Advance Access publication June 24, 2009

Political Analysis (2009) 17:236–260
doi:10.1093/pan/mpp009

Dealing with Weak Instruments: An Application to the Protection for Sale Model

Kishore Gawande
Bush School of Government and Public Service, Texas A&M University, College Station, TX 77843-4220
e-mail: kgawande@tamu.edu (corresponding author)

Hui Li
Department of Economics, Eastern Illinois University, 600 Lincoln Avenue, Charleston, IL 61920
e-mail: hli@eiu.edu

Endogeneity of explanatory variables is now receiving the concern it deserves in the empirical political science literature. Instrumental variables (IVs) estimators, such as two-stage least squares (2SLS), are the primary means for tackling this problem. These estimators solve the endogeneity problem by ''instrumenting'' the endogenous regressors using exogenous variables (the instruments). In many applications, a problem that the IV approach must overcome is that of weak instruments (WIs), where the instruments only weakly identify the regression coefficients of interest. With WIs, the infinite-sample properties (e.g., consistency) used to justify the use of estimators like 2SLS are on thin ground because these estimators have poor small-sample properties. Specifically, they may suffer from excessive bias and/or Type I error. We highlight the WI problem in the context of empirical testing of ''protection for sale'' model that predicts the cross-sectional pattern of trade protection as a function of political organization, imports and output. These variables are endogenous. Importantly, the instruments used to solve the endogeneity problem are weak. A method better suited to exact inference with WIs is the limited information maximum likelihood (LIML) estimator. Censoring in the dependent variable in the application requires a nonlinear Tobit LIML estimator.

1 Introduction

Controlled experiments in political science in which covariates are exogenously varied in order to observe the response on the dependent variables are rare. Endogeneity of explan-

Authors' note: We thank two anonymous referees, the editor (Christopher Zorn), and an associate editor (Robert Franzese) for insightful comments that have improved the paper. Responsibility for any remaining errors is ours. Kishore Gawande acknowledges financial support from the Helen and Roy Ryu Chair at the Bush School. Data set and programs are available on the POLMETH Web site.

atory variables should be recognized as a major problem in empirical work in all fields of political science. The use of methods to deal with endogeneity in empirical work in the influential journals indicates that this problem is beginning to receive the attention it should.

In the presence of endogeneity, ordinary least squares (OLS) are biased in small samples, and this bias persists in large samples. The primary means of tackling the endogeneity problem in the economics and political science literature is to use instrumental variables (IVs) estimators such as two-stage least squares (2SLS). In this method, exogenous variables (instruments) are used to rid the regressors of the endogeneity problem. These instruments "identify" the coefficient of interest in a regression model (more on this below). It is a safe prediction that the use of IVs in empirical work in political science will escalate as endogeneity becomes widely recognized as a serious impediment to making credible empirical inferences. Our survey in Section 2 indicates that this process is underway.

To begin, consider the simple regression model:

$$y = Y\beta + \in, \tag{1}$$

where y is the dependent variable, Y is the single regressor, and β is the single parameter of interest. There are available T observations on y and Y. Thus, they are $T \times 1$ vectors, as is the error vector ϵ. Each element of ϵ has the usual statistical properties—it is independently, identically, normally distributed. In this "structural" model, Y is an endogenous regressor. Endogeneity of the regressor means that Y and the error term ϵ are correlated. Thus, a random shock to the dependent variable y also affects Y. As the model (1) stands, the parameter β is "unidentified" in the absence of any further information. In order to identify β, suppose a reduced-form model for the endogenous regressor that predicts Y using exogenous variables Z may be specified as:

$$Y = Z\Pi + \upsilon, \tag{2}$$

where Z is a $T \times K$ matrix consisting of T observations on K variables. These K variables Z are uncorrelated with the error term υ in equation (2), making them exogenous in equation (2), and also uncorrelated with the error term ϵ, clearing the way for their use as exogenous instruments to solve the endogeneity problem in equation (1).[1]

[1]The textbook specification of a full simultaneous equation system may be used to motivate the reduced form. But that is not necessary, so long as it is convincingly argued that the variables Z do not explain y, that is, they do not have any role as regressors in equation (1). A simultaneous equation system would comprise equation (1) and another equation in which the dependent variable Y is "explained" by the endogenous regressor y and the exogenous regressors Z:

$$Y = ya + Z\Delta + \mu,$$

where Z is the same matrix of exogenous variables uncorrelated with the error term μ, and α and Δ are coefficients, respectively, on the regressors y and Z. Substituting out y and solving for Y, we get the reduced form (equation 2). In equation (2), Π is a function of the structural equation parameters β, α, and Δ.

Since Y and ϵ are correlated, OLS estimates of β using equation (1) are biased in small samples and inconsistent in large samples. In contrast, instrumental variables estimators such as the 2SLS estimator are consistent. Hahn and Hausman (2002b, 238) provide expressions for the expected bias of OLS and 2SLS estimators. The sample counterparts of those expressions may be used to compute the bias of these estimators in finite samples.[2]

The identification issue is standard fare in econometrics textbooks (e.g., Greene 2000), but its textbook solution is limited to rank and order conditions, which are both necessary and sufficient for uniquely recovering β from the reduced-form parameters Π. In equations (1) and (2), for example, the existence of a single instrument solves the identification problem, at least in theory. In practice, for instruments to be valid, they must be defended as being exogenous, that is, uncorrelated with the error term ϵ. The main concern in this paper goes beyond identification and exogeneity of the instruments. We are concerned with their small-sample bias, a property that used be brushed away using asymptotics. The simple truth is that the majority, if not most, samples in political science and economics are finite and usually small. Making inference from small samples by appeal to large-sample methods is hopeful at best and usually a leap in the dark. In order to make credible inferences using finite samples, it is important for estimators to possess good small-sample properties, primarily low bias. To achieve this, instruments (e.g., Z in equation [2]) must be strongly correlated with the endogenous regressor. In practice, however, instruments are likely to be "weak," which leads to questionable inferences and invalid testing. Weakness of instruments refers to the *relevance* of instruments, as measured by the correlations of the instruments Z with the endogenous regressor Y. We say that Z fails to identify β when the reduced-form parameter $\Pi = 0$. This is the case of nonidentification. *Weak* identification is the case when Π is "close" to zero or approaches zero as the sample increases (see, e.g., Bartels 1991; Hahn and Hausman 2002b). It is this case with which we are concerned.

If the correlations of Z with Y are weak, their use afflicts testing and inference about β and may render them invalid. Bound et al. (1995) made this point forcefully by showing that replacing a measured instrument by a randomly generated instrument led to the same inference about the structural parameter in the influential study by Angrist and Krueger (1991). That essentially rendered the original inference

[2]In the framework of equations (1) and (2), the expected bias of the 2SLS and OLS estimators may be approximated as (Hahn and Hausman (2002b, equations [7] and [8]):

$$E(\text{Bias}_{\text{2SLS}}) \approx \frac{K.\sigma_{\in\upsilon}}{R_f^2 \cdot \sum_{i=1}^{T} Y^2},$$

and

$$E(\text{Bias}_{\text{OLS}}) \approx \frac{\text{cov}(Y,\in)}{\text{var}(Y)}.$$

As T increases, $\sum_{i=1}^{T} Y^2$ becomes large and the expected bias of the 2SLS estimator decreases (to zero in the limit). This is not true for the expected bias of the OLS estimator since the denominator remains unchanged as T increases. Even so, the small-sample bias of the 2SLS estimator may be substantial. The greater the first-stage R_f^2 (of the reduced-form model [equation 2]), the lower is the expected bias, whereas the higher is the covariance between ϵ and υ or the number of instruments K, the greater is the expected bias (for any given data realization of Y).

invalid. The problem, of course, was that the measured instrument was irrelevant in the first place.[3]

What, then, is to be done? The intent of this paper is to demonstrate that (1) the weakness of instruments is diagnosed easily, and in fact should be required of empirical studies that feature structural models, (2) the limitations of standard estimators like 2SLS should be clarified in the presence of the weakness of instruments (as demonstrated by Bartels 1991), and (3) the weak instrument (WI) problem may be greatly reduced by using estimators that are more robust to the problem.

The diagnosis of WIs and statistical properties of estimators using WIs were first set forth in Anderson and Rubin (1949). The newer literature that develops and advances those ideas, and from which we borrow, includes Stock and Yogo (2004), Moreira (2003), Stock, Wright, and Yogo (2002), Kleibergen (2002), Hahn and Hausman (2002a), Staiger and Stock (1997), and Nelson and Startz (1990). This literature contends that when the correlation between the endogenous regressors and instruments is low, the asymptotic approximation of the limiting distribution of an IV estimator may be substantially different from its small-sample distribution. Simulations convincingly demonstrate that conventionally used asymptotics (i.e., large-sample theory used to justify the use of IVs) generally yield extremely poor approximations to the exact (i.e., small sample) distributions. That the divergence of the asymptotic approximation from the exact distribution may lead to highly misleading inference about the structural parameters is the main thrust of the argument to use methods that (1) yield better approximations, (2) allow transparency of inferences in terms of the expected bias involved in the use of one method versus another, and (3) allow the use of diagnostics, for example, size of the test that one is willing to tolerate with WI when testing using an asymptotic approximation. An emerging consensus from this literature is the preference for the use of limited information maximum likelihood (LIML) estimators over 2SLS. The case for alternate estimators is even more compelling if policy were to be based on inferences from small samples generated by quasi-natural political and economic experiments.

The vehicle we use to highlight and solve the WI problem is testing the ‘‘protection for sale’’ model by Grossman and Helpman (1994, henceforth GH). It predicts the cross-sectional pattern of protection as a function of measurable political and economic variables. The Grossman-Helpman (GH) model delivers a clearly testable prediction about the cross-sectional structure of tariffs. It has, thus, attracted much empirical attention. Among the most direct studies of the GH models are Goldberg and Maggi (1999), Feenstra and Branstetter (2002), Mitra, Thomakos, and Ulubasolglu (2002), Eicher and Osang (2002), Facchini, Van Biesebroeck, and Willmann (2003), Bombardini (2008), McCalman (2004), and Gawande and Bandopadhyay (2000). The GH model formalizes special interest behavior in a specific manner, but the relevance of special interests is already well explored in empirical work in political science. McGillivray (1997), Hiscox (2002), and Reinhardt and Busch(1999), for example, have used trade policy data

[3]The *validity* or exogeneity of instruments is presumed here, so we can focus on the WI problem. That is, prerequisite for any structural estimation and the validity of instruments must be established by authors of empirical work using political logic. That is, the burden is on the authors to convince the readers that the instruments are uncorrelated with the error u. Solving this problem is no mean task and requires care.

to investigate the influence of special interests in complex institutional settings. Their empirical explorations are not restricted by any one formal model but rather motivated by more than one theory. In this sense, their specifications are *ad hoc* but the lessons they teach us are rich with evidence and exploratory inference. It is neither the intention here to weigh the pros and cons of the two exercises nor is it our claim that the GH94 model is representative of the broader political economy of trade policy literature. On the contrary, they are both valuable. The validity of the GH model is sought by modelers wishing to apply it theoretically and empirically in other contexts—the GH94 model has become the basis for positive theories in a variety of trade policy contexts including regional trade agreements, trade bargaining, and trade institutions. The validity of less formally derived hypothesis is valuable precisely because the complexity of institutions they study are difficult to model formally, and the exploratory hypotheses they empirically precede, indeed, motivate new theoretical advances.

The GH model is apt for studying the WI problem. Crucial to the estimation of the key model parameter is solving an endogeneity problem. The problem is that the theoretically valid instruments that are available to solve the endogeneity problem are empirically weak. A more reliable method than 2SLS—which suffers from a potentially unacceptable degree of bias—is the LIML estimator. While the LIML estimator is standard in econometric software, since our protection data are censored, they must be modeled nonlinearly. A methodological contribution of the paper is to estimate a Tobit LIML model. Our hope is that this paper motivates LIML estimation of nonlinear probit, logit, count data, and multinomial models in relevant applications, so that their use becomes routinized as they become recognized to be robust alternatives to their two-stage counterparts. A lesser objective, but one that may appeal to consumers of this literature, is to update the results from an older benchmarked data set to one from the 1990s that better represents the current structure of trade barriers in the United States.

The paper proceeds as follows. Section 2 motivates the WI problem by indicating a developing trend in the flagship political science journals toward recognizing the endogeneity problem and taking the necessary steps to solve it. Section 3 lays out the GH model and its predictions. Section 4 describes the new data used to test this model. In Section 5, the problem of WI is explained and resolved. Since the available instruments are weak (but strong enough to allow inferences with the appropriate methods), the structural parameter in the GH model is estimated using the LIML estimator, inference from which is less sensitive to the WIs problem than it is from the 2SLS estimator. In Section 5, we discuss these results. Section 6 concludes.

2 The State of Structural Modeling in Political Science

An encouraging trend has been developing in empirical work published in the top political science journals toward recognizing and dealing with the endogeneity problem. Between 2002 and 2005, papers recognizing endogeneity were few and far between. In order to assess the quality of their instruments, we requested many of those authors for their data. Lassen (2005), Rudra (2005), and Calvo and Murillo (2004) generously complied with our request. In all three studies, we find the quality of their instruments to be fair-to-very good. We take from this that users of IV estimators (at least until 2005) failed to report quality-of-instrument diagnostics not because their instruments were weak but because they did not recognize it as a potential problem.[4]

In the 2006 volumes of *American Journal of Political Science*, only two papers—Bawn and Rosenbluth (2006) and Prakash and Potoski (2006)—took steps to solve the endogeneity problem. Even though a number of papers explicitly mentioned endogeneity to be a problem, they used OLS to estimate their model parameters anyway. An argument in favor of doing so is that instruments are hard to find. This is particularly true for studies that use primary surveys of individual attitudes and behavior but is also a problem for studies using well-established databases like censuses. A learning from the literature on endogeneity in general, and the WI problem in particular, is that the least that those studies should do is to indicate the extent of the bias in their OLS estimates, making the imperfection transparent to the reader. Some articles indicate that it is better to estimate using OLS rather than using poor (i.e., weak) instruments due to the bias they impart to the IV estimates. That is precisely a point this paper makes. The extent of that bias should be made clear (using, e.g., the Hahn-Hausman approximations).

Beginning in 2007 are signs of change. The 2007 and 2008 (until June) volumes of the *American Journal of Political Science* contains papers by Gabel and Scheve (2007), Konisky (2007), Lebo, McGlynn, and Koger (2007), and Lewis-Beck, Nadeau, and Elias (2008) that tackle endogeneity using IVs methods. Although Lebo, McGlynn, and Koger (2007, footnote 21) leave it to the reader to gauge the weakness of instruments from the Anderson test statistic, their attempt to convey that information is commendable. Although Lewis-Beck et al. (2008, 90) indicate that their instrument correlates "well" with the endogenous regressor, reporting the *partial* correlation would have been a far superior

[4]We computed the first-stage F-statistics for each of the three papers. We will show in Section 4.2 below how the first-stage F-statistic proves to be a useful diagnostic of the WI problem. Intuitively, it measures the strength of the correlation between the instruments and the endogenous regressor. Staiger and Stock's thumb rule of $F > 10$ is an informal cutoff rule for demonstrating the strength of instruments. Note again that our intent is not to critique the choice of instruments. The performance of that important role is the prerogative of researchers specialized in those areas and must be undertaken if research is to move theory and learning forward. Our intent is to judge the quality of instruments on the face of evidence presented.

Lassen's (2005) first-stage F-statistic for his second (first) model shows that his single (four) instrument has a partial R^2 of just .0099 (.0175), but since his sample consists of over 2000 observations, it yields an F-statistic of 18.53 (8.81), engendering confidence in the second-stage structural estimates.

Calvo and Murillo (2004) estimate their model using three-stage least squares (3SLS), a full-systems method. Our estimation of their models by 2SLS yields a partial R^2 for their 10 instruments of 0.46. With a sample consisting of 96 observations, this yields a first-stage F-statistic of around 8. The statistics imply that the instruments are not weak (but not too strong either). We note that our 2SLS estimates of the structural parameters were not as strong or precise as the authors find from their 3SLS procedures. Using 3SLS requires a serious commitment to the exclusion restrictions in each equation in the system and must be theoretically motivated and justified in order to prefer their use over a 2SLS procedure. Calvo and Murillo do justify the use of their recursive system, and we leave it to future research to corroborate their inference.

Rudra (2005) uses panel data and fixed-effects IV methods to make inferences about the structural parameter in her model. The instrument used is actually the squared deviation of the endogenous regressor itself from its overall mean. This is justified using Lewbel (1997). A problem with using the overall sample mean (as opposed to *country* means) to construct the instrument is that the instrument can then have a few very influential points (namely countries with very low and very high per capita incomes), which then determines the fitted values used in the second stage. If the fitted values also inherit those influential values, the estimate on the structural parameter is then determined by a handful of extreme observations, not representative of the sample as a whole. In Rudra's case, this is ameliorated by the presence of (1) other variables in the first-stage regressions and (2) a large sample. This problem may be solved using the squared deviation of the endogenous regressor from the *country* means. However, this instrument turns out to be quite weak by our calculations. Although theoretically (i.e., econometrically) valid for identification, Lewbell's procedure must deal with the extreme values issue on the implementation side. Perhaps, for that reason, this (easy) solution has not caught on, at least in the economics literature. That said, the Lewbell instrument (using overall mean deviation) did yield strong first-stage statistics. Since we did not have Rudra's code, we did not attempt to replicate her second-stage results.

indicator of instrument quality. Konisky's paper is notable for indicating the correct WI diagnostics from the first stage: the partial R^2 of the instruments as well as the first-stage F-statistic. Although he uses fixed effects, his instruments turn out to be strong and afford sturdy inferences. The Gabel-Scheve paper is a model that future papers should emulate in carefully thinking through the identification strategy. Their choice of instruments is econometrically sound (whether it is politically sound is left to specialists in the area to assess). They report the first-stage F-statistic, which is quite weak (Gabel and Scheve 2007; Table 1).[5] They recognize the weakness of their instruments (see their footnote 16) and indicate that results from using other stronger instruments are similar to what they report in the paper.

What is notable about this 2007–08 set of papers is that they report a number of diagnostics about their instruments, particularly overidentification tests. We encourage reporting about the weakness of instruments as well. With the availability of more panel data, fixed-effects estimators are becoming prevalent. Instruments can be particularly weak in panel data when fixed effects are used in estimation since fixed effects allow the use of only within variation in the data to estimate the structural parameters. Decomposing the variation in the instruments often shows that the lion's share is accounted by the cross-sectional variation, which is absorbed by the fixed effects. The remaining within variation may be too small to make for strong instruments.[6] This does not mean that panel methods that use cross-sectional variation as well, for example random-effects, should be preferred (that choice must be grounded in an argument that has nothing to do with the quality of instruments). Rather, our point is to encourage researchers to spend more effort on thinking through the endogeneity problem, to take time and effort constructing the right instruments, and regardless of the result, to report the strength or weakness of their instruments. If instruments are weak, then the study should indicate the extent of bias in the estimates. The Gabel-Scheve paper is a model in this regard.

Seriously consideration of the WI problem requires techniques that are robust to the problem. The remainder of the paper is used to demonstrate the use of a robust technique in the presence of WI. We will test the GH model using new data from the 1990s. The results are themselves of interest to readers of trade politics. The bigger message is methodological. The application is meant to exemplify the endogeneity problem and demonstrate (1) the diagnosis of the quality of instruments and (2) a solution to the problem of WI. Our solution—to use limited information methods—is generally valid for a wide range of empirical applications in political science.

3 The Econometric Model

The Grossman and Helpman (1994) model is a theory of how trade policy is set by a government that maximizes a weighted sum of welfare (W) and lobbying contributions (C)

[5]Using the diagnostics in Stock and Yogo (2004), we surmise that the F-statistic of 3.08 with four instruments leaves their IV estimates with a bias that is approximately 40% of the bias of the OLS estimate.

[6]Take, for example, a study that uses repeated observation over time for each country. Including country-fixed effects exploits variation within each country over the time period of the study to make inferences. Instruments such as *per capita* income, for example (assuming *per capita* income does not belong in the structural model [equation 1]), may prove weak since the within variation may be low for the period of the study. Acemoglu et al. (2008) show that country-fixed effects eliminate the statistical significance of income as a cause of democracy. If this finding is due to the fact that controlling for fixed effects reduces the variation in the income data, it implies that income may prove to be a poor instrument in panel studies with country-fixed effects.

Table 1 Variables and descriptive statistics

	Description	*Mean*	*SD*
$N/(1+N)$	Dependent variable: N = NTB coverage ratio	0.186	0.194
I_{10} H (z/e)	I_p = 1 if PAC spending is in top (100!p)th percentile of sample; z = Output/imports; e = absolute import elasticity	6.717	8.360
I_{25} H (z/e)	I_p = 1 if PAC spending is in top (100!p)th percentile of sample; z = Output/imports; e = absolute import elasticity	5.867	8.15
I_{50} H (z/e)	I_p = 1 if PAC spending is in top (100!p)th percentile of sample; z = Output/imports; e = absolute import elasticity	3.772	7.167
HERF	Herfindal index of firm concentration	0.078	0.066
HERFSQ	Square term of VAHERF	0.010	0.016
PRODWORKER	Fraction of employee whose occupation is production, 2000 Census	47.72	11.36
PRODWORKERSQ	Square term of PRODUCT	2412	1114
FOOD	Dummy: Food Manufacturing and Beverage and Tobacco Product Manufacturing	0.105	0.308
TEXTILES	Dummy: Textile Mills and Textile Product Mills	0.055	0.228
APPAREL	Dummy: Apparel Manufacturing and Leather and Allied Product Manufacturing	0.086	0.281
WOOD	Dummy: Wood Product Manufacturing and Paper Manufacturing	0.059	0.235
PETR&CHEM	Dummy: Petroleum and Coal Products Manufacturing and Chemical Manufacturing	0.078	0.269
NONMETALS	Dummy: Plastics and Rubber Products Manufacturing and Nonmetallic Mineral Product Manufacturing	0.090	0.287
METALS	Dummy: Primary Metal Manufacturing and Fabricated Metal Product Manufacturing	0.121	0.327
MACH	Dummy: Machinery Manufacturing	0.156	0.364
ELECTRONIC	Dummy: Computer and Electronic Product Manufacturing	0.090	0.287
ELECTRICAL	Dummy: Electrical Equipment, Appliance, and Component Manufacturing	0.047	0.212
AUTOS	Dummy: Transportation Equipment Manufacturing	0.055	0.228
MISC	Dummy: Miscellaneous Manufacturing	0.059	0.236
WAGE	Annual compensation per employee ($ million)	0.039	0.011
K/L	Capital-labor ratio ($ million per worker)	0.082	0.101

Note. Sample has 256 NAICS six-digit industries. All data are from 1995. NTBs (N) include price control, quantity control, and technical control measures.

made by industry lobbies.[7] Human capital is specific to a sector, so that the return to human capital increases with the price of the good produced by that sector. Since protection assures higher prices than would prevail under free trade, owners of sector-specific human

[7]Although the policy in question is trade tariffs (and subsidies), the model is applicable to a number of redistributive polices.

capital organize into lobbies with the purpose of bending policy in order to increase the price of the goods they produce. The government is cognizant of the welfare costs (deadweight losses) protectionist policy will impose on the public. At the same time, lobbying contributions help government defray a variety of expenditures designed to keep it in power.

Thus, government maximizes the political welfare function $aW + C$ when deciding how much protection to supply to a sector that lobbies for it. W is the public's total welfare, whereas C is the money contributions that the government gets from lobbies. The parameter a is the terms at which the government trades off a dollar of welfare loss for a dollar of contributions. This parameter is the focus of interest. It measures the susceptibility of the government to "sell out" its public and the terms at which it does so. The GH model has proved attractive for empirical research because it delivers a clear structural prediction that may be used to estimate the parameter a.

The model predicts that the cross-industry pattern of protection, when the government unilaterally and optimally chooses tariffs based on such a political welfare function, is a function of three variables: the inverse of the import-penetration ratio, the (absolute) elasticity of import demand, and a variable indicating whether an industry is politically organized. Specifically (see, e.g., Grossman and Helpman 1994, Goldberg and Maggi 1999, or Gawande and Bandopadhyay 2000 for the analytical derivation) the prediction is given by

$$\frac{t_i}{1+t_i} = \frac{I_i - \alpha_L}{a + \alpha_L}\left(\frac{X_i/M_i}{e_i}\right). \; i = 1, \ldots, n. \tag{3}$$

In equation (3), the subscript i denotes industry i and the dependent variable $t_i = \left(p_i - p_i^0\right)/p_i^0$ is the *ad valorem* tariff for good i, where p_i is the domestic price for good i and p_i^0 its world price. In the first term on the right hand side I_i is an indicator variable that equals one if producers of good i are politically organized. α_L is the proportion of the population that is organized into lobbies. X_i/M_i is the equilibrium ratio of output to imports (the inverse import-penetration ratio) and $e_i = -M_i^{'} p_i / M_i$ is the absolute elasticity of import demand. Thus, if producers of good i are organized ($I_i > 0$), they are able to "buy" protection ($t_i > 0$). Industry i is protected only if it is organized, but not otherwise. Since $\alpha_L < 1$, political organization confers positive protection.

At issue is the recovery of the political economy parameter a—the relative weight the government places on a dollar of lobbying contributions versus a dollar of welfare loss that protection inflicts on the public. The greater is a, the more welfare minded is the government; the lower is a, the more cheaply its redistributive powers are "bought" by special interests.

The intuition behind equation (3) is attractively simple. Industry output X_i captures the size of rents from protection. Imports determine the extent of welfare losses from protection, so the smaller are imports and the higher is the tariff. In any general equilibrium model of tax policy, the Ramsey pricing rule is that the lower is the demand elasticity, the higher should be the tax since they are less distortionary. This logic is inherent in equation (3)—the lower is the absolute elasticity e_i the higher is the tariff predicted to be, all else equal. The cross-sectional structure of protection in equation (3) has been the subject of empirical inquiry in a series of recent studies, including Goldberg and Maggi (1999), Mitra, Thomakos, and Ulubasolglu (2002), McCalman (2004), Eicher and Osang (2002), and Gawande and Bandopadhyay (2000).

In what follows, we will assume that a negligible fraction of the population is organized as lobbies. This simplifies the model's predictions. The main implication of this assumption is that lobbies lobby for protection of the good they produce but not against protection of other good that they consume. This accords well with reality.[8] With this assumption, the prediction simplifies to

$$\frac{t_i}{1+t_i} = \frac{I_i}{a}\left(\frac{X_i/M_i}{e_i}\right), \; i = 1, \ldots, n. \tag{4}$$

A burden that the use of social science data imposes on econometric analysis is the problem of regressor endogeneity. In this instance, the joint determination of protection and the regressor $I_i \times \frac{X_i/M_i}{e_i}$ raises concerns about reverse causality, for example. Clearly tariffs reduce imports and increase domestic production, thereby decreasing import penetration (or increasing z). The vast literature on the political economy of protection predating the GH model has attempted to deal with this problem. Political organization I_i may be endogenously determined with protection as well. Mitra's (1999) theoretical model, for example, puts this endogeneity at the center of his argument about lobbying organization.

In the next section, we attempt, in the context of the protection for sale model, a comprehensive solution, the endogeneity problem. Our treatment is relevant to a broad class of applications in political science and political economy.

4 Data and Methodology

Empirical estimation proceeds with a stochastic version of equation (4)

$$\frac{t_i}{1+t_i} = \beta\left(I_i \times \frac{X_i/M_i}{e_i}\right) + \in_i, \; i = 1, \ldots, n, \tag{5}$$

where $\in_i$ is independently and identically distributed across all i. The parameter of interest is $\beta = 1/a$. From the estimate of β, a can be recovered as $1/\beta$.

4.1 *Data*

4.1.1 Dependent variable

Cross-industry data from 1995 at the six-digit North American Industrial Classification System (NAICS) level of aggregation is used for 256 manufacturing industries in this study. Although past studies of the GH model using U.S. data have all relied on the well-benchmarked 1983 data set (e.g., Gawande and Bandopadhyay 2000; Goldberg and Maggi 1999), this is the first study to test the GH model against more recent U.S. data. As argued in preceding papers, unilaterally imposed protection is the appropriate data used to test the GH model. Tariffs in manufacturing have been multilaterally determined since the Kennedy round cuts of the 1960s. As a result, *ad valorem* tariffs are unsuitable for testing a prediction from a model in which protection is unilaterally imposed. Further, as tariff levels have declined in manufacturing, nontariff barrier (NTB) have mushroomed. For these reason, as in past studies, we use the NTB coverage ratio to measure protection. That is, we measure the proportion of imports covered by all NTBs. Admittedly an imperfect measure of the restrictiveness of NTB protection, it is nevertheless the best available measure at the scope of this study. Computing *ad valorem* equivalents of NTBs is

[8]At the same time, this assumption does not preclude lobbying by downstream users against protection to their upstream suppliers since protection raises their input costs.

no easy task, for the heterogeneity in the types of NTBs is considerable. Further, computing such equivalents requires making questionable assumptions about parameters that govern demand and supply. Coverage ratios are, thus, a practical alternative and have been used in many new studies of protectionism. Thus, the dependent variable in (5) $t_i/(1 + t_i)$ is measured as $N_i/(1 + N_i)$, where N_i is the NTB coverage ratio.

The NTB coverage ratios were computed for the year 1995 from the Trade Analysis and Information System (TRAINS) database put out by United Nations Conference on Trade and Development (UNCTAD).[9] TRAINS contains indicators of bilateral NTBs at the eight-digit Harmonized System (HS) level of over 6000 commodities. It identifies seven types of NTBs. The overall NTB coverage ratio of U.S. imports from all other countries is based on the union of the indicators across these NTBs.[10]

4.1.2 Regressor

Measures of political organization by producers, I_i, have been carefully constructed using lobbying data from the 1991 to 92 and 1993 to 94 election cycles. A mapping from firms to industry constructed by Beaulieu and Magee (2004) was used to aggregate corporate political action committee (PAC) contribution data (downloaded from the Federal Election Commission Web site, http://www.fec.org) to standard industrial classification (SIC) industry-level contributions, which was then concorded into the NAICS level of this study. PAC contributions by multiproduct firms who are active in more than one industry were fractionally mapped equally across the industries in which they produced. Thus, the mapping is sum preserving.

All industries made positive contributions to congressional election campaigns. However, since PAC contributions target a basket of policy instruments, of which trade policy may be crucially important for some sectors but not for others, it may be erroneous to let $I_i = 1$ for all industries. We solve this problem in the manner of Goldberg and Maggi, who present their results across a set of measures of I_i using thresholds. Three thresholds are used as follows: The percentile distribution of PAC spending was first determined. Three percentile thresholds in increasing order of expenditures per unit value added, at the 10th percentile, 25th percentile, and 50th (median) percentile, were used. For a given threshold, say the median, sector i was assigned $I_i = 1$, if that sector was above the median. Three sets of regressions corresponding to these thresholds are, thus, reported.[11]

The inverse import-to-output ratio $z_j = x_j/m_j$ is measured using data from the Annual Survey of Manufactures for 1995 on domestic production and imports. Import demand elasticities are taken from Gallaway, McDaniel, and Rivera (2003); 309 short-run elasticity estimates at four-digit SIC (1987 basis) level are concorded into the six-digit NAICS (1997 basis) level using the method described in Appendix A.

When imports are zero (largely the result of concording from trade-based data-keeping systems into domestic production-based data-keeping systems), the ratio $X_i/$

[9]A parallel set of coverage ratios has been computed by Jon Haveman, which we used to check our calculations. The ones used in this paper have been computed by us. Although there are some differences, they are minor.

[10]The U.S. Census Bureau concordance available at http://www.census.gov/foreign-trade/reference/codes/index.html#concordance was applied to aggregate the HS-level NTB indicators down to the six-digit NAICS lines, weighting by imports.

[11]It would be beneficial to amend the model so that in equation (5) the binary indicator of political organization is replaced with a continuous variable that measures the intensity of lobbying. This extension is not pursued in this paper.

M_i is undefined, which are dropped in the sample. Equally important are the cases where this ratio takes on extremely large values due to very small M_i, either because of the imperfect concordance between different data-keeping systems (trade is at the HS level and production at the NAICS level) or because these are predominantly exporting industries that are not interested in tariffs anyway.[12] These influential values can distort inference about β in equation (5). Our solution is to drop observations for which the ratio $X_i/M_i > 100$.[13] The availability of the elasticity and other manufacturing data yield a sample of 256 NAICS industries. They accounted for two-thirds of value added in U.S. manufacturing in 1995.

4.2 *Methodology: Endogeneity and WIs*

4.2.1 Endogeneity and instruments

The main challenge confronting consistent estimation of the parameter β in equation (5) is the endogeneity of the regressor $\left(I_i \times \frac{X_i/M_i}{e_i}\right)$. As described above, import penetration is endogenous as is the indicator for political organization. We instrument $\left(I_i \times \frac{X_i/M_i}{e_i}\right)$ using a set of variables (IVs) constructed from the 1994 and 1995 Annual Survey of Manufactures and the Census of Population. The Herfindahl index (HERF) is the IV used to identify the impact of domestic political organization, based on the theoretically and empirically well-established idea that the smaller the free-rider problem of organizing coalitions the more effective the coalition. The human capital embodied in direct labor, measured by the fraction of workers in the industry who are production workers (PRODWORKER), is used to instrument imports. Theoretically, the human capital measure is taken to be exogenous because it is technologically constant. The ratio of production-to-nonproduction workers changes mainly in response to technological innovations (e.g., a labor saving innovation that keeps overhead labor constant but reduces production workers per unit of output). If shocks to technology are uncorrelated with shocks to trade protection (the dependent variable), then the human capital measure is exogenous. Moreover, the human capital measure is a source of productivity and comparative advantage for manufacturing industries and is therefore correlated with imports M. The Herfindahl index is fairly constant over time and responds to technology shake-ups that alter market structure (e.g., new patents, inventions, regulations). If these shocks are uncorrelated with trade protection shocks, then the Herfindahl index is exogenous. Further, industry concentration is an important determinant of collective action (Olson 1965) and, hence, correlated with political organization I. For reasons described below, we also include the squares of these two variables (HERFSQ and PRODWORKERSQ) as instruments.

In addition to these four variables, we include the set of 12 industry dummies defined at the three-digit NAICS level of aggregation. The industry controls are, by definition, exogenous. We experimented with two specifications, one in which the three-digit dummies were included only as instruments and excluded from the structural equation and another in which the three-digit dummies appeared in both the structural and reduced-form equations. Both specifications yield qualitatively similar results of the structural equation parameters.

[12]Theoretically, they may be interested in lobbying for export subsidies, but we do not have that subsidy data. Exporting industries should therefore be dropped if the dependent variable measures only protection. See Gawande and Hoekman (2006) for tests of the model using agricultural tariffs and subsidies.

[13]This drops 31 observations from the sample.

In the paper, we report the results of the first experiment, where the industry dummies are treated as instruments.[14,15]

The NTB coverage ratios are nonnegative, but protection can in theory and practice be negative. For example, if foreign subsidies are not countervailed, then protection is effectively negative since domestic prices are driven below the domestic producer price. The theory does not rule out subsidization of imports, which implies negative protection. Thus, a Tobit specification that models a cut-point in the measured outcome (at zero) for a latent outcome that is continuous over the real line is appropriate for the NTB data. A 2SLS method for the Tobit model with endogenous regressors is described in Smith and Blundell (1986). This is implemented by including the residual from the (linear) first-stage regression in the structural model and then estimating the structural model as a Tobit. If the residual is statistically significant, then it indicates that the endogeneity problem is a concern. Including the residual term purges the endogeneity in the regressor and ensures the consistency of the coefficient on the endogenous regressor.

A complication with the use of this procedure (and other limited information procedures) is that the individual endogenous variables, specifically import penetration *M*/*X* and political organization *I*, are embodied nonlinearly in the endogenous regressor (equation 5). In order to consistently estimate β, we follow Kelejian (1971) and use higher order terms of the instruments in the reduced-form equation for the endogenous regressor $\left(I_i \times \frac{X_i/M_i}{e_i}\right)$. It is for this reason that HERFSQ and PRODWORKERSQ are used as additional instruments. The 2SLS estimate of β is based on this extension of the Smith-Blundell method.

4.2.2 Weak instrument diagnostics

The quality of instruments has important consequences for inference in structural models due to potential problems concerning identification and testability. Identification is the problem of distinguishing parameter values (β in equation [1]) on the basis of the data. Testability is concerned with designing procedures for clearly separating different subsets of parameter values (see, e.g., Greene 2000, chap. 16) for a textbook treatment of these issues). When parameters are close to regions where they are no longer identifiable, large-sample distribution theory breaks down and the use of large-sample approximations results in highly flawed tests. This situation is likely when instruments that are used to solve the identification problem are weak. Thus, for example, the 2SLS estimator, which is consistent and asymptotically efficient when instruments are adequate, is not only strongly biased in the direction of OLS but its distribution is far from normal when instruments are weak. Beginning with Anderson and Rubin (1949), the seriousness of this problem is underscored in a number of papers (e.g., Stock, Wright, and Yogo 2002, Stock and Yogo 2004, Moreira

[14]The other set of results is available from the authors. Our data, code, and results are also publicly available at the *POLMETH* data site.

[15]The Sargan test of overidentifying restrictions (OIRs), which tests for exogeneity of instruments (that is, $E(Z\epsilon) = 0$, where Z is the set of instruments and ϵ is the structural error term—see equations [1] and [2]), rejects the exogeneity of instruments when the industry dummies are treated as instruments. However, the test of overidentifying restrictions does not reject exogeneity of errors when the industry dummies are included in the structural form as well. That is, the four main instruments HERF, HERFSQ, PRODWORKER, and PRODWORKERSQ pass the instrument exogeneity test. However, the industry dummies are by definition endogenous, and we are unable to make decide on the basis of the Sargan tests. We choose to follow the theoretical GH specification that has no variables in the structural form other than $\left(I_i \times \frac{X_i/M_i}{e_i}\right)$. As mentioned, both specifications produce similar results (see the previous footnote).

2003, Kleibergen 2002, Hahn and Hausman 2002a, Staiger and Stock 1997, and Nelson and Startz 1990).

We borrow from Stock, Wright, and Yogo (2002) to describe the problem of WIs and robust estimation in their presence. Consider again the structural model of Section 1

$$y = Y\beta + \in \tag{6}$$

and the reduced form for Y given by

$$Y = Z\Pi + \upsilon. \tag{7}$$

In the structural equation, the error term ϵ has variance $\sigma^2_{\in}$. The covariance of the errors $(\in, \upsilon)$ is given by the 2×2 matrix Σ with diagonal elements $\sigma^2_{\in}$ and σ^2_{υ} and off-diagonal element $\sigma^2_{e\upsilon}$. The validity of 2SLS or other two-stage methods requires that instruments not be weakly related to y (given the exogenous variables in the structural equation if any). More formally, the focus of attention in the WI setting is the concentration parameter μ^2,

$$\mu^2 = \Pi' Z' Z \Pi \big/ \sigma^2_{\upsilon}. \tag{8}$$

μ^2 is closely related to the first-stage F-statistic. Specifically, when μ^2/K is large, it is approximately distributed as $F - 1$. Only if μ^2 is fairly large is the distribution of the 2SLS estimator well approximated by the normal distribution (with mean β and variance $\sigma^2_{\upsilon}/\sigma^2_{\upsilon}$). Simulations by Nelson and Startz (1990) show that for small values of μ^2/K, the distribution of the 2SLS estimator is very nonnormal and possibly even bimodal.

Intuitively, if the instruments are weakly related to the endogenous regressor (conditional on any exogenous variable in the structural equation), then t-tests based on conventional asymptotic approximations may be quite inaccurate in a small-sample sense. That is, asymptotic approximations are problematic with small samples, and small samples are the rule in political science and economics. Although large-sample approximations make things simpler, they brush under the carpet small-sample problems (e.g., bias) associated with making inferences using these approximations. On the other hand, constructing the exact small-sample distribution of estimators is an imposing, and sometimes impossible, task.[16]

We investigate weakness in the instruments on the basis of two diagnostics.

- The bias of the 2SLS estimator *relative* to the large-sample bias of the OLS estimator: $E\left(\hat{\beta}^{2SLS} - \beta\right) \Big/ \text{plim}\left(\hat{\beta}^{OLS} - \beta\right)$. If $\mu^2 > 0$, then the instruments are relevant but may be weak.[17] If μ^2 is small, the 2SLS bias is not very different from the large-sample bias of the OLS estimator, and this relative bias is close to one. As μ^2 becomes large, the relative bias is approximately inversely proportional to $\mu^2/(K - 2)$. Whether the

[16] See, for example, Phillips (1983). Bayesians (and frequentists) have accomplished the art of making exact inferences using computational techniques such as Monte Carlo simulations to solve the problem of integration in high dimensions. Even so, the endogeneity problem has received little attention in the Bayesian literature since Rothenberg (1975) and the WI problem even less. We note that while strict Bayesians do not see identification as a problem, the school of practical Bayesians, following in the tradition of Rothenberg, might find solutions to the WI problem useful.

[17] If $\mu^2 = 0$, then the instruments are not only weak but also irrelevant.

instruments are too weak for reliable inference from 2SLS estimation depends on the maximum relative bias that one is willing to tolerate.

- The size of the LIML estimator. The size of a test or the significance level α[18] is the maximal probability of rejecting a true null hypothesis (Type I error). The usual asymptotic approximations are inaccurate with WI—a test at, say, the 5% level of significance based on the Wald asymptotic approximation may have small sample (i.e., *actual*) size that is much larger than 5%. Whether the instruments are too weak for reliable inference from LIML estimation may be based on this "size distortion." It may be stated as the maximum actual size that one is willing to tolerate when testing at, say, a 5% level of significance using an asymptotic approximation.

Stock and Yogo (2004) provide tables of critical values with which to compare first-stage F-statistics, making these diagnostics easy to apply. Because some methods of estimation are more robust to WI than other methods, the detection of WI can indicate which is the more robust method. In our case, the diagnostics indicate that the LIML estimator is more robust to WI than the 2SLS estimator.

5 Results

Table 1 presents descriptive statistics for the dependent variable, the regressor $I \times (X/M)/e$ defined at three cutoffs for I, and the set of instruments used to identify the structural model. As described above, the Herfindahl index of industry concentration (HERF), its squared term (HERFSQ), the percentage of employees that are production workers (PRODWORKER), its squared term (PRODWORKERSQ), and 12 three-digit NAICS industry dummies are used as instruments. The 2SLS estimates of β using the Smith-Blundell-Kelejian method are presented in Table 2. The three cutoffs over which the political organization indicator I are defined in the three models should cover the spectrum of possible patterns of political organization in the trade arena. Thus, the statistically significant positive estimates for β across the three models indicate a robust qualitative affirmation of the GH model. The estimates of β range from 0.008 to 0.013.[19]

Reported below the structural coefficient estimates are the implied values of the parameter a. They indicate that the government weights a dollar of welfare anywhere between 77 and 125 times as much as a dollar of campaign contributions.[20] They are quantitatively similar to what we have come to expect about the size of this coefficient from the Goldberg-Maggi and Gawande-Bandopadhyay studies using data from the early 1980s, McCalman's Australia study, and the Mitra-Thomakos-Ulubasoglu Turkey study. The GH model appears to hold up over time.

[18]see, for example, Greene (2000, chap. 4).

[19]Smith and Blundell (1986) show that the t-test of statistical significance of the first-stage residual from the regression of $I \times (X/M)/e$ on the instruments (the residual enters as an explanatory variable in the Tobit model) is equivalent to a Lagrange multiplier test of the weak exogeneity of $I \times (X/M)/e$. The estimates indicate that the hypothesis of weak exogeneity is rejected. The inclusion of the residual term is, thus, necessary to purge the endogeneity in the regressor and ensure consistent estimates of β.

[20]The direct (uninstrumented) Tobit estimates range between 0.0026 (a = 379.1) and 0.0038 (a = 261.5). Since the a's implied by the Tobit estimates are larger than our two-stage Tobit estimates, and since the bias of the latter is probably smaller than the bias of the former (this is true of OLS versus 2SLS—see, e.g., Hahn and Hausman 2002b), the true a should be even smaller than those implied by the two-stage Tobit (i.e., the Smith-Blundell-Kelejian) estimates.

Table 2 Smith-Blundell-Kelejian two-stage least squares estimates of Tobit structural equation (3)

	I defined at					
	10% cutoff (I = I_{10})		*25% cutoff* ($I = I_{25}$)		*50% cutoff* ($I = I_{50}$)	
Structural equation for NTBs						
	est	t	est	t	est	t
$I \times (X/M)/e$	0.008	4.940***	0.009	5.290***	0.013	4.470***
N	256		256		256	
–2Hln(L)						
	205.48		202.02		208.22	
Implied values of a						
a	125		111.1		76.9	
Smith–Blundell test of weak exogeneity						
Coefficient on first-stage residual	−0.009	−4.540***	−0.011	−4.710***	−0.013	−3.850***
First-stage statistic						
F	3.88		4.82		4.29	

Note. In the Smith–Blundell 2SLS procedure, the first-stage residuals are statistically significant, rejecting the hypothesis of exogeneity of $I \times (X/M)/e$. The presence of the residual corrects for endogeneity. First-stage estimates are available from authors. ** and *** indicate, respectively, statistical significance at 5% and 1%.

These inferences presume that the instruments are up to the task. To investigate the quality of instruments, the last row of Table 2 reports the F-statistic from the first-stage regression of $I \times (X/M)/e$ on the IVs described in Table 1. The value of F between 3.88 and 4.82 indicates a potential WI problem. In order to formally assess whether this is so, we use Table 3. It contains extracts from tables provided in Stock and Yogo (2004) expressly for the purpose of using the F-statistic to detect WI. The table indicates three diagnostics—two for the 2SLS estimator and one for the LIML estimator. The first diagnostic is used to test the hypothesis that the bias of the 2SLS estimator relative to the (large sample) OLS bias is greater than a specific acceptable level. The table indicates that in order to ensure that the 2SLS relative bias is no greater than 30%, the F-statistic must be greater than 4.59 (we use the $K = 16$ row since that is the closest K reported in the Stock-Yogo tables to the number of instruments we use). Thus, inferences from our Smith-Blundell-Kelejian estimator in Table 2 require us to be willing to tolerate a bias in those estimates of at least 30%, and probably closer to 40%, relative to the OLS bias.

The second diagnostic is used to test the hypothesis that the actual size of the 2SLS t-test at 5% can exceed a specific acceptable tolerance level. The critical values with $k = 16$ reject the hypothesis that the size of the 5% test is actually less than 25% (for which the critical value of F with 16 instruments is 15.19), an unacceptably high level of Type I error in most applications.[21] The third diagnostic is used to test the same size hypothesis but for the LIML estimator. The much smaller critical values in Table 3 compared with the

[21]Our F-statistics are far lower than the critical F value of 15.19, so the actual size of our test is much worse than 25%. The Stock-Yogo tables do not provide the critical F values beyond the 25% size, which stretches the norms of acceptable Type I error anyway.

Table 3 stock and Yogo (2004) critical values for tests of WIs based on two-stage least squares bias, two-stage least squares size, and LIML size; significance level is 5%

TSLS relative bias				
K	0.05	0.10	0.20	0.30
3	13.91	9.08	6.46	5.39
16	21.28	11.50	6.39	**4.59**
TSLS size				
K	0.10	0.15	0.20	0.25
1	16.38	8.96	6.66	**5.53**
2	19.93	11.59	8.75	7.25
3	22.30	12.83	9.54	7.80
16	52.77	27.99	19.51	15.19
LIML size				
K	0.10	0.15	0.20	0.25
1	16.38	8.96	6.66	5.53
2	8.68	5.33	4.42	3.92
3	6.46	4.36	3.69	3.32
16	**3.27**	2.48	2.18	2.00

corresponding values for the 2SLS size test show that LIML is more robust to size distortions with the same instruments. The first-stage F-statistics are all greater than critical value of 3.27, thus rejecting the hypothesis that the actual size exceeds 10%. We, thus, undertake LIML estimation of the structural model.[22] We have derived the likelihood function for LIML estimation of the Tobit model in Appendix B.

The LIML estimates of β are presented in Table 4. They are larger than their 2SLS counterparts, so that the implied estimates of a are smaller. The estimates of β are all statistically significant at the conventional 5% level of significance. According to the third Stock-Yogo size test, if we are willing to accept an actual size of 0.10, inferences about β at the usual 5% level are robust to the WI problem. The implied values of a range between 37.8 and 70.8.

5.1 *Interpreting the Estimates*

We have questioned neither the assumptions of the GH model nor the lobbying construct that is central to it. Taking the predictions at face value, the estimates imply that the U.S. government is close to a welfare maximizer.[23] Although our estimates of a are lower than those in previous tests of the model, they are still quite high. The LIML results indicate that as recently as the mid-1990s, the U.S. government put a weight of between 38 and 71 on

[22]LIML may additionally be preferred due to its smaller bias relative to 2SLS. In their seminal numerical examinations of the exact distributions of the 2SLS and LIML estimators, Anderson and Sawa (1982) and Anderson, Kunimoto, and Sawa (1982) found the LIML estimator to be median unbiased (the moments may not exist for the original LIML estimator—the Fuller-corrected LIML estimator is mean unbiased). Others have found the bias of the Fuller-corrected LIML estimator to be smaller than that of the 2SLS estimator. Monte Carlo experiments of small-sample properties have also confirmed the superiority of LIML over 2SLS under normal and certain nonnormal distributions (see, e.g., Gao and Lahiri 2000).

[23]Note that the model presumes equal marginal utility of income for the public, so "fairness" considerations do not apply.

Table 4 LIML estimates of Tobit structural equation (3)

	I defined at					
	10% cutoff ($I = I_{10}$)		*25% cutoff* ($I = I_{25}$)		*50% cutoff* ($I = I_{50}$)	
Structural equation for NTBs						
	est	*t*	est	*t*	est	*t*
$I \times (X/M)/e$	0.014	5.80***	0.016	5.93***	0.026	5.21***
N	256		256		256	
−2Hln(L)	2252.2		2194.6		2089.8	
Implied values of *a*						
a	70.8		64.3		37.8	
Reduced form for ($I \times z/e$)						
	est	*t*	est	*t*	est	*t*
HERF	29.735	0.9	9.875	0.33	26.921	1.24
HERFSQ	−225.831	−1.73*	−166.020	−1.42	−160.071	−1.87*
PRODWORKER	−2.084	−2.68***	−2.017	−2.89***	−2.126	−3.63***
PRODWORKERSQ	0.022	2.95***	0.022	3.21***	0.021	3.88***
FOOD	16.123	2.98***	15.895	3.26***	12.981	3.17***
TEXTILES	0.938	0.13	−0.803	−0.13	6.128	1.25
APPAREL	−6.996	−0.9	−8.175	−1.18	0.397	0.08
WOOD	4.660	0.76	0.160	0.03	5.132	1.23
NONMETALS	−2.781	−0.52	−2.620	−0.55	1.902	0.52
METALS	4.128	0.63	−0.009	0	4.631	1.05
MACH	0.626	0.11	0.266	0.05	5.110	1.28
ELECTRONIC	−3.621	−1.06	−3.453	−1.13	−0.061	−0.03
ELECTRICAL	2.638	0.39	1.915	0.32	7.825	1.71*
AUTOS	10.393	1.71*	9.838	1.81*	11.274	2.7***
MISC	−2.807	−0.43	−3.969	−0.69	2.868	0.64
Intercept	53.846	3.24***	52.715	3.52***	50.067	3.86***
First-stage *F* (from Table 2)	3.88		4.82		4.29	

Note. LIML likelihood function derived in Appendix B. PETR&CHEM dummy dropped since intercept term is present. *, **, and *** indicate, respectively, statistical significance at 10%, 5%, and 1%.

a dollar of public welfare relative to the weight it put on a dollar of campaign contributions. Thus, net welfare loss of a dollar from protection is valued far more by the U.S. government than is $1 of campaign contribution. These estimates of the parameter *a* are at odds with the fact that industrial protection imposed tens of billions of dollars in welfare losses per annum during that period, whereas political contributions were manyfold lower.[24] The implication is that *a* is greatly overestimated by us and previous researchers. Solving this problem requires taking a harder look at the model and what it may be missing.

Here we suggest extensions of the GH model that may solve the problem of overestimating *a*. First, the GH model presumes government to be a unitary actor. It does not

[24]See, for example, Hufbauer et al. (1986) and de Melo and Tarr (1990) for estimates of (net) welfare losses from computational general equilibrium models.

model trade policy as the outcome of interactions among several political actors as is the case in the real world. In the U.S. Congress, the decision to initiate new legislation or make changes to existing statutes or continue with the status quo usually starts in a committee. Thus, committees set agendas. The role of the House Ways and Means committee looms large in trade issues. Trade policy in agriculture is the purview of the House Agriculture committee. Several politicians become influential on the margin at various stages of the process. Thus, it is not possible in the real world for a singular "government" to guarantee a specific policy with certainty *ex ante*, as assumed in the GH model.

One implication is that protection is for sale but uncertainly. The greater this *ex ante* uncertainty, the lower the money contributions (since they are made *ex ante* as well). If uncertainty were built into the model then it can be theoretically shown that the estimate of a should be scaled down by the magnitude of the uncertainty. For example, if lobbies are risk neutral and assess their chances of obtaining protection to be 10%, then the estimates of a are 1/10 of those reported or between 4.77 and 7.14. These are probably much closer to the true a's.[25]

A promising direction of research toward scaling down the a estimates is to build more institutional detail into the GH model. This is beginning to be done, but we are as yet unaware of papers that have presented a modified version of equation (4) in the presence of institutional detail. A good example is the model of two-party *electoral competition* by Grossman and Helpman (1995) with similar objective functions as the one used here. This model introduces a number of new parameters that may bring the a estimates more in line with reality. An attractive feature of models in this genre is that they introduce the lobbying of legislatures, as opposed to unitary entities. The next step would be to introduce legislative processes such as legislative bargaining into the GH model.

5.2 *Sensitivity Analysis*

Although Table 2 has remained strictly true to the theory, the prolific literature on the political economy of trade protection has experimented with a range of *ad hoc* variables (see, e.g., Baldwin 1985). In Table 5, we admit three such out-of-model explanatory variables that we consider to be weakly exogenous. They are as follows: WAGE is the average annual wage in manufacturing in 1995 (in million dollars). Its coefficient is expected to be negative if, as the first-generation literature suggests, trade protection is used to also protect unskilled worker wages. K/L measures the capital-labor ratio (in million dollars per manufacturing worker). Its coefficient is expected to be negative since capital intensive industries enjoy a comparative cost advantage and should not need protection. Table 5 indicates that in the extended model the coefficients on WAGE and K/L have the expected signs, but the coefficient on WAGE is not estimated with statistical significance. The main result in Table 5 is the robustness of the estimates on the issue variable $I \times z/e$. Conditioning out these out-of-model variables raises the coefficient values above the corresponding coefficients in Table 4. However, even though they result in smaller estimates of a (between 16 and 30), they still portray the U.S. government as welfare oriented. The theoretical extensions suggested above appear to be the best way to solve the puzzle of the high a.

[25]The use of PAC data is in line with this extended theory. When PACs contribute to House and Senate races, they must assess the certainty with which these contributions may translate into favorable policy. They are not contributions made to a unitary policymaker who can deliver a promised policy with certainty.

Table 5 LIML estimates of structural equation with out-of-model explanatory variables

	I defined at:					
	10% cutoff ($I = I_{10}$)		*25% cutoff* ($I = I_{25}$)		*50% cutoff* ($I = I_{50}$)	
Structural equation for NTBs						
	est	*t*	est	*t*	est	*t*
$I \times (z/e)$	0.034	2.85***	0.027	3.47***	0.062	2.23**
WAGE	−3.764	−1.31	−1.277	−0.73	−1.578	−0.56
K/L	−0.704	−1.63*	−0.699	−2.02**	−1.507	−2.14**
N	256		256		256	
−2Hln(L)	2234.8		2182.2		2073.1	
Implied values of						
a	29.16		36.76		16.19	
First-stage *F*	3.57		4.43		3.66	

Note. See Notes to Table 4.

Finally, we estimated the baseline and extended models using LIML but as linear, not Tobit, models. The results reported are qualitatively similar to those from the linear models.

6 Conclusion

Using social science data to find causal relationships must contend with the problem of endogeneity of the ''independent'' variables or regressors. Technically, since the regressors are correlated with the error term, using standard estimators like OLS yield biased estimates. The bias may be unacceptably large when precision in the estimates are desired, for example, when estimating structural models to inform or perform policy. The most popular solution to the problem is to ''instrument'' the endogenous regressors using IVs. In theory, adequate IVs are those that are strongly correlated with the endogenous regressors but not with the error term. In practice, finding adequate instruments is difficult. Highly correlated instruments are more than likely to be themselves endogenous. Variables that are found to be acceptable instruments on theoretical grounds and empirical tests are more than likely to be *weakly* correlated with regressors. This problem of WIs that is so relevant for data generated by political quasi-experiments is the subject of this paper.

Specifically, with WIs, using the 2SLS estimator may not resolve the original problem of bias, which motivated the use of 2SLS over OLS in the first place. Another small-sample problem with the 2SLS estimator is that a hypothesis test at, say, the 5% level may have an exact (small sample) size that is greater than 5%. An emerging consensus from the recent literature on the WI problem is that the LIML estimator may have better small-sample properties than 2SLS. Although the LIML method is the staple of most econometrics texts, it is underused in empirical work.

We suspect that many, if not most, studies in political science and economics suffer from the WI problem. We demonstrate the diagnosis and a solution to the WI problem in the context of testing the Grossman and Helpman (1994) protection-for-sale political economy model. The model has been tested using an older benchmark data set from around 1980. We update those estimates using data from the late 1990s, which better represent the current structure of protection in the United States. Since protection data are censored, we estimate a Tobit model using LIML estimator. We expect this Tobit LIML estimator to be useful in other political science applications. We hope this study also encourages the use of LIML estimators for other popular nonlinear models with qualitative and discrete dependent variables.

A main message for empirical researchers is that it is important to diagnose the extent of the WI problem. That is easily accomplished using simple diagnostics like the first-stage F-statistic. The first-stage F-statistic may be used to indicate the maximum bias in the 2SLS estimator relative to the OLS estimator. If that bias is tolerable, then simply stating the relative bias effectively conveys the limitations of the estimator. Another use of the first-stage F-statistic is to indicate the exact (small sample) Type I error when testing a hypothesis at, say, the 5% level. If the amount of Type I error in the 2SLS estimator is unacceptably large, then a better alternative may be to use the LIML estimator. Thus, inferences are possible in the presence of WIs. What is required is to explicitly communicate the limitations of the inference (in terms of bias and/or exact size) and to use estimators like LIML that have good small-sample properties.

Appendix

Appendix A: Data

The NTB coverage ratios are constructed from the UNCTAD database TRAINS. TRAINS has nontariff measures (NTMs) and imports at the 10-digit HS level. Examples of the types of NTB in the data are tariff quota measures antidumping measures, countervailing measures, prohibition, licensing, authorization, product characteristic requirement, product marking requirement, product labeling requirement, and product inspection requirement. In order to construct the coverage ratio, a binary indicator is assigned to each 10-digit commodity depending on whether any of these NTMs exist. This is done bilaterally for each trading partner, and the coverage ratio at the 10-digit level is constructed. The HS10-to-NAICS (1997 basis, six-digit level) concordance file (downloaded from Jon Haveman's Web site at http://www.macalester.edu/research/economics/PAGE/HAVEMAN/Trade.Resources/TradeConcordances.html) is used to aggregate up to the NAICS level of this study.

Manufacturing data from the Annual Survey of Manufactures (ASM) for 1994–96 are taken from the report "Statistics for Industry Groups and Industries" by the Census. These data are at the four-digit 1987 version of US SIC level (the directly available data from the ASM are incomplete at the NAICS level). The mapping from SIC to NAICS requires matching (1) new industries in NAICS that did not exist in SIC, (2) existing SIC industries that are separated and reorganized in NAICS, and (3) SIC codes that just do not have matching NAICS codes. Thus, several current concordances for SIC and NAICS are limited. The fractional mapping from SIC to NAICS, thus, uses the fraction of an SIC's shipments that is classified in its associated NAICS codes as the bridge. This bridge is available at the Census Bureau's Web site (http://www.census.gov/epcd/ec97brdg). Using this approach, we were able to map over 95% of the SIC-level data in terms of value added for each of the 1994–96 years.

Elasticity data are taken from Gallaway, McDaniel, and Rivera (2003). Although the long-run elasticity is more appropriate in the study, they are quite incomplete, and so we use short-run elasticity estimates in our study. The elasticities are adjusted using Fuller's (1986) method to correct for errors-in-variables. A similar concordance into NAICS as for the ASM data is used to concord the 297 (positive and thus usable) elasticity estimates at SIC level into 313 six-digit NAICS industries.

The lobbying data at the PAC level of each lobbying firm are collected for every 2-year election cycle and is available from the Federal Election Commission at their Web site http://www.fec.org. The lobbying data at the firm level were concorded to the four-digit SIC level. We are grateful to Chris Magee for supplying the concordance used in Beaulieu and Magee (2004). For our study of 1995 NTBs, we sum the lobbying spending during the 1991–92 and 1993–94 election cycles. The same method for the ASM data is used to concord the lobbying data into the NAICS.

Appendix B: Tobit LIML function

Consider the structural equation for the variable y with one endogenous regressor Y and no exogenous variables given by

$$y_i = Y_i\beta + \in_j,\ i = 1, \ldots, T,$$

where the error term ϵ_i is identically and independently distributed normally, with mean zero and variance $\sigma^2_{\in}$. The coefficient β is identified by a fixed set of K instruments Z. The reduced form for Y_i is given by

$$Y_i = Z_i\Pi + \upsilon_i. \tag{10}$$

Substituting from reduced form into the structural form, the limited information two-equation system may be written as

$$\begin{aligned} y_i^* &= Z_i\Pi\beta + u_i \\ Y_i &= Z_i\Pi + \upsilon_i \end{aligned}, \tag{11}$$

where $u_i = \beta\upsilon_\iota + \epsilon_i$. Thus, the errors terms in the limited information system are correlated. The structural equation in (3) is written in terms of the latent variable y_i^* which is censored below zero. The observed data y_i equals zero whenever $y_i^* < 0$, but equals y_i^* whenever $y_i^* > 0$. The likelihood function described below is used to account for the truncation within the LIML context.

We assume the error vector $(u_i, \upsilon_i)'$ in equation (11) is iid bivariate normal with variance-covariance matrix Ω,

$$\Omega = \begin{pmatrix} \omega_u & \omega_{u\upsilon} \\ \omega_{\upsilon u} & \omega_\upsilon \end{pmatrix}. \tag{12}$$

We write the likelihood function in terms of the elements of Ω^{-1} denoted

$$\Omega^{-1} = \begin{pmatrix} \omega^u & \omega^{u\upsilon} \\ \omega^{\upsilon u} & \omega^\upsilon \end{pmatrix}. \tag{13}$$

The joint probability density function (pdf) of the data, without censoring is given by

$$f(y_i^*, Y_i) = \frac{\sqrt{|\Omega^{-1}|}}{2\pi}\exp\left[-\frac{\omega^u(y_i^* - Z_i\Pi\beta)^2 + \omega^\upsilon(Y_i - Z_i\Pi)^2 + 2\omega^{u\upsilon}(y_i^* - Z_i\Pi\beta)(Y_i - Z_i\Pi)}{2}\right]. \tag{14}$$

Since $y_i = y_i{*}$ if $y_i{*} > 0$, the conditional pdf $f(y_i, Y_i|y_i > 0)$ is

$$f(y_i, Y_i|y_i{>}0) = \frac{1}{P(y_i{>}0, Y_i)} \times f(y_i, Y_i). \tag{15}$$

For the censored values of y_i all we know is that $y_i{*} < 0$, and hence the conditional probability when we have censoring in the structural equation is

$$P(y_i{=}0, Y_i) = P(y_i^*{<}0, Y_i) = \int_{-\infty}^{0} f(y_i^*,\ Y_i)dy_i^*. \tag{16}$$

Thus,

$$P(y_i{=}0, Y_i) = \sqrt{\frac{|\Omega^{-1}|}{2\pi\omega^u}}\exp\left[-\frac{|\Omega^{-1}|(Y_i - Z_i\Pi)^2}{2\omega^u}\right]F\left(-\sqrt{\omega^u}Z_i\Pi\beta + \frac{\omega^{u\upsilon}}{\sqrt{\omega^u}}(Y_i - Z_i\Pi)\right), \tag{17}$$

where $F(.)$ is the standard normal cumulative density function.

Define the dummy variable D_i that takes the value 0 if observation i has censoring in y_i and 1 if there is no censoring. The log-likelihood function can then be written as

$$L_i = (1 - D_i)\ln\{P(y_i{=}0, Y_i)\} + D_i\ln\{P(y_i{>}0, Y_i) \times f(y_i, Y_i|y_i{>}0)\} \tag{18}$$

Maximizing the log-likelihood simultaneously yields the LIML structural and reduced-form parameters (β, Π, Ω) for the Tobit model.

References

Acemoglu, Daron, Simon Johnson, James A. Robinson, and Pierre Yared. 2008. Income and democracy. *American Economic Review* 98:808–42.

Anderson, T. W., N. Kunitomo, and T. Sawa. 1982. Evaluation of the distribution function of the limited information maximum likelihood estimator. *Econometrica* 50:1009–27.

Anderson, T. W., and H. Rubin. 1949. Estimators of the parameters of a single equation in a complete set of stochastic equations. *The Annals of Mathematical Statistics* 21:570–82.

Anderson, T. W., and T. Sawa. 1979. Evaluation of the distribution function of the two-stage least squares estimate. *Econometrica* 47:163–82.

Angrist, Joshua D., and Alan B. Krueger. 1991. Does compulsory school attendance affect schooling and earnings. *Quarterly Journal of Economics* 106:979–1014.

Baldwin, Robert. 1985. *The political economy of US import policy*. Cambridge, MA: MIT Press.

Bartels, Larry M. 1991. Instrumental and 'quasi-instrumental' variables. *American Journal of Political Science* 35:777–800.

Bawn, Kathleen, and Frances Rosenbluth. 2006. Short versus long coalitions: Electoral accountability and the size of the public sector. *American Journal of Political Science* 50:251–65.

Beaulieu, Eugene, and Chris Magee. 2004. Campaign contributions and trade policy: New tests of Stolper-Samuelson. *Economics and Politics* 16:163–87.

Bombardini, Matilde. 2008. Firm heterogeneity and lobby participation. *Journal of International Economics* 75:329–48.

Bound, J., D. A. Jaeger, and R. Baker. 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak. *Journal of the American Statistical Association* 90:443–50.

Calvo, Ernesto, and Maria V. Murillo. 2004. Who delivers? Partisan clients in the Argentine electoral market. *American Journal of Political Science* 48:742–57.

de Melo, Jaime, and David Tarr. 1990. Welfare costs of U.S. quotas in textiles, steel and autos. *Review of Economics and Statistics* 489–97.

Eicher, Theo, and Thomas Osang. 2002. Protection for sale: An empirical investigation: Comment. *American Economic Review* 92:1702–10.

Facchini, Giovanni, Johannes Van Biesebroeck, and Gerald Willmann. 2003. *Protection for sale with imperfect rent capturing*. University of Kiel Economics Working Paper.

Feenstra, Robert, and Lee Branstetter. 2002. Trade and foreign direct investment in China: A political economy approach. *Journal of International Economics* 58:335–58.

Fuller, Wayne A. 1986. *Measurement error models*. New York: John Wiley.

Gabel, Matthew, and Kenneth Scheve. 2007. Estimating the effect of elite communications on public opinion using instrumental variables. *American Journal of Political Science* 51:1013–28.

Gallaway, Michael P., Christine A. McDaniel, and Sandra A. Rivera. 2003. Short-run and long-run industry-level estimates of U.S. Armington elasticities. *North American Journal of Economics and Finance* 14: 49–68.

Gao, Chuanming, and Kajal Lahiri. 2000. Further consequences of viewing LIML as an iterated Aitken estimator. *Journal of Econometrics* 98:187–202.

Gawandem, Kishore, and Bernard Hoekman. 2006. Lobbying and agricultural policy in the United States. *International Organization* 2006:527–61.

Gawande, Kishore, and Pravin Krishna. 2003. The political economy of trade policy: Empirical approaches. In *Handbook of international trade*, eds. J. Harrigan and E. Kwan Choi. Oxford: Blackwell Publishers.

Gawande, Kishore, and Usree Bandyopadhyay. 2000. Is protection for sale? A test of the Grossman-Helpman theory of endogenous protection. *Review of Economics and Statistics* 89:139–52.

Greene, William H. 2000. *Econometric analysis*. 4th ed. New York: Prentice Hall.

Goldberg, Pinelopi, and Giovanni Maggi. 1999. Protection for sale: An empirical investigation. *American Economic Review* 89:1135–55.

Grossman, Gene M., and Elhanan Helpman. 1995. Protection for sale. *American Economic Review* 4: 833–50.

———. 1995. Electoral competition and special interest politics. *Review of Economic Studies* 63:265–86.

Hahn, Jinyong, and Jerry Hausman. 2002a. A new specification test for the validity of instrumental variables. *Econometrica* 70:163–89.

———. 2002b. Notes on bias in estimators for simultaneous equation models. *Economics Letters* 75:237–41.

Hiscox, Michael J. 2002. Commerce, coalitions, and factor mobility: Evidence from congressional votes on trade legislation. *American Political Science Review* 96:593–608.

Hufbauer, Gary C., Diane T. Berliner, and Kimberly A. Elliott. 1986. Trade protection in the United States: 31 case studies. Washington, DC: Institute of International Economics.

Kelejian, Harry H. 1971. Two stage least squares and econometric systems linear in parameters but non-linear in the endogenous variables. *Journal of the American Statistical Association* 66:373–4.

Kleibergen, Frank. 2002. Pivotal statistics for testing structural parameters in instrumental variables regressions. *Econometrica* 70:1781–804.

Konisky, David M. 2007. Regulatory competition and environmental enforcement: Is there a race to the bottom? *American Journal of Political Science* 51:853–72.

Lassen, David D. 2005. The effect of information on voter turnout: Evidence from a natural experiment. *American Journal of Political Science* 49:103–18.

Lebo, Matthew J., Adam J. McGlynn, and Gregory Koger. 2007. Strategic party government: Party influence in congress, 1789–2000. *American Journal of Political Science* 51:464–81.

Lewbel, Arthur. 1997. Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D. *Econometrica* 65:1201–14.

Lewis-Beck, Michael S., Richard Nadeau, and Angelo Elias. 2008. Economics, party, and the vote: Causality issues and panel data. *American Journal of Political Science* 52:84–95.

McCalman, Phillip. 2004. Protection for sale and trade liberalization: An empirical investigation. *Review of International Economics* 12:81–94.

McGillivray, Fiona. 1997. Party discipline as a determinant of the endogenous formation of tariffs. *American Journal of Political Science* 41:584–607.

Mitra, Devashish. 1999. Endogenous lobby formation and endogenous protection: A long-run model of trade policy determination. *American Economic Review* 89:1116–34.

Mitra, Devashish, Dimitrios D. Thomakos, and Mehmet A. Ulubasoglu. 2002. Protection for sale' in a developing country: Democracy versus dictatorship. *Review of Economics and Statistics* 84:497–508.

Moreira, Marcelo J. 2003. A conditional likelihood ratio test for structural models. *Econometrica* 71:1027–48.

Nelson, Charles R., and Richard Startz. 1990. Some further results on the exact small sample properties of the instrumental variables estimator. *Econometrica* 58:967–76.

Olson, Mancur. 1965. *The logic of collective action: Public goods and the theory of groups*. Cambridge: Harvard University Press.

Phillips, Peter C. B. 1983. Exact small sample theory in the simultaneous equations model. In *Handbook of econometrics*. Vol. 1, eds. Z. Griliches and M. D. Intrilligator, 449–516. Amsterdam, The Netherlands: North-Holland.

Prakash, Aseem, and Matthew Potoski. 2006. Racing to the bottom? Trade, environmental governance, and ISO 14001. *American Journal of Political Science* 50:350–64.

Reinhardt, Eric, and Marc Busch. 1999. Industrial location and protection: The political and economic geography of U.S. nontariff barriers. *American Journal of Political Science* 43:1028–50.

Rothenberg, T. 1975. Bayesian analysis of simultaneous equations models. In *Studies in Bayesian econometrics and statistics*, eds. S. E. Fienberg and A. Zellner. Amsterdam, The Netherlands: North-Holland.

Rudra, Nita. 2005. Globalization and the strengthening of democracy in the developing world. *American Journal of Political Science* 49:704–30.

Smith, R. J., and R. W. Blundell. 1986. An exogeneity test for a simultaneous equation Tobit model with an application to labor supply. *Econometrica* 54:679–86.

Staiger, Douglas, and James H. Stock. 1997. Instrumental variables regression with weak instruments. *Econometrica* 65:557–86.

Stock, James H., J. H. Wright, and Motohiro Yogo. 2002. A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics* 20:518–29.

Stock, James H., and Motohiro Yogo. 2004. Testing for weak instruments in linear IV regression. In *Identification and inference in econometric models: Essays in honor of Thomas J. Rothenberg*, eds. D. W. K. Andrews and J. H. Stock. Cambridge: Cambridge University Press.