THE ENVIRONMENTAL KUZNETS CURVE: REAL PROGRESS OR MISSPECIFIED MODELS?

Daniel L. Millimet, John A. List, and Thanasis Stengos*

Abstract—We explore the importance of modeling strategies when estimating the emissions-income relationship. Using U.S. state-level panel data on nitrogen oxide and sulfur dioxide emissions, we estimate several environmental Kuznets curves using the standard parametric framework as well as a more flexible semiparametric alternative. Formal statistical comparisons of the results overwhelmingly reject the parametric approach. Moreover, the differences, particularly for sulfur dioxide, are economically significant.

I. Introduction

RECENTLY, considerable attention has been given to empirically modeling the relationship between economic growth and the environment (for example, Grossman and Krueger, 1995; Schmalensee et al., 1998). Although there have been some notable disparities across the empirical estimates from different studies, the typical shape of the income-pollution relationship has been an inverted U. In light of this evidence, some commentators have argued that the "way to attain a decent environment in most countries is to become rich" (Beckerman, 1992). Although this premise is appealing, within this literature little attention has been given to the econometric properties of the various reduced-form models employed.

Our goal in this study is to formally test the appropriateness of the traditional parametric regression specifications commonly used in the Environmental Kuznets curve (EKC) literature against the alternative of the semiparametric partially linear regression (PLR) model. To carry out our objective we employ a panel data set on U.S. state-level sulfur dioxide (SO_2) and nitrogen oxide (NO_x) emissions from 1929 to 1994. Our empirical results provide several important insights. First, regardless of empirical technique, pollution projections across U.S. states appear to follow the inverted-U shape commonly found (at the country level) between emissions or ambient pollution levels and a measure of income. Second, concerning the exact nature of the emissions-income relationship, modeling assumptions matter. Utilizing a recent specification test developed by Zheng (1996) and Li and Wang (1998), we find that parametric and semiparametric models provide statistically different estimates of the EKC, especially in the SO_2 models. In particular, we find that the parametric approach typically produces estimates of the EKC turning points that are too

pessimistic. Third, we specifically test—and fail to reject—the assumption of no serial correlation in the PLR model.[1]

On the one hand, these results represent good news in that the received estimates in the literature may be considered conservative. Alternatively, since optimal regulatory policies critically depend on an accurate estimate of the peak of the Kuznets curve, these findings should send a warning that in some cases more flexible empirical techniques provide much different policy conclusions. The remainder of the study proceeds as follows. Section II describes the data and empirical methods. Section III presents the empirical results. Section IV lays out a specification test, allowing us to discern between the parametric and semiparametric models, as well as a test for serial correlation. Section V concludes.

II. Data and Empirical Methods

A. Data

Data on the emissions of the two criterion air pollutants, SO_2 and NO_x, were originally published in the U.S. Environmental Protection Agency's (EPA) *National Air Pollutant Emission Trends, 1900–1994,* and cover fiscal years 1929–1994. The emission estimation methodologies fall into two major regimes: 1929–1984 and 1985–1994. Emission estimates from 1929 to 1984 are derived using an approach whereby information at the national level (such as activity indicators, material flows, control efficiencies, and fuel property values) is used to create nationwide emissions estimates. Each state is then allotted a certain portion based on its production activities. Emissions for the years 1985–1994 are estimated using a methodology whereby emissions are derived at the plant or county level and aggregated to the state level.

As List and Gallet (1999) point out, these data are particularly useful for modeling the EKC in that they offer important advantages absent in most other data sets. One major advantage of using U.S. data is that they probably yield more reliable estimates than the Global Environmental Monitoring System (GEMS) data used in many cross-country studies (for example, Grossman, 1995; Stern et al.,

---

[1] Li and Hsiao (1998) recently developed tests for serial correlation in the context of a panel data semiparametric partially linear model. The statistics are developed assuming large numbers of individuals and a small time horizon. In that case the asymptotic analysis is based on $N \rightarrow \infty$ for a fixed value of T. Li and Stengos (2001) extend their results to the pure time series case under certain mixing conditions that characterize the data dependence, and perform the asymptotic analysis allowing $T \rightarrow \infty$. The tests they derive generalize the Durbin h-statistic and the usual LM statistics to test the null of no serial correlation in a semiparametric partially linear time series model.

TABLE 1.—SUMMARY STATISTICS, 1929–1994[a]

Variable	Mean	Std. Dev.	Minimum	Maximum
Per capita sulfur dioxide emissions	0.16	0.21	0.00	1.62
Per capita nitrogen oxide emissions	0.09	0.07	0.02	1.14
Per capita state income	9.09	4.24	1.16	22.46
Observations		3168		

[a] Emissions measured in thousands of short tons. Income measured in thousands of 1987 U.S.$.

1996). Another advantage is the sheer length and breadth of the U.S. emissions data set. Since the data set encompasses 48 states and a reasonably long time period (1929–1994), there is a greater chance that it will capture the entire EKC, alleviating some of the concerns about making out-of-sample inferences.

Table 1 presents summary statistics for our variables of interest. It is important to note that, consistent with previous EKC studies, we use variables in per capita terms. A second observation from table 1 is that both per capita SO_2 and NO_x have a good deal of variation. This is intuitively appealing, given that a variety of temporally variant factors account for state emissions activities, such as motor vehicle characteristics (fuel type, vehicle type, technology, and extent of travel), industry emissions, and other off-road activities. Equally important, real per capita income levels over the 1929–1994 period vary significantly, from a low of $1,160 to a high of $22,460.[2]

B. Empirical Methods

The empirical analysis proceeds by examining the emissions data to determine if estimates from simple reduced-form models differ across parametric and semiparametric models and, if so, analyzing whether the differences are significant and robust to changes in sample composition. The methodology employed involves estimating EKCs for subsamples of the data set [namely, state-specific EKCs as in List and Gallet (1999)] as well as for the entire sample (referred to as the *pooled model*). In addition, where all states are pooled together, we estimate both the parametric and semiparametric model twice: once using the full data set, 1929–1994, and once using a partial data set, 1985–1994. Estimates based on the partial data set are obtained for two reasons. First, we want to assess the effect, if any, of the change in measurement scheme employed by the EPA

beginning in 1985 (discussed in the previous section). Second, utilizing the full data set—spanning 66 years—restricts the parameters of any model to be constant over this period. Comparisons based on the restricted sample will help ensure the robustness of our results.

Two parametric approaches are used. First, as is frequently done in the literature, we estimate a two-way fixed-effects panel data model using a cubic specification:

$$P_{it} = \alpha_i + \varphi_t + Y_{it}\delta + \epsilon_{it}, \quad i = 1, 2, \ldots, 48,$$
$$t = 1, 2, \ldots, 66, \tag{1}$$

where P_{it} represents pollutant emissions in state i at time t; Y_{it} is a vector that includes (per capita) income, income squared, and income cubed for state i at time t ($Y_{it} = [y_{it} \ y_{it}^2 \ y_{it}^3]$); δ is a vector of slope coefficients; α_i are state fixed effects, which control for unobserved state factors that affect emission rates; φ_t are time effects, which capture relevant factors that are equivalent across the U.S. states, such as macroeconomic effects and federal environmental legislation; and ϵ_{it} is a contemporaneous error term. Second, following Schmalensee et al. (1998), we replace the cubic specification in (1) with a spline (piecewise linear) function of income based on 10 segments, each containing an equal number of observations.[3] Equation (1), using the cubic and the linear spline specifications, are estimated separately for both pollutant types.

The semiparametric PLR model can be written as

$$P_{it} = \alpha_i + \varphi_t + g(y_{it}) + u_{it}, \quad i = 1, 2, \ldots, 48,$$
$$t = 1, 2, \ldots, 66, \tag{2}$$

where y_{it} is income, $g(\cdot)$ is an unknown function, u_{it} is a mean zero residual assumed to be uncorrelated with $g(\cdot)$, and everything else is as previously defined. Robinson (1988) outlined a kernel-based method for obtaining $\sqrt{n}$-consistent estimates of β (the other parameters that enter the PLR model; in this case, α_i and φ_τ) by conditioning on the variables which enter (2) nonlinearly (in this case, y_{it}). Whereas in the empirical literature the emphasis is on concentrating out the effect of the nonlinear variable(s) by suppressing $g(\cdot)$, in the present paper we provide estimates of $g(\cdot)$ and analyze its behavior.[4] The results presented here are based on standard local constant kernel estimation using the standard Gaussian density as the kernel and applying cross-validation to obtain the smoothing parameter. We also obtained local polynomial regression estimates (Fan, 1993). However, since the results were very similar to the standard kernel estimates, we only report the latter.

---

[2] One potential shortcoming of the data is that the lowest income observed is approximately $1,200 per capita (table 1). This is roughly equivalent to the per capita GDP in Guatemala in 1999. However, because many developing countries currently have per capita incomes below this level, our results may not capture any peculiarities in the income-pollution relationship that occur exclusively in extremely low income settings. A quick calculation using data from the World Bank indicates that approximately 38% of the world's population in 2000 resided in countries with a per capita GDP less than $1,200 (in 1987 U.S.$). Future research into the effect of modeling assumptions on the shape of EKC at this end of the income distribution appears warranted.

[3] As in Schmalensee et al. (1998), we originally used 20 segments, each with an equal number of observations. We were, however, unable to reject the 10-segment spline model at standard significance levels ($p = 0.24$ for NO_x; $p = 0.26$ for SO_2).

[4] Stock (1989) also focuses on estimating the nonlinear portion of the PLR model, treating fixed effects as nuisance parameters.

TABLE 2.—PARAMETRIC (CUBIC SPECIFICATION) RESULTS: POOLED MODEL, FULL AND PARTIAL SAMPLE[a]

	Dependent Variable			
	Full Sample (1929–1994)		Partial Sample (1985–1994)	
Variable	NO_x	SO_2	NO_x	SO_2
Income	3.0E−05 (9.5)	1.1E−04 (13.3)	6.2E−05 (0.8)	2.5E−04 (5.5)
Income²	−2.4E−09 (−8.2)	−5.6E−09 (−7.3)	−4.2E−09 (−0.8)	−1.4E−08 (−5.0)
Income³	4.8E−14 (5.5)	1.1E−13 (4.6)	8.0E−14 (0.8)	2.7E−13 (4.7)
Estimated peak[b]	8,657 (1,850)		10,570 (18,439)	16,417 (1,863)

[a] Emissions measured in per capita levels. *t*-ratios in parentheses. "Pooled model" refers to the models estimated using data from all states. The pooled specification includes state and time fixed effects.
[b] Standard errors in parentheses; obtained via the delta method.

Prior to continuing, it is worth noting that the PLR estimates are computed using standardized data due to the nonidentification of an unrestricted intercept term. This is very similar to the conditional scatterplots (which remove state and time effects) that are presented in figures 1 and 2 in the next section, which are also centered around 0 and of roughly the same scale. However, although these conditional scatterplots are overall in agreement with the PLR estimates, the nonidentification leads to a scaling issue when comparing the parametric and semiparametric results. The scaling issue becomes less important for the middle range of the distribution of the data, corresponding to per capita incomes between approximately $15,000 and $20,000. For further details on the estimation of (2), the reader is referred to Härdle (1990).

III. Empirical Results

The parametric results from the cubic specification using the full and partial pooled samples are given in table 2. In the full sample, we find evidence consistent with an inverted-U EKC for the U.S. states for NO_x. The coefficients on the individual income terms in both the NO_x and SO_2 model are significantly different from 0 at the $p < 0.01$ level. Additionally, the income coefficients are jointly significant in each regression at the $p < 0.01$ level. These findings are consistent with coefficient estimates presented in List and Gallet (1999). For the partial sample, however, the individual estimates are only significant in the SO_2 model, although the pattern of alternating signs still holds in the NO_x model. Moreover, the cubic specification now indicates an inverted-U emissions-income path for SO_2.

For ease of comparison, we present sketches of the estimated parametric (cubic and linear spline specifications) and semiparametric paths in figures 1–4. Figures 1 and 2 illustrate the results for nitrogen oxide and sulfur dioxide emissions using the full and partial pooled samples. Vertical lines are added at the predicted peak of the parametric EKC (as estimated by the cubic specification), if such a peak exists, and at the upper and lower limit of the 95% confidence interval for the estimated parametric peak (if the

bounds are within the range of the graph).[5] As stated previously, it is important to note that the apparent flatness of the parametric EKCs in the upper right-hand panel does not imply that after a critical level of income is reached per capita income and per capita emissions fail to be negatively related; rather, the different scales across the parametric and semiparametric models render the curvature of the parametric EKCs minute. Nonetheless, coefficient estimates in table 2 and the other panels in figures 1 and 2, which separate out the parametric plots, confirm that the inverted U holds for both pollutant types in the parametric models as well.[6] Finally, specification tests presented in the next section provide formal tests for differences between the parametric and semiparametric specifications.

Overall, the empirical results in figures 1 and 2 are quite striking. First, there is little difference between the two parametric models. Second, according to the full-sample estimates, whereas the estimated NO_x peaks obtained via the parametric cubic specification and semiparametric models do not differ substantially (roughly between $8,000 and $12,000 in 1987 dollars), the estimation strategies do yield drastically disparate results in the SO_2 model.[7] In both parametric models, SO_2 emissions are a positive monotonic function of income. However, according to the semiparametric model, emissions decline with incomes over approximately $7,000–$9,000. As the PLR results appear more in line with conditional scatterplots in figure 2, it appears that the parametric results may be overly driven by a few outliers, or by a few observations in the tail of the income distribution. From a policy perspective, this difference is quite important, for the semiparametric estimates suggest that by 1977 all states were on the downward-sloping portion of the EKC (and all states except for Mississippi

---

[5] Standard errors of the estimated peaks based on the cubic specification are obtained via the delta method (Greene, 1993, p. 297).
[6] Although not presented, the slopes of the linear segments in the spline model are statistically different from 0 in the majority of segments as well.
[7] No formal statistical test exists for testing the equality of the peaks across the parametric and semiparametric specifications. In the next section, however, we do test for overall differences in the two functional forms.

FIGURE 1.—POOLED NO_x RESULTS

FULL SAMPLE

PARAMETRIC: FULL SAMPLE

PARAMETRIC: FULL SAMPLE
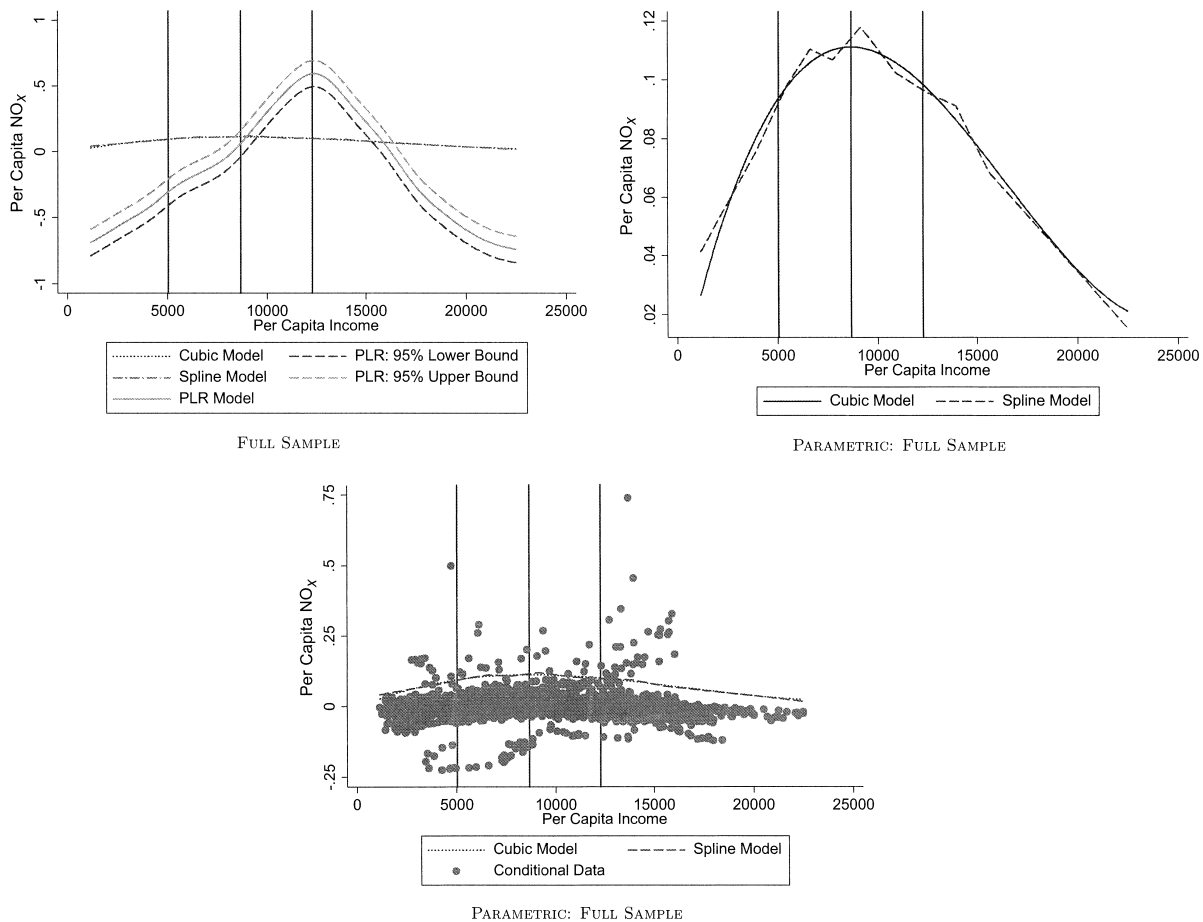
were on it by 1972), whereas the parametric estimates are much more pessimistic.

Since the data utilized in this study span a long time period and two distinct measurement schemes, as a check for robustness we reestimate the models using data only from 1985 to 1994. The emission-income relationships, also plotted in figures 1 and 2, are not overly different from those based on the full sample, with the exception that the partial sample plots cover a much smaller range of incomes. As a result, the plots primarily capture the downward-sloping portion of the EKCs documented using the full sample. Two minor differences do emerge, however. First, NO_x emissions are now invariant with respect to income according to the parametric cubic specification (see table 2). The cubic model indicates a peak at an income of roughly $11,000, but the standard error is over $18,000. Second, the parametric cubic specification now yields a peak for the SO_2 EKC. The peak occurs between approximately $13,000 and $20,000. The linear spline model, however, indicates that SO_2 emissions remain monotonically increasing in income. Despite these changes, the parametric models still paint a much more pessimistic model than the semiparametric specification.

Finally, in line with List and Gallet (1999), we allow each state to follow its own EKC. In the interest of brevity, we select nine different states and estimate state-specific parametric and semiparametric EKCs.[8] The states were chosen somewhat arbitrarily, although with an eye on the results found in List and Gallet. Estimated plots—analogous to those presented in figures 1 and 2—are presented in figures 3 and 4. The NO_x plots in figure 3 show that an inverted-U relationship typically exists between emissions and income for this particular subsample of states. In addition, the turning-point estimates are generally similar (between $10,000 and $15,000) across the states and across the various estimation methods. The results for SO_2 (figure 4), on the other hand, are quite variable across the set of states examined. For example, whereas the semiparametric EKCs for Nevada and Utah show a downward trend in emissions at higher income levels, the parametric models provide a stark contrast—suggesting that the prospect for declines in SO_2 emissions is bleak. Moreover, for states such as Missouri and Oklahoma, the parametric models continue to

[8] In the state-specific specifications the time fixed effects are replaced with a linear time trend.

FIGURE 1.—(CONTINUED)

PARTIAL SAMPLE

PARAMETRIC: PARTIAL SAMPLE

PARAMETRIC: PARTIAL SAMPLE
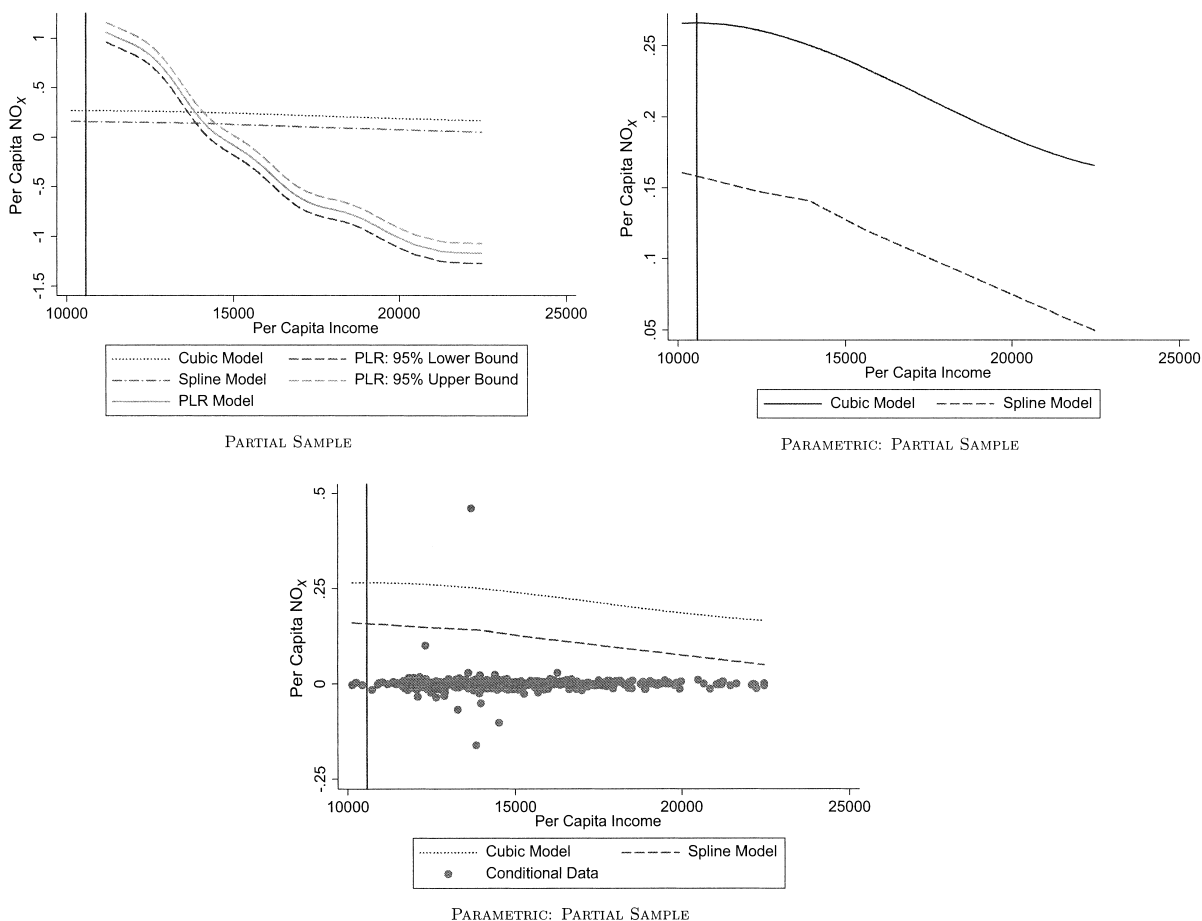
"Pooled model" refers to the models estimated using data from all states and includes state and time fixed effects. Vertical lines represent the peaks (if there is one) and 95% confidence interval (if the bounds lie within the range of the graph) according to the cubic specification model. Spline results based on 10 piecewise linear segments. Full sample uses all data, 1929–1994; partial sample only 1985–1994.

yield more pessimistic results than the semiparametric model.

IV.   Specification Tests

A.   Parametric versus Semiparametric

To formally test if the parametric and PLR models yield statistically different results, we employ a recently proposed specification test by Zheng (1996) and Li and Wang (1998). The test compares the null hypothesis of a linear regression model (cubic and spline specifications) against a PLR alternative formulation such as that suggested in Robinson (1988). Continuing with the notation from the previous section and omitting the time subscripts, the data are given by $\{P_i, X_i, y_i\}_{i=1}^n$, distributed as an i.i.d. process, where X_i includes the variables entering linearly (in our case, the state and time fixed effects). The dimension of X_i is q, and in the present case y_i is of dimension 1 (although in general may be of higher dimension). The null hypothesis is given by

$$H_0 : P_i = X_i\beta + Y_i\delta + u_i, \qquad i = 1, \ldots, n, \qquad (3)$$

where Y_i is a vector of transformed values of y_i (such as a cubic specification), and the alternative by

$$H_1 : P_i = X_i\beta + g(y_i) + u_i, \qquad i = 1, \ldots, n. \qquad (4)$$

Let $\hat{E}(P_i|y_i)$ and $\hat{E}(X_i|y_i)$ be the nonparametric kernel estimates of $E(P_i|y_i)$ and $E(X_i|y_i)$, respectively. The kernel method employed to estimate $E(P_i|y_i)$, for example, relies on a weighting scheme—the kernel function—that gives more weight to observations local to the one that defines the regression function. Specifically, an estimate of $E(P_i|y_i)$ is given by

$$E(P_i|y_i) = \frac{\frac{1}{nh^p} \sum_{j=1}^n K\left(\frac{y_i - y_j}{h}\right)P_i}{\frac{1}{nh^p} \sum_{j=1}^n K\left(\frac{y_i - y_j}{h}\right)}, \qquad (5)$$

where $K(\cdot)$ denotes the kernel function. We can use as the kernel function the product of univariate kernel functions (see Härdle, 1990), and we have opted for the product of the

FIGURE 2.—POOLED SO$_2$ RESULTS

FULL SAMPLE

PARAMETRIC: FULL SAMPLE

PARAMETRIC: FULL SAMPLE
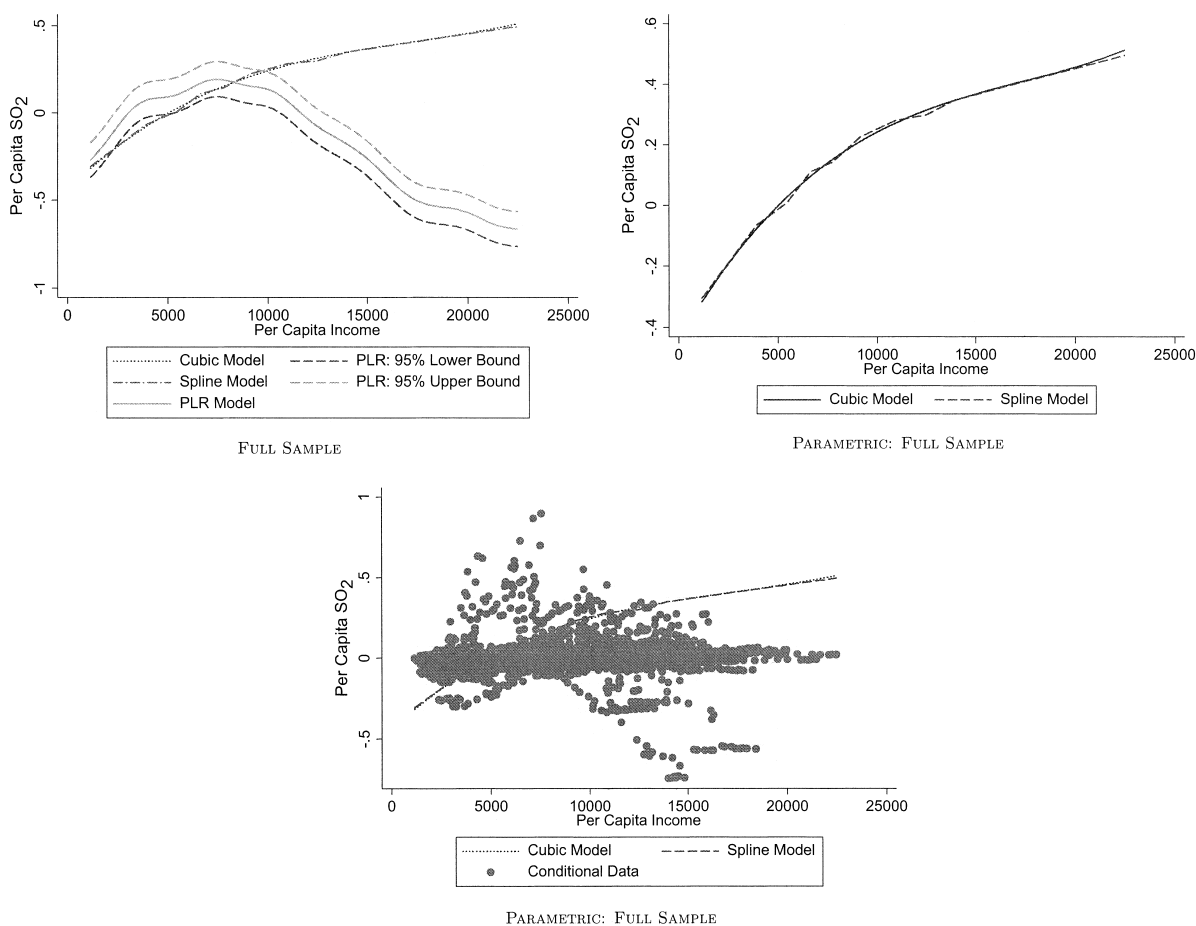
univariate Gaussian density function. The parameter h denotes the bandwidth, which is chosen to balance the bias and variance terms that enter the approximate mean squared error (MSE) expression for $\hat{E}(P_i|y_i)$. We chose h using cross-validation. Though it is well known that cross-validation is slow to converge, it is nonetheless an acceptable and widely used method; alternative methods (such as the plug-in rule) have their drawbacks as well (see for example Park and Turlach, 1992; Park and Marron, 1988).

The null and alternative hypotheses are given by equations (3) and (4) above. Under the null hypothesis, $E(u_i|X_i, y_i) = 0$ for $i = 1, \ldots, n$. Therefore, a consistent test statistic can be constructed based on $E\{u_i E(u_i|y_i)\}$, since $E\{u_i E(u_i|y_i)\} = E\{E(u_i|X_i)^2\} \geq 0$ with equality holding if and only if H_0 is true. To obtain a feasible test statistic one replaces u_i with $\hat{u}_i$, the OLS residuals from the linear regression model given by equation (3). In that case, $E(\hat{u}_i|y_i)$ can be consistently estimated using nonparametric kernel estimation techniques. The test statistic is given by

$$J_n = \frac{nh^{p/2}I_n}{\sqrt{\hat{\Omega}}}, \qquad (6)$$

where $I_n = [1/n(n-1)h^p] \sum_{i=1}^n \sum_{j=1, j\neq i}^n \hat{u}_i \hat{u}_j K_{i,j}$, $K_{i,j} = K((y_i - y_j)/h)$ is the kernel function, h is the smoothing parameter, and $\hat{\Omega} = [2/n(n-1)h^p] \sum_{i=1}^n \sum_{j=1, j\neq i}^n \hat{u}_i^2 \hat{u}_j^2 K_{i,j}^2$. The test statistic is shown by Zheng (1996) and Li and Wang (1998) to have an asymptotic standard normal distribution under H_0. However, in small samples its distribution appears to be skewed to the left, and Li and Wang suggest bootstrapping to obtain its distribution and critical values. This is the route we follow here, using 999 bootstraps to obtain the critical values.

Table 3 presents the test statistics and 95% critical values, comparing the null hypotheses of the cubic specifications with the semiparametric PLR model. Table 4 presents the analogous results for the spline model compared with the PLR model. In all cases, we clearly reject the null hypothesis of the parametric model.[9] This only confirms what we suspected from visual inspection of the graphs in the previous section; estimation strategy plays a vital role in

[9] We also conducted the same specification tests using a linear and a quadratic specification as the null. Consistent with table 3, we reject the null in all cases at conventional levels of significance.

FIGURE 2.—(CONTINUED)

PARTIAL SAMPLE

PARAMETRIC: PARTIAL SAMPLE

PARAMETRIC: PARTIAL SAMPLE
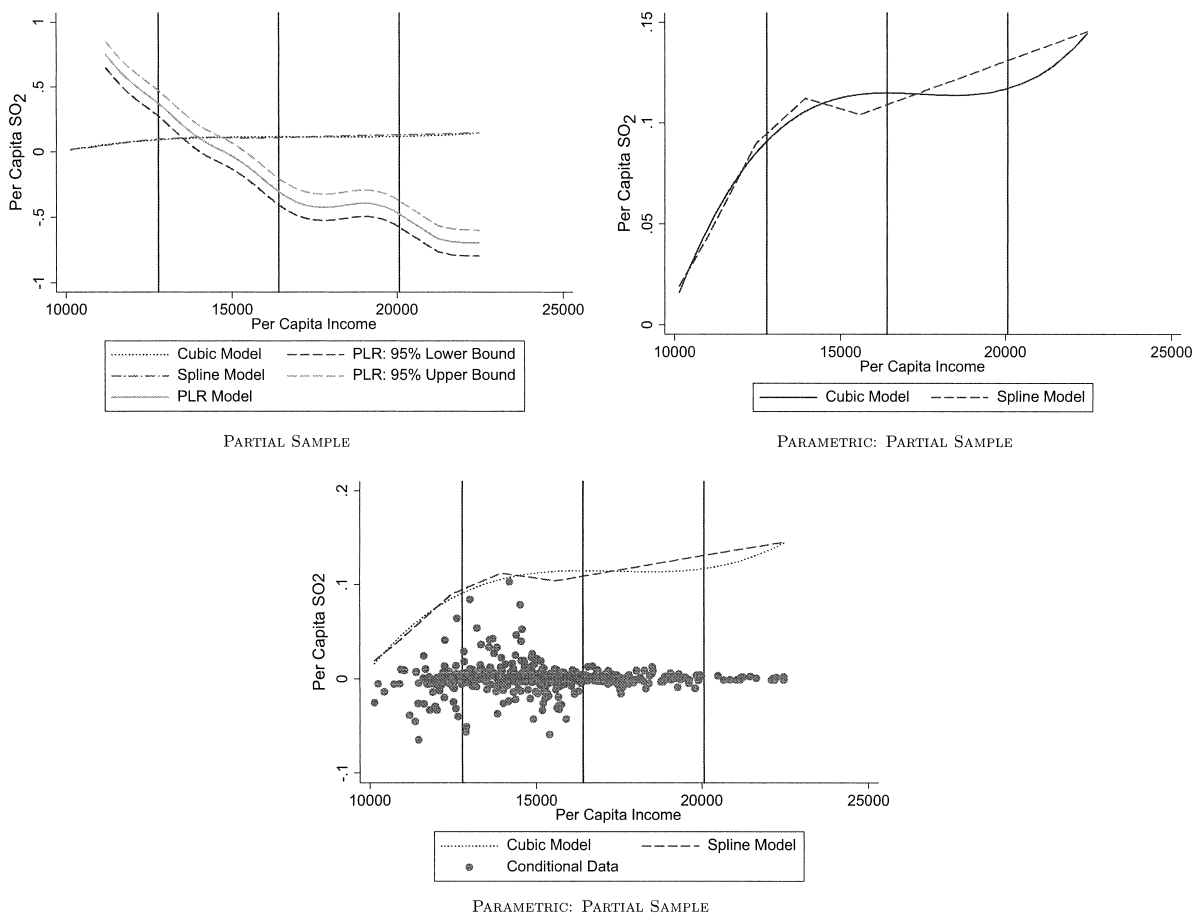
See note to figure 1.

understanding the emissions-income relationship for nitrogen oxide and sulfur dioxide emissions.

B. Serial Correlation

The assumption of i.i.d. errors in the various models is convenient and typically employed in the EKC literature. However, failure of this assumption to hold in practice may have unanticipated consequences for the reliability of the specification tests in section IV A. To test this assumption in the modeling of state-specific EKCs, we utilize a time series test of serial correlation for the PLR model recently developed by Li and Stengos (2001). Consider the time series counterpart of equation (4),

$$P_t = X_t\beta + g(y_t) + u_t,$$
$$t = -L + 1, \ldots, 0, 1, \ldots, T \tag{7}$$

where X_t and y_t are of dimensions q and 1 as before (although again y_t in general may be of higher dimension) and $g(\cdot)$ is an unknown smooth function. The hypotheses of interest are whether the error term u_t exhibits zero first-order serial correlation, or zero finite-order serial correlation.

To test zero (finite-order) serial correlation of u_t, we need an estimator of u_t. To avoid the random denominator problem associated with nonparametric kernel estimation, we estimate a density-weighted version of equation (7) above (see for example Powell, Stock, & Stoker, 1989). Let $f(\cdot)$ be the density function of Z_t, and $\hat{f} = \hat{f}(y_t)$ be its kernel estimate. Li and Stengos (2001) suggest a density-weighted test statistic for testing the absence of finite-order serial correlation, and they derive its asymptotic distribution under the stronger condition that the error u_t follows a martingale difference process. Under this condition we have $\rho_f = E(u_t f_t u_{t-1} f_{t-1}) = 0$ and $\rho = E(u_t u_{t-1}) = 0$. The construction of the test statistic for testing first-order serial correlation can be based on either of the sample analogs of ρ_f or ρ. However, it is easier to construct a test statistic based on a sample analogue of ρ_f, as opposed to ρ, because the former avoids the technical difficulty associated with the random denominator in the kernel estimation. We will use ρ_f as the basis to construct our test statistic.

FIGURE 3.—NITROGEN OXIDE EMISSIONS: SELECTED STATES
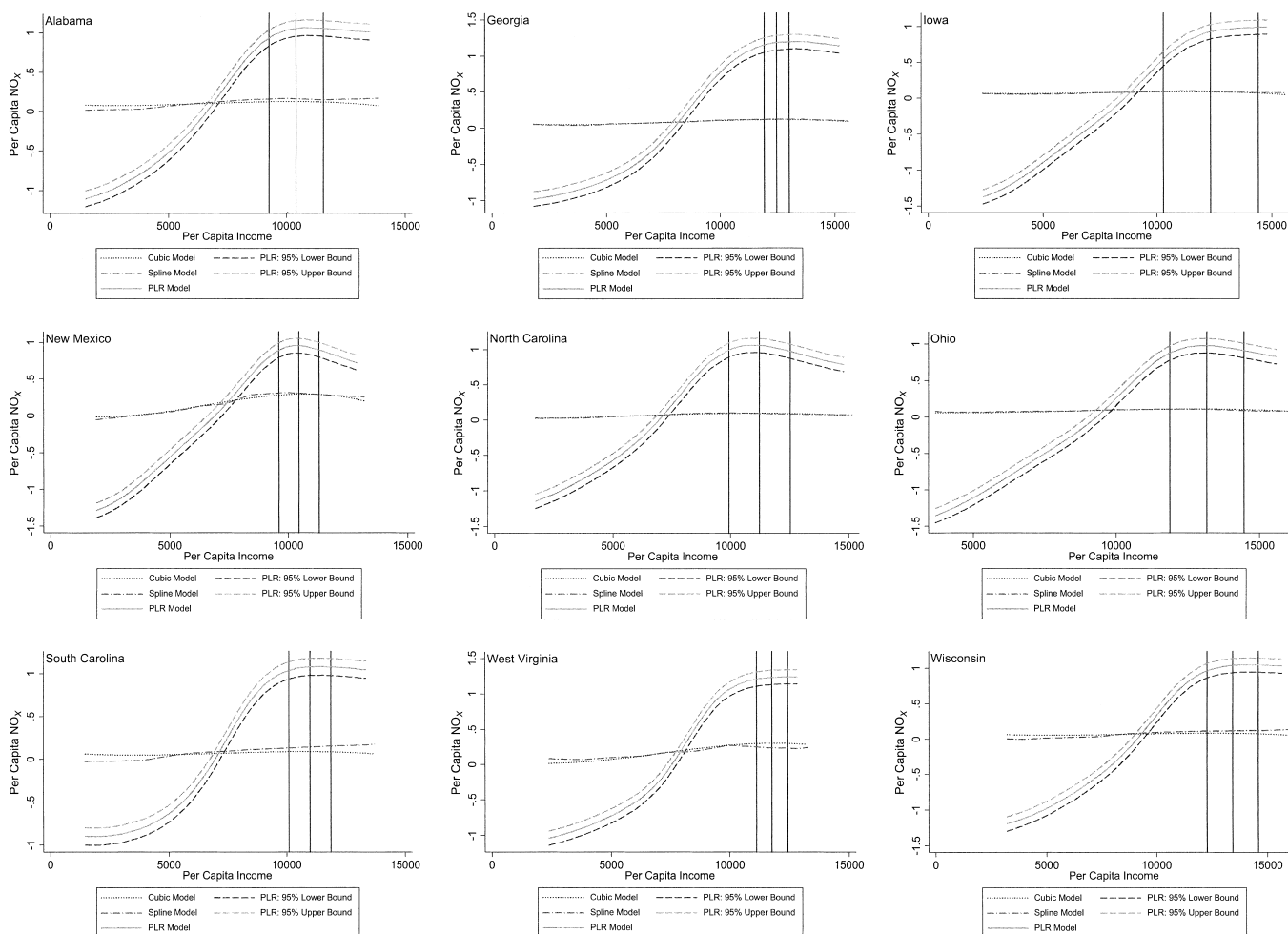
The vertical lines correspond to the peaks and 95% confidence intervals as estimated by the cubic specification. Results based on the full sample (1929–1994). Spline results based on 10 piecewise linear segments.

Therefore, our test statistic for testing zero first-order serial correlation is given by

$$I_T = \sqrt{T}\, S_{\tilde{u}\hat{f}\tilde{u}_{-1}\hat{f}_{-1}} \overset{\text{def}}{=} \frac{1}{\sqrt{T}} \sum_t \tilde{u}_t \tilde{u}_{t-1} \hat{f}_t \hat{f}_{t-1}, \qquad (8)$$

where $\tilde{u}_t$ is the residual from the semiparametric estimation of equation (7), defined as

$$\tilde{u}_t = P_t - \hat{E}(P_t|y_t) - [\hat{X}_t - \hat{E}(X_t|y_t)]\hat{\beta}$$

It can be shown that under certain mixing and regularity conditions, I_T is asymptotically normally distributed.

The above may be generalized to the case of higher-order serial correlation. In that case, if we define $\rho_s = E(u_t f_t u_{t-s} f_{t-s})$ $(s = 1, 2, \ldots, p)$, the test considers the joint null hypothesis $\rho_1 = \rho_2 = \cdots = \rho_p = 0$. Denoting $\hat{\rho}_s = 1/T \sum_t \tilde{u}_t \tilde{u}_{t-1} \hat{f}_t \hat{f}_{t-s}$ and $\hat{\rho} = (\hat{\rho}_1, \ldots, \hat{\rho}_p)^T$, the test

statistic is based on a quadratic form of $\sqrt{T}\,\hat{\rho}$ with a $\chi^2(p)$ asymptotic distribution.

Li and Hsiao (1998) considered the problem of testing zero-error serial correlation in a partially linear panel data model with a large N (N being the number of cross-sectional observations) and a fixed value of T. Li and Stengos (2001) derive the asymptotic distributions of the test statistics for testing zero first-order (or finite-order) serial correlation when T is large (N can be large or small) provided that for each i, $\{P_{it}, X_{it}, y_{it}\}$ satisfy certain mixing conditions. The test statistic for zero first-order serial correlation is a simple adaptation of the test statistic in (8) to the case of panel data:

$$I_n = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{u}_{it} \tilde{u}_{i,t-1} \hat{f}_{it} \hat{f}_{i,t-1}, \qquad (9)$$

where $\tilde{u}_{it}$ and $\hat{f}_{it}$ are defined similarly to $\tilde{u}_t$ and $\hat{f}_t$ above. We

FIGURE 4.—SULFUR DIOXIDE EMISSIONS: SELECTED STATES
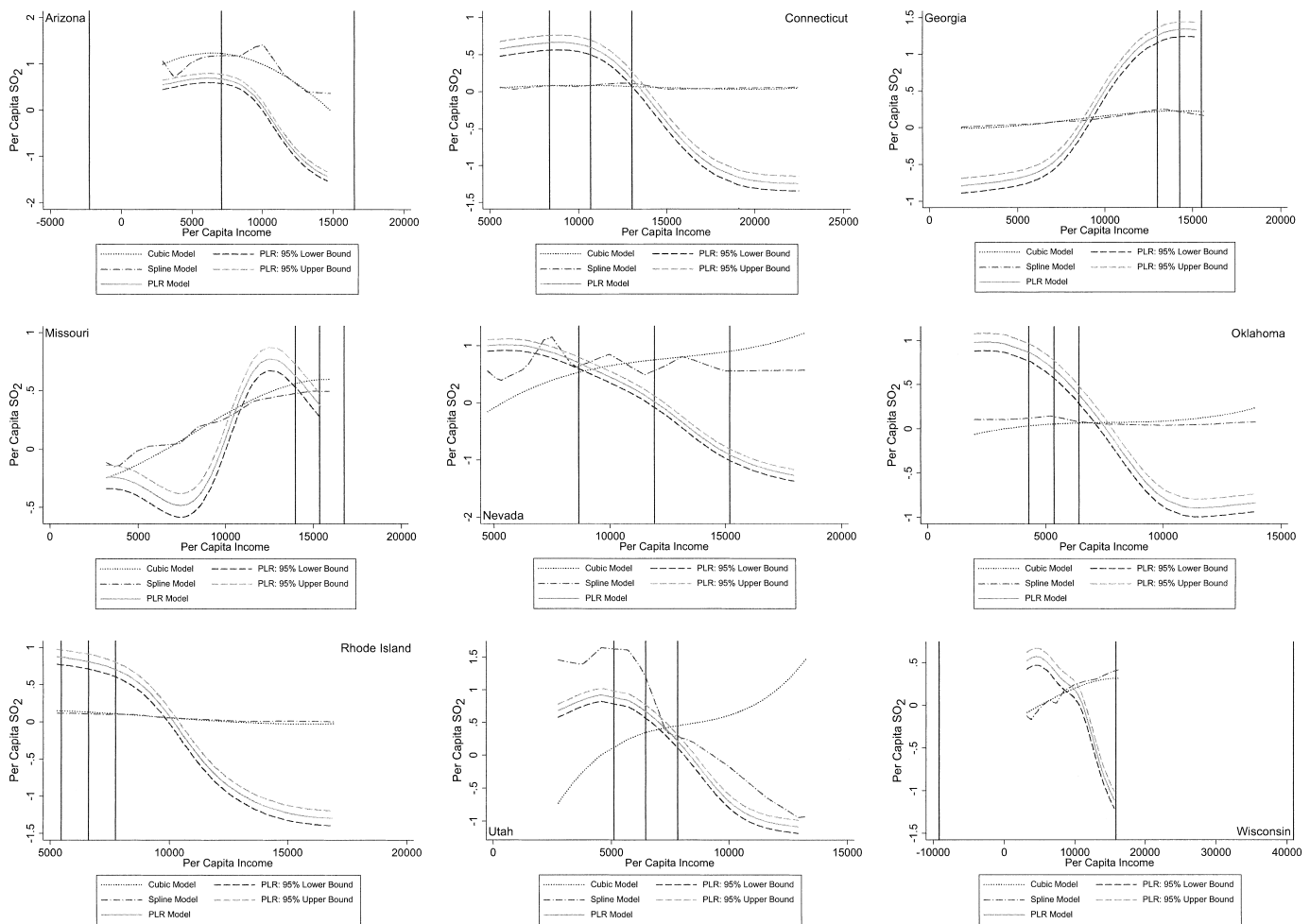
See note to figure 3.

TABLE 3.—SPECIFICATION TEST RESULTS: CUBIC SPECIFICATION VERSUS PLR[a]

| State (NO$_x$ Sample) | Dependent Variable | | State (SO$_2$ Sample) |
	NO$_x$	SO$_2$	
Pooled[b] (1929–1994)	1.8495 (0.3822)	1.7434 (0.3556)	Pooled[b] (1929–1994)
Pooled[b] (1985–1994)	1.6787 (0.3872)	1.9552 (0.4135)	Pooled[b] (1985–1994)
Alabama	3.3220 (0.4731)	2.1239 (0.4431)	Arizona
Georgia	4.3893 (0.4435)	1.1926 (0.4325)	Connecticut
Iowa	3.0176 (0.3918)	1.4175 (0.4155)	Georgia
New Mexico	1.0501 (0.4011)	1.7827 (0.4730)	Missouri
North Carolina	1.3475 (0.4222)	2.3774 (0.4053)	Nevada
Ohio	3.5470 (0.4077)	2.2324 (0.4053)	Oklahoma
South Carolina	5.7276 (0.4875)	1.7049 (0.4435)	Rhode Island
West Virginia	1.0750 (0.4522)	5.6161 (0.3988)	Utah
Wisconsin	3.3073 (0.4658)	1.8813 (0.4600)	Wisconsin

[a] Table reports the test statistic given in (6) along with the 95% critical value in parentheses.
[b] "Pooled" refers to the models estimated using data from all states. The pooled models include state and time fixed effects.

utilize (9) to test for first-order serial correlation in the pooled PLR models.

Table 5 presents the p-values associated with the null hypotheses of no first-order serial correlation in the PLR models. In all 22 cases analyzed, we fail to reject the null hypothesis of no serial correlation at conventional levels.

V. Conclusion

The literature on the environmental Kuznets curve has developed rapidly over the past five years. This paper takes a step back from this burgeoning body and examines whether traditional econometric modeling assumptions have

TABLE 4.—SPECIFICATION TEST RESULTS: SPLINE SPECIFICATION VERSUS PLR[a]

| State (NO$_x$ Sample) | Dependent Variable | | State (SO$_2$ Sample) |
	NO$_x$	SO$_2$	
Pooled (1929–1994)	1.7592 (0.4436)	1.7481 (0.4935)	Pooled (1929–1994)
Pooled (1985–1994)	1.7873 (0.4066)	1.6464 (0.4711)	Pooled (1985–1994)
Alabama	1.0892 (0.4411)	1.2897 (0.4511)	Arizona
Georgia	1.2273 (0.3922)	1.3310 (0.3991)	Connecticut
Iowa	1.4678 (0.4835)	1.2650 (0.4751)	Georgia
New Mexico	1.4308 (0.4125)	1.2105 (0.4810)	Missouri
North Carolina	0.9177 (0.3788)	0.8327 (0.5013)	Nevada
Ohio	1.3421 (0.4125)	0.9911 (0.4318)	Oklahoma
South Carolina	0.8969 (0.4611)	1.2927 (0.4091)	Rhode Island
West Virginia	0.9083 (0.3828)	1.1170 (0.4438)	Utah
Wisconsin	1.1691 (0.4017)	1.0855 (0.4663)	Wisconsin

[a] Spline model based on 10 piecewise linear segments. See table 3.

TABLE 5.—TEST RESULTS FOR FIRST-ORDER SERIAL CORRELATION[a]

| State (NO$_x$ Sample) | Dependent Variable | | State (SO$_2$ Sample) |
	NO$_x$	SO$_2$	
Pooled[b] (1929–1994)	0.8018	0.4839	Pooled[b] (1929–1994)
Pooled[b] (1985–1994)	0.1645	0.2327	Pooled[b] (1985–1994)
Alabama	0.7717	0.7550	Arizona
Georgia	0.5834	0.2067	Connecticut
Iowa	0.4578	0.2517	Georgia
New Mexico	0.3653	0.1122	Missouri
North Carolina	0.4398	0.3221	Nevada
Ohio	0.2493	0.1841	Oklahoma
South Carolina	0.2293	0.1229	Rhode Island
West Virginia	0.2941	0.2839	Utah
Wisconsin	0.2426	0.8746	Wisconsin

[a] Table reports the p-value associated with the test given in (8). The null hypothesis is no first-order serial correlation.
[b] "Pooled" refers to the models estimated using data from all states. The pooled models include state and time fixed effects.

been unduly restrictive. In the spirit of the advances made in Schmalensee et al. (1998), we use a semiparametric partially linear regression model to present three new findings. First, we find that regardless of modeling assumptions, the data follow the inverted-U shape commonly found (at the country level) between emissions or ambient pollution levels and a measure of income. Second, utilizing the test of Li and Stengos (2001) for serial correlation in the PLR model, we fail to reject the null hypothesis of no first-order autocorrelation. Finally, based on recently developed specification tests due to Zheng (1996) and Li and Wang (1998), we reject in all cases the null hypothesis of the parametric model—either a cubic or a piecewise linear spline specification—in favor of the more flexible semiparametric alternative. In particular, we find that for sulfur dioxide emissions, parametric EKC modeling is especially problematic, in that the location of the estimated peak of the EKC is quite sensitive to modeling assumptions. Given that properly estimating the peak of the EKC is invaluable (see for example Cropper and Griffiths, 1994), our findings suggest that a closer look at a good deal of the empirical modeling in this area is warranted.

REFERENCES

Beckerman, Wilfred, "Economic Growth and the Environment: Whose Growth? Whose Environment?" *World Development* 20 (1992), 481–496.

Cropper, Maureen, and Charles Griffiths, "The Interaction of Population Growth and Environmental Quality," *American Economic Review* 84 (1994), 250–254.

Fan, Jianqing, "Local Linear Regression Smoothers and Their Minimax Efficiencies," *Annals of Statistics* 21 (1993), 196–216.

Greene, William H., *Econometric Analysis, 2nd ed.* (Englewood Cliffs, NJ: Prentice-Hall, 1993).

Grossman, Gene, "Pollution and Growth: What Do We Know?" (pp. 19–50), in I. Goldin and A. Winters (Eds.), *Sustainable Economic Development: Domestic and International Policy* (Cambridge, U.K.: Cambridge University Press, 1995).

Grossman, Gene, and Alan B. Krueger, "Economic Growth and the Environment," *Quarterly Journal of Economics* 33 (1995), 53–77.

Härdle, Wolfgang, *Applied Nonparametric Statistics* (Cambridge, U.K.: Cambridge University Press, 1990).

Li, Dingding, and Thanasis Stengos, "Testing Serial Correlation in Semiparametric Time Series Models," *Journal of Time Series Analysis,* forthcoming.

Li, Qi, and Cheng Hsiao, "Testing Serial Correlation in Semiparametric Panel Data Models," *Journal of Econometrics* 87 (1998), 207–237.

Li, Qi, and Suojin Wang, "A Simple Consistent Bootstrap Test for a Parametric Regression Function," *Journal of Econometrics* 87 (1998), 145–151.

List, John A., and Craig A. Gallet, "The Environmental Kuznets Curve: Does One Size Fit All?" *Ecological Economics* 31 (1999), 409–423.

Park, Byeong U., and James S. Marron, "Comparison of Data-Driven Bandwidth Selectors," Universität Bonn Sonderforschungsbereich 303 discussion paper no. A-215 (1988).

Park, Byeong U., and Berwin A. Turlach, "Practical Performance of Several Data Driven Bandwidth Selectors," Université Catholique de Louvain CORE discussion paper no. 9205 (1992).

Powell, James L., James H. Stock, and Thomas M. Stoker, "Semiparametric Estimation of Index Coefficients," *Econometrica* 57 (1989), 1403–1430.

Robinson, Peter M., "Root-N-Consistent Semiparametric Regression," *Econometrica* 56 (1988), 931–954.

Schmalensee, Richard, Thomas M. Stoker, and Ruth A. Judson, "World Carbon Dioxide Emissions: 1950–2050," this REVIEW 80:1 (1998), 15–27.

Stern, David I., Michael S. Common, and Edward B. Barbier, "Economic Growth and Environmental Degradation: The Environmental Kuznets Curve and Sustainable Development," *World Development* 24 (1996), 1151–1160.

Stock, James H., "Nonparametric Policy Analysis," *Journal of the American Statistical Association* 84 (1989), 567–575.

U.S. Environmental Protection Agency, *National Air Pollutant Emission Trends, 1900–1994* (Washington, DC: Office of Air Quality Planning and Standards, 1995).

Zheng, John X., "A Consistent Test of Functional Form via Nonparametric Estimation Techniques," *Journal of Econometrics* 75 (1996), 263–268.