

REVIEW

Open Access

A cursory review of the identification strategies



Fabio Gaetano Santeramo

Correspondence:
fabio.gaetano.santeramo@gmail.com
University of Foggia, Via Napoli 27,
Foggia, Italy

Abstract

The article provides a literature review on the topic of identification of supply and demand. In particular, it discusses the identification problem, that is the issue of having to solve for unique values of the parameters of the structural model from the values of the parameters of the reduced form of the model. We summarize several methodologies employed in the literature to solve this problem and gives practical examples. These solutions include, but are not limited to, using instrumental variables, adopting a recursive structure, holding demand constant, and imposing inequality constraints in order to restrict the domain of estimates. We also discuss on two major recent contributions in agricultural economics. The review will guide researchers in selecting the most suited approach to identify demand and supply.

Keywords: Econometrics; Identification; Supply; Demand

Introduction

Identification is a main issue in econometrics, the branch of economics that aims to answer to empirical questions based on economic models. Econometrics models are always based on assumptions, not always testable or falsifiable. In this framework, identification deals with the relationship between the assumptions of an econometric model and the possibility of answering or not, an empirical question using that model.

In applied economics the identification problem is a major challenge in many situations. An emblematic is the estimation of the supply and demand equations. Faced for the first time back in the late '20s (Wright 1928), the identification problem is still very modern and debated (Roberts and Schlenker 2013). A century of research on this topic has lead scholars to propose a myriad of approaches including the use of instrumental variables, the adoption of recursive structures, or the imposition of inequality constraints.

The present note aims at reviewing the status of art of identification in applied economics with particular emphasis to agricultural economics. The remainder of the note is as follows: section two summarizes the identification problem providing several definitions, the subsequent paragraph reviews the solutions that have been proposed in a century of research, and finally we conclude with final remarks¹.

On the identification problem

The area of identification studies the necessary and sufficient conditions to estimate (consistently) parameters of interest². From a different perspective, the identification problem in econometrics is the issue of having to solve for unique values of the parameters of the

structural model from the values of the parameters of the reduced form of the model (i.e. a single estimate of the structural parameters from the reduced form parameters for each structural equation, *cfr.* Maddala 1992)³. Therefore, if there are multiple solutions which make the reduced form coefficients compatible with the structural coefficients, the model is underidentified. Instead if there are no compatible solutions, the model is said to be overidentified. Finally, if a solution exists and is unique, the model is said to be just identified or exactly identified⁴.

All in all, the identification problem can be viewed as the (unresolved) dilemma of economists to make (correct) inference by reducing at most the number and strength of (necessary) assumptions. A major criticism related to this puzzle is the well known *Law of Decreasing Credibility* (Manski 2003) which states that “the credibility of inference decreases with the strength of the assumptions maintained”. Let us provide a practical example of the identification problem: the estimation of a system of demand and supply equations (Koopmans 1949).

Consider a linear model for the supply and demand: the former will be upward sloping with respect to price and the latter is expected to be downward sloping. We observe data on both the price (P) and the traded quantity (Q) of this good for several years. Unfortunately this information does not suffice to identify both demand and supply by using regression analysis on observations of Q and P. In fact it is impossible to estimate a downward slope and an upward slope with one linear regression line involving only two variables. Indeed, additional variables solve this issue and help to identify the individual relations. Put differently, by observing shifts in the demand (supply) curve, due to an exogenous variable, it is possible to identify the positive (negative) slope of the supply (demand) equation.

For instance, while we need demand shifters to estimate the slope of the supply, we need supply shifters to estimate the slope of the demand. More generally, we are able to identify the parameters of the equation (in our case the supply) not affected by the exogenous variable (Z). In order to identify both the supply and the demand equation, we would need both a variable (or *shifter*) Z entering the demand equation but not the supply equation (e.g. in agricultural economics it is common to use weather variables), and X entering the supply equation but not the demand equation (e.g. in agricultural economics a common approach is to introduce income as demand shifter). In other words, we need to introduce Z, a demand shifter (e.g. income) and X, a supply shifter (e.g. weather variable):

$$\text{demand: } Q = a_1 + b_1P + c_1Z \ ; \ \text{supply: } Q = a_2 + b_2P + c_2X$$

with positive b_2 and negative b_1 . Here both equations are identified if c_1 and c_2 are nonzero. Solving for P and Q we obtain the reduced-form equations:

$$Q = \frac{a_1b_2 - a_2b_1}{b_2 - b_1} + c_1 \frac{b_2}{b_2 - b_1} Z - c_2 \frac{b_1}{b_2 - b_1} X$$

$$P = \frac{a_1 - a_2}{b_2 - b_1} + \frac{c_1}{b_2 - b_1} Z - \frac{c_2}{b_2 - b_1} X$$

where $\pi_1 = \frac{a_1b_2 - a_2b_1}{b_2 - b_1}$, etc. are the reduced-form parameters. Suppose we observe Z, but not X. In this case we have two estimates for b_2 , and a_2 : the supply is said over-identified; the demand is under-identified. When we have unique estimates for the structural parameters, the equations are said exactly identified; multiple estimates imply over-identification; no estimates imply under-identification.

Proposed solutions: a cursory review

As previously described, the identification problem arises when we try to identify parameters using a reduced form. In the example of supply and demand, we may solve the problem by using an instrumental variable. Few points need to be recalled. More precisely, an instrument will be valid if the variable is correlated with the endogenous regressor and uncorrelated with the regression error.

Maddala (1977) pointed it is very difficult to have such kind of a variable, and econometrics textbooks do not provide clear guidelines. Angrist and Krueger (2001), p. 73, argue that “good instruments often come from detailed knowledge of the economic mechanism and institutions determining the regressor of interest”. For example, a valid instrument shifts only one “curve” (e.g. supply, but not demand). In agricultural markets, the instrument may be rainfall or weather shocks.

Wright (1928) has pioneered the use of instrumental variables. He estimated they supply and demand for flaxseed and used prices of substituted goods as instrumental variables for demand, and yield per acre as instruments for supply. He averaged out the estimates obtained using different instruments. Current researches have shown that a more efficient way to rely on multiple instruments is to use a two-stage least squares (2SLS) procedure. The method is described below.

First we provide a chronological review of the solutions have been proposed to solve the identification problem.

A simple, probably too naive, solution is to ignore the problem. Indeed this solution is not lacking of a theoretical justification. As Wright (1929) pointed in JASA, ignoring the issue is a valid solution if “it may be assumed that the dynamic forces will continue to operate thereafter in the same manner as they have been operating during that period”.

Another solution is to adopt a recursive structure:

$$(i) p_t = \beta_1 q_t + u^D ; (ii) q_t = \beta_2 p_{t-1} + u^S.$$

In this formulation p_{t-1} is exogenous in the supply equation, u^S is uncorrelated with u^D (therefore there are no common shifters), and q_t is exogenous in the demand equation with p_t on the left hand side.

Frisch and Waugh (1933) have proposed another approach. They suggested to hold demand constant. Given that the observed quantity demanded differs from the true (or latent) demand, the approach consists of estimating the observed demand and correcting for the bias. We clarify with an example. Suppose that quantity is measured with error ϵ_t , that is:

$$\text{(true demand) } q_t^* = \beta p_t + \gamma W_t \quad \text{(observed quantity) } q_t = q_t^* + \epsilon_t$$

where W_t represents all determinants of demand and ϵ_t is pure independent measurement error. Solving for observed demand:

$$\text{(observed demand) } q_t = \beta p_t + \gamma W_t + \epsilon_t$$

where $E(p_t \epsilon_t) = 0$. The approach suggested by Frisch and Waugh (1933) is to adjust for the bias, given the “known” γ and W_t . In this case, as they prove, OLS estimates are consistent.

Another approach is to use an instrumental variables (IV) regression. In the case of a single equation, the Limited Information Maximum Likelihood method (LIML) is a valid

alternative. The method has been proposed by Anderson and Rubin (1949), and has been popular until the introduction of the 2SLS by Theil (1965)⁵. The LIML consists in minimizing the residual sum of squares (RSS) to select the regressors. More precisely, the LIML minimize the ratio of RSS under the restricted and unrestricted model (Maddala 1992). The analogy with the 2SLS is very strong in that the latter minimize the difference of RSS under the restricted and unrestricted model. As a consequence, if the model is exactly identified the 2SLS and LIML provide identical estimates. Sargan (1958) has extended the IV approach to multiple instruments through the 2SLS method.

In a nutshell, the approach is as follows. In the first stage, each explanatory variable that is an endogenous covariate in the equation of interest is regressed on all of the exogenous variables in the model (including both exogenous covariates in the equation of interest and the excluded instruments). This first stage allows us to obtain the predicted values. In the second stage, the regression of interest is estimated as usual, except that in this stage each endogenous covariate is replaced with the predicted values from the first stage (Wooldridge 2010).

Empirically, the 2SLS is performed as follows. Let y be the dependent variable, x_1, \dots, x_{k-1} the explanatory variables, x_k the endogenous regressor, z_1, \dots, z_M the set of instruments.

(I) First stage: compute \hat{x}_k regressing x_k on regressors and instruments.

$$x_k = \alpha + \sum_{i=1}^{k-1} \beta_i x_i + \sum_{j=1}^N \gamma_j z_j \quad (1)$$

(II) Second stage: estimate the model replacing x_k with \hat{x}_k .

$$y = \alpha + \sum_{i=1}^{k-1} \beta_i x_i + \beta_k \hat{x}_k \quad (2)$$

From an empirical point of view, it is worth recalling the pitfalls of instrumental variables approach. The 2SLS provides consistent, but not unbiased estimates, therefore researchers that use this approach should always aspire to use large datasets. Moreover, an instrumental variable correlated with omitted variables can lead to biased estimates that is much greater than the bias in ordinary least squares estimates. However, the bias is proportional to the degree of overidentification, hence using fewer instruments would reduce the bias. Moreover, it is wise to test for the validity of instruments. Many tests have been proposed and some are implement in common packages (see Berkowitz et al. 2012)⁶.

For the above mentioned approaches we have implicitly assumed to deal only with a single equation. Special attention needs the case in which we consider a simultaneous equation model. An efficient way to estimate a full system of equations is to use Generalized Method of Moments (GMM) estimation. Unfortunately, GMM is usually unfeasible, unless the system covariance matrix (Σ) is known. Alternative approaches consist in estimating the system by using a three stage least squares (3SLS) procedure, or by adopting a full information maximum likelihood (FIML) estimator. The former consists in estimating a 2SLS (or equation-by-equation) and then using the residuals to compute Σ . Using $\hat{\Sigma}$ the estimation of the third stage will be consistent. Alternatively a FIML estimator can be adopted. The estimator uses information about all the equations in the system, providing consistent estimates. Although asymptotically equivalent, the FIML is not equal to the

continuously updated 3SLS estimator (unless the system is just-identified). Empirically, the 3SLS estimator is much easier to be computed than the FIML estimator (Davidson and MacKinnon 2004).

Alternative approaches have been proposed. Leamer (1981) has suggested to solve the identification problem by imposing inequality constraints in order to restrict the domain of estimates. His words are self-explanatory: “when the regression of quantity on price yields a positive estimate, we may assume that this is an attenuated estimate of the supply curve and that the data contain no useful information about the demand curve.

If the estimate is negative, the number may be treated as an attenuated estimate of the demand slope, and we may assert that the data contain no useful information about the supply curve” (Leamer 1981), p. 321. Thurman and Wohlgenant (1989) provide an empirical application of Leamer’s method in agricultural markets for the estimation of demand, whereas Renuka and Kalirajan (2002) applied the method to the demand for services. More recently, Garnache and Mérel (2015) use a mathematical programming framework, and a set of constraints to identify crop supply elasticities.

Rigobon (2003) exploits the intuition in Wright (1928) suggesting to restrict the parametric space using the information provided by the heteroskedasticity in the data (e.g. due to crises, policy shifts, changes in collecting the data, etc.). He provides necessary and sufficient conditions for identification of a system of simultaneous equations. In particular, Rigobon suggests to use the second moments to increase the number of relations between the parameters in the reduced and structural forms. An appealing feature of his approach is that it only requires the existence of heteroskedasticity in that the direct modeling of the source of heteroskedasticity can be ignored for the identification purpose. The approach is as follows. First, Rigobon (2003) estimated a vector autoregressive model of interest rates (prices may be used for agricultural markets); second, he defined subsamples according to different volatility; finally he computed the covariances matrices that have been used in the GMM estimation of contemporaneous shocks. Although the intuition to use the variance of the shocks to reduce the bias in OLS estimates has to be attributed to Wright (1928), Rigobon (2003) generalized the intuition and provided the conditions to identify the system⁷.

Roberts and Schlenker (2013) have revisited the problem of identification of supply and demand for agricultural commodities. The authors use theory of storage to derive the following empirical model:

$$\text{(Supply)} \quad \log(s_t) = \alpha_s + \beta_s \log(E[p_t | \Omega_{t-1}]) + \gamma_s w_t + f(t) + u_t \quad (3)$$

$$\text{(Demand)} \quad \log(c_t) = \alpha_d + \beta_d \log(p_t) + g(t) + v_t \quad (4)$$

and $c_t = s_t - z_t$ (consumption, c_t , is the difference of supply, s_t , and storage, z_t), α_s and α_d are intercepts for supply and demand, the Ω is the information set, w_t stands for the random weather-induced yield shocks, $f(t)$ and $g(t)$ capture time trends in supply and demand, u_t and v_t are the error terms. The rationale for (24) and (25) is that weather-induced shocks (current and lagged) are expected to shift only the supply curve, and to leave the demand unchanged. The model is solved in two stages. The first stage consists

in estimating $\log(p_t)$ and $\log(E[p_t|\Omega_{t-1}])$. The authors suggest to use a distributed lag model of yield shocks and a polynomial time trend. The reduced forms are as follows:

$$\log(p_t) = \pi_{d0} + \sum_{k=1}^{K-1} \mu_d w_{t-k} + f(t) + \epsilon_{dt} \quad (5)$$

$$\log(E[p_t|\Omega_{t-1}]) = \pi_{s0} + \sum_{k=1}^K \mu_s w_{t-k} + f(t) + \epsilon_{st} \quad (6)$$

where $f(t)$ and $g(t)$ represent the polynomial time trend functions, ϵ_{dt} and ϵ_{st} are the error terms. In the second stage the lagged yield shocks are used as instruments. In particular the supply is estimated as follows:

$$\log(s_t) = \alpha_s + \beta_s \log(E[\widehat{p_t}|\Omega_{t-1}]) + \lambda_{s0} \gamma_{s0} w_t + f(t) + u_t \quad (7)$$

and demand is obtained as follows:

$$\text{(Demand)} \log(c_t) = \alpha_d + \beta_d \widehat{\log(p_t)} + g(t) + v_t \quad (8)$$

The novelty of this approach is that Roberts and Schlenker (2013) have considered simultaneously four commodities that are substitutes in supply and demand, and have instrumented supply by using weather shocks⁸.

More recently, Steinwender (2014) has proposed a novel approach to identify the demand equation. Starting from a simple two markets model, and allowing for trade and storage, the identification problem may occur if the unobserved demand shocks are positively correlated with change in stock and exports. Put differently, because quantities and prices are determined contemporaneously, we need a valid instrument to estimate them correctly. Steinwender (2014) proposed to use the fact that exports (which take k periods to reach destination) are predetermined at destination as instruments to identify demand shocks. The demand equation takes the following form:

$$\text{(Demand)} p_{t+k} = \alpha_d + \beta_d (x_t - \Delta s_{t+k}) + v_{t+k} \quad (9)$$

where the price p_{t+k} is function of exports (x_t) that reach location at time k , and the change in stock (Δs_{t+k}) at time $t+k$. The approach is interesting in that it does not require other data than exports, stock changes, and prices. A drawback is that stock data are usually not available at the same time frequency as trade and price data: price and trade data are usually at monthly, weekly, and also daily frequency, whereas stock data are rarely available at such a high frequency.

Conclusions

Estimating supply and demand equations is a challenge that has puzzled applied economists for decades. The old-dated identification problem is still very actual and debated. We have provided a cursory review of the approaches that have been adopted since the early '20s till third millennium.

As it has emerged by our note, several approaches are feasible, none exempt from limitations. Applied economists should prefer feasible and clear approaches. In this perspective it is very likely that the use of instrumental variables will play a major role in the next decades: its long tradition on one hand, and the requirement of available data such as weather variables and stock data, play in favor of this approach. While from a practical point of view it seems advisable to solve the identification issue starting from a careful

evaluation of available data, the novel approach proposed by Roberts and Schlenker (2013) and Steinwender (2014) prove that innovative solutions are still emerging.

In order to explore novel ways to identify demand and supply in agricultural economics, researchers should carefully evaluate the assumptions that tie these fundamentals together. Exploring the role of expectations, unpredictable events, biological and physical constraints is important to disentangle the different drivers of demand and supply.

The economists toolkit includes a large variety of alternative techniques and that, indeed, deepening on feasible techniques to solve the identification problem represents a promising area of research.

Endnotes

¹The interested reader may refer to an earlier version of the present paper (Santeramo 2014) for further details.

²A formal, and simple, definition is provided by A. M. Shaikh. See <http://home.uchicago.edu/~amshaikh/webfiles/ident.pdf>. We provide a shorter version.

Let P denote the true distribution of the observed data X . Denote by $\mathbf{P} = \{P_\theta : \theta \in \Theta\}$ a model for the distribution of the observed data, assuming correct specification, that is $P \in \mathbf{P}$. We know that $\theta \in \Theta_0(P)$, where $\Theta_0(P) = \{\theta \in \Theta : P_\theta = P\}$ where $\Theta_0(P)$ is the identified set and θ is identified if the identified set is a singleton for all $P \in \mathbf{P}$.

³The reduced form of a model is the one in which the endogenous variables are expressed as functions of the exogenous variables.

⁴In general, a linear system of M equations, with $M > 1$, cannot be identified from the data if less than $M - 1$ variables are excluded from that equation. This is a particular form of the order condition for identification (the exclusion criterion), which is necessary but not sufficient for identification. The rank condition is a necessary and sufficient condition for identification. In the case of only exclusion restrictions, it must "be possible to form at least one nonvanishing determinant of order $M - 1$ from the columns of A corresponding to the variables excluded a priori from that equation," where A is the matrix of coefficients of the equations (Fisher 1966), p. 40. We are grateful to the referee for his suggestion.

⁵Interestingly, IV are still applied in empirical papers (e.g. Hendricks et al. 2015)

⁶The FAR test, recently developed, does not overreject the null hypothesis when we use half of the sample without replacement. The test is implemented in STATA.

⁷See Okumura (2011), and Lütkepohl and Netsunajev (2014) for recent applications.

⁸Their approach has been already applied in several papers (Auffhammer et al. 2013; Haile et al. 2015)

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The author gratefully acknowledges financial support from the International Food Policy Research Institute, AGRODEP Project. An earlier version of this paper circulated as IFPRI report under the title "On the Estimation of Supply and Demand Elasticities of Agricultural Commodities." Any opinions expressed are those of the author and not those of the International Food Policy Research Institute.

Received: 31 December 2014 Accepted: 16 September 2015

Published online: 28 September 2015

References

- Anderson TW, Rubin H (1949) Estimation of the parameters of a single equation in a complete system of stochastic equations. *Ann Math Stat* 20(1):46–63
- Angrist JD, Krueger AB (2001) Instrumental variables and the search for identification: from supply and demand to natural experiments. *J Econ Perspect* 15(4):69–85
- Auffhammer M, Hsiang SM, Schlenker W, Sobel A (2013) Using weather data and climate model output in economic analyses of climate change. *Rev Environ Econ Policy* 7(2):181–198. [ret016](#)
- Berkowitz D, Caner M, Fang Y (2012) The validity of instruments revisited. *J Econ* 166:255–266

- Davidson R, MacKinnon JG (2004) *Econometric theory and methods*. Oxford University Press, New York
- Fisher FM (1966) *The identification problem in econometrics*. RE Krieger Publishing Company, McGraw-Hill
- Frisch R, Waugh FV (1933) Partial time regression as compared with individual trends. *Econometrica* 1:221–223
- Garnache C, Mérel PR (2015) What can acreage allocations say about crop supply elasticities? A convex programming approach to supply response disaggregation. *J Agric Econ* 66(1):236–256
- Haile MG, Kalkuhl M, Von Braun J (2015) Worldwide acreage and yield response to international price change and volatility: a dynamic panel data analysis for wheat, rice, corn, and soybeans. *Am J Agric Econ*:1–9. doi:10.1093/ajae/AAV013
- Hendricks NP, Janzen JP, Smith A (2015) Futures prices in supply analysis: are instrumental variables necessary *Am J Agric Econ* 97(1):22–39
- Koopmans TC (1949) Identification problems in economic model construction. *Econometrica (Econometric Soc)* 17(2):125–144
- Leamer EE (1981) Is it a demand curve, or is it a supply curve? Partial identification through inequality constraints'. *Rev Econ Stat* 63(3):319–327
- Lütkepohl H (2014) Disentangling demand and supply shocks in the crude oil market: How to check sign restrictions in structural VARs. *J Appl Econ* 29(3):479–496
- Maddala GS (1977) *Econometrics*. McGraw-Hill, New York
- Maddala, GS (1992) *Introduction to Econometrics*. MacMillan Publishing Company, New York
- Manski CF (2003) *Partial identification of probability distributions*. Springer Science & Business Media, New York
- Renuka M, Kalirajan KP (2002) How IncomeElastic Is The consumers demand for services in Singapore *Int Econ J* 16(1):95–104
- Rigobon R (2003) Identification through heteroskedasticity. *Rev Econ Stat* 85:777–792
- Roberts MJ, Schlenker W (2013) Identifying Supply and Demand Elasticities of Agricultural Commodities: Implications for the US Ethanol Mandate. *Am Econ Rev Am Econ Assoc* 103(6):2265–95
- Santeramo FG (2014) On the Estimation of Supply and Demand Elasticities of Agricultural Commodities. AGRODEP Technical Note 10. International Food Policy Research Institute, Washington, DC. <http://www.agrodep.org/fr/resource/no-10-estimation-supply-and-demand-elasticities-agricultural-commodities>
- Sargan JD (1958) The estimation of economic relationships using instrumental variables. *Econometrica: J Econom Soc*:393–415
- Steinwender C (2014) Information frictions and the law of one price: when the States and the Kingdom became United (No. ERSD-2014-12). WTO Staff Working Paper. https://www.wto.org/english/res_e/reser_e/ersd201412_e.pdf
- Theil H (1965) The information approach to demand analysis. *Econom: J Econom Soc*:67–87
- Thurman WN, Wohlgenant MK (1989) Consistent estimation of general equilibrium welfare effects. *Am J Agric Econ* 71(4):1041–1045
- Wooldridge JM (2010) *Econometric analysis of cross section and panel data*. The MIT Press, Cambridge, Massachusetts
- Wright PG (1928) *The tariff on animal and vegetable oils*. Macmillan, New York
- Wright, PG (1929). *J Am Stat Assoc* 24:207–15

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Immediate publication on acceptance
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ springeropen.com
