




## Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models

Arthur Lewbel


To cite this article: Arthur Lewbel (2012) Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models, Journal of Business & Economic Statistics, 30:1, 67-80, DOI: [10.1080/07350015.2012.643126](https://doi.org/10.1080/07350015.2012.643126)

To link to this article: <https://doi.org/10.1080/07350015.2012.643126>

 View supplementary material 

 Published online: 22 Feb 2012.

 Submit your article to this journal 

 Article views: 2500

 View related articles 

 Citing articles: 184 View citing articles 

# Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models

Arthur LEWBEL

Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467  
([lewbel@bc.edu](mailto:lewbel@bc.edu))

This article proposes a new method of obtaining identification in mismeasured regressor models, triangular systems, and simultaneous equation systems. The method may be used in applications where other sources of identification, such as instrumental variables or repeated measurements, are not available. Associated estimators take the form of two-stage least squares or generalized method of moments. Identification comes from a heteroscedastic covariance restriction that is shown to be a feature of many models of endogeneity or mismeasurement. Identification is also obtained for semiparametric partly linear models, and associated estimators are provided. Set identification bounds are derived for cases where point-identifying assumptions fail to hold. An empirical application estimating Engel curves is provided.

KEY WORDS: Endogeneity; Heteroscedastic errors; Identification; Measurement error; Partly linear model; Simultaneous system.

## 1. INTRODUCTION

This article provides a new method of identifying structural parameters in models with endogenous 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 heteroscedastic 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. Even when this main identifying assumption does not hold, it is still possible to obtain set identification, specifically bounds on the parameters of interest.

For the main model, estimators take the form of modified two-stage least squares or generalized method of moments (GMM). Identification of semiparametric partly linear triangular and simultaneous systems is also considered. In an empirical application, this article's methodology is applied to deal with measurement error in total expenditures, resulting in Engel curve estimates that are similar to those obtained using a more standard instrument. A literature review shows similarly satisfactory empirical results obtained by other researchers using this article's methodology, based on earlier working article versions of this article.

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

$$Y_1 = X'\beta_1 + Y_2\gamma_1 + \varepsilon_1 \quad (1)$$

$$Y_2 = X'\beta_2 + Y_1\gamma_2 + \varepsilon_2. \quad (2)$$

This system of equations is triangular when  $\gamma_2 = 0$ , otherwise it is fully simultaneous (if it is known that  $\gamma_1 = 0$ , then renumber the equations to set  $\gamma_2 = 0$ ). The errors  $\varepsilon_1$  and  $\varepsilon_2$  may be correlated with each other.

Assume  $E(\varepsilon X) = 0$ , which is the standard minimal regression assumption for the exogenous regressors  $X$ . This permits identification of the reduced form, but is of course not sufficient to identify the structural model coefficients. Typically, structural model identification is obtained by imposing equality constraints on some coefficients, such as assuming that some elements of  $\beta_1$  or  $\beta_2$  are zero, which is equivalent to assuming the availability of instruments. This article instead obtains identification by restricting correlations of  $\varepsilon\varepsilon'$  with  $X$ . The resulting identification is based on higher moments and so is likely to provide less reliable estimates than identification based on standard exclusion restrictions, but may be useful in applications where traditional instruments are not available or could be used along with traditional instruments to increase efficiency.

Restricting correlations of  $\varepsilon\varepsilon'$  with  $X$  does not automatically provide identification. In particular, the structural model parameters remain unidentified under the standard homoskedasticity assumption that  $E(\varepsilon\varepsilon' | X)$  is constant, and more generally, are not identified when  $\varepsilon$  and  $X$  are independent.

However, what this article shows is that the model parameters may be identified given some heteroscedasticity. In particular, identification is obtained by assuming that  $\text{cov}(X, \varepsilon_j^2) \neq 0$  for  $j = 2$  in a triangular system (or for both  $j = 1$  and  $j = 2$  in a fully simultaneous system) and assuming that  $\text{cov}(Z, \varepsilon_1\varepsilon_2) = 0$  for an observed  $Z$ , where  $Z$  can be a subset of  $X$ . If  $\text{cov}(Z, \varepsilon_1\varepsilon_2) \neq 0$ , then set identification, specifically bounds on

Later, the identification results will be extended to cases where  $X'\beta_1$  and  $X'\beta_2$  are replaced by unknown functions of  $X$ .

parameters, can still be obtained as long as this covariance is not too large.

The remainder of this section provides examples of models where these identifying assumptions hold and comparisons to related results in the literature.

### 1.1 Independent Errors

For the simplest possible motivating example, let Equations (1) and (2) hold. Suppose  $\varepsilon_1$  and  $\varepsilon_2$  have the standard model error property of being mean zero and are conditionally independent of each other, so  $\varepsilon_1 \perp \varepsilon_2 \mid Z$  and  $E(\varepsilon_1) = 0$ . It would then follow immediately that the key identifying assumption  $\text{cov}(Z, \varepsilon_1 \varepsilon_2) = 0$  holds, because then  $E(\varepsilon_1 \varepsilon_2 Z) = E(\varepsilon_1)E(\varepsilon_2 Z) = 0$ . This, along with ordinary heteroscedasticity of the errors  $\varepsilon_1$  and  $\varepsilon_2$ , then suffices for identification.

More generally, independence or uncorrelatedness of  $\varepsilon_1$  and  $\varepsilon_2$  is not required, for example, it is shown below that the identifying assumptions still hold if  $\varepsilon_1$  and  $\varepsilon_2$  are correlated with each other through a factor structure, and they hold in a classical measurement error framework.

### 1.2 Classical Measurement Error

Consider a standard linear regression model with a classically mismeasured regressor. Suppose we do not have an outside instrument that correlates with the mismeasured regressor, which is the usual method of identifying this model. It is shown here that we can identify the coefficients in this model just based on heteroscedasticity. The only nonstandard assumption that will be needed for identification is the assumption that the errors in a linear projection of the mismeasured regressor on the other regressors be heteroscedastic, which is more plausible than homoskedasticity in most applications.

The goal is estimation of the coefficients  $\beta_1$  and  $\gamma_1$  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

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

Here,  $U$  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  $Y_1$ . So far, all of these assumptions are exactly those of the classical linear regression mismeasured regressor model.

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

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

Substituting out the unobservable  $Y_2^*$  yields the familiar triangular system associated with measurement error models

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

where the  $Y_1$  equation is the structural equation to be estimated, the  $Y_2$  equation is the instrument equation, and  $\varepsilon_1$  and  $\varepsilon_2$  are unobserved errors.

The standard way to obtain identification in this model is by an exclusion restriction, that is, by assuming that one or more elements of  $\beta_1$  equal zero and that the corresponding elements of  $\beta_2$  are nonzero. The corresponding elements of  $X$  are then instruments, and the model is estimated by linear 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  $\hat{Y}_2$  and the subset of  $X$  that has nonzero coefficients.

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

However, in this mismeasured regressor model, there is no reason to believe that  $V_2$ , the error in the  $Y_2$  equation, would be independent of  $X$ , because the  $Y_2$  equation (what would be the first-stage regression in two-stage least squares) is just the linear projection of  $Y_2$  on  $X$ , not a structural model motivated by any economic theory.

The perhaps surprising result, which follows from Theorem 1 below, is that if  $V_2$  is heteroscedastic (and hence not independent of  $X$ , as expected), then the structural model coefficients in this model are identified and can be easily estimated. The above assumptions yield a triangular model with  $E(X\varepsilon) = 0$ ,  $\text{cov}(X, \varepsilon_2^2) \neq 0$ , and  $\text{cov}(X, \varepsilon_1 \varepsilon_2) = 0$  and hence satisfy this article's required conditions for identification.

The classical measurement error assumptions are used here by way of illustration. They are much stronger than necessary to apply this article's methodology. For example, identification is still possible when the measurement error  $U$  is correlated with some of the elements  $X$  and the error independence assumptions given above can be relaxed to restrictions on just a few low-order moments.

### 1.3 Unobserved Single-Factor Models

A general class of models that satisfy this article'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:

$$Y_1 = X'\beta_1 + Y_2\gamma_1 + \varepsilon_1, \quad \varepsilon_1 = \alpha_1 U + V_1 \quad (3)$$

$$Y_2 = X'\beta_2 + Y_1\gamma_2 + \varepsilon_2, \quad \varepsilon_2 = \alpha_2 U + V_2, \quad (4)$$

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 (3) with  $\alpha_1 = -\gamma_1$  and Equation (4) with  $\gamma_2 = 0$  and  $\alpha_2 = 1$ . The unobserved common factor  $U$  is the measurement error in  $Y_2$ .

*Supply and Demand:* Equations (3) and (4) 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 (3) and (4) 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, 1979).

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  $\beta_1$  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.

Assume we have no ordinary instruments and no equality constraints on the parameters. Let  $Z$  be a vector of observed exogenous variables, in particular,  $Z$  could be a subvector of  $X$ , or  $Z$  could equal  $X$ . Assume  $X$  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  $V_1^2$ . An alternative set of stronger but more easily interpreted sufficient conditions are that one or both of the idiosyncratic errors  $V_j$  be heteroscedastic,  $\text{cov}(Z, V_1V_2) = 0$ , and that the common factor  $U$  be conditionally independent of  $Z$ . These are all standard assumptions, except that one usually either imposes homoscedasticity or allows for heteroscedasticity, rather than requiring heteroscedasticity.

Given these assumptions,

$$\begin{aligned} \text{cov}(Z, \varepsilon_1\varepsilon_2) &= \text{cov}(Z, \alpha_1\alpha_2U^2 + \alpha_1UV_2 \\ &\quad + \alpha_2UV_1 + V_1V_2) = 0 \\ \text{cov}(Z, \varepsilon_2^2) &= \text{cov}(Z, \alpha_2^2U^2 + 2\alpha_2UV_2 + V_2^2) \\ &= \text{cov}(Z, V_2^2) \neq 0, \end{aligned}$$

which are the requirements for applying this article's identification theorems and associated estimators.

To apply this article's estimators, it is not necessary to assume that the errors are actually given by a factor model such as  $\varepsilon_j = \alpha_jU + V_j$ . In particular, 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

$$E(X\varepsilon_1) = 0, \quad E(X\varepsilon_2) = 0, \quad \text{cov}(Z, \varepsilon_1\varepsilon_2) = 0, \quad (5)$$

along with some heteroscedasticity of  $\varepsilon_j$ . The moments (5) provide identification whether or not  $Z$  is subvector of  $X$ .

## 1.4 Empirical Examples

Based on earlier working versions of this article, a number of researchers apply this article's identification strategy and associated estimators to a variety of settings where ordinary instruments are either weak or difficult to obtain.

Giambona and Schwiendbacher (2007) applied the method in a model relating the debt and leverage ratios of firms to the tangibility of their assets. Emran and Hou (2008) applied it to a model of household consumption in China based on distance to domestic and international markets. Sabia (2007) used the method to estimate equations relating body weight to academic performance, and Rashad and Markowitz (2007) used it in a similar application involving body weight and health insurance. Finally, in a later section of this article, I report results for a model of food Engel curves where total expenditures may be mismeasured. All of these studies report that using this article's estimator yields results that are close to estimates based on traditional instruments (though Sabia 2007 also noted that his estimates are closer to ordinary least squares). Taken together, these studies provide evidence that the methodology proposed in this article may be reliably applied in a variety of real data settings where traditional instrumental variables are not available.

## 1.5 Literature Review

Surveys of methods of identification in simultaneous systems include Hsiao (1983), Hausman (1983), and Fuller (1987). Roehrig (1988) provided a useful general characterization of identification in situations where nonlinearities contribute to identification, as is the case here. Particularly relevant for this article is previous work that obtains identification based on variance and covariance constraints. With multiple-equation systems, various homoscedastic factor model covariance restrictions are used along with exclusion assumptions in the LISREL class of models (Joreskog and Sorbom 1984). The idea of using heteroscedasticity in some way to help estimation appears in Wright (1928) and so is virtually as old as the method of instrumental variables itself. Recent articles that use general restrictions on higher moments instead of outside instruments as a source of identification include Dagenais and Dagenais (1997), Lewbel (1997), Cragg (1997), and Erickson and Whited (2002).

A closely related result to this article's is Rigobon (2002, 2003), which uses heteroscedasticity based on discrete, multiple regimes instead of regressors. Some of Rigobon's identification results can be interpreted as special cases of this article's models in which  $Z$  is a vector of binary dummy variables that index regimes and are not included among the regressors  $X$ . Sentana (1992) and Sentana and Fiorentini (2001) employed a similar idea for identification in factor models. Hogan and Rigobon (2003) propose a model that, like this article's, involves decomposing the error term into components, some of which are heteroscedastic.

Klein and Vella (2010) also used heteroscedasticity restrictions to obtain identification in linear models without exclusion restrictions (an application of their method is Rummery, Vella, and Verbeek 1999), and their model also implies restrictions on how  $\varepsilon_1^2$ ,  $\varepsilon_2^2$ , and  $\varepsilon_1\varepsilon_2$  depend on regressors, but not the same restrictions as those used in the present article. The method proposed here exploits a different set of heteroscedasticity restrictions from theirs, and as a result, this article's estimators have many features that estimators in Klein and Vella (2010) do not have, including the following: This article's assumptions nest standard mismeasured regressor models and unobserved factor models, unlike theirs. This article's estimator extends to fully simultaneous systems, not just triangular systems, and

extends to a class of semiparametric models. Klein and Vella (2003) assumed a multiplicative form of heteroscedasticity that imposes strong restrictions on how all higher moments of errors depend on regressors, while this article's model places no restrictions on third and higher moments of  $\varepsilon_j$  conditional on  $X, Z$ . Finally, this article provides some set identification results, yielding bounds on parameters, that hold when point-identifying assumptions are violated.

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

Some articles have exploited GARCH system heteroscedastic specifications to obtain identification, including King, Sentana, and Wadhvani (1994) and Prono (2008). Other articles that exploit heteroscedasticity in some way to aid identification include Leamer (1981) and Feenstra (1994).

Variables that in past empirical applications have been proposed as instruments for identification might more plausibly be used as this article's  $Z$ . For example, in the returns to schooling model, Card (1995) 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 may be independent of unobserved ability (though see Carneiro and Heckman 2002) and may 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 type or quality of education one actually receives or may be correlated with proximity to locations where good jobs are available (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 article's  $Z$  to identify returns to schooling, given by  $\gamma_1$  in the triangular model, where  $Y_1$  is wages and  $Y_2$  is schooling.

The next section describes this article's main identification results for triangular and then fully simultaneous systems. This is followed by a description of associated estimators and an empirical application to Engel curve estimation. Later sections provide extensions, including set identification (bounds) for when the point-identifying assumptions do not hold and identification results for nonlinear and semiparametric systems of equations.

## 2. POINT 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 the relevant moments with probability limits of sample moments and sample projections.

### 2.1 Triangular Model Identification

First, consider the linear triangular model:

$$Y_1 = X'\beta_{10} + Y_2\gamma_{10} + \varepsilon_1 \quad (6)$$

$$Y_2 = X'\beta_{20} + \varepsilon_2. \quad (7)$$

Here,  $\beta_{10}$  indicates the true value of  $\beta_1$ , and similarly for the other parameters. Traditionally, this model would be identified by imposing equality constraints on  $\beta_{10}$ . Alternatively, if the errors  $\varepsilon_1$  and  $\varepsilon_2$  were uncorrelated, this would be a recursive system and so the parameters would be identified. Identification conditions are given here that do not require uncorrelated errors or restrictions on  $\beta_{10}$ . Example applications include unobserved factor models such as 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.  $E(XY')$ ,  $E(XY_1Y_1')$ ,  $E(XY_2Y_2')$ , 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  $Z$ ,  $\text{cov}(Z, \varepsilon_1\varepsilon_2) = 0$ .

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  $X$ . Sections 1.1, 1.2, and 1.3 provide examples of models satisfying these assumptions.

Define matrices  $\Psi_{ZX}$  and  $\Psi_{ZZ}$  by

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

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

and let  $\Psi$  be any positive definite matrix that has the same dimensions as  $\Psi_{ZZ}$ .

*Theorem 1.* Let Assumptions A1 and A2 hold for the model of Equations (6) and (7). Assume  $\text{cov}(Z, \varepsilon_2^2) \neq 0$ . Then, the structural parameters  $\beta_{10}, \beta_{20}, \gamma_{10}$ , and the errors  $\varepsilon$  are identified, and

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

$$\begin{pmatrix} \beta_{10} \\ \gamma_{10} \end{pmatrix} = (\Psi'_{ZX}\Psi\Psi_{ZX})^{-1}\Psi'_{ZX}\Psi E \left[ \begin{pmatrix} X \\ [Z - E(Z)]\varepsilon_2 \end{pmatrix} Y_1 \right]. \quad (8)$$

Proofs are in the Appendix. For  $\Psi = \Psi_{ZZ}^{-1}$ , Theorem 1 says that the structural parameters  $\beta_{10}$  and  $\gamma_{10}$  are identified by an 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. The assumption that  $Z$  is uncorrelated with  $\varepsilon_1\varepsilon_2$  means that  $(Z - \bar{Z})\varepsilon_2$  is a valid instrument for  $Y_2$  in Equation (6) since it is uncorrelated with  $\varepsilon_1$ , with the strength of the instrument (its correlation with  $Y_2$  after controlling for the other instruments  $X$ ) being proportional to the covariance of  $(Z - \bar{Z})\varepsilon_2$  with  $\varepsilon_2$ , which corresponds to the degree of heteroscedasticity of  $\varepsilon_2$  with respect to  $Z$ .

Taking  $\Psi = \Psi_{ZZ}^{-1}$  corresponds to estimation based on ordinary linear two-stage least squares. Other choices of  $\Psi$  may be preferred for increased efficiency, accounting for error heteroscedasticity. Efficient GMM estimation of this model is discussed later.

The requirement that  $\text{cov}(Z, \varepsilon_2^2)$  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  $X$ . For example, we may apply a Breusch and Pagan (1979) test for this form of heteroscedasticity to Equation (7). Also, if  $\text{cov}(Z, \varepsilon_2^2)$  is close to or equal to zero, then  $(Z - \bar{Z})\varepsilon_2$  will be a weak or useless instrument, and this problem will be evident in the form of imprecise estimates with large standard errors. Hansen (1982) type tests of GMM moment restrictions can also be implemented to check validity of the model's assumptions, particularly Assumption A2.

## 2.2 Fully Simultaneous Linear Model Identification

Now consider the fully simultaneous model

$$Y_1 = X'\beta_{10} + Y_2\gamma_{10} + \varepsilon_1 \quad (9)$$

$$Y_2 = X'\beta_{20} + Y_1\gamma_{20} + \varepsilon_2, \quad (10)$$

where the errors  $\varepsilon_1$  and  $\varepsilon_2$  may be correlated, and again no equality constraints are imposed on the structural parameters  $\beta_{10}$ ,  $\beta_{20}$ ,  $\gamma_{10}$ , and  $\gamma_{20}$ .

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 1 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  $\gamma_{20}$  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 1 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:

$$W_j = Y_j - X'E(XX')^{-1}E(XY_j). \quad (11)$$

*Assumption A3.* Define  $W_j$  by Equation (11) for  $j = 1, 2$ . The matrix  $\Phi_W$ , defined as the matrix with columns given by the vectors  $\text{cov}(Z, W_1^2)$  and  $\text{cov}(Z, W_2^2)$ , has rank 2.

Assumption A3 requires  $Z$  to contain at least two elements (though sometimes one element of  $Z$  can be a constant; see Corollary 1 later). If  $E(\varepsilon_1\varepsilon_2 | \tilde{Z}) = E(\varepsilon_1\varepsilon_2)$  for some scalar  $\tilde{Z}$ , as would arise if the common unobservable  $U$  is independent of  $\tilde{Z}$ , then Assumptions A2 and A3 might be satisfied by letting  $Z$  be a vector of different functions of  $\tilde{Z}$ , for example, defining  $Z$  as the vector of elements  $\tilde{Z}$  and  $\tilde{Z}^2$  (as long as  $\tilde{Z}$  is not binary).

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  $\text{cov}(Z, W_j^2)$ . A Breusch and Pagan (1979) test may be applied to each of these reduced-form regressions. An estimated matrix rank test such as Cragg and Donald (1997) could be applied to the resulting estimated matrix  $\Phi_W$ , or perhaps more simply test if the determinant of  $\Phi'_W\Phi_W$  is zero, since rank 2 requires that  $\Phi'_W\Phi_W$  be nonsingular.

*Assumption A4.* Let  $\Gamma$  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 (9) for  $Y_2$  and Equation (10) 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  $\beta_{10}$  and  $\beta_{20}$ , 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. In other models, the choice of  $\Gamma$  may be determined by context, for example, many economic models (like those requiring stationary dynamics or decreasing returns to scale) require coefficients such as  $\gamma_1$  and  $\gamma_2$  to be less than 1 in absolute value, which then defines a set  $\Gamma$  that satisfies Assumption A4. In a supply-and-demand model,  $\Gamma$  may be defined by downward sloping demand and upward sloping supply curves, since in that case,  $\Gamma$  only includes elements  $\gamma_1, \gamma_2$  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's (1976) finding 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 (9) and (10). Then, the structural parameters  $\beta_{10}$ ,  $\beta_{20}$ ,  $\gamma_{10}$ , and  $\gamma_{20}$  and the errors  $\varepsilon$  are identified.

## 2.3 Additional Simultaneous Model Results

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

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  $\gamma_{20} = 0$ . Having  $\gamma_{10}\gamma_{20} = 1$  also cannot occur in the supply-and-demand application, because the slopes of supply and demand curves make  $\gamma_{10}\gamma_{20} \leq 0$ . 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  $\Phi_W$  to have rank less than 2. However, Theorem 1 can be relaxed to allow  $\gamma_{10}\gamma_{20} = -1$ , by replacing Assumption A3 with the assumption that  $\Phi_\varepsilon$  has rank 2, because then Equation (27) in the proof still holds and identifies  $\gamma_{10}/\gamma_{20}$ , which along with  $\gamma_{10}\gamma_{20} = -1$  and some sign restrictions could identify  $\gamma_{10}$  and  $\gamma_{20}$  in this case. However, Assumption A3 has the advantage of being empirically testable.

In either case, Theorem 2 requires both  $\varepsilon_1$  and  $\varepsilon_2$  to be heteroscedastic with variances that depend upon  $Z$ , since otherwise the vectors  $\text{cov}(Z, \varepsilon_1^2)$  and  $\text{cov}(Z, \varepsilon_2^2)$  will equal zero. Moreover, the variances of  $\varepsilon_1$  and  $\varepsilon_2$  must be different functions of  $Z$  for the rank of  $\Phi_\varepsilon$  to be 2.

*Corollary 1.* Let Assumptions A1, A2, A3, and A4 hold in the model of Equations (9) and (10), replacing  $\text{cov}(Z, \varepsilon_1\varepsilon_2)$  in Assumption A2 with  $E(Z\varepsilon_1\varepsilon_2)$  and replacing  $\text{cov}(Z, W_j^2)$  with  $E(ZW_j^2)$  in Assumption A3, for  $j = 1, 2$ . Then, the structural parameters  $\beta_{10}$ ,  $\beta_{20}$ ,  $\gamma_{10}$ , and  $\gamma_{20}$  and the errors  $\varepsilon$  are identified.

Corollary 1 can be used in applications where  $E(\varepsilon_1\varepsilon_2) = 0$ . Theorem 2 could also be used in this case, but Corollary 1

provides additional moments. In particular, if only a scalar  $\tilde{Z}$  is known to satisfy  $\text{cov}(\tilde{Z}, \varepsilon_1 \varepsilon_2) = 0$ , then identification by Theorem 2 will fail because the rank condition in Assumption A3 is violated with  $Z = \tilde{Z}$ , but identification may still be possible using Corollary 1 because there we may let  $Z = (1, \tilde{Z})$ .

*Corollary 2.* Let Assumptions A1 and A2 hold for the model of Equations (9) and (10). Assume  $\text{cov}(Z, \varepsilon_2^2) \neq 0$ , that some element of  $\beta_{20}$  is known to equal zero and the corresponding element of  $\beta_{10}$  is nonzero. Then, the structural parameters  $\beta_{10}$ ,  $\beta_{20}$ ,  $\gamma_{10}$ , and  $\gamma_{20}$  and the errors  $\varepsilon$  are identified.

Corollary 2 is like Theorem 1, except that it assumes an element of  $\beta_{20}$  is zero instead of assuming  $\gamma_{20}$  is zero to identify Equation (10). Then, as in Theorem 1, Corollary 2 uses  $\text{cov}(Z, \varepsilon_1 \varepsilon_2) = 0$  to identify Equation (9) without imposing the rank 2 condition of Assumption A3 and the inequality constraints of Assumption A4. Only a scalar  $Z$  is needed for identification using Theorem 1 or Corollary 1 or 2.

### 3. ESTIMATION

#### 3.1 Simultaneous System Estimation

Consider estimation of the structural model of Equations (9) and (10) based on Theorem 2. Define  $S$  to be the vector of elements of  $Y$  and  $X$ , and the elements of  $Z$  that are not already contained in  $X$ , if any.

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

$$\begin{aligned} Q_1(\theta, S) &= X(Y_1 - X'\beta_1 - Y_2\gamma_1) \\ Q_2(\theta, S) &= X(Y_2 - X'\beta_2 - Y_1\gamma_2) \\ Q_3(\theta, S) &= Z - \mu \\ Q_4(\theta, S) &= (Z - \mu)(Y_1 - X'\beta_1 - Y_2\gamma_1) \\ &\quad \times (Y_2 - X'\beta_2 - Y_1\gamma_2). \end{aligned}$$

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

*Corollary 3.* Assume Equations (9) and (10) hold. Define  $\theta$ ,  $S$ , and  $Q(\theta, S)$  as above. Let Assumptions A1, A2, A3, and A4 hold. Let  $\Theta$  be the set of all values  $\theta$  might take on, and let  $\theta_0$  denote the true value of  $\theta$ . 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  $E(\varepsilon_1 \varepsilon_2) = 0$ , then  $\mu$  can be dropped from  $\theta$ , 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 (9) and (10) is completely straightforward. With a sample of  $n$  observations  $S_1, \dots, S_n$ , the standard Hansen (1982) GMM estimator is

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{i=1}^n Q(\theta, S_i)' \Omega_n^{-1} \sum_{i=1}^n Q(\theta, S_i) \quad (12)$$

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

resulting estimator is efficient GMM with

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N \left( 0, E \left( \frac{\partial Q(\theta_0, S)}{\partial \theta'} \right) \Omega_0^{-1} \times E \left( \frac{\partial Q(\theta_0, S)}{\partial \theta'} \right)' \right). \quad (13)$$

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 (as might occur if the errors are close to homoskedastic), 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.

The standard regularity conditions for the large-sample properties of GMM impose compactness of  $\Theta$ . 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  $\Theta$  so that the product of the first two elements of every  $\theta \in \Theta$  is finite, nonpositive, and excludes an open neighborhood of  $-1$ . 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 1, 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  $\beta_2$  and  $\gamma_2$  accordingly.

Based on the proof of Theorem 2, a numerically simpler but possibly less efficient estimator would be the following. First, let  $\hat{W}_j$  be the vector of residuals from linearly regressing  $Y_j$  on  $X$ . Next, let  $\hat{C}_{jkh}$  be the sample covariance of  $\hat{W}_j \hat{W}_k$  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 (27), estimate  $\gamma_1$  and  $\gamma_2$  by:

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

where  $\Gamma$  is a compact set satisfying Assumption A4. The above estimator for  $\gamma_1$  and  $\gamma_2$  is numerically equivalent to an ordinary nonlinear least squares regression over  $K$  observations, where  $K$  is the number of elements of  $Z$ . Finally,  $\beta_1$  and  $\beta_2$  may be estimated by linearly regressing  $Y_1 - Y_2 \hat{\gamma}_1$  and  $Y_2 - Y_1 \hat{\gamma}_2$  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 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  $\mu$  as before, and now let  $\theta = \{\gamma_1, \beta_1, \beta_2, \mu\}$

and

$$\begin{aligned} Q_1(\theta, S) &= X(Y_1 - X'\beta_1 - Y_2\gamma_1) \\ Q_2(\theta, S) &= X(Y_2 - X'\beta_2) \\ Q_3(\theta, S) &= Z - \mu \\ Q_4(\theta, S) &= (Z - \mu)(Y_1 - X'\beta_1 - Y_2\gamma_1) \\ &\quad \times (Y_2 - X'\beta_2). \end{aligned}$$

Let  $Q(\theta, S)$  be the vector obtained by stacking the above four vectors into one long vector, and we immediately obtain Corollary 4.

*Corollary 4.* Assume Equations (6) and (7) hold. Define  $\theta$ ,  $S$ , and  $Q(\theta, S)$  as above. Let Assumptions A1 and A2 hold with  $\text{cov}(Z, W_2^2) \neq 0$ . Let  $\Theta$  be the set of all values  $\theta$  might take on, and let  $\theta_0$  denote the true value of  $\theta$ . Then, the only value of  $\theta \in \Theta$  that satisfies  $E[Q(\theta, S)] = 0$  is  $\theta = \theta_0$ .

The GMM estimator (12) and limiting distribution (13) then follow immediately.

Based on Theorem 1, a simpler estimator of the triangular system of Equations (6) and (7) is as follows. With  $\gamma_{20} = 0$ ,  $\beta_{20}$  can be estimated by linearly regressing  $Y_2$  on  $X$ . Then, letting  $\widehat{\varepsilon}_{2i}$  be the residuals from this regression,  $\beta_{10}$  and  $\gamma_{10}$  can be estimated by an ordinary linear two-stage least squares regression of  $Y_1$  on  $Y_2$  and  $X$ , using  $X$  and  $(Z - \bar{Z})\widehat{\varepsilon}_2$  as instruments, where  $\bar{Z}$  is the sample mean of  $Z$ . Letting overbars denote sample averages, the resulting estimators are:

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

$$\begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\gamma}_1 \end{pmatrix} = (\widehat{\Psi}'_{ZX} \widehat{\Psi}^{-1}_{ZZ} \widehat{\Psi}_{ZX})^{-1} \widehat{\Psi}'_{ZX} \widehat{\Psi}^{-1}_{ZZ} \begin{pmatrix} \overline{X'Y_1} \\ (Z - \bar{Z})\widehat{\varepsilon}_2, Y_1 \end{pmatrix}, \quad (14)$$

where  $\widehat{\Psi}_{ZX}$  replaces the expectation defining  $\Psi_{ZX}$  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  $\Psi_{ZZ}^{-1}$ . 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 instruments  $(Z - \bar{Z})\widehat{\varepsilon}_2$ . Using the standard theory of two-step estimators (see, e.g., Newey and McFadden 1994), with independent, identically distributed observations, this gives:

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

where

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

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

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.

### 3.3 Extension: Additional Endogenous Regressors

We consider two cases here: additional endogenous regressors for which we have ordinary outside instruments, and additional endogenous regressors to be identified using heteroscedasticity.

In the triangular system, the estimator can be described as a linear two-stage least squares regression of  $Y_1$  on  $X$  and on  $Y_2$ , using  $X$  and an estimate of  $[Z - E(Z)]\varepsilon_2$  as instruments. Suppose now that in addition to  $Y_2$ , one or more elements of  $X$  are also endogenous. Suppose for now that we also have a set of ordinary instruments  $P$  (so  $P$  includes all the exogenous elements of  $X$  and enough additional outside instruments so that  $P$  has at least the same number of elements as  $X$ ). It then follows that estimation could be done by a linear two-stage least squares regression of  $Y_1$  on  $X$  and on  $Y_2$ , using  $P$  and an estimate of  $[Z - E(Z)]\varepsilon_2$  as instruments. Note, however, that it will now be necessary to also estimate the  $Y_2$  equation by two-stage least squares, that is, we must first regress  $Y_2$  on  $X$  by two-stage least squares using instruments  $P$  to obtain the estimated coefficient  $\widehat{\beta}_2$ , before constructing  $\widehat{\varepsilon}_2 = Y_2 - X'\widehat{\beta}_2$ . Then, as before, the estimate of  $[Z - E(Z)]\varepsilon_2$  is  $(Z - \bar{Z})\widehat{\varepsilon}_2$ . Alternatively, the GMM estimator now has  $Q_1(\theta, S)$  and  $Q_2(\theta, S)$  given by  $Q_1(\theta, S) = P(Y_1 - X'\beta_1 - Y_2\gamma_1)$  and  $Q_2(\theta, S) = P(Y_2 - X'\beta_2)$ , while  $Q_3(\theta, S)$  and  $Q_4(\theta, S)$  are the same as before.

Similar logic extends to the case where we have more than one endogenous regressor to be identified from heteroscedasticity. For example, suppose we have the model:

$$Y_1 = X'\beta_{10} + Y_2\gamma_{10} + Y_3\delta_{10} + \varepsilon_1$$

$$Y_2 = X'\beta_{20} + \varepsilon_2, \quad Y_3 = X'\beta_{30} + \varepsilon_3.$$

So, now we have two endogenous regressors,  $Y_2$  and  $Y_3$ , with no available outside instruments or exclusions. If our assumptions hold both for  $\varepsilon_2$  and for  $\varepsilon_3$  in place of  $\varepsilon_2$ , then the model for  $Y_1$  can be estimated by two-stage least squares, using  $X$  and estimates of both  $[Z - E(Z)]\varepsilon_2$  and  $[Z - E(Z)]\varepsilon_3$  as instruments.

## 4. ENGEL CURVE ESTIMATES

An Engel curve for food is empirically estimated, where total expenditures may be mismeasured. Total expenditures are subject to potentially large measurement errors, due in part to infrequently purchased items (see, e.g., Meghir and Robin 1992). 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 and the possibly mismeasured regressor  $Y_2$  is log real total expenditures. Sample means are  $\bar{Y}_1 = 0.285$  and  $\bar{Y}_2 = 0.599$ . The other regressors  $X$  are a constant, age, spouse's



age, squared ages, seasonal dummies, and dummies for spouse working, gas central heating, ownership of a washing machine, one car, and two cars. There are 854 observations.

The model is  $Y_1 = X'\beta_1 + Y_2\gamma_1 + \varepsilon_1$ . This is the Working (1943) and Leser (1963) functional form for Engel curves. Non-parametric and parametric regression analyses of these 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).

Table 1 summarizes the empirical results. Ordinary least squares, which does not account for mismeasurement, has an estimated log total expenditure coefficient of  $\hat{\gamma}_1 = -0.127$ . Ordinary two-stage least squares, using log total income as an instrument, substantially reduces the estimated coefficient to  $\hat{\gamma}_1 = -0.086$ . This is model TSLS 1 or equivalently GMM 1 in Table 1. TSLS1 and GMM 1 are exactly identified and so are numerically equivalent.

If we did not observe income for use as an instrument, we might instead apply the GMM estimator based on Corollary 4, using the moments  $\text{cov}(Z, \varepsilon_1\varepsilon_2) = 0$ . 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 1, which yields  $\hat{\gamma}_1 = -0.078$ . This is relatively close to the estimate based on the external instrument log income, as would be expected if income is a valid instrument and if this article's methodology for identification and estimation without external instruments is also valid. The standard errors in GMM 2 are a good bit higher than those of GMM 1, suggesting that not having an external 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  $(Z - \bar{Z})\varepsilon_2$  as instruments (Equation (14)). The GMM 2 estimates are closer than TSLS 2 to the income-instrument-based estimates GMM 1 and have smaller standard errors, which shows that the increased asymptotic 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 1 also reports estimates obtained using both moments based on the external instrument, log income, and on  $\text{cov}(Z, \varepsilon_1\varepsilon_2) = 0$ . The results in TSLS 3 and GMM 3 are very similar to TSLS 1 and GMM 1, which just use the external instrument. This is consistent with validity of both sets of identifying moments, but with the outside instrument being much stronger or more informative, as expected. The Hansen test also fails to reject this joint set of overidentifying moments, with a  $p$ -value of 12.5%.

To keep the analysis simple, possible mismeasurement of the food budget share arising from mismeasurement of total expenditures, as in Lewbel (1996), has been ignored. This is not an uncommon assumption, for example, 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). However, as a check, the Engel curves were reestimated in the form of quantities of food regressed on levels of total expenditures. The

results were more favorable than those reported in Table 1. In particular, the ordinary least squares estimate of the coefficient of total expenditures was 0.124, the two-stage least squares estimate using income as an instrument was 0.172, and the two-stage least squares estimate using this article's moments was 0.174, nearly identical to the estimate based on the outside instrument.

One may question the validity of the assumptions for applying Theorem 1 in this application. Also, although income is commonly used as an outside instrument for total expenditures, it could still have flaws as an instrument (e.g., it is possible for reported consumption and income to have common sources of measurement errors). In particular, the estimates show a reversal of the usual attenuation direction of measurement error bias, which suggests some violation of the assumptions of the classical measurement error model, for example, it is possible that the measurement error could be negatively correlated with instruments or with other covariates.

Still, it is encouraging that this article's methodology for obtaining estimates without external instruments yields estimates that are close to (though not as statistically significant as) estimates that are obtained by using an ordinary external instrument, and the resulting overidentifying moments are not statistically rejected.

In practice, this article's estimators will be most useful for applications where external instruments are either weak or unavailable. The reason for applying it here in the Engel curve context, where a strong external instrument exists, is to verify that the method works in real data, in the sense that this article's estimator, applied without using the external instrument, produces estimates that are very close to those that were obtained when using the outside instrument. The fact that the method is seen to work in this context where the results can be checked should be encouraging for other applications where alternative strong instruments are not available.

## 5. SET IDENTIFICATION-RELAXING IDENTIFYING ASSUMPTIONS

This article's methodology is based on three assumptions—namely, regressors  $X$  uncorrelated with errors  $\varepsilon$ , heteroscedastic errors  $\varepsilon$ , and  $\text{cov}(Z, \varepsilon_1\varepsilon_2) = 0$ . As shown earlier, this last assumption arises from classical measurement error and omitted factor models, but one may still question whether it holds exactly in practice. Theorem 3 shows that one can still identify sets, specifically interval bounds, for the model parameters when this assumption is violated by assuming this covariance is small rather than zero. Small here means that this covariance is small relative to the heteroscedasticity in  $\varepsilon_2$ ; specifically, Theorem 3 assumes that the correlation between  $Z$  and  $\varepsilon_1\varepsilon_2$  is smaller (in magnitude) than the correlation between  $Z$  and  $\varepsilon_2^2$ .

For convenience, Theorem 3 is stated using a scalar  $Z$ , but given a vector  $Z$ , one could exploit the fact that Theorem 3 would then hold for any linear combination of the elements of  $Z$ , and one could choose the linear combination that minimizes the estimated size of the identified set.

Define  $W_j$  by Equation (11) for  $j = 1, 2$ . Given a random scalar  $Z$  and a scalar constant  $\zeta$  define  $\Gamma_1$  as the set of all values of  $\gamma_1$  that lie in the closed interval bounded by the two roots (if

Table 1. Engel Curve Estimates

	OLS	TSLs 1	TSLs 2	TSLs 3	GMM 1	GMM 2	GMM 3
$\beta_{11}$	0.361 (0.0056)	0.336 (0.012)	0.318 (0.035)	0.336 (0.011)	0.336 (0.012)	0.332 (0.028)	0.337 (0.011)
$\gamma_1$	-0.127 (0.0083)	-0.086 (0.020)	-0.055 (0.058)	-0.086 (0.018)	-0.086 (0.020)	-0.078 (0.047)	-0.087 (0.018)
$\chi^2$						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 external instrument. TSLs 2 is this article's heteroscedasticity-based estimator, Equation (14), which uses  $(Z - \bar{Z})\varepsilon_2$  as instruments, where  $Z$  is all the regressors  $X$  except the constant. TSLs 3 uses both  $(Z - \bar{Z})\varepsilon_2$  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  $\beta_{11} = \bar{X}\beta$ , which is the Engel curve intercept at the mean of the  $X$  regressors, and  $\gamma_1$ , 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.

they are real) of the quadratic equation:

$$\frac{[\text{cov}(W_1 W_2, Z)]^2}{[\text{cov}(W_2^2, Z)]^2} - \frac{\text{var}(W_1 W_2)}{\text{var}(W_2^2)} \tau^2 + 2 \left( \frac{\text{cov}(W_1 W_2, W_2^2)}{\text{var}(W_2^2)} \tau^2 - \frac{\text{cov}(W_1 W_2, Z)}{\text{cov}(W_2^2, Z)} \right) \gamma_1 + (1 - \tau^2) \gamma_1^2 = 0. \quad (15)$$

Also, define  $B_1$  as the set of all values of  $\beta_1 = E(XX')^{-1}E[X(Y_1 - Y_2\gamma_1)]$  for each  $\gamma_1 \in \Gamma_1$ .

*Theorem 3.* Let Assumption A1 hold for the model of Equations (6) and (7). Assume  $E(X\varepsilon_1) = 0$ ,  $E(X\varepsilon_2) = 0$ , and for some observed random scalar  $Z$  and some nonnegative constant  $\tau < 1$ ,  $|\text{corr}(Z, \varepsilon_1\varepsilon_2)| \leq \tau |\text{corr}(Z, \varepsilon_2^2)|$ . Then the structural parameters  $\gamma_{10}$  and  $\beta_{10}$  are set identified by  $\gamma_{10} \in \Gamma_1$ ,  $\beta_{10} \in B_1$ , and  $\beta_{20}$  is point identified by  $\beta_{20} = E(XX')^{-1}E(XY_2)$ .

Note that an implication of Theorem 3 is that Equation (15) has real roots whenever  $|\text{corr}(Z, \varepsilon_1\varepsilon_2)| < |\text{corr}(Z, \varepsilon_2^2)|$ , and  $\tau$  is defined as an upper bound on the ratio of these two correlations. The smaller the value of  $\tau$  is, the smaller will be the identified sets  $\Gamma_1$  and  $B_1$ , and hence the tighter will be the bounds on  $\gamma_{10}$  and  $\beta_{10}$  given by Theorem 3. One can readily verify that the sets  $\Gamma_1$  and  $B_1$  collapse to points, corresponding to Theorem 1, when  $\tau = 0$ .

An obvious way to construct estimates based on Theorem 3 is to substitute  $W_j = Y_j - X'E(XX')^{-1}E(XY_j)$  into Equation (15), replace all the expectations in the result with sample averages, and then solve for the two roots of the resulting quadratic equation given  $\tau$ . These roots will then be consistent estimates of the boundary of the interval that brackets  $\gamma_{10}$ .

To illustrate the size of the bounds implied by Theorem 3, consider the model:

$$Y_1 = \beta_{11} + X\beta_{12} + Y_2\gamma_1 + \varepsilon_1, \quad \varepsilon_1 = U + e^X S_1 \quad (16)$$

$$Y_2 = \beta_{21} + X\beta_{22} + \varepsilon_2, \quad \varepsilon_2 = U + e^{-X} S_2, \quad (17)$$

where  $X$ ,  $U$ ,  $S_1$ , and  $S_2$  are independent standard normal scalars,  $Z = X$ , and  $\beta_{11} = \beta_{12} = \beta_{21} = \beta_{22} = \gamma_1 = 1$ . A supplemental appendix to this article includes a Monte Carlo analysis of the estimator using this design. It can be shown by tedious but straightforward algebra that for this design, Equation (15) re-

duces to:

$$1 - \frac{12 + 12e^2 - e^4 + 3e^8}{2 + 4e^2 - e^4 + 3e^8} \tau^2 + 2 \left( \frac{5 + 7e^2 - e^4 + 3e^8}{2 + 4e^2 - e^4 + 3e^8} \tau^2 - 1 \right) \gamma_{10} + (1 - \tau^2) \gamma_{10}^2 = 0. \quad (18)$$

Evaluating these equations for various values of  $\tau$  shows that the identified region  $\Gamma_1$  for  $\gamma_{10}$  is quite narrow unless  $\tau$  is very close to its upper bound of 1. In this design, the true value is  $\gamma_{10} = 1$ , which equals the identified region when  $\tau = 0$ . For  $\tau = 0.1$ , the identified interval based on Equation (18) is [0.995, 1.005], and for  $\tau = 0.5$ , the identified interval is [0.973, 1.023]. Even for the loose bound on  $\text{cov}(Z, \varepsilon_1\varepsilon_2)$  given by  $\tau = 0.9$ , the identified interval is still the rather narrow range [0.892, 1.084].

## 6. NONLINEAR MODEL 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

$$Y_1 = g_1(X) + Y_2\gamma_{10} + \varepsilon_1 \quad (19)$$

$$Y_2 = g_2(X) + Y_1\gamma_{20} + \varepsilon_2, \quad (20)$$

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  $E(Y | X)$  and  $E(Y Y' | X)$  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.*  $E(\varepsilon_1 | X) = 0$ ,  $E(\varepsilon_2 | X) = 0$ , and for some random vector  $Z$ ,  $\text{cov}(Z, \varepsilon_1\varepsilon_2) = 0$ .

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  $Z$ ),  $\text{cov}(Z, \varepsilon_1\varepsilon_2) = 0$  if  $Z$  is uncorrelated with  $U^2$ .

*Assumption B3.* Define  $W_j = Y_j - E(Y_j | X)$  for  $j = 1, 2$ . The matrix  $\Phi_W$ , defined as the matrix with columns given by the vectors  $\text{cov}(Z, W_1^2)$  and  $\text{cov}(Z, W_2^2)$ , has rank 2.

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  $\Phi_W$ , and more generally heteroscedasticity of  $W_1$  and  $W_2$  implies heteroscedasticity of  $\varepsilon$ .

*Theorem 4.* Let Equations (19) and (20) hold. If Assumptions B1 and B2 hold,  $\text{cov}(Z, \varepsilon_2^2) \neq 0$ , and  $\gamma_{20} = 0$ , then the structural parameter  $\gamma_{10}$ , the functions  $g_1(X)$  and  $g_2(X)$ , and the variance of  $\varepsilon$  are identified. If Assumptions B1, B2, B3, and A4 hold, then the structural parameters  $\gamma_{10}$  and  $\gamma_{20}$ , the functions  $g_1(X)$  and  $g_2(X)$ , and the variance of  $\varepsilon$  are identified.

An immediate corollary of Theorem 4 is that the partly linear simultaneous system

$$Y_1 = h_1(X_1) + X_2\beta_{10} + Y_2\gamma_{10} + \varepsilon_1 \quad (21)$$

$$Y_2 = h_2(X_1) + X_2\beta_{20} + Y_1\gamma_{20} + \varepsilon_2, \quad (22)$$

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

$$Y_1 = G_1(X, \beta_0) + Y_2\gamma_{10} + \varepsilon_1 \quad (23)$$

$$Y_2 = G_2(X, \beta_0) + Y_1\gamma_{20} + \varepsilon_2, \quad (24)$$

where the functions  $G_j(X, \beta_0)$  are known and the parameter vector  $\beta_0$ , which could include  $\gamma_1$  and  $\gamma_2$ , is unknown. This generalizes Equations (3) and (4) by allowing nonlinear functions of  $X$ . Letting  $g_j(X) = G_j(X, \beta_0)$ , Theorem 4 provides sufficient conditions for identification of this model, assuming that  $\beta_0$  is identified given identification of the functions  $g_j(X) = G_j(X, \beta_0)$ . The immediate analog to Corollary 3 is then that  $\beta_0$ ,  $\gamma_{10}$ ,  $\gamma_{20}$ , and  $\mu_0$  can be estimated from the moment conditions:

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

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

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

$$E[(Z - \mu_0)[Y_1 - G_1(X, \beta_0) - Y_2\gamma_{10}]$$

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

For efficient estimation in this case, where some of the moments are conditional, see, for example, 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, for example, 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 (19) and (20), where the functions  $g_j(X)$  are not parameterized. We now have identification based on the moments:

$$E[Y_1 - g_1(X) - Y_2\gamma_{10} | X] = 0$$

$$E[Y_2 - g_2(X) - Y_1\gamma_{20} | X] = 0$$

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

$$E[(Z - \mu_0)(Y_1 - g_1(X) - Y_2\gamma_{10})$$

$$\times (Y_2 - g_2(X) - Y_1\gamma_{20})] = 0.$$

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  $\hat{H}_j(X)$  be a uniformly consistent estimator of  $H_j(X) = E(Y_j | X)$ , for example, 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  $\hat{W}_{ji} = Y_{ji} - \hat{H}_j(X_i)$  for each observation  $i$ . Next, let  $\hat{C}_{jkh}$  be the sample covariance of  $\hat{W}_j\hat{W}_k$  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 (27), estimate  $\gamma_1$  and  $\gamma_2$  by

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

where  $\Gamma$  is a compact set satisfying Assumption A4. The above estimator for  $\gamma_1$  and  $\gamma_2$  is numerically equivalent to an ordinary nonlinear least squares regression over  $K$  observations of data, where  $K$  is the number of elements of  $Z$ . In a triangular system, that is, with  $\gamma_2 = 0$ , this step reduces to a linear regression for estimating  $\gamma_1$ . Finally, estimates of the functions  $g_1(X)$  and  $g_2(X)$  are obtained by nonparametrically regressing  $Y_1 - Y_2\hat{\gamma}_1$  and  $Y_2 - Y_1\hat{\gamma}_2$  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 4.

This estimator of  $\hat{\gamma}_1$  and  $\hat{\gamma}_2$  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  $\hat{\gamma}_1$  and  $\hat{\gamma}_2$ . However, assuming that  $\hat{\gamma}_1$  and  $\hat{\gamma}_2$  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  $Y_1 - Y_2\gamma_{10}$  and  $Y_2 - Y_1\gamma_{20}$  on  $X$ , respectively.

Further extension to estimation of the partly linear system of Equations (21) and (22) is immediate. For this model, the Assumption B2 moments  $E(\varepsilon_1 | X) = 0$ ,  $E(\varepsilon_2 | X) = 0$ , and  $\text{cov}(Z, \varepsilon_1\varepsilon_2) = 0$  are:

$$\begin{aligned} E[Y_1 - h_1(X_1) - X_2\beta_{10} - Y_2\gamma_{10} | X] &= 0 \\ E[Y_2 - h_2(X_1) - X_2\beta_{20} - Y_1\gamma_{20} | X] &= 0 \\ E(Z - \mu_0) &= 0 \\ E[(Z - \mu_0)[Y_1 - h_1(X_1) - X_2\beta_{10} - Y_2\gamma_{10}] \\ &\quad \times [Y_2 - h_2(X_1) - X_2\beta_{20} - Y_1\gamma_{20}]] = 0, \end{aligned}$$

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 article describes a new method of obtaining identification in mismeasured regressor models, triangular systems, simultaneous equation systems, and some partly linear semiparametric systems. The identification comes from observing a vector of variables  $Z$  (which can equal or be a subset of the vector of model regressors  $X$ ) that are uncorrelated with the covariance of heteroscedastic errors. The existence of such a  $Z$  is shown to be a feature of many models in which error correlations are due to an unobserved common factor, including mismeasured regressor models. Associated two-stage least squares and GMM estimators are provided.

The proposed estimators appear to work well in both a small Monte Carlo study (provided as a supplemental appendix to this article) and in an empirical application. Citing working paper versions of the present article, some articles by other researchers listed earlier include empirical applications of the proposed estimators and find them to work well in practice.

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 of 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  $\tilde{Z}$  satisfies the conditions to be an element of  $Z$ , but is only observed with classical measurement error, then this mismeasured  $\tilde{Z}$  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 Hansen's (1982) test, can be applied.

The identification here is based on higher moments and so is likely to give noisier, less reliable estimates than identification

based on standard exclusion restrictions, but may be useful in applications where traditional instruments are weak or nonexistent. This article's moments based on  $\text{cov}(Z, \varepsilon_1\varepsilon_2) = 0$  can be used along with traditional instruments to increase efficiency and provide testable overidentifying restrictions.

This article also shows that bounds on estimated parameters can be obtained when the identifying assumption  $\text{cov}(Z, \varepsilon_1\varepsilon_2) = 0$  does not hold, provided that this covariance is not too large relative to the heteroscedasticity in the errors. In a numerical example, these bounds appear to be quite narrow.

The identification scheme in the article requires the endogenous regressors to appear additively in the model. A good direction for future research would be searching for ways to extend the identification method to allow for including the endogenous regressors nonlinearly. Perhaps it would be possible to replace linearity in endogenous regressors with local linearity, applying this article's methods and assumptions to a kernel weighted locally linear representation of the model.

It would also be worth considering whether additional moments for identification could be obtained by allowing for more general dependence between  $Z$  and  $\varepsilon_2^2$  and corresponding zero higher moments. One simple example is to let the assumptions of Theorems 1 and 2 hold using  $\varpi(Z)$  in place of  $Z$  for different functions  $\varpi$ , such as higher moments of  $Z$ , thereby providing additional instruments for estimation.

## ACKNOWLEDGMENTS

I thank Roberto Rigobon, Frank Vella, Todd Prono, Susanne Schennach, Jerry Hausman, Raffaella Giacomini, Tiemen Woutersen, Christina Gathmann, Jim Heckman, and anonymous referees for helpful comments. Any errors are my own.

## APPENDIX

*Proof of Theorem 1.* Define  $W_j$  by Equation (11) for  $j = 1, 2$ . These  $W_j$  are identified by construction. Using the Assumptions, substituting Equations (6) and (7) 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  $\text{cov}(Z, \varepsilon_1\varepsilon_2) = 0$  is equivalent to  $\text{cov}[Z, (W_1 - \gamma_{10}W_2)W_2] = 0$ . Solving for  $\gamma_{10}$  shows that  $\gamma_{10}$  is identified by  $\gamma_{10} = \text{cov}(Z, W_1W_2)/\text{cov}(Z, W_2^2)$ . Given identification of  $\gamma_{10}$ , the coefficients  $\beta_{10}$  and  $\beta_{20}$  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  $E(X\varepsilon_j) = 0$ . Also,  $\varepsilon$  is 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 (8), observe that  $\Psi_{ZX}$  simplifies to:

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

which spans the same column space as

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

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

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

which then gives Equation (8).

*Proof of Theorem 2.* Substituting Equations (9) and (10) for  $Y_1$  and  $Y_2$  in the definitions of  $W_1$  and  $W_2$  shows that:

$$W_1 = \frac{\varepsilon_1 + \varepsilon_2\gamma_{10}}{1 - \gamma_{10}\gamma_{20}}, \quad W_2 = \frac{\varepsilon_2 + \varepsilon_1\gamma_{20}}{1 - \gamma_{10}\gamma_{20}}, \quad (\text{A.1})$$

and solving these equations for  $\varepsilon$  yields:

$$\varepsilon_1 = W_1 - \gamma_{10}W_2, \quad \varepsilon_2 = W_2 - \gamma_{20}W_1. \quad (\text{A.2})$$

Note that  $\gamma_{10}\gamma_{20} \neq 1$  by Assumption A4. Using Equation (A.3), the condition  $\text{cov}(Z, \varepsilon_1\varepsilon_2) = 0$  is equivalent to:

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

$$(1 + \gamma_{10}\gamma_{20})\text{cov}(Z, W_1W_2) - \gamma_{10}\text{cov}(Z, W_2^2) - \gamma_{20}\text{cov}(Z, W_1^2) = 0. \quad (\text{A.3})$$

Now  $1 + \gamma_{10}\gamma_{20} \neq 0$ , since otherwise, it would follow from Equation (A.3) that the rank of  $\Phi_W$  is less than 2. Define

$$\lambda_1 = \frac{\gamma_{10}}{1 + \gamma_{10}\gamma_{20}}, \quad \lambda_2 = \frac{\gamma_{20}}{1 + \gamma_{10}\gamma_{20}} \quad (\text{A.4})$$

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

$$\text{cov}(Z, W_1W_2) = \lambda_1\text{cov}(Z, W_2^2) + \lambda_2\text{cov}(Z, W_1^2) = \Phi_W\lambda, \quad (\text{A.5})$$

so  $\lambda$  is identified by:

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

and  $\Phi'_W\Phi_W$  is not singular because  $\Phi_W$  is rank 2. Solving Equation (28) for  $\gamma_{10}$  gives:

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

The above quadratic in  $\gamma_{10}$  has at most two roots, and for each root, the corresponding value for  $\gamma_{20}$  is given by  $\gamma_{20} = \gamma_{10}\lambda_2/\lambda_1$ . Let  $(\gamma_1^*, \gamma_2^*)$  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  $\lambda_1$  and  $\lambda_2$ . One of these solutions must be  $(\gamma_{10}, \gamma_{20})$ , and by Assumption A4, the other solution is not an element of  $\Gamma$ , so  $(\gamma_{10}, \gamma_{20})$  is identified. Note that the conditions required for the quadratic to have real rather than complex or imaginary roots are automatically satisfied, because  $(\gamma_{10}, \gamma_{20})$  is real.

Given identification of  $\gamma_{10}$  and  $\gamma_{20}$ , the coefficients  $\beta_{10}$  and  $\beta_{20}$  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,  $\varepsilon$  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 (A.1) in Theorem 2 was derived using only Assumptions A1 and A2. Evaluating  $\text{cov}(Z, W_j^2)$  using Equation (A.1) and the assumption that  $\text{cov}(Z, \varepsilon_1\varepsilon_2) = 0$  gives, for each element  $Z_k$  of  $Z$

$$\begin{pmatrix} \text{cov}(Z_k, W_1^2) \\ \text{cov}(Z_k, W_2^2) \end{pmatrix} = \left(\frac{1}{1 - \gamma_{10}\gamma_{20}}\right)^2 \begin{bmatrix} 1 & \gamma_{10}^2 \\ \gamma_{20}^2 & 1 \end{bmatrix} \begin{pmatrix} \text{cov}(Z_k, \varepsilon_1^2) \\ \text{cov}(Z_k, \varepsilon_2^2) \end{pmatrix}, \quad (\text{A.6})$$

so  $\Phi_W$  is rank 2 if and only if  $\Phi_\varepsilon$  is rank 2 and the matrix relating the two above is nonsingular, which requires  $|\gamma_{10}\gamma_{20}| \neq 1$ .

*Proof of Corollary 1.* Using Equation (A.2) and following the same steps as the proof of Theorem 2, the condition  $E(Z\varepsilon_1\varepsilon_2) = 0$  yields

$$E(ZW_1W_2) = \lambda_1E(ZW_2^2) + \lambda_2E(ZW_1^2) = \Phi_W\lambda$$

instead of Equation (A.5). This identifies  $\lambda$  and the rest of the proof is the same.

*Proof of Corollary 2.*  $\beta_{20}$  and  $\gamma_{20}$ , and hence  $\varepsilon_2$ , 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 (A.1),  $\text{cov}(Z, \varepsilon_1\varepsilon_2) = 0$  implies  $\text{cov}[Z, (W_1 - \gamma_{10}W_2)\varepsilon_2] = 0$ , which when solved for  $\gamma_{10}$  gives

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

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

*Proof of Corollaries 3 and 4.* By Equations (9) and (10),  $Q_1 = X\varepsilon_1$ ,  $Q_2 = X\varepsilon_2$  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  $\text{cov}(Z, \varepsilon_1\varepsilon_2) = 0$ . 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  $\theta = \theta_0$ .

*Proof of Theorem 3.* First observe that if  $\text{cov}(Z, \varepsilon_2^2) = 0$ , then this fact along with the other assumptions would imply that the conditions of Theorem 1 hold, giving point identification, which is a special case of the statement of Theorem 3. So, for the remainder of the proof, assume the case in which  $\text{cov}(Z, \varepsilon_2^2) \neq 0$ . Note this means also that  $\text{var}(\varepsilon_2^2) \neq 0$  and  $\text{var}(Z) \neq 0$ , because  $\text{var}(\varepsilon_2^2) = 0$  or  $\text{var}(Z) = 0$  would imply  $\text{cov}(Z, \varepsilon_2^2) = 0$ . These inequalities will ensure that the denominators in the fractions given below are nonzero.

By the definition of  $\tau$ :

$$\begin{aligned} [\text{corr}(\varepsilon_1\varepsilon_2, Z) / \text{corr}(W_2^2, Z)]^2 &\leq \tau^2 \\ \frac{[\text{cov}(\varepsilon_1\varepsilon_2, Z)]^2 \text{var}(W_2^2) \text{var}(Z)}{\text{var}(\varepsilon_1\varepsilon_2) \text{var}(Z) [\text{cov}(W_2^2, Z)]^2} &\leq \tau^2 \\ \frac{[\text{cov}(\varepsilon_1\varepsilon_2, Z)]^2}{[\text{cov}(W_2^2, Z)]^2} &\leq \frac{\text{var}(\varepsilon_1\varepsilon_2)}{\text{var}(W_2^2)} \tau^2. \end{aligned}$$

Now by Assumption A1 and Equation (11),  $\varepsilon_1 = W_1 - W_2\gamma_{10}$  and  $W_2 = \varepsilon_2$ , so

$$\begin{aligned} \frac{[\text{cov}(W_1 W_2 - W_2^2 \gamma_{10}, Z)]^2}{[\text{cov}(W_2^2, Z)]^2} &\leq \frac{\text{var}(W_1 W_2 - W_2^2 \gamma_{10})}{\text{var}(W_2^2)} \tau^2 \\ \frac{[\text{cov}(W_1 W_2, Z) - \text{cov}(W_2^2, Z) \gamma_{10}]^2}{[\text{cov}(W_2^2, Z)]^2} &\leq \frac{\text{var}(W_1 W_2) - 2\text{cov}(W_1 W_2, W_2^2 \gamma_{10}) + \text{var}(W_2^2 \gamma_{10})}{\text{var}(W_2^2)} \tau^2 \\ \frac{[\text{cov}(W_1 W_2, Z)]^2 - 2\text{cov}(W_1 W_2, Z) \text{cov}(W_2^2, Z) \gamma_{10} + [\text{cov}(W_2^2, Z)]^2 \gamma_{10}^2}{[\text{cov}(W_2^2, Z)]^2} &\leq \frac{\text{var}(W_1 W_2) - 2\text{cov}(W_1 W_2, W_2^2) \gamma_{10} + \text{var}(W_2^2) \gamma_{10}^2}{\text{var}(W_2^2)} \tau^2 \\ \left( \frac{\text{cov}(W_1 W_2, Z)}{\text{cov}(W_2^2, Z)} \right)^2 - 2 \frac{\text{cov}(W_1 W_2, Z)}{\text{cov}(W_2^2, Z)} \gamma_{10} + \gamma_{10}^2 &\leq \frac{\text{var}(W_1 W_2)}{\text{var}(W_2^2)} \tau^2 - 2 \frac{\text{cov}(W_1 W_2, W_2^2)}{\text{var}(W_2^2)} \tau^2 \gamma_{10} + \tau^2 \gamma_{10}^2, \end{aligned}$$

and moving all the terms to the left gives:

$$\begin{aligned} \frac{\text{cov}(W_1 W_2, Z)^2}{\text{cov}(W_2^2, Z)^2} - \frac{\text{var}(W_1 W_2)}{\text{var}(W_2^2)} \tau^2 + 2 \left[ \frac{\text{cov}(W_1 W_2, W_2^2)}{\text{var}(W_2^2)} \tau^2 \right. \\ \left. - \frac{\text{cov}(W_1 W_2, Z)}{\text{cov}(W_2^2, Z)} \right] \gamma_{10} + (1 - \tau^2) \gamma_{10}^2 \leq 0. \end{aligned}$$

For  $0 \leq \tau < 1$ , the coefficient of  $\gamma_{10}^2$  is positive, so this inequality holds for all  $\gamma_1$  that lie between the roots of the corresponding equality given by Equation (15).

*Proof of Theorem 4.* Like Theorem 1, substituting Equations (19) and (20) for  $Y_1$  and  $Y_2$  in the Assumption B3 definitions of  $W_1$  and  $W_2$  shows that Equations (A.1) and (A.2) hold in this model. Identification of  $\gamma_{10}$  and  $\gamma_{20}$  then follows exactly as in the Proof of Theorem 1. Given identification of  $\gamma_{10}$  and  $\gamma_{20}$ , the functions  $g_1(X)$  and  $g_2(X)$  are identified by  $g_1(X) = E(Y_1 | X) - E(Y_2 | X)\gamma_{10}$  and  $g_2(X) = E(Y_2 | X) - E(Y_1 | X)\gamma_{20}$ , both of which follow from  $E(\varepsilon_j | X) = 0$ . Finally,  $\varepsilon$  is now identified by  $\varepsilon_1 = Y_1 - g_1(X) - Y_2\gamma_{10}$  and  $\varepsilon_2 = Y_2 - g_2(X) - Y_1\gamma_{20}$ .

## SUPPLEMENTAL MATERIALS

**Appendix:** The supplemental online appendix to this paper contains a Monte Carlo analysis of the proposed estimators.

[Received April 2010. Accepted December 2010.]

## REFERENCES

- Ai, C., and Chen, X. (2003), "Efficient Estimation of Models With Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1844. [76]
- Banks, J., Blundell, R., and Lewbel, A. (1997), "Quadratic Engel Curves and Consumer Demand," *Review of Economics and Statistics*, 79, 527–539. [73]
- Breusch, T., and Pagan, A. (1979), "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica*, 47, 1287–1294. [71]
- Card, D. (1995), "Geographic Variations in College Proximity to Estimate the Returns to Schooling," in *Aspects of Labor Market Behavior: Essays in Honor of John Vanderkamp*, eds. L. N. Christofides, E. K. Grand, and R. Swidinsky, Toronto: University of Toronto Press, pp. 201–222. [70]
- Carneiro, P., and Heckman, J. (2002), "The Evidence on Credit Constraints in Post Secondary Schooling," *Economic Journal*, 112, 705–734. [70]
- Chamberlain, G. (1987), "Asymptotic Efficiency in Estimation With Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334. [76]
- Cragg, J. (1997), "Using Higher Moments to Estimate the Simple Errors-in-Variables Model," *Rand Journal of Economics*, 28, S71–S91. [69]
- Cragg, J., and Donald, S. (1997), "Inferring the Rank of a Matrix," *Journal of Econometrics*, 76, 223–250. [71]
- Dagenais, M. G., and Dagenais, D. L. (1997), "Higher Moment Estimators for Linear Regression Models with Errors in the Variables," *Journal of Econometrics*, 76, 193–222. [69]
- Emran, M. S., and Hou, Z. (2008), "Access to Markets and Household Consumption: Evidence from Rural China," unpublished manuscript, George Washington University. [69]
- Erickson, T., and Whited, T. (2002), "Two-Step GMM Estimation of the Errors-in-Variables Model Using High Order Moments," *Econometric Theory*, 18, 776–799. [69]
- Feenstra, B. (1994), "New Product Varieties and the Measurement of International Prices," *American Economic Review*, 84, 157–177. [70]
- Fuller, W. A. (1987), *Measurement Error Models*, New York: Wiley. [69]
- Giambona, E., and Schwiendbacher, A. (2007), "Debt Capacity of Tangible Assets: What Is Collateralizable in the Debt Market?" unpublished manuscript, University of Amsterdam. [69]
- Hansen, L. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054. [71]
- Hausman, J. (1983), "Specification and Estimation of Simultaneous Models," in *Handbook of Econometrics*, Vol. 1, eds. Z. Griliches and M. Intriligator, Amsterdam: North-Holland, pp. 391–448. [69]
- Hausman, J. A., Newey, W. K., and Powell, J. L. (1995), "Nonlinear Errors in Variables: Estimation of Some Engel Curves," *Journal of Econometrics*, 65, 205–253. [74]
- Heckman, J. (1974), "Shadow Prices, Market Wages, and Labor Supply," *Econometrica*, 42, 679–693. [69]
- (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161. [69]
- Heckman, J., and Vytalil, E. (1998), "Instrumental Variables Methods for the Correlated Random Coefficient Model," *Journal of Human Resources*, 33, 974–987. [70]
- Hogan, V., and Rigobon, R. (2003), "Using Unobserved Supply Shocks to Estimate the Returns to Educations," Unpublished manuscript. [69]
- Hsiao, C. (1983), "Identification," in *Handbook of Econometrics*, (vol. 1), eds. Z. Griliches and M. Intriligator, Amsterdam: North-Holland, pp. 223–283. [69]
- Joreskog, K. G., and Sorbom, D. (1984), *Advances in Factor Analysis and Structural Equation Models*, Lanham, MD: Rowman and Littlefield. [69]
- King, M., Sentana, E., and Wadhvani, S. (1994), "Volatility and Links Between National Stock Markets," *Econometrica*, 62, 901–933. [70]
- Kitamura, Y., Tripathi, G., and Ahn, H. (2003), "Empirical Likelihood Based Inference in Conditional Moment Restriction Models," unpublished manuscript. [76]
- Klein, R., and Vella, F. (2010), "Estimating a Class of Triangular Simultaneous Equations Models without Exclusion Restrictions," *Journal of Econometrics*, 154, 154–164. [69]
- Leamer, E. (1981), "Is It a Demand Curve or Is It a Supply Curve? Partial Identification Through Inequality Constraints," *Review of Economics and Statistics*, 63, 319–327. [70]
- Leser, C. E. V. (1963), "Forms of Engel Functions," *Econometrica*, 31, 694–703. [74]
- Lewbel, A. (1996), "Demand Estimation With Expenditure Measurement Errors on the Left and Right Hand Side," *Review of Economics and Statistics*, 78, 718–725. [74]
- (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. [74]
- Meghir, C., and Robin, J.-M. (1992), "Frequency of Purchase and the Estimation of Demand Systems," *Journal of Econometrics*, 53, 53–86. [73]
- Newey, W. (1993), "Efficient Estimation of Models With Conditional Moment Restrictions," in *Handbook of Statistics*, Vol. 11, eds. G. S. Maddala, C. R. Rao, and H.D. Vinod, Amsterdam: North Holland, pp. 419–454. [76]
- Newey, W. K., and McFadden, D. (1994), "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. R. F. Engle and D. L. McFadden, Amsterdam: Elsevier, pp. 2111–2245. [73]
- Newey, W., and Powell, J. (2003), "Instrumental Variables Estimation of Nonparametric Models," *Econometrica*, 71, 1557–1569. [76]
- Newey, W., and Smith, R. J. (2004), "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, 72, 219–255. [72]

- Otsu, T. (2003), "Penalized Empirical Likelihood Estimation of Conditional Moment Restriction Models With Unknown Functions," unpublished manuscript, University of Wisconsin. [76]
- Prono, T. (2008), "GARCH-Based Identification and Estimation of Triangular Systems," Working Paper No. QAU08-4, Federal Reserve Bank of Boston. [70]
- Rashad, I., and Markowitz, S. (2007), "Incentives in Obesity and Health Insurance," NBER Working Paper No. W13113, NBER. [69]
- Rigobon, R. (2002), "The Curse of Non Investment Grade Countries," *Journal of Development Economics*, 69, 423–449. [69]
- (2003), "Identification Through Heteroscedasticity," *Review of Economics and Statistics*, 85, 419–454. [69]
- Roehrig, C. S. (1988), "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica*, 56, 433–447. [69]
- Rummery, S., Vella, F., and Verbeek, M. (1999), "Estimating the Returns to Education for Australian Youth via Rank-Order Instrumental Variables," *Labour Economics*, 6, 491–507. [69]
- Sabia, J. J. (2007), "The Effect of Body Weight on Adolescent Academic Performance," *Southern Economic Journal*, 73, 871–900. [69]
- Sentana E. (1992), "Identification of Multivariate Conditionally Heteroscedastic Factor Models," FMG discussion paper number 139, LSE. [69]
- Sentana E., and Fiorentini, G. (2001), "Identification, Estimation, and Testing of Conditional Heteroscedastic Factor Models," *Journal of Econometrics*, 102, 143–164. [69]
- Staiger, D., and Stock, J. H. (1997), "Instrumental Variable Regression With Weak Instruments," *Econometrica*, 65, 557–586. [72]
- Stock, J. H., Wright, J. H., and Yogo, M. (2002), "A Survey of Weak Instruments and Weak Identification in Generalized Methods of Moments," *Journal of Business and Economic Statistics*, 20, 518–529. [72]
- Working, H. (1943), "Statistical Laws of Family Expenditure," *Journal of the American Statistical Association*, 38, 43–56. [74]
- Wright, P. (1928), *The Tariff on Animal and Vegetable Oils*, New York: Macmillan. [69]