Functional Form of the Environmental Kuznets Curve

Hector O. Zapata
Department of Agricultural Economics & Agribusiness
Louisiana State University
Baton Rouge, Louisiana 70803-5604
(225) 578-2766
hzapata@agcenter.lsu.edu

Krishna Paudel
Department of Agricultural Economics & Agribusiness
Louisiana State University
Baton Rouge, Louisiana 70803-5604
(225) 578-7363
kpaudel@agcenter.lsu.edu

Charles B. Moss
Department of Food and Resource Economics
University of Florida
Gainesville, Florida 32611-0240
(352) 392-1845 Ext. 404
cbmoss@ifas.ufl.edu

November 12, 2008

Abstract
This paper provides an overview of recent econometric advances on testing of functional form with semiparametric methods and their application to the estimation of the environmental Kuznets curve (EKC). The EKC in environmental economics is estimated using panel data models and provides a fertile ground for the adoption of recent panel data developments. The paper provides a discussion of recent, and perhaps future, contributions in panel data analysis with semiparametric econometrics and highlights their usefulness in deciphering the existence of the EKC.
Introduction

This study provides an overview of the literature on the effect of economic growth on environmental quality using semiparametric and nonparametric techniques. The relationship between economic growth and environmental quality became an increasingly important point in economic development since the mid 1990s. Grossman and Krueger (1995) examine the relationship between economic growth and environmental quality during the North American Free Trade Agreement debate of the 1990s. Their major conclusion was that increased development initially led to environmental deterioration, but this deterioration started to decline as some level of economic prosperity was obtained. While the turning point varied by pollutant, the relative reduction in pollution started at income levels of less than $8,000 (in 1985 dollars). Given the similarity between the accepted relationship between income inequality and economic growth (typically referred to as the Kuznets’s Curve after Simon Kuznets) this inverted-U relationship (where the level of pollution increased until some level of prosperity is obtained) has been labeled as the Environmental Kuznets’s Curve (EKC). The formal use of the EKC first appears in the literature in Patel, Pinckney, and Jaeger (1995). Current development issues such as alternative sources of energy (biofuels, solar, wind) and global warming re-emphasize the importance of environmental quality in the pursuit of economic development.

The literature on the subject is voluminous and continues to grow, and so do the controversial findings. One issue of controversy in the existing literature is the sensitivity of the relationship between economic growth and environmental quality to individual
specific factors. Different countries may experience different stages of development and
the point at which environmental quality begins to improve may vary accordingly.
Similarly, some countries may be have been slow in monitoring environmental
degradation that data may not be available for a period long enough to reveal any
significant relationship. From an econometric perspective, new insights on the EKC
relationship may emerge from model specifications that are sufficiently flexible to allow
data properties to manifest themselves via nonparametric and semiparametric
comparisons. For instance Millimet et al., 2003 and Paudel et al., 2005 provide empirical
support for nonlinear effects between pollution and income for some pollutants but not
for others, thus finding support for more flexible semiparametric functional forms of the
EKC. Yet, it is difficult to generalize such findings without repeated samples in
experimental or simulation data. Fortunately, new Monte Carlo evidence on the spatial
and temporal dimensions of this problem has been recently published which shed light on
specification issues and that may be useful in empirical EKC research. This paper has
two objectives. First, to summarize the existing literature on model specification tests that
has been or could be implemented in EKC research, and second, to provide a discussion
of EKC research questions that can be addressed via advances in semiparametric
econometric methods.

**EKC Models**

As a starting point, Grossman and Krueger estimate the quantity of air and water
pollution \((Y_i)\) at station \(i\) at time \(t\) as

\[
Y_i = \beta_1 G_i + \beta_2 G_i^2 + \beta_3 G_i^3 + \beta_4 G_i^2 - \beta_5 G_i^2 - \beta_6 G_i^2 + \beta_7 X_i + \epsilon_i
\]

(1)
where $G_{it}$ is the gross domestic product (GDP) per capita in the country where station $i$ is located, $\bar{G}_{it-3}$ is the average GDP per capita for the previous three years, $X_{it}$ is a vector of other covariates, and $\epsilon_{it}$ is an error term. This parametric specification is sufficiently flexible to allow for the hypothesized inverted-U formulation, but it also places several significant restrictions the estimated relationship. Intuitively, the inverted-U shape results from environmental quality as a superior good. In the initial stage of develop, each individual in the society is unwilling to pay the direct cost of reducing emissions (i.e., the marginal utility of income based on other goods is higher than the marginal utility of environmental quality). However, as income grows the marginal utility of income based on other goods falls as the marginal utility of environmental quality increases. Hence, the linear specification presented in Equation 1 provides for a reduced form expression of these changes.

The most general specification of the EKC that appears in the literature is the two-way fixed effects panel data model

$$p_{jit} = \alpha_i + \phi_t + X_{it}\delta + u_{jit} \tag{2}$$

where $p_{jit}$ is concentration of a pollutant (e.g., $j=SO_2,NO_x$) in state or county $i$ in time $t$, $\alpha_i$ are specific state/country fixed effects that control location specific factors that affect emission rates; $\phi_t$ are time effects such as the common effect of environmental or other policies; $X_{it}$ is CPI adjusted per capita income in state/county $i$ in time $t$ and $Y$ is vector containing polynomial effects up to order three on per capita income (i.e., $X_{it} = (x_{it} x_{it}^2 x_{it}^3)$), $\delta$ is the associated vector of slope coefficients; and $u_{jt}$ is a contemporaneous error term. A variation of the Equation 2 is one where the polynomial
income effect is replaced with a spline function of income based on a number of preselected knots $K$ (e.g. Millimet et al. 2003; Schmalensee et al. 1998). As articulated in List and Gallet (1999), Equation 1 is a reduced form model that does not lend itself to the inclusion of endogenous characteristics of income or to causality inferences; its specification is general enough to allow for individual-specific effects (heterogeneous $\alpha$ and $\delta$), thus avoiding heterogeneity bias; lastly, state-specific time trends can capture a number of implied effects related to technology, population changes, regulations, and pollution measurement.

The hypothesis of an inverted-U relationship between economic growth and environmental quality is by definition nonlinear in income. Implicitly this nonlinearity can be approximated with a Taylor series expansion based on a low order polynomial in income, one question is whether these parametric restrictions adequately represent the nonlinearity of the EKC relationship. One alternative is to model the nonlinear effects using a nonparametric component on income while permitting fixed and time effects to enter the model

$$p_{j\mu} = \alpha + g(X_{\alpha}) + f(.) + u_{\mu}$$

where all previous definitions hold and $f(.)$ represents other variables such as population density and other social and country characteristics; a nonparametric structure for income is indicated by $g(.)$ which replaces the polynomial component in Equation 2, and $u_{\mu}$ is an error component which can take different structures. The specification of error components can depend solely on the cross section to which the observation belongs or on both the cross section and time series. If the specification depends on the
cross section then we have \( u_{it} = v_i + \varepsilon_{it} \), and if the specification assumed to be dependent on both cross section and time series then the error components follow \( u_{it} = v_i + e_t + \varepsilon_{it} \). Here \( \varepsilon_{it} \) is assumed to be classical error term with zero mean and homoscedastic covariance matrix, \( v_i \) represents heterogeneity across individuals, and \( e_t \) represents the heterogeneity over time. The nature of the error structures leads to different estimation procedures, and this is also true in the parametric specification of Equation 1.

**Estimating the EKC Relationship Semiparametrically**

A special issue of Ecological Economics (1998) reports a complete account of previous parametric EKC studies. The interest of the present survey is identifying econometric advances in the estimation of the EKC that fall mainly into the subject of semiparametric modeling and that have been published after the special issue.

List and Gallet (1999) used U.S. state level sulfur dioxide and nitrogen oxide emission to income data from 1929 through 1994 to test and demonstrate the importance of using more general functional forms of the EKC that allow for heterogeneity between per capita emissions and incomes for U.S. states. It is argued that a major advantage of using US data is that it is considered more reliable than data such as the Global Environmental Monitoring System (GEMS) often used in many cross-country studies. They also allude to the importance of using a long data period in the analysis in order to better capture upward and downward portions of the estimated EKC, which are useful for the estimation of out-of-sample turning points. Initial Hausman (1978) tests led to the rejection of random effects in favor of a fixed effects model. Using \( F \)-tests of slope homogeneity across states, List and Gallet showed that state level EKCs vary from one
State to another, suggesting that slope heterogeneity should be controlled in econometric estimations to mitigate biased and inconsistent parameter estimates. They are able to illustrate that turning point estimates from a heterogeneous panel model are much different from those obtained from a traditional model where only the intercepts are allowed to vary (isomorphic mode). This finding allows them to categorize States according to whether their per capita income turning points fall below, within, or above the 95% confidence interval relative to estimates of turning points based on the traditional (isomorphic) model. Heterogeneity and nonlinearities have been the subject of recent research work reported below.

Nonpoint source water pollutants in Louisiana watersheds were studied in Paudel et al. (2005) and turning points were estimated for nitrogen (N), phosphorus (P), and dissolved oxygen (DO) at the watershed level for 53 parishes for the period 1985 to 1999 using data collected by the Department of Environmental Quality. Parametric and semiparametric models as in Equations 2 and 3 were estimated. The parametric model is similar to Equation 2 except that population density and a weighted income variable to represent spillover effect were added to the model. One way and two way fixed and random effects models were estimated and a Hausman test used to evaluate the appropriateness of the model specifications. The best parametric model is set up as the null hypothesis and tested against a semiparametric model, that is,

\[ H_0 : \text{Equation 2} \]
\[ H_a : \text{Equation 3} \]  \hspace{1cm} (4)

Hausman (1978) is a general form of specification tests in econometrics. The fundamental idea for developing the test was that under the null hypothesis \((H_0)\) of no misspecification, there will exit a consistent, asymptotically normal and asymptotically
efficient estimator, but under the alternative hypothesis \( (H_a) \) of misspecification, such an estimator will be biased and inconsistent. Hausman developed a test based on the difference between an estimator under \( H_0 \) and another consistent estimator under \( H_0 \) and \( H_a \). Hausman proved that under \( H_0 \), the test statistic (he labeled it \( m \) ) is distributed as Chi-squared with \( K \) degrees of freedom, where \( K \) is the number of parameters in the model under \( H_0 \). The power of this test was also approximated in large samples for alternatives close to the null hypothesis (local power).

After Hausman’s contribution to specification tests, White (1982) provided a unified framework for studying the consequences and detection of model misspecification in the context of maximum likelihood estimation. He proved that a quasi-maximum likelihood estimator (QMLE) converged to a well defined limit and proposed more general statistics for robust inference. White’s sequential procedure for detecting misspecification included the popular information matrix test, which is sensitive to model misspecification, and a Hausman test based on the distance between a maximum likelihood estimator (MLE) and an alternative consistent QMLE; both tests have an asymptotic Chi-squared distribution. White also introduced a gradient test that explored the inconsistency of a supposed MLE for the parameters of interest by observing that under \( H_0 \), the gradient has zero expectation.

New more generalized methods were being developed around the idea that it is possible to formulate more flexible specifications of departures from the null hypothesis. This led to the introduction of a conditional moment (CM) test of functional form (Newey, 1985; Tauchen, 1985; among others). Newey showed that most model specification tests are special forms of the CM test, noting, however, that the power of the
CM test relies on the choice of a weighting function. Thus, the CM test is not consistent against all possible alternatives, which led Bierens (1990) to introduce a consistent CM (CCM) test. Bierens work proved that any CM test of functional forms of nonlinear regression models can be converted into a chi-square test that is consistent against all deviations from the null. The CCM test has the additional property of not depending on the randomization of test parameters that CM tests depend on. Bierens formulates a null hypothesis that the parametric specification is correct, meaning that the data generation process (GDP) characterized by the probability distribution function $F$ is such that

$$P\left( E\left( y_j \mid x_j \right) = f\left( x_j \mid \theta_0 \right) \right) = 1 \text{ for some } \theta_0 \text{ that belongs to a parameter space in } R^n. $$

The alternative hypothesis is that $F$ belongs to the class of distributions for which the above probability is less than one for all $\theta$ that belong to the parameter space. In a particular setting, the practical implementation of Bierens’ test statistic can be cumbersome and is sensitive to the choice of moment conditions, which can lead to different conclusions on model acceptance. As will be summarized in a later section, nonparametric methods provided a solution to such ambiguity and have lead to significant contributions in consistent model specification tests. Deeper and more comprehensive accounts of model specification tests in a parametric setting can be found in Godfrey (1990), Bera and Yoon (1993), Bera (2000), and Bera et al. (2001).

**Model Specification**

Most empirical EKC econometric research has been based on parametric models of either individual time series (the EKC is estimated for each unit in a cross section over time), or for panel data models. It is well known that parametric functional forms are very restrictive and that nonparametric procedures provide a means for adding flexibility to
functional form specification. The main interest of this section is to summarize the literature on model specification tests that use a parametric null hypothesis versus a semiparametric alternative. By casting the alternative into a semiparametric framework, some parametric components can remain in the model while the functional form is relaxed with nonparametric components. In addition to flexibility, semiparametric specifications offer greater estimation precision than fully nonparametric methods (Horowitz and Lee, 2002, referred to as HL hereafter). Other practical considerations in applied fields such as environmental economics are that economic models usually contain a large number of variables, and this runs directly into the curse of dimensionality problem inherent to nonparametric methods. Semiparametric models are effective in reducing the dimension of nonparametric models. Although there is a growing literature on multivariate nonparametric analysis (Scott 1992), dimension reduction simplifies estimation and testing problems and leads to more interpretable results. One last practical consideration is that sample sizes of 30 to 50 observations are not uncommon in applied work, and significant progress has been made in developing semiparametric testing procedures for small samples. This is not to suggest that semiparametric models will be the best solution for a parametric misspecification problem. As pointed out by Horowitz and Lee, care is needed in choosing the semiparametric specification that best represents a conditional mean function.

The simple linear regression model with classical assumptions is given by

\[ y_i = x_i \beta + u_i \]  

(5)
(as depicted in the original formulation of Grossman and Krueger). A related semiparametric extension considered by Robinson (1988, 1989) and Yatchew (2003) is given by:

$$y_i = x_i' \beta + f(z_i) + u_i$$

where $f(z_i)$ is the nonparametric component. It is assumed that $f(.)$ is a smooth function, and for a scalar representation, $E(u|x,z) = 0$ and the variance $V(u|z,x) = \sigma^2$. In practice, this model is estimated by removing the nonparametric effect $f(.)$ and then analyzing the parametric portion of the model in the usual way (econometric programs such as SHAZAM and GAUSS have routines for the estimation of this model via Robinson’s (1988a) approach). An example of a difference estimator for this model is found in Yatchew (2003, p. 3). Equation 6 is commonly referred to as a partial linear model and is the focus of this paper.

A summary of somewhat recent developments in specification tests in semiparametric econometric models can be found in Yatchew (2003). The book contains summaries that facilitate test implementation in applied work, and includes the work of Bierens (1990), Hardle and Mammen (1993), Hong and White (1995), Li (1994) and Zheng (1996). As stated earlier, the interest in this paper is on specification tests that have a parametric model as the specification under the null hypothesis and a semiparametric model as the alternative. The empirical interest in this approach to specification testing lies on the assumption that most econometric research in

---

1 Ellison and Ellison (2000) provide a Monte Carlo simulation comparison of various nonparametric specification tests; their suggested test statistics are based on quadratic forms in the residuals of a parametric (null) model versus a nonparametric alternative.
environmental economics, for example, is founded on economic theory and, that although
the specific functional form may not be known, the set of endogenous and exogenous
variables is known. A recent account of specification tests with these null and alternative
hypotheses may prove useful to applied researchers in revisiting previous parametric
analyses in light of these new more flexible functional forms and to others who may be
interested in generalizing some of the testing procedures.

Bierens (1982) identified two important trends in model specification testing that
were of interest at the time: a) tests using one or more well-defined non-nested alternative
specifications (e.g., Davidson and MacKinnon (1981)), and b) tests of the orthogonality
condition that the expectation of the errors conditional on the regressors equals zero
almost surely without employing a well-specified alternative hypothesis (e.g. Hausman
(1978); White (1981); Ramsey (1974)). Bierens proposed two tests for the functional
form of nonlinear regression models, without employing specified alternative hypotheses,
using the orthogonality condition and basing the tests on a Fourier transform
coloration of conditional expectations. The null hypothesis to be tested was
formulated as:

\[ H_0 : P\left[ f(x_j, \theta_0) = g(x_j) \right] = 1 \text{ for some } \theta_0 \in \Theta \]  

(7)

against the alternative hypothesis that \( H_0 \) is false:

\[ H_a : P\left[ f(x_j, \theta_0) = g(x_j) \right] < 1 \text{ for all } \theta \in \Theta . \]  

(8)

A brief history of nonparametric functional estimation techniques such as kernel and series
methods that have been developed to construct consistent model specification tests can be found
in the paper by Fan and Li (1996).
Bierens introduced two tests using Chebyshev’s inequality, with one of the tests converging in distribution to $\chi^2$ if $H_0$ is true and certain assumptions hold. Bierens conducted a limited Monte Carlo simulation to study the performance of the test. When the $H_0$ is false, both tests were sensitive to the choice of the interval length ($\varepsilon$ neighborhood), but when the null was true, the test statistics were less sensitive to the choice of $\varepsilon$. For cases of varying parameters, the $\chi^2$ test was found to be more sensitive to both $\varepsilon$ and $\beta_n$, especially when $H_0$ is false. This paper had a significant impact on the literature using consistent conditional moment model specification tests (e.g. Bierens (1990); Bierens and Ploberger (1997), among others). Bierens (1990), for instance, shows that any CM test of functional form of nonlinear regression models can be converted to a Chi-square test that is consistent against all deviations from the null hypothesis (without requiring randomization). Some of the recent work extends Bierens’ idea by using alternative consistent estimators with nonparametric methods and comparing the nonparametric model with the parametric one.

New specification tests for parametric and semiparametric models were introduced by Whang and Andrews (1993). They devote an entire section to testing a parametric null versus a semiparametric alternative (their approach is general enough to include a variety of model specifications in cross-sectional, time series, time varying, and sample selection models).\(^3\) Their results can be simply put as follows:

\[
\begin{align*}
H_0 & : \text{Parametric Model (1)} \\
H_a & : \text{Semiparametric Model (2)}
\end{align*}
\]

\(^3\) For similar null and alternative hypothesis see Hong and White (1995) and a summary of this test in Yatchew (2003, Ch.6, p. 123).
(as depicted in Equation 4). They show that under suitable conditions for these two models under $H_0$, the statistic:

$$LPT_T^d \rightarrow \chi^2_q$$

The estimation of this test statistic requires an OLS estimator for $\beta$ under $H_0$, a nonparametric function estimator ($\pi$), a sample average (indicator) of misspecification (that goes to zero under $H_0$), and an estimator of the asymptotic covariance matrix ($\Phi$) of the sample estimator (Whang and Andrews (1993, p. 304)). Whang and Andrews proved that the sample average estimator of misspecification converges to a $N(0, \Phi)$ under $H_0$ and certain sufficient conditions hold. The generalized strategy in this paper also introduces tests under autocorrelation and heteroskedasticity. Specification tests for semiparametric null models versus nonparametric alternative models are also provided in the paper.

Nonlinear regression models with nonnested alternatives are tackled in Delgado and Stengos (1994); they also discuss earlier developments, including Robinson’s (1988) semiparametric method. Delgado and Stengos developed a specification test of a parametrically specified nonlinear model against a weakly specified non-nested alternative for i.i.d. data. The competing hypotheses are formulated as:

$$H_0: E[Y|Z = z, X = x] = f(\beta_0, z) \text{ versus } H_a: E[Y|Z = z, X = x] = E[Y|X = x]$$

In this setup, the null hypothesis is a parametric model (i.e., the usual multiple regression model) while the alternative hypothesis is a nonparametric model; these competing

---

4 For an example using single index models (SIM) see Härdle et al. (2004, pp. 183-184), with extensions to other models in Chapters 7-9.
models are non-nested. To develop a test, they proposed the use of an artificial nesting procedure in terms of a parametric test given by:

\[ H_0 : \delta_0 = 0 \text{ versus } H_a : \delta_0 = 1 \] (12)

where the parameter \( \delta_0 \) artificially links the two hypotheses by means of the following artificial regression:

\[ H_i : E[Y|Z = z, X = x] = (1 - \delta_0) f(\beta_0, z) + \delta_0 g(x) \] (13)

with \( g(.) \) being an unknown function. Delgado and Stengos first formulate a case where the null hypothesis is a linear model and introduce a \( J \)-test statistic that in practice can be calculated as a \( t \)-ratio from an artificial regression as in Davidson and MacKinnon (1981). For a nonlinear null hypothesis case, they develop a \( P \)-test that can also be calculated as the \( t \)-ratio from an OLS regression. A Monte Carlo simulation was designed to study size and power properties of the tests in small samples \( (n = 25, 100, 500) \); it was found that the test performed well in terms of size and power in \( iid \) data, leaving dependent data cases to future research.

A significant new concept in the development of consistent specification tests with semiparametric models was that of Fan and Li (1996), who advanced the use of Central Limit Theorems for degenerate \( U \)-statistics of order higher than 2 for the construction of such tests. Fan and Li argue that previous tests used ad hoc modifications to overcome the “degeneracy problem.” The null and alternative hypotheses for a test of a semiparametric partially linear model are given by:

\[ H_0 : g(x) = z\gamma + \theta(w) \text{ a.e. for some } \gamma \in \mathbb{R}^q \text{ and some } \theta(.) : \mathbb{R}^q \to \mathbb{R} \]

\[ H_a : g(x) = z\gamma + \theta(w) \text{ for all } \gamma \in \mathbb{R}^q \text{ and } \theta(.) : \mathbb{R}^q \to \mathbb{R}, \text{ and } x = (w', z') \] (14)
Fan and Li adopt a two-step estimation procedure similar to the one in Robinson (1988) to derive the following statistic under the null:

\[ T^b = \frac{n h^{3/2} I_n^{d, d} }{ \sqrt{2\hat{\sigma}^2} } \to N(0,1) \]  

(15)

where \( \hat{\sigma}^2 \) is a consistent estimator of the variance of \( I_n^b \) which converges to a normal distribution with mean 0 and variance \( 2\sigma_b^2 \). In practice, the null hypothesis would be rejected if \( T_b \) were greater than the upper \( \alpha \)-percentile of the standard normal distribution.

Li and Stengos (1996) considered the estimation of a general partially linear (Robinson, 1988) semiparametric panel data model, when regressors are correlated with the errors, via instrumental variable estimation. The model is given by

\[ y_{it} = x_{it}' \beta + \theta(z_{it}) + u_{it}, \quad (i = 1, \ldots, N; t = 1, \ldots, T) \]  

(16)

where \( x_{it} \) and \( z_{it} \) are of dimension \( p \) and \( q \) respectively, with an unknown coefficient vector and functional form \( \beta \) and \( \theta(.) \), respectively. Li and Stengos give this model empirical flavor by allowing some or all the components \( x_{it} \) to be correlated with the error \( u_{it} \) and the data to be large in \( N \) and small in \( T \). By adopting Robinson’s (1988) estimator, they obtain a consistent (infeasible) instrumental variable estimator for \( \beta \), which is made feasible by nonparametric estimation (kernel method via density-weighted) of the involved conditional expectations.

If a set of assumptions hold, then various specifications (error components, serial correlation, conditional heteroskedasticity of unknown form) are allowed. Also, conditions are defined for root-\( N \)-consistency of the nonparametric estimator for \( \beta \).
Simultaneously, Li and Stengos (1995) followed up the above work by introducing a \( J \)-
type test for non-nested panel data models with a semiparametric structure (see Li and
Stengos (1996)) as the alternative hypothesis. The null model is given by:

\[
y_{it} = \lambda y_{i,t-1} + x_{it}' \beta + u_{it}, (i = 1, \ldots N; t = 1, \ldots T) \tag{17}
\]

Under the alternative hypothesis of \( H_1 \), the model is specified as:

\[
y_{it} = \lambda y_{i,t-1} + \theta(z_{it}) + \nu_{it} \tag{18}
\]

with \( x \) and \( z \) non-nested and the unknown parameter vector \( \beta \) is of dimension \( p \times 1 \).

One assumption is that \( N \) is large and \( T \) is small or of moderate size so that the
asymptotic results are for the situation where \( N \rightarrow \infty \) with a fixed value of \( T \). As in
Delgado and Stengos, Li and Stengos formulate a linear combination of the two
hypotheses to obtain:

\[
y_{it} = \lambda y_{i,t-1} + x_{it}' \delta + \alpha \theta(z_{it}) + u_{it} \tag{19}
\]

where \( \delta = (1 - \alpha) \beta \) so that by restricting \( \alpha \) to either 0 or 1 one obtains the null or the
alternative model. Under some regularity conditions, they show that their test statistic \( \hat{I} \)
is a consistent test against the alternative \( H_a \) and converges to a standard normal
distribution. A Monte Carlo simulation for small samples \( (N = 50,100 \text{ and } T = 4) \), with
2000 replications, showed that the test had good estimated size and power for the DGPs
that were simulated.

A computationally appealing consistent test that combines moment conditions as in
Bierens (1990) and nonparametric (kernel) methods was introduced by Zheng (1996).
Zheng’s test is a residual regression test of specification and is based on a \( U \)
statistic
(Fan and Li (1996)) that can be transformed to a standard normal distribution for one-
sided testing (Yatchew, 2003, p.123). Zheng performed a Monte Carlo experiment to assess small sample \((n = 100, 200, \ldots 700\) observations) for a linear null model versus a linear one \((H_0\) true for this model), a linear model with interactions to check test power for higher order terms, and two nonlinear models (power check against nonlinearity). Because the simulation models are low order, a bivariate standard normal density function is used as the kernel function. It was found that the test has adequate size, the size is not very sensitive to the choice of bandwidth, and is more powerful than Bierens’ (1990) consistent test and many of the nonparametric cases. Zheng points out that the test can be used in a semiparametric setting. It can also be used to test distributional assumptions in the parametric binary choice, censored regression, truncated regression, and sample selection models. Extensions to heterogeneity, autocorrelation, and some other time series problems were highlighted also.

The consistent test by Zheng was considered and a bootstrap method proposed to approximate its null distribution in Li and Wang (1998). This approach did not require the estimation of the alternative model, and no consistent estimators for the variance of the error terms in the regression model under the alternative hypothesis was required. This paper is perhaps the first analysis that considered testing a parametric null model against a semiparametric partially linear model that is in tune with most empirical modeling in environmental economics, that is, models with many regressors and small sample sizes. This is also a fairly complete paper in terms of references for readers interested in expanding on nonparametric and semiparametric testing. A consistent test for testing a null hypothesis such as (1) versus an alternative hypothesis such as (2) is
constructed as follows.\footnote{Li, Hsiao and Zinn (2003) provide an illustration of hypothesis testing with a parametric null and a semiparametric alternative based on series estimation methods in partially linear models. They claim that series estimation of their tests and bootstrapping are simpler to compute than in the case of kernel methods.} Under the null hypothesis, since \( E(\varepsilon_i | z_i, x_{2i}) = 0 \), for \( i = 1, 2, \ldots n \), we have that:

\[
E\left[ \varepsilon_i E(\varepsilon_i | x_{2i}) f(x_{2i}) \right] = E\left[ E\left( \varepsilon_i z_i \right)^2 f(x_{2i}) \right] = E\left( \left[ f(x_{2i}) - f(x_{2i}, \theta_0) \right]^2 f(x_{2i}) \right) \geq 0 \tag{20}
\]

with equality holding iff \( H_0 \) is true. A feasible tests statistic can be obtained by replacing \( \varepsilon_i \) with \( e_i \), the OLS residuals from the linear regression in model (1), in which case, a nonparametric kernel estimation method can be used to consistently estimate \( E(\varepsilon_i | x_{2i}) \).

Thus, a consistent statistic can be constructed based on the \( e_i \). The test statistic is given by:

\[
J_n = \frac{nh^{p/2}I_n}{\sqrt{\Omega}} \sim N(0,1) \text{ under } H_0 \tag{21}
\]

where \( n \) is the number of observations, \( h \) is the bandwidth (usually chosen using cross-validation), \( p \) is the dimension of the vector of unknown parameters \( \beta \), and the denominator is the square root of a consistent estimator of \( \Omega \), the asymptotic variance of \( nh^{p/2}I_n \), and \( I_n \) is given by:

\[
I_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \frac{1}{h^p} e_i e_j K \left( \frac{x_{2i} - x_{2j}}{h} \right) \tag{22}
\]

with \( K(.) \) the kernel function. The Monte Carlo simulation in Zheng supports good finite sample properties, but better results were associated with samples of size 300 or greater.
$J_n$ was found to have small sample skewness, and therefore, Li and Wang (1998) suggest using bootstrapping to obtain its distribution and critical values. They found that the bootstrap method provides a much more accurate approximation to the test than the asymptotic theory result even for samples as small as 50 observations. It was observed, as expected from bootstrap theory (e.g., Yatchew, 2003), that when the error is conditionally homoskedastic, the naïve and the wild bootstrap work well. When the error is conditionally heteroskedastic, however, wild bootstrap outperforms the naïve bootstrap significantly. It was also found that this test has a quite stable estimated size over a wide range of smoothing parameter choices, suggesting that in practice, general cross-validation or similar criteria can be used to choose $h$. However, they point out that choosing $h$ optimally in the sense of maximizing the power of the test under the alternative hypothesis is left for future research.\(^6\)

An extension of this model to a semiparametric partially linear panel data model (Equation 2) with serially correlated errors is proposed in Li and Hsiao (1998). Again, the case considered is that for which $N$ is large and $T$ small, and the asymptotics are derived consistent with this assumption. The proposed approach using Robinson’s (1988) kernel estimation for the conditional expectations (on $x_2$) and subsequently estimate the unknown coefficients on the parametric part by eliminating the estimated nonparametric

\(^6\) A recent evaluation of nonparametric specification testing with bootstrap methods is found in Lee and Ullah (2001). They consider three nonparametric tests for functional form: 1) sums of squared residuals from the null and alternative models, 2) fitted values of the null and alternative models, and 3) nonparametric conditional moment tests. One of the main conclusions is that the Li and Wang (1998) and Zheng (1996) tests work best; these tests had good size and power properties in the DGPs they simulated. The Effect of optimal bandwidth choice on these tests performance was not evaluated. See also the specification testing section in Ullah and Roy (1998), and Baltagi’s (1995, 1996) work.
component. As opposed to Robinson (1988), who proposed trimming small values in density estimation, Li and Hsiao use a density-weighted approach to avoid the issue of using a random denominator associated kernel estimation. They propose a root-$N$-consistent estimator as in Robinson using all observations ($n = NT$) and proceed to developing the statistic under the null hypothesis:

$$H_0: \text{The error in the SPLR follows a martingale difference process.} \quad (23)$$

Li and Hsiao show that under certain conditions the test statistic for zero first-order correlation is given by

$$I_n = \frac{1}{\sqrt{NT}} \sum \sum e_{i,t} e_{i,t-1} \hat{f}_{i,t} \hat{f}_{i,t-1}$$ \quad (24)

where $e_{i,t}$ are the estimated residuals from the semiparametric model in Equation 6 using the adapted Robinsons’ (1988) procedure, and the estimated $f_{i,t}$ functions are the cross-sectional data (for a fixed value of $t$) kernel densities of $x_{2,i}$. Note that to be able to calculate the residuals, the conditional expectations of $y$ and $z$ are also needed (Li and Hsiao (1998)). It can be shown that:

$$I_n \overset{d}{\rightarrow} N\left(0, \sigma_0^2\right), \text{ and } S_n = \frac{I_n}{\hat{\sigma}_0} \overset{d}{\rightarrow} N(0,1) \text{ given a consistent estimator of } \hat{\sigma}_0^2 \quad (25)$$

Li and Hsiao (1998) provide the formula for the consistent estimator for the variance of $I_n$ and expand the test to higher order serial correlation. It can also be shown that the asymptotic distributions of this tests statistic remain unchanged when $T$ is large (Li and Stengos (2003)). The $I_n$ test for zero first-order serial correlation was generalized by Li and Hsiao to test for the absence of individual effects in panel data. They introduce a $J_n$ statistic which is a density weighted version of the parametric counterpart (Breusch-
Pagan (1980)), and they show that under similar conditions used to derive the distribution of \( I_n, T_n = J_n / s \) (\( s^2 \) consistent estimator of the variance of \( J_n \) and \( s \) is the square-root of \( s^2 \)) converges in distribution to a standard normal (i.e., \( N(0,1) \)). Simulation results in Li and Hsiao showed that both tests performed well in small samples (\( N = 50,100,200 \)) and had good size when the errors are homoskedastic. The \( S_n \) test overestimates the sizes for certain DGPs when errors are conditionally heteroskedastic. The \( T_n \) test, however, was powerful in detecting one-way error structure but not powerful against an AR(1) error process. Simulation results in Li and Stengos suggest that for \( T = 100,200, \) and \( 500 \) the \( I_n \) test has quite good size, approximating their nominal size as \( T \) increases in DGPs that included white noise (the null), AR(1), MA(1) and a conditionally heteroskedastic process. Only slight changes in size were found to a 20% increase (decrease) in smoothing parameters. Although the test was less powerful when conditional heteroskedasticity was present, they still were powerful in detecting first-order serial correlation in the DGPs used in the Monte Carlo simulation. In changing the sample conditions to large \( T \) and \( N \) small, Li and Stengos (2003) show that \( J_n \) no longer has an asymptotically normal distribution.

General regression models with time varying coefficients have been of interest to applied researchers for some time. A prime example would be the estimation of stochastic frontier models used to study technical efficiency. Li et al. (2002) introduced a
semiparametric smooth coefficient model, the smooth function of which can be estimated via local least squares with a kernel weight function. The model is given by: \(^7\)

\[ y_i = f(z_i) + x_i'\beta(z_i) + u_i \]  

(26)

By setting \( \beta(z_i) = \beta \) this model reduces to the same form as Equation 2, that is, Robinson’s (1988) semiparametric model. The parametric part of the model (model (1)) can be set as the null model and the above model as the alternative. These hypotheses can be formulated as:

\[ H_0 : \delta(z) - \delta_0(z) = 0 \text{ a.e. versus } H_1 : \delta(z) - \delta_0(z) \neq 0 \]  

(27)

on a set of positive measure.

Li et al. use the concept of integrated squared differences \( I \) as the basis for the test. This concept is defined as the integral of the squared differences between the two estimators in the hypotheses. Under \( H_0 \), \( I = 0 \) and \( I > 0 \) under \( H_1 \). When the null model is a simple linear regression, least squares can be used to estimate \( \delta_0(z) \). Li et al. propose using local least squares to estimate \( \delta(z) \). To avoid the random denominator problem associated with this estimator, they introduce a weighted version of \( I \) with a kernel function \( D_n(z) \) as the weight function. The test statistic so derived \( (I_n) \) involves a \( q \)-dimensional integral and a nonzero center term under \( H_0 \). Li et al. propose a solution that eliminates the need for integration and removes the nonzero center term to obtain the following test statistic:

\(^7\) A test of a parametric (or semiparametric) null hypothesis against a nonparametric alternative can be found in Aït-Sahalia, Bickel and Stoker (2001). The authors indicate how the test statistic can be written as the difference in residual sums-of-squares under the null and alternative hypotheses.
\[
\hat{I}_n = \frac{1}{n^q h^q} \sum_{i} \sum_{j \neq i} x_i' x_j e_i e_j K \left( \frac{z_i - z_j}{h} \right),
\]

where \( K(\cdot) \) is a second order kernel function and the residuals \( e_i = y_i - d_0(z_i) \). In the case of a simple linear regression model, \( d_0 \) is the least squares estimator of \( \delta_0 \). Under certain conditions, Li et al. show that:

\[
J_n = \frac{n^{q/2} \hat{I}_n}{\sigma_0} \xrightarrow{d} N(0,1)
\]

In the denominator of \( J_n \) a consistent estimator of the true variance of \( \hat{I}_n \) is used (Li et al., p. 415). In practice, \( H_0 \) is rejected when \( J_n \) is greater than upper \( \alpha^{th} \) percentile from a standard normal distribution. Note that this test is very similar to that in Zheng (1996), although Zheng uses a conditional moment method in the derivation. Li et al. also note that \( J_n \) is sensitive to the choice of the bandwidth. The application of this model and test are illustrated via a production function example, using 1,406 firms, where two parametric null models are considered, a Cobb-Douglas production function and a translog model that allows for interactions. The application favors the semiparametric model over the parametric null models. Estimation results and tests can be applied to time series data provided the regressors are nonnested and the data are \( \beta \)-mixing with certain decay rates.

**Poolability**

The question of whether a fixed effects panel data model (pooling) is appropriate has received limited attention in the EKC literature. Criado (2008) argues that in most applications, no formal tests of the homogeneity assumption is conducted on the time
(stability of the cross sectional regressions over time) and space (stability of the cross sectional regressions over individual units). Existing literature on the subject has generated mixed results. Criado tests poolability in the EKC by testing the adequacy of such an assumption on both dimensions via nonparametric tests robust to functional misspecification using models similar to those in equations (1) and (2). The data set is a balanced panel of 48 Spanish provinces over the 1990-2002 period and the pollutants include methane, carbon monoxide and dioxide, nitrous oxide, ammonia, non methane volatile organic compounds, nitrogen and sulphur oxides. Poolability tests on the spatial dimension (spatial heterogeneity) reject it, particularly for nonparametric specifications. Time poolability (temporal homogeneity) results were mixed; it holds for three of four air pollutants in Spanish provinces and the estimated pooled nonparametric functions reflected inverted U shapes. It was also pointed out that the parametric and nonparametric tests overwhelmingly rejected the null hypothesis of spatial homogeneity and fixed effects, and that failure to recognize this property of EKC panel data would lead to mixed findings. The work suggested future EKC research with advances in parametric and nonparametric quantile regression, random coefficient modeling, and panel heterogeneity. In similar research, Azomahou et al. (2006) use the local linear kernel regression to estimate $W(x_u)$ with $x_u = (x_u, x_{i,t-1})$. They claim that the local linear (polynomial of order 1) kernel estimator performs better than the local constant (polynomial of order 0) kernel estimator or Nadaraya–Watson estimator, since it is less affected by the bias resulting from data asymmetry (notably at the boundaries of the sample). They use standard univariate Gaussian kernel and marginal integration to estimate the nonparametric method. To select the bandwidth in the nonparametric
regression, they used a least squares cross validation method. To develop the confidence interval of the estimated function, they used a wild bootstrap method. To test for the suitability of nonparametric vs parametric functional form, they used a specification test developed by Li and Wang (1998).

**Redefining the EKC Model**

The traditional dependent variable in EKC model is pollutants. Lately Bugliani et al. (2008) indicate that this consumption based measure of EKC may not be a good variable to use. They use ecological footprint data instead of pollution and estimated an EKC model for year 2001 using cubic and quadratic functional forms. Canas et al. (2003) use direct material input as a dependent variable instead of pollutant to find EKC of 160 countries in panel parametric model with quadratic and cubic functional forms. Another alternative to pollution as a dependent variable has been replaced by efficiency scale. For example, Zaim and Taskin (2000) calculate efficiency score assuming desirable output and simultaneous production of undesirable output in the production function. They then use this efficiency score to find if EKC exist for CO₂. Similar to a concern on what should be used as a dependent variable, several authors have raised concern about what should be used as independent variable in the regression model. The major concern has been the reduced form nature of the equation. Stern (2004), Copeland and Taylor (2004) and Auci and Bechhetti (2006) have shown this concern. These authors have emphasized on a need to use a decomposition approach where by scale effect, composition effect and abatement effect should be recognized properly. Those who have used panel data have also recognized the need to do panel unit root and panel cointegration tests so that the estimated parameters hold desired properties.
One simple way of testing these hypotheses was proposed by Hong and White (1995), denoted at $T_n$ which converges to a standard normal distribution under $H_0$ and is rejected for large values of $T_n$. This test is similar to the $J_n$ test of Li and Wang and Zhen previously discussed. The main finding of this study was that the semiparametric model seems to work best for phosphorous for not for nitrogen and dissolved oxygen. For some nonmetropolitan locations, pollution continues to increase signaling a need for stricter environmental control. Turning point estimates appeared sensitive to parametric and semiparametric specifications. It was surprising that such a disaggregated analysis would generate results that are consistent with the mixed findings reported in cross country analyses.

Taskin and Zaim (2000) estimated nonparametric relationship between GDP and CO2 emission efficiency parameter using Kernel nonparametric regression techniques, where no a priori restriction on the functional form and the degree of polynomial is imposed on the structure of the model. Here, the assumption of a particular form for the conditional mean in the parametric estimation is replaced with the assumption that the conditional mean comes from a dense class of functional forms consisting of twice continuously differentiable function. The authors used the Nadaraya–Watson nonparametric kernel estimator to depict the functional relationship between the environmental efficiency index and per capita income. They chose Epanechnikov kernel for the kernel function. Results obtained from nonparametric analysis between the relationship between CO2 efficiency and GDP were similar to a cubic functional form in parametric functional form.
Frazer (2006) estimates overlapping nonparametric regression to describe the relationship between inequality and GDP. Suppose inequality is function of GDP, \( \text{inequality} = f(\text{GDP}) \). Further they use \( f(x) = f(y|x) \) function and estimated \( f(x) \) equation using a local linear least square regression. At each point of \( x \), the author ran a weighted linear regression of the inequality measure on the income per capita measure, with the weights chosen to be large for sample points that are close to \( x \), and the weights chosen to be very small for points further from \( x \). The estimate is then given by the constant term that results from such a regression. Specifically, the weights used are from a normal (Gaussian) density function. The bandwidth parameter used determines the amount to diminish the weight given to distant points in the regression which they have selected based on Silverman (1986). Then, symmetric confidence intervals are calculated using a bootstrap method. They claim Kuznets curve is an unconditional relationship between inequality and income level so they use partially linear model per Robinson’s specification (1988). The conditional expectation is calculated using a linear nonparametric least square regression for variables other than the income variable.

**Partial Linear Model and Serial Correlation**

Rather than specifying a panel data model of heterogeneity on income as in List and Gallet, Millimet et al. (2003) advance that the appropriateness of a parametric specification of the EKC should be based on the formulation of an alternative hypothesis of a semiparametric partial linear model (PLR). This idea is pursued using the same panel data as in List and Gallet, and estimations are reported for sulfur dioxide and nitrogen oxide for the entire sample (1929-1994) and for a partial sample (1985-1994). A model specification test of Zheng (1996) and Li and Wang (1998) was used to test parametric
(equation 1) and semiparametric (Equation 2) models of the EKC. The parametric specification is a two-way fixed effects panel data model. The semiparametric model follows Robinson (1988) root-N consistent estimates of the intercepts and time effects in equation 2, conditional on the nonlinear income variables; the standard Gaussian density was used in local constant kernel estimation and cross-validation generated the smoothing parameters. As in List and Gallet, individual State EKCs were calculated for cubic parametric and semiparametric models. Convincing results were reported in favor of adopting model specification tests of the EKC to decipher whether the implications from parametric models were statistically different from those generated from semiparametric EKCs. The hypotheses of interest are given by equation 26 and the test statistic is $J_n$ which has an asymptotic normal distribution under $H_0$. Because of small sample skewness, bootstrapping of critical values is usually required. Millimet et al. provided results for the PLR and a spline model (for $H_a$) and the conclusion favors the semiparametric model of the EKC over the parametric one. State specific EKCs are based on time series data; thus, Li and Stengos test for first order serial correlation in a PLR was estimated using a density-weighted version of Equation 3 above (this avoids the random denominator problem associated with nonparametric kernel estimation), and it was adapted to a panel data model (Li and Hsiao, 1998) to the statistic $I_n$ discussed previously. The results favor the null model of no serial correlation in this data set. An relevant policy finding of this study is that the location of the peak of the EKC is sensitive to modeling assumptions, a finding consistent with the heterogeneity results in List and Gallet.
Partial Linear Model with Heteroskedasticity

Roy et al. (2004) used a semiparametric model to examine the EKC for carbon monoxide (CO), ozone (O₃) and nitrogen oxide. The estimation technique in this application adjusted the standard PLR to allow for heteroskedasticity (Robinson, 1988) and tested a quadratic parametric model against the semiparametric model using the Li and Wang (1998) test. As opposed to most previous applications, the variables are expressed as the natural log of a particular pollutant and income. Because this is a panel data specification, a generalized local linear estimator (Henderson and Ullah, 2005) is used. Roy et al. started the analysis by first considering linear, quadratic, and cubic models of income for each pollutant and analyzed the statistical significance of income; they found that income was significant in some models but not in others. This led to the specification of the semiparametric model as a more flexible specification alternative. The main result of this study is that the quadratic model is strongly rejected in favor of the semiparametric specification, and similar results are obtained for estimates of the income elasticities.

Partial Linear Model with Smooth Coefficients

Deforestation can quickly deteriorate the quality of the environment, and in the process of economic development, most developing countries must confront local (loss in biodiversity) and global (carbon sequestration) dimensions of such environmental degradation. Van and Azomahou (2007) investigated nonlinearities and heterogeneity in the deforestation process with parametric and semiparametric EKCs, and their focus is on whether the EKC exists and identify the determinants of deforestation. The data set was a panel of 59 developing countries over the period 1972-1994. The EKC if first estimated as a quadratic parametric model with deforestation rate as the dependent variable and
GDP per capita and other variables as independent variables. F tests of fixed time and
country effects tested supported a fixed country effects model. A Hausman test supported
the existence of a random effects model relative to a fixed effects specification; however,
the overall specification was insignificant. In order to check the robustness of the
functional form between deforestation rate and GDP per capita, a semiparametric fixed
effects model was estimated (as in Paudel et al.). The salient finding was the
nonexistence of an EKC for the deforestation process. The analysis was extended to
investigate whether other variables (e.g., population growth rate, trade ratio
((imports+exports)/GDP), population density, the literacy rate, and political institutions)
may be more relevant in the determination of deforestation and a model similar to
Equation 2 was estimated. Contrary to the previous case, the data supported a fixed
effects model, with many of the new variables were significant, and a within estimator
was preferred to a first difference estimator. A semiparametric model such as Equation 3
was specified, with GDP assumed to enter nonlinearly in the nonparametric function
$g(\cdot)$. The method of Robinson (1988) was used to estimate a first difference
representation of Equation 3 but the results did not support the existence of an EKC. It
was hypothesized that perhaps modeling bias could be reduced by specifying a smooth
coefficient model (e.g., Li et al. 2002) that captures the influence of GDP on
deforestation rates depending upon the state of development of each country. The model
is given by:

$$p_{it} = \alpha(z_{it}) + x_{it}'\beta(z_{it}) + u_{it}$$

(30)

where $\beta(x_{it})$ is a smooth function of $x_{it}$. Note that when $\beta(x_{it}) = \beta$ the model reduces
to a standard PLR (Equation 3). Having a nonparametric effect ($z_{it} = GDP$) on the
deforestation rate and varying coefficients on other determinants of deforestation \( (x_n) \) allows the assumption that GDP per capita can have a direct effect and a nonneutral effect, respectively, on the deforestation rate. The model specification tests \( (H_0 \text{ vs. } H_a) \) in Li et al. is similar to the \( J_n \) statistic discussed earlier which follows a standard normal distribution under \( H_a \). One finding from smooth coefficients for the growth rate of GDP per capita was that for developing countries at a higher stage of economic development, the growth rate of GDP per capita accelerates the deforestation process and deteriorates environmental quality. The results from a \( J_n \) test supported the parametric model to the semiparametric one at the 5% significance level; in fact, Equation 1 with a quadratic polynomial in GDP was preferred to all other models. Heterogeneity due to the economic development process, however, could not be ascertained with these data and the authors suggested that further work is needed on this research question.

**Discussion**

The emphasis of this survey paper was on recent developments in semiparametric econometric methods and their application to the study of the pollution-economic growth trade off, commonly referred to as the environmental Kuznets curve. The applied econometric essays on the EKC are vast and have used techniques located close to the econometric methods production frontier with standard panel data models. The list of papers reviewed included the standard heterogeneous panel data model which is the typical general structure used as the null model in semiparametric model specification evaluations. Variations of this parametric structure include the standard partial linear regression (PLR) of Robinson (1988) and extensions thereof, including a PLR with
heterogeneity, serial correlation, and heteroskedasticity; the poolability of panel data models has been assessed in parametric and nonparametric settings. The direct and nonneutral effects of economic growth were also studied in a more generalized version of the PLR with smooth coefficients.

Recent developments in econometrics have contributed to a better understanding of economic relationships using panel data models. Absent from the existing EKC literature reviewed above are applications that evaluate functional form of Bayesian models. In some case, a researcher may have parametric knowledge that could be used as prior information on parametric models. In the context of the EKC, prior information could be built around diffuse, independent, priors on the parametric function and partially informative priors on the nonparametric function (e.g., Koop and Poirier, 2004; Huang and Lin, 2007). Another natural extension of future EKC research would relate to the estimation of semiparametric models that contain continuous and discrete regressors. The nonparametric cross-validation technique introduced by Hsiao et al. (2007) is applicable to the case where the EKC contains dummy variables; one point of appeal of this estimator is that its superior performance carries over to model specification tests (see also Racine and Li, 2004) such as the $J_n$ test in equation (29).

Research from standard EKC parametric panel data models previously published typically start by applying a Hausman for fixed versus random effects. Subsequently, the best parametric structure is set up as the null model and a semiparametric model as the alternative (as in equations (1 and 2), respectively). The PLR smooth coefficient model of Li et al. (and other recent applications such as Henderson and Ullah, 2005; Lin and Carroll, 2006; and Henderson et al., 2008) has been revisited by Sun and Carroll (2008),
with the random effects and fixed effects as the null and alternative hypotheses, respectively (recall that in Li et al. the null hypothesis is a parametric model smooth coefficient model whereas the alternative is a semiparametric smooth coefficient model). They propose an estimator that is consistent when there is an additive intercept term (case in which the conventional first difference model fails to generate a consistent estimator). They show the inconsistency of random effects estimators if the true model is one with fixed effects and that fixed effects estimators are consistent under both random and fixed effects panel data models. It is concluded that estimation of a random effects model is appropriate only when the individual effect is independent of the regressors. They also introduce a $J_n$ statistic (similar to equation (29)) for the above hypotheses that, under asymptotic normality of the proposed estimator, converges to a standard normal distribution. The test is one sided and rejects the random effects model for large values at some significance level. Sun and Carroll provide Monte Carlo evidence that supports the satisfactory finite sample performance of the estimator and test statistic and suggest bootstrapping critical values for future research. Given that the question of random effects often plays out in EKC applications (and often rejected), the estimator and statistic introduced in Sun and Carroll should shed brighter light on heterogeneity properties of EKC panels with semiparametric varying coefficient models.

One of the most promising econometric advances, and an area that is still emerging, is the estimation of nonstationary semiparametric panel data models. There is considerable empirical evidence on the existence of unit roots in per capita pollutants and income variables (e.g., Romero-Avila, 2008 inter alia; Liu et al. 2006). This evidence points to the adequacy of vector autoregression and error correction models (ECM) for
some nonstationary panels, and mixed results for others. The failure of many of these previous studies in finding an inverted U shaped EKC in nonstationary panel data consistent with the data generation process led Romero-Avila to design a study that jointly controlled for structural breaks and cross-sectional dependence; the main finding was one of mixed unit roots for the emissions and income relationship of the EKC, putting to question findings that support ECM in world or specific country groups. There seems to be much room to improving existing parametric methods with panel data in a way that are functionally more flexible and consistent with a variety of data generation processes. Perhaps the most challenging finding is the mixed unit roots in panels and the ensuing interpretation of estimated parameters. Baltagi and Kao (2000) is a comprehensive survey of the literature on panel unit roots, cointegration, dynamic panels and heterogeneity in a parametric setting (see also the March 2007 Special Issue of the Journal of Applied Econometrics). Ullah and Roy (1998) provide a fairly complete analysis of the developments in nonparametric analysis with panel data. Extensions of such work to semiparametric nonstationary panels should enhance the empirical understanding of the tradeoff between pollution and growth in environmental economics and the practice of semiparametric econometrics in general.

References


Frazer, G. 2006. Inequality and development across and within countries. World Development, 34, 1459-1481.


