A Note on Limited versus Full Information Estimation in Non-Linear Models*

J. Miguel Villas-Boas
(University of California, Berkeley)

December, 2007

* Address for correspondence: Haas School of Business, University of California at Berkeley, Berkeley, CA 94720-1900. E-mail: villas@haas.berkeley.edu.

A Note on Limited versus Full Information Estimation in Non-Linear Models

Abstract

This paper compares limited with full information estimation in non-linear models. The main result is that, in a non-linear model, under general monotonicity conditions, there is a specification of the full information model which is consistent with an arbitrarily specified limited information model. In particular, there is a specification of the full information model that is consistent with a typical linear model of the endogenous variable in a limited information approach. In general, in a non-linear model, a full information and limited information specifications do not represent the same model, but the limited information model covers a "larger" range of the space of potential models than a full information model. The paper presents an application to monopoly pricing with non-linear demand, and shows some simulation results of the relationship between the limited and full information model in that application.

1. Introduction

Data from economic environments come most often from the interaction between different economic agents. The behavior of one type of economic agents depends on the behavior of other economic agents, which may be based on information that is not available to the researcher. That is, when modeling the behavior of some type of agents we need to account for the endogeneity of the behavior of the other agents. The approaches that have been used to deal with this issue can be classified in two basic camps: a full information approach and a limited information approach (e.g., Hausman, 1984). Roughly speaking, the full information approach models all the economic agents' interaction in the system, explicitly assuming how the behavior of some agents depends on the behavior of other agents (constraints of parameters across equations). On the other hand, the limited information approach focusses on modeling of the behavior of one type of agents and uses some "limited" information about the behavior of the other agents to identify the behavior of the target type of agents.

It is well known that in linear models the full information estimation approach, although more efficient, puts more constraints on the model than a limited information approach. That is, if the assumptions made in the full information model are not true, then the full information approach may yield inconsistent estimates for the parameters, while the limited information approach may still yield consistent estimates if the "limited" information is true (e.g., Hausman, 1984). In other words, in a full information approach the mis-specification of one behavior equation affects the other behavior equations, while the limited information approach is not as prone to mis-specification (due to fewer assumptions).

However, the relationship between the full information and limited information approach in non-linear models is not as well understood.[1] This paper studies this relationship when the limited information approach uses an auxiliary equation.

The main result is that, under general monotonicity conditions, in a non-linear model there is a specification of the full information model which is consistent with an arbitrarily specified limited information model (that is, an auxiliary equation). That is, consider an economic system where the relationship between the explanatory variables and the behavior of the economic agents is monotonic. Then, if we specify a limited information model, there is a specification of the full information model where the monotonic assumptions hold and which is consistent with the limited information model. The proof is by construction of the specification of the full-information

[1] For estimation approaches of non-linear systems of equations see, for example, Amemiya, 1985, Ch. 8.

model that would make the limited information model true. This construction involves the inverse function of the behavior equation of the full information model, which is guaranteed to exist given the monotonicity assumption. In particular, this result means that there is a specification of the full information model that is consistent with a typical linear model of the endogenous variable in a limited information approach (e.g., Newey 1987, Rivers and Vuong 1988, Villas-Boas and Winer 1999, Blundell and Powell, 2004, Petrin and Train 2006).[2]

In general, in a non-linear model, a full information and a limited information specifications do not represent the same model (unlike in the linear-model case). However, we can recover the idea of the linear models that full information estimation requires more assumptions than limited information estimation. In a non-linear model this idea is represented by the result that the limited information model covers a "larger" range of the space of potential models than a full information model does, in a sense defined below. This result can be seen as just formalizing some intuition that one might have from the linear models, but it has important implications about when one needs to fully model structurally an economic system, or when we can use auxiliary equations to consider the endogeneity of some variables in part of the economic system.

As discussed below, a well-known "alternative" limited information approach is to invert, if possible, the target behavior equation(s) to obtain the error term as a function of the observable variables, create a moment condition with the orthogonality of the error term and some instruments, and finally estimate the parameters with the general method of moments. However, in some cases either the inversion of the target behavior equation is not possible (for example, in the case of limited dependent variables), or is computationally difficult. One may also have concerns about the small sample statistical properties of the method of moments in the particular problem being studied. The researcher may therefore want to use in some cases a limited information approach with an auxiliary equation, and this paper discusses the relationship between this approach and a full information approach.

The interest in non-linear models has recently increased because of the interest in more careful modeling of the behavior of economic agents in ways that are consistent with first principles (for example, assuming utility maximization as a basis for consumer choice), while continuing to have models that are easy to work with. This paper studies how the limited and full information approaches apply to such models.

The next section presents the model set-up, the question of the relationship between limited and full information estimation, and derives the main results. Section 3 presents an application to

[2]See also Wooldridge (2002), pp. 472-477.

monopoly pricing with non-linear demand, and shows simulation results of the relationship between the limited and full information model in that application. Finally, Section 4 presents concluding remarks.

2. Model and Results

In many circumstances one can write a non-linear model of the behavior of economic agents that has the form

$$y = h(x, \varepsilon) \tag{1}$$

where y is the behavior, x is some observable variables that affect that behavior, ε are some unobservable variables that affect that behavior, and where the function $h()$ is known/assumed. An example of this is where y is the choice of a consumer or the market share of a firm, x represents the prices charged by firms, ε represents some unobservable characteristics of the consumers, and $h()$ is obtained, for example, from some utility maximization by consumers (for example, the demand model in Berry et al. 1995).[3]

The linear example of equation (1) is just

$$y = x\beta + \varepsilon.$$

This is an example that we will always come back to, and compare results with.

One crucial issue in the estimation of a model of this type is that x may be set as a function of ε. That is, x is endogenous. In the example above, firms may set prices as a function of characteristics of the consumer preferences that are not observed by the researcher.

In general, one might also have some theory of how x is determined. In the example above, one could, for instance, have the theory that firms fully observe ε and each other's costs, are choosing prices to maximize their own profits, and behaving as in a Nash equilibrium.

From this theory, one might have

$$x = f(g(z, \eta), \varepsilon) \tag{2}$$

in which one knows the function $f()$ from the theory, and assumes a function $g()$ of some exogenous observables z and unobservables η, representing some construct that also affects x according to the theory. Assume that both $f()$ and $g()$ are monotonic in its arguments. In particular, assume

[3] In terms of the example above one might have to define one of the x as minus the price of that firm.

$\frac{\partial f}{\partial g}, \frac{\partial f}{\partial \varepsilon}, \frac{\partial g}{\partial z}, \frac{\partial g}{\partial \eta} > 0$. In terms of the example above, $g()$ could represent the marginal cost of production, z some input prices, and η some unobservables that affect the marginal cost of production. The price being set, according to the theory, is a function $f()$ of the marginal cost of production and of the demand unobservables.[4] The error term ε is assumed independent of z.

In a linear example we would have

$$g(z, \eta) = z\gamma + \eta$$

and

$$x = z\lambda\gamma + \lambda\eta + \lambda'\varepsilon$$

where λ and λ' would be some function of β.

Estimating (1) and (2) together, one can, in general, obtain consistent estimates of the parameters in $h()$ and in $g()$. This has been called the full information approach in the sense that one uses the information from the theory, which is equation (2), to estimate equation (1). However, one potential problem with this approach is that if the theory is incorrect one brings the mis-specification of equation (2) into the estimation of (1), which results in inconsistent estimates of the parameters of equation (1).

In the linear case one well-known approach is to use z as instruments for x in equation (1) to obtain an estimate of β. This produces consistent estimates of β, which are not affected by any possible assumed relationship between λ or λ' and β.

Consider now the non-linear case. One approach, if the function $h(x, \varepsilon)$ is invertible, is to obtain $\varepsilon = h_\varepsilon^{-1}(x, y)$, and use some instruments z with the property that $E[\varepsilon|z] = 0$, to construct moment conditions, from which consistent estimates can be obtained with a general method of moments (GMM) estimator. For example, if $h(x, \varepsilon)$ is parametrically specified as $h((x, \varepsilon, \theta)$, then a consistent GMM estimator of θ could be obtained from the moment conditions $E[h_\varepsilon^{-1}(x, y, \theta)m(z)] = 0$ where $m(z)$ is a vector of nonlinear functions of the instruments (e.g., using only the demand moment in Berry et al. 1995). This approach achieves consistent estimates of the parameters of the target behavior equation (1) which is robust to any mis-specification in (2), and is consistent with any

[4] To simplify the presentation we consider y, x, and z to be scalars. The extension to the multi-dimensional case is straightforward, but instead of a monotonicity condition we now need that $f()$ be invertible. One sufficient condition for $f()$ to be invertible is that g, having the same dimension as x, belongs to a convex set, and the Jacobian of $f()$ with respect to g is negative (or positive) semidefinite (Gale and Nikaido, 1965). If there is a dominant diagonal condition (for example, under price competition with products sufficiently differentiated, then the Nash equilibrium price of one firm may be more affected by its marginal cost than by the marginal costs of the other firms) then the condition above is satisfied. See also Vives (1999, pp.47-48).

specification of (2).

However, in some cases, the inversion of the function $h(x, \varepsilon)$ is not possible (for example, with limited dependent variables), or is computationally difficult, which may rule out the possibility of using that method of moments approach. In other cases the researcher may not be fully satisfied with the small sample statistical properties of the method of moments estimator in the particular problem being studied.

In those cases, another possibility may be to use an auxiliary equation of the endogenous variable(s) x without the constraints of equation (2). The purpose of this paper is to identify conditions under which one can perform this limited information approach with an auxiliary equation in a non-linear model, that is consistent with equation (2), and still obtain consistent estimates of the parameters of the target behavior equation (1) even if (2) is mis-specified. For estimation procedures with such auxiliary equations see, for example, Newey (1987), Rivers and Vuong (1988), or Blundell and Powell (2004).

In particular, suppose that one wants to substitute equation (2) with a linear equation

$$x = z\alpha + \widetilde{\eta} \tag{3}$$

where the parameters α are not constrained to be related to the parameters in equation (1), and where we would like to estimate the parameters in equation (1) by estimating (1) and (3) together. For this to be correct it must be that equation (3) is true given equation (2). In particular, given some assumed distribution of $\widetilde{\eta}$, are there a function $g()$ and a distribution of η such that equation (3) is true? Moreover, is there an infinite number of triples of f, g and the distribution over η such that equation (3) is true, so that we can say that estimating (1) with the help of (3) produces consistent estimates for the parameters in (1) without being affected by some mis-specification in (2)?

Given a cumulative distribution of $(\widetilde{\eta}, \varepsilon)$ given z, and a function f increasing in the first argument, we are looking for the existence of a function $g()$ and a probability distribution of η such that

$$z\alpha + \widetilde{\eta} =^d f(g(z, \eta), \varepsilon), \tag{4}$$

where $=^d$ means that "it has the same probability distribution given z."[5]

[5] The presentation here is based on the complete probability distributions of the error terms. One should be able to obtain similar results when one only considers a set of moment conditions for the probability distributions of the error terms. Note that the same questions about equation (3) can also be raised if one takes a semiparametric approach with respect to the $\widetilde{\eta}$ error (Blundell and Powell, 2004).

Note that for the linear example the answer to this question is obvious, because, if $f()$ is linear, and we make $g()$ linear, having η distributed as a linear combination of $\widetilde{\eta}$ and ε makes the equality hold.[6]

For the case where $f()$ is non-linear, note that we can obtain the distribution of $g(z,\eta)$, given z, and given the probability distribution of $(\widetilde{\eta},\varepsilon)$ and the function f, by rewriting (4) as

$$g(z,\eta) =^d f^{-1}(z\alpha + \widetilde{\eta}, \varepsilon). \tag{5}$$

We can, similarly, obtain the joint distribution of $(g(z,\eta),\varepsilon)$. In general, $g(z,\eta)$ is not independent of ε. Denote the cumulative probability distribution of $g(z,\eta)$ given z as $\Phi(g(z,\eta);z)$. Note that, given the monotonicity of $f()$, this also means that $g(z,\eta)$ is stochastically increasing in z (that is, the probability distribution of $g(z_1,\eta)$ first order stochastically dominates the probability distribution of $g(z_2,\eta)$ if and only if $z_1 > z_2$).

Having obtained the distribution of $g(z,\eta)$ given z, we then have immediately the result that, given a distribution of $(\widetilde{\eta},\varepsilon)$, there is a specification of the full information model that represents the same model as the limited information model represented by (1) and (3).[7] We state this result in the following proposition.

Proposition 1: *Consider a non-linear economic system composed of the behavior equations (1) and (2), the full information model, where function f is given and monotonic. Then, for every limited information model, (1) and (3), there is a probability distribution over η and a function $g(z,\eta)$, such that the specification of the full information model represents the same model as the limited information model.*

The implication of this proposition is that, in general, the researcher can focus on a primary equation (e.g., equation (1)) with a limited information model, and not have to worry about a full information model of the economic system. More specifically, the functional form of (3) is not inconsistent with the known properties of the full information model, because there is a probability distribution over η and a function $g(z,\eta)$ that can make it consistent.

We now turn to the idea that a full information model requires more assumptions than a limited information model. One first point to make is that in the typical specification of a full information

[6] For example, if $(\widetilde{\eta},\varepsilon)$ are distributed jointly normal, then η's probability distribution is a normal distribution.

[7] To simplify notation (and corresponding to what is typically presented in the literature), we present the discussion above with equation (3) linear in z and $\widetilde{\eta}$. Exactly the same result holds if equation (3) is non-linear and monotonic in z and $\widetilde{\eta}$.

model one specifies both the probability distribution of η and the functional form $g(z, \eta)$. From above, it is clear that for a full information model to be completely specified we only need to specify the probability distribution of $g(z, \eta)$ given z.

More interestingly, note that from equation (5), for each function $f()$ there is a probability distribution of $g(z, \eta)$ that satisfies the equality. That is, the limited information model represented by equation (3) can be true for an infinite number of full information models (and infinite-dimensional), represented by the function $f()$. Because of the importance of the result we state it in the following proposition.

PROPOSITION 2: *For every limited information model, equations (1) and (3), there is an infinite number of full information models (a function f and a probability distribution of $g(z, \eta)$ given z in equations (1) and (2)) that represent the same model as the limited information model. For every specification of a full information model, there is only one limited information model that represents the same model as that full information model.*

This means that in order to obtain consistent estimates of the parameters, the full information model requires more assumptions than the limited information model. A limited information model is more robust to mis-specification than a full information model. Note, however, that arbitrary specifications of a limited and full information model may not represent the same model. This is also immediate from equation (5). If the full information model specifies a probability distribution for $g(z, \eta)$ that is different than the one implied by (5) then the limited and full information models represent different models. And it could be that the full information model is true, which means that the limited information model is mis-specified. Note then that this means that although one may interpret the limited information approach with the auxiliary equation as requiring less assumptions than the full information model, this limited information approach requires still more assumptions than the method of moments described above, as that method obtains consistent estimates of the parameters for any specification of the full information model (the function f and the probability distribution of $g(z, \eta)$ given z).

In order to gain further insight into the relationship between the full information model and the limited information with an auxiliary equation, consider now the question of whether there is a function $g(z, \eta)$ and a probability distribution of η given z that is independent of z, such that equation (5) holds. Such probability distribution of η given z, independent of z, is often considered in full information specifications.

From equation (5), we know that the cumulative probability distribution of η given z is $\Phi(g(z,\eta);z)$. One can then obtain that for this to be independent of z we need

$$\frac{\partial g}{\partial z} = -\frac{\Phi_1}{\Phi_2} \tag{6}$$

where Φ_i represents the partial derivative of Φ with respect to its ith argument. That is, there is a function $g()$ and a probability distribution of η given z, independent of z, such that the limited information model is true. Note that from equation (6), we can see that, in general, the function $g(z,\eta)$ will not be linear in z and η, which is a typical specification in full information models. That is, if the true full information model is the one with the typical specification of a linear $g(z,\eta)$, then the typical specification of the limited information model (the linear equation (3)) is not true. Equation (6), in general, specifies a function $g(z,\eta)$ that is not linear in z and η as, in general, Φ_1 and Φ_2 will also be functions of η.

However, the function $g(z,\eta)$ can be approximated by a Taylor expansion of equation (5) to a linear function as

$$g(z,\eta) =^d \frac{\partial f^{-1}}{\partial z\alpha + \tilde{\eta}}(z_0\alpha, 0) \cdot ((z - z_0)\alpha + \tilde{\eta}) + \frac{\partial f^{-1}}{\partial \varepsilon}(z_0\alpha, 0) \cdot \varepsilon. \tag{7}$$

3. Example of an Application

Consider as a simple example the case of demand for a monopoly product, and in which price may be set strategically. In order to make the presentation clearer, denote the endogenous variable x as P for price. Suppose that the demand for the product takes the form

$$y = h(P,\varepsilon) = \frac{1}{1 + e^{\beta_0 + \beta_1 P + \varepsilon}} \tag{8}$$

where the parameters are β_0 and β_1 are to be estimated from data. The researcher observes the price P and demand y in each period, but does not observe a demand shock per period ε. Denote the marginal costs of production as $g(z,\eta)$, where z represents some exogenous observable input prices, and η are some shocks that are not observed by the researcher.

In the full information model, the monopolist is maximizing her profit with respect to price after observing the shocks ε and η. The first order of such maximization results in

$$1 + e^{\beta_0 + \beta_1 P + \varepsilon} - \beta_1 e^{\beta_0 + \beta_1 P + \varepsilon}(P - g(z,\eta)) = 0 \tag{9}$$

which is the implicit version of equation (2), and defines the price P as a function of z, ε, and η.

Consider now the limited information equation corresponding to equation (3) as

$$P = \alpha_0 + \alpha_1 z + \widetilde{\eta}. \tag{10}$$

The question in the previous section was whether, given a joint probability distribution for $(\varepsilon, \widetilde{\eta})$, is there a function $g(z, \eta)$, such that there is a probability distribution over $g(z, \eta)$ given z, such that equation (10) is true given equation (9). In order to see this, we can use equations (9) and (10) to obtain equation (5) in this example as

$$g(z, \eta) =^d \alpha_0 + \alpha_1 z + \widetilde{\eta} - \frac{1 + e^{\beta_0 + \beta_1(\alpha_0 + \alpha_1 z + \widetilde{\eta}) + \varepsilon}}{\beta_1 e^{\beta_0 + \beta_1(\alpha_0 + \alpha_1 z + \widetilde{\eta}) + \varepsilon}}. \tag{11}$$

It is then immediate to obtain the distribution of $g(z, \eta)$ given z, given the joint distribution over $(\varepsilon, \widetilde{\eta})$.

In order to get some sense of such distributions I run the following simulation, with $\beta_0 = 0, \beta_1 = .3, \alpha_0 = 3, \alpha_1 = 1$. I drew 5,000 pairs $(\varepsilon, \widetilde{\eta})$ from a joint normal distribution with variance .2 for each variable, and covariance $-.1$, and looked at $1,000$ observations each for the values of $z = 3, 3.2, 3.4, 3.6, 3.8$. The kernel density of the distribution of $g(z, \eta)$ is presented in Figures 1-5 for each z.[8] Other parameter values were tried and led to similar kernel densities.

The linear regression of the price P on z yielded an $R^2 = .29$. The mean demand across all observations was .14 with a standard deviation of .05, a minimum of .03, and a maximum of .39. In order to check the approximation (7) for this example I also run the linear regression of $g(z, \eta)$ on z, which yielded a $R^2 = .34$.[9] In order to have a sense whether the additive separable $g(z, \eta)$ under the linear model yielded a distribution for η that was "approximately" independent of z, I run Kolmogorov-Smirnov tests comparing the distributions of the residuals of the regression of $g(z, \eta)$ on z for each z. The results of these tests are presented in Table 1. It can be seen that the tests do not reject the hypothesis that the distributions of the residuals in the linear model across z are the same. However, this is just a small simulation for some parameter values, and these tests of the additive separable approximation (7) could lead to different results for other parameter values

Table 1
Kolmogorov-Smirnov Tests of Residuals
of Linear Model across z

z	3.00	3.20	3.40	3.60	3.80
3.00		.032	.044	.040	.030
		(.67)	(.27)	(.38)	(.74)
3.20			.040	.041	.032
			(.38)	(.35)	(.67)
3.40				.028	.041
				(.81)	(.35)
3.60					.044
					(.27)

The Table presents the value of the Kolmogorov-Smirnov statistic D for each pair of distributions of residuals determined by a pair of z's. The P-value is presented in parenthesis.

or other functional forms.

4. Conclusion

This paper presents some results on the relationship between limited and full information estimation in non-linear models. Given general monotonicity conditions, any limited information model with an auxiliary equation can be shown to be true with a specification of the full information model for non-linear models. Furthermore, it can be argued that a limited information model makes less assumptions for estimation than a full information model, as for any limited information model there is an infinite number (infinite-dimensional) of full information models for which the limited information model is true. However, the typical applications of the limited information model (additive separable error term) and full information model (additive separable error, in which the distribution is independent of the exogenous variables) cannot both be true in general in a non-linear model. An application of static monopoly was presented, where the complete specification of the full information model was recovered given a limited information specification.

[8] The kernel density is obtained with the Epanechnikov kernel with bandwidth equal to $.9\sqrt{\mathrm{Var}(g(z,\eta))}/1000^{1/5}$, where $\mathrm{Var}(g(z,\eta))$ is the variance of $g(z,\eta)$ given z (see, e.g., Tapia and Thompson, 1978). The bandwidth for each z was approximately .10.

[9] The parameter on z was estimated at 1.15 with a standard error of .02. The constant was estimated at -1.39 with a standard error of .08.

REFERENCES

AMEMIYA, T. (1985), *Advanced Econometrics,* Harvard University Press: Cambridge, Massachusetts.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995), "Automobile Prices in Market Equilibrium," *Econometrica,* **63**, 841-889.

BLUNDELL, R.W., AND J.L. POWELL (2004), "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies,* **71**, 655-679.

GALE, D., AND H. NIKAIDO (1965), "The Jacobian Matrix and Global Univalence of Mappings," *Mathematische Annalen,* **159**, 81-93.

HAUSMAN, J. (1984), "Specification and Estimation of Simultaneous Equation Models," *in Handbook of Econometrics,* Vol. 1, Eds. Z. Griliches and M. Intriligator, Elsevier: Amsterdam.

NEWEY, W.K. (1987), "Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables," *Journal of Econometrics,* **36**, 231-250.

PETRIN, A., AND K. TRAIN (2006), "Control Function Corrections for Unobserved Factors in Differentiated Product Models," *working paper,* University of Minnesota and University of California, Berkeley.

RIVERS, D., AND Q.H. VUONG (1988), "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics,* **39**, 347-366.

TAPIA, R.A., AND J.R. THOMPSON (1978), *Nonparametric Probability Density Estimation,* Baltimore: Johns Hopkins University Press.

VILLAS-BOAS, J.M., AND R.S. WINER (1999), "Endogeneity in Brand Choice Models," *Management Science,* **45**, 1324-1338.

VIVES, X. (1999), *Oligopoly Pricing: Old Ideas and New Tools,* Cambridge, Massachusetts: The MIT Press.

WOOLDRIDGE, J.M. (2002), *Econometric Analysis of Cross Section and Panel Data,* Cambridge, Massachusetts: The MIT Press.

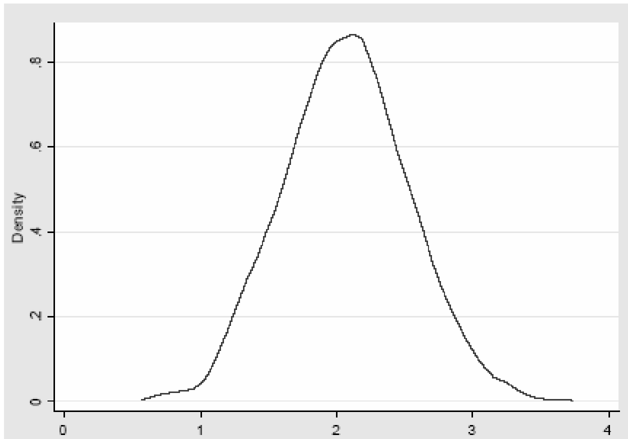

Figure 1: Kernel density of $g(z,\eta)$ for z=3.0 in simulation.

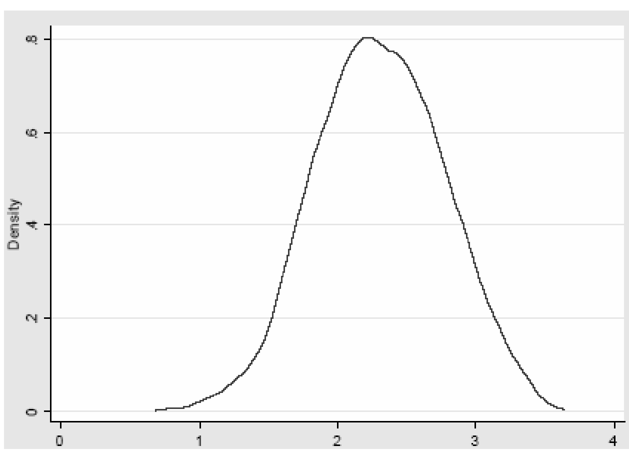

Figure 2: Kernel density of $g(z,\eta)$ for z=3.2 in simulation.

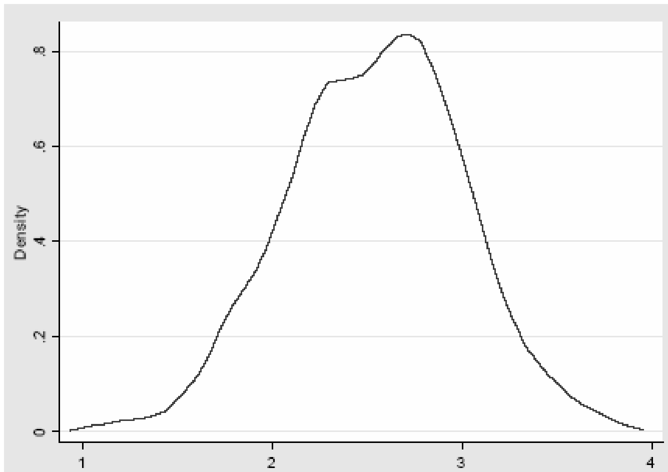

Figure 3: Kernel density of $g(z,\eta)$ for z=3.4 in simulation.

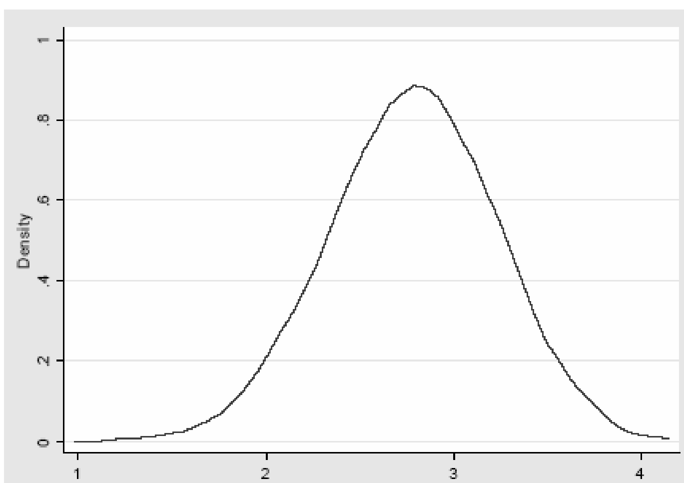

Figure 4: Kernel density of $g(z,\eta)$ for z=3.6 in simulation.

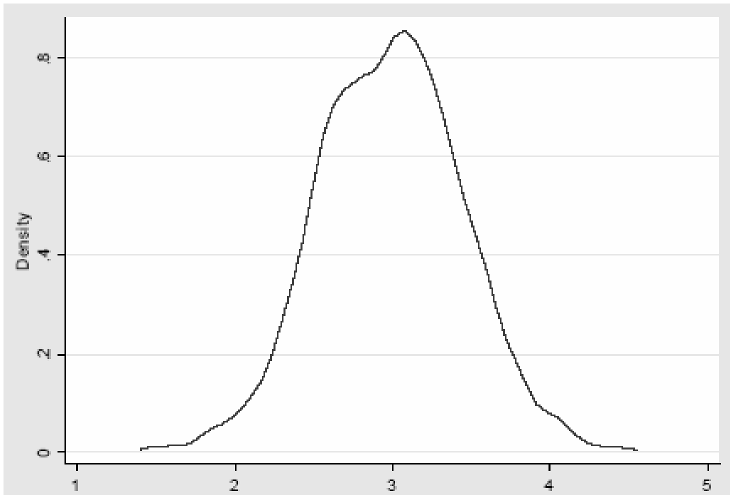

Figure 5: Kernel density of $g(z,\eta)$ for z=3.8 in simulation.