Identification and Estimation Using Heteroscedasticity Without Instruments: The Binary Endogenous Regressor Case

Arthur Lewbel*

Boston College

Original December 2016, revised July 2017

Abstract

Lewbel (2012) provides an estimator for linear regression models containing an endogenous regressor, when no outside instruments or other such information is available. The method works by exploiting model heteroscedasticity to construct instruments using the available regressors. Some authors have considered the method in empirical applications where an endogenous regressor is binary (e.g., endogenous Diff-in-Diff or endogenous binary treatment models). The present paper shows that the assumptions required for Lewbel's estimator can indeed be satisfied when an endogenous regressor is binary. Caveats regarding application of the estimator are discussed.

JEP codes: C35, C36, C30, C13

Keywords: Simultaneous systems, linear regressions, endogeneity, identification, heteroscedasticity, binary regressors

*Corresponding address: Arthur Lewbel, Dept of Economics, Maloney Hall, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA, 02467, USA. (617)-552-3678, lewbel@bc.edu, https://www2.bc.edu/arthur-lewbel/

1 Introduction

Linear regression models containing endogenous regressors are generally identified using outside information such as exogenous instruments, or by parametric distribution assumptions. Some papers obtain identification without outside instruments by exploiting heteroscedasticity, including Rigobon (2003), Klein and Vella (2010), Lewbel (2012), and Prono (2014). See also Lewbel (2016).

Some authors, including include Emran, Robano, and Smith (2014) and Hoang, Pham, and Ulubaşoğlu (2014), have questioned whether the Lewbel (2012) estimator can be used when the endogenous regressor is binary. Others, including Le Moglie, Mencarini, and Rapallini (2015), have applied the Lewbel (2012) estimator with a binary endogenous variable, though without verifying if the assumptions hold.

Examples of such applications would include Diff-in-Diff models with endogenous fixed effects, or models with binary endogenous treatment indicators. Binary endogenous regressors are a natural case to consider in part because they imply that the instrument equation will automatically have heteroscedastic errors, which is one of the requirements of the estimator.

This paper shows validity of the Lewbel (2012) estimator when an endogenous regressor is binary. So, e.g., the estimator might be applied to estimate a (homogeneous) treatment effect when the binary treatment is not randomly assigned and when exogenous instruments are not available. However, the sufficient conditions given here do impose strong restrictions on the error term of the model.

2 The Model and Estimator

Assume a sample of observations of endogenous variables Y_1 and Y_2, and a vector of exogenous covariates X. We wish to estimate γ and the vector β in the model

$$
\begin{aligned}
Y_1 &= X'\beta + Y_2\gamma + \varepsilon_1 \\
Y_2 &= X'\alpha + \varepsilon_2
\end{aligned}
$$

where the errors ε_1 and ε_2 may be correlated. As in Lewbel (2012), we also consider the more general case where $Y_2 = g(X) + \varepsilon_2$ for some nonlinear, possibly unknown function g.

Standard instrumental variables estimation depends on having an element of X that appears in the Y_2 equation but not in the Y_1 equation, and uses that excluded regressor as an instrument for Y_2. The problem considered here is that perhaps no element of X is excluded from the Y_1 equation, or equivalently, we're not sure that any element of β is zero. Lewbel (2012) provides identification and a corresponding very simple linear two stage least squares estimator for β and γ, in this case where no element of X can be used as an instrument for Y_2. The method consists of constructing valid instruments for Y_2 by exploiting information contained in heteroscedasticity of ε_2.

The Lewbel (2012) estimator can be summarized as the following two steps.

1. Estimate $\widehat{\alpha}$ by an ordinary least squares regression of Y_2 on X, and obtain estimated residuals $\widehat{\varepsilon}_2 = Y_2 - X'\widehat{\alpha}$.

2. Let Z be some or all of the elements of X. Estimate β and γ 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.

This is implemented in the STATA module IVREG2H by Baum and Schaffer (2012).

In addition to the standard exogenous X assumptions that $E(X\varepsilon_1) = 0$, $E(X\varepsilon_2) = 0$, and $E(XX')$ is nonsingular, the key additional assumptions required for applying this estimator are that $Cov(Z, \varepsilon_1\varepsilon_2) = 0$ and $Cov(Z, \varepsilon_2^2) \neq 0$, where either $Z = X$ or Z is a subset of the elements of X. Lewbel (2012) shows that a variety of standard econometric models satisfy these assumptions. For example, the assumptions hold when the errors ε_1 and ε_2 satisfy the factor structure $\varepsilon_1 = cU + V_1$, and $\varepsilon_2 = U + V_2$ for some constant c, where U and V_1 are unobserved homoscedastic errors, V_2 is an unobserved heteroscedastic error, and U, V_1, and V_2 are mutually independent conditional on Z. Examples where these conditions can hold are when Y_2 is endogenous due to classical measurement error, or because of the presence of some underlying unobservable factor U that affects both Y_1 and Y_2 (e.g., U could be unobserved ability in a model where Y_2 is education and Y_1 is a labor market outcome).

Lewbel (2012) doesn't explicitly assume that Y_2 is continuous. However, that paper doesn't show that its identifying assumptions can be satisfied when Y_2 is not continuous. For example, if Y_2 was binary then U could not be independent of V_2 in the above factor structure example. This paper shows that the identifying assumptions can be satisfied when Y_2 is binary.

3 A Binary Endogenous Regressor

Suppose that Y_2 is binary. Then $Y_2 = X'\alpha + \varepsilon_2$ is a linear probability model. But we also wish to allow for more general models, so let $Y_2 = g(X) + \varepsilon_2$ where $g(X) = E(Y_2 \mid X)$. Here $g(X)$ is possibly nonlinear and possibly unknown. For example, if Y_2 satisfies a probit or logit model, then $g(X) = F\left(X'\alpha\right)$ where F is the cumulative normal or logistic distribution function. Also included are nonparametric models, where $g(X)$ is estimated by a nonparametric regression of Y_2 on X.

If the Y_2 equation is a linear probability model, then $\widehat{g}(X) = X'\widehat{\alpha}$ where $\widehat{\alpha}$ is obtained by ordinary least squares. Alternatively, the estimate of g could be $\widehat{g}(X) = F\left(X'\widehat{\alpha}\right)$ where $\widehat{\alpha}$ is obtained by a logit, probit, or other threshold crossing model estimator. Or $\widehat{g}(X)$ could be a nonparametric kernel or sieve regression of Y_2 on X. Whatever estimator is used for $\widehat{g}(X)$, step 1 of the estimator is then to construct the residuals $\widehat{\varepsilon}_2$ by $\widehat{\varepsilon}_2 = Y_2 - \widehat{g}(X)$, and step 2 of the estimator described above remains unchanged.

Maintain the usual linear model assumptions for the exogenous regressors X, that X is uncorrelated with ε_1 and ε_2, and that $E(XX')$ is nonsingular. If g is nonlinear then also assume consistency of $\widehat{g}$. We now show how the key additional assumptions required for the Lewbel (2012) estimator can be satisfied with Y_2 binary. For simplicity, the result is derived taking $Z = X$, which then implies that the restrictions can also hold when Z is any subset of X.

ASSUMPTION A1: Let $g(X) = E(Y_2 \mid X)$ and define $\varepsilon_2 = Y_2 - g(X)$. Assume the random variable $g(X)$ is bounded and that $Cov\left[X, g(X)(1 - g(X))\right] \neq 0$.

ASSUMPTION A2: Assume $Y_1 = X'\beta + Y_2\gamma + \varepsilon_1$ with $\varepsilon_1 = Y_2U + V$ for some unobserved random errors U and V, where the vector (U, V) is independent of Y_2, conditioning on X. Assume $E(U \mid X) = c(X) / (g(X)(1 - g(X)))$ and $E(V \mid X) = -c(X) / (1 - g(X))$, where $c(X)$ is any function such that $Cov(X, c(X)) = 0$.

Assumption A1 imposes minimal restrictions on Y_2 and X, and hence on the error ε_2. Boundedness of $g(X)$ is just for simplicity and could be relaxed. The covariance condition in Assumption A1 is testable, since it can be estimated as the sample covariance between X and $\widehat{g}(X)(1-\widehat{g}(X))$. Assumption A2 places strong distributional restrictions on ε_1, specifically, on the conditional means of the component latent errors U and V. It should be stressed that these are not necessary conditions. Rather, they're just one set of assumptions are shown to work.

The covariance condition in Assumption A2 will automatically hold if $c(X)$ is any constant. However, it's also easy to find functions of X that can work. For example if Z is any symmetrically distributed element of X that is independent of the other elements of X, then $c(X)$ could equal $(Z-E(Z))^k$ for any even integer k.

THEOREM 1: Let Assumptions A1 and A2 hold. Then $E(\varepsilon_1 \mid X)=0$, $E(\varepsilon_2 \mid X)=0$, $Cov(X,\varepsilon_1\varepsilon_2)=0$ and $Cov(X,\varepsilon_2^2)\neq 0$.

Using the same types of derivations as in the proof of Theorem 1, one can also readily verify that $E(\varepsilon_1 Y_2)=E(c(X))$ so Y_2 is indeed an endogenous regressor as long as $E(c(X))\neq 0$. In a supplemental online appendix, I show that the required conditions can also be satisfied if both Y_1 and Y_2 are binary.

PROOF of Theorem 1: Verifying each of the conditions in turn, we have

$$\begin{aligned} E(\varepsilon_1 \mid X) &= E(Y_2U+V \mid X)=g(X)E(U \mid X)+E(V \mid X) \\ &= g(X)\frac{c(X)}{g(X)(1-g(X))}+\frac{-c(X)}{(1-g(X))}=0. \end{aligned}$$

$$E(\varepsilon_2 \mid X)=E(Y_2-g(X) \mid X)=g(X)-g(X)=0.$$

$$\begin{aligned} E(\varepsilon_1\varepsilon_2 \mid X) &= E(Y_2U\varepsilon_2+V\varepsilon_2 \mid X)=E(Y_2U\varepsilon_2 \mid X)+E(V \mid X)E(\varepsilon_2 \mid X) \\ &= E(Y_2U(Y_2-g(X)) \mid X)=E(U(Y_2-Y_2g(X)) \mid X) \\ &= E(U \mid X)g(X)(1-g(X))=\frac{c(X)}{g(X)(1-g(X))}g(X)(1-g(X))=c(X) \end{aligned}$$

so

$$Cov(X,\varepsilon_1\varepsilon_2)=Cov(X,E(\varepsilon_1\varepsilon_2 \mid X))=Cov(X,c(X))=0.$$

Next

$$\begin{aligned} E\left(\varepsilon_2^2 \mid X\right) &= E\left((Y_2 - g(X))^2 \mid X\right) = E\left(\left(Y_2 - 2Y_2 g(X) + g(X)^2\right) \mid X\right) \\ &= g(X) - 2g(X)^2 + g(X)^2 = g(X)(1 - g(X)) \end{aligned}$$

so

$$Cov(X, \varepsilon_2^2) = Cov\left(X, E\left(\varepsilon_2^2 \mid X\right)\right) = Cov\left[X, g(X)(1 - g(X))\right] \neq 0.$$

4 Conclusions and Caveats

Theorem 1 shows that the assumptions required to apply the Lewbel (2012) can be satisfied when Y_2 is binary, and a supplemental appendix to this paper provides a different way to satisfy these assumptions when both Y_1 and Y_2 are binary. So, e.g., the STATA module IVREG2H by Baum and Schaffer (2012) can be used without change when just Y_2 is binary, or when both Y_1 and Y_2 are binary.

A drawback of these results is that there are no obvious behavioral models that directly imply Assumption A2. This is in sharp contrast to the case for continuous Y_2, where as Lewbel (2012) shows, the assumptions are satisfied when Y_2 suffers from standard, classical measurement error, or when Y_1 and Y_2 are correlated due to a standard single factor structure (e.g., where Y_1 is wage and Y_2 is education, which is endogenous due to scalar unobserved ability). There may well exist more plausible or better motivated alternative constructions to Theorem 1 that would also work with a binry Y_2. Searching for such alternatives would be a useful direction for future research. Similar constructions to Theorem 1 should also be possible when Y_2 is discrete with more support points, or is a censored variable.

Theorem 1 works regardless of the specification of the function $g(X)$, and so in particular can be used with the linear probability model that assumes $g(X) = X'\alpha$. See, however, Lewbel, Dong, and Yang (2012) for warnings regarding the linear probability model. When $g(X)$ is nonlinear (e.g., if it's given by a logit or probit model), then alternative estimators may exist. For example, in that case one might directly use $\widehat{g}(X)$ as an instrument for Y_2. See, e.g., Dong (2010) and Escanciano, Jacho-Chávez, and Lewbel (2016). If a binary Y_2 is endogenous specifically because of measurement error, then other possible estimators are given by Chen, Hu, and Lewbel (2008a, 2008b).

Identification based on constructed instruments as in Lewbel (1997), Erickson and Whited (2002), or Lewbel (2012) depend on strong modeling assumptions. It's generally preferable to instead use instruments that are excluded and exogenously determined based on randomization or economic theory. But in practice one is often not sure if a given variable is truly a valid instrument. In these cases, constructed instruments can be used to provide overidentifying information for model tests and for robustness checks. In particular, one might estimate the model using both outside instruments and constructed instruments, and then test jointly for validity of all the instruments, using e.g., a Sargan (1958) or Hansen (1982) J-test. If validity is rejected, then either the model is misspecified or at least one of these instruments is invalid. If validity is not rejected, it's still possible that the model is wrong or the instruments are invalid, but one would at least have increased confidence in both the outside instrument and the constructed instrument. Both might then be used in estimation to maximize efficiency.

One could also estimate the model separately using outside instruments and constructed instruments. If the estimates are similar across these different sets of identifying assumptions, then that provides support for the model and evidence that the results are not just artifacts of one particular set of identifying assumptions. More generally, identification based on functional form or constructed instruments is preferably not used in isolation, but rather is ideally employed in conjunction with other means of obtaining identification, both as a way to check robustness of results to alternative identifying assumptions and to increase efficiency of estimation.

References

[1] Baum, C. F. and M. E. Schaffer, (2012) "IVREG2H: Stata module to perform instrumental variables estimation using heteroskedasticity-based instruments," Statistical Software Components S457555, Boston College Department of Economics.

[2] Chen, X., Y. Hu, and A. Lewbel (2008a), “A Note on the Closed-form Identification of Regression Models with a Mismeasured Binary Regressor,” Statistics and Probability Letters, 78, pp. 1473-1479.

[3] Chen, X., Y. Hu, and A. Lewbel (2008b), “Nonparametric Identification of Regression Models Containing a Misclassified Dichotomous Regressor Without Instruments,” Economics Letters, 100, pp. 381-384.

[4] Dong, Y. (2010), "Endogenous Regressor Binary Choice Models Without Instruments, With an Application to Migration," Economics Letters, 107(1), 33-35.

[5] Emran, M. S., V. Robano, and S. C. Smith, (2014) "Assessing the Frontiers of Ultrapoverty Reduction: Evidence from Challenging the Frontiers of Poverty Reduction/Targeting the Ultra-poor, an Innovative Program in Bangladesh," Economic Development and Cultural Change, 62(2), pp. 339-380.

[6] Erickson, T. and T. M. Whited (2002), "Two-step GMM estimation of the errors-in-variables model using high-order moments" Econometric Theory, 18(3), pp. 776-799.

[7] Escanciano, J-C, D. Jacho-Chávez, and A. Lewbel (2016), “Identification and Estimation of Semi-parametric Two Step Models,” Quantitative Economics, 7, pp. 561-589.

[8] Hansen, L. P. (1982). "Large Sample Properties of Generalized Method of Moments Estimators". Econometrica 50(4), pp. 1029–1054.

[9] Hoang, T. X., C. S. Pham, and M. A. Ulubaşoğlu, (2014) "Non-Farm Activity, Household Expenditure, and Poverty Reduction in Rural Vietnam: 2002–2008," World Development, 64, pp. 554–568.

[10] Klein, R. and F. Vella (2010), "Estimating a class of triangular simultaneous equations models without exclusion restrictions," Journal of Econometrics 154(2), pp. 154–164.

[11] Le Moglie, M., L. Mencarini, and C. Rapallini, (2015), "Is it just a matter of personality? On the role of subjective well-being in childbearing behavior," Journal of Economic Behavior and Organization, 117, pp. 453–475.

[12] 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.

[13] Lewbel, A. (2012), “Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models.” Journal of Business and Economic Statistics, 30, 67-80.

[14] Lewbel, A. (2016), “The Identification Zoo - Meanings of Identification in Econometrics," Unpublished Mauscript, Boston College.

[15] Lewbel, A., Y. Dong, and T. T. Yang (2012), “Viewpoint: Comparing Features of Convenient Estimators for Binary Choice Models With Endogenous Regressors,” Canadian Journal of Economics, 45, 809-829.

[16] Prono, T. (2014), "The Role of Conditional Heteroskedasticity in Identifying and Estimating Linear Triangular Systems, with Applications to Asset Pricing Models that Include a Mismeasured Factor," Journal of Applied Econometrics, 29(5), pp. 800-824.

[17] Rigobon, R. (2003), ”Identification Through Heteroskedasticity,” Review of Economics and Statistics, 85(4), pp. 777-792.

[18] Sargan, J. D. (1958), "The Estimation of Economic Relationships Using Instrumental Variables". Econometrica 26(3), pp. 393–415.

SUPPLEMENTAL APPENDIX to: Identification and Estimation Using Heteroscedasticity Without Instruments: The Binary Endogenous Regressor Case

Arthur Lewbel
Boston College

July 2017

This supplemental Appendix provides a companion result to Theorem 1 in the main text, for when both Y_1 and Y_2 are binary.

Theorem 1 establishes that the conditions required to apply the Lewbel (2012) estimator can be satisfied with Y_2 binary. But can the conditions also hold when Y_1 is binary? The construction used in Theorem 1 will not generally hold in that case, because if Y_1 is binary, then V must always equal either $-\left(X'\beta + Y_2\left(\gamma + U\right)\right)$ or $1 - \left(X'\beta + Y_2\left(\gamma + U\right)\right)$, which generally means V will be correlated with Y_2, violating Assumption A2.

Theorem 2 below shows that it is still possible to satisfy the conditions needed to apply Lewbel (2012) when both Y_1 and Y_2 are binary, but we require a different construction from that used in Theorem 1. Still, in this case the estimation of $\widehat{g}\left(X\right)$ and therefore of the residuals $\widehat{\varepsilon}_2$ remains exactly the same as in Theorem 1.

ASSUMPTION A3: Assume Y_1 and Y_2 are binary. Let $p_{ab}\left(X\right) = \Pr\left(Y_1 = a, Y_2 = b \mid X\right)$ where a and b each equal zero or one.

Note that the functions $p_{11}\left(X\right)$, $p_{01}\left(X\right)$, and $p_{10}\left(X\right)$ in Assumption A3 completely define the joint distribution of Y_1 and Y_2 conditional on X. The fourth probability $p_{00}\left(X\right)$ is determined from the other three by $p_{00}\left(X\right) = 1 - p_{11}\left(X\right) - p_{10}\left(X\right) - p_{01}\left(X\right)$.

ASSUMPTION A4: For some vector β, scalar γ, and function $g(X)$, define $\varepsilon_2 = Y_2 - g(X)$ and $\varepsilon_1 = Y_1 - X'\beta - Y_2\gamma$. Assume 1. $Cov\left[X, g(X)(1-g(X))\right] \neq 0$, 2. $p_{11}(X) + p_{01}(X) - g(X) = 0$, 3. $p_{11}(X) + p_{10}(X) - X'\beta - g(X)\gamma = 0$, and 4. $Cov(X, c(X)) = 0$ where the function $c(X)$ is defined by $c(X) = p_{11}(X) - \left(X'\beta + g(X)\gamma\right) g(X)$.

THEOREM 2: Let Assumptions A3 and A4 hold. Then $E(\varepsilon_1 \mid X) = 0$, $E(\varepsilon_2 \mid X) = 0$, $Cov(X, \varepsilon_1\varepsilon_2) = 0$ and $Cov(X, \varepsilon_2^2) \neq 0$.

To see how Theorem 2 shows that the necessary assumptions can be satisfied, suppose we start with some chosen function $g(X)$, a vector β, and a scalar γ. By the definition of these functions and parameters, $g(X)$ and $X'\beta + g(X)\gamma$ must lie between zero and one, so their product must also lie in that range. Then let the probability $p_{11}(X)$ equal this product plus a small constant, or plus a small function $c(X)$ that satisfies $Cov(X, c(X))$. We can then let $p_{01}(X)$ be given by $p_{01}(X) = g(X) - p_{11}(X)$ and let $p_{10}(X)$ be given by $p_{10}(X) = X'\beta + g(X)\gamma - p_{11}(X)$. The only constraint here is that the range of values that $g(X)$ and $X'\beta + g(X)\gamma$ can take on must be sufficiently small to allow the functions $p_{11}(X)$, $p_{01}(X)$, and $p_{10}(X)$ defined in this way to be nonnegative and sum to less than one. It then follows that Y_1 and Y_2 are drawn from the conditional distribution given by $p_{11}(X)$, $p_{01}(X)$, and $p_{10}(X)$, then they will satisfy the required assumptions for the given $g(X)$, β, and γ.

PROOF of Theorem 2: As in the proof of Theorem 1, considering each of the required conditions in turn, we have

$$E(Y_2 \mid X) = \Pr(Y_2 = 1 \mid X) = p_{01}(X) + p_{11}(X) = g(X)$$

where the last equality is by A4.2. It then follows that $E(\varepsilon_2 \mid X) = E(Y_2 - g(X) \mid X) = 0$.

Next,

$$E(Y_1 \mid X) = \Pr(Y_1 = 1 \mid X) = p_{10}(X) + p_{00}(X) = X'\beta + g(X)\gamma$$

where the last equality is by A4.3. Therefore

$$
\begin{aligned}
E\left(\varepsilon_1 \mid X\right) &= E\left(Y_1 - X'\beta - Y_2\gamma \mid X\right) = \\
&= E\left(Y_1 \mid X\right) - X'\beta - E\left(Y_2 \mid X\right)\gamma \\
&= X'\beta + g\left(X\right)\gamma - X'\beta - g\left(X\right)\gamma = 0
\end{aligned}
$$

Turning to the $Cov(X, \varepsilon_1\varepsilon_2)$ term, first observe

$$
\begin{aligned}
E\left(\varepsilon_1\varepsilon_2 \mid X\right) &= E\left(Y_1\varepsilon_2 \mid X\right) = E\left(Y_1Y_2 - Y_1 g\left(X\right) \mid X\right) \\
&= E\left(Y_1Y_2 \mid X\right) - E\left(Y_1 \mid X\right) g\left(X\right) \\
&= \Pr\left(Y_1 = 1, Y_2 = 1 \mid X\right) - \left(X'\beta + g\left(X\right)\gamma\right) g\left(X\right) \\
&= p_{11}\left(X\right) - \left(X'\beta + g\left(X\right)\gamma\right) g\left(X\right) = c\left(X\right)
\end{aligned}
$$

so

$$
Cov(X, \varepsilon_1\varepsilon_2) = Cov\left(X, E\left(\varepsilon_1\varepsilon_2 \mid X\right)\right) = Cov\left(X, c\left(X\right)\right) = 0.
$$

Finally, we have $Cov(X, \varepsilon_2^2) \neq 0$ by the same derivation as in Theorem 1.