A PARTIALLY PARAMETRIC MODEL

DANIEL J. HENDERSON AND CHRISTOPHER F. PARMETER

Abstract. In this paper we propose a model which includes both a known (potentially) nonlinear parametric component and an unknown nonparametric component. This approach is feasible given that we estimate the finite sample parameter vector and the bandwidths simultaneously. We show that our objective function is asymptotically equivalent to the individual objective criteria for the parametric parameter vector and the nonparametric function. In the special case where the parametric component is linear in parameters, our single-step method is asymptotically equivalent to the two-step partially linear model estimator in Robinson (1988). Monte Carlo simulations support the asymptotic developments and show impressive finite sample performance. We apply our method to the case of a partially constant elasticity of substitution production function for an unbalanced sample of 134 countries from 1955-2011 and find that the parametric parameters are relatively stable for different nonparametric control variables in the full sample. However, we find substantial parameter heterogeneity between developed and developing countries which results in important differences when estimating the elasticity of substitution.

1. Introduction

A notable development in applied economic research over the last twenty years is the use of the partially linear regression model (Robinson 1988) to study a variety of phenomena. The enthusiasm for the partially linear model (PLM) is not confined to a specific application domain. A few examples include, Athey & Levin (2001), who study the role of private information in timber auctions; Banerjee & Duflo (2003), who look at the nonlinear relationship between income inequality and economic growth; Blundell & Windmeijer (2000), who identify the determinants of demand for health services utilizing differences in average waiting times; Carneiro, Heckman & Vytlacil (2011), who estimate marginal returns to education; Finan, Sadoulet & de Janvry (2005), who show that access to even small plots of land in rural

Date: June 7, 2016.

Daniel J. Henderson, Department of Economics, Finance and Legal Studies, University of Alabama; e-mail: djhender@cba.ua.edu. Christopher F. Parmeter, Department of Economics, University of Miami; e-mail: cparmeter@bus.miami.edu.

*All errors are ours alone. All R code in this paper is available upon request.

Mexico can raise household welfare significantly; Gorton & Rosen (1995), who study U.S. bank failures in the 1980s; Lyssiotou, Pashardes & Stengos (2002), who study heterogeneous age effects on consumer demand; Millimet, List & Stengos (2003), who reject a parametric environmental Kuznets curve specification and Yatchew & No (2001), who study gasoline demand in Canada.

Formally, the PLM specifies the conditional mean of y as two separate components, one which is parametric and another which is nonparametric. The model is given as

$$y_i = \boldsymbol{x}_i\beta + m(\boldsymbol{z}_i) + u_i, \quad i = 1, 2, \ldots, n \tag{1}$$

where $\boldsymbol{x}$ is a $1 \times p$ vector of regressors (which does not include a column of ones) and β is a $p \times 1$ vector of parameters. The $1 \times q$ vector of regressors $\boldsymbol{z}$ enter solely through the unknown smooth function $m(\cdot)$. u_i is a conditional mean (on $\boldsymbol{z}$ and $\boldsymbol{x}$) 0 random variable capturing noise. Robinson (1988) outlined a two-step method, and demonstrated that estimation of the finite dimensional parameter vector β at the parametric rate ($\sqrt{n}$) is attainable while the nonparametric component (the infinite dimensional parameter) is estimated at the standard nonparametric rate. The ability to allow a low dimensional number of covariates to enter the model in an unspecified manner is appealing from a practical matter, and the ability to recover β with no loss in rate makes the partially linear model an attractive tool.

Estimation of β and $m(\cdot)$ requires a degree of smoothing. A simple approach is to deploy kernel smoothing to construct unknown conditional means of y and each element of $\boldsymbol{x}$ on $\boldsymbol{z}$ which is the method championed by Robinson (1988). However, the Robinson (1988) approach identifies the parameter vector β via conditionally demeaning on $\boldsymbol{z}$ and then performing ordinary least squares using these demeaned objects. Therefore, most nonlinear (in parameter) parametric functional forms will be infeasible with the standard two-stage approach.

Nonlinear (in parameter) functional forms are also prominent in applied economic research and include Duffy, Papageorgiou & Perez-Sebastian (2004), who estimate a two-level constant elasticity of substitution production function to examine capital-skill complementarity; Haley (1976), who estimates earnings profiles via nonlinear least-squares; Karabarbounis & Neiman (2014), who consider nonlinear estimation methods to examine the global decline in labor share; Laffont, Ossard & Vuong (1995), who use simulated nonlinear least-squares to estimate first price auction models; Polachek, Das & Thamma-Apiroam (2016), who estimate a highly nonlinear function in order to look at heterogeneity in ability measures; Santos Silva & Tenreyro (2006) who advocate estimating gravity models of international trade in levels and Sauer (1988), who estimates models that are nonlinear in parameters to determine the returns to quality and co-authorship in academia. However, when researchers gravitate towards semiparametric models, they implicitly assume that the parametric component is linear in parameters. This may violate economic theory (and potentially lead to biased and inconsistent estimators); here we hope to remedy this problem.

If control variables were to be included in a nonlinear, parametric modeling framework, the two-stage conditioning approach of Robinson (1988) would no longer apply given that conditional expectations do not commute through nonlinear operations. In these instances, an alternative approach is needed to account for the dual, parametric and nonparametric structure. Motivated by the cross-validation function for the single-index model in Härdle, Hall & Ichimura (1993), we propose a method which estimates simultaneously a parameter vector from a possibly nonlinear parametric function and a bandwidth vector for a nonparametric component. Härdle, Liang & Gao (2000, pp. 26) develop a similar cross-validation function for a partially linear model, but do not study the properties of the estimators from this optimization, see also, Gao (1998) and Gao & Anh (1999). Simultaneous estimation allows the user to dispense with the standard conditional demeaning (and the requisite bandwidth selection) and can seamlessly handle parametric functional forms that are nonlinear in parameters. In the case where the parametric component is linear in parameters, the single-step approach is asymptotically equivalent to Robinson's (1988) two-step estimator.

Monte Carlo simulations reveal impressive finite sample performance of our single-step estimator. First, in the case of a nonlinear parametric function, the simulations show that the estimates of the both the parametric and nonparametric functions are consistent. Second, in the case of a linear in parameters parametric function, we find many cases where our model performs better than the standard two-stage approach. If we judge by the current usage of the PLM by applied economists, our simulations suggest that our one-step procedure may have widespread appeal.

We apply our estimator to an unbalanced sample of 134 countries over the years 1955-2011 to estimate a partially constant elasticity of substitution production (CES) function with nonparametric control variables. Specifically, we take the CES production function with both physical and human capital as inputs as given (as argued by Masanjala & Papageorgiou (2004) and others) and allow for controls in a nonparametric fashion (a variety of continuous variables along with country and time effects). We find, for the full sample, that the parametric estimates are relatively stable and there is some evidence of an elasticity of substitution greater than unity (Duffy & Papageorgiou 2000). However, as in Duffy & Papageorgiou (2000), we find substantial heterogeneity between developed and lesser developed countries. When splitting the sample, we find strong evidence for an elasticity of substitution parameter near 1 in each sub-sample.

The rest of this paper proceeds as follows: Section 2 outlines the Robinson (1988) method, while Section 3 presents the proposed estimator, discusses the asymptotic behavior of our objective criterion, and establishes asymptotic equivalence to the Robinson (1988) estimator in the case of a linear in parameters model. The simulation results are provided in Section 4 and our empirical application in given in Section 5. Section 6 concludes.

2. The Partially Linear Model

In this section we outline the Robinson (1988) estimator. We do so both to familiarize the reader with the approach as well as help us make direct comparisons of this estimator (both theoretically and in simulations) from ours in the case where the parametric component is linear in parameters.

2.1. Estimation of the Parametric Component.

Robinson (1988) estimates β via a two-step process by first isolating the parametric component. Once this is achieved, OLS can be used to estimate the parameter vector β. Mechanically, take the conditional expectation of each side of Equation (1) with respect to $\boldsymbol{z}$. This leads to

$$\begin{aligned} E(y_i|\boldsymbol{z}_i) &= E\left(\boldsymbol{x}_i\beta + m(\boldsymbol{z}_i) + u_i|\boldsymbol{z}_i\right) \\ &= E(\boldsymbol{x}_i\beta|\boldsymbol{z}_i) + E\left(m(\boldsymbol{z}_i)|\boldsymbol{z}_i\right) + E(u_i|\boldsymbol{z}_i) \\ &= E(\boldsymbol{x}_i|\boldsymbol{z}_i)\beta + m(\boldsymbol{z}_i) \end{aligned}$$

which is subtracted from Equation (1) to obtain

$$\begin{aligned} y_i - E(y_i|\boldsymbol{z}_i) &= \boldsymbol{x}_i\beta + m(\boldsymbol{z}_i) + u_i - [E(\boldsymbol{x}_i|\boldsymbol{z}_i)\beta + m(\boldsymbol{z}_i)] \\ &= [\boldsymbol{x}_i - E(\boldsymbol{x}_i|\boldsymbol{z}_i)]\beta + u_i. \end{aligned} \tag{2}$$

If $E(y_i|\boldsymbol{z}_i)$ and $E(\boldsymbol{x}_i|\boldsymbol{z}_i)$ are known, OLS can be used to estimate β.

In practice, these conditional means are unknown and must be estimated. Robinson (1988) suggests use of local-constant least-squares (LCLS) to estimate each conditional mean separately. As shown in Henderson & Parmeter (2015), these conditional expectations are constructed as

$$\widehat{E}(y_i|\boldsymbol{z}_i) = \frac{\sum_{i=1}^n K_{h_z}(\boldsymbol{z}_i, \boldsymbol{z})y_i}{\sum_{i=1}^n K_{h_z}(\boldsymbol{z}_i, \boldsymbol{z})}; \quad \widehat{E}(x_{li}|\boldsymbol{z}_i) = \frac{\sum_{i=1}^n K_{h_z}(\boldsymbol{z}_i, \boldsymbol{z})x_{li}}{\sum_{i=1}^n K_{h_z}(\boldsymbol{z}_i, \boldsymbol{z})},$$

for $l = 1, \ldots, p$, where

$$K_{h_z}(\boldsymbol{z}_i, \boldsymbol{z}) = \prod_{d=1}^{q} k\left(\frac{z_{id} - z_d}{h_{zd}}\right)$$

is the product kernel function and the bandwidth vector h_z uses the subscript term to note that the bandwidths are for the regressors in $\boldsymbol{z}$. The q bandwidths for the conditional expectation of a random variable w given $\boldsymbol{z}$ can be calculated via the cross-validation function

$$CV(h_{wz}) = \sum_{i=1}^{n} [w_i - \widehat{E}_{-i}(w_i|\boldsymbol{z}_i)]^2,$$

where $\widehat{E}_{-i}(w_i|\boldsymbol{z}_i) = \frac{\sum_{j=1, j\neq i}^{n} K_{h_{wz}}(\boldsymbol{z}_j, \boldsymbol{z}_i) w_j}{\sum_{j=1, j\neq i}^{n} K_{h_{wz}}(\boldsymbol{z}_j, \boldsymbol{z}_i)}$ is the leave-one-out estimator of $E(w_i|\boldsymbol{z}_i)$, with $w = y$ or x_l, with x_l being the l^{th} element of $\boldsymbol{x}$, in our setting. We use h_{wz} to signify that the bandwidths used to smooth the different variables can differ.

Once these conditional means have been estimated, they are plugged into Equation (2) to obtain

$$y_i - \widehat{E}(y_i|\boldsymbol{z}_i) = [\boldsymbol{x}_i - \widehat{E}(\boldsymbol{x}_i|\boldsymbol{z}_i)]\beta + u_i.$$

The estimator of the parameter vector β is obtained by OLS regression of $y_i - \widehat{E}(y_i|\boldsymbol{z}_i)$ on $\boldsymbol{x}_i - \widehat{E}(\boldsymbol{x}_i|\boldsymbol{z}_i)$ as

$$\widehat{\beta}_2 = \left\{\sum_{i=1}^{n} \left[\boldsymbol{x}_i - \widehat{E}(\boldsymbol{x}_i|\boldsymbol{z}_i)\right]' [\boldsymbol{x}_i - \widehat{E}(\boldsymbol{x}_i|\boldsymbol{z}_i)]\right\}^{-1} \sum_{i=1}^{n} \left[\boldsymbol{x}_i - \widehat{E}(\boldsymbol{x}_i|\boldsymbol{z}_i)\right]' \left[y_i - \widehat{E}(y_i|\boldsymbol{z}_i)\right], \tag{3}$$

where we use the subscript 2 to denote estimators stemming from Robinson's (1988) two-step method.

2.2. **Estimation of the Nonparametric Component.** The most common approach to estimating the unknown function $m(\cdot)$ is to replace β with $\widehat{\beta}_2$ in Equation (1). Doing this yields

$$y_i = \boldsymbol{x}_i\widehat{\beta}_2 + m(\boldsymbol{z}_i) + u_i,$$

and since $\boldsymbol{x}_i\widehat{\beta}_2$ is observed, it can be subtracted from each side,

$$y_i - \boldsymbol{x}_i\widehat{\beta}_2 = m(\boldsymbol{z}_i) + u_i.$$

Next, nonparametrically regress $y_i - \boldsymbol{x}_i\widehat{\beta}_2$ on $\boldsymbol{z}_i$ to obtain the estimator of the unknown function. For example, LCLS leads to the estimator of the conditional mean as

$$\widehat{m}_2(\boldsymbol{z}) = \frac{\sum_{i=1}^n K_{h_z}(\boldsymbol{z}_i, \boldsymbol{z})(y_i - \boldsymbol{x}_i\widehat{\beta}_2)}{\sum_{i=1}^n K_{h_z}(\boldsymbol{z}_i, \boldsymbol{z})}.$$

An appropriate bandwidth vector for estimation of $m_2(\cdot)$ can be selected via the cross-validation function

$$CV(h_z) = \sum_{i=1}^n [y_i - \boldsymbol{x}_i\widehat{\beta}_2 - \widehat{E}_{-i}(y_i - \boldsymbol{x}_i\widehat{\beta}_2|\boldsymbol{z}_i)]^2.$$

Given that nonparametric estimators converge at much slower rates than parametric estimators and since Robinson's (1988) estimator of β is $\sqrt{n}$-consistent, we can treat it as if it were known when determining the rate of convergence of $\widehat{m}(\boldsymbol{z})$. Thus, what we are left with (asymptotically) is a nonparametric regression of a 'known' value on $\boldsymbol{z}$.

3. The Partially Nonlinear Model

A key feature of Robinson's (1988) estimation approach is that β can be identified through conditional (on $\boldsymbol{z}$) demeaning $\boldsymbol{x}$ and y. However, this framework succeeds only when the parametric component of the conditional mean is linear in parameters. Alternatively, if we instead allowed for arbitrary nonlinearities, such as

$$y_i = g(\boldsymbol{x}_i; \beta) + m(\boldsymbol{z}_i) + u_i, \tag{4}$$

then no conditional demeaning approach is available as $E[y_i|\boldsymbol{z}_i] = E[g(\boldsymbol{x}_i;\beta)|\boldsymbol{z}_i] + m(\boldsymbol{z}_i)$ does not produce a simple estimand upon subtraction from (4), which produces

$$y_i - E[y_i|\boldsymbol{z}_i] = g(\boldsymbol{x}_i;\beta) - E[g(\boldsymbol{x}_i;\beta)|\boldsymbol{z}_i] + u_i.$$

The conditional expectation of $g(\boldsymbol{x}_i;\beta)$ naturally does not readily suggest an estimator.

To overcome this problem, we consider selecting $\widehat{\beta}$ and the bandwidth vector h (for the unknown function) at the same time. Formally, the objective function to estimate β and h simultaneously is:

$$\min_{h,\widehat{\beta}} \frac{1}{n}\sum_{i=1}^{n}\left[y_i - g(\boldsymbol{x}_i;\widehat{\beta}) - \widehat{m}_{-i}(\boldsymbol{z}_i)\right]^2, \tag{5}$$

where

$$\widehat{m}_{-i}(\boldsymbol{z}_i) = \frac{\sum_{j=1,j\neq i}^{n} K_h(\boldsymbol{z}_j,\boldsymbol{z}_i)(y_j - g(\boldsymbol{x}_i;\widehat{\beta}))}{\sum_{j=1,j\neq i}^{n} K_h(\boldsymbol{z}_j,\boldsymbol{z}_i)}$$

is the leave-one-out estimator of $m\left(\boldsymbol{z}_i\right)$.

The minimization routine should choose a value $\widehat{\beta}$ which is close to β and bandwidth parameters, h, which are close to their optimal values. We hypothesize that minimizing Equation (5) over both parameter vectors, simultaneously, achieves this goal. We will refer to the (single-step) estimators of β and $m(\boldsymbol{z})$ from Equation (5) as $\widehat{\beta}_1$ and $\widehat{m}_1(\boldsymbol{z})$, respectively.

3.1. **Decomposition of the Objective Function.**

We conjecture that, as in Härdle et al. (1993), our objective function can be shown to be equivalent to the individual objective criteria for β and $m(\boldsymbol{z})$. That is,

$$
\begin{aligned}
\widehat{S}(\beta, h) &= \sum_{i=1}^{n}\left[y_i - g(\boldsymbol{x}_i;\widehat{\beta}) - \widehat{m}_{-i}(\boldsymbol{z}_i)\right]^2 \\
&= \sum_{i=1}^{n}\left\{y_i - g(\boldsymbol{x}_i;\widehat{\beta}) - m(\boldsymbol{z}_i) - [\widehat{m}_{-i}(\boldsymbol{z}_i) - m(\boldsymbol{z}_i)]\right\}^2 \\
&= \sum_{i=1}^{n}\left[y_i - g(\boldsymbol{x}_i;\widehat{\beta}) - m(\boldsymbol{z}_i)\right]^2 + \sum_{i=1}^{n}[\widehat{m}_{-i}(\boldsymbol{z}_i) - m(\boldsymbol{z}_i)]^2 \\
&\quad -2\sum_{i=1}^{n}\left[y_i - g(\boldsymbol{x}_i;\widehat{\beta}) - m(\boldsymbol{z}_i)\right][\widehat{m}_{-i}(\boldsymbol{z}_i) - m(\boldsymbol{z}_i)] \\
&= \widetilde{S}(\widehat{\beta}) + T(h) + \text{remainder term}, \qquad (6)
\end{aligned}
$$

where the first term ($\widetilde{S}$) is the nonlinear least-squares problem for an unknown parametric vector (with known nonparametric function) and the second term ($T(h)$) is the cross-validation function for an unknown nonparametric function. Minimizing $\widehat{S}(\beta, h)$ is quite similar to minimizing $\widetilde{S}(\widehat{\beta})$ with respect to β and $T(h)$ with respect to h. Further, from Härdle et al. (1993), this will produce a $\sqrt{n}$-consistent estimator of β and an asymptotically optimal estimator of h.

The final term, which we label as the remainder term, can be decomposed as

$$
\begin{aligned}
\sum_{i=1}^{n}\left[y_i - g(\boldsymbol{x}_i;\widehat{\beta}) - m(\boldsymbol{z}_i)\right][\widehat{m}_{-i}(\boldsymbol{z}_i) - m(\boldsymbol{z}_i)] =& \sum_{i=1}^{n}\left\{y_i - g(\boldsymbol{x}_i;\beta_0) - m(\boldsymbol{z}_i) - \left[g(\boldsymbol{x}_i;\widehat{\beta}) - g(\boldsymbol{x}_i;\beta_0)\right]\right\} \\
& [\widehat{m}_{-i}(\boldsymbol{z}_i) - m(\boldsymbol{z}_i)] \\
\approx& \sum_{i=1}^{n} u_i\,[\widehat{m}_{-i}(\boldsymbol{z}_i) - m(\boldsymbol{z}_i)] \\
& - \sum_{i=1}^{n}\left[\frac{\partial g(\boldsymbol{x}_i;\beta)}{\partial\beta}(\widehat{\beta} - \beta_0)\right][\widehat{m}_{-i}(\boldsymbol{z}_i) - m(\boldsymbol{z}_i)] \\
=& R_2(h) + R_1(\beta, h) \qquad (7)
\end{aligned}
$$

and will be shown to be asymptotically negligible.

3.2. **Main Result.** We will maintain the following regularity conditions. Assume that $\mathcal{A} \subseteq \mathbb{R}$ is a convex set and denotes the support of $\boldsymbol{z}$. We have the following conditions

Assumption 3.1. *The density of $\boldsymbol{z}$, $f(\boldsymbol{z})$ is bounded away from 0 on $\mathcal{A}$ and possesses two bounded derivatives.*

Assumption 3.2. *i) $m(\boldsymbol{z})$ has two bounded, continuous derivatives. ii) $g(\boldsymbol{x}, \beta)$ has at least one continuous and bounded derivative with respect to β.*

Assumption 3.3. *The kernel $k(\cdot)$ is a symmetric probability density function supported on the interval* $(-1, 1)$ *with a bounded derivative.*

Assumption 3.4. $E(u_i|\boldsymbol{z}_i, \boldsymbol{x}_i) = 0$, $E(u_i^2|\boldsymbol{z}_i, \boldsymbol{x}_i) = \sigma^2(\boldsymbol{z}_i, \boldsymbol{x}_i)$ *where $\sigma^2(\cdot)$ is continuous and bounded everywhere and* $\sup_i E|u_i|^c < \infty\ \forall c$.

The requirement of two derivatives for both $f(\cdot)$ and $m(\cdot)$ in Assumptions 3.1 and 3.2 are due to the second order kernel we are smoothing the data with. Further, the restriction that $f(\boldsymbol{z})$ be bounded away from 0 on $\mathcal{A}$ ensures that with probability 1, the denominators in the definition of our estimator for $\widehat{m}(\boldsymbol{z})$ are bounded away from 0. We could remove the requirement that the kernel is bounded, but this only serves to add complications to the requisite theory. Lastly, Assumption 3.4 imposes the condition that all of the moments of the u's are bounded. While this is restrictive, it can be relaxed to only a finite number of moments. However, as our derivations follow from Härdle et al. (1993), their method of proof does not provide an estimate of a minimum moment condition on ε and so augmenting their theory is not pursued here. Lastly, let $||\cdot||$ denote the Euclidean norm.

Theorem 3.1. *Under assumptions 3.1 to 3.4, for* $\mathcal{B}_n = \left\{\beta \in \mathcal{B} : ||\beta - \beta_0|| \leq Cn^{-1/2}\right\}$ *and* $\mathcal{H}_n = \left\{h : C_1 n^{-1/5} \leq h \leq C_2 n^{-1/5}\right\}$, *we have*

$$\widehat{S}(\beta, h) = \widetilde{S}(\widehat{\beta}) + T(h) + R_1(\beta, h) + R_2(h),$$

where $\sup_{\beta\in\mathcal{B}_n,h\in\mathcal{H}_n}|R_1(\beta,h)| = o_p(n^{1/5})$ *and* $\sup_{h\in\mathcal{H}_n}|R_2(h)| = o_p(1)$.

This provides a complete description of our objective function for the partially parametric model. Further, it can be shown that with probability 1, as $n\to\infty$, the minimum of $\widehat{S}(\beta,h)$, within a ball of radius $n^{-1/2}$ of β_0 for $\boldsymbol{x}$ and of $n^{-1/5}$ for $\boldsymbol{z}$, satisfies $\widehat{h}/h_{opt}\to 1$ and $n^{1/2}(\widehat{\beta}-\beta_0)\to N(0,\mathcal{V})$.

3.3. **Comparison to the Partially Linear Model.** We know (given the correct rate on the bandwidths) that the Robinson (1988) estimator of β is semiparametric efficient, for a homoskedastic error term. Here we will show that, when the parametric function is linear in parameters, our single-step estimator is asymptotically equivalent to the Robinson (1988) estimator.

Formally, the objective function we use to estimate β and h simultaneously is given as

$$\min_{h,\widehat{\beta}}\frac{1}{n}\sum_{i=1}^{n}[y_i-\boldsymbol{x}_i\widehat{\beta}-\widehat{m}_{-i}(\boldsymbol{z}_i)]^2, \tag{8}$$

where

$$\widehat{m}_{-i}(\boldsymbol{z}_i)=\frac{\sum_{j=1,j\neq i}^{n}K_h(\boldsymbol{z}_j,\boldsymbol{z}_i)(y_j-\boldsymbol{x}_j\widehat{\beta})}{\sum_{j=1,j\neq i}^{n}K_h(\boldsymbol{z}_j,\boldsymbol{z}_i)}$$

is the leave-one-out estimator of $m\left(\boldsymbol{z}_i\right)$.

Again, the minimization routine should choose a value for $\widehat{\beta}$ which is close to β and bandwidth parameters (h) which are close to their optimal values (see also Härdle et al. (2000, pp. 144)). We will continue to refer to the estimators of β and $m(\boldsymbol{z})$ from Equation (8) as $\widehat{\beta}_1$ and $\widehat{m}_1(\boldsymbol{z})$, respectively.

Given that the Robinson (1988) estimator is semiparametric efficient, the single-stage estimator presented here will not have superior performance in large samples. However, we can establish asymptotic equivalence with the two-stage estimator. For simplicity, we consider the case of a single parametric (x) regressor. We solve (analytically) for the single-step estimator of β. This requires solving for β in the first-order condition of our quadratic objective function:

$$\frac{\partial}{\partial\beta}\frac{1}{n}\sum_{i=1}^{n}[y_i - x_i\beta - m_{-i}(\boldsymbol{z}_i)]^2 = \frac{2}{n}\sum_{i=1}^{n}[y_i - x_i\beta - m_{-i}(\boldsymbol{z}_i)]\left[-x_i - \frac{\partial m_{-i}(\boldsymbol{z}_i)}{\partial\beta}\right],$$

where

$$\frac{\partial m_{-i}(\boldsymbol{z}_i)}{\partial\beta} = -\left(\sum_{\substack{j=1\\ j\neq i}}^{n} x_j K_h(\boldsymbol{z}_j, \boldsymbol{z}_i)\right) \Bigg/ \left(\sum_{\substack{j=1\\ j\neq i}}^{n} K_h(\boldsymbol{z}_j, \boldsymbol{z}_i)\right).$$

Setting the first-order condition equal to zero and solving for β yields

$$\widehat{\beta}_1 = \left\{\sum_{i=1}^{n}\left[x_i - \frac{\sum_{j=1,j\neq i}^{n} x_j K_h(\boldsymbol{z}_j, \boldsymbol{z}_i)}{\sum_{j=1,j\neq i}^{n} K_h(\boldsymbol{z}_j, \boldsymbol{z}_i)}\right]^2\right\}^{-1} \sum_{i=1}^{n}\left[x_i - \frac{\sum_{j=1,j\neq i}^{n} x_j K_h(\boldsymbol{z}_j, \boldsymbol{z}_i)}{\sum_{j=1,j\neq i}^{n} K_h(\boldsymbol{z}_j, \boldsymbol{z}_i)}\right]\left[y_i - \frac{\sum_{j=1,j\neq i}^{n} y_j K_h(\boldsymbol{z}_j, \boldsymbol{z}_i)}{\sum_{j=1,j\neq i}^{n} K_h(\boldsymbol{z}_j, \boldsymbol{z}_i)}\right].$$

This is the two-step estimator of β in Equation (3) except that we have a single bandwidth vector (h) and the estimated conditional expectations of x and y are replaced with leave-one-out versions. Assuming that the rates on the bandwidths are the same across the two methods, these two estimators are asymptotically equivalent.

There are several interesting comparisons that can be drawn here. First, the single-step estimator behaves like Robinson's (1988) two-step estimator with the *same* bandwidth used to estimate all of the conditional means. This may or may not be desirable in practice. For example, it may be of some benefit if the two-step estimator produced a data-driven bandwidth which led to substantial undersmoothing of one of the covariates which enters in parametrically, leading to overfitting and residuals which displayed little variation, making it difficult to identify β. Secondly, if one component of $\boldsymbol{z}$ had no relation to a given component of $\boldsymbol{x}$, then, using a local-constant least-squares estimator, having the ability to have a specific bandwidth be set to ∞ may be useful given that in this case that component of $\boldsymbol{z}$ should not enter into the conditional mean. With a single bandwidth vector, where the focus is on y, this may not be warranted. One can envision small sample settings where either approach would be expected to outperform the other depending upon the relationship between the variables which enter in nonparametrically and those that enter in parametrically. This is a primarily a finite sample tradeoff and one which we study in our simulations.

4. Simulations

In this section we hope to achieve two goals. First, we want to show that our estimators of the parametric and nonparametric functions (as well as the conditional mean) are consistent. Second, even though we established asymptotical equivalence to the Robinson (1988) estimator, we are curious as to the practical gains or losses in finite samples. Here we conduct an array of Monte Carlo simulations to determine the performance of our single-step estimator as well as try to provide a benchmark where possible (linear in parameters setting) relative to the two-stage estimator.

4.1. **Set-up.**

For our purposes, we consider both a nonlinear in parameters model

$$y_i = x_i^{\beta} + m(z_i) + u_i,$$

with $\beta = 2$ and a linear in parameters model

$$y_i = x_i\beta + m(z_i) + u_i,$$

with $\beta = 1$, where in each case x_i and z_i are scalars. We consider samples sizes of $n =$ 100, 200 and 400 over $S =$ 1,000 Monte Carlo simulations. For all smoothing we use a second-order Gaussian kernel (Li 1996). All bandwidths for Robinson's (1988) approach are determined via least-squares cross-validation. We generate x_i and z_i following Martins-Filho & Yao (2012).[1] In this setting we have $x_i = w_{1i}^2 + w_{2i}^2 + \varepsilon_{1i}$ and $z_i = w_{1i} + w_{2i} + \varepsilon_{2i}$, where w_1, w_2, ε_1 and ε_2 are all generated as independent standard normal random variables. u is also generated as a standard normal random variable.

[1] In Section 4.4, we consider the case where the regressors are linearly related.

We consider eight different functional form specifications for $m(\cdot)$ to determine the impact that curvature has on performance. The different functional forms that we deploy are

DGP_1**:** $m(z) = 0.8 + 0.7z$;

DGP_2**:** $m(z) = 2 + 1.8\sin(1.5z)$;

DGP_3**:** $m(z) = 2.75\frac{e^{-3z}}{1+e^{-3z}} - 1$;

DGP_4**:** $m(z) = 0.7z + 1.4e^{-16z^2}$;

DGP_5**:** $m(z) = \sqrt{z+10}$;

DGP_6**:** $m(z) = |z|$;

DGP_7**:** $m(z) = z + sin(z)$;

DGP_8**:** $m(z) = cos(z)$.

To assess performance, we consider three different criteria. For the parametric component, we consider the average squared error of the parameter estimator $\widehat{\beta}$, (ASE)

$$ASE\left(\widehat{\beta}\right) = S^{-1}\sum_{s=1}^{S}\left[\widehat{\beta}_s - \beta\right]^2,$$

where s indexes the S simulations. Second, we consider the ASE of the estimator of the unknown function,

$$ASE\left[\widehat{m}(\cdot)\right] = (nS)^{-1}\sum_{s=1}^{S}\sum_{i=1}^{n}\left[\widehat{m}_s(z_i) - m(z_i)\right]^2,$$

where $ASE\left[\widehat{m}(\cdot)\right]$ is evaluated at the sample points for each simulation. Third, we consider the ASE of the unknown conditional mean of y,

$$ASE\left[\widehat{E}(y|x,z)\right] = (nS)^{-1}\sum_{s=1}^{S}\sum_{i=1}^{n}\left[\widehat{E}_s(y_i|x_i,z_i) - E(y_i|x_i,z_i)\right]^2,$$

where again $\widehat{E}(y|x,z)$ is the estimated conditional mean of y. As with $ASE\left[\widehat{m}(\cdot)\right]$, $ASE\left[\widehat{E}(y|x,z)\right]$ is evaluated at the sample points for each simulation.

4.2. **Nonlinear in Parameters.** The main results are given in Table 1. Here we show the performance of our estimator with respect to several forms of the nonparametric function for different sample sizes. For each of the 8 DGPs, we give the average square error of the parametric component, the nonparametric component and the conditional mean. The results are similar across DGPs and so we will discuss them jointly.

We see that the average square error for each estimator falls with the sample size and that the parametric function is estimated more precisely than the nonparametric function, as expected. We also see that the average square errors of the estimator of the conditional mean fall with the sample size as hypothesized.

We tried other parametric functional forms as well as other methods to generate the regressors and/or errors and found very similar results. These are available upon request.

Table 1. Simulation Performance of the Partially Parametric Model (nonlinear relationship between the regressors), 1000 Simulations. Each column represents the average square error over the simulations for the parametric parameter, the nonparametric function and conditional mean, respectively.

	ASE_β	ASE_m	ASE_y	ASE_β	ASE_m	ASE_y
		DGP_1			DGP_5	
$n = 100$	0.054	0.157	0.132	0.048	0.124	0.103
$n = 200$	0.035	0.093	0.084	0.027	0.077	0.070
$n = 400$	0.037	0.074	0.062	0.024	0.053	0.050
		DGP_2			DGP_6	
$n = 100$	0.068	0.213	0.188	0.060	0.172	0.143
$n = 200$	0.041	0.127	0.119	0.034	0.114	0.097
$n = 400$	0.035	0.089	0.081	0.024	0.072	0.063
		DGP_3			DGP_7	
$n = 100$	0.065	0.180	0.155	0.066	0.174	0.158
$n = 200$	0.037	0.117	0.104	0.040	0.115	0.103
$n = 400$	0.028	0.081	0.073	0.037	0.100	0.082
		DGP_4			DGP_8	
$n = 100$	0.054	0.198	0.184	0.063	0.165	0.138
$n = 200$	0.035	0.134	0.125	0.035	0.093	0.083
$n = 400$	0.028	0.089	0.081	0.017	0.057	0.052

4.3. **Linear in Parameters.** Table 2 presents our results for the eight different DGPs. We present the median relative ASE for each object across the 1,000 simulations for the two estimators. We take the median ASE for each estimator (e.g., $m(\boldsymbol{z})$) and report the ratio of the single-step over the two-step. For the parametric estimates, we report the ASE across the 1,000 simulations for each estimator and compare the ratio. We use the proposed single-step estimator as the benchmark. Reported tabular entries that are greater than one indicate superior performance of our approach relative to the two-step approach of Robinson (1988). Given that we have relative ASE metrics, in the table we refer to them as $RASE$ for each object of interest.

Table 2. Simulation Performance of the Partially Linear Model (nonlinear relationship between regressors), 1000 Simulations. All columns are relative performance of single-step partially linear estimator to the two-step partially linear estimator. Tabular entries greater than 1 indicate superior performance of the single-step partially linear estimator.

	$RASE_\beta$	$RASE_m$	$RASE_y$	$RASE_\beta$	$RASE_m$	$RASE_y$
		DGP_1			DGP_5	
$n = 100$	1.480	1.365	1.313	2.019	1.443	1.293
$n = 200$	1.227	1.280	1.246	1.308	1.478	1.347
$n = 400$	1.155	1.237	1.198	1.297	1.404	1.322
		DGP_2			DGP_6	
$n = 100$	2.120	2.031	1.871	1.773	1.188	1.148
$n = 200$	1.777	2.091	1.999	1.546	1.161	1.099
$n = 400$	1.670	2.068	1.955	1.232	1.158	1.165
		DGP_3			DGP_7	
$n = 100$	1.592	1.413	1.341	1.992	1.601	1.504
$n = 200$	1.256	1.359	1.314	1.466	1.517	1.490
$n = 400$	1.132	1.343	1.327	1.155	1.526	1.466
		DGP_4			DGP_8	
$n = 100$	1.588	1.348	1.320	1.413	2.080	1.875
$n = 200$	1.549	1.446	1.415	1.252	1.695	1.604
$n = 400$	1.056	1.592	1.565	1.174	1.560	1.494

The table shows substantial finite sample gains for each metric for each DGP. Although the relative performance varies across DGPs, we typically find the ratios to be much larger than unity for each sample size. It appears that, at least for these cases, that the finite sample gains do not fade quickly.

At least two features are apparent from Table 2. First, for estimation of the parametric coefficient, the unknown function and the unknown conditional mean, the single-stage estimator always produces a lower median ASE. Second, as expected, $RASE_\beta$ is decreasing as n increases.

4.4. **Linearly Related Regressors.** As noted earlier, we considered a nonlinear relationship between the regressors. It could be argued that this would paint the Robinson (1988) estimator unfavorably given the conditional demeaning. We therefore decided to perform the entire set of simulations where the regressors had a linear relationship (multivariate normal with mean 0, variance 1 and covariance 0.1). We omit the analogous simulations to Table 1 as they do not bring about any substantive changes, but report the analogous simulation results of Table 2 in Table 3.

These results show improved performance for the two-step estimator. Most of the ratios are now closer to 1 and generally appear to decrease with the sample size (especially for the β). We now actually find several cases (DGPs 2 and 4) where the two-step estimator outperforms our estimator with respect to the parametric parameter.

We note that while it is likely feasible to find a scenario whereby the Robinson (1988) estimator outperforms our estimator, both estimators perform well in finite sample simulations and hence either is feasible in practice when the parametric component is linear in parameters. It may be useful to try each estimator in practice to show robustness.

5. Application: A Partially Constant Elasticity of Substitution Production Function

One of the more debated topics in macroeconomics (both theoretical and empirical) has to do with the form of the production function and more specifically, the elasticity of substitution parameter. Many economists have been happy to accept the Cobb-Douglas specification

TABLE 3. Simulation Performance of the Partially Linear Model (linear relationship between the regressors), 1000 Simulations. All columns are relative performance of single-step partially linear estimator to the two-step partially linear estimator. Tabular entries greater than 1 indicate superior performance of the single-step partially linear estimator.

	$RASE_\beta$	$RASE_m$	$RASE_y$	$RASE_\beta$	$RASE_m$	$RASE_y$
		DGP_1			DGP_5	
$n = 100$	1.122	1.112	1.084	1.195	1.267	1.223
$n = 200$	1.086	1.114	1.077	1.079	1.091	1.135
$n = 400$	1.084	1.076	1.069	1.040	1.048	1.044
		DGP_2			DGP_6	
$n = 100$	1.339	1.289	1.275	1.121	1.320	1.242
$n = 200$	1.042	1.353	1.252	1.125	1.339	1.264
$n = 400$	0.756	1.321	1.165	1.081	1.339	1.305
		DGP_3			DGP_7	
$n = 100$	6.0025	1.583	1.477	1.279	1.195	1.183
$n = 200$	1.167	1.626	1.525	1.160	1.139	1.119
$n = 400$	1.151	1.726	1.590	1.126	1.131	1.125
		DGP_4			DGP_8	
$n = 100$	0.832	1.342	1.237	2.733	1.260	1.193
$n = 200$	0.701	1.368	1.247	1.085	1.333	1.234
$n = 400$	0.395	1.256	1.149	1.055	1.335	1.296

which implies an elasticity of substitution parameter of unity as this value of the parameter aligns with Kaldor's (1960) stylized facts regarding factor shares over time. A further benefit of the Cobb-Douglas specification to empirical researchers is that it is linear in parameters when measured in logarithms.

Although Solow (1957) first employed the Cobb-Douglas specification, he noted that there was no reason that this need be *the* specification and that a CES production function is more flexible and still accommodates Kaldor's (1960) stylized facts (Solow 1958). This difference is important theoretically as the elasticity of substitution being greater than, less than, or equal to unity has important consequences for policy and our understanding of long term growth.

Here we will take the CES specification as given (considering the results in Sun, Henderson & Kumbhakar (2011), Duffy & Papageorgiou (2000) and others), but add additional control variables for which we are unsure of how they should enter the model. We therefore look at a partially CES production function with nonparametric controls (both alternative inputs as well as country and time controls).

5.1. **Specification.** Output (Y) of country i in time period t is determined by a production function

$$Y_{it} = F\left(K_{it}, H_{it}L_{it}, Z_{it}, \mu_i, \lambda_t\right),$$

where $F(\cdot)$ is the production function, K is physical capital, HL is human capital augmented labor, Z are other (potential) inputs, and μ and λ represent country and time effects. The arguments put forth by Duffy & Papageorgiou (2000) and others regarding the use of a constant returns-to-scale, CES production function with physical and human capital, and labor input are convincing, but little is known about how other inputs, if any, should be included, as well as how country and time effects should be included in the model.

Given this uncertainty, we take their (constant returns-to-scale) CES production function (in per worker terms) as given, but allow for other inputs as well as country and time effects to enter our conditional mean nonparametrically.[2] This leads to our main equation of interest as

$$y_{it} = A\left[\delta k_{it}^{\rho} + (1-\delta) H_{it}^{\rho}\right]^{1/\rho} + m\left(Z_{it}, i, t\right) + u_{it}, \tag{9}$$

where $y_{it} = Y_{it}/L_{it}$, $k_{it} = K_{it}/L_{it}$, A is our technology parameter assumed to be greater than zero, δ (assumed to lie between zero and one) is the parameter used to help determine shares (only directly interpretable in the case of a Cobb-Douglas, constant returns-to-scale production function), ρ (assumed less than or equal to unity) helps determine the elasticity of substitution ($\sigma = 1/(1-\rho)$), while $m(\cdot)$ is our unknown smooth function with arguments

[2]We felt it prudent to check to see if the constant returns-to-scale assumption was appropriate. Specifically, we estimated each of the models allowing for variable returns-to-scale. While we found some evidence of increasing returns-to-scale, none of the parameter estimates were significantly different from unity and hence we impose constant returns-to-scale for the remainder of the paper.

Z, i and t, which represent additional control variables (e.g., energy or debt), country and time effects,[3] respectively. u_{it} is the usual idiosyncratic error term.

In what follows, we plan to estimate Equation (9) using five different variables for Z (as well as without a Z variable). We will do so for both an unbalanced and balanced panel. Further, we will look for parameter heterogeneity with respect to level of development.

We note here that we use a Gaussian kernel function for our continuous variables and Li and Racine kernels (Li & Racine 2007) for our unordered and ordered variables: country and time, respectively.

5.2. **Data.** Our data primarily come from the Penn World Tables and the World Bank.[4] The full sample has a total of 134 countries (measured yearly) over the period 1955–2011. All variables are averaged over five year intervals to reduce the impact of macroeconomic cycles on our results. The largest number of observations for any regression is 1383 country-time points. The smallest sample is for the developed balanced sample at 299 country-time observations.

Output (CGDPE), physical capital (CK) and labor (EMP) are obtained via the Penn World Tables, Version 8.1 (Feenstra, Inklaar & Timmer 2016). Human capital (HC), also obtained from the Penn World Tables, uses the average years of schooling data from Barro & Lee (2010) along with the Psacharopoulos (1994) returns to education. Specifically, let ϵ_{it} represent the average number of years of education of the adult population in country i at time t and this our allows our human capital augmented labor to be defined as

$$H_{it}L_{it} = h(\epsilon_{it})L_{it} = \exp\left[\phi(\epsilon_{it})\right] L_{it}$$

where ϕ is a piecewise linear function, with a zero intercept and a slope of 0.134 through the fourth year of education, 0.101 for the next four years, and 0.068 for education beyond the

[3]We drop μ and λ in favor of i and t as our nonparametric function does not require them to enter linearly as is most common in the literature.
[4]All of our data is available upon request.

eighth year. Clearly, the rate of return to education (where ϕ is differentiable) is

$$\frac{\partial \ln h(\epsilon_{it})}{\partial \epsilon_{it}} = \phi'(\epsilon_{it})$$

and $h(0) = 1$.

Data from the World Bank are primarily used to fill the additional control variables (Z). Arable land (hectares per person) has been used as an input in production at least as early as Ricardo (1937). Emissions data has been used as an input by Brock (1997) and Brock & Taylor (2010) and we propose to use CO_2 emissions (metric tons per capita). Energy has long been considered an input in production (e.g., Berndt & Wood (1979), Helliwell, Sturm & Salou (1985) and Maddison (1987)) and we use kilogram of oil equivalent per capita. Public debt has long been seen as a deterrence to economic growth (e.g., Diamond (1965) or Ludvigson (1996)) and we employ the measure of external debt stocks (public and publicly guaranteed). Finally, we use a measure of imports (imports of goods and services) which has been argued to be especially important as intermediate inputs in developing countries (e.g., Amiti & Konings (2007) or Hentschel (1992)).

Each of these additional inputs has been argued to impact output and we do not take a stance as to the importance of one over another at this point. Here we simply plan to estimate each model specification and then use nonparametric model averaging methods (Henderson & Parmeter 2016) to determine the importance of each in terms of the prediction of output per worker.

5.3. **Results.** After testing for correct parametric specification, we focus on the robustness of the parametric parameters with respect to different nonparametric controls, sample sizes as well as subsets of the data. Specifically, we look at the total amount of data available to us for each nonparametric control individually, as well as a balanced sample where we have the same country-time observations for each control variable. In the balanced data case, we are able to directly compare across models both via the parameter estimates as well as measures of fit. Further, we use a model averaging approach similar to that in Henderson & Parmeter (2016) as a method of comparison. After looking at the full sample, we follow the lead of Duffy & Papageorgiou (2000) and look for parameter heterogeneity with respect to the level of development. We note here that in order to obtain enough (balanced) observations for the smaller (developed) sample, we dropped the debt variable for those results. Finally, we consider a Kmenta (1967) approximation to obtain a functional form that is linear in parameters in order to compare our approach to that of the single-step method.

5.3.1. *Testing for Correct Parametric Specification.* Before getting to the main results, we test for correct specification of a fully parametric model where the regressor Z as well as the individual effects enter linearly. Formally, we consider the fully parametric specification

$$y_{it} = A\left[\delta k_{it}^{\rho} + (1-\delta) H_{it}^{\rho}\right]^{1/\rho} + Z_{it}\gamma + \mu_i + \lambda_t + u_{it}$$

and estimate it via nonlinear least-squares. We perform the trivial extension of allowing for a semiparametric specification (à la Section 2.2 of Li and Wang, 1998) of the Hsiao, Li and Racine (2007) correct functional form test for mixed discrete and continuous data (as suggested in their conclusion). Using a wild bootstrap (with 999 resamples) version of the test, in each case, we reject the parametric specification with p-values that are zero to at least three decimal places.

5.3.2. *Full Unbalanced Sample.* The first set of partially nonlinear results are given in Table 4. Each column represents a separate regression. The first column is the regression where we do not include a Z variable and the subsequent columns each introduce a separate nonparametric control variable (in addition to the country and time effects controlled for in each column). The upper rows represent the parametric parameter estimates with 399 wild bootstrapped standard errors beneath each estimate. The lower rows give the pseudo R^2 measure (squared correlation between the true and predicted values of output per worker), the minimized value of the objective function in Equation (5) and the number of country-time observations (n).

The parameter estimates are similar to what is typically found in the literature (Duffy & Papageorgiou 2000). We find both cases where the estimate of δ is significantly above zero and cases where δ is not significantly different from zero. In the cases where this parameter is small, we also find smaller values for the technology parameter (A). Although this is consistent with the literature, when we move to the balanced samples, we will see the estimate of δ is at the more economically intuitive level.

The most studied parameter in the literature is σ (derived from ρ). This elasticity of substitution parameter has many implications in macroeconomic theory. In a majority of cases, the estimated elasticity is greater than one and in two cases we find it significantly greater than one (CO_2 and debt). There is mixed evidence here as to whether or not the CES production function is more suitable than the Cobb-Douglas production function.

The measures of fit here are relatively good and we see that the R^2 is improved as we add the nonparametric controls (Z). Further, we see the objective function is smaller for the land and imports controls, but hold off at this point given that we are not looking at the same samples of data.

5.3.3. *Full Balanced Sample.* While the results in Table 4 are interesting. It is difficult to compare across models because we have different samples. In Table 5, we look at a balanced sample. The first thing we notice is that the parameter estimates are far more similar here as compared to Table 4. This should not come as much of a surprise given that we are now using the same observations for each column. We see that the coefficients for the share and technology parameters are now much larger (and similar to the others) for both the land and debt controls.

The most interesting result here is perhaps the elasticity of substitution estimates. Our balanced samples each have estimates of σ which are greater than unity and we see that they are significantly greater than 1 in five of the six cases. Taking these results as given,

TABLE 4. Finite dimensional parameter estimates for each set of nonparametric control variables (full unbalanced panel). The columns represent each continuous Z variable (discrete controls for country and time also included). 399 wild bootstrapped standard errors are below each parameter estimate. R^2 is calculated as the square of the correlation coefficient between the actual and fitted values, whereas "obj" is the minimized value of the objective function in Equation (5).

	no Z	land	CO_2	energy	debt	imports
A	4.5660	0.0107	5.1298	6.1389	5.2419	0.3502
	0.7792	0.3841	1.0979	0.9276	0.3297	0.4229
δ	0.4166	0.0105	0.3923	0.5253	0.2490	0.0101
	0.0448	0.0610	0.0415	0.0385	0.0222	0.2048
ρ	0.0217	0.6455	0.0388	−0.0107	0.1016	0.1852
	0.0179	0.3881	0.0157	0.0140	0.0120	0.2380
σ	1.0222	2.8214	1.0404	0.9894	1.1131	1.2273
	0.0195	0.9587	0.0168	0.0144	0.0149	0.3236
R^2	0.5770	0.6006	0.9414	0.7430	0.5175	0.3967
obj	2.6648	0.7603	1.1121	1.7173	3.0586	0.9901
n	1383	1252	1331	1035	715	1272

we suggest a rejection of the Cobb-Douglas model and the possibility for endogenous growth models (at least the necessary condition).

The goodness-of-fit measures vary between the two tables, but are still generally large. The objective function points to the model which uses imports as a nonparametric control.

A more sophisticated method to compare across methods is given in the final two columns. Here we include both the cross-validated (CV) and least-squares (LS) weights from a nonparametric model averaging estimator (Henderson & Parmeter 2016). In short, the nonparametric model averaging method picks weights from each model such that the out of sample prediction of the left-hand-side variable is best. The type of weighting (as is common in the model averaging literature) does not appear to matter much as weights are the same (for each individual model) to four decimal places. We see that the model with imports as a control has the largest weight, followed by CO_2. The other four models have negative weights. We could have constrained the weights to go between zero and one, but this would simply lead to weights near zero and less predictive power.

TABLE 5. Finite dimensional parameter estimates for each set of nonparametric control variables (full balanced panel). The columns represent each continuous Z variable (discrete controls for country and time also included). 399 wild bootstrapped standard errors are below each parameter estimate. R^2 is calculated as the square of the correlation coefficient between the actual and fitted values, whereas "obj" is the minimized value of the objective function in Equation (5). CV weight and LS weight are the cross-validated and least-squares model averaging weights of each model, respectively.

	no Z	land	CO_2	energy	debt	imports
A	5.1415	6.0421	6.1677	5.9700	6.3465	2.3240
	0.6700	0.5521	0.8881	0.6544	0.3939	1.5608
δ	0.3269	0.2815	0.4538	0.3721	0.2484	0.2973
	0.0328	0.0238	0.0497	0.0228	0.0217	0.0561
ρ	0.0684	0.0904	0.0168	0.0508	0.1008	0.0845
	0.0163	0.0129	0.0200	0.0101	0.0128	0.0276
σ	1.0735	1.0994	1.0171	1.0535	1.1121	1.0923
	0.0195	0.0162	0.0215	0.0114	0.0165	0.0339
R^2	0.6868	0.6536	0.8225	0.5111	0.4356	0.7514
obj	2.4098	3.0588	2.1498	2.1201	4.1176	0.3719
n	523	523	523	523	523	523
CV weight	−0.0524	−0.0120	0.0423	−0.1023	−0.1022	1.1084
LS weight	−0.0524	−0.0120	0.0423	−0.1023	−0.1022	1.1084

5.3.4. *Developed versus Lesser Developed Sample.* While we may believe in the CES production function, we do not necessarily believe in parameter homogeneity. We therefore cut the sample based on the level of development. Here we simply separated the sample via status in the OECD. However, as mentioned earlier, doing so significantly decreased the sample size and hence we were forced to omit one variable (debt) in order to get a reasonable sample size for estimation. The omission of the debt variable model left us with 299 balanced OECD country-time observations and 623 balanced non-OECD country-time observations.

The results for the OECD regressions are given in Table 6 and the non-OECD regressions are given in Table 7. We find that the parameters are similar between the tables. However, now we only find a single case where the elasticity of substitution parameter is significantly greater than unity (OECD no Z column) and now we have a single case where the elasticity of substitution parameter is significantly below unity (non-OECD imports column). We

TABLE 6. Finite dimensional parameter estimates for each set of nonparametric control variables (OECD balanced panel). The columns represent each continuous Z variable (discrete controls for country and time also included). Debt is excluded to increase the sample size. 399 wild bootstrapped standard errors are below each parameter estimate. R^2 is calculated as the square of the correlation coefficient between the actual and fitted values, whereas "obj" is the minimized value of the objective function in Equation (5). CV weight and LS weight are the cross-validated and least-squares model averaging weights of each model, respectively.

	no Z	land	CO_2	energy	imports
A	6.1149	1.1708	5.2134	5.7407	1.0750
	0.4689	0.0784	0.9164	0.6637	0.0818
δ	0.4086	0.4586	0.5162	0.5122	0.5165
	0.0291	0.0364	0.0483	0.0321	0.0360
ρ	0.0318	0.0153	−0.0057	−0.0042	−0.0055
	0.0111	0.0134	0.0176	0.0117	0.0130
σ	1.0328	1.0155	0.9943	0.9958	0.9945
	0.0118	0.0138	0.0179	0.0119	0.0131
R^2	0.5884	0.9823	0.9305	0.8102	0.9946
obj	1.0742	0.0084	0.1501	0.2257	0.0062
n	299	299	299	299	299
CV weight	0.0209	0.1123	−0.0185	−0.0606	0.9468
LS weight	0.0209	0.1123	−0.0185	−0.0606	0.9468

must note that direct comparisons with Table 5 may not be fair because we do not have the same sample observations. That being said, not allowing for parameter heterogeneity may falsely lead us to believe the elasticity of substitution parameter is greater than one.

Other interesting results can be seen with respect to the goodness-of-fit measures. The R^2 estimates are larger for a majority of the nonparametric controls in Tables 6 and 7 as compared to Table 5. With respect to the model averaging weights, we see more cases of positive weights (three being positive in the OECD sample and four in the non-OECD sample). The OECD sample still weights imports highly whereas the non-OECD sample now points to CO_2.

The sum of these results suggests, as widely believed, that parameter heterogeneity may be an issue in cross-country analysis. Further, other inputs appear to be relevant in the

TABLE 7. Finite dimensional parameter estimates for each set of nonparametric control variables (OECD balanced panel). The columns represent each continuous Z variable (discrete controls for country and time also included). Debt is excluded to increase the sample size. 399 wild bootstrapped standard errors are below each parameter estimate. R^2 is calculated as the square of the correlation coefficient between the actual and fitted values, whereas "obj" is the minimized value of the objective function in Equation (5). CV weight and LS weight are the cross-validated and least-squares model averaging weights of each model, respectively.

	no Z	land	CO_2	energy	imports
A	4.1431	1.4632	1.8185	6.3927	1.6360
	1.0750	0.1934	0.8141	0.8956	0.1550
δ	0.4226	0.4468	0.5373	0.4894	0.5910
	0.0486	0.0822	0.0442	0.0292	0.0502
ρ	0.0230	0.0213	−0.0125	0.0031	−0.0356
	0.0196	0.0341	0.0164	0.0108	0.0200
σ	1.0236	1.0217	0.9877	1.0032	0.9656
	0.0215	0.0349	0.0167	0.0109	0.0188
R^2	0.7090	0.7542	0.9806	0.8984	0.8854
obj	1.2495	0.5433	0.3613	1.7948	0.5675
n	623	623	623	623	623
CV weight	0.0509	−0.0386	0.8485	0.1097	0.0338
LS weight	0.0509	−0.0386	0.8485	0.1097	0.0338

prediction of output per worker, but they do not carry the same level of importance, nor does one input impact different groups of countries the same.

5.3.5. *Kmenta Approximation.* For sake of comparison, despite the warnings (regarding the bias of the estimates) of Sun et al. (2011) and Thursby & Lovell (1978), we consider a Kmenta (1967) approximation of the CES production function. Our Kmenta (1967) approximation log-linearizes the CES production function about $\rho = 0$ and results in a production function that is linear in parameters. This allows us to compare the results of our estimator to that of Robinson (1988). Formally, log-linearizing our CES specification about $\rho = 0$ leads to

$$\ln A + \delta \ln k_{it} + (1-\delta) \ln H_{it} - \frac{1}{2}\delta (1-\delta) \rho \left(\ln \frac{k_{it}}{H_{it}} \right)^2 ,$$

whereby we estimate the equation

$$\ln y_{it} = \beta_1 \ln k_{it} + \beta_2 \ln H_{it} + \beta_3 \left(\ln \frac{k_{it}}{H_{it}} \right)^2 + m\left(\ln Z_{it}, i, t\right) + u_{it}$$

where the parameters of interest can be obtained from estimates of the model above.[5]

The results of this experiment gave values for δ that ranged from over 0.3 to less than 0.6 for the single-step method. The results for the elasticity of subsitution were very close to unity except for full balanced sample model with energy as the Z variable, where the estimate was significantly less than unity. The fit of each model was in excess of 0.87. For the Robinson (1988) method, the fit of each model was substantially improved (with many psuedo-R^2 values in excess of 0.99). However, this perceived overfitting leads to estimates for the elasticity of substitution parameter that are often very close to zero, despite the CES specification being log-linearized about $\rho = 0$. The full set of results are available upon request.

6. Conclusion

In this paper we proposed an estimator which can simultaneously estimate both a (potentially) nonlinear in parameter parametric function and an unknown nonparametric function. We were able to accomplish this by minimizing a cross-validation function with respect to the bandwidth parameter and the parametric parameters simultaneously. We further showed that our objective function is asymptotically equivalent to minimizing the individual objective criteria for the parametric parameter vector and the nonparametric function. In the special case where the parametric function is linear in parameters, our estimator is asymptotically equivalent to the partially linear model proposed by Robinson (1988).

Our simulations and empirical examples showed that our method works well in finite samples. The Monte Carlo simulations supported the asymptotic development as well as

[5]Noting that the intercept term cannot be identified without restricting the mean of the nonparametric component to be equal to zero, we can obtain our parameters of interest via $\delta = \beta_1$, $\rho = -\frac{2\beta_3}{\beta_1\beta_2}$, and $\sigma = \frac{1}{(1-\rho)}$.

showed that gains were feasible relative to the two-step estimator in the linear in parameters case.

We employed our estimator to study a partially constant elasticity of substitution production function for 134 countries over the period 1955-2011. We showed that the parametric parameters were relatively stable with respect to nonparametric control variables. When we looked at a balanced sample, we found evidence for an elasticity of substitution greater than unity, but those results went away when we allowed for parameter heterogeneity (primarily finding the elasticity of substitution not significantly different from one).

References

Amiti, M. & Konings, J. (2007), 'Trade liberalization, intermediate inputs, and productivity: Evidence from Indonesia', *American Economics Review* **97**, 1611–1638.

Athey, S. & Levin, J. (2001), 'Information and competition in U.S. forest service timber auctions', *Journal of Political Economy* **109**(2), 375–416.

Banerjee, A. V. & Duflo, E. (2003), 'Inequality and growth: What can the data say?', *Journal of Economic Growth* **8**, 267–299.

Barro, R. & Lee, J.-W. (2010), 'A new data set on educational attainment in the world, 1950-2010', *Journal of Development Economics* **104**, 184–198.

Berndt, E. R. & Wood, D. O. (1979), 'Engineering and econometric interpretations of energy-capital complementarity', *American Economic Review* **69**, 342–354.

Blundell, R. W. & Windmeijer, F. (2000), 'Identifying demand for health resources using waiting times information', *Health Economics* **9**(6), 465–474.

Brock, W. A. (1997), A polluted golden age, *in* V. L. Smith, ed., 'Economics of natural and environmental resources', Gordon and Brech Science Publishers.

Brock, W. A. & Taylor, M. S. (2010), 'The green Solow model', *Journal of Economic Growth* **15**, 127–153.

Carneiro, P., Heckman, J. J. & Vytlacil, E. J. (2011), 'Estimating marginal returns to education', *American Economic Review* **101**, 2754–2781.

Diamond, P. A. (1965), 'National debt in a neoclassical growth model', *American Economic Review* **55**, 1126–1150.

Duffy, J. & Papageorgiou, C. (2000), 'The Solow model with CES technology: nonlinearities and parameter heterogeneity', *Journal of Economic Growth* **5**, 87–120.

Duffy, J., Papageorgiou, C. & Perez-Sebastian, F. (2004), 'Capital-skill complementarity? Evidence from a panel of countries', *Review of Economics and Statistics* **86**, 327–344.

Feenstra, Robert, C., Inklaar, R. & Timmer, M. P. (2016), 'The next generation of the Penn World Table', *American Economic Review* .

Finan, F., Sadoulet, E. & de Janvry, A. (2005), 'Measuring the poverty reduction potential of land in rural Mexico', *Journal of Development Economics* **77**, 27–51.

Gao, J. (1998), 'Semiparametric regression smoothing of non-linear time series', *Scandinavian Journal of Statistics* **25**, 521–539.

Gao, J. & Anh, V. V. (1999), 'Semiparametric regression under long-range dependent errors', *Journal of Statistical Planning and Inference* **80**, 37–57.

Gorton, G. & Rosen, R. (1995), 'Corporate control, portfolio choice and the decline of banking', *Journal of Finance* **50**(5), 1377–1420.

Haley, W. J. (1976), 'Estimation of the earnings profile from optimal human capital accumulation', *Econometrica* **44**, 1223–1238.

Härdle, W., Hall, P. & Ichimura, H. (1993), 'Optimal smoothing in single-index models', *Annals of Statistics* **21**(1), 157–178.

Härdle, W., Liang, H. & Gao, J. (2000), *Partially Linear Models*, Physica-Verlag.

Helliwell, J. F., Sturm, P. H. & Salou, G. (1985), 'International comparison of the sources of productivity slowdown 1973-1982', *European Economic Review* **28**, 157–191.

Henderson, D. J. & Parmeter, C. F. (2015), *Applied Nonparametric Econometrics*, Cambridge University Press.

Henderson, D. J. & Parmeter, C. F. (2016), Model averaging over nonparametric estimators, *in* R. C. Hill, G. Gonzalez-Rivera & T.-H. Lee, eds, 'Advances in Econometrics', Vol. 36, Emerald.

Hentschel, J. (1992), *Imports and growth in highly indebted countries*, Springer-Verlag.

Kaldor, N. (1960), Capital accumulation and economic growth, *in* F. A. Lutz & D. C. Hague, eds, 'The theory of capital', St. Martin's Press.

Karabarbounis, L. & Neiman, B. (2014), 'The global decline of labor share', *Quarterly Journal of Economics* **129**, 61–103.

Kmenta, J. (1967), 'On estimation of the CES production function', *International Economic Review* **8**, 180–189.

Laffont, J.-J., Ossard, H. & Vuong, Q. (1995), 'Econometrics of first-price auctions', *Econometrica* **63**, 953–980.

Li, Q. (1996), 'On the root-n-consistent semiparametric estimation of partially linear models', *Economics Letters* **51**, 277–286.

Li, Q. & Racine, J. S. (2007), *Nonparametric econometrics: Theory and practice*, Princeton.

Ludvigson, S. (1996), 'The macroeconomic effects of government debt in a stochastic growth model', *Journal of Monetary Economics* **38**, 25–45.

Lyssiotou, P., Pashardes, P. & Stengos, T. (2002), 'Age effects on consumer demand: An additive partially linear regression model', *Canadian Journal of Economics* **35**(1), 153–165.

Maddison, A. (1987), 'Growth and slowdown in advanced capitalist economies: Techniques of quantitative assessment', *Journal of Economic Literature* **25**, 649–698.

Martins-Filho, C. & Yao, F. (2012), 'Kernel-based estimation of semiparametric regression in triangular systems', *Economics Letters* **115**, 24–27.

Masanjala, W. & Papageorgiou, C. (2004), 'A cross-country empirical investigation of the aggregate production function specification', *Journal of Applied Econometrics* **19**, 171–201.

Millimet, D. L., List, J. A. & Stengos, T. (2003), 'The environmental Kuznets curve: Real progress or misspecified models?', *Review of Economics and Statistics* **85**(4), 1038–1047.

Polachek, S. W., Das, T. & Thamma-Apiroam, R. (2016), 'Micro and macro implications of heterogeneity in the production of human capital', *Journal of Political Economy* .

Psacharopoulos, G. (1994), 'Returns to investment in education: A global update', *World Development* **22**, 1325–1343.

Ricardo, D. (1937), *The principles of political economy and taxation*, London.

Robinson, P. M. (1988), 'Root-n-consistent semiparametric regression', *Econometrica* **56**, 931–954.

Santos Silva, J. M. C. & Tenreyro, S. (2006), 'The log of gravity', *Review of Economics and Statistics* **88**, 641–658.

Sauer, R. D. (1988), 'Estimates of returns to quality and coauthorship in economic academia', *Journal of Political Economy* **96**, 855–866.

Solow, R. M. (1957), 'Technical change and the aggregate production function', *Review of Economics and Statistics* **39**, 312–320.

Solow, R. M. (1958), 'A skeptical note on the constancy of relative shares', *American Economic Review* **48**, 618–631.

Sun, K., Henderson, D. J. & Kumbhakar, S. K. (2011), 'Biases in estimating log production', *Journal of Applied Econometrics* **26**, 708–714.

Thursby, J. G. & Lovell, C. A. K. (1978), 'An investigation of the Kmenta approximation to the CES function', *International Economic Review* **19**, 363–377.

Yatchew, A. & No, J. A. (2001), 'Household gasoline demand in Canada', *Econometrica* **69**(6), 1697–1709.

Proof of Theorem 3.1

To come ...

University of Alabama, University of Miami