visiting the Center for Economic Institutions at Hitotsubashi University, and he is grateful for their hospitality and generous support.

*Address: Research School of Economics, Australian National University, Canberra ACT 0200, Australia. E-mail: tue.gorgens@anu.edu.au.

†Address: CREATES and School of Economics and Management, University of Aarhus, DK-8000 Aarhus C, Denmark. E-mail: awurtz@econ.au.dk.

1 Introduction

Generalized method of moments (GMM) and its special cases instrumental variables (IV) and two-stage least squares (2SLS) are frequently used to estimate parametric models in econometrics. These models specify moments as functions of data and a finite-dimensional parameter vector. The functional form is assumed to be known, apart from the parameters. In many applications, it is desirable to test the validity of the assumed functional form. In some cases there may be an obvious alternative model to test against. Often, however, there are no obvious alternatives. In this paper, we develop a test of functional form, which has power against models which specify the moments as functions of data, a finite-dimensional parameter vector, and a real function (an infinite-dimensional parameter vector).

Our test is based on the ideas of Aerts, Claeskens, and Hart (1999). They considered testing a parametric fit against a nonparametric alternative within several estimation frameworks: maximum likelihood, quasi-maximum likelihood, and general estimating equations. Their test is based on a sequence of LM test statistics, each designed to test against a specific parametric alternative. The sequence nests the null model, and in the limit it spans the class of models which can be written as functions of data, a finite-dimensional parameter vector and a real function. The LM statistics are divided by their degrees of freedom, and a single test statistic is constructed as the largest of these weighted LM statistics.

In this paper we extend these ideas to the testing of models which are formulated as restrictions on moment functions. Such models include regression models, models estimated by IV and, more generally, models estimated by GMM. In particular, our extension is applicable in overidentified models. There are two important new issues to consider when extending the original test to a GMM framework, namely identification of the model under the alternative and the selection of moment restrictions to use in the construction of the LM statistics. We discuss two approaches to the selection issue. For simplicity we shall refer to our extension as the GMM-ACH test.

Although the GMM-ACH approach is flexible and can be tailored to test against misspecification in any specific direction, importantly it can also be used to test the overall validity of a conditional moment restriction. There is a large literature on testing the validity of conditional moment restrictions. The central idea of the Integrated Conditional Moment (ICM) tests developed by Bierens (1982), Bierens (1990), and Bierens and Ploberger (1997) is to replace the conditional moment restriction with an equivalent set of unconditional moment restrictions based on, e.g., exponential weight functions. The ICM tests were developed for regression models, but can easily be adapted to test for functional-form specification in models with endogenous explanatory variables. In the tests considered by Donald et al. (2003), the conditional moment restriction is replaced with an equivalent sequence of unconditional moment restrictions based on series (see also Newey, 1985, and de Jong and Bierens, 1994). The GMM-ACH test uses the same setup as their GMM-test, but where the latter is based on Hansen's J-statistic the GMM-ACH tests is based on a sequence of LM statistics. Other approaches to testing the validity of a conditional moment restriction considers tests based on a marked empirical process and Kolmogorov-Smirnov or Cramér-von Mises statistics, see Stute (1997), Andrews (1997), Whang (2001), and Van Keilegom et al. (2008), and tests based on nonparametric estimation, see Tripathi and Kitamura (2003).[1]

Finally, the test proposed by Horowitz (2006) has a form similar to the ICM test, but uses a particular class of density functions for weighting instead of exponential functions. Horowitz proved that his test has better power properties than previously considered tests. Moreover, his simulation evidence suggests that his test has significantly better finite-sample power than the tests proposed by Bierens (1990), Tripathi and Kitamura (2003) and Donald et al. (2003). However, implementing Horowitz's test can be nontrivial, in part because it is not asymptotically pivotal. This implies that the critical values must be computed specifically for each application.

We compare the performance of the GMM-ACH test to some of the existing tests in a Monte Carlo study. Given the similarities with Donald et al.'s (2003) test and

[1] The nonparametric smoothing approach is somewhat hampered in IV models by ill-posed inverse problems; see e.g. Horowitz and Spokoiny (2001) for references to tests for models with no endogenous regressors. For testing conditional moment restrictions with dependent data, see e.g. Escanciano (2007).

given Horowitz's (2006) power results, we focus on these competing tests. The results confirm that the test by Horowitz tends to have the better power. The GMM-ACH test, however, has power close to (and some cases better than) that of Horowitz's test in these simulations. The test by Donald et al. has the lowest power of the three. A second Monte Carlo study compares testing against misspecification in a specific direction with testing against overall misspecification. The results show that the potential power gains of a specific test can be substantial.

In comparison with some of the other tests in the literature, including the test by Horowitz (2006), the GMM-ACH test is relatively simple to implement. In particular, the asymptotic distribution and hence the asymptotic critical values of the test are known. Moreover, since the test is based on LM statistics it is not necessary to estimate any alternative models, which is an advantage in some applications. We anticipate that in most applications performing the test involves, in principle, nothing more complicated than taking derivatives and inverting matrices.

The paper is structured as follows. Section 2 introduces the GMM-ACH test and explains the mechanics of the test in a simple and familiar IV setting. Section 3 considers the general GMM setting. We focus on the case where the infinite-dimensional parameter vector is an unknown function of a real variable, but discuss the extension to functions of several variables at the end. Section 4 presents examples of GMM-ACH tests in linear multiple regression models with endogenous explanatory variables, including an empirical example based on an Engel curve model. Section 5 concludes.

Throughout the paper, $0_{(a\times b)}$ denotes an $a\times b$-dimensional matrix of 0s and $I_{(j)}$ denotes the j-dimensional identity matrix. The symbol 0 is also used to denote a function which maps the real line to the number 0.

2 A simple IV model

In this section, we use a simple IV setup to explain how the GMM-ACH test is constructed. In the first subsection, we consider a version of the GMM-ACH test which uses the minimum number of moment restrictions required for each LM statistic. In the second subsection, we discuss a version which uses the same set of moment restrictions for all LM statistics. In the last subsection we present the results of a Monte Carlo study.

2.1 Minimum number of moment restrictions

The objective is to test a given parametric model against a nonparametric alternative model. Using subscript i to indicate a generic observation, let y_i be a scalar left-hand side variable, let x_i be a scalar right-hand side variable, and let z_i be a scalar instrument. Assume n independent observations are available. In this section the parametric model of interest is

$$y_i = x_{0i}'\beta^* + u_i, \quad \mathsf{E}(u_i|z_i) = 0, \quad \beta^* \in \mathbb{R}^2, \tag{1}$$

where $x_{0i} = (1, x_i)'$, β^* is an unknown two-dimensional parameter vector and u_i is an unobserved random variable. The nonparametric alternative model is the nonlinear model given by

$$y_i = x_{0i}'\beta^* + \gamma^*(x_i) + u_i, \quad \mathsf{E}(u_i|z_i) = 0, \quad \beta^* \in \mathbb{R}^2, \quad \gamma^* \in \Gamma, \tag{2}$$

where $\gamma^* : \mathbb{R} \to \mathbb{R}$ is an unknown function and Γ is a set of square integrable real functions. We assume that functions of the form $x_{0i}'\beta$ are excluded from Γ. We also assume that $0 \in \Gamma$, so that model (2) nests model (1). In terms of (2), the null hypothesis is that $\gamma^* = 0$ and the alternative hypothesis is that $\gamma^* \neq 0$.

The GMM-ACH test is based on four steps. The first step is to construct a sequence of nested parametric alternative models which approximate the nonparametric model (2). A series expansion of γ^* is used for this purpose. Let b_1, b_2, ... be a sequence of basis functions ($b_k : \mathbb{R} \to \mathbb{R}$ for $k = 1, 2, \ldots$) and assume that for each $\gamma \in \Gamma$ there are coefficients $\alpha_{\gamma 1}$, $\alpha_{\gamma 2}$, ... such that $\sum_{k=1}^{j} \alpha_{\gamma k} b_k$ converges to γ as $j \to \infty$. The sequence of nested parametric alternative models is based on the partial sums of the series expansion.

Define the functions g_1, g_2, ... by

$$g_j(e, \theta_j) = \sum_{k=1}^{j} \theta_{jk} b_k(e), \quad j = 1, 2, \ldots, \tag{3}$$

where $\theta_j = (\theta_{j1}, \ldots, \theta_{jj})' \in \mathbb{R}^j$ is a j-dimensional parameter vector and $e \in \mathbb{R}^1$. Under suitable regularity conditions, γ^* can be approximated arbitrarily well by $g_j(\cdot, \theta_j)$ by taking j large enough and choosing the appropriate θ_j.[2] Let θ_1^*, θ_2^*, ... denote these "pseudo-true" parameter vectors. A sequence of approximate alternative models can therefore be constructed as[3]

$$y_i = x_{0i}'\beta^* + g_j(x_i, \theta_j^*) + u_{ji}, \quad \mathsf{E}(u_{ji}|z_i) \to 0 \text{ as } j \to \infty,$$
$$\beta^* \in \mathbb{R}^2, \quad \theta_j^* \in \mathbb{R}^j, \quad j = 1, 2, \ldots. \tag{4}$$

In terms of (4), the null hypothesis is that $\theta_j^* = 0_{(j\times1)}$ for all $j = 1, 2, \ldots$ and the alternative hypothesis is that $\theta_j^* \neq 0_{(j\times1)}$ for some $j = 1, 2, \ldots$.

The second step in the GMM-ACH test concerns the identification of the parameters in the null model and in the approximate alternative models. In this section, we have chosen to specify the models using the conditional moment restriction $\mathsf{E}(u_i|z_i) = 0$. We assume that this conditional moment restriction identifies the parameters under the null as well as under the alternative. In practice, if x_i is continuously distributed, then it is convenient to base estimation and testing on unconditional moment restrictions. Since $\mathsf{E}(u_i|z_i) = 0$ implies $\mathsf{E}(u_i t(z_i)) = 0$ for any choice of function $t : \mathbb{R} \to \mathbb{R}$, arbitrarily many unconditional moment restrictions can easily be constructed. At least 2 moment restrictions are needed to identify and estimate β^*, and at least $2 + j$ moment restrictions are needed to identify and test hypotheses about $(\beta^{*\prime}, \theta_j^{*\prime})'$. A natural choice of additional instrument for identifying the coefficient on $b_k(x_i)$ is $b_k(z_i)$.

[2]To establish equivalence between (2) and (4) using Lemma 2.1 of Donald et al. (2003), the regularity conditions include that $\mathsf{E}(\gamma(x_i)^2) < \infty$ and that $\mathsf{E}[(\gamma(x_i) - \sum_{k=1}^{j} \alpha_{\gamma k} b_k(x_i))^2] \to 0$ as $j \to \infty$ for all $\gamma \in \Gamma$. For an introduction to the use of series in econometrics, see for example Pagan and Ullah (1999).

[3]In practice, it may happen that x_{0i} and the basis functions used in the construction of g_j are collinear. Indeed, this happened in the power function basis example offered just above. Since we are not interested in the latter per se, the offending terms may simply be omitted from g_j.

We proceed here by constructing an GMM-ACH test based on using the minimum number of moment restrictions required in each calculation. In the next section we discuss a version of the test which uses the same set of moment restrictions in all calculations. Under the null, assume β^* in (1) is identified by the two unconditional moment restrictions,

$$\mathsf{E}\big(z_{0i}(y_i - x_{0i}'\beta^*)\big) = 0_{(2\times 1)}, \quad \beta^* \in \mathbb{R}^2, \tag{5}$$

where $z_{0i} = (1, z_i)'$. Under the alternatives, assume that $(\beta^{*\prime}, \theta_j^{*\prime})'$ is identified by the $2 + j$ unconditional moment restrictions

$$\mathsf{E}\big(z_{ji}(y_i - x_{0i}'\beta^* - g_j(x_i, \theta_j^*))\big) \to 0_{(2+j\times 1)} \text{ as } j \to \infty,$$
$$\beta^* \in \mathbb{R}^2, \quad \theta_j^* \in \mathbb{R}^j, \quad j = 1, 2, \ldots, \tag{6}$$

where $z_{ji} = (1, z_i, b_1(z_i), \ldots, b_j(z_i))'$ for $j = 1, 2, \ldots$. Since the number of moment restrictions equal the number of parameters, the parameters are exactly identified for each j.

The third step in the GMM-ACH test is to calculate j statistics, one for testing the null model (1) against each of the approximate alternative models given in (4). There are several statistics which can be used. Here we follow Aerts et al. (1999) and use LM statistics. First we estimate the model under the null by solving the empirical analogues of (5). That is, the estimator, $\tilde{\beta} = \big(n^{-1}\sum_{i=1}^n z_{0i}x_{0i}'\big)^{-1}\big(n^{-1}\sum_{i=1}^n z_0 y_i\big)$, is the solution in β to

$$n^{-1}\sum_{i=1}^{n} z_{0i}(y_i - x_{0i}'\beta) = 0_{(2\times 1)}. \tag{7}$$

For given j, the first-order conditions for joint estimation of β and θ_j are similarly

$$n^{-1}\sum_{i=1}^{n} z_{ji}(y_i - x_{ji}'\delta_j) = 0_{(2+j\times 1)}, \quad j = 1, 2, \ldots, \tag{8}$$

where $\delta_j = (\beta', \theta_j')' \in \mathbb{R}^{2+j}$ and $x_{ji} = (1, x_i, b_1(x_i), \ldots, b_j(x_i))'$ for $j = 1, 2, \ldots$. We then construct an LM test based on the fact that if the null is true, then the first-order conditions should be (approximately) satisfied when evaluated at the parameter estimate obtained under the null; that is, at $\beta = \tilde{\beta}$ and $\theta_j = 0_{(j\times 1)}$. We show in Section 3 and Appendix A that the LM statistic can be calculated as[4]

$$R_j = M_j'(A_j^{-1})'H_j'\big(H_jA_j^{-1}B_j(A_j^{-1})'H_j'\big)^{-1}H_jA_j^{-1}M_j, \quad j = 1, 2, \ldots, \tag{9}$$

where[5]

$$M_j = n^{-1}\sum_{i=1}^{n} z_{ji}(y_i - x_{0i}'\tilde{\beta}), \quad j = 0, 1, \ldots. \tag{10}$$

$$A_j = -n^{-1}\sum_{i=1}^{n} z_{ji}x_{ji}', \quad j = 0, 1, \ldots, \tag{11}$$

$$B_j = \frac{1}{n}\left(n^{-1}\sum_{i=1}^{n}(y_i - x_{0i}'\tilde{\beta})^2 z_{ji}z_{ji}'\right), \quad j = 0, 1, \ldots, \tag{12}$$

$$H_j = \begin{bmatrix} 0_{(j\times 2)} & I_{(j)} \end{bmatrix}, \quad j = 1, 2, \ldots. \tag{13}$$

Given j, the LM statistic R_j has an asymptotic χ_j^2-distribution under the null.

Although perhaps not obvious from (9), note that the LM statistic has the form $R_j = M_j'\mathrm{Var}(M_j)^- M_j$, where $\mathrm{Var}(M_j)^-$ is a generalized inverse of the variance matrix of M_j or an estimate of that matrix. A generalized inverse is required since by construction the first two components of M_j are 0. Calculation of $\mathrm{Var}(M_j)^-$ is complicated by the fact that M_j depends on the estimated parameter $\tilde{\beta}$, and the nesting properties of M_0 and M_j and of A_0 and A_j, which follow from the definition of x_{0i}, x_{1i}, … and z_{0i}, z_{1i}, …, are crucial in deriving (9). We discuss the nesting property in more detail in Section 3.

The fourth and final step in the GMM-ACH test is to construct an overall test statistic

[4]The simple IV setup with exact identification is almost a special case of the GEE setup considered by Aerts et al. (1999). Their LM statistic is valid only if A_j is symmetric (e.g. if $z_i = x_i$). They stated the LM statistic in a different form. Let $[X]_j$ denote the lower right $j \times j$-submatrix of the $(2+j)\times(2+j)$-matrix X or the last j elements of the $(2+j)$-vector X, then R_j can be expressed as $R_j = [M_j]_j'[A_j^{-1}]_j\big([A_j^{-1}B_jA_j^{-1}]_j\big)^{-1}[A_j^{-1}]_j[M_j]_j$.

[5]For simplicity the dependence of R_j, M_j, A_j, B_j (and other random matrices defined below) on n is suppressed in the notation.

by taking the maximum over a sequence of weighted LM statistics. The weights are the reciprocal of the degrees of freedom of the individual statistics. Specifically, the GMM-ACH test statistic is

$$S_r = \max_{1 \leq j \leq r} (R_j / j), \tag{14}$$

where r is an upper bound which should be increasing in n.[6] In Section 3, we argue that the distribution of S_r under the null converges, as $r \to \infty$ and $n \to \infty$, to a distribution which does not depend on any unknown population characteristics. Hart (1997, p178) tabulated this distribution, and the 1%, 5% and 10% critical values are 6.75, 4.18 and 3.22. The requirement that $r \to \infty$ is not important for picking the critical value; Aerts, Claeskens, and Hart (1999, p872) claimed that the asymptotic approximation is usually fine for critical values less than 10% as long as $r > 5$.

It is worth noting that if an appropriate sequence of instruments is used, then the GMM-ACH test described in this section is in fact a test of the conditional moment restriction $\mathsf{E}(y_i - x_{0i}'\beta^* | z_i) = 0$ against the negation $\mathsf{E}(y_i - x_{0i}'\beta | z_i) \neq 0$ for all $\beta \in \mathbb{R}^2$. That is, the alternative hypothesis is broader than suggested by (2). Intuitively, this follows because the sequence of unconditional moment restrictions $\mathsf{E}((y_i - x_{0i}'\beta^*) b_j^z(z_i)) = 0$ for $j = 1, 2 \ldots$ is equivalent to the original conditional moment restriction under appropriate regularity conditions.[7] Therefore, if the null is false, there is some j^* such that $n^{-1} \sum_{i=1}^n b_{j^*}^z(z_i)(y_i - x_{0i}'\tilde{\beta})$ converges to a nonzero constant and hence $R_{j^*} \to^p \infty$ as $n \to \infty$. Since R_{j^*} defined in this section remains unchanged as $r \to \infty$, it follows that $S_r \to^p \infty$ as $r \to \infty$ and $n \to \infty$.

Finally, note that the LM and GMM-ACH statistics presented in this subsection do not depend on the specification of the alternative model. That is, despite the facts that

[6] In a likelihood framework, rejecting the null if S_r is large is equivalent to rejecting the null if the Akaike Information Criterion (AIC) of one of the alternative models is sufficiently larger than the AIC of the null model. For further discussion of the connection between the GMM-ACH and the AIC statistics, see Aerts et al. (1999).

[7] See e.g. Bierens (1982, Theorem 2) and Donald et al. (2003, Lemma 2.1). The latter requires that for each t with $\mathsf{E}(t(z_i)^2) < \infty$ there are coefficients $\alpha_{t1}, \alpha_{t2}, \ldots$ such that $\mathsf{E}\big[\big(t(z_i) - \sum_{k=1}^j \alpha_{tk} b_k^z(z_i)\big)^2\big] \to 0$ as $j \to \infty$. The instrument sequence used in the text is $b_1^z(z_i) = 1$, $b_2^z(z_i) = z_i$, $b_3^z(z_i) = b_1(z_i)$, $b_4^z(z_i) = b_2(z_i)$ etc. where $b_k : \mathbb{R} \to \mathbb{R}$ for $k = 1, 2, \ldots$ and $\mathsf{E}(b_k(x_i) b_{k'}(x_i)) < \infty$.

A_j appears in (9) and that A_j depends on x_{ji}, the LM statistics depend numerically only on x_{0i}. The derivations in Appendix A.2 show that the columns of A_j corresponding to the derivatives of the moments with respect to the parameters θ_j in the approximate alternative model drop out of the formula in (9).[8] This is a consequence of using the minimum number of moment restrictions, and holds both in the simple IV model presented in this section and in the setup discussed in Section 3.3. In the next subsection, we present a version of the GMM-ACH test where there are more moment restrictions than parameters, and here the alternative model plays a substantive role.

2.2 Same set of moment restrictions

The version of the GMM-ACH test presented above is based on using the minimum number of moment restrictions required to identify the parameters under the null and the alternative hypotheses. The literature on hypothesis testing in IV and GMM settings (see e.g. Engle, 1984; Newey and McFadden, 1994) usually recommends using the same set of moment restrictions under both the null and the alternative. For notational simplicity, we consider the case where there are $2+r$ moment restrictions in this section. The general case presented in Section 3 allows for an arbitrary number (equal to or larger than $2+r$). In this case, (5) and (6) are replaced by

$$\mathsf{E}\big(z_{ri}(y_i - x_{0i}'\beta^*)\big) = 0_{(2+r\times 1)}, \quad \beta^* \in \mathbb{R}^2, \tag{15}$$

and

$$\mathsf{E}\big(z_{ri}(y_i - x_{0i}'\beta^* - g_j(x_i,\theta_j^*))\big) \to 0_{(2+r\times 1)} \text{ as } j \to \infty \text{ and } r \to \infty,$$
$$\beta^* \in \mathbb{R}^2, \quad \theta_j^* \in \mathbb{R}^j, \quad j = 1,\dots,r. \tag{16}$$

Except in the case where $j = r$, there are more equations than unknown parameters in (15) and (16).

[8]See equations (66) and (67) in Appendix A.2. When the test statistic does not depend on the specification of an alternative model, the convergence conditions imposed on the series expansion in the beginning of this section are redundant.

Since the parameters are overidentified, we estimate β^* using 2SLS. The 2SLS estimator, $\tilde{\beta}$, based on (15) is

$$\tilde{\beta} = (A_0' W_r A_0)^{-1} A_0' W_r \left(n^{-1} \sum_{i=1}^{n} z_{ri} y_i \right), \tag{17}$$

where $A_0 = -n^{-1} \sum_{i=1}^{n} z_{ri} x_{0i}'$, and the weight matrix is $W_r = \left(n^{-1} \sum_{i=1}^{n} z_{ri} z_{ri}' \right)^{-1}$. For given j, the 2SLS first-order conditions for estimation of β and θ_j are

$$A_j' W_r \left(n^{-1} \sum_{i=1}^{n} z_{ri} (y_i - x_{ji}' \delta_j) \right) = 0_{(2+j \times 1)}, \quad j = 1, \ldots, r, \tag{18}$$

where $A_j = -n^{-1} \sum_{i=1}^{n} z_{ri} x_{ji}'$ and $\delta_j = (\beta', \theta_j')' \in \mathbb{R}^{2+j}$ for $j = 1, \ldots, r$. The LM statistic is constructed from (18) evaluated at $\beta = \tilde{\beta}$ and $\theta_j = 0_{(j \times 1)}$. As in Section 2.1, calculation of the variance of these first-order conditions is complicated by the randomness of $\tilde{\beta}$, but facilitated by the nesting properties imposed on A_0 and A_j via the construction of the approximate alternative models. We show in Section 3 and Appendix A that, for each j, LM statistics for testing $\theta_j^* = 0_{(j \times 1)}$ against $\theta_j^* \neq 0_{(j \times 1)}$ using (16) can be constructed as

$$R_j = M_r' W_r A_j J_j A_j' W_r M_r, \quad j = 1, 2, \ldots, \tag{19}$$

where A_j for $j = 0, \ldots, r$ are defined above and

$$M_r = n^{-1} \sum_{i=1}^{n} z_{ri} (y_i - x_{0i}' \tilde{\beta}), \tag{20}$$

$$B_r = \frac{1}{n} \left(n^{-1} \sum_{i=1}^{n} (y_i - x_{0i}' \tilde{\beta})^2 z_{ri} z_{ri}' \right), \tag{21}$$

$$C_j = (A_j' W_r A_j)^{-1} A_j' W_r B_r W_r A_j (A_j' W_r A_j)^{-1}, \quad j = 1, \ldots, r, \tag{22}$$

$$J_j = (A_j' W_r A_j)^{-1} H_j' \left(H_j C_j H_j' \right)^{-1} H_j (A_j' W_r A_j)^{-1}, \quad j = 1, \ldots, r, \tag{23}$$

and H_j is defined in (13). Finally, the GMM-ACH statistic is $S_r = \max_{1 \leq j \leq r} (R_j / j)$, as before. The asymptotic distributions, as $r \to \infty$ and $n \to \infty$, of the LM statistics and the GMM-ACH statistic are the same as in the previous section.

Intuitively, when the minimum number of moment restrictions are used, the LM statistics are large if the additional instruments in z_{ji} are correlated with the residuals from the null model. When the same set of moment restrictions are used, the LM statistics are large if the columns in A_j corresponding to the additional regressors in x_{ji} are not orthogonal to the weighted empirical moments from the null model, $W_r M_r$. Since the first depends on additional instruments and the other on additional regressors, the two versions of the test may have different power properties. The next subsection presents a small Monte Carlo study which compares the two versions of the GMM-ACH test.

2.3 A small Monte Carlo study

In the remainder of this section we present and discuss simulation results on the finite-sample behavior of several versions of the GMM-ACH test for the simple IV model. We consider both the test based on the minimum number and on the same set of moment restrictions, and we calculate the tests using both power and Fourier flexible form bases in the series approximation. We compare the GMM-ACH tests with the tests developed by Donald et al. (2003) and Horowitz (2006), as well as with simple ad hoc t and LM tests.

The setup considered by Donald et al. (2003) is similar to ours, but their test is based on the J-statistic for overidentifying restrictions. In general, the J-test does not have power against nonparametric alternatives. Donald et al. modified the J-test by letting the number of overidentifying restrictions depend on the sample size. As the sample size increases, the test gains power against a larger set of alternatives. The additional moment restrictions are generated from a conditional moment restriction, as described in Section 2.1.

As explained in the Introduction, Horowitz (2006) developed a test similar in form to the ICM test by Bierens (1982). Horowitz proved that the power of his test is arbitrarily close to 1 uniformly over a class of alternatives whose distance from the null hypothesis is of order $n^{-1/2}$. He compared several specification tests in a simulation study and found that they have inferior power properties compared to his own test. For simplicity, we focus on comparing the GMM-ACH test with the tests by Donald et al. (2003) and Horowitz (2006) in this section.

As a benchmark, we report a simple t test based on the model obtained by adding one additional term to the null model. Since in most cases this alternative coincides with the data-generating process, we expect this t test to have very good power properties. In practice, the data-generating process is likely to be more complicated and we would then expect a t test to have less favorable power properties.

Finally, to illustrate the effect of taking the maximum of weighted LM test statistics against a sequence of parametric alternatives, we also report on the properties of an ordinary LM test against the largest (rth) parametric alternative.

The designs, and some of the results, are taken from Horowitz (2006). The data-generating process for all these experiments is

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + u_i, \tag{24}$$

$$x_i = \Phi\big(\rho v_{1i} + (1-\rho^2)^{1/2} v_{2i}\big), \tag{25}$$

$$z_i = \Phi(v_{1i}), \tag{26}$$

$$u_i = 0.2(\eta v_{2i} + (1-\eta^2)^{1/2} v_{3i}), \tag{27}$$

where Φ denotes the standard normal distribution function, v_{1i}, v_{2i} and v_{3i} are independent standard normal random variables, and β_0, β_1, β_2, β_3, ρ and η are scalar parameters which vary across designs.

The results are shown in Table 1. Technical details of the implementation are given in the table notes.[9] The results in the first part of the table show that the different versions of the GMM-ACH test have good level control. The only exception is the design where the GMM-ACH test is based on 2SLS and a power function basis. In that design the GMM-ACH test rejects too much and, perhaps surprisingly, so does the t test. In most

[9] Horowitz (2006) held v_{1i} and v_{2i} constant in repeated samples for the HOR test, but not for the DIN test. We use exactly the same simulated data (same random seed, v_{1i} and v_{2i} constant) as Horowitz used for the HOR test.

other cases the level is correct within Monte Carlo sampling error (± 1.4 percentage point for a 5% test) or it is too low. This may cause lower power.

The second part of Table 1 shows that the GMM-ACH tests have power comparable to Horowitz's test in these designs, and in some cases even better power. The test by Donald et al. has significantly lower power in most of the designs. Notice also that the idea of combining a sequence of LM test statistics into the GMM-ACH test generally has a positive effect on power. For many of the designs, there is a power loss of about 20 percentage points when doing a single LM test rather than doing the GMM-ACH test.

In sum, it appears that the GMM-ACH test has good properties. The level is well controlled, and the power is close to that of Horowitz's test and much better than Donald et al.'s test. A power basis seems to yield better power than a Fourier flexible form basis. However, this is not surprising given that the data generating process is polynomial. The simulations do not show a clear favorite between using the minimum number or the same set of moment restrictions in the GMM-ACH test. Finally, we note that the power of the GMM-ACH test is generally higher than the power of the ad hoc LM test.

3 A GMM-based specification test

The previous section presented the main ideas of the GMM-ACH test in the context of a simple linear IV model. In this section we develop the GMM-ACH test for a general nonlinear model identified by moment restrictions. Our framework includes many models of interest in economics such as system of equations models (typically estimated by two-stage least squares) and dynamic panel data models with fixed effects (typically estimated by GMM). When the parameters are overidentified, these models are not included in the frameworks discussed by Aerts et al. (1999).

The presentation is divided into several subsections. The first sets up the null and the alternative hypotheses. The second subsection reviews GMM estimation and LM testing and defines the GMM-ACH statistic. Subsections three and four considers two consistent versions of the test, one using the minimum number of moment restrictions and the other using the same set of moment restrictions in all calculations. The final subsection offers remarks.

3.1 Model and hypotheses

Some econometric models are stated in terms of conditional moment restrictions (e.g. simple IV models, financial time series models) and others in terms of unconditional moment restrictions (e.g. dynamic panel data models). Ultimately the estimation of most models is based on unconditional moment restrictions, and we therefore specify the general model in terms of unconditional moment restrictions. We discuss how the GMM-ACH approach can be used to test conditional moment restrictions in Section 3.5.

The setting is the following. Assume n independent observations are available for analysis. Let v_i be a generic random vector of data, let β^* be an unknown h-vector of parameters, and let $\gamma^* : \mathbb{R}^d \to \mathbb{R}$ be an unknown function. Let F be a known infinite-dimensional vector of functions of these three quantities. The econometric model is cast in terms of a vector of moment restrictions,

$$\mathsf{E}\big(F(v_i, \beta^*, \gamma^*)\big) = 0, \quad \beta^* \in \mathbb{R}^h, \quad \gamma^* \in \Gamma, \tag{28}$$

where 0 here represents an infinite-dimensional vector of 0s and where Γ is set of square integrable real functions. We assume that the null function is in Γ; i.e. $0 \in \Gamma$. We also assume that (28) identifies β^* and γ^*. As in Section 2, this may require exclusion of certain (e.g. linear) functions from Γ. In general it is not possible to identify a function (equivalent to an infinitely-dimensional parameter) such as γ^* from a finite set of moment restrictions, which is why we allow F to be infinitely-dimensional. In terms of (28), the null hypothesis is that $\gamma^* = 0$. The alternative hypothesis is that $\gamma^* \neq 0$.

The range of null and alternative models which can be cast in the form of (28) is very wide. We provide some examples in Section 4. The generality of (28) and the fact that we have made few assumptions about γ^* and how γ^* interacts with v_i and β^* are strengths of the GMM-ACH approach. Often, γ^* will simply be a function of one of the components of v_i. In multiple-equation models such as dynamic panel data models, γ^* may be a function of a different component of v_i in each equation. In general, the argument of γ^* may be a function involving both v_i and β^* as in single-index models.

The GMM-ACH approach to testing the null against the nonparametric alternative is based on approximating the unknown γ^* with a sequence of nested parametric alternatives, $g_1,\, g_2,\, \ldots$. The construction of this sequence is explained in Section 2.1 for the case of a scalar argument ($d = 1$). A sequence of approximating functions with a multivariate argument ($d \geq 1$) can be constructed similarly. Specifically, let $c_1,\, c_2,\, \ldots$ denote an ordered sequence of basis functions ($c_k : \mathbb{R}^d \to \mathbb{R}$ for $k = 1, 2, \ldots$) such that for each $\gamma \in \Gamma$ there are coefficients $\alpha_{\gamma 1},\, \alpha_{\gamma 2},\, \ldots$ such that $\sum_{k=1}^{j} \alpha_{\gamma k} c_k$ converges to γ as $j \to \infty$. A sequence of approximating functions $g_1,\, g_2,\, \ldots$ can be defined as

$$g_j(e, \theta_j) = \sum_{k=1}^{\lambda_j} \theta_{jk} c_j(e), \quad j = 1, 2, \ldots, \tag{29}$$

where $\theta_j = (\theta_{j1}, \ldots, \theta_{j\lambda_j})' \in \mathbb{R}^{\lambda_j}$ and $\lambda_j \geq \lambda_{j-1} + 1$ with $\lambda_0 = 0$. Here λ_j denotes the number of parameters in (or terms in the partial sum of) the jth approximating function. For reasons discussed in Section 3.5, it is desirable to allow λ_j to be larger than j.

Since only a finite number of parameters are unknown under the null and the parametric alternatives, they may be identified from a finite set of moment restrictions. For $j = 1, 2, \ldots$, let F_j denote the first l_j components of F. Under the null, assume without loss of generality that β^* is identified (possibly overidentified) by the l_0 moment restrictions

$$\mathsf{E}\big(F_0(v_i, \beta^*, 0)\big) = 0_{(l_0 \times 1)}, \quad \beta^* \in \mathbb{R}^h. \tag{30}$$

Under the parametric alternatives $j = 1, 2, \ldots$, let $\theta_1^*,\, \theta_2^*,\, \ldots$ denote the "pseudo-true" values, and assume similarly that β^* and θ_j^* are identified (possibly overidentified) by the

l_j moment restrictions[10]

$$\mathsf{E}\big(F_j(v_i,\beta^*,g_j(\cdot,\theta_j^*))\big) \to 0_{(l_j\times 1)} \text{ as } j\to\infty,$$

$$\beta^*\in\mathbb{R}^h,\quad \theta_j^*\in\mathbb{R}^{\lambda_j},\quad j=1,2,\ldots. \quad (31)$$

In terms of the parameters of the approximating models, the null hypothesis can be restated as $\theta_j^* = 0_{(\lambda_j\times 1)}$ for all $j = 1,2,\ldots$, while the alternative hypothesis is that $\theta_j^* \neq 0_{(\lambda_j\times 1)}$ for at least one of $j = 1,2,\ldots$.

3.2 Test statistics

We now review GMM estimation and LM testing. For convenience, define $\delta_0 = \beta$ and $\delta_j = (\beta',\theta_j')'$ for $j = 1,2,\ldots$. Then define $f_0(\cdot,\delta_0) = F_0(\cdot,\beta,0)$ and $f_j(\cdot,\delta_j) = F_j(\cdot,\beta,g_j(\cdot,\theta_j))$ for $j = 1,2,\ldots$. The GMM criterion functions, q_j, are

$$q_j(\delta_j) = (1/2)m_j(\delta_j)'W_jm_j(\delta_j),\quad j=0,1,\ldots, \quad (32)$$

where W_j are some $l_j\times l_j$ symmetric weight matrices, and m_j are estimators of the moments $\mathsf{E}\big(f_j(v_i,\delta_j)\big)$, as a function of δ_j, defined by

$$m_j(\delta_j) = n^{-1}\sum_{i=1}^{n} f_j(v_i,\delta_j),\quad j=0,1,\ldots. \quad (33)$$

For each $j = 0,1,\ldots$, the first-order condition for a minimum at $\tilde{\delta}_j$ is $\mathsf{D}q_j(\tilde{\delta}_j) = 0_{(h+\lambda_j\times 1)}$. The derivatives of q_j with respect to δ_j are

$$\mathsf{D}q_j(\delta_j) = a_j(\delta_j)'W_jm_j(\delta_j),\quad j=0,1,\ldots, \quad (34)$$

[10]To establish equivalence between (28) and (31) using Lemma 2.1 of Donald et al. (2003) certain regularity conditions must be satisfied. The exact requirements will depend on the details of how γ^* interacts with v_i and β^*. For example, suppose the argument of γ^* can be written $w_i = \psi(v_i)$ for some known ψ. Then the conditions include that $\mathsf{E}(\gamma(w_i)^2) < \infty$ and that $\mathsf{E}[(\gamma(w_i) - \sum_{k=1}^{\lambda_j}\alpha_{\gamma k}b_k(w_i))^2] \to 0$ as $j\to\infty$ for all $\gamma\in\Gamma$. Also, as noted in Footnote 3, basis functions which are collinear with the null model must be omitted from the series.

where a_j are the gradients of m_j,

$$a_j(\delta_j) = n^{-1} \sum_{i=1}^{n} \mathsf{D} f_j(v_i, \delta_j), \quad j = 0, 1, \ldots. \tag{35}$$

Here $\mathsf{D} f_j$ denotes the $l_j \times h + \lambda_j$-matrix of partial derivative functions of f_j with respect to δ_j. In many applications, the moment functions are linear in the parameters and the first-order conditions can be solved analytically for $\tilde{\delta}_j$.

Define the "pseudo-true" parameter vector $\delta_j^* = (\beta^{*\prime}, \theta_j^{*\prime})'$ and define the restricted estimator as $\tilde{\delta}_{0j} = (\tilde{\delta}_0', 0'_{(\lambda_j \times 1)})'$ for $j = 1, 2, \ldots$. Define the matrices

$$H_j = \begin{bmatrix} 0_{(\lambda_j \times h)} & I_{(\lambda_j)} \end{bmatrix}, \quad j = 1, 2, \ldots. \tag{36}$$

With this notation, the null hypothesis can then be expressed as $H_j \delta_j^* = 0_{(\lambda_j \times 1)}$ for $j = 1, 2, \ldots$, while the alternative is that $H_j \delta_j^* \neq 0_{(\lambda_j \times 1)}$ for some $j = 1, 2, \ldots$.

LM statistics are based on the fact that if the null is true, then the derivative of the GMM criterion function for model j should be close to $0_{(h+\lambda_j \times 1)}$ when evaluated at $\tilde{\delta}_{0j}$. For each j, LM statistics for testing $H_j \delta_j^* = 0_{(\lambda_j \times 1)}$ against $H_j \delta_j^* \neq 0_{(\lambda_j \times 1)}$ have the form

$$R_j = \mathsf{D} q_j(\tilde{\delta}_{0j})' \mathrm{Var}(\mathsf{D} q_j(\tilde{\delta}_{0j}))^{-} \mathsf{D} q_j(\tilde{\delta}_{0j}), \quad j = 1, 2, \ldots, \tag{37}$$

where $\mathrm{Var}(\mathsf{D} q_j(\tilde{\delta}_{0j}))^{-}$ is a generalized inverse of the variance matrix of the gradient $\mathsf{D} q_j(\tilde{\delta}_{0j})$ or an estimate of that matrix. Note that the rank of $\mathrm{Var}(\mathsf{D} q_j(\tilde{\delta}_{0j}))$ is λ_j. We discuss estimation of $\mathrm{Var}(\mathsf{D} q_j(\tilde{\delta}_{0j}))$ and $\mathrm{Var}(\mathsf{D} q_j(\tilde{\delta}_{0j}))^{-}$ below.

The GMM-ACH statistic, S_r, is the maximum of a sequence of weighted LM statistics for testing the null hypothesis against the alternatives in the sequence, where the weights are the reciprocal of the statistic's degrees of freedom. Specifically,[11]

$$S_r = \max_{1 \leq j \leq r} (R_j / \lambda_j), \tag{38}$$

[11] While the LM statistic is convenient, alternatively one could base the GMM-ACH test on Wald or distance metric tests.

where r is some appropriately large integer (with $r \to \infty$ as $n \to \infty$).

In the theorems below we describe two cases where S_r is asymptotically pivotal; that is, under the null its asymptotic distribution does not depend on any unknown population quantities. Specifically, the asymptotic distribution is a transformation of the generalized arc-sine distribution, namely

$$\mathsf{P}(S_r \leq s) \to \exp\left(-\sum_{k=1}^{\infty} \frac{\mathsf{P}(\chi_k^2 > ks)}{k}\right) \quad \text{as } r \to \infty \text{ and } n \to \infty, \tag{39}$$

where χ_k^2 has a chi-square distribution with k degrees of freedom. As mentioned in Section 2.1, asymptotic critical values have been tabulated by Hart (1997).

The nesting properties of the moment restrictions are important in the derivation of the asymptotic distribution of the LM statistics and the GMM-ACH statistic. In particular, the nesting properties are used to ensure that each LM statistic is asymptotically $\chi^2_{\lambda_j}$-distributed and the differences between R_{j-1} and R_j for $j = 2, 3, \ldots$ are asymptotically uncorrelated. For ease of reference, we state them as Assumption 1.

Assumption 1 *Let $l_0 \leq l_1 \leq \cdots$. For $j = 1, 2, \ldots$, the first l_{j-1} components of $f_j(v_i, \delta_j)$ equal $f_{j-1}(v_i, \delta_{j-1})$ for all (v_i, δ_j) such that $\delta_j = (\delta'_{j-1}, 0'_{\lambda_j - \lambda_{j-1} \times 1})'$, and the restricted estimator is $\tilde{\delta}_{0j} = (\tilde{\delta}'_0, 0'_{(\lambda_j \times 1)})'$.*

The theorems below require that each LM statistic is asymptotically $\chi^2_{\lambda_j}$-distributed under the null. Using Assumption 1, an estimator of $\mathrm{Var}(\mathsf{D}q_j(\tilde{\delta}_{0j}))$ is derived in Appendix A.1. Our setup is not quite standard and we have been unsuccessful in finding the necessary results in the literature. Standard treatments of LM statistics assume that W_j is an estimate of the optimal weight matrix and that the restricted estimator is obtained from minimizing $\mathsf{D}q_j$ with respect to δ_j subject to the restrictions $H_j \delta_j^* = 0_{(\lambda_j \times 1)}$ (see e.g. Newey and McFadden, 1994, Section 9). In the present case, the weight matrix, W_j, is arbitrary and the LM statistic is evaluated at $\tilde{\delta}_{0j}$, which is obtained from solving a different problem, namely the unrestricted minimization of $\mathsf{D}q_0$ with respect to δ_0.

3.3 Minimum number of moment restrictions

The first case we consider is where the number of moment restriction used under the null and each parametric alternative equals the number of parameters in the corresponding model. The simple IV model discussed in Section 2.1 is an example of such a setup. In this case, the parameters are exactly identified both under the null and approximate alternative hypotheses. Only a minimum number of moment restrictions are used in each step. Define $M_j = m_j(\tilde{\delta}_{0j})$ and $A_j = a_j(\tilde{\delta}_{0j})$. When $l_j = h + \lambda_j$ for all $j = 0, 1, \ldots$, then $\mathsf{D}q_j(\tilde{\delta}_{0j})'\mathrm{Var}(\mathsf{D}q_j(\tilde{\delta}_{0j}))^- \mathsf{D}q_j(\tilde{\delta}_{0j})$ is the same as $M_j'\mathrm{Var}(M_j)^- M_j$. Define

$$B_j = \frac{1}{n}\left(n^{-1}\sum_{i=1}^{n} f_j(v_i, \tilde{\delta}_{0j}) f_j(v_i, \tilde{\delta}_{0j})'\right), \quad j = 0, 1, \ldots, \tag{40}$$

$$C_j = A_j^{-1} B_j (A_j')^{-1}, \quad j = 1, 2, \ldots, \tag{41}$$

and

$$J_j = (A_j^{-1})' H_j' \big(H_j C_j H_j'\big)^{-1} H_j A_j^{-1}, \quad j = 1, 2, \ldots, \tag{42}$$

then J_j is an estimator of $\mathrm{Var}(M_j)^-$. In Appendix A.2, we show that the LM statistics simplify to[12]

$$R_j = M_j' J_j M_j, \quad j = 1, 2, \ldots. \tag{43}$$

Note that, this is the same formula as (9) used in Section 2.1. IV estimators are invariant to the choice of weight matrix, which also drops out of the formula for the LM statistics.

Theorem 1 below states sufficient conditions for R_j to be asymptotically $\chi^2_{\lambda_j}$-distributed and provides the corresponding asymptotic distribution of S_r.

Theorem 1 *Assumption 1 holds and regularity conditions are satisfied.*[13] *For each* $j =$

[12]This formula has the same form as the LM statistic based on the quasi-maximum likelihood estimator given in Theorem 3.5 in the article by White (1982).

[13]For simplicity we do not spell out the standard regularity conditions required for Taylor expansions to be valid, central limit theorems to hold, etc. As indicated in (39), the limiting distribution is valid for $r \to \infty$ as $n \to \infty$. To bound the behavior of the test statistic as $r \to \infty$, it is assumed that, for given

$1, 2, \ldots$, suppose $l_j = h+\lambda_j$ and A_j is invertible. For each $j = 1, 2, \ldots$, suppose there exists a nonstochastic matrix, Σ_j, such that $n^{1/2} M_j \to^d N(0_{(h+\lambda_j\times 1)}, \Sigma_j)$ and $\mathrm{Var}(M_j) \to^p \Sigma_j$ as $n \to \infty$ and such that the first h rows and columns of Σ_j consist of 0s and the lower right $\lambda_j \times \lambda_j$ submatrix of Σ_j is positive definite. Then under the null the asymptotic distribution of S_r as $r \to \infty$ and $n \to \infty$ is given in (39).

3.4 Same set of moment restrictions

The second case we consider is where the same set of moment restrictions and weight matrix are used to calculate all LM statistics. That is, $l_j = l_0$ and $W_j = W_0$ for all $j = 1, 2, \ldots$. This is the case usually considered in the literature on hypothesis testing in IV and GMM settings (see e.g. Engle, 1984; Newey and McFadden, 1994). The 2SLS setup in Section 2.2 provides an example. As in the previous subsection, define $M_j = m_j(\tilde{\delta}_{0j})$ and $A_j = a_j(\tilde{\delta}_{0j})$. Using Assumption 1, we show in Appendix A.3 that J_j is an estimator of $\mathrm{Var}(\mathsf{D}q_j(\tilde{\delta}_{0j}))^-$, where

$$J_j = (A_j' W_r A_j)^{-1} H_j' \big(H_j C_j H_j'\big)^{-1} H_j (A_j' W_r A_j)^{-1}, \quad j = 1, 2, \ldots, \tag{44}$$

$$C_j = (A_j' W_r A_j)^{-1} A_j' W_r B_r W_r A_j (A_j' W_r A_j)^{-1}, \quad j = 1, 2, \ldots, \tag{45}$$

and B_j is defined in (40). The LM statistics simplify to[14]

$$R_j = M_r' W_r A_j J_j A_j' W_r M_r, \quad j = 1, 2, \ldots, \tag{46}$$

which is the same formula as (19) used in Section 2.2. The theorem below is the equivalent of Theorem 1.

Theorem 2 *Assumption 1 holds and regularity conditions are satisfied.*[15] *For each $j = 1, 2, \ldots$, suppose $l_j = l_0$ and $W_j = W_0$. For each $j = 1, 2, \ldots$, suppose there exists a non-*

$\pi > 1$ and for every $\epsilon > 0$, there is a positive integer j_0 such that $\mathsf{P}\big(\max_{j_0 \le j \le r} R_j/\lambda_j \le (\pi+1)/2\big) < \epsilon$ for all sufficiently large n. Here π denotes the critical value used in the test.

[14] If an optimal weight matrix is used, so W_r and B_r^{-1} are equivalent, then R_j in (46) is the same as LM_{2n} in Table 2 in the article by Newey and McFadden (1994). In this case J_j simplifies to $J_j = (A_j' W_r A_j)^{-1} H_j' (H_j (A_j' W_r A_j)^{-1} H_j')^{-1} H_j (A_j' W_r A_j)^{-1}$ for $j = 1, 2, \ldots$.

[15] See footnote 13.

stochastic matrix, Σ_j, such that $n^{1/2}\mathsf{D}q_j(\tilde{\delta}_{0j}) \to^d N(0_{(h+\lambda_j\times 1)}, \Sigma_j)$ and $\mathrm{Var}(\mathsf{D}q_j(\tilde{\delta}_{0j})) \to^p \Sigma_j$ as $n \to \infty$ and such that the first h rows and columns of Σ_j consist of 0s and the lower right $\lambda_j \times \lambda_j$ submatrix of Σ_j is positive definite. Then under the null the asymptotic distribution of S_r as $r \to \infty$ and $n \to \infty$ is given in (39).

The proofs of the theorems are omitted, since they are similar to the proof of Theorem 3 by Aerts et al. (1999).

3.5 Remarks

We conclude this section with some remarks. First, because of the LM approach, parameter estimates need only be calculated once. In some applications, not having to estimate the model under the alternative is an advantage. For example, it is often difficult to estimate models when the first-order conditions are nonlinear in the parameters.

Second, note that essentially the same assumptions underpin both Theorem 1 and Theorem 2. In practice, one therefore has a choice of whether to implement the test using the minimum number of moment restrictions or using the same set of moment restrictions.

Third, the weighting of the LM statistics means that the ordering of the terms in the series approximation matters for the numerical value of the GMM-ACH test statistic. This issue also arises in nonparametric estimation based on series. The advice from that literature is to ensure that "important terms" are at the beginning of the series (see e.g. Gallant, 1981).

Fourth, a sequence of approximating functions with a multivariate argument ($d > 1$) can be constructed by interacting d univariate bases (see e.g. Donald et al., 2003, p59; Aerts et al., 2000, p413). To formalize this, associate with each j a d-tuple of nonnegative integers $(\tau_{j1}, \dots, \tau_{jd})$ for $j = 1, 2, \dots$. Let $e = (e_1, \dots, e_d)' \in \mathbb{R}^d$ and define $b_j(e_k) = 1$ for $j = 0$. Then define $c_j = \prod_{k=1}^d b_{\tau_{jk}}(e_k)$ for $j = 1, 2, \dots$. For c_1, c_2, $\dots$ to be a basis, it is important that all possible d-tuples are included in the sequence. That is, given any $(\tau_1, \dots, \tau_d)$ there is a j^* such that $c_{j^*} = \prod_{k=1}^d b_{\tau_k}(e_k)$. The exception from this rule is basis functions which are omitted because of collinearity with the null model.

Fifth, the ordering is particularly ambiguous in applications where γ is a function of a vector. Let Λ_j denote the set of d-tuples $(\tau_1, \ldots, \tau_d)$ which correspond to the sequence of basis functions $c_1, \ldots, c_{\lambda_j}$. Note that $\Lambda_1 \subset \Lambda_2 \subset \cdots$ by construction. Aerts et al. (2000) proposed to use index sets in which each argument is treated symmetrically. That is, if $(\tau_1, \ldots, \tau_d)$ is in Λ_j then all permutations are also in Λ_j. Also, in order to preserve power, they preferred sequences in which at most $d!$ terms are added in each step. The desired sequence can be constructed, for example, by sorting the indices first by their sum and second by their maximum value; the Λ_js are defined by selecting the subsequences which satisfy the symmetry requirement.[16]

Sixth, the GMM-ACH test can be used as a test of overall misspecification of conditional moment restrictions. Suppose for concreteness that v_i consists of x_i and z_i, which may have common elements, and that the null model is $\mathsf{E}\big(\rho(x_i, \beta^*)|z_i\big) = 0$, where β^* is an unknown parameter vector and ρ is a known vector-valued function. To test the null against its negation, let d be the dimension of x_i, p the dimension of z_i, and take

$$F_j(v_i, \beta, g_j(\cdot, \theta_j)) = \big(\rho(x_i, \beta) + g_j(x_i, \theta_j)\big) \otimes \big(1, c_1^z(z_i), \ldots, c_{l_j-1}^z(z_i)\big)', \tag{47}$$

where $1, c_1^z, c_2^z, \ldots$ is a sequence of basis functions ($c_k^z : \mathbb{R}^p \to \mathbb{R}$ for $k = 1, 2, \ldots$) such that (47) with $\theta_j = 0_{(\lambda_j \times 1)}$ is equivalent to the null model. Equivalence is not difficult to achieve (see e.g. Bierens, 1982; Bierens, 1990; de Jong and Bierens, 1994; Bierens and Ploberger, 1997; Stinchcombe and White, 1998; Donald et al., 2003; Escanciano, 2009). If the basis functions are chosen appropriately, $l_j = h + \lambda_j$, and the test is implemented using the minimum number of moment conditions, consistency follows by an argument parallel to that outlined for the simple IV model at the end of Section 2.1.[17]

Seventh, consistency of the GMM-ACH test in general can be proved using arguments similar to those needed to prove Theorem 4 of Aerts et al. (1999), which concerns consistency of a score-based test in likelihood models. We offer the following comments. As mentioned in Section 2.1, when A_j is invertible then the LM statistic does not directly

[16]For $d = 3$ let $P(\tau_1, \tau_2, \tau_3)$ denote the set of permutations of (τ_1, τ_2, τ_3), then $\Lambda_1 = P(0, 0, 1)$ with $\lambda_1 = 3$, $\Lambda_2 = \Lambda_1 \cup P(0, 1, 1)$ with $\lambda_2 = 6$, $\Lambda_3 = \Lambda_2 \cup P(0, 0, 2)$ with $\lambda_3 = 9$, $\Lambda_4 = \Lambda_3 \cup (1, 1, 1)$ with $\lambda_4 = 10$, $\Lambda_5 = \Lambda_4 \cup P(0, 1, 2)$ with $\lambda_5 = 16$, $\Lambda_6 = \Lambda_5 \cup P(0, 0, 3)$ with $\lambda_6 = 19$, etc.

[17]Since g_j drops out of the LM statistics when the minimum number of moment restrictions are used (see Section 2.1), it does not matter in (47) that ρ is a vector while g_j is a real function.

depend on how the alternative model is specified. However, it depends indirectly on the alternative through the moment restrictions used to identify the parameters of the alternative model. When there are more moment restrictions than parameters under the alternative, then A_j is not invertible and, in general, the LM statistic will depend directly on the specification of the alternative model. By explicitly specifying an alternative model and choosing moment restrictions accordingly, one can direct the power towards alternatives of particular interest. We provide Monte Carlo evidence in support of this idea in Section 4.3.

Eighth, when the same set of moment restrictions and the same weight matrix are used under the null as well as under the parametric alternatives, then the estimator, with 0s appended as appropriate, computed by solving the unrestricted problem of minimizing $q_0(\delta_0)$ with respect to δ_0 is identical to the estimators obtained by solving the restricted problem of minimizing $q_j(\delta_j)$ with respect to δ_j subject to $H_j\delta_j = 0_{(\lambda_j \times 1)}$ for $j = 1, 2, \ldots$. This can be seen from inspecting the first-order conditions.

Ninth, it is possible that there are other cases where S_r is asymptotically pivotal. A key property of the LM statistics under Theorem 2 is that the first λ_{j-1} components of $\mathsf{D}q_j(\tilde{\delta}_{0j})$ equal $\mathsf{D}q_{j-1}(\tilde{\delta}_{0,j-1})$ for all $j = 1, \ldots, r$. Mathematically, there are ways of achieving this which do not require using the same set of moment restrictions for all $j = 0, 1, \ldots$. Examining the first-order conditions, (34), reveals that the key property is also satisfied if the partial derivatives of the last $l_j - l_{j-1}$ components of the empirical moment function with respect to the first λ_{j-1} components of the parameter vector are all 0 and the weight matrix is block-diagonal with 0s in the first l_{j-1} rows (columns) of the last $l_j - l_{j-1}$ last columns (rows). The first requirement means that the additional $l_j - l_{j-1}$ moment restrictions must not depend on the previous λ_{j-1} parameters. If the moment restrictions are constructed by multiplying instruments and "residuals", then the additional $l_j - l_{j-1}$ instruments must be orthogonal to the partial derivatives of the residuals with respect to the previous λ_{j-1} parameters.[18] Thus, while it may be possible to construct other LM-based GMM-ACH test statistics, the requirements are complicated

[18] If the weight matrix is constructed using the second moments of a set of instruments, then the additional $l_j - l_{j-1}$ instruments must also be orthogonal to the previous λ_{j-1} instruments.

and seem less generalizable. Hence, we do not further pursue this possibility.

4 Examples and a second Monte Carlo study

In this section we consider single-equation linear models with endogenous right-hand side variables. The first two subsections explain how the GMM-ACH test can be used to test against misspecification in a specific direction. We consider a test against misspecification with respect to a single regressor and with respect to a linear combination (an index) of regressors. The potential gains from using a directed test instead of a test of overall misspecification is explored in a second Monte Carlo study. The final subsection contains an empirical application.

The setup is the following. Let y_i be a scalar random variable as in Section 2, but now let x_i and z_i be random vectors. Also, partition $x_i = (w_{1i}, w_{2i}')'$ where w_{1i} is scalar. A constant may be included in x_i and z_i. Suppose the parametric model of interest is

$$y_i = w_{1i}\beta_1^* + w_{2i}'\beta_2^* + u_i, \quad \mathsf{E}(u_i|z_i) = 0, \quad \beta^* \in \mathbb{R}^h, \tag{48}$$

where $\beta^* = (\beta_1^*, \beta_2^{*\prime})'$ is an unknown parameter vector and u_i is an unobserved random variable. As before, assume that n independent observations are available.

Equations of this form arises often in economics. For example, let (48) represent an Engel curve where y_i is the share of total expenditure spent on certain items in household i, w_{1i} is the log of total expenditure (as an indicator of permanent income), w_{2i} represents household characteristics, and z_i includes the variables in w_{2i} as well as household income as the instrument for total expenditure. Then this is the well-known Working-Leser specification of the Engel curve relationship.

As another example, consider a simultaneous equation system representing demand and supply of a certain good. Let y_i be the log of the total (equilibrium) quantity of the good traded in market i, let w_{1i} be the log of the (equilibrium) price of the good, let w_{2i} represent the characteristics of buyers in market i, and let z_i include the variables in w_{2i} as well as characteristics of suppliers. Then (48) represents the structural demand equation.

4.1 Nonlinear effect of a single regressor

The first alternative specification we consider allows for a nonlinear effect in w_{1i}. In the Engel curve example, the alternative model represents a nonlinear permanent income effect. In the market demand example, the alternative model allows for a nonlinear price elasticity. Formally, the nonparametric alternative model is

$$y_i = w_{1i}\beta_1^* + w_{2i}'\beta_2^* + \gamma^*(w_{1i}) + u_i, \quad \mathsf{E}(u_i|z_i) = 0, \quad \beta^* \in \mathbb{R}^h, \quad \gamma^* \in \Gamma. \tag{49}$$

The approximate alternative models are

$$y_i = w_{1i}\beta_1^* + w_{2i}'\beta_2^* + g_j(w_{1i}, \theta_j^*) + u_{ji}, \quad \mathsf{E}(u_{ji}|z_i) \to 0 \text{ as } j \to \infty,$$
$$\beta^* \in \mathbb{R}^h, \quad \theta_j^* \in \mathbb{R}^j, \quad j = 1, 2, \ldots, \tag{50}$$

where $g_j(e, \theta_j) = \sum_{k=1}^{j} \theta_{jk} b_k(e)$ for $j = 1, 2, \ldots$ are the partial sums from the series approximations of γ and θ_1^*, θ_2^*, ... are pseudo-true values as defined earlier. (For simplicity, we have set $\lambda_j = j$.)

The main issue in applying the GMM-ACH test is to choose moment restrictions to estimate β^* under the null and to identify θ_1^*, θ_2^*, ... under the alternative. There are many potential restrictions to choose from in this model, since the conditional moment restriction implies an infinite number of unconditional moment restrictions which can be used for estimation and testing. In practice, under the null, the model is virtually always estimated using the restrictions

$$\mathsf{E}\big(z_i(y_i - w_{1i}\beta_1^* - w_{2i}'\beta_2^*)\big) = 0_{(l_0 \times 1)}, \quad \beta^* \in \mathbb{R}^h, \tag{51}$$

where l_0 is the dimension of z_i. Section 3 shows that there are two ways to proceed under the alternative.

If the number of instruments is equal to the number of endogenous variables and the null model is exactly identified, it is natural to base the test on Theorem 1. Define $z_{0i} = z_i$ and $z_{ji} = (z_i, b_1(z_i^1), \ldots, b_j(z_i^1))'$ for $j = 1, 2, \ldots$, where z_i^1 denotes one of the instruments and b_1, b_2, ... are the basis functions used in the series approximation (with any linear term removed). If w_{1i} is exogenous, the natural choice for z_i^1 is w_{1i} itself. If w_{1i} is endogenous, the natural choice is one of the variables excluded from x_i.[19] The moment restrictions are then

$$\mathsf{E}\big(z_{ji}(y_i - w_{1i}\beta_1^* - w_{2i}'\beta_2^* - g_j(w_{1i}, \theta_j^*))\big) \to 0_{(l_j \times 1)} \text{ as } j \to \infty,$$
$$\beta^* \in \mathbb{R}^h, \quad \theta_j^* \in \mathbb{R}^j, \quad j = 1, 2, \ldots, \quad (52)$$

where the number of moment restrictions is $l_j = h + j$ for $j = 0, 1, \ldots$.

Define $x_{0i} = (w_{1i}, w_{2i}')'$ and $x_{ji} = (w_{1i}, w_{2i}', b_1(w_{1i}), \ldots, b_j(w_{1i}))'$ for $j = 1, 2, \ldots$. Formally the matrices which are used in the LM statistics and the GMM-ACH test statistic are exactly as given in (9)–(13) in Section 2.1, with the symbols x_{ji} and z_{ji} as defined in the present section and with $\tilde{\beta}$ being the usual IV estimator. To base this test on Theorem 2 instead of Theorem 1, simply use formulae (19)–(23) in Section 2.2.

If the null model is overidentified, it is most natural to base testing on Theorem 2. In this case, the need to choose which moment restrictions to use to identify θ_1^*, θ_2^*, ... is perhaps even more apparent. At one extreme one can use a basis based on a single variable as in the previous case. At the other extreme one can use a basis based on all of the available instrumental variables. In the latter case, z_{ri} is redefined as $z_{ri} = (c_1(z_i), \ldots, c_{\lambda_r}(z_i))$, where c_1, c_2, ... is a basis with a vector argument as in Section 3.1. Which is better is likely to depend on the strength of the instruments in the particular application. In either case, the test is calculated using formulae (19)–(23).

[19] It is possible to derive optimal instruments when the unconditional moment restrictions are based on a conditional moment restriction, see e.g. Newey and McFadden (1994, Sections 5.3–5.4).

4.2 Nonlinear effect of an index

The second alternative specification we consider allows for a nonlinear effect in the index $w_{2i}'\beta^*$. Specifically,

$$y_i = w_{1i}\beta_1^* + w_{2i}'\beta_2^* + \gamma^*(w_{2i}'\beta^*) + u_i, \quad \mathsf{E}(u_i|z_i) = 0, \quad \beta^* \in \mathbb{R}^h, \quad \gamma^* \in \Gamma. \tag{53}$$

In the Engel curve and the market demand examples, one might consider this alternative in order to check the robustness of $\tilde{\beta}_1$ to misspecification of the influence of household characteristics or buyer characteristics.

The approximate alternative models are

$$y_i = w_{1i}\beta_1^* + w_{2i}'\beta_2^* + g_j(w_{2i}'\beta^*, \theta_j^*) + u_{ji}, \quad \mathsf{E}(u_{ji}|z_i) \to 0 \text{ as } j \to \infty,$$
$$\beta^* \in \mathbb{R}^h, \quad \theta_j^* \in \mathbb{R}^j, \quad j = 1, 2, \ldots, \tag{54}$$

where the partial sums g_1, g_2, $\ldots$ may be defined as in the previous subsection. Since there are no obvious single candidate instruments for the index, Theorem 2 may be better suited than Theorem 1.

Define $x_{0i} = (w_{1i}, w_{2i}')'$ and $x_{ji} = (w_{1i}, w_{2i}', b_1(w_{2i}'\tilde{\beta}), \ldots, b_j(w_{2i}'\tilde{\beta}))'$ for $j = 1, 2, \ldots$. If a power basis is used, then $x_{ji} = (w_{1i}, w_{2i}', (w_{2i}'\tilde{\beta})^2, \ldots, (w_{2i}'\tilde{\beta})^{1+j})'$, which shows a similarity with the well-known RESET test for functional form. Let z_{ji} be one of vectors discussed in Section 4.1. The test statistics based on Theorem 2 is given in (19).

Note that this test differs from the one discussed in the last paragraph of Section 4.1. The instrument vector z_{ri}, the vector of empirical moments restrictions evaluated under the null M_r, the weight matrix W_r, and the moment variance matrix B_r are all the same, but the matrix A_j is different because x_{ji} is different. When the same set of moment restrictions is used throughout, the LM statistics are functions of reweighted empirical moment restrictions. The reweighting is determined by the derivative of the moment restrictions with respect to the parameters θ_j under the alternative (i.e. by A_j). Since the alternative hypotheses are different here and in Section 4.1, the LM statistics are based on different reweightings of the (same) empirical moment restrictions. Intuitively, the resulting tests are powerful against the particular alternative of interest.

4.3 Monte Carlo study with multiple regressors

In this section, we present simulation results for a linear model with two endogenous regressors and two instruments. The purpose is to investigate the finite-sample properties of the GMM-ACH in multiple regression and to investigate the value of prior knowledge about restrictions on the alternative models.

It is clear that the dimension of the parameter in the approximating functions can increase very quickly with j, even for moderate-sized d. The ability of the test to detect smaller "high-frequency" deviations from the null is therefore decreasing in d. In some applications, it may be possible to alleviate this problem by restricting the kind of alternatives considered, such as γs which are additive in its arguments (see Aerts et al., 2000, for a discussion and simulation results). We report on two tests in this section. For the first test, the alternative model is a function of the second regressor only, and for the second test, the alternative model is a function of both regressors.

The data-generating process for the experiments presented in this section is

$$y_i = \beta_0 + \beta_{11}x_{1i} + \beta_{21}x_{2i} + \beta_{22}x_{2i}^2 + \beta_{23}x_{2i}^3 + u_i, \tag{55}$$

where $(u_i, x_{1i}, x_{2i}, z_{1i}, z_{2i})'$ is jointly normally distributed with mean 0 and correlation matrix

$$\begin{bmatrix} 1 & \sigma_{ux} & \sigma_{ux} & 0 & 0 \\ \sigma_{ux} & 1 & 0 & \sigma_{xz} & 0 \\ \sigma_{ux} & 0 & 1 & 0 & \sigma_{xz} \\ 0 & \sigma_{xz} & 0 & 1 & 0 \\ 0 & 0 & \sigma_{xz} & 0 & 1 \end{bmatrix} \tag{56}$$

Apart from introducing a second regressor and instrument, this design differs from the previous (see Section 2.3) by omitting the transformation of regressors and instruments to the $[0, 1]$ range. The values of the variance matrix are similar to those used in the previous experiments (prior to transformation by Φ). Specifically, $\text{Var}(u_i) = 0.2^2$, $\text{Var}(x_{1i}) = 1$, $\sigma_{xz} = \rho$ and $\sigma_{ux} = \eta(1 - \rho^2)^{1/2}$ with the same values of ρ and η as before.[20] The values of the βs are given in the table of results.

We consider eight different versions of the GMM-ACH test. Most importantly, we consider tests where the alternative is misspecification with respect to x_{2i} only and tests where the alternative is misspecification with respect to both x_{1i} and x_{2i}. The null model is correctly specified with respect to x_{1i} in all cases, so the former test is expected to have higher power due to the conservation of degrees of freedom. As in Section 2.3, we also consider both the version of the test which uses a minimum set of moment restrictions and the version which uses the same set of moment restrictions, and we consider both a power basis and a Fourier flexible form basis.

The results are shown in Table 2. Technical details of the implementation are given in the table notes. The five panels correspond to different assumptions about the null and the data-generating process. In the first two panels, the null is true. The level control is good, although in the some (quadratic) cases a bit low for the test which tests against misspecification with respect to x_{2i} only. The next three panels show the power of the test. The power is 100% or near 100% in all cases for the test against misspecification with respect to x_{2i} only. The power of the test against misspecification with respect to both x_{1i} and x_{2i} is 100% or near 100% when the null is linear and the data-generating process is quadratic, but fall dramatically in the simulations where the data-generating process is cubic. The power of the test which uses the minimum number of moment restrictions is never higher and in many cases lower than the power of test which uses the same set of moment restrictions in all calculations. There is no clear pattern when comparing the results for the tests based on the power basis and the Fourier flexible form basis.

[20]The combinations of $(\rho, \eta, \sigma_{ux})$ are (0.7,0.1,0.07), (0.8,0.1,0.06) and (0.8,0.5,0.30).

4.4 Empirical example

In this subsection, we apply the GMM-ACH test to the Engel curve model described earlier. We use the same data as Blundell et al. (1998); BDP henceforth.[21] The data come from the 1980–1982 British Family Expenditure Survey. The extract is limited to married or cohabiting couples with one or two children, living in Greater London or south-east England, where the head of the household is currently employed. For further details about the sample, including summary statistics, please see BDP's article.

One of the models considered by BDP has the form (48). In our notation, y_i is the share of total expenditure spent on certain items, w_{1i} is log of total expenditure, w_{2i} is a dummy for having two instead of one child in the family, and z_i includes w_{2i} as well as total disposable income. The alternative specification is given in (49). (Since w_{2i} is a dummy, the alternative given in (53) is not relevant.)

Table 3 shows estimation results using different parametric specifications and different estimation methods. The OLS estimates are similar to those reported in Tables II-VII by BDP, although not identical. The GMM-ACH tests reject the linear specification for fuel, transport and (marginally) for other goods. To help understand the outcome of the GMM-ACH tests, the last panel of Table 3 shows IV estimates for a model which is quadratic in the log of total expenditure. The statistical significance of the t-statistics for the coefficients on the squared terms agree with the GMM-ACH tests in all cases (the marginal case of other goods is only significant at the 5.7% level).

BDP also tested the linear model against a nonparametric alternative. Their approach is much more complicated than ours and involves estimating the model under the non-parametric alternative, a notoriously difficult problem. Their conclusions are different from ours. They rejected the linear specification for alcohol and other goods and no other categories. While the differences in conclusions are interesting, further investigation is beyond the scope of the present paper.

[21] These data are available from the Journal of Applied Econometrics' data archive.

5 Concluding remarks

Inspired by Aerts et al. (1999), we suggest an GMM-ACH specification test of a parametric function against a nonparametric alternative. The test is developed for models which are identified by moment restrictions. The test requires only estimation under the null, and hence nonparametric estimation is not involved. The GMM-ACH test is asymptotically pivotal, which makes it easy to obtain critical values.

In a small Monte Carlo study, the GMM-ACH test has good level and power properties compared to existing tests. The test developed by Horowitz (2006) tends to have the best power of all, but it is difficult to perform. The GMM-ACH test has power that is close to that of Horowitz's test, and it is easy to carry out. The simulations also show that the GMM-ACH test has substantially higher power than an LM test of the null against a single, high-order parametric alternative. Hence, the idea of combination of test statistics against a sequence of parametric alternatives proves to be valuable.

Originally, our interest in testing for functional form in GMM settings was motivated by dynamic panel data models with fixed effects. This particular application is relatively complex, partly because these models have several equations per subject and each equation has its own set of instruments, and partly because GMM estimation of these models in practice is often troubled by weak instruments. We intend to publish our results for this case separately.

A Estimating the variance of the GMM gradient

In this appendix, we derive the estimators of $\mathrm{Var}(\mathsf{D}q_j(\tilde{\delta}_{0j}))^-$ given in Section 3. Section A.1 shows that $\mathrm{Var}(\mathsf{D}q_j(\tilde{\delta}_{0j}))$ can be estimated consistently. Section A.2 shows that the LM statistic (37) simplifies to (43) in the case where $l_j = h + \lambda_j$ for all $j = 1, 2, \ldots$. Section A.3 establishes (44) for the case where $l_j = l_0$ for all $j = 1, 2, \ldots$. Throughout this appendix j is a fixed integer.

Our arguments in Section A.1 are similar to those given by Newey and McFadden (1994, Section 9). The main differences are that we consider the case where the restricted estimator may be based on a subset of the moment restrictions and where the weight matrix, W_j, is arbitrary. Newey and McFadden considered the case where the moment restrictions are identical under the null and the alternative and where W_j is an estimate of the optimal weight matrix.

A.1 The general case

In this section, we show that $\mathrm{Var}(\mathsf{D}q_j(\tilde{\delta}_{0j}))$, which appears in (37) in Section 3, can be estimated as

$$\mathrm{Var}(\mathsf{D}q_j(\tilde{\delta}_{0j})) = T_j B_j T_j', \quad j = 0, 1, \ldots, \tag{57}$$

where B_j is given in (40) and T_j are the matrices defined by

$$T_j = A_j' W_j \left[I_{(l_j)} - A_j N_{1j} \left(A_0' W_0 A_0 \right)^{-1} A_0' W_0 N_{2j} \right], \quad j = 0, 1, \ldots, \tag{58}$$

with

$$N_{1j} = \begin{bmatrix} I_{(h)} \\ 0_{(\lambda_j \times h)} \end{bmatrix}, \quad j = 0, 1, \ldots, \tag{59}$$

and

$$N_{2j} = \begin{bmatrix} I_{(l_0)} & 0_{(l_0 \times l_j - l_0)} \end{bmatrix}, \quad j = 0, 1, \ldots. \tag{60}$$

We use two key properties of the testing problem set up in Section 3, namely that the restricted estimator is $\tilde{\delta}_{0j} = (\tilde{\delta}_0', 0'_{(\lambda_j \times 1)})'$ where $\tilde{\delta}_0$ is the solution to the unrestricted minimization problem $\mathsf{D}q_0(\tilde{\delta}_0) = 0_{(h \times 1)}$, and that the first l_0 components of $m_j(\tilde{\delta}_{0j})$ equal $m_0(\tilde{\delta}_0)$. These properties are implied by Assumption 1.

In general, the GMM gradient evaluated at the unrestricted estimator is identically equal to $0_{(h+\lambda_j \times 1)}$. However, this is not the case when evaluated at the restricted estimator.

A Taylor series expansion of $m_j(\tilde{\delta}_{0j})$ around the "pseudo-true" value, δ_j^*, implies

$$n^{1/2}\mathsf{D}q_j(\tilde{\delta}_{0j}) = a_j(\tilde{\delta}_{0j})'W_j\big[n^{1/2}m_j(\delta_j^*) + a_j(\tilde{\delta}_{0j})n^{1/2}(\tilde{\delta}_{0j} - \delta_j^*)\big] + o_p(1), \tag{61}$$

where $\mathsf{D}q_j$, m_j and a_j are defined in (32), (33) and (35) and W_j is a given weight matrix. Under standard regularity conditions, $a_j(\tilde{\delta}_{0j})$ and W_j converge in probability to matrices of (finite) constants. Therefore the main sources of variation for $n^{1/2}\mathsf{D}q_j(\tilde{\delta}_{0j})$ are the empirical moments, $n^{1/2}m_j(\delta_j^*)$, and the estimated parameters, $n^{1/2}(\tilde{\delta}_{0j} - \delta_j^*)$.

In the present context, the restricted estimator has the form $\tilde{\delta}_{0j} = (\tilde{\delta}_0', 0'_{(\lambda_j\times 1)})'$, where $\tilde{\delta}_0$ is the solution to the (unrestricted) estimation problem, $\mathsf{D}q_0(\tilde{\delta}_0) = 0_{(h\times 1)}$. Under the null, the "pseudo-true" value can be similarly partitioned, $\delta_j^* = (\delta_0^{*\prime}, 0'_{(\lambda_j\times 1)})'$. To derive the distribution of $n^{1/2}(\tilde{\delta}_0 - \delta_0^*)$, note that a Taylor series expansion similar to (61) yields

$$n^{1/2}\mathsf{D}q_0(\tilde{\delta}_0) = a_0(\tilde{\delta}_0)'W_0\big[n^{1/2}m_0(\delta_0^*) + a_0(\tilde{\delta}_0)n^{1/2}(\tilde{\delta}_0 - \delta_0^*)\big] + o_p(1). \tag{62}$$

Since $\mathsf{D}q_0(\tilde{\delta}_0) = 0_{(h\times 1)}$, it follows that

$$n^{1/2}(\tilde{\delta}_0 - \delta_0^*) = -\big(a_0(\tilde{\delta}_0)'W_0a_0(\tilde{\delta}_0)\big)^{-1}a_0(\tilde{\delta}_0)'W_0n^{1/2}m_0(\delta_0^*) + o_p(1). \tag{63}$$

An approximation for $n^{1/2}(\tilde{\delta}_{0j} - \delta_j^*)$ follows by appending rows of zeros. With N_{1j} as defined in (59), $n^{1/2}(\tilde{\delta}_{0j}-\delta_j^*) = N_{1j}n^{1/2}(\tilde{\delta}_0-\delta_0^*)$. Before inserting into (61), it is convenient to express $m_0(\delta_0^*)$ in terms of $m_j(\delta_j^*)$. This will facilitate keeping track of the covariance between the empirical moments and the estimated parameters in (61). By construction, estimation under the null is based on the first l_0 moment restrictions out of a total of l_j restrictions under alternative j. This means that if $\delta_{0j} = (\delta_0', 0'_{(\lambda_j\times 1)})'$, then $m_0(\delta_0) = N_{2j}m_j(\delta_{0j})$, where N_{2j} is defined in (60). It follows that

$$n^{1/2}(\tilde{\delta}_{0j} - \delta_j^*) = -N_{1j}\big(a_0(\tilde{\delta}_0)'W_0a_0(\tilde{\delta}_0)\big)^{-1}a_0(\tilde{\delta}_0)'W_0N_{2j}n^{1/2}m_j(\delta_j^*) + o_p(1). \tag{64}$$

Inserting (64) into (61) then gives

$$n^{1/2}\mathsf{D}q_j(\tilde{\delta}_{0j}) = T_j n^{1/2} m_j(\delta_j^*) + o_p(1), \tag{65}$$

where T_j is defined in (58).

A central limit theorem implies $n^{1/2}m_j(\delta_j^*) \to^d N(0_{(h+\lambda_j\times 1)}, \Omega_j)$, where Ω_j is defined by $\Omega_j = \mathsf{E}\big(f_j(v_i,\delta_j^*)f_j(v_i,\delta_j^*)'\big)$ and f_j is defined in Section 3. It follows that the asymptotic variance of $\mathsf{D}q_j(\tilde{\delta}_{0j})$ can be estimated by $T_j\Omega_j T_j'$. Replacing Ω_j with the estimator B_j defined in (40) yields (57).

A.2 The case of $l_j = h + \lambda_j$

This section shows that the LM statistic (37) with variance estimator (57) simplifies to (43) in the case where $l_j = h+\lambda_j$, A_j is invertible and its upper left submatrix is A_0, and W_j is any nonsingular matrix.

When A_j and W_j are invertible, the LM statistic simplifies to

$$\begin{aligned} R_j &= M_j' W_j A_j \big(T_j B_j T_j'\big)^- A_j' W_j M_j \\ &= M_j' \big(U_j B_j U_j'\big)^- M_j, \end{aligned} \tag{66}$$

where

$$U_j = I_{(l_j)} - A_j N_{1j} A_0^{-1} N_{2j}. \tag{67}$$

Partition A_j and A_j^{-1} as

$$A_j = \begin{bmatrix} A_{00} & A_{0j} \\ A_{j0} & A_{jj} \end{bmatrix} \quad \text{and} \quad A_j^{-1} = \begin{bmatrix} A^{00} & A^{0j} \\ A^{j0} & A^{jj} \end{bmatrix}, \tag{68}$$

where A_{00} and A^{00} are h-dimensional and A_{jj} and A^{jj} are λ_j-dimensional square matrices. Assumption 1 implies that $A_{00} = A_0$. (Generally $A^{00} \neq A_0^{-1}$.) Using this result, U_j can be written

$$U_j = \begin{bmatrix} 0_{(h\times h)} & 0_{(h\times\lambda_j)} \\ -A_{j0}A_{00}^{-1} & I_{(\lambda_j)} \end{bmatrix}. \tag{69}$$

Partition B_j similarly to A_j. Then

$$U_jB_jU_j' = \begin{bmatrix} 0_{(h\times h)} & 0_{(h\times\lambda_j)} \\ 0_{(\lambda_j\times h)} & K_j \end{bmatrix}, \tag{70}$$

where K_j is defined as

$$K_j = A_{j0}A_{00}^{-1}B_{00}(A_{00}^{-1})'A_{j0}' - B_{j0}(A_{00}^{-1})'A_{j0}' - A_{j0}A_{00}^{-1}B_{j0}' + B_{jj}. \tag{71}$$

Partition M_j as

$$M_j = \begin{bmatrix} M_0 \\ M_{\star j} \end{bmatrix}, \tag{72}$$

where $M_0 = 0_{(h\times 1)}$ by the definition of the IV estimator and $M_{\star j}$ is a λ_j-vector. Since K_j is nonsingular, then

$$R_j = M_{\star j}'K_j^{-1}M_{\star j}. \tag{73}$$

Rules for inverting partitioned matrices imply that $A_{j0}A_{00}^{-1} = (A^{jj})^{-1}A^{j0}$. Substituting this into (71) and rearranging yield

$$\begin{aligned} K_j &= (A^{jj})^{-1}A^{j0}B_{00}(A^{j0})'(A^{jj})^{-1\prime} - B_{j0}(A^{j0})'(A^{jj})^{-1\prime} - (A^{jj})^{-1}A^{j0}B_{0j} + B_{jj} \\ &= (A^{jj})^{-1}\Big(A^{j0}B_{00}(A^{j0})' - A^{jj}B_{j0}(A^{j0})' - A^{j0}B_{0j}(A^{jj})' + A^{jj}B_{jj}(A^{jj})'\Big)(A^{jj})^{-1\prime}. \end{aligned} \tag{74}$$

Recall that $H_j = [0_{(\lambda_j\times h)}\, I_{(\lambda_j)}]$. From the last line in the previous expression it follows that

$$K_j = (A^{jj})^{-1} H_j A_j^{-1} B_j (A_j^{-1})' H_j' (A^{jj})^{-1\prime}. \tag{75}$$

Finally, noting that $H_j A_j^{-1} M_j = A^{jj} M_{\star j}$ yields

$$\begin{aligned} R_j &= M_{\star j}' \big((A^{jj})^{-1} H_j A_j^{-1} B_j (A_j^{-1})' H_j' ((A^{jj})^{-1})' \big)^{-1} M_{\star j} \\ &= M_{\star j}' (A^{jj})' \big(H_j A_j^{-1} B_j (A_j^{-1})' H_j' \big)^{-1} A^{jj} M_{\star j} \\ &= M_j' (A_j^{-1})' H_j' \big(H_j A_j^{-1} B_j (A_j^{-1})' H_j' \big)^{-1} H_j A_j^{-1} M_j. \end{aligned} \tag{76}$$

The last line is identical to (43).

A.3 The case of $l_j = l_0$

As noted by e.g. Engle (1984, p795), the form of LM statistics simplifies when the same set of moment restrictions is used both under the null and under the alternative; that is, when $l_0 = l_j$ and $W_0 = W_j$. Define $E_j = A_j' W_j A_j$ and $G_j = I_{(h+\lambda_j)} - E_j^{-1/2} H_j' (H_j E_j^{-1} H_j')^{-1} H_j E_j^{-1/2}$. Assumption 1 implies that the first h columns of A_j equal A_0. Using this fact and formulae for inverting partitioned matrices, it can be verified that

$$N_{1j} \big(A_0' W_0 A_0 \big)^{-1} A_0' W_0 N_{2j} = E_j^{-1/2} G_j E_j^{-1/2} A_j' W_j. \tag{77}$$

This result can be used to simplify the expression for T_j in (58),

$$\begin{aligned} T_j &= A_j' W_j \big[I_{(l_j)} - A_j E_j^{-1/2} G_j E_j^{-1/2} A_j' W_j \big] \\ &= \big[I_{(h+\lambda_j)} - A_j' W_j A_j E_j^{-1/2} G_j E_j^{-1/2} \big] A_j' W_j \\ &= \big[I_{(h+\lambda_j)} - E_j^{1/2} G_j E_j^{-1/2} \big] A_j' W_j \\ &= \big[I_{(h+\lambda_j)} - \big(I_{(h+\lambda_j)} - H_j' (H_j E_j^{-1} H_j')^{-1} H_j E_j^{-1} \big) \big] A_j' W_j \\ &= H_j' (H_j E_j^{-1} H_j')^{-1} H_j E_j^{-1} A_j' W_j. \end{aligned} \tag{78}$$

It follows that $T_j B_j T_j'$ simplifies to

$$T_j B_j T_j' = H_j'(H_j E_j^{-1} H_j')^{-1} H_j E_j^{-1} A_j' W_j B_j W_j A_j E_j^{-1} H_j'(H_j E_j^{-1} H_j')^{-1} H_j. \tag{79}$$

Using the definition of a generalized inverse, it is straightforward to verify that the expression $E_j^{-1} H_j' \big(H_j E_j^{-1} A_j' W_j B_j W_j A_j E_j^{-1} H_j'\big)^{-1} H_j E_j^{-1}$ is a generalized inverse of $T_j B_j T_j'$. The resulting estimator of $\mathrm{Var}(\mathsf{D}q_j(\tilde{\delta}_{0j}))$ is J_j given in (44). The corresponding LM statistic is given in (46).

References

Aerts, M., G. Claeskens, and J. D. Hart (1999). Testing the fit of a parametric function. *Journal of the American Statistical Association 94*(447), 869–879.

Aerts, M., G. Claeskens, and J. D. Hart (2000). Testing lack of fit in multiple regression. *Biometrika 87*(2), 405–424.

Andrews, D. W. K. (1997). A conditional Kolmogorov test. *Econometrica 65*(5), 1097–1128.

Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics 20*, 105–134.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica 58*(6), 1443–1458.

Bierens, H. J. and W. Ploberger (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica 65*, 1129–1151.

Blundell, R., A. Duncan, and K. Pendakur (1998). Semiparametric estimation and consumer demand. *Journal of Applied Econometrics 13*, 435–461.

de Jong, R. M. and H. J. Bierens (1994). On the limit behavior of a chi-square type test if the number of conditional moments tested approaches infinity. *Econometric Theory 10*(1), 70–90.

Donald, S. G., G. W. Imbens, and W. K. Newey (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics 117*, 55–93.

Engle, R. F. (1984). Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. In Z. Griliches and M. D. Intriligator (Eds.), *Handbook of Econometrics*, Volume II, pp. 775–826. Amsterdam: Elsevier Science Publishers BV.

Escanciano, J. C. (2007). Model checks using residual marked empirical processes. *Statistica Sinica 17*, 115–138.

Escanciano, J. C. (2009). On the lack of power of omnibus specification tests. *Econometric Theory 25*(1), 162–194.

Gallant, A. R. (1981). On the bias in flexible functional forms and an essentially unbiased form: The Fourier flexible form. *Journal of Econometrics 15*, 211–245.

Hart, J. D. (1997). *Nonparametric smoothing and lack-of-fit tests.* New York: Springer Verlag.

Horowitz, J. L. (2006). Testing a parametric model against a nonparametric alternative with identification through instrumental variables. *Econometrica 74*(2), 521–538.

Horowitz, J. L. and V. G. Spokoiny (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica 69*(3), 599–631.

Newey, W. K. (1985). Maximum likelihood specification testing and conditional moment tests. *Econometrica 53*(5), 1047–1070.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of econometrics. Volume 4. Handbooks in Economics, vol. 2.* Amsterdam; London and New York: Elsevier, North-Holland.

Pagan, A. and A. Ullah (1999). *Nonparametric econometrics.* Cambridge: Cambridge University Press.

Stinchcombe, M. B. and H. White (1998). Consistent specification testing with nuisance parameters present only under the alternative. *Econometric Theory 14*(3), 295–325.

Stute, W. (1997). Nonparametric model checks for regression. *Annals of Statistics 25*(2), 613–641.

Tripathi, G. and Y. Kitamura (2003). Testing conditional moment restrictions. *Annals of Statistics 31*(6), 2059–2095.

Van Keilegom, I., C. S. Sellero, and W. G. Manteiga (2008). Empirical likelihood based testing for regression. *Electronic Journal of Statistics 2*, 581–604.

Whang, Y.-J. (2001). Consistent specification testing for conditional moment restrictions. *Economics Letters 71*, 299–306.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica 50*(1), 1–26.

Table 1: Monte Carlo results for simple IV model (nominal size 5%)

ρ	η	HOR	t	DIN	ACH Min		ACH Same		LM Min		LM Same	
					P	F	P	F	P	F	P	F
					Null is true							
Null: linear; DGP: $y_i = 0.5x_i + u_i$												
0.8	0.1	5.1	5.2	4.8	5.2	5.2	5.6	5.0	4.1	4.2	4.2	4.2
0.8	0.5	3.0	3.4	4.3	4.1	3.4	3.5	3.1	3.6	5.6	3.7	5.8
0.7	0.1	4.9	5.2	4.5	5.1	5.2	5.4	4.6	4.3	4.1	4.3	4.1
Null: quadratic; DGP: $y_i = 0.5x_i - 0.5x_i^2 + u_i$												
0.8	0.1	5.3	4.0	4.8	5.0	4.3	5.1	4.2	4.4	4.1	4.5	4.2
0.8	0.5	4.6	7.7	5.0	7.5	3.4	3.8	2.0	5.7	5.7	7.2	6.1
0.7	0.1	5.6	3.6	4.3	5.6	4.3	5.4	4.5	4.5	4.2	4.9	4.3
					Null is false							
Null: linear; DGP: $y_i = 0.5x_i - 0.5x_i^2 + u_i$												
0.8	0.1	65.8	71.4	44.7	69.2	69.8	71.1	70.0	39.3	39.3	39.7	39.4
0.8	0.5	72.1	82.7	45.9	78.4	78.2	81.0	79.1	49.9	50.3	50.2	50.4
0.7	0.1	42.1	44.4	25.9	42.1	42.7	45.2	46.9	22.8	22.1	22.8	22.3
Null: linear; DGP: $y_i = 0.5x_i - x_i^2 + x_i^3 + u_i$												
0.8	0.1	68.4	67.1	49.8	64.0	62.7	65.1	64.1	40.3	39.0	40.4	39.0
0.8	0.5	66.3	58.0	48.0	56.6	52.6	55.6	54.5	30.7	34.1	32.0	35.4
0.7	0.1	42.4	41.2	26.2	36.2	36.1	38.3	37.9	17.8	17.1	18.4	16.8
Null: quadratic; DGP: $y_i = 0.5x_i - x_i^2 + 4x_i^3 + u_i$												
0.8	0.1	89.0	90.0	72.2	86.8	56.8	93.4	74.8	68.3	65.0	69.1	65.8
0.8	0.5	97.2	98.7	68.5	98.0	82.3	97.7	80.4	83.8	78.0	85.3	79.3
0.7	0.1	52.7	59.0	29.8	49.1	18.2	67.1	34.8	27.6	25.9	29.5	27.2

Legend: HOR: test by Horowitz (2006); t: ordinary t test for adding one additional term to the null model; DIN: the IV test by Donald, Imbens, and Newey (2003); ACH Min: implemented as in Section 2.1; ACH Same: implemented as in Section 2.2; LM Min: the rth LM statistic from the ACH Min calculations; LM Same: the rth LM statistic from the ACH Same calculations; P: based on power basis; F: based on Fourier flexible form basis; null linear: $y_i = \beta_0 + \beta_1 x_i + u_i$; null quadratic: $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + u_i$. *Notes:* HOR, t and DIN quoted from Horowitz (2006). There are 500 observations in each sample and 1000 samples per experiment. In the calculations of the GMM-ACH tests, $r = 6$ and all additional terms under the alternative are orthogonalized to reduce multicollinearity. For the last set of experiments, the dgp process is incorrectly stated in Horowitz's article with the term $2x_i^3$ instead of $4x_i^3$.

Table 2: Monte Carlo results for multi-regressor IV model (nominal size 5%)

		Testing against mis-specification wrt x_{2i} only				Testing against mis-specification wrt (x_{1i}, x_{2i})			
σ_{xz}	σ_{xu}	ACH Min		ACH Same		ACH Min		ACH Same	
		P	F	P	F	P	F	P	F
				Null is true					
Null: linear; DGP: $y_i = 0.5x_{1i} + 0.5x_{2i} + u_i$									
0.7	0.07	4.9	4.9	4.8	4.6	4.8	4.8	4.1	4.8
0.8	0.06	4.6	4.8	4.4	4.5	5.7	5.7	5.7	5.0
0.8	0.30	5.2	5.0	4.5	4.0	5.6	5.6	5.5	5.2
Null: quadratic; DGP: $y_i = 0.5x_{1i} + 0.5x_{2i} - 0.5x_{2i}^2 + u_i$									
0.7	0.07	3.6	4.4	5.0	4.1	5.0	5.0	5.2	4.7
0.8	0.06	3.8	4.2	3.3	3.0	6.1	6.1	5.6	5.2
0.8	0.30	3.5	3.4	4.5	3.5	5.4	5.4	5.6	5.4
				Null is false					
Null: linear; DGP: $y_i = 0.5x_{1i} + 0.5x_{2i} - 0.5x_{2i}^2 + u_i$									
0.7	0.07	100.0	100.0	100.0	100.0	98.3	99.1	97.2	99.1
0.8	0.06	100.0	100.0	100.0	100.0	99.9	100.0	99.6	100.0
0.8	0.30	100.0	100.0	100.0	100.0	99.9	100.0	99.7	100.0
Null: linear; DGP: $y_i = 0.5x_{1i} + 0.5x_{2i} - x_{2i}^2 + x_{2i}^3 + u_i$									
0.7	0.07	91.6	93.1	92.2	94.5	19.7	22.2	19.6	24.8
0.8	0.06	99.9	100.0	99.9	100.0	33.2	51.4	34.8	52.3
0.8	0.30	99.9	100.0	99.9	100.0	33.7	49.3	35.4	50.6
Null: quadratic; DGP: $y_i = 0.5x_{1i} + 0.5x_{2i} - x_{2i}^2 + 4x_{2i}^3 + u_i$									
0.7	0.07	90.8	73.8	90.7	79.1	6.4	5.1	11.4	5.9
0.8	0.06	100.0	99.8	100.0	99.8	16.2	5.1	22.0	6.1
0.8	0.30	100.0	99.7	99.9	99.8	17.1	6.8	20.5	6.5

Legend: ACH Min: implemented using minimum number of moment restrictions; ACH Same: implemented using same set of moment restrictions; P: based on power basis; F: based on Fourier flexible form basis; null linear: $y_i = \beta_0 + \beta_{11}x_{1i} + \beta_{21}x_{2i} + u_i$; null quadratic: $y_i = \beta_0 + \beta_{11}x_{1i} + \beta_{21}x_{2i} + \beta_{22}x_{2i}^2 + u_i$; wrt: with respect to. *Notes:* There are 500 observations in each sample and 1000 samples per experiment. In the calculations of the GMM-ACH tests, $r = 6$, the ordering of basis functions is as described in Section 3.5, and no orthogonalization is carried out.

Table 3: Engel curve estimates

	Share of total expenditures					
	Food	Fuel	Clothes	Alcohol	Transport	Other
Summary statistics for dependent variable						
Mean	.3565	.0910	.1072	.0606	.1324	.2523
Simple linear model, OLS estimates						
$\tilde{\beta}_1$	$-.1338^*$	$-.0472^*$	$.0813^*$	$.0198^*$	$.0394^*$	$.0406^*$
	(.0060)	(.0032)	(.0059)	(.0041)	(.0069)	(.0067)
Linear model with demographics, OLS estimates						
$\tilde{\beta}_1$	$-.1384^*$	$-.0474^*$	$.0819^*$	$.0216^*$	$.0411^*$	$.0412^*$
	(.0060)	(.0032)	(.0059)	(.0042)	(.0069)	(.0068)
$\tilde{\beta}_2$	$.0338^*$	.0012	$-.0045$	$-.0129^*$	$-.0130^*$	$-.0047$
	(.0048)	(.0026)	(.0047)	(.0033)	(.0055)	(.0054)
Linear model with demographics, IV estimates						
$\tilde{\beta}_1$	$-.1412^*$	$-.0274^*$	$.0473^*$	.0156	$.0295^*$	$.0762^*$
	(.0122)	(.0067)	(.0123)	(.0085)	(.0142)	(.0140)
$\tilde{\beta}_2$	$.0341^*$	$-.0005$	$-.0015$	$-.0124^*$	$-.0119^*$	$-.0077$
	(.0048)	(.0026)	(.0049)	(.0034)	(.0056)	(.0055)
GMM-ACH test of the linear model with demographics						
ACH Min	0.719	6.556^*	2.145	0.530	14.268^*	3.950
ACH Same	1.200	15.594^*	1.013	0.531	16.243^*	5.033^*
Quadratic model with demographics, IV estimates						
$\tilde{\beta}_1$	$-.0618$	-2.1008^*	.9794	$-.0855$	2.7383^*	-1.4708
	(.6782)	(.5065)	(.6938)	(.4740)	(.9295)	(.8135)
$\tilde{\beta}_2$	$.0336^*$	$.0112^*$	$-.0068$	$-.0119^*$	$-.0273^*$	.0011
	(.0063)	(.0047)	(.0064)	(.0044)	(.0086)	(.0075)
$\tilde{\beta}_3$ (w_{1i}^2)	$-.0086$	$.2256^*$	$-.1014$	.0110	$-.2947^*$	.1683
	(.0736)	(.0549)	(.0752)	(.0514)	(.1008)	(.0882)

Legend: $\tilde{\beta}_1$: coefficient on log total expenditure; $\tilde{\beta}_2$: coefficient on indicator of two children; $\tilde{\beta}_3$: coefficient on the square of log total expenditure; ACH Min: implemented using the minimum number of moment restrictions; ACH Same: implemented using the same set of moment restrictions; standard errors in ( ) parentheses; *: statistical significance at the 5% level. *Notes:* Constant included in all models, but not reported. GMM-ACH tests based on a power basis, $r = 6$, and all additional terms under the alternative are orthogonalized to reduce multicollinearity. Data from Blundell, Duncan, and Pendakur (1998). Number of observations: 1519.