Economics Department

Working Papers in Economics

Boston College *Year* 2004

Identification of Endogenous Heteroskedastic Models

Arthur Lewbel
Boston College,

Identification of Endogenous Heteroskedastic Models

Arthur Lewbel   Boston College

Revised March 2004

Abstract

This paper proposes a new method of obtaining identification in mismeasured regressor models, triangular systems, linear simultaneous equation systems, and structural vector autoregressions. Associated estimators take the form of ordinary two stage least squares or generalized method of moments. The method may be used in applications where other sources of identification such as instrumental variables or repeated measurements are not available. Identification comes from a heteroskedastic covariance restriction that is shown to be a feature of many endogeneity and measurement error models. Identification is also obtained in some semiparametric partly linear models. An empirical application and a Monte Carlo study are provided.

Keywords: Simultaneous systems, endogeneity, identification, heteroscedasticity, measurement error, partly linear models. JEL codes: C3, C13, C14, D12

This research was supported in part by the National Science Foundation through grant SES-9905010. I would like to thank Roberto Rigobon, Frank Vella, Todd Prono, Susanne Schennach, Raffaella Giacomini, Tiemen Woutersen, Christina Gathmann, and Jim Heckman for helpful comments. Any errors are my own.

Corresponding Author: Arthur Lewbel, Department of Economics, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA, 02467, USA. (617)-552-3678, lewbel@bc.edu, http://www2.bc.edu/~lewbel/

1 Introduction

This paper provides a new method of identifying structural parameters in models with endogeneous or mismeasured regressors. The method may be used in applications where other sources of identification such as instrumental variables, repeated measurements, or validation studies are not available. The identification comes from having regressors uncorrelated with the product of heteroskedastic errors, which is shown to be a feature of many models in which error correlations are due to an unobserved common factor, such as unobserved ability in returns to schooling models, or the measurement error in mismeasured regressor models.

Estimators take the form of modified two stage least squares or generalized method of moments. Identification of semiparametric partly linear triangular and simultaneous systems are also considered. A Monte Carlo study is provided. In an empirical application, this paper's methodology is applied to deal with measurement error in total expenditures, resulting in Engel curve estimates that are similar to those obtained using an external instrument.

Let Y_1 and Y_2 be observed endogenous variables, let X be a vector of observed exogeneous regressors, and let $\varepsilon = (\varepsilon_1, \varepsilon_2)$ be unobserved errors. Consider structural models of the form

$$\begin{aligned} Y_1 &= X'\beta_1 + Y_2\gamma_1 + \varepsilon_1 \\ Y_2 &= X'\beta_2 + Y_1\gamma_2 + \varepsilon_2 \end{aligned}$$

This system of equations is triangular when $\gamma_2 = 0$, otherwise it is fully simultaneous (if it is known that $\gamma_1 = 0$, renumber the equations to set $\gamma_2 = 0$) The errors ε_1 and ε_2 may be correlated with each other.

Assume $E(\varepsilon X) = 0$. This is of course not sufficient to identify the structural model's coefficients. Typically, identification is obtained by imposing equality constraints on the coefficients, such as that some elements of β_1 or β_2 be zero, or equivalently by assuming the availability of instruments. This paper instead considers obtaining identification by restricting correlations of the second moments of the errors with X. This does not trivially provide identification. In particular, the structural model parameters remain unidentified under the standard homoskedasticity assumption that $E(\varepsilon\varepsilon' \mid X)$ is constant, and more generally are not identified when ε and X are independent.

In contrast, this paper shows that the model parameters may be identified given heteroskedasticity. In particular, identification is obtained by assuming that $E(X\varepsilon) = 0$, that $Cov(X, \varepsilon_j^2) \neq 0$ for $j = 2$ in a triangular system (or for both

$j = 1$ and $j = 2$ in a fully simultaneous system) and that $Cov(Z, \varepsilon_1\varepsilon_2) = 0$, where Z can be a subset of X. Assuming that the reduced form equations are linear projections of Y on X, there errors ε will satisfy $E(X\varepsilon) = 0$ by construction, and typically such errors, defined only by projections, can be expected to be heteroskedastic, making $Cov(X, \varepsilon_j^2) \neq 0$. Therefore, for applying this paper's estimator, the main issue is having $Cov(Z, \varepsilon_1\varepsilon_2) = 0$ hold. Consider some examples.

1.1 Classical Measurement Error

As a motivating example, consider the mismeasured regressor model. The goal is estimation of the coefficients β_1 and γ_1 in

$$Y_1 = X'\beta_1 + Y_2^*\gamma_1 + V_1$$

where the regression error V_1 is mean zero and independent of the covariates X, Y_2^*. However the scalar regressor Y_2^* is mismeasured, and we instead observe Y_2 where

$$Y_2 = Y_2^* + U, \qquad E(U) = 0, \qquad U \perp X, Y_1, Y_2^*$$

where U is classical measurement error, so U is mean zero and independent of the true model components X, Y_2^*, and V_1, or equivalently, independent of X, Y_2^*, and Y_1.

Define V_2 as the residual from a linear projection of Y_2^* on X, so by construction

$$Y_2^* = X'\beta_2 + V_2, \qquad E(XV_2) = 0$$

Substituting out the unobservable Y_2^* yields the triangular system

$$\begin{aligned} Y_1 &= X'\beta_1 + Y_2\gamma_1 + \varepsilon_1, \quad & \varepsilon_1 = -\gamma_1 U + V_1 \\ Y_2 &= X'\beta_2 + \varepsilon_2, & \varepsilon_2 = U + V_2 \end{aligned}$$

where ε_1 and ε_2 are unobserved errors.

The standard way to obtain identification in this model is by an exclusion restriction, that is, one or more elements of β_1 are assumed to equal zero and the corresponding elements of β_2 are assumed to be nonzero. The corresponding elements of X are then instruments, and the model is estimated by two stage least squares, with $Y_2 = X'\beta_2 + \varepsilon_2$ being the first stage regression and the second stage is the regression of Y_1 on the subset of X that has nonzero coefficients and on $\widehat{Y}_2$.

Assume now that we have no exclusion restriction and hence no instrument, so there is no covariate that we are confident affects Y_2 without also affecting Y_1. In that case, the structural model coefficients cannot be identified in the usual way, and in particular, are not identifed when U, V_1, and V_2 are normal and independent of X

However, in this mismeasured regressor model, there is no reason to believe that V_2 would be independent of X, since the Y_2 equation (the first stage equation) is just the linear projection of Y_2 on X, not a structural model motivated by economic theory. The perhaps surprising result, which follows from Theorem 1 below, is that, *if V_2 is heteroskedastic (and hence not independent of X), then the structural model coefficients in this model are identified.* This is because the above assumptions yield a triangular model with $E(X\varepsilon) = 0$, $Cov(X, \varepsilon_2^2) \neq 0$, and $Cov(X, \varepsilon_1\varepsilon_2) = 0$, and hence satisfy this paper's required conditions for identification with $Z = X$, (or more precisely, with Z equal to all the elements of X except the constant).

The classical measurement error assumptions above are used here by way of illustration. They are much stronger than necessary to apply this paper's methodology. For example, by taking Z to be a subset of elements of X, we may permit the measurement error U to be correlated with the other elements of the vector of regressors X. Also, the various error independence assumptions given above may be relaxed to restrictions on just a few first and second moments, as described below.

1.2 Unobserved Single Factor Models

A general class of models that satisfy this paper's assumptions are systems in which the correlation of errors across equations are due to the presence of an unobserved common factor U, that is

$$Y_1 = X'\beta_1 + Y_2\gamma_1 + \varepsilon_1, \qquad \varepsilon_1 = \alpha_1 U + V_1 \tag{1}$$

$$Y_2 = X'\beta_2 + Y_1\gamma_2 + \varepsilon_2, \qquad \varepsilon_2 = \alpha_2 U + V_2 \tag{2}$$

where U, V_1, and V_2 are unobserved variables that are uncorrelated with X and are conditionally uncorrelated with each other, conditioning on X. Here V_1 and V_2 are idiosyncratic errors in the equations for $Y_1$and Y_2, respectively, while U is an omitted variable or other unobserved factor that may directly influence both Y_1 and Y_2.

Examples:

MEASUREMENT ERROR. The mismeasured regressor model described above yields equation (1) with $\alpha_1 = -\gamma_1$ and equation (2) with $\gamma_2 = 0$ and $\alpha_2 = 1$. The unobserved common factor U is the measurement error in Y_2.

SUPPLY AND DEMAND. Equations (1) and (2) are supply and (inverse) demand functions, with Y_1 being quantity and Y_2 price. V_1 and V_2 are unobservables that only affect supply and demand, respectively, while U denotes an unobserved factor that affects both sides of the market, such as the price of an imperfect substitute.

RETURNS TO SCHOOLING. Equations (1) and (2) with $\gamma_2 = 0$ are models of wages Y_1 and schooling Y_2, with U representing an individual's unobserved ability or drive (or more precisely the residual after projecting unobserved ability on X), which affects both her schooling and her productivity (Heckman 1974, 1976).

STRUCTURAL VECTOR AUTOREGRESSION. Y_1 and Y_2 are stationary processes jointly determined by equations (1) and (2) with $U = 0$ and X consisting of lagged values of Y_1 and Y_2.

In each of these examples, some or all of the structural parameters are not identified without additional information. Typically, identification is obtained by imposing equality constraints on the coefficients of X. In the measurement error and returns to schooling examples, assuming that one or more elements of β_1 equal zero permits estimation of the Y_1 equation using two stage least squares with instruments X. For supply and demand, the typical identification restriction is that each equation possess this kind of exclusion assumption. In the structural VAR, having $U = 0$ makes the errors ε_1 and ε_2 independent, and at least one additional restriction on coefficients is typically used to obtain identification, such as constraining some element of β_1 or β_2 to equal zero or imposing the Blanchard and Quah (1989) long run zero restriction on coefficients.

A new way of obtaining identification in these models is proposed here. Assume we have no ordinary instruments and no equality constraints on the parameters. Let Z be a vector of observed exogeneous variables, in particular, Z could be a subvector of X, or Z could equal X. Assume (X, Z) is uncorrelated with (U, V_1, V_2). Assume also that Z is uncorrelated with (U^2, UV_j, V_1V_2) and that Z is correlated with V_2^2. If the model is simultaneous assume that Z is also correlated with V_1^2. Then

$$cov(Z, \varepsilon_1\varepsilon_2) = cov(Z, \alpha_1\alpha_2U^2 + \alpha_1UV_2 + \alpha_2UV_1 + V_1V_2) = 0$$

$$cov(Z, \varepsilon_2^2) = cov(Z, \alpha_2^2U^2 + 2\alpha_2UV_2 + V_2^2) = cov(Z, V_2^2) \neq 0$$

which are the requirements for this paper's identification theorems.

Stronger but more easily interpreted sufficient conditions to yield the identifying assumptions $cov(Z, \varepsilon_1\varepsilon_2) = 0$ and $cov(Z, \varepsilon_j^2) \neq 0$ are that one or both of the idiosyncratic errors V_j be heteroscedastic and that the common factor U be conditionally independent of Z. These conditions are discussed further below.

1.3 Overview

This paper's main result is that, in the model

$$\begin{aligned} Y_1 &= X'\beta_1 + Y_2\gamma_1 + \varepsilon_1 \\ Y_2 &= X'\beta_2 + Y_1\gamma_2 + \varepsilon_2 \end{aligned}$$

The parameters are identified if $E(X\varepsilon) = 0$, $Cov(Z, \varepsilon_1\varepsilon_2) = 0$, and $Cov(Z, \varepsilon_j^2) \neq 0$ for $j = 2$ in a triangular system (or for both $j = 1$ and $j = 2$ in a fully simultaneous system) where Z may (though need not be) a subset of X. The models described in the last two subsections are examples of models in which these assumptions are satisfied.

In triangular systems like the wage equation or measurement error models, it will be shown that β_1 and γ_1 can be consistently estimated without outside instruments by an ordinary linear two stage least squares regression of Y_1 on X and Y_2, using X and $(Z - \overline{Z})\widehat{\varepsilon}_2$ as instruments, where $\overline{Z}$ is the sample mean of Z and $\widehat{\varepsilon}_2$ is the residual from linearly regressing Y_2 on X. No outside instruments are required, because all of the elements of Z can also be elements of X.

More generally, for either triangular or simultaneous systems, estimation may be done using ordinary GMM (generalized method of moments; see Hansen 1982) based on the moments

$$E(X\varepsilon_1) = 0, \qquad E(X\varepsilon_2) = 0, \qquad Cov(Z, \varepsilon_1\varepsilon_2) = 0. \tag{3}$$

It is not necessary to assume that the errors are actually given by a factor model like $\varepsilon_j = \alpha_j U + V_j$. In particular, possible third and higher moment implications of factor model or classical measurement error constructions are not imposed. All that is required for identification and estimation are the moments (3) along with some heteroskedasticity of ε_j.

The moments (3) provide identification whether or not Z is subvector of X, but in the latter case Z might alternatively suffice as an ordinary outside instrument. It is perhaps more surprising that identification still obtains when Z is a subvector of X and no outside instruments and no restrictions on β_1 and β_2 are available.

An unusual feature of this identification result is that it only holds if the model errors are heteroskedastic. If the errors are homoskedastic then the instruments $(Z - \overline{Z})\widehat{\varepsilon}_2$ in the triangular model case or more generally the moments $cov(Z, \varepsilon_1\varepsilon_2) = 0$, will not provide enough additional information to yield identification.

The assumption used here that a product of errors be uncorrelated with covariates has occasionally been exploited in other contexts, e.g., to aid identification in a correlated random coefficients model, Heckman and Vytlacil (1998) assume covariates are uncorrelated with the product of a random coefficient and a regression model error.

Variables that in the past have been proposed as excluded instruments for identification might more plausibly be used as this paper's Z. For example, in the returns to schooling model Card (1995, 2002) and others propose using measures of access to schooling, such as distance to or cost of colleges in one's area, as wage equation instruments. Access measures are plausibly assumed to be independent of unobserved ability (though see Carneiro and Heckman 2002) and to affect the schooling decision. However, access may not be appropriate as an excluded variable in wage (or other outcome) equations because access may correlate with the resulting type or quality of education one actually receives, or may be correlated with proximity to urban areas and hence to availability of jobs. See, e.g., Hogan and Rigobon (2003). Therefore, instead of excluding measures of access to schooling or other proposed instruments from the outcome equation, it may be more appropriate to include them as regressors in both equations, and use them as this paper's Z to identify returns to schooling γ_1.

As shown earlier, in the classical measurement error model, this paper's assumptions hold with $Z = X$. The only nonstandard assumption that is needed to identify the coefficients in a classically mismeasured regressor model without outside instruments is the assumption that V_2 be heteroskedastic, which is both empirically testable and more plausible than homoskedasticity in most applications.

Similarly, structural vector autoregressions are typically estimated with macroeconomic data, where time varying volatility is the norm rather than the exception, which along with the standard assumption in such models that ε_1 and ε_2 are independent provides a natural justification for assuming both that $cov(Z, \varepsilon_1\varepsilon_2) = 0$ (where Z, like X, now equals lags of Y_1 and Y_2) and that ε_1 and ε_2 are heteroskedastic.

In addition to the above models, this paper also considers partly linear specifications, where both $X'\beta_1$ and $X'\beta_2$ are replaced with unknown functions $g_1(X)$

and $g_2(X)$. These g_j functions will be nonparametrically identified and can be estimated along with γ_1 and γ_2. In this case, instead of a simple GMM, estimation may take the form of a two step estimator with a nonparametric first step.

The topic of identification in simultaneous systems has a long history. See, e.g., the surveys Hsiao (1983), Hausman (1983), and Fuller (1987). Roehrig (1988) provides a useful general characterization of identification in situations where nonlinearities contribute to identification, as is the case here. Particularly relevant for this paper is previous work that obtains identification based on variance and covariance constraints. With multiple equation systems, various homoskedastic factor model covariance restrictions are used along with exclusion assumptions in the LISREL class of models (Joreskog and Sorbom 1984).

A closely related result to this paper's is Rigobon (2002, 2003). Relative to the above model, Rigobon uses heteroskedasticity based on discrete, multiple regimes instead of regressors. Some of Rigobon's identification results can be interpreted as special cases of this paper's models in which Z is a vector of binary dummy variables that index regimes and are not included amongst the regressors X. Sentana and Fiorentini (2001) employ a similar idea for identication in a factor model. Hogan and Rigobon (2003) propose a model that, like this paper's, involves decomposing the error term into components, some of which are heteroskedastic. Another related model is Klein and Vella (2003), who also obtain identification by restrictions on second moments, in their case using a specific semiparametric functional form of multiplicative heteroskedasticity. An example of using GARCH system heteroskedastic specifications to obtain identification is King, Sentana, and Wadhwani (1994).

Other recent papers that use restrictions on higher moments instead of outside instruments as a source of identification include Vella and Verbeek (1997), Dagenais and Dagenais (1997), Lewbel (1997), Cragg (1997), and Erickson and Whited (2002).

2 Identification

For simplicity it is assumed that the regressors X are ordinary random variables with finite second moments, so results are easily stated in terms of means and variances. However, it will be clear from the resulting estimators that this can be relaxed to handle cases such as time trends or deterministic regressors by replacing these moments with probability limits of sample moments and sample projections.

2.1 Triangular Model Identification

First consider the linear triangular model

$$Y_1 = X'\beta_{10} + Y_2\gamma_{10} + \varepsilon_1 \tag{4}$$

$$Y_2 = X'\beta_{20} + \varepsilon_2 \tag{5}$$

Here β_{10} indicates the true value of β_1, and similarly for the other parameters. Traditionally, this model would be identified by assuming some element of β_{10} were known to equal zero. Alternatively, if the errors ε_1 and ε_2 were uncorrelated, this would be a recursive system and so the parameters would be identified without exclusion assumptions on β_{10}. Identification conditions are given here that do not require uncorrelated errors or restrictions on β_{10}. Example applications include the mismeasured regressor model and the returns to schooling model described in the introduction.

ASSUMPTION A1: $Y = (Y_1, Y_2)'$ and X are random vectors. $E(XY')$, $E(XY_1Y')$, $E(XY_2Y')$, and $E(XX')$ are finite and identified from data. $E(XX')$ is nonsingular.

ASSUMPTION A2: $E(X\varepsilon_1) = 0$, $E(X\varepsilon_2) = 0$, and, for some random vector Z, $cov(Z, \varepsilon_1\varepsilon_2) = 0$.

The elements of Z can be discrete or continuous, and Z can be a vector or a scalar. Some or all of the elements of Z can also be elements of X. As described in the introduction, $cov(Z, \varepsilon_1\varepsilon_2) = 0$ holds if $\varepsilon_j = U\alpha_j + V_j$ where Z is uncorrelated with U^2.

Define matrices Ψ_{ZX} and Ψ_{ZZ} by

$$\Psi_{ZX} = E\left[\begin{pmatrix} X \\ [Z - E(Z)]\varepsilon_2 \end{pmatrix}\begin{pmatrix} X \\ Y_2 \end{pmatrix}'\right], \quad \Psi_{ZZ} = E\left[\begin{pmatrix} X \\ [Z - E(Z)]\varepsilon_2 \end{pmatrix}\begin{pmatrix} X \\ [Z - E(Z)]\varepsilon_2 \end{pmatrix}'\right]$$

and let Ψ be any positive definite matrix that has the same dimensions as Ψ_{ZZ}.

THEOREM 1. Let Assumptions A1 and A2 hold for the model of equations (4) and (5). Assume $cov(Z, \varepsilon_2^2) \neq 0$. Then the structural parameters β_{10}, β_{20}, γ_{10}, and the errors ε are identified, and

$$\beta_{20} = E(XX')^{-1}E(XY_2)$$

$$\begin{pmatrix} \beta_{10} \\ \gamma_{10} \end{pmatrix} = \left(\Psi'_{ZX}\Psi\Psi_{ZX}\right)^{-1}\Psi'_{ZX}\Psi E\left[\begin{pmatrix} X \\ [Z - E(Z)]\varepsilon_2 \end{pmatrix} Y_1\right] \tag{6}$$

Proofs are in the Appendix. For $\Psi = \Psi_{ZZ}^{-1}$, Theorem 1 says that the structural parameters β_{10} and γ_{10} are identified by the ordinary linear two stage least squares regression of Y_1 on X and Y_2 using X and $[Z - E(Z)]\varepsilon_2$ as instruments. Other choices of Ψ may be used if Ψ_{ZZ} is singular, or may be preferred for efficiency to account for error heteroskedasticity. Efficient GMM estimation of this model is discussed later.

The requirement that $cov(Z, \varepsilon_2^2)$ be nonzero can be empirically tested, because this covariance can be estimated as the sample covariance between Z and the squared residuals from linearly regressing Y_2 on X. Specifically, we may apply a Breusch and Pagan (1979) test for this form of heteroskedasticity to equation (5).

Klein and Vella (2003) identify the structural parameters in equations (4) and (5) by assuming a specific multiplicative heteroskedastic model for ε_1 and ε_2. Their model has the feature that the correlation coefficient between ε_1 and ε_2 is independent of X, and so in that sense is similar to Theorem 1.

2.2 Fully Simultaneous Linear Model Identification

Now consider the fully simultaneous model

$$Y_1 = X'\beta_{10} + Y_2\gamma_{10} + \varepsilon_1 \tag{7}$$

$$Y_2 = X'\beta_{20} + Y_1\gamma_{20} + \varepsilon_2 \tag{8}$$

where the errors $\varepsilon_1$and ε_2 may be correlated, and again no equality constraints are imposed on the structural parameters β_{10}, β_{20}, γ_{10}, and γ_{20}.

In some applications it is standard or convenient to normalize the second equation so that, like the first equation, the coefficient of Y_1 is set equal to one and the coefficient of Y_2 is to be estimated. An example is supply and demand, with Y_1 being quantity and Y_2 price. The identification results derived here immediately extend to handle that case, because identification of γ_{20} implies identification of $1/\gamma_{20}$ and vice versa when $\gamma_{20} \neq 0$, which is the only case in which one could normalize the coefficient of Y_1 to equal one in the second equation.

Some assumptions in addition to A1 and A2 are required to identify this fully simultaneous model. Given Assumption A2, reduced form errors W_j are

$$W_j = Y_j - X'E(XX')^{-1}E(XY_j) \tag{9}$$

ASSUMPTION A3: Define W_j by equation (9) for $j = 1, 2$. The matrix Φ_W, defined as the matrix with columns given by the vectors $cov(Z, W_1^2)$ and $cov(Z, W_2^2)$, has rank two.

Assumption A3 requires Z to contain at least two elements. If $E(\varepsilon_1\varepsilon_2 \mid \widetilde{Z}) = E(\varepsilon_1\varepsilon_2)$ for some scalar $\widetilde{Z}$, as would arise if the common unobservable U is independent of $\widetilde{Z}$, then Assumptions A2 and A3 might be satisfied by letting Z be a vector of different functions of $\widetilde{Z}$, for example defining Z as the vector of elements $\widetilde{Z}$ and $\widetilde{Z}^2$.

Assumption A3 is testable, because one may estimate W_j as the residuals from linearly regressing Y_j on X, and then use Z and the estimated W_j to estimate $cov(Z, W_j^2)$. A Breusch and Pagan (1979) test may be applied to each of these reduced form regressions. An estimated matrix rank test like Cragg and Donald (1996) could be applied to the resulting estimated matrix Φ_W, or perhaps more simply test if the determinant of $\Phi_W'\Phi_W$ is zero, since rank two requires that $\Phi_W'\Phi_W$ be nonsingular.

ASSUMPTION A4: Let Γ be the set of possible values of $(\gamma_{10}, \gamma_{20})$. If $(\gamma_1, \gamma_2) \in \Gamma$, then $(\gamma_2^{-1}, \gamma_1^{-1}) \notin \Gamma$.

Given any nonzero values of $(\gamma_{10}, \gamma_{20})$, solving equation (7) for Y_2 and equation (8) for Y_1 yields another representation of the exact same system of equations, but having coefficients $(\gamma_{20}^{-1}, \gamma_{10}^{-1})$ instead of $(\gamma_{10}, \gamma_{20})$. As long as $(\gamma_{10}, \gamma_{20}) \neq (1, 1)$ and no restrictions are placed on β_{10}, β_{20}, Assumption A4 simply says that we have chosen (either by arbitrary convenience or external knowledge) one of these two equivalent representations of the system. Assumption A4 is not needed for models that break this symmetry either by being triangular as in Theorem 1, or through an exclusion assumption as in Corollary 2 below. In other models the choice of Γ may be determined by context, e.g., typical structural VAR models constrain γ_1 and γ_2 to be less than one in absolute value, which defines a set Γ that satisfies Assumption A4. In a supply and demand model Γ may be defined by downward sloping demand and upward sloping supply curves, since in that case Γ only includes elements γ_1, γ_2 where $\gamma_1 \geq 0$ and $\gamma_2 \leq 0$, and any values that violate Assumption A4 would have the wrong signs. This is related to Fisher (1976), who showed that sign constraints in simultaneous systems yield regions of admissible parameter values.

THEOREM 2. Let Assumptions A1, A2, A3, and A4 hold in the model of equations (7) and (8). Then the structural parameters β_{10}, β_{20}, γ_{10}, γ_{20}, and the errors ε are identified.

2.3 Additional Simultaneous Model Results

LEMMA 1: Define Φ_ε to be the matrix with columns given by the vectors $cov(Z, \varepsilon_1^2)$ and $cov(Z, \varepsilon_2^2)$. Let Assumptions A1 and A2 hold, and assume $|\gamma_{10}\gamma_{20}| \neq 1$. Then Assumption A3 holds if and only if Φ_ε has rank two.

Lemma 1 assumes $\gamma_{10}\gamma_{20} \neq 1$ and $\gamma_{10}\gamma_{20} \neq -1$. The case $\gamma_{10}\gamma_{20} = 1$ is ruled out by Assumption A4 in Theorem 2. This case cannot happen in the returns to schooling or measurement error applications because triangular systems have $\gamma_{20} = 0$. It also cannot occur in the supply and demand application, because $\gamma_{10}\gamma_{20} \leq 0$ in that case. As shown in the proof of Theorem 2, the case of $\gamma_{10}\gamma_{20} = -1$ is ruled out by Assumption A3, because it causes Φ_W to have rank less than two. However, Theorem 1 can be relaxed to allow $\gamma_{10}\gamma_{20} = -1$, by replacing Assumption A3 with the assumption that Φ_ε has rank two, because then equation (23) in the proof still holds and identifies γ_{10}/γ_{20}, which along with $\gamma_{10}\gamma_{20} = -1$ and some sign restrictions could identify γ_{10} and γ_{20} in this case. However, Assumption A3 has the advantage of being empirically testable.

In either case, Theorem 2 requires both ε_1 and ε_2 to be heteroskedastic with variances that depend upon Z, since otherwise the vectors $cov(Z, \varepsilon_1^2)$ and $cov(Z, \varepsilon_2^2)$ will equal zero. Moreover, the variances of ε_1 and ε_2 must be different functions of Z for the rank of Φ_ε to be two.

COROLLARY 1. Let Assumptions A1, A2, A3, and A4 hold in the model of equations (7) and (8), replacing $cov(Z, \varepsilon_1\varepsilon_2)$ in Assumption A2 with $E(Z\varepsilon_1\varepsilon_2)$ and replacing $cov(Z, W_j^2)$ with $E(ZW_j^2)$ in Assumption A3, for $j = 1, 2$. Then the structural parameters β_{10}, β_{20}, γ_{10}, γ_{20}, and the errors ε are identified.

Corollary 1 can be used in applications where $E(\varepsilon_1\varepsilon_2) = 0$, as in the structural VAR model. Theorem 2 could also be used in this case, but Corollary 1 provides additional moments. In particular, if only a scalar $\widetilde{Z}$ is known to satisfy $cov(\widetilde{Z}, \varepsilon_1\varepsilon_2) = 0$, then identification by Theorem 2 will fail because the rank condition in Assumption A3 is violated with $Z = \widetilde{Z}$, but identification may still be possible using Corollary 1 because there we may let $Z = (1, \widetilde{Z})$.

COROLLARY 2. Let Assumptions A1 and A2 hold for the model of equations (7) and (8). Assume $cov(Z, \varepsilon_2^2) \neq 0$, that some element of β_{20} is known to equal zero and the corresponding element of β_{10} is nonzero. Then the structural parameters β_{10}, β_{20}, γ_{10}, γ_{20}, and the errors ε are identified.

Corollary 2 is like Theorem 1, except that it assumes an element of β_{20} is zero instead of assuming γ_{20} is zero as its exclusion restrction to identify equation (8). Then, as in Theorem 1, Corollary 2 uses $cov(Z, \varepsilon_1\varepsilon_2) = 0$ to identify equation (8) without imposing the rank two condition of Assumption A3 and the inequality constraints of Assumption A4. Only a scalar Z is needed for identification using Theorem 1 or Corollaries 1 or 2.

3 Estimation

3.1 Simultaneous System Estimation

Consider estimation of the structural model of equations (7) and (8) based on Theorem 2. Define S to be the vector of elements of Y, X, and the elements of Z that are not already contained in X, if any.

Let $\mu = E(Z)$ and let θ denote the set of parameters $\{\gamma_1, \gamma_2, \beta_1, \beta_2, \mu\}$. Define the vector valued functions

$$Q_1(\theta, S) = X(Y_1 - X'\beta_1 - Y_2\gamma_1)$$

$$Q_2(\theta, S) = X(Y_2 - X'\beta_2 - Y_1\gamma_2)$$

$$Q_3(\theta, S) = Z - \mu$$

$$Q_4(\theta, S) = (Z - \mu)\,(Y_1 - X'\beta_1 - Y_2\gamma_1)(Y_2 - X'\beta_2 - Y_1\gamma_2)$$

Define $Q(\theta, S)$ to be the vector obtained by stacking the above four vectors into one long vector.

COROLLARY 3: Assume equations (7) and (8) hold. Define θ, S, and $Q(\theta, S)$ as above. Let Assumptions A1, A2, A3, and A4 hold. Let Θ be the set of all values θ might take on, and let θ_0 denote the true value of θ. Then the only value of $\theta \in \Theta$ that satisfies $E[Q(\theta, S)] = 0$ is $\theta = \theta_0$.

A simple variant of Corollary 3 is that if $E(\varepsilon_1\varepsilon_2) = 0$, as in the structural VAR model, then μ can be dropped from θ, with Q_3 dropped from Q, and the $Z - \mu$ term in Q_4 replaced with just Z.

Given Corollary 3, GMM estimation of the model of equations (7) and (8) is completely straightforward. With a sample of n observations $S_1,...,S_n$, the standard GMM estimator is

$$\widehat{\theta} = \arg\min_{\theta\in\Theta} \sum_{i=1}^{n} Q(\theta, S_i)'\Omega_n \sum_{i=1}^{n} Q(\theta, S_i) \tag{10}$$

for some sequence of positive definite Ω_n. If the observations S_i are independently and identically distributed and if Ω_n is a consistent estimator of $\Omega_0 = E\left[Q(\theta_0, S)Q(\theta_0, S)'\right]$, then efficient GMM has

$$\sqrt{n}(\widehat{\theta} - \theta_0) \rightarrow^d N\left(0, E\left(\frac{\partial Q(\theta_0, S)}{\partial\theta'}\right)\Omega_0^{-1} E\left(\frac{\partial Q(\theta_0, S)}{\partial\theta'}\right)'\right) \tag{11}$$

More generally, with dependent data, standard time series versions of GMM would be directly applicable. Alternative moment based estimators with possibly better small sample properties, such as Generalized Empirical Likelihood, could be used instead of GMM. See, e.g., Newey and Smith (2004). Also, if these moment conditions are weak, then alternative limiting distribution theory based on weak instruments, such as Staiger and Stock (1997), would be immediately applicable. See Stock, Wright, and Yogo (2002) for a survey of such estimators.

Standard GMM assumes Θ is compact. When $\gamma_{20} \neq 0$ this must be reconciled with Assumption A4 and with Lemma 1. For example, in the supply and demand model we might define Θ so that the product of the first two elements of every $\theta \in \Theta$ is finite, nonpositive, and excludes an open neighborhood of minus one. This last constraint could be relaxed as discussed after Lemma 1.

If one wished to normalize the second equation so that the coefficient of Y_1 equaled one, as might be more natural in a supply and demand system, then the same GMM estimator could be used just by replacing $Y_2 - X'\beta_2 - Y_1\gamma_2$ in the Q_2 and Q_4 functions with $Y_1 - X'\beta_2 - Y_2\gamma_2$, redefining β_2 and γ_2 accordingly.

Based on the proof of Theorem 2, a numerically simpler but less efficient estimator would be the following. First, let $\widehat{W}_j$ be the vector of residuals from linearly regressing Y_j on X_j. Next, let $\widehat{C}_{jkh}$ be the sample covariance of $\widehat{W}_j\widehat{W}_k$ with Z_h, where Z_h is the $h'th$ element of the vector Z. Assume Z has a total of H elements. Based on equation (23), estimate γ_1 and γ_2 by

$$(\widehat{\gamma}_1, \widehat{\gamma}_2) = \arg\min_{(\gamma_1\gamma_2)\in\Gamma} \sum_{h=1}^{H} \left((1+\gamma_1\gamma_2)\widehat{C}_{12h} - \gamma_1\widehat{C}_{22h} - \gamma_2\widehat{C}_{11h}\right)^2$$

where Γ is a compact set satisfying Assumption A4. The above estimator for γ_1 and γ_2 is numerically equivalent to an ordinary nonlinear least squares regression over a small number of observations of data, that is, H observations. Finally, β_1 and β_2 may be estimated by linearly regressing $Y_1 - Y_2\widehat{\gamma}_1$ and $Y_2 - Y_1\widehat{\gamma}_2$ on X, respectively. The consistency of this procedure follows from the consistency of each step, which in turn is based on the steps of the identification proof of Theorem 1 and the consistency of linear regressions and sample covariances.

In practice, this simple procedure might be useful for generating consistent starting values for efficient GMM.

3.2 Triangular System Estimation

The GMM estimator used for the fully simultaneous system can be applied to the triangular system of Theorem 1 by setting $\gamma_2 = 0$. Define S and μ as before, and now let $\theta = \{\gamma_1, \beta_1, \beta_2, \mu\}$ and

$$Q_1(\theta, S) = X(Y_1 - X'\beta_1 - Y_2\gamma_1)$$
$$Q_2(\theta, S) = X(Y_2 - X'\beta_2)$$
$$Q_3(\theta, S) = Z - \mu$$
$$Q_4(\theta, S) = (Z - \mu)(Y_1 - X'\beta_1 - Y_2\gamma_1)(Y_2 - X'\beta_2).$$

Let $Q(\theta, S)$ be the vector obtained by stacking the above four vectors into one long vector, and we immediately obtain

COROLLARY 4: Assume equations (4) and (5) hold. Define θ, S, and $Q(\theta, S)$ as above. Let Assumptions A1 and A2 hold with $cov(Z, W_2^2) \neq 0$. Let Θ be the set of all values θ might take on, and let θ_0 denote the true value of θ. Then the only value of $\theta \in \Theta$ that satisfies $E[Q(\theta, S)] = 0$ is $\theta = \theta_0$.

The GMM estimator (10) and limiting distribution (11) then follow immediately.

Based on Theorem 1, the triangular system of equations (4) and (5) can be more easily estimated as follows. When $\gamma_{20} = 0$, β_{20} can be estimated by linearly regressing Y_2 on X. Then, letting $\widehat{\varepsilon}_{2i}$ be the residuals from this regression, β_{10} and γ_{10} can be estimated by an ordinary linear two stage least squares regression of Y_1 on Y_2 and X, using X and $(Z - \overline{Z})\widehat{\varepsilon}_2$ as instruments, where $\overline{Z}$ is the sample mean of Z. Letting overbars denote sample averages, the resulting estimators are

$$\widehat{\beta}_2 = \overline{XX'}^{-1}\overline{XY_2}, \quad \widehat{\varepsilon}_2 = Y_2 - X'\widehat{\beta}_2$$

$$\begin{pmatrix} \widehat{\beta}_1 \\ \\ \widehat{\gamma}_1 \end{pmatrix} = \left(\widehat{\Psi}'_{ZX}\widehat{\Psi}^{-1}_{ZZ}\widehat{\Psi}_{ZX}\right)^{-1}\widehat{\Psi}'_{ZX}\widehat{\Psi}^{-1}_{ZZ}\begin{pmatrix} \overline{XY_1} \\ \overline{\qquad\qquad\qquad} \\ (Z-\overline{Z})\widehat{\varepsilon}_2 Y_1 \end{pmatrix} \tag{12}$$

where $\widehat{\Psi}_{ZX}$ replaces the expectation defining Ψ_{ZX} with a sample average, and similarly for $\widehat{\Psi}$, in particular, for ordinary two stage least squares $\widehat{\Psi}$ would be a consistent estimator of $\widehat{\Psi}^{-1}_{ZZ}$. The limiting distribution for $\widehat{\beta}_2$ is standard ordinary least squares. The distribution for $\widehat{\beta}_1$ and $\widehat{\gamma}_1$ is basically that of ordinary two stage least squares, except account must be taken of the estimation error in the instuments $(Z-\overline{Z})\widehat{\varepsilon}_2$. Using the standard theory of two step estimators (see, e.g., Newey and McFadden 1994). With independent, identically distributed observations this gives

$$\sqrt{n}\left[\begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\gamma}_1 \end{pmatrix} - \begin{pmatrix} \beta_{10} \\ \gamma_{10} \end{pmatrix}\right] \rightarrow^d N\left(0, \left(\Psi'_{ZX}\Psi\Psi_{ZX}\right)^{-1}\Psi'_{ZX}\Psi var\begin{pmatrix} X\varepsilon_1 \\ R \end{pmatrix}\Psi\Psi_{ZX}\left(\Psi'_{ZX}\Psi\Psi_{ZX}\right)^{-1}\right)$$

where

$$R = [Z - E(Z)]\varepsilon_2\varepsilon_1 - cov(Z, X')E(XX')^{-1}X\varepsilon_2$$

is the influence function associated with $\overline{(Z-\overline{Z})\widehat{\varepsilon}_2\varepsilon_1}$.

While numerically simpler, since no numerical searching is required, this two stage least square estimator could be less efficient than GMM. It will be numerically identical to GMM when the parameters are exactly identified rather than overidentified, that is, when Z is a scalar. More generally this two stage least squares estimator could be used for generating consistent starting values for efficient GMM estimation.

4 Monte Carlo

Monte Carlo simulations draw data from the reduced form of the structural model

$$Y_1 = \beta_{11} + X\beta_{12} + Y_2\gamma_1 + \varepsilon_1, \qquad \varepsilon_1 = U + e^X S_1 \tag{13}$$

$$Y_2 = \beta_{21} + X\beta_{22} + Y_1\gamma_2 + \varepsilon_2, \qquad \varepsilon_2 = U + e^{-X} S_2 \tag{14}$$

where X, U, S_1, and S_2 are independent standard normal scalars and $\beta_{11} = \beta_{12} = \beta_{21} = \beta_{22} = \gamma_1 = 1$. The triangular design sets $\gamma_2 = 0$ and $Z = X$. The fully simultaneous design sets $\gamma_2 = -.5$ and $Z = (X, X^2)$. With these choices of Z the model parameters in each design are exactly identified by Theorems 1 and 2. The parameters in equation (13) for the triangular design, and all of the parameters in both equations in the simultaneous design, are not identified using traditional exclusion assumptions. Table 1 reports results of 10, 000 simulations of each design, with sample size $n = 500$.

The triangular design is estimated using the two stage least squares estimator in equation (12), which is numerically identical to GMM because the model is exactly identified. The simultaneous system design is estimated using GMM as in equation (10). Ignoring Assumption A4, no inequality constraints on the parameters were imposed on the estimates, though for GMM the true values of the parameters were used as starting values for the optimizing iterations in each simulation.

The triangular model estimates are quite accurate, with less than one percent mean bias and root mean squared errors under .275. The simultaneous system parameters have biases of a few percent, but much larger root mean squared errors. These are largely due to a very small number of extreme estimates, as can be seen by median absolute errors that are only modestly larger than in the triangular model case, and virtually the same interquartile ranges.

5 Engel Curve Estimates

An Engel curve for food is empirically estimated, where total expenditures may be mismeasured. The data consist of the same set of demographically homogeneous households that were used to analyze Engel curves in Banks, Blundell and Lewbel (1997). These are all households in the United Kingdom Family Expenditure Survey 1980-1982 composed of two married adults without children, living in the Southeast (including London). The dependent variable Y_1 is the food budget share, the possibly mismeasured regressor Y_2 is log real total expenditures, and the other regressors X are a constant, age, spouses age, squared ages, seasonal dummies, and dummies for spouse working, gas central heating, ownership of a washing machine, one car, and two cars. Sample means are $\overline{Y}_1 = .285$ and $\overline{Y}_2 = .599$. There are 854 observations.

The model is $Y_1 = X'\beta_1 + Y_2\gamma_1 + \varepsilon_1$. This is the Working (1943) - Leser (1963) functional form for Engel curves. Nonparametric and parametric regression analyses on this data show that this functional form fits food (though not other) budget shares quite well. See, e.g., Banks, Blundell and Lewbel (1997), figure 1A.

Total expenditures are subject to potentially large measurement errors, due in part to infrequently purchased items. See, e.g., Meghir and Robin (1992). To keep the analysis simple, possible mismeasurement of the food budget share arising from mismeasurement of total expenditures, as in Lewbel (1996), is ignored, though it might be handled by modeling levels of food expenditures instead of shares, and applying nonlinear extensions discussed later. Hausman, Newey, and Powell (1995) is a prominent example of Engel curve estimation assuming that budget shares are not mismeasured and log total expenditures suffer classical measurement error (though with the complication of a polynomial functional form).

Table 2 summarizes the empirical results. Ordinary least squares, which does not account for mismeasurement, has an estimated log total expenditure coefficient of $\widehat{\gamma}_1 = -.127$. Ordinary two stage least squares, using log total income as an outside instrument, substantially reduces the estimated coefficient to $\widehat{\gamma}_1 = -.086$. This is model TSLS 1 or equivalently GMM 1 in Table 2. TSLS1 and GMM 1 are exactly identified, and so are numerically equivalent. Log total expenditures are correlated with other regressors such as age and durables ownership indicators, which in this application resulted in a reversal of the usual attenuation direction of measurement error bias.

If we did not observe income for use as an outside instrument, we could instead apply the GMM estimator based on Corollary 4, using the moments $cov(Z, \varepsilon_1\varepsilon_2) = 0$. As discussed in the introduction, with classical measurement error we may let Z equal all the elements of X except the constant. The result is model GMM 2 in Table 2, which yields $\widehat{\gamma}_1 = -.078$. This is relatively close to the estimate based on the outside instrument, as would be expected if the outside instrument is valid and if this paper's methodology for identification and estimation without external instruments is also valid. However, the standard errors in GMM 2 are a good bit higher than those of GMM 1, suggesting that not having an outside instrument hurts efficiency.

The estimates based on Corollary 4 are overidentified, so the GMM 2 estimates differ numerically from the two stage least squares version of this estimator, reported as TSLS 2, which uses $(Z - \overline{Z})\widehat{\varepsilon}_2$ as instruments as in equation (12). The GMM 2 estimates are closer to the outside instrument estimates GMM 1, and have smaller standard errors, which shows that the increased aymptotic efficiency of GMM is valuable here. A Hansen (1982) test fails to reject the overidentifying moments in this model at the 5% level, though the p-value of 6.5% is close to rejecting.

Table 2 also reports estimates obtained combining both moments based on both the outside instrument, log income, and on $cov(Z, \varepsilon_1\varepsilon_2) = 0$. The results, in TSLS 3 and GMM 3, are very similar to TSLS 1 and GMM 1, which just use the outside instrument. This is consistent with validity of both sets of identifying moments, but with the outside instrument being much stronger or more informative. The Hansen test also fails to reject this joint set of overidentifying moments, with a p-value of 12.5%.

One may question the assumptions for applying Theorem 1 in this application. Similarly, the validity of income as an outside instrument for total expenditures is questionable (e.g., they could both have common sources of measurement errors). However, the results that are obtained here are consistent with these assumptions holding. In particular, this paper's methodology for obtaining estimates without outside instruments yields estimates that are close to, but not as statistically significant as, estimates that are obtained by using an outside instrument, and the resulting overidentifying moments are not statistically rejected.

6 Extensions

This section considers extending the model to allow for nonlinear functions of X. Details regarding regularity conditions and limiting distributions for associated estimators are not provided, because they are immediate applications of existing estimators once the required identifying moments are established.

6.1 Semiparametric Identification

Consider the model

$$Y_1 = g_1(X) + Y_2\gamma_{10} + \varepsilon_1 \tag{15}$$

$$Y_2 = g_2(X) + Y_1\gamma_{20} + \varepsilon_2 \tag{16}$$

where the functions $g_j(X)$ are unknown. In this simultaneous system, each equation is partly linear as in Robinson (1988).

ASSUMPTION B1: $Y = (Y_1, Y_2)'$, where Y_1 and Y_2 are random variables. For some random vector X, the functions $E(Y \mid X)$ and $E(YY' \mid X)$ are finite and identified from data.

Given a sample of observations of Y and X, the conditional expectations in Assumption B1 could be estimated by nonparametric regressions, and so would be identified. These conditional expectations are the reduced form of the underlying structural model.

ASSUMPTION B2: $E(\varepsilon_1 \mid X) = 0$, $E(\varepsilon_2 \mid X) = 0$, and for some random vector Z, $cov(Z, \varepsilon_1 \varepsilon_2) = 0$.

As before, the elements of Z can all be elements of X also, so no outside instruments are required. No exclusion assumptions are imposed, so all of the same regressors X that appear in g_1 can also appear in g_2, and vice versa. If $\varepsilon_j = U\alpha_j + V_j$, where U, V_1, and V_2 are mutually uncorrelated (conditioning on Z), $cov(Z, \varepsilon_1 \varepsilon_2) = 0$ if Z is uncorrelated with U^2.

ASSUMPTION B3: Define $W_j = Y_j - E(Y_j \mid X)$ for $j = 1, 2$. The matrix Φ_W, defined as the matrix with columns given by the vectors $cov(Z, W_1^2)$ and $cov(Z, W_2^2)$, has rank two.

Assumption B3 is analogous to Assumption A3, but employs a different definition of W_j. These definitions will coincide if the conditional expectation of Y given X is linear in X. Lemma 1 continues to hold with this new definition of W_j and hence of Φ_W, and more generally heteroskedasticity of W_1 and W_2 implies heteroskedasticity of ε.

THEOREM 3: Let equations (15) and (16) hold. If Assumptions B1 and B2 hold, $cov(Z, \varepsilon_2^2) \neq 0$, and $\gamma_{20} = 0$ then the structural parameter γ_{10}, the functions $g_1(X)$ and $g_2(X)$, and the variance of ε are identified. If Assumptions B1, B2, B3, and A4 hold then the structural parameters γ_{10} and γ_{20}, the functions $g_1(X)$ and $g_2(X)$, and the variance of ε are identified.

An immediate corollary of Theorem 3 is that the partly linear simultaneous system

$$Y_1 = h_1(X_1) + X_2\beta_{10} + Y_2\gamma_{10} + \varepsilon_1 \tag{17}$$

$$Y_2 = h_2(X_1) + X_2\beta_{20} + Y_1\gamma_{20} + \varepsilon_2 \tag{18}$$

where $X = (X_1, X_2)$ will also be identified, since $g_j(X) = h_j(X_1) + X_2\beta_{j0}$ is identified.

6.2 Nonlinear Model Estimation

Consider the model

$$Y_1 = G_1(X, \beta_0) + Y_2\gamma_{10} + \varepsilon_1 \tag{19}$$

$$Y_2 = G_2(X, \beta_0) + Y_1\gamma_{20} + \varepsilon_2 \tag{20}$$

where the functions $G_j(X, \beta_0)$ are known and the parameter vector β_0, which could include γ_1 and γ_2, is unknown. This generalizes equations (1) and (2) by allowing nonlinear functions of X. Letting $g_j(X) = G_j(X, \beta_0)$, Theorem 3 provides sufficient conditions for identification of this model, assuming that β_0 is identified given identifcation of the functions $g_j(X) = G_j(X, \beta_0)$. The immediate analog to Corollary 3 is then that β_0, γ_{10}, γ_{20}, and μ_0 can be estimated from the moment conditions

$$E[(Y_1 - G_1(X, \beta_0) - Y_2\gamma_{10}) \mid X] = 0$$

$$E[(Y_2 - G_2(X, \beta_0) - Y_1\gamma_{20}) \mid X] = 0$$

$$E(Z - \mu_0) = 0$$

$$E[\left(Z - \mu_0\right)[Y_1 - G_1(X, \beta_0) - Y_2\gamma_{10}][Y_2 - G_2(X, \beta_0) - Y_1\gamma_{20}] = 0$$

For efficient estimation in this case where some of the moments are conditional see, e.g., Chamberlain (1987), Newey (1993), and Kitamura, Tripathi, and Ahn (2003). Ordinary GMM can be used for estimation by replacing the first two conditional moments above with unconditional moments

$$E[\zeta(X)(Y_1 - G_1(X, \beta_0) - Y_2\gamma_{10})] = 0$$

$$E[\zeta(X)(Y_2 - G_2(X, \beta_0) - Y_1\gamma_{20})] = 0$$

For some chosen vector valued function $\zeta(X)$. Asymptotic efficiency may be obtained by using an estimated optimal $\zeta(X)$; see, e.g., Newey (1993) for details.

As in the linear model, some of these moments may be weak, which would suggest the use of weak instrument limiting distributions in the GMM estimation. See Stock, Wright, and Yogo (2002) for a survey of applicable weak moment procedures.

6.3 Semiparametric Estimation

Consider estimation of the partly linear system of equations (15) and (16), where the functions $g_j(X)$ are not parameterized. We now have identification based on the moments

$$E[Y_1 - g_1(X) - Y_2\gamma_1 \mid X] = 0$$

$$E[Y_2 - g_2(X) - Y_2\gamma_1 \mid X] = 0$$

$$E(Z - \mu_0) = 0$$

$$E[(Z - \mu_0)\,(Y_1 - g_1(X) - Y_2\gamma_{10})(Y_2 - g_2(X_1) - Y_1\gamma_{20})] = 0$$

These are conditional moments containing unknown parameters and unknown functions, and so general estimators for these types of models may be applied. Examples include Ai and Chen (2003), Otsu (2003), and Newey and Powell (2003).

Alternatively, the following estimation procedure could be used, analogous to the numerically simple estimator for linear simultaneous models described earlier. Assume we have n independent, identically distributed observations. Let $\widehat{H}_j(X)$ be a uniformly consistent estimator of $H_j(X) = E(Y_j \mid X)$, e.g., a kernel or local polynomial nonparametric regression of Y_j on X. Now, as defined by Assumption B3, $W_j = Y_j - H_j(X)$, so let $\widehat{W}_{ji} = Y_{ji} - \widehat{H}_j(X_i)$ for each observation i. Next, let $\widehat{C}_{jkh}$ be the sample covariance of $\widehat{W}_j\widehat{W}_k$ with Z_h, where Z_h is the $h'th$ element of the vector Z. Assume Z has a total of K elements. Based on equation (23), estimate γ_1 and γ_2 by

$$(\widehat{\gamma}_1, \widehat{\gamma}_2) = \arg\min_{(\gamma_1\gamma_2)\in\Gamma} \sum_{h=1}^{K} \left((1+\gamma_1\gamma_2)\widehat{C}_{12h} - \gamma_1\widehat{C}_{22h} - \gamma_2\widehat{C}_{11h}\right)^2$$

where Γ is a compact set satisfying Assumption A4. The above estimator for γ_1 and γ_2 is numerically equivalent to an ordinary nonlinear least squares regression over a small number of observations of data, that is, K observations. Finally, estimates of the functions $g_1(X)$ and $g_2(X)$ are obtained by nonparametrically regressing $Y_1 - Y_2\widehat{\gamma}_1$ and $Y_2 - Y_1\widehat{\gamma}_2$ on X, respectively. The consistency of this procedure follows from the consistency of each step, which in turn is based on the steps of the identification proof of Theorem 3.

This estimator of $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ is an example of a semiparametric estimator with nonparametric plug-ins. See, e.g., section 8 of Newey and McFadden (1994). Unlike Ai and Chen (2003), this numerically simple procedure might not yield efficient estimates of $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$. However, assuming that $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ converge at a faster rate than nonparametric regressions, the limiting distributions of the estimates of the functions $g_1(X)$ and $g_2(X)$ will be the same as for ordinary nonparametric regressions of $Y_1 - Y_2\gamma_1$ and $Y_2 - Y_1\gamma_2$ on X, respectively.

Further extension to estimation of the partly linear system of equations (17) and (18) is immediate. For this model the Assumption B2 moments $E(\varepsilon_1 \mid X) = 0$, $E(\varepsilon_2 \mid X) = 0$, and $cov(Z, \varepsilon_1\varepsilon_2) = 0$ are

$$E[Y_1 - h_1(X_1) - X_2\beta_{10} - Y_2\gamma_1 \mid X] = 0$$

$$E[Y_2 - h_2(X_1) - X_2\beta_{20} - Y_2\gamma_1 \mid X] = 0$$

$$E(Z-\mu_0)=0$$

$$E[\left(Z-\mu_0\right)[Y_1-h_1(X_1)-X_2\beta_{10}-Y_2\gamma_{10}][Y_2-h_2(X_1)-X_2\beta_{20}-Y_1\gamma_{20}]=0$$

which could again be consistently estimated by the above described procedure, replacing the nonparametric regression steps with partly linear nonparametric regression estimators such as Robinson (1988), or by directly applying an estimator such as Ai and Chen (2003) to these moments.

7 Conclusions

This paper describes a new method of obtaining identification in mismeasured regressor models, triangular systems, simultaneous equation systems, structural vector autoregressions, and some partly linear semiparametric systems. Associated estimators are provided. The proposed estimators appear to work well in both a small Monte Carlo study and in an empirical application. The identification comes from observing a vector of variables Z that are uncorrelated with the covariance of heteroskedastic errors, which is shown to be a feature of many models in which error correlations are due to an unobserved common factor.

Unlike ordinary instruments, identification is obtained even when all the elements of Z are also regressors in every model equation. However, Z shares many of the convenient features instruments in ordinary two stage least squares models. As with ordinary instrument selection, given a set of possible choices for Z, the estimators remain consistent if only a subset of the available choices are used, so variables that one is unsure about can be safely excluded from the Z vector, with the only loss being efficiency. Similarly, as with ordinary instruments, if some variable $\widetilde{Z}$ satisfies the conditions to be an element of Z, but is only observed with classical measurement error, then this mismeasured $\widetilde{Z}$ can still be used as an element of Z. If Z has more than two elements (or more than one element in a triangular system) then the model parameters are overidentified and standard tests of overidentifying restrictions, such as Hansens (1982) test, can be applied. Similarly, moments based on $cov(Z,\varepsilon_1\varepsilon_2)=0$ can be used along with ordinary valid outside instruments to improve efficiency and provide testable overidentifying restrictions.

8 Appendix

PROOF OF THEOREM 1: Define W_j by equation (9) for $j = 1, 2$. These W_j are identified by construction. Using the Assumptions, substituting equations (4) and (5) for Y_1 and Y_2 in the definitions of W_1 and W_2 shows that $W_1 = \varepsilon_1 + \varepsilon_2 \gamma_{10}$ and $W_2 = \varepsilon_2$, so $cov(Z, \varepsilon_1 \varepsilon_2) = 0$ is equivalent to $cov[Z, (W_1 - \gamma_{10} W_2) W_2] = 0$. Solving for γ_{10} shows that γ_{10} is identified by $\gamma_{10} = cov(Z, W_1 W_2)/cov(Z, W_2^2)$. Given identification of γ_{10}, the coefficients β_{10} and β_{20} are identified by $\beta_{10} = E(XX')^{-1} E[X(Y_1 - Y_2 \gamma_{10})]$ and $\beta_{20} = E(XX')^{-1} E(XY_2)$, which follow from $E(X\varepsilon_j) = 0$. Finally ε is now identified by $\varepsilon_1 = Y_1 - X'\beta_{10} - Y_2 \gamma_{10}$ and $\varepsilon_2 = Y_2 - X'\beta_{20}$. Finally, to show equation (6), observe that Ψ_{ZX} simplifies to

$$\Psi_{ZX} = \begin{pmatrix} E(XX') & E(XX')\beta_{20} \\ E(ZX\varepsilon_2) & E(ZX\varepsilon_2)\beta_{20} + cov(Z, \varepsilon_2^2) \end{pmatrix}$$

which spans the same column space as

$$\begin{pmatrix} E(XX') & 0 \\ E(ZX\varepsilon_2) & cov(Z, \varepsilon_2^2) \end{pmatrix}$$

and so has rank equal to the number of columns, which makes $\Psi_{ZX} \Psi \Psi_{ZX}$ nonsingular. Also

$$E\left[\begin{pmatrix} X \\ [Z - E(Z)]\varepsilon_2 \end{pmatrix} Y_1\right] = \Psi_{ZX} \begin{pmatrix} \beta_{10} \\ \gamma_{10} \end{pmatrix} + \begin{pmatrix} 0 \\ cov(Z, \varepsilon_1 \varepsilon_2) \end{pmatrix}$$

which then gives equation (6).

PROOF OF THEOREM 2: Substituting equations (7) and (8) for Y_1 and Y_2 in the definitions of W_1 and W_2 shows that

$$W_1 = \frac{\varepsilon_1 + \varepsilon_2 \gamma_{10}}{1 - \gamma_{10}\gamma_{20}}, \quad W_2 = \frac{\varepsilon_2 + \varepsilon_1 \gamma_{20}}{1 - \gamma_{10}\gamma_{20}} \tag{21}$$

and solving these equations for ε yields

$$\varepsilon_1 = W_1 - \gamma_{10} W_2, \quad \varepsilon_2 = W_2 - \gamma_{20} W_1 \tag{22}$$

Note that $\gamma_{10}\gamma_{20} \neq 1$ by Assumption A4. Using equation (22), the condition $cov(Z, \varepsilon_1 \varepsilon_2) = 0$ is equivalent to

$$cov[Z, (W_1 - \gamma_{10} W_2)(W_2 - \gamma_{20} W_1)] = 0$$

$$(1+\gamma_{10}\gamma_{20})cov(Z,W_1W_2)-\gamma_{10}cov(Z,W_2^2)-\gamma_{20}cov(Z,W_1^2)=0 \quad (23)$$

Now $1+\gamma_{10}\gamma_{20}\neq 0$, since otherwise it would follow from equation (23) that the rank of Φ_W is less than two. Define

$$\lambda_1=\frac{\gamma_{10}}{1+\gamma_{10}\gamma_{20}},\quad \lambda_2=\frac{\gamma_{20}}{1+\gamma_{10}\gamma_{20}} \quad (24)$$

and $\lambda=(\lambda_1,\lambda_2)'$ We then have

$$cov(Z,W_1W_2)=\lambda_1 cov(Z,W_2^2)+\lambda_2 cov(Z,W_1^2)=\Phi_W\lambda \quad (25)$$

so λ is identified by

$$\lambda=(\Phi_W'\Phi_W)^{-1}\Phi_W' cov(Z,W_1W_2)$$

and $\Phi_W'\Phi_W$ is not singular because Φ_W is rank two. Solving equation (24) for γ_{10} gives

$$0=\lambda_2\gamma_{10}^2-\gamma_{10}+\lambda_1$$

The above quadratic in γ_{10} has at most two roots, and for each root the corresponding value for γ_{20} is given by $\gamma_{20}=\gamma_{10}\lambda_2/\lambda_1$. Let (γ_1^*,γ_2^*) denote one of these solutions. It can be seen from

$$\lambda_1=\left(\frac{1}{\gamma_{10}}+\gamma_{20}\right)^{-1},\quad \lambda_2=\left(\frac{1}{\gamma_{20}}+\gamma_{10}\right)^{-1}$$

that the other solution must be $(\gamma_2^{*-1},\gamma_1^{*-1})$, since that yields the same values for λ_1 and λ_2. One of these solutions must be $(\gamma_{10},\gamma_{20})$, and by Assumption A4 the other solution is not an element of Γ, so $(\gamma_{10},\gamma_{20})$ is identified.

Given identification of γ_{10} and γ_{20}, the coefficients β_{10} and β_{20} are identified by $\beta_{10}=E(XX')^{-1}E[X(Y_1-Y_2\gamma_{10})]$ and $\beta_{20}=E(XX')^{-1}E[X(Y_2-Y_1\gamma_{20})]$, which follow from $E(X\varepsilon_j)=0$. Finally ε is now identified by $\varepsilon_1=Y_1-X'\beta_{10}-Y_2\gamma_{10}$ and $\varepsilon_2=Y_2-X'\beta_{20}-Y_1\gamma_{20}$.

PROOF OF LEMMA 1: Equation (21) in Theorem 2 was derived using only Assumptions A1 and A2. Evaluating $cov(Z,W_j^2)$ using equation (21) and the assumption that $cov(Z,\varepsilon_1\varepsilon_2)=0$ gives, for each element Z_k of Z,

$$\begin{pmatrix} cov(Z_k,W_1^2)\\ cov(Z_k,W_2^2)\end{pmatrix}=\left(\frac{1}{1-\gamma_{10}\gamma_{20}}\right)^2\begin{bmatrix}1 & \gamma_{10}^2\\ \gamma_{20}^2 & 1\end{bmatrix}\begin{pmatrix} cov(Z_k,\varepsilon_1^2)\\ cov(Z_k,\varepsilon_2^2)\end{pmatrix} \quad (26)$$

So Φ_W is rank two if and only if Φ_ε is rank two and the matrix relating the two above is nonsingular, which requires $|\gamma_{10}\gamma_{20}| \neq 1$.

PROOF OF COROLLARY 1: Using equation (22) and following the same steps as the proof of Theorem 2, the condition $E(Z\varepsilon_1\varepsilon_2) = 0$ yields

$$E(ZW_1W_2) = \lambda_1 E(ZW_2^2) + \lambda_2 E(ZW_1^2) = \Phi_W \lambda$$

instead of equation (25). This identifies λ and the rest of the proof is the same.

PROOF OF COROLLARY 2: β_{20} and γ_{20}, and hence ε_2, are identified from the usual moments that permit two stage last squares estimation. Each W_j is identified as in Theorem 1, and by equation (21), $cov(Z, \varepsilon_1\varepsilon_2) = 0$ implies $cov[Z, (W_1 - \gamma_{10}W_2)\varepsilon_2 = 0$, which solving for γ_{10} gives

$$\gamma_{10} = cov(Z, W_1\varepsilon_2)/cov(Z, W_2\varepsilon_2)$$

and $cov(Z, W_2\varepsilon_2) = cov(Z, \varepsilon_2^2) \neq 0$, so γ_{10} is identified. The rest of the proof is the same as the end of the proof of Theorem 2.

PROOF OF THEOREM 3: Like Theorem 1, substituting equations (15) and (16) for Y_1 and Y_2 in the Assumption B3 definitions of W_1 and W_2 shows that equations (21) and (22) hold in this model. Identification of γ_{10} and γ_{20} then follows exactly as in the Proof of Theorem 1. Given identification of γ_{10} and γ_{20}, the functions $g_1(X)$ and $g_2(X)$ are identified by $g_1(X) = E(Y_1 \mid X) - E(Y_2 \mid X)\gamma_{10}$ and $g_2(X) = E(Y_2 \mid X) - E(Y_1 \mid X)\gamma_{20}$, both of which follow from $E(\varepsilon_j \mid X) = 0$. Finally ε is now identified by $\varepsilon_1 = Y_1 - g_1(X) - Y_2\gamma_{10}$ and $\varepsilon_2 = Y_2 - g_2(X) - Y_1\gamma_{20}$.

PROOF OF COROLLARIES 3 and 4: By equations (7) and (8), $Q_1 = X\varepsilon_1$, $Q_2 = X\varepsilon_1$ and $Q_4 = (Z - \mu)\varepsilon_1\varepsilon_2$, and $E(Q_3) = 0$ makes $\mu = E(Z)$, so $E(Q) = 0$ is equivalent to $E(X\varepsilon_1) = 0$, $E(X\varepsilon_2) = 0$, and $cov(Z, \varepsilon_1\varepsilon_2) = 0$. It then follows from Theorem 2, or from Theorem 1 when $\gamma_{20} = 0$, that the only $\theta \in \Theta$ that satisfies $E[Q(\theta, S)] = 0$ is $\theta = \theta_0$.

References

[1] AI, C. AND X. CHEN (2003), "Efficient Estimation of Models With Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795-1844.

[2] BANKS, J., R. BLUNDELL, AND A. LEWBEL, (1997), "Quadratic Engel Curves and Consumer Demand," with James Banks and Richard Blundell, *Review of Economics and Statistics*, 79, 527-539.

[3] BLANCHARD, O. AND D. QUAH (1989), "The dynamic effects of aggregate demand and supply disturbances," *American Economic Review*, 79, 655-73.

[4] BREUSCH, T. AND A. PAGAN (1979), "A Siimple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica*, 47, 1287-1294.

[5] CARD, D. (1995), "Using Geographic Variations in College Proximity to Estimate the Returns to Schooling," in *Aspects of Labor Market Behavior: Essays in Honor of John Vanderkamp*," L. N. Christofides, E. K. Grand, and R. Swidinsky, eds., Toronto: University of Toronto Press.

[6] CARNEIRO, P. AND J. HECKMAN (2002), "The Evidence on Credit Constraints in Post Secondary Schooling," *Economic Journal*, 112, 705-734.

[7] CHAMBERLAIN, G. (1987), "Asymptotic Efficiency in Estimation With Conditional Moment Restrictions," *Journal of Econometrics* 34, 305-334.

[8] CRAGG, J. (1997), "Using Higher Moments to Estimate the Simple Errors-in-Variables Model," *Rand Journal of Economics* 28, S71-S91.

[9] CRAGG, J. AND S. DONALD (1997), "Inferring the Rank of a Matrix," *Journal of Econometrics* 76, 223-250.

[10] DAGENAIS, M. G. AND D. L. DAGENAIS (1997), "Higher Moment Estimators for Linear Regression Models with Errors in the Variables," *Journal of Econometrics* 76, 193-222.

[11] ERICKSON, T. AND T. WHITED (2002), "Two-Step GMM Estimation of the Errors-in-Variables Model Using High Order Moments," *Econometric Theory*, 18, 776-799.

[12] FULLER, W. A. (1987), Measurement Error Models, New York: Wiley.

[13] HANSEN, L., (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

[14] HAUSMAN, J. (1983), ”Specification and Estimation of Simultaneous Models,” in Z. Griliches and M. Intriligator (eds.), Handbook of Econometrics, Vol. 1, North-Holland, 391-448.

[15] HAUSMAN, J.A., W. K. NEWEY, AND J. L. POWELL (1995), ”Nonlinear Errors in Variables: Estimation of Some Engel Curves,” *Journal of Econometrics*, 65, 205-253.

[16] HECKMAN, J. (1974), ”Shadow Prices, Market Wages, and Labor Supply,” *Econometrica*, 42, 679-693.

[17] HECKMAN, J. (1976), ”Sample Selection Bias as a Specification Error,” *Econometrica*, 47, 153-161.

[18] HECKMAN, J. AND E. VYTLACIL (1998). "Instrumental Variables Methods for the Correlated Random Coefficient Model", *Journal of Human Resources*, 33, 974-987.

[19] HOGAN, V. AND R. RIGOBON (2003), ”Using Unobserved Supply Shocks to Estimate the Returns to Educations,” Unpublished manuscript.

[20] HSIAO, C. (1983), “Identification,” in Z. Griliches and M. Intriligator (eds.), Handbook of Econometrics, Vol. 1, North-Holland, 223-283.

[21] JORESKOG, K. G. AND D. SORBOM (1984), Advances in Factor Analysis and Structural Equation Models, Rowman and Littlefield: Maryland.

[22] KING, M., E. SENTANA, AND S. WADHWANI (1994), ”Volatility and Links Between National Stock Markets,” *Econometrica,* 62, 901-933.

[23] KITAMURA, Y., G. TRIPATHI, AND H. AHN (2003), ”Empirical Likelihood Based Inference in Conditional Moment Restriction Models,” unpublished manuscript.

[24] KLEIN, R. AND F. VELLA (2003), ”Identification and Estimation of the Triangular Simultaneous Equations Model in the Absence of Exclusion Restrictions Through the Presence of Heteroskedasticity,” unpublished manuscript.

[25] LESER, C. E. V. (1963), "Forms of Engel Functions," *Econometrica*, 31, 694-703.

[26] LEWBEL, A. (1996), ”Demand Estimation With Expenditure Measurement Errors on the Left and Right hand side,” *Review of Economics and Statistics,* 78, 718-725.

[27] LEWBEL, A. (1997), ”Constructing Instruments for Regressions With Measurement Error When No Additional Data are Available, With an Application to Patents and R&D,” *Econometrica,* 65, 1201-1213.

[28] MEGHIR, C. AND J.-M. ROBIN (1992), ”Frequency of Purchase and the Estimation of Demand Systems,” *Journal of Econometrics*, 53, 53-86.

[29] NEWEY, W. (1993), ”Efficient Estimation of Models With Conditional Moment Restrictions,” in *Handbook of Statistics*, vol. 11, ed. by G. S. Maddala, C. R. Rao, and H. D. Vinod, Amsterdam: North Holland, chapter 16.

[30] NEWEY, W. AND J. POWELL (2003), ”Instrumental Variables Estimation of Nonparametric Models,” *Econometrica*, 71, 1557-1569.

[31] NEWEY, W. AND R. J. SMITH (2004), ”Higher Order Properties of Gmm and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219-255.

[32] OTSU, TAISUKE, (2003), ”Penalized Empirical Likelihood Estimation of Conditional Moment Restriction Models With Unknown Functions," University of Wisconsin unpublished manuscript.

[33] RIGOBON, R. (2002), ”The Curse of Non-investment Grade Countries,” *Journal of Development Economics*, 69, 423-449.

[34] RIGOBON, R. (2003), ”Identification Through Heteroskedasticity,” *Review of Economics and Statistics*, 85,

[35] ROEHRIG, C. S. (1988), ”Conditions for Identification in Nonparametric and Parametric Models,” *Econometrica*, 56, 433-447.

[36] SENTANA E. AND G. FIORENTINI (2001), ”Identification, Estimation, and Testing of Conditional Heteroskedastic Factor Models,” *Journal of Econometrics*, 102, 143-164.

[37] STAIGER, D. AND J. H. STOCK (1997), ”Instrumental Variable Regression With Weak Instruments,” *Econometrica*, 65, 557-586.

[38] STOCK, J. H., J. H. WRIGHT, AND M. YOGO (2002), ”A Survey of Weak Instruments and Weak Identification in Generalized Methods of Moments” *Journal of Business and Economic Statistics*, 20, 518-529

[39] VELLA, F. AND M. VERBEEK (1997), ”Order as an Instrumental Variable: An Application to the Return to Schooling,” unpublished manuscript.

[40] WORKING, H., (1943), "Statistical Laws of Family Expenditure", *Journal of the American Statistical Association*, 38, 43-56.

Table 1. Simulation Results

Triangular Model Two Stage Least Squares

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
β_{11}	1.00	1.00	.134	.915	1.00	1.09	.134	.105	.087
β_{12}	1.00	1.01	.273	.835	1.00	1.17	.273	.209	.168
γ_1	1.00	.999	.036	.980	1.00	1.02	.036	.026	.019
β_{21}	1.00	1.00	.129	.917	1.00	1.08	.129	.102	.084
β_{22}	1.00	1.00	.275	.830	.996	1.17	.275	.209	.168

Simultaneous System GMM

	TRUE	MEAN	SD	LQ	MED	UQ	RMSE	MAE	MDAE
β_{11}	1.00	1.03	2.75	.918	1.00	1.09	2.75	.130	.085
β_{12}	1.00	.999	.667	.833	1.00	1.17	.667	.218	.170
γ_1	1.00	1.01	1.26	.974	.999	1.02	1.26	.047	.025
β_{21}	1.00	1.02	3.55	.913	1.00	1.09	3.55	.162	.090
β_{22}	1.00	1.05	6.53	.830	.998	1.17	6.53	.299	.172
$-\gamma_2$	.500	.504	1.63	.527	.501	.477	1.63	.059	.025

Notes: The reported statistics are as follows. TRUE is the true value of the parameter, MEAN and SD are the mean and standard deviation of the estimates across the simulations. LQ, MED, and UQ are the 25% (lower) 50% (median) and 75% (upper) quartiles. RMSE, MAE, and MDAE are the root mean squared error, mean absolute error and median absolute error of the estimates.

Table 2. Engel Curve Estimates

	OLS	TSLS 1	TSLS 2	TSLS 3	GMM 1	GMM 2	GMM 3
β_{11}	0.361	0.336	0.318	0.336	0.336	0.332	0.337
	(.0056)	(0.012)	(0.035)	(0.011)	(0.012)	(0.028)	(0.011)
γ_1	-0.127	-0.086	-0.055	-0.086	-0.086	-0.078	-0.087
	(.0083)	(0.020)	(0.058)	(0.018)	(0.020)	(0.047)	(0.018)
χ^2						18.8	17.7
d.f.						11	12
p-value						0.065	0.125

Notes: OLS is an ordinary least squares regression of food share Y_1 on household characteristics X and log total expenditures Y_2. TSLS 1 is this regression estimated using two stage least squares with log real income as an ordinary outside instrument. TSLS 2 is this paper's heteroskedasticity based estimator, equation (12), which uses $(Z - \overline{Z})\widehat{\varepsilon}_2$ as instruments, where Z is all the regressors X except the constant. TSLS 3 uses both $(Z - \overline{Z})\widehat{\varepsilon}_2$ and the outside variable log real income as instruments. GMM 1, GMM 2, and GMM 3 are the same three models estimated by efficient GMM, based on Corollary 4.

Reported above are $\beta_{11} = \overline{X}'\beta$, which is the Engel curve intercept at the mean of the regressors, and γ_1, which is the Engel curve slope coefficient of Y_2. Standard errors are in parentheses. Also reported is the Hansen (1982) specification test chi squared statistic for the overidentified GMM models 2 and 3, along with its degrees of freedom and p-value.