ELSEVIER

Economics Letters 54 (1997) 119–126

economics letters

Generated regressors in linear and nonlinear models

Kishore Gawande*

Department of Economics, University of New Mexico, Albuquerque, NM 87131, USA

Received 11 July 1996; accepted 18 December 1996

Abstract

The need to use estimates from separate studies as regressors often arises. Such regressors need to be modelled as variables measured with error. We demonstrate the use of generated regressors in linear and nonlinear models that arise from the predictions in Grossman and Helpman (*American Economic Review*, 1994, 84 (4), 833–850) and Grossman and Helpman (*Journal of Political Economy*, 1995, 103, 675–708).

Keywords: Generated regressors; Errors in variables

JEL classification: C20; F13

1. Introduction

It is often the case that economic variables employed in econometric analyses themselves need to be estimated before they can be used. Since generated regressors often have large sample variation only because they have large variation in their standard errors across observations, they need to be modelled as variables measured with error. The main intent of the paper is to demonstrate ways to deal with generated regressors in linear and nonlinear models which practitioners in all fields should find useful. A case in point is the innovative and important models in the literature of the political economy of protection advanced by Grossman and Helpman (1994), (1995). In the next section we consider two hypotheses from their models and in the third section we perform the estimations using generated regressors.[1]

2. Two models of the political economy of protection

Hypothesis 1. (Grossman and Helpman, 1994; restricted version of Proposition 2, p. 842.) The government chooses trade taxes and subsidies that satisfy

*Tel.: +1 505 277-1965; fax: +1 505 277-9445; e-mail: gawande@unm.edu

[1]While the analysis here is too brief to constitute a test, the data is carefully constructed and does provide some empirical evidence about the Grossman–Helpman models.

PII S0165-1765(97)00014-1

$$\frac{t_i}{1+t_i} = \beta\left(\frac{z_i}{e_i}\right), \quad i = 1, \ldots, n, \tag{1}$$

where t_i is the ad valorem trade taxes or subsidies for good i; z_i is the ratio of domestic output to imports and e_i is the absolute own price elasticity of import demand.[2] Consider the denominator in the right-hand side expression, e_i. All else equal, industries with high absolute import demand elasticities or export supply elasticities will have smaller deviations from free trade. The reason is two-fold. First, the social cost of protection is relatively higher in such industries and the government will be relatively more averse to protecting these industries. Second, lobbies in industries other than the one seeking protection who share the deadweight losses, especially upstream industries, will lobby against protection to that industry. Now consider the numerator, z_i. All else equal, industries with great political power, here reflected in a high ratio of domestic output to imports, z_i, will secure greater protection because in high value-added industries the stakes from protection are greater, inducing larger political contributions from factors specific to these industries, while low import volume (which raise z_i) does not impose enough of a social cost that would lead to organized opposition against protection for those industries. Hence the prediction that $\beta > 0$.

Hypothesis 2. (Grossman and Helpman, 1995; restatement of Eq. (25).) Let t_i be the home tariff rate and t_i^* be the foreign tariff rate on industry i's imports. Then in a 'trade talks' equilibrium the ratio (τ_i/τ_i^*) where $\tau_i = 1 + t_i$ and $\tau_i^* = 1 + t_i^*$, can be written as a function of home and foreign output-to-import ratios (z_i and z_i^*, respectively) and absolute home and foreign import elasticities (e_i, e_i^*, respectively) as follows:

$$\ln\left(\frac{\tau_i}{\tau_i^*}\right) = \ln\left[1 - \xi^*\left(\frac{z_i^*}{e_i^*}\right)\right] - \ln\left[1 - \xi\left(\frac{z_i}{e_i}\right)\right]. \tag{2}$$

Eq. (2) is derived from Eq. (25) in Grossman and Helpman and written in a form suitable for the empirical analysis.[3] The predictive content of this theory is contained in the hypothesis that for industries that are organized into lobbies in both countries, $\xi > 0$ and $\xi^* > 0$. The pattern of protection reflects equilibrium at two levels. At the domestic level equilibrium is determined by the weight government places on a marginal dollar of PAC lobbying for protection (which is itself determined within the Grossman and Helpman (1994) framework of inter-firm political competition) versus a marginal dollar of welfare loss imposed on the public due to the protection. At the international level a bargaining equilibrium is determined between two countries conditionally on the sets of (t, t^*) allowed by their domestic equilibria. This is the most theoretically rigorous model in this literature and makes the most precise prediction about the cross-industry pattern of protection. The prediction that $\xi > 0$ and $\xi^* > 0$ means that even though the two governments grant protection to those industries

[2] The full proposition in Grossman and Helpman (1994) includes industries that are net exporters, but since we have no data on export supply elasticities, those industries are excluded from the analysis. Since our data is for US trade barriers against Japan, we do not lose many observations from dropping US (bilateral) net export industries from the US–Japan cross-section.

[3] In their Eq. (24) substitute $M_i' = e_i/(p_i/M_i)$ and $-M_i^{*\prime} = e_i^*/(p_i^*/M_i^*)$ and, noting that $p_i = \tau_i^O \pi_i$ and $p_i^* = \tau_i^{*O} \pi_i^*$, simplify the right-hand side. Divide both sides of the equation by τ_i^{*O} and express the ratio (τ_i^O/τ_i^{*O}) as a function of the other variables. Taking logs we get the expression in Eq. (2).

that lobby intensively, where lobbying intensity drops off they are each willing to lower protection relative to the partner country. The positive coefficient on ξ thus conveys the message that the lower the domestic cost to the US (incorporated in the rhs variable without asterisks in Eq. (2)), the lower its level of protection relative to its partner. Further, when industries are organized politically on both sides, it is the difference in the political pressures they are able to apply to their governments that determine their relative levels of protection.

3. Model estimation

3.1. Methodology with generated regressors

In Eqs. (1) and (2) the key import elasticity variable needs to be separately estimated. Fortunately, these elasticities have been estimated for each of the 3-digit SIC industries by Sheills et al. (1986), which we borrow for our purpose.[4] Let the variable E (own price elasticity of imports) denote those elasticity estimates. For some industries i, E_i is quite imprecisely estimated while for others it is sharply estimated, with sample values in $[-23.85, -0.042]$,[5] and standard errors in $[0.070, 118.0]$. Clearly, the direct use of the unadjusted E variable will lead to quite erroneous and unpredictable results. We propose to use an errors-in-variables correction on E based on the methodology in Fuller (1987). E_i is modelled as the observed value of the true (unobserved) own price elasticity of imports, e_i, but which is measured with error:

$$E_i = e_i + u_i, \tag{3}$$

where u_i is the measurement error in E_i with mean 0 and known variance $\sigma^2_{u,i}$.[6] This variance is equal to the square of the estimated standard errors reported on the estimates E_i in Sheills et al. (1986).

First consider model (1). It is linear in parameters although nonlinear in the variable measured with error. This requires a specification of the error distribution on the right-hand side variable (z_i/E_i) or an approximation, under the presumption z_i (the value added-to-imports ratio) is measured accurately. If we use a Taylor approximation on $E_i^{-1}z_i$, the error variance is[7]

$$V(E_i^{-1}z_i - e_i^{-1}z_i) \doteq V(-e_i^{-2}z_i u_i) = e_i^{-4}z_i^2\sigma^2_{u,i} \doteq (E_i^2 - \sigma^2_{u,i})^{-2} z_i^2 \sigma^2_{u,i}. \tag{4}$$

[4] Since all other data is at the 4-digit level, we replicate these estimates at the 4-digit level.

[5] A few import elasticities have a contrary positive sign and are discarded.

[6] Normality is not needed for consistent estimation, only the weaker conditions in Eqs. (3.1.3) and (3.1.4) of Fuller. However, normality of Eq. (4) is needed far statistical testing and inference.

[7] Consider the measurement error model

$$y_i = \beta(z_i/e_i) + w_i\gamma + \varepsilon_i,$$

$$E_i = e_i + u_i,$$

where the (scalar) coefficient β is the issue of interest, γ is a $k \times 1$ parameters on the control variables w_i, and ε_i is distributed iid normally. The measurement error u_i has variance $\sigma^2_{u,i}$. Expanding $E_i^{-1}z_i$ around $e_i^{-1}z_i$ we get

$$E_i^{-1}z_i = e_i^{-1}z_i - e_i^{-2}z_i(E_i - e_i).$$

The error variance of the function $E_i^{-1}z_i$ is $V(E_i^{-1}z_i - e_i^{-1}z_i)$ which can be approximated as in Eq. (4).

To evaluate this expression we use $\max\{E_i^2 - \sigma^2_{u,i}, \sigma^2_{u,i}\}$ to avoid small numbers. Now Eq. (1) can be analyzed as a model linear in (z_i/e_i) which is measured as (z_i/E_i), a variable that has measurement error with mean zero and variance given by Eq. (4). We can then estimate the model using the method of moments estimator in Eq. (3.1.19) of Fuller (1987, p. 193).

Now consider model (2). It is nonlinear in parameters and in the variable measured with error. The method of moments estimator is difficult to apply here. An alternative, simpler, method (see, e.g., Fuller, 1987, Ch. 3) is to replace e_i in Eq. (2) by the prediction $\hat{E}_i$ constructed as follows. Denote the sample variance of E by $\bar{\sigma}^2_E$ and the mean of the measurement error variances by $\bar{\sigma}^2_u$. Let $\hat{\sigma}^2_e = \bar{\sigma}^2_u - \bar{\sigma}^2_E$ and $\bar{E}$ denote the sample mean. Now construct the predictor

$$\hat{E}_i = \bar{E} + \frac{\hat{\sigma}^2_e}{\sigma^2_{u,i}}(E_i - \bar{E}). \tag{5}$$

Thus, whenever E_i has measurement error variance exactly equal to $\hat{\sigma}^2_e$ (an estimate for the sample variance of e_i had we been able to measure it exactly) it is presumed to be measured without error. Otherwise it is scaled down or scaled up according to Eq. (5).[8] The model in Eq. (2) is then estimated by nonlinear least squares or ML with the predictor in Eq. (5) used in place of E_i.

3.2. Empirical results

Two differences between the theoretical prediction in Eq. (1) and the estimating equations above should be noted. First, while Grossman and Helpman derive their equation for ad valorem tariffs, we use NTB coverage ratios as of 1983, since the practice of protection after the Tokyo round has primarily involved NTBs (see, e.g., Leamer, 1990). Second, own import elasticities are available only for the US. In order to proceed to estimate model (2) we must make the assumption that the elasticities (true and measured) are the same in the US and its partner countries, that is, $E_i = E_i^*$. While this would be quite incorrect if the partner countries were developing countries, this assumption is not unreasonable for similar countries who mostly trade in similar goods. The alternative of performing the Sheills et al. study for the partner country is a costly one. Construction of the NTB data and other variables used in the study is detailed in Gawande (1995).

A stochastic version of Eq. (1) is used to investigate the Grossman and Helpman (1994) hypothesis of protection for sale. In Table 1 two models are estimated based on Eq. (1): a simple linear model with just the issue variable z_i/e_i together with a constant term, and a model with ten political-economic control variables in addition to the variables in the simple model. These variables are motivated by Baldwin (1986). In order to group NTBs into somewhat homogeneous categories we run the models with price NTB coverage ratios (P_i) that include and quantitative NTB coverage ratios

[8] Very large estimated standard errors on E_i, can raise $\bar{\sigma}^2_E$ so high as to lead to a negative value for $\hat{\sigma}^2_e$. This points out that estimates of E_i in Sheills et al. with very high standard errors are 'inadmissible'. We drop E_i estimates with standard error exceeding 9. After dropping these poorly measured elasticities and applying the correction in Eq. (5), values of $\hat{E}_i$ lie in the interval [−2.356, −0.524] which are realistic elasticity values. By contrast the uncorrected E_i values were contained in the large interval [−23.85, −0.042]. Taking account of the positive values of E_i that were dropped from consideration earlier, we have a sample size of 247.

Table 1
Hypothesis 1. Grossman and Helpman (1994): $\beta>0$ U.S. bilateral NTB coverage of 4-digit SIC imports from Japan

$\frac{t_i}{1+t_i} = \beta(z_i/e_i) + X\gamma + \varepsilon_i$, (regression model)

$E_i = e_i + u_i$ (measurement error model)

	With errors in variables correction				No errors in variables correction			
	Dependent variable: US price NTBs: $P_i/(1+P_i)$		Dependent variable: US quant NTBs: $Q_i/(1+Q_i)$		Dependent variable: US price NTBs: $P_i/(1+P_i)$		Dependent variable: US quant NTBs: $Q_i/(1+Q_i)$	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
β	0.221	0.365	−0.225	−0.154	0.166*	0.253*	−0.168	−0.107
	(0.358)	(0.455)	(0.361)	(0.193)	(1.620)	(2.293)	(1.489)	(1.047)
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
N	247		247		247		247	
k (incl. constant)	2	16	2	16	2	16	2	16
R^2	0.011	0.095	0.009	0.361	0.011	0.096	0.009	0.363

Notes:
(1) NTBs measured as coverage ratios. Construction is described in Leamer (1990).
(2) z_i is value added in industry i divided by imports from Japan. e_i is absolute value of own import price elasticity. As a measure of e_i we use the import elasticities estimated in Sheills et al. (1986), denoted E_i. The measurement error variance of u_i are the squares of the standard errors reported in Sheills et al.
(3) EIV corrected estimates are by method of moments based on Eq. (3.1.19) in Fuller (1987). Uncorrected estimates are by OLS.
(4) Absolute t-values in parentheses. ** and * signify statistical significance at 1% and 5% (one-tailed test), respectively. k includes constant.
(5) The control variables in Model 2 are U.S. exports scaled by consumption, corporate PAC spending per firm, percent unionization, proportion of workers that are scientists and engineers, proportion managerial, proportion unskilled, average earnings, output per firm, 4-firm industry concentration, geographical spread (number of states in which production is located), number employed, employment growth, change in import penetration between 1979 and 1982, labor intensity. See Gawande (1995) for construction of variables.

(Q_i) separately.[9] In Table 1, the columns under the heading "With errors in variables correction" report the method of moments estimates using Eq. (3.1.19) in Fuller (1987). In the columns under "No errors in variables correction" are reported OLS estimates. Before examining the effect of measurement errors, consider the nature of the inferences about the Grossman and Helpman hypothesis. The main implication of the Grossman and Helpman hypothesis is that $\beta>0$; industries with greater stakes (higher value added-to-imports ratio) and lower potential for deadweight losses (smaller absolute import elasticities) will succeed in obtaining greater protection. This is borne out by US price NTBs on imports from Japan, but not by US quantitative NTBs on imports from Japan. Price NTBs are similar to ad valorem tariffs in their effects than are quantitative restrictions, and therefore

[9]Price NTBs include, for example, antidumping and countervailing duties, and quantitative NTBs include quotas and voluntary export restraint agreements. Hence P_i and Q_i replace t_i in the Grossman and Helpman model. We ran the models with post-Tokyo round tariffs, but found no support for the theory from that data.

are closer in spirit to the Grossman and Helpman theory which is framed in terms of tariffs.[10] If the theory is judged in terms of the price NTB results, the estimates provide some support for the theory. However, that support is statistically weak. When there is no correction for measurement errors in E_i, and E_i is used in place of e_i, the estimate on β of 0.166 from the simple model is statistically significant at 5% (right-tailed test) and the estimate of 0.253 from the general model is statistically significant at 1%. However the method of moments estimates after the EIV corrections, while larger in magnitude than the OLS estimates (0.221 from the simple model and 0.365 from the general model) are not statistically significant. The increased standard error of estimate(s), which is a well-known result in the EIV literature, is the main effect of the EIV correction; if we remove the variation in E that arises from measurement error from the precision matrix $X'X$, this has the effect of making the matrix of regressors X more collinear. This reduces the information contained in the data about individual coefficient, which is then reflected in higher standard errors or lower t-values. Hence, if we wish to correct for EIV and want sharper inferences, the only recourse is through the use of personal prior information, as suggested by Klepper and Leamer (1984). Another effect of the EIV correction, but not obvious from Table 1 since the full results are not reported, is that variables other than the variable measured with error are affected, usually adversely, upon making the required EIV correction. This point is underscored in the application of Klepper et al. (1993).

In order to examine Hypotheses 2 about the bargaining equilibrium postulated by Grossman and Helpman (1995), we estimate a stochastic version of Eq. (2) which we report in Table 2. The Grossman–Helpman hypothesis is the joint hypothesis $\xi>0$ and $\xi^*>0$. A weaker hypothesis is that either one of the inequality holds. For our controls, we use just four industry-group dummies (Food, Resource-intensive, Manufacturing, Capital-intensive), described in Note 4 of Table 2, so as to give equal and symmetric treatment to US and Japan NTBs.[11] First consider the nature of the inferences. The only result that conforms to the GH prediction is the positive estimate on ξ^* from Model 2 (with controls) with dependent variable $(1+P_i)/(1+N_i^*)$. Otherwise the results go against the Grossman and Helpman predictions. What are we to infer about the Grossman and Helpman theory upon which the prediction is based? Probably not much. The results may mostly indicate that as of 1983, US–Japan NTBs did not display the characteristics of a bargaining equilibrium. That is probably true since US–Japan trade relations at the time, as now, could be better described as a trade war and NTB escalation rather than a period of compromise. A test of the theory will probably have to wait until after the effects of the 1988 Omnibus Trade Act, which allows the US greater bargaining strength through the delivery of stronger and more credible threats, work themselves out. Now consider the effect of the EIV correction using Eq. (4). Here, the effect of the EIV correction is different from the effect on the method of moment (MOM) estimates of Table 1. In Table 2 the use of the EIV-corrected variable leads to larger coefficients with higher t-values. The methodology used for the nonlinear model is due to the fact that the method for Hypothesis 1 cannot be employed here. Since this model is nonlinear in parameters, the MOM computations are difficult, sometimes even impossible, to perform. The price paid in order to use this simpler methodology is that the estimates are biased. The

[10] Perhaps the results based on quantitative NTBs may be an indication that coverage ratios are not adequate measures of the restrictiveness of quotas and voluntary export restrictions that are the main quantitative instruments used by the US. The costly construction of tariff equivalents for individual industries may be required if more precise inferences are required.

[11] The set of political economic control variables is available only for the US and not Japan. Hence industry dummies are used. The measures of fit indicate an adequate fit for the cross-sectional data.

Table 2
Hypothesis 2. Grossman and Helpman (1995): $\xi>0$, $\xi^*>0$ US-Japan bilateral NTB coverages of their 4-digit SIC imports

$\ln(\tau_i/\tau_i^*) = \ln[1 - \xi^*(z_i^*/e_i^*)] - \ln[1 - \xi(z_i/e_i)] + D\delta + \xi_i$ (regression model)
$E_i = e_i + u_i$ (measurement error model)

	With errors in variables correction				No errors in variables correction			
	US price NTBs-to-Japan's ALL NTBs: $\ln[(1+P_i)/(1+N_i^*)]$		US quant NTBs-to-Japan's ALL NTBs: $\ln[(1+Q_i)/(1+N_i^*)]$		US price NTBs-to-Japan's ALL NTBs: $\ln[(1+P_i)/(1+N_i^*)]$		US QUANT NTBs-to-Japan's ALL NTBs: $\ln[(1+Q_i)/(1+N_i^*)]$	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
ξ^*	−2.98*	−0.941	−8.420**	−5.791**	−1.626	−0.151	−5.912	−3.131**
	(1.740)	(0.682)	(3.644)	(2.962)	(1.566)	(0.194)	(1.436)	(2.889)
ξ	−3.432**	0.668*	−6.131**	−2.297**	−2.035**	0.042	−3.668**	−0.964**
	(4.989)	(1.609)	(7.038)	(2.821)	(4.591)	(0.156)	(6.312)	(2.281)
Constant	Yes	No	Yes	No	Yes	No	Yes	No
Dummies	No	Yes	No	Yes	No	Yes	No	Yes
N	247		247		247		247	
k (incl. constant)	3	6	3	6	3	6	3	6
R^2	0.115	0.355	0.258	0.398	0.109	0.345	0.218	0.393

Notes:
(1) Asterisks on variables denote Japan variables. See Notes 1–3 below Table 1.
(2) Nonlinear least-squares (NLS) estimates. EIV correction to e_i as in Eq. (5) done before NLS estimation.
(3) Absolute t-values in parentheses. ** and * signify statistical significance at 1% and 5% (one-tailed test), respectively. k includes constant.
(4) Four industry group dummies are included in D as follows: *Food Processing* [SIC=21], *Resource-intensive* [SIC=21 (Tobacco), 22 (Textiles), 23 (Apparel), 24 (Wood), 25 (Furniture), 26 (Paper), 27 (Printing), 31 (Leather), 32 (Glass)]; *Manufacturing* [SIC=33 (Primary metal), 34 (Fabricated metal), 35 (Machinery), 36 (Electrical), 37 (Transport), 38 (Instruments), 39 (Misc.)]; *Capital-intensive* [SIC=28 (Chemical), 29 (Petroleum Refining), 30 (Rubber)].

argument for its use is that this bias can be significantly lower than if the correction were not made.[12] Given the large range of sample values for E_i, we presume that the bias is substantially reduced due to the correction.

4. Conclusion

The need to use estimates from separate studies as regressors often arises. Such regressors need to be modelled as variables measured with error. The innovative models of Grossman and Helpman (1994), (1995) make simple testable predictions, but since import elasticities across industries figure prominently in their predictions elasticity estimates from a separate study need to be used in place of the true elasticities. We demonstrate how to deal with this problem of using generated regressors in (i) a linear setting and (ii) a nonlinear setting.

[12] Fuller (1987, p. 268) provides an estimator for the bias.

Acknowledgments

I am indebted to Wayne Fuller for his generous and helpful suggestions. All remaining errors are mine.

Appendix 1

Data

For data construction and sources refer to the appendix in Gawande (1995).

References

Baldwin, R.E., 1986. The Political Economy of US Import Policy. MIT Press, Cambridge, MA.
Fuller, W.A., 1987. Measurement Error Models. Wiley, New York.
Gawande, K., 1995. Are US nontariff barriers retaliatory? An application of extreme bounds analysis in the Tobit model. Review of Economics and Statistics, forthcoming.
Grossman, G.M., Helpman, E., 1994. Protection for Sale. American Economic Review 84 (4), 833–850.
Grossman, G.M., Helpman, E., 1995. Trade wars and trade talks. Journal of Political Economy 103, 675–708.
Klepper, S., Kamlet, M.S., Frank, R.G., 1993. Regressor diagnostics for the errors-in-variables model – An application to the health effects of pollution. Journal of Environmental Economics and Management 24, 190–211.
Klepper, S., Leamer, E.E., 1984. Consistent sets of estimates for regression with error in all variables. Econometrica 55, 163–184.
Leamer, E.E., 1990. The structure and effects of tariff and nontariff barriers in 1983. In: Jones, R.W., Krueger, A. (Eds.), The Political Economy of International Trade: Essays in Honor of Robert E. Baldwin. Basil Blackwell, Cambridge, MA.
Sheills, C.R., Stern, R.F., Deardorff, A.V., 1986. Estimates of the elasticities of substitution between imports and home goods for the United States. Weltwirtschaftliches Archiv 122, 497–519.