SINGLE-STEP ESTIMATION OF A PARTIALLY LINEAR MODEL

DANIEL J. HENDERSON AND CHRISTOPHER F. PARMETER

ABSTRACT. In this paper we propose an asymptotically equivalent single-step alternative to the two-step partially linear model estimator in Robinson (1988). The estimator not only has the potential to decrease computing time dramatically, it shows substantial finite sample gains in Monte Carlo simulations.

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 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

\[
y_i = x_i' \beta + m(z_i) + u_i, \quad i = 1, 2, \ldots, n
\]

where \( x \) is a \( 1 \times p \) vector of regressors (which does not include a column of ones) and \( \beta \) is a \( p \times 1 \) vector of parameters. The \( 1 \times q \) vector of regressors \( z \) enter solely through the unknown smooth function \( m(\cdot) \). \( u_i \) is a conditional mean (on \( z \) and \( x \)) \( 0 \) random variable capturing noise. Robinson (1988) outlined a two-step method, and demonstrated that estimation of the finite dimensional parameter vector \( \beta \) 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 $\beta$ with no loss in rate makes the partially linear model an attractive tool.

Estimation of $\beta$ 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 $x$ on $z$ which is the method championed by Robinson (1988). However, the performance of the
method may depend significantly on the bandwidths chosen in this first stage. Further, even
having estimated $\beta$, bandwidths are still required for computing an estimator of $m(\cdot)$.

Motivated by the cross-validation function for the single-index model in Härdle, Hall &
Ichimura (1993), we propose a method which estimates $\beta$ and the bandwidth vector used
to estimate the nonparametric component $m(z)$ simultaneously. In our new approach, it
is possible that computation time can diminish drastically as we only need to estimate one
set of bandwidths as opposed to the $p+2$ sets in the standard approach.\footnote{Note that since we optimize over $h$ and $\beta$ simultaneously, it is feasible in a given situation that this could lead to a lengthier search.} We further show
that our approach is asymptotically equivalent to that in Robinson (1988). Monte Carlo
simulations reveal impressive finite sample gains of our single-step estimator. The findings
presented here are important for applied research as little attention has been paid to the
appropriate selection of bandwidths in practice when using the PLM. Simply judging by the
current usage of the PLM by applied economists, our simulations suggest that our one-step
procedure will have widespread appeal.

2. Estimation

2.1. The Two-Step Method.

process by first isolating the parametric component. Once this is achieved, OLS can be used
to estimate the parameter vector $\beta$. Mechanically, take the conditional expectation of each
side of Equation (1) with respect to $z$. This leads to

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

which is subtracted from Equation (1) to obtain

$$y_i - E(y_i|z_i) = x_i\beta + m(z_i) + u_i - [E(x_i\beta|z_i) + m(z_i)] = [x_i - E(x_i|z_i)]\beta + u_i.$$ If $E(y_i|z_i)$ and $E(x_i|z_i)$ are known, OLS can be used to estimate $\beta$.

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

$$\hat{E}(y_i|z_i) = \frac{\sum_{i=1}^{n} K_{h_i}(z_i, z)y_i}{\sum_{i=1}^{n} K_{h_i}(z_i, z)}$$

$$\hat{E}(x_i|z_i) = \frac{\sum_{i=1}^{n} K_{h_i}(z_i, z)x_{ii}}{\sum_{i=1}^{n} K_{h_i}(z_i, z)}$$
for \( l = 1, \ldots, p \), where
\[
K_{h_z}(z_i, z) = \prod_{d=1}^q k \left( \frac{z_{id} - z_d}{h_{zd}} \right)
\]
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 \( z \). The \( q \) bandwidths for the conditional expectation of \( y \) given \( z \) can be calculated via the cross-validation function
\[
CV(h_z) = \sum_{i=1}^n [y_i - \hat{E}_{-i}(y_i|z_i)]^2,
\]
where \( \hat{E}_{-i}(y_i|z_i) = \frac{\sum_{j=1, j \neq i}^n K_{h_z}(z_j, z_i)y_j}{\sum_{j=1, j \neq i}^n K_{h_z}(z_j, z_i)} \) is the leave-one-out estimator of \( E(y_i|z_i) \). Similarly, the \( q \) bandwidths for each of the \( p \) conditional expectations of \( x \) given \( z \) are calculated (separately and uniquely) via
\[
CV(h_z) = \sum_{i=1}^n [x_{i1} - \hat{E}_{-i}(x_{i1}|z_i)]^2
\]
where \( \hat{E}_{-i}(x_{i1}|z_i) = \frac{\sum_{j=1, j \neq i}^n K_{h_z}(z_j, z_i)x_{ij}}{\sum_{j=1, j \neq i}^n K_{h_z}(z_j, z_i)} \) is the leave-one-out estimator of \( E(x_{i1}|z_i) \) for the \( l^{th} \) element of \( x \).

Once these conditional means have been estimated, they are plugged into Equation (2) to obtain
\[
y_i - \hat{E}(y_i|z_i) = [x_i - \hat{E}(x_i|z_i)]\beta + u_i.
\]
The estimator of the parameter vector \( \beta \) is obtained by OLS regression of \( (y_i - \hat{E}(y_i|z_i)) \) on \( (x_i - \hat{E}(x_i|z_i)) \) as
\[
(3) \hat{\beta}_2 = \left\{ \sum_{i=1}^n [x_i - \hat{E}(x_i|z_i)]' [x_i - \hat{E}(x_i|z_i)] \right\}^{-1} \sum_{i=1}^n [x_i - \hat{E}(x_i|z_i)]' [y_i - \hat{E}(y_i|z_i)],
\]
where we use the subscript 2 to denote estimators stemming from the two-step method.

### 2.1.2. Estimation of the Nonparametric Component

The most common approach to estimating the unknown function \( m(\cdot) \) is to replace \( \beta \) with \( \hat{\beta} \) in Equation (1). Doing this yields
\[
y_i = x_i \hat{\beta} + m(z_i) + u_i,
\]
and since \( x_i \hat{\beta} \) is known, it can be subtracted from each side,
\[
y_i - x_i \hat{\beta} = m(z_i) + u_i.
\]
Next, nonparametrically regress \( y_i - x_i \hat{\beta} \) on \( z_i \) to obtain the estimator of the unknown function. For example, LCLS leads to the estimator of the conditional mean as
\[
\hat{m}_z(z) = \frac{\sum_{i=1}^n K_{h_z}(z_i, z)(y_i - x_i \hat{\beta})}{\sum_{i=1}^n K_{h_z}(z_i, z)}.
\]
An appropriate bandwidth vector for the estimation of \( m_2(\cdot) \) can be selected via the cross-validation function

\[
CV(h_z) = \sum_{i=1}^{n} [y_i - x_i \hat{\beta} - \hat{E}_z(y_i - x_i \hat{\beta} | z_i)]^2.
\]

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

2.2. The Single-Step Method. While we know (given the correct rate on the bandwidths) that the Robinson (1988) estimator of \( \beta \) is semiparametric efficient, the cross-validation routines are designed for the best out-of-sample predictions of \( y \) and each element of \( x \), and not necessarily the best estimates of \( \beta \) or \( m(\cdot) \). Here we consider selecting \( \hat{\beta} \) and the bandwidth vector \( h \) (for the unknown function) at the same time.

Formally, the objective function we use to estimate \( \hat{\beta} \) and \( h \) simultaneously is as follows:

\[
\min_{h, \beta} \frac{1}{n} \sum_{i=1}^{n} [y_i - x_i \hat{\beta} - \hat{m}_{-i}(z_i)]^2,
\]

where

\[
\hat{m}_{-i}(z_i) = \frac{\sum_{j=1, j \neq i}^{n} K_h(z_j, z_i)(y_j - x_j \hat{\beta})}{\sum_{j=1, j \neq i}^{n} K_h(z_j, z_i)}
\]

is the leave-one-out estimator of \( m(z_i) \).

The minimization routine should choose a value for \( \hat{\beta} \) which is close to \( \beta \) and bandwidth parameters which are close to their optimal values. We hypothesize that minimizing Equation (4) over both parameter vectors, simultaneously, achieves this goal. We will refer to the estimators of \( \beta \) and \( m(z) \) from Equation (4) as \( \hat{\beta}_1 \) and \( \hat{m}_1(z) \), 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 \( \beta \). This requires solving for \( \beta \) in the first-order condition of our quadratic objective function:

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

where

\[
\frac{\partial m_{-i}(z_i)}{\partial \beta} = -\left( \sum_{j=1}^{n} x_j K_h(z_j, z_i) \right) \left/ \left( \sum_{j=1}^{n} K_h(z_j, z_i) \right) \right.
\]
Setting the first-order condition equal to zero and solving for $\beta$ yields

$$
\hat{\beta}_1 = \left\{ \sum_{i=1}^{n} \left[ x_i - \sum_{j=1, j \neq i}^{n} x_j K_h(z_j, z_i) \right] \right\}^{-1} \sum_{i=1}^{n} \left[ x_i \sum_{j=1, j \neq i}^{n} K_h(z_j, z_i) - y_i \sum_{j=1, j \neq i}^{n} K_h(z_j, z_i) \right].
$$

This is the two-step estimator of $\beta$ in Equation (3) except that we have a single bandwidth vector $(h)$ and the conditional means 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.

### 3. Simulations

Even though we have established the asymptotic equivalence of our single-stage partially linear estimator to the two-stage estimator of Robinson (1988), it remains to be seen if there are any practical gains in finite samples. Here we conduct an array of Monte Carlo simulations to determine the performance of our single-step estimator relative to the benchmark two-stage estimator.

For our purposes, we focus on the simple setting of

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

where $x_i$ and $z_i$ are each scalar and we set $\beta = 1$. 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). In this setting we have $x_i = w_{i1}^2 + w_{i2}^2 + u_{i1}$ and $z_i = w_{i1} + w_{i2} + u_{i1}$, where $w_{i1}, w_{i2}, u_{i1}$ and $u_{i2}$ are all generated as independent standard normal random variables. $\varepsilon$ is generated as a standard normal random variable.

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

- $DGP_1$: $m(z) = 0.8 + 0.7z$;
- $DGP_2$: $m(z) = 2 + 1.8 \sin(1.5z)$;
- $DGP_3$: $m(z) = 2.75 \frac{e^{-3z}}{1+e^{-3z}} - 1$;
- $DGP_4$: $m(z) = 0.7z + 1.4e^{-16z^2}$;
- $DGP_5$: $m(z) = \sqrt{z} + 10$;
- $DGP_6$: $m(z) = |z|$;
- $DGP_7$: $m(z) = z + \sin(z)$;
- $DGP_8$: $m(z) = \cos(z)$.

These DGPs cover an array of shapes of the unknown function and many have been used in Monte Carlo simulations in other settings for nonparametric estimation.

To assess performance, we consider three different criteria. For the parametric component, since both methods produce unbiased estimators of $\beta$, we consider the average squared error
(ASE) of $\hat{\beta}_1$ and $\hat{\beta}_2$

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

for $j = 1$ or $2$, and $s$ indexes the $S$ simulations. Second, we consider the ASE of the unknown function,

$$ASE\left[\hat{m}_j(\cdot)\right] = n^{-1} \sum_{i=1}^{n} [\hat{m}_j(z_i) - m(z_i)]^2.$$

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

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

where again $\hat{E}_j(y|x, z)$ is the conditional mean of $y$ for $j = 1$ or $2$. As with $ASE\left[\hat{m}_j(\cdot)\right]$, $ASE\left[\hat{E}_j(y|x, z)\right]$ is evaluated at the sample points for each simulation.

Table 1 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., $m(z)$) 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.

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. Three features are apparent from Table 1. First, for both estimation of the unknown function and the unknown conditional mean, the single-stage estimator always produces a lower median $ASE_m$ and $ASE_E$. Second, there are finite sample gains from our approach for the parametric component, as $RASE_\beta$ is also always greater than one. Third, as expected, $RASE_\beta$ is decreasing as $n$ increases.

4. Conclusion

This paper has proposed an asymptotically equivalent estimator of the PLM to that proposed by Robinson (1988). This estimator chooses the smoothing parameters along with the parametric component in a single-step, thus circumventing the need to estimate additional sets of bandwidths in the two-step approach. Impressive finite sample gains across a range of shapes of the unknown function appeared. Given the appeal of the PLM in applied economic research, the new estimator should engender confidence in use of this model as it pertains to the selection of bandwidths.
Table 1. Simulation Performance of Partially Linear Model, 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.

<table>
<thead>
<tr>
<th></th>
<th>$\text{RASE}_\beta$</th>
<th>$\text{RASE}_{m}$</th>
<th>$\text{RASE}_E$</th>
<th>$\text{RASE}_\beta$</th>
<th>$\text{RASE}_{m}$</th>
<th>$\text{RASE}_E$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$DGP_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 100$</td>
<td>1.480</td>
<td>1.365</td>
<td>1.313</td>
<td>2.019</td>
<td>1.443</td>
<td>1.293</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>1.227</td>
<td>1.280</td>
<td>1.246</td>
<td>1.308</td>
<td>1.478</td>
<td>1.347</td>
</tr>
<tr>
<td>$n = 400$</td>
<td>1.155</td>
<td>1.237</td>
<td>1.198</td>
<td>1.297</td>
<td>1.404</td>
<td>1.322</td>
</tr>
<tr>
<td>$DGP_2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 100$</td>
<td>2.120</td>
<td>2.031</td>
<td>1.871</td>
<td>1.773</td>
<td>1.188</td>
<td>1.148</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>1.777</td>
<td>2.091</td>
<td>1.999</td>
<td>1.546</td>
<td>1.161</td>
<td>1.099</td>
</tr>
<tr>
<td>$n = 400$</td>
<td>1.670</td>
<td>2.068</td>
<td>1.955</td>
<td>1.232</td>
<td>1.158</td>
<td>1.165</td>
</tr>
<tr>
<td>$DGP_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 100$</td>
<td>1.592</td>
<td>1.413</td>
<td>1.341</td>
<td>1.992</td>
<td>1.601</td>
<td>1.504</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>1.256</td>
<td>1.359</td>
<td>1.314</td>
<td>1.466</td>
<td>1.517</td>
<td>1.490</td>
</tr>
<tr>
<td>$n = 400$</td>
<td>1.132</td>
<td>1.343</td>
<td>1.327</td>
<td>1.155</td>
<td>1.526</td>
<td>1.466</td>
</tr>
<tr>
<td>$DGP_4$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 100$</td>
<td>1.588</td>
<td>1.348</td>
<td>1.320</td>
<td>1.413</td>
<td>2.080</td>
<td>1.875</td>
</tr>
<tr>
<td>$n = 200$</td>
<td>1.549</td>
<td>1.446</td>
<td>1.415</td>
<td>1.252</td>
<td>1.695</td>
<td>1.604</td>
</tr>
<tr>
<td>$n = 400$</td>
<td>1.056</td>
<td>1.592</td>
<td>1.565</td>
<td>1.174</td>
<td>1.560</td>
<td>1.494</td>
</tr>
</tbody>
</table>
References


University of Alabama, University of Miami