Finite Sample Performance of Semiparametric Estimation Methods for Partially Linear Models Under Nonparametric Endogeneity

David Anderson
David.J.Anderson@Colorado.EDU

Follow this and additional works at: http://scholar.colorado.edu/honr_theses
Part of the Econometrics Commons

Recommended Citation
**Finite Sample Performance of Semiparametric Estimation Methods for Partially Linear Models Under Nonparametric Endogeneity**

**David Anderson**  
Department of Economics  
University of Colorado  
Boulder, CO 80309-0256, USA  
email: David.J.Anderson@Colorado.edu  
Voice: + 1 408 348 8725

**Abstract.** In this paper we discuss the derivation, and use a Monte Carlo study to examine the finite sample performances of select estimators put forth in Martins-Filho et al. (2015) for partially linear semiparametric models under nonparametric endogeneity. We find that the selected estimators sufficiently account for the explicit nonparametric endogeneity of the underlying model in finite samples, and conclude that the proposed estimators $\hat{m}(x)$ and $\beta_p$ are more efficient than their counterparts $\tilde{m}(x)$ and $\beta_{wp}$. Our findings support the assertion that the estimators for $m(x)$ are oracle efficient.

**Keywords and phrases.** Semiparametric regression, Nonparametric endogeneity, Monte Carlo

**JEL classifications.** C14, C15

Defended March 31, 2015

*Thesis Advisor*  
Carlos Martins-Filho | Department of Economics

*Defense Committee*  
Martin Boileau | Department of Economics | Honors Council Representative  
Jem Corcoran | Department of Applied Mathematics
1 Introduction

Semiparametric estimation methods combine the precision of parametric methods and the flexibility of non-parametric estimation. Although quite desirable for their minimal assumptions, nonparametric estimation methods suffer from the curse of dimensionality and are not feasible for large dimensional regressors. As a result, it may be desirable to combine parametric and nonparametric assumptions to construct flexible yet easy to estimate models.

The curse of dimensionality in nonparametric regression is most easily illustrated by examining the optimal convergence rates put forth in Stone (1982). He derives the optimal rate of convergence, \( n^{-r} \), of an estimator \( \hat{\theta}(X) \) to the true \( p \)-times differentiable regression function \( \theta(X) \) as \( r = p/(2p + D) \) where \( X \in \mathbb{R}^D \). Clearly the rate of convergence of nonparametric estimation has an inverse relationship with the dimensionality of \( X \), requiring a much larger sample size \( n \) or strong assumptions about the differentiability of the underlying functional form for the same precision obtainable from a lower dimensional regressor.

One restriction we can impose on a fully nonparametric model to solve the curse of dimensionality is additivity, i.e. some object of estimation \( f(X) \) where \( X \in \mathbb{R}^D \) can be written as \( \sum_{d=1}^{D} g_d(X_d) \), where \( X_d \) is the \( d \)-th component of \( X \). As illustrated above, if we can impose this restriction, the convergence rates estimators is much faster in this additive structure than we would have otherwise. The convergence rate of this specific additive structure is \( r = p/(2p + 1) \), and is examined further in Stone (1985).

In the model we examine in this paper, we restrict the object of estimation to be partially linear and additive, in a combination of parametric and nonparametric assumptions mentioned earlier. These restrictions result in the semiparametric model \( Y = Z^\prime \beta + m(X) + \varepsilon \) for some conformable vectors \( Z^\ast, X \) and some scalars \( Y, \varepsilon \). This is the standard semiparametric model first popularized by Robinson (1988), and has been considered under a variety of relaxed assumptions and additional structure.

For the model examined in this paper, we add the an auxiliary equation as in Newey et al. (1999) to account for the explicit nonparametric endogeneity assumption \( E(\varepsilon|X) \neq 0 \). We add the assumption that \( X = g(Z) + U \) where \( Y \in \mathbb{R}, X \in \mathbb{R}^D, Z \in \mathbb{R}^L \) and the previously defined \( Z^\ast \) is a subvector of \( Z, Z^\ast \in \mathbb{R}^{L_1} \).
and $L_1 < L$. Estimators for models with this structure have been examined by Martins-Filho and Yao (2012) who examined the case where both the parametric and nonparametric regressors are endogenous, and Gao and Phillips (2013) who consider the model where the parametric portion is endogenous and the nonparametric portion is strictly non stationary.

In this paper we present the derivation of some selected estimators for $\beta$ and $m(X)$ proposed by Martins-Filho et al. (2015), and use a Monte Carlo study to examine the finite sample performance under a varying number data generating procedure designs.

2 Literature Review

Our interest begins with Robinson (1988) which outlays a $\sqrt{n} -$ consistent estimator for $\beta$ for the model $Y = Z'\beta + m(X) + \varepsilon$ where $E(\varepsilon|X) = 0$. For some observed sample $\{(Y_i, Z_{i1}, \cdots, Z_{iL_i}, X_{i1}, \cdots, X_{iD})\}_{i=1}^n$ of some sample size $n$, denote $Y' = (Y_1 \cdots Y_n)$, $Z = (Z_{i1})_{i=1}^{n,L_i}$ and $X = (X_{i1})_{i=1}^{n,D}$ and let $A_i$ denote the $i$th row of some matrix $A$. His method estimated $\beta$ as

$$\hat{\beta} = \left[ \frac{1}{n} \sum_{i=1}^{n} (Z_i - \hat{Z}_i)'(Z_i - \hat{Z}_i) \right]^{-1} \left[ \frac{1}{n} \sum_{i=1}^{n} (Z_i - \hat{Z}_i)'(Y_i - \hat{Y}_i) \right]$$

where $\hat{Z}_i$ and $\hat{Y}_i$ are Nadaraya-Watson kernel estimators defined as

$$\hat{Z}_i = \left[ \frac{1}{nh_x} \sum_{j=1}^{n} K\left( \frac{X_{ij} - X_{ij}}{h_x} \right) \right]^{-1} \left[ \frac{1}{nh_x} \sum_{j=1}^{n} K\left( \frac{X_{ij} - X_{ij}}{h_x} \right) Z_{ij} \right]$$

$$\hat{Y}_i = \left[ \frac{1}{nh_x} \sum_{j=1}^{n} K\left( \frac{X_{ij} - X_{ij}}{h_x} \right) \right]^{-1} \left[ \frac{1}{nh_x} \sum_{j=1}^{n} K\left( \frac{X_{ij} - X_{ij}}{h_x} \right) Y_{ij} \right]$$

for some choice of kernel $K: \mathbb{R}^D \rightarrow \mathbb{R}$ and bandwidth $h_x$. Robinson assumed that $Y$ is conditionally homoskedastic, i.e. $V(Y|X, Z) = \sigma^2$. The model examined in our paper can be seen as a relaxation of that in Robinson (1988), as the additional assumption of $E(\varepsilon|X) = 0$ onto our model results in the estimator he proposed.

In general, the Nadaraya-Watson kernel estimator estimates $E(Y|X)$ as

$$\hat{m}(x) = \left[ \frac{1}{nh} \sum_{j=1}^{n} K\left( \frac{X_i - x}{h} \right) \right]^{-1} \left[ \frac{1}{nh} \sum_{j=1}^{n} K\left( \frac{X_i - x}{h} \right) Y_i \right]$$
where $K$ is a symmetric kernel function with bandwidth $h$. $K$ is any real valued integrable function such that $\int K(\cdot) = 1$ and $\int |K(\cdot)| < \infty$. The bandwidth $h$ does not need to be a constant and can vary with $x$.

We use constant bandwidths and symmetric nonnegative kernels in the estimators examined in this paper, i.e. $K(-a) = K(a)$ and $K(a) \geq 0 \forall a \in \mathbb{D}$ where $\mathbb{D}$ is the domain of the kernel. This method of estimation is asymptotically unbiased but has finite sample bias. The variance converges to 0 at the rate $n^{\frac{2}{5}}$ when $D = 1$ and $h \propto n^{1/5}$; slower than the estimation method for the parametric portion proposed by Robinson.

Newey et al. (1999) considered a fully nonparametric model $Y = m(X, Z^s) + \varepsilon$ and added a reduced form equation for $X$, that $X = g(Z) + U$ where $X, Z^s, Z$ are defined as in the introduction, and some unobserved error $U$. This structural equation is necessary to account for endogenous regressors and is estimated nonparametrically because structural models do not generally have tight functional forms.

The partially linear triangular system model we are testing in this paper can be seen as a restriction of the fully nonparametric form considered in Newey et al. (1999) for our partially linear structure. Their estimation procedure focuses on the use of series estimators, where as ours is based around kernel estimation and explicitly allows for endogenous variables to enter the model nonparametrically. Our model is also similar to the model examined in Martins-Filho and Yao (2012) as we restrict the regressors to a partially linear additive form, but where ours only has non-parametric endogeneity.

Gao and Phillips (2013) considered a similar model to ours, but under the assumptions that the parametric regressors were endogenous and the nonparametric regressors were exogenous but non stationary. They were able to show that due to the parametric endogeneity in their model, semiparametric least squares estimation for $\beta$ was not consistent, but proposed an instrumental variable method which was, and may even be $\sqrt{n}$, consistent even with nonstationary nonparametric regressors.
3 Model

The model we are examining is

\[ Y = Z' \beta + m(X) + \varepsilon \] (1)

\[ X = g(Z) + U \] (2)

where \( Y \in \mathbb{R}, X \in \mathbb{R}^D, Z \in \mathbb{R}^L \) are observed random variables and \( Z^s \in \mathbb{R}^{L_1} \) is a subvector of \( Z \) where \( L_1 < L \). We define \( g(Z) \equiv (g_1(Z) \cdots g_D(Z)) \) where \( g_d(Z) : \mathbb{R}^L \to \mathbb{R} \) for \( d = 1, \ldots, D, \beta \in \mathbb{R}^{L_1} \), \( m(X) : \mathbb{R}^D \to \mathbb{R} \) are unknown parameters. As per usual \( \varepsilon \) and \( U \) are unobserved errors. We specifically assume that \( X \) is endogenous, i.e. \( E(X'\varepsilon) \neq 0 \), and

\[ E(U|Z) = 0, \ E(\varepsilon|Z,U) = E(\varepsilon|U) \] (3)

We define \( \lambda(U) \equiv E(\varepsilon|U) \) and by the law of iterated expectation, along with (2) and (3),

\[ E(\varepsilon|X,U) = E(E(\varepsilon|Z,X,U)|X,U) \] (4)

\[ = E(E(\varepsilon|Z,U)|X,U) \] (5)

\[ = E(E(\varepsilon|U)|X,U) \] (6)

\[ = \lambda(U) \] (7)

Then,

\[ E(Y - Z^s' \beta|X,U) = m(X) + \lambda(U) \] (8)

Let \( f_{X,U}, f_X, \) and \( f_U \) denote the joint densities of \( (X' U')', X \) and \( U \) respectively, and define the function \( r(x,u) = \frac{f_X(x)f_U(u)}{f_{X,U}(x,u)} \forall x,u \) such that \( f_{X,U}(x,u) \neq 0 \). From the definition of \( \lambda(U) \) observe that if we further assume \( E(\varepsilon) = 0 \), then by the law of iterated expectations \( E(\lambda(u)) = 0 \). It can be easily verified that \( E((Y - Z^s')r(x,u)|X) = m(X) + E(\lambda(u)) \). Then,

\[ E((Y - Z^s' \beta)r(x,u)|X) = m(x) \] (9)

If we define \( E(m(x)) \equiv \alpha_m \), then we also have that

\[ E((Y - Z^s' \beta)r(x,u)|U) = \alpha_m + \lambda(U) \] (10)
If we define $V \equiv \varepsilon - \lambda(U)$, we can write (1) as

$$Y - Z^t \beta = m(X) + \lambda(U) + V \quad (11)$$

Then, if we take the expectation of both sides of (10), we get that

$$E(E((Y - Z^t \beta)r(x,u)|U)) = E((Y - Z^t \beta)r(x,u))$$

$$= E(Yr(x,u)) - E(Z^t r(x,u))\beta \quad (13)$$

$$= \alpha_m \quad (14)$$

Given (1), (9), (10), and (11), we can write,

$$Y - E(Yr(X,U)|X) - E(Yr(X,U)|U) = (Z^t - E(Z^t r(X,U)|X)) - E(Z^t r(X,U|U))\beta - \alpha_m + V \quad (15)$$

Then substituting in (14) we have

$$W(Y,X,U) = R(Z^*,X,U)\beta + V \quad (16)$$

Where $W(Y,X,U) \equiv Y - E(Yr(X,U)|X) - E(Yr(X,U)|U) + E(Yr(X,U))$ and $R(Z^*,X,U) \equiv Z^t - E(Z^t r(X,U)|X) - E(Z^t r(X,U)|U) + E(Z^t r(X,U))$. Observe that since $E(V|Z,X,U) = 0$, $E(R(Z^*,X,U)^tV) = 0$ by the law of iterated expectations. Then we can write that

$$E(R(Z^*,X,U)^t W(Y,X,U)) = E(R(Z^*,X,U)^t R(Z^*,X,U))\beta \quad (17)$$

Assuming that $det(E(R(Z^*,X,U)^t R(Z^*,X,U))) \neq 0$,

$$\beta = E(R(Z^*,X,U)^t R(Z^*,X,U))^{-1} E(R(Z^*,X,U)^t W(Y,X,U)) \quad (18)$$

Note that for any measurable nonzero function $\omega(X,U) : \mathbb{R}^D \times \mathbb{R}^D \to \mathbb{R}$

$$\omega(X,U)W(Y,X,U) = \omega(X,U)R(Z^*,X,U)\beta + \omega(X,U)V \quad (19)$$

where $E(\omega(X,U)^t V) = 0$ and $E(\omega(X,U)^2 R(Z^*,X,U)^t V) = 0$

So if $det(E(\omega(X,U)^2 R(Z^*,X,U)^t R(Z^*,X,U))) \neq 0$, we have that

$$\beta = E(\omega(X,U)^2 R(Z^*,X,U)^t R(Z^*,X,U))^{-1} E(\omega(X,U)^2 R(Z^*,X,U)^t W(Y,X,U)) \quad (20)$$
4 Estimation

Suppose we observe a random sample \( \{(Y_i, Z_{i1}, \ldots, Z_{iL}, X_{i1}, \ldots, X_{iD})\}_{i=1}^n \) of some sample size \( n \). Define the following notation \( Y' = (Y_1 \cdots Y_n) \), \( Z = (Z_{i1})_{i=1,l=1}^n \) and \( X = (X_{i1},d)_{i=1,d=1}^n \). For any arbitrary matrix \( A \) of dimension \( n \times P \), \( A_{ip} \) denotes the \( p \)-th column of \( A \) and \( A_{il} \) denotes the \( i \)-th row of \( A \). For a conformable vector \( a \), define

\[
C(A, a) = \begin{pmatrix}
1 & A_{11} - a_1 & A_{12} - a_2 & \ldots & A_{1P} - a_P \\
1 & A_{21} - a_1 & A_{22} - a_2 & \ldots & A_{2P} - a_P \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & A_{n1} - a_1 & A_{n2} - a_2 & \ldots & A_{nP} - a_P
\end{pmatrix}
\]

In addition, given a multivariate kernel \( K(a) : \mathbb{R}^P \rightarrow \mathbb{R} \) with some bandwidth \( h > 0 \) denote \( \mathcal{K}(a, h) = \text{diag}\{K(A_{il} - a)\}_{i=1}^n \) then for some vector \( y \) denote the local linear smooth evaluated at \( a \) by

\[
\hat{\pi}_K(a; A, h, y) = s_K(a; A, h)
\]

where

\[
s_K(a; A, h) = e_p(C(A, a)' \mathcal{K}(a, h)C(A, a))^{-1}(C(A, a)' \mathcal{K}(a, h)\mathbf{A})
\]

and \( e_p \) is a \( P + 1 \times 1 \) dimensional vector where the first element is 1, and the rest are 0. The local constant smooth results when \( C(A, a)' = (1 \cdots 1) \), denoted by \( \hat{\pi}_K(a; A, h, y) \).

For \( z \in \mathbb{R}^L \) we define \( \hat{g}_d(z) = \hat{\pi}_K(z; \mathbf{A}, h, \mathbf{X}) \), \( \hat{g}(z)' = (\hat{g}_1(z) \cdots \hat{g}_D(z)) \) and the residual vector \( \hat{U}_i = \mathbf{X}_i - \hat{g}(\mathbf{Z}_i)' \) for \( i = 1, \ldots, n \) where \( \hat{U}_i \) is the \( i \)-th row of the matrix \( \hat{U} \).

Next, define \( \hat{r}(\mathbf{X}_j, \hat{U}_j') \equiv \hat{r}_j = \frac{f_{\hat{U}}(\mathbf{X}_j, \hat{U}_j')}{f_{\mathbf{X}}(\mathbf{X}_j, \hat{U}_j')} \) where

\[
\hat{f}_{\mathbf{X}}(\mathbf{X}_j', \hat{U}_j') = \frac{1}{nh_Ph_U^2} \sum_{i=1, i \neq j}^n K_1 \left( \frac{\hat{U}_i - \hat{U}_j}{h_U} \right) K_2 \left( \frac{\mathbf{X}_i' - \mathbf{X}_j'}{h_X} \right)
\]

\[
\hat{f}_{\hat{U}}(\hat{U}_j') = \frac{1}{nh_U^2} \sum_{i=1, i \neq j}^n K_1 \left( \frac{\hat{U}_i - \hat{U}_j}{h_U} \right) , \hat{f}_{\mathbf{X}}(\mathbf{X}_j') = \frac{1}{nh_X^2} \sum_{i=1, i \neq j}^n K_2 \left( \frac{\mathbf{X}_i' - \mathbf{X}_j'}{h_X} \right)
\]

for some kernels \( K_1, K_2 \) and bandwidths \( h_X, h_U \).

Let \( \mathbf{A} \circ \mathbf{B} \) denote the Hadamard product of two conformable matrices and let \( \hat{r}' = (\hat{r}_1 \cdots \hat{r}_n) \) and \( \mathbf{Z}' = (Z_{i1})_{i=1,l=1}^n \). We define the following estimators for \( W_i \equiv W(Y_i, X_{i1}', \hat{U}_{i1}') \) and \( R_i \equiv R(Z_{i1}', X_{i1}', \hat{U}_{i1}') \) that are defined in (16).
\[
\tilde{W}_i = Y_i - \pi_{K_2}^1(X_i', : X, h_X, Y \circ \hat{r}) - \pi_{K_1}^1(U_i', : \hat{U}, h_U, Y \circ \hat{r}) + \frac{1}{n} Y \circ \hat{r}
\]

\[
\tilde{R}_i = Z_i' - (\pi_{K_2}^1(X_i', : X, h_X, Z_1 \circ \hat{r}) \cdots \pi_{K_2}^1(X_i', : X, h_X, Z_{L_1} \circ \hat{r})
- (\pi_{K_1}^1(U_i', : \hat{U}, h_U, Z_1 \circ \hat{r}) \cdots \pi_{K_1}^1(U_i', : \hat{U}, h_U, Z_{L_1} \circ \hat{r})
+ \frac{1}{n} Y \circ \hat{r} \cdots Z_{L_1} \circ \hat{r})
\]

for \(i = 1, \ldots, n\)

Then, based on (20), let \(\omega(X, U)^2\) be estimated at all points by \(\hat{r}_i\), define the weighted pilot estimator for \(\beta\) as

\[
\beta_{wp} = \left(\frac{1}{n} \sum_{i=1}^{n} \tilde{R}_i' \tilde{R}_i\right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \tilde{R}_i' \tilde{W}_i
\]

We will also test an unweighted estimator for \(\beta\) for comparison purposes, defined as

\[
\beta_p = \left(\frac{1}{n} \sum_{i=1}^{n} \tilde{R}_i' \tilde{R}_i\right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \tilde{R}_i' \tilde{W}_i
\]

Based on (9), (10), and (14), we use \(\beta_{wp}\), to define the following pilot estimators,

\[
\tilde{m}(X_i') = \pi_{K_2}^1(X_i', : X, h_X, (Y - Z \beta_{wp}) \circ \hat{r})
\]

\[
\tilde{\lambda}(U_i') = \pi_{K_1}^1(U_i', : \hat{U}, h_U, (Y - Z \beta_{wp}) \circ \hat{r}) - \tilde{\alpha}_m
\]

where

\[
\tilde{\alpha}_m = \frac{1}{n} Y \circ \hat{r} - \frac{1}{n} Z \circ \hat{r} \cdots Z \circ \hat{r})\beta_{wp}
\]

Alternatively, we can estimate \(m\) with a one step backfitting procedure. Let

\[
\hat{\lambda}' = (\hat{\lambda}(U_1'), \ldots, \hat{\lambda}(U_n'))
\]

and \(Y^m = Y - Z \beta_{wp} - \hat{\lambda}\). For \(x \in \mathbb{R}^D\), define

\[
\tilde{m}(x) = \pi_{K_2}^1(x, : X, h_x, Y)
\]

We also estimate (23), (24), (25), and (26) using \(\beta_p\), rather than \(\beta_{wp}\), in our estimates of \(m\) to assess the validity of the assertion in Martins-Filho et al. (2015) that the estimators for \(m\) are oracle efficient, as we use the same DGP and they did not test this case.
5 Monte Carlo study

To test this estimation procedure, we conducted a Monte Carlo experiment. For the computational feasibility, we set \( D = 1 \), \( L = 2 \), and \( L_1 = 1 \), and consider the following regressions

\[
DGP_1 : Y_i = Z_{i,1} \beta + \ln(|X_i - 1| + 1) \text{sgn}(X_i - 1) + \varepsilon_i, \quad (27)
\]

\[
DGP_2 : Y_i = Z_{i,1} \beta + 2 + \cos(X_i) + \varepsilon_i, \quad (28)
\]

\[
DGP_3 : Y_i = Z_{i,1} \beta + \frac{3 \exp(X_i)}{1 + \exp(X_i)} + \varepsilon_i \quad (29)
\]

for \( i = 1, \ldots, n \). We test \( n = 200, 400, \) and \( 600 \), and we set \( \beta = .5 \) and \( 1 \). In all three cases, \( X_i = Z_{i1}^2 + Z_{i2}^2 + U_i \), where \( Z_{i1}, Z_{i2}, \varepsilon_i \) and \( U_i \) are generated respectively as

\[
\begin{pmatrix}
Z_{i1} \\
Z_{i2}
\end{pmatrix}
\sim iid \ N\left( \begin{pmatrix} 0 \\
0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\
0.5 & 1 \end{pmatrix} \right)
\]

and

\[
\begin{pmatrix}
\varepsilon_i \\
U_i
\end{pmatrix}
\sim iid \ N\left( \begin{pmatrix} 0 \\
0 \end{pmatrix}, \begin{pmatrix} 1 & \theta \\
\theta & 1 \end{pmatrix} \right)
\]

where \( \theta = .3, .6, \) or \( .9 \) representing weak, medium, and strong endogeneity. We perform each parameter set 1000 times. We note that similar \( DGP \)s are used in Ai and Chen (2003), Su (2008), Martins-Filho and Yao (2012), and Martins-Filho et al. (2015). We chose the second order univariate Epanechnikov kernel for \( K_1 \) and \( K_2 \), and the product of two univariate Epanechnikov kernels for \( K \) where applicable. We used the rule of thumb bandwidth \( \hat{2} \hat{\sigma}(W)n^{-1/k} \) where \( k = 5 \) for the univariate Epanechnikov kernel, and \( k = 10 \) for the bivariate case. We use a bivariate local linear estimator \( \hat{g}(Z_{i1}, Z_{i2}) \) for \( g(Z_{i1}, Z_{i2}) \) to obtain the residuals \( \hat{U} \). We are testing both the weighted pilot estimator for \( \beta \) as seen in (21), and an unweighted pilot estimator for \( \beta \) for comparison purposes as seen in (22). For the estimators of \( m(x) \) we test the simple pilot estimator defined in (23), as well as the one step back fit estimator in (26). We test each estimation method using both \( \beta_{wp} \) and \( \beta_p \) so we can assess the claim in Martins-Filho et al. (2015) that the \( m(x) \) estimators are oracle efficient.

It is worth noting the short computational time required for this estimation method. For example, for a sample size of 200, one design run takes roughly 7 minutes in MATLAB on a 1.8 GHz Intel Core i5 processor.
We compare the bias ($B$), standard deviation ($S$) and root mean square error ($R$) of the estimates $\beta_p$ and $\beta_{wp}$, and use the mean of the root mean square error ($M$) to gauge the efficiency of the two estimates for $m(X)$ given by (23) and (26). For examining the efficiency of the $\beta$ estimates, we separate the (B), (S), and (R) values into tables by which DGP was chosen for $m(X)$. We separate rows by the true value chosen for $\beta$ and sample size ($n$) and columns by the level of endogeneity ($\theta$). The last table is the mean RMSE values for the two estimates of $m(X)$ calculated with both estimates for $\beta$ are separated vertically by sample size and horizontally by the level of endogeneity $\theta$.

We see that the estimates for $\beta$ are increasingly efficient as the sample size grows, and not particularly sensitive to the level of endogeneity, especially at larger sample sizes. In the smallest sample size, there is a slight observable increase in $S$ and $R$ as theta increases, but this increase is on the order of one tenth of the values for $S$ and $R$. With only a few exceptions, $\beta_p$ dominates $\beta_{wp}$ for all 3 of the metrics at every DGP and endogeneity level. For the estimates of $\beta$ we also see that the RMSE is dominated by the standard deviation, as the bias is consistently an order of magnitude smaller than the standard deviation across DGPs, with the larger sample sizes showing an even larger magnitude difference. We note that there is no clear impact on the magnitude of $\beta$ on the efficiency of the estimators.

For the estimates of $m(x)$, we see that the one step backfitting method $\hat{m}(x)$ dominates the pilot estimator $\tilde{m}(x)$ across all data designs and sample sizes, using either estimate for $\beta$. We notice that M actually decreases as endogeneity increases almost everywhere, and in some cases this decrease is quite large. Like with the estimates for $\beta$, the estimates of $m$ do not seem to be sensitive to the magnitude of $\beta$, as $M$ is consistent across fixed DGPs and sample size, but varying values for $\beta$.

Comparing across DGPs, DGP$_1$ has consistently lower values for $S$ and $R$ than DGP$_2$, which in turn is lower than DGP$_3$. For the estimate of $m(x)$, DGP$_1$ again is consistently easiest to estimate, where as DGP$_3$ has slightly smaller values compared to DGP$_2$. Then, overall DGP$_1$ is easiest to estimate for both estimators of $\beta$ and both of $m$.

Martins-Filho et al. (2015) showed that $\beta_{wp}$ was consistent in the presence of endogeneity, and given our finite sample results we propose that $\beta_p$ is also a consistent estimator for $\beta$. If this is true, $\beta_p$ could
be used as a pilot estimator to construct a semiparametric efficient estimator for $\beta$, a process discussed in Martins-Filho et al. (2015). Our results for $\hat{m}(x)$ and $\tilde{m}(x)$ using $\beta_{wp}$ and $\beta_p$ supports the assertion that $\hat{m}(x)$ and $\tilde{m}(x)$ are oracle efficient and only a consistent estimator for $\beta$ is needed. There seems to be a general trend that $\hat{m}(x)$ preforms better with $\beta_{wp}$ for $DGP_3$, while in $DGP_1, DGP_2$ it has a very slight edge with $\beta_p$ under the two higher levels of endogeneity. We do not observe this same kind of difference with $\tilde{m}(x)$.

From these results, we can conclude that these various estimators suitably account for the endogeneity of $X$, in general $\beta_p$ out preforms $\beta_{wp}$ in estimating $\beta$, and $\hat{m}(x)$ estimated with either estimate for $\beta$ outperforms $\tilde{m}(x)$, and both $\hat{m}(x)$ and $\tilde{m}(x)$ are relatively insensitive to the estimation method of $\beta$ used in the calculations.

We include in figure demonstrating the estimators for the design $DGP_2, \beta = 1, \theta = .9$ and $n = 200$. It is visually evident that while both estimators match the curvature of the underlying structure, $\hat{m}(x)$ provides a much better goodness of fit.

6 Summary and Closing Remarks

In this paper we examine the derivation, computation, and asymptotic properties of a set of estimation procedures put forth by Martins-Filho et al. (2015), and test their finite sample performances with a Monte Carlo study. Under various $DGP$ designs, we conclude that the estimators sufficiently account for endogenous variables entering the model nonparametrically, where the one step backfitting procedure and unweighted pilot estimators are more efficient in finite samples than their counterparts for estimating $m(x)$ and $\beta$ respectively.

In the same spirit, we could test estimators for alternative data structures than we had here. We could impose a different structure to the error, perhaps one which was conditionally heteroskedastic with or without stationarity similar to what was considered in Gao and Phillips (2013). Whatever the structure of the data, Monte Carlo studies will be perpetually useful in discerning the finite sample performance of various estimators, and will continue to be a mainstay in econometric literature, so there will always be work
of this nature to be done.

To expand the scope of this specific Monte Carlo study, we could examine the derivation of the semi-parametric efficient estimator for $\beta$ proposed in Martins-Filho et al. (2015) and compare the finite sample performance of an efficient estimator for $\beta$ to the others examined earlier in this paper. Alternatively, we could explicitly build on this work by examining the asymptotic properties of $\beta_p$ to confirm that it is $\sqrt{n}$ consistent, and therefore suitable as the first stage in efficiently estimating $\beta$, rather than just postulate this based on the finite sample properties.
References


7 Appendix - Tables and Figures

| Table 1: Finite sample performance of β estimates for DGP 1 |
|-----------------|-----------------|-----------------|
|                 | θ = .3          | θ = .6          | θ = .9          |
|                 | B               | S               | R               | B               | S               | R               | B               | S               | R               |
| β = .5          | n = 200         |                 |                 |                 |                 |                 |                 |                 |                 |
| β_p             | -0.0039         | 0.2422          | 0.2421          | -0.0105         | 0.2278          | 0.2279          | 0.0064          | 0.2235          | 0.2234          |
| β_wp            | -0.0046         | 0.3456          | 0.3455          | -0.0178         | 0.3732          | 0.3734          | 0.0237          | 0.4625          | 0.4629          |
| β = 1           | n = 400         |                 |                 |                 |                 |                 |                 |                 |                 |
| β_p             | -0.0027         | 0.1597          | 0.1596          | 0.0010          | 0.1722          | 0.1721          | -0.0053         | 0.1397          | 0.1397          |
| β_wp            | -0.0028         | 0.2516          | 0.2515          | 0.0121          | 0.3092          | 0.3093          | -0.0086         | 0.2525          | 0.2525          |
| β = 1           | n = 600         |                 |                 |                 |                 |                 |                 |                 |                 |
| β_p             | 0.0000          | 0.1398          | 0.1397          | -0.0001         | 0.1186          | 0.1185          | -0.0023         | 0.1019          | 0.1019          |
| β_wp            | -0.0003         | 0.2336          | 0.2335          | -0.0049         | 0.2167          | 0.2167          | -0.0071         | 0.2047          | 0.2047          |
| β = 1           | n = 200         |                 |                 |                 |                 |                 |                 |                 |                 |
| β_p             | -0.0109         | 0.2374          | 0.2375          | 0.0004          | 0.2301          | 0.2300          | 0.0094          | 0.2238          | 0.2239          |
| β_wp            | -1.9472         | 0.3791          | 0.3794          | 0.0027          | 0.3432          | 0.3430          | 0.0059          | 0.5133          | 0.5131          |
| β = 1           | n = 400         |                 |                 |                 |                 |                 |                 |                 |                 |
| β_p             | -0.0018         | 0.1957          | 0.1956          | -0.0109         | 0.1518          | 0.1521          | -0.0014         | 0.1508          | 0.1507          |
| β_wp            | -0.0030         | 0.3069          | 0.3067          | -0.0129         | 0.2606          | 0.2608          | 0.0015          | 0.2692          | 0.2690          |
| β = 1           | n = 600         |                 |                 |                 |                 |                 |                 |                 |                 |
| β_p             | 0.0013          | 0.1460          | 0.1459          | 0.0036          | 0.1202          | 0.1202          | -0.0005         | 0.1082          | 0.1081          |
| β_wp            | 0.0031          | 0.2456          | 0.2455          | 0.0067          | 0.2273          | 0.2273          | -0.0019         | 0.2234          | 0.2233          |
Table 2. Finite sample performance of $\beta$ estimates for $DGP_2$

<p>| $\theta = .3$ | $\theta = .6$ | $\theta = .9$ |</p>
<table>
<thead>
<tr>
<th>B</th>
<th>S</th>
<th>R</th>
<th>B</th>
<th>S</th>
<th>R</th>
<th>B</th>
<th>S</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta = .5$</td>
<td>$n = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_p$</td>
<td>-0.0012</td>
<td>0.3162</td>
<td>0.3161</td>
<td>-0.0017</td>
<td>0.2804</td>
<td>0.2803</td>
<td>-0.0134</td>
<td>0.2414</td>
</tr>
<tr>
<td>$\beta_{wp}$</td>
<td>-0.0078</td>
<td>0.5232</td>
<td>0.5230</td>
<td>-0.0149</td>
<td>0.5156</td>
<td>0.5156</td>
<td>-0.0315</td>
<td>0.4609</td>
</tr>
<tr>
<td>$n = 400$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_p$</td>
<td>0.0055</td>
<td>0.2008</td>
<td>0.2008</td>
<td>-0.0018</td>
<td>0.1717</td>
<td>0.1716</td>
<td>0.0117</td>
<td>0.1867</td>
</tr>
<tr>
<td>$\beta_{wp}$</td>
<td>0.0094</td>
<td>0.3749</td>
<td>0.3749</td>
<td>0.0043</td>
<td>0.3442</td>
<td>0.3441</td>
<td>0.0299</td>
<td>0.3674</td>
</tr>
<tr>
<td>$n = 600$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_p$</td>
<td>0.0112</td>
<td>0.1908</td>
<td>0.1910</td>
<td>-0.0030</td>
<td>0.1423</td>
<td>0.1423</td>
<td>0.0012</td>
<td>0.1213</td>
</tr>
<tr>
<td>$\beta_{wp}$</td>
<td>0.0297</td>
<td>0.4609</td>
<td>0.4616</td>
<td>-0.0065</td>
<td>0.3001</td>
<td>0.3000</td>
<td>0.0029</td>
<td>0.2526</td>
</tr>
<tr>
<td>$\beta = 1$</td>
<td>$n = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_p$</td>
<td>-0.0133</td>
<td>0.2973</td>
<td>0.2975</td>
<td>-0.0025</td>
<td>0.2884</td>
<td>0.2882</td>
<td>-0.0029</td>
<td>0.2429</td>
</tr>
<tr>
<td>$\beta_{wp}$</td>
<td>-0.0128</td>
<td>0.5650</td>
<td>0.5649</td>
<td>-0.0063</td>
<td>0.4951</td>
<td>0.4949</td>
<td>-0.0005</td>
<td>0.4262</td>
</tr>
<tr>
<td>$n = 400$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_p$</td>
<td>-0.0075</td>
<td>0.2177</td>
<td>0.2177</td>
<td>0.0038</td>
<td>0.1981</td>
<td>0.1980</td>
<td>0.0102</td>
<td>0.1964</td>
</tr>
<tr>
<td>$\beta_{wp}$</td>
<td>-0.0074</td>
<td>0.4191</td>
<td>0.4190</td>
<td>0.0054</td>
<td>0.4270</td>
<td>0.4268</td>
<td>0.0149</td>
<td>0.3697</td>
</tr>
<tr>
<td>$n = 600$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_p$</td>
<td>0.0000</td>
<td>0.1472</td>
<td>0.1471</td>
<td>-0.0085</td>
<td>0.1553</td>
<td>0.1554</td>
<td>0.0068</td>
<td>0.1262</td>
</tr>
<tr>
<td>$\beta_{wp}$</td>
<td>-0.0032</td>
<td>0.3070</td>
<td>0.3069</td>
<td>-0.0274</td>
<td>0.4020</td>
<td>0.4027</td>
<td>0.0103</td>
<td>0.2650</td>
</tr>
</tbody>
</table>

Table 3. Finite sample performance of $\beta$ estimates for $DGP_3$

<p>| $\theta = .3$ | $\theta = .6$ | $\theta = .9$ |</p>
<table>
<thead>
<tr>
<th>B</th>
<th>S</th>
<th>R</th>
<th>B</th>
<th>S</th>
<th>R</th>
<th>B</th>
<th>S</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta = .5$</td>
<td>$n = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_p$</td>
<td>-0.0008</td>
<td>0.4053</td>
<td>0.4051</td>
<td>-0.0111</td>
<td>0.3463</td>
<td>0.3464</td>
<td>0.0091</td>
<td>0.3415</td>
</tr>
<tr>
<td>$\beta_{wp}$</td>
<td>0.0041</td>
<td>0.6685</td>
<td>0.6682</td>
<td>-0.0241</td>
<td>0.6322</td>
<td>0.6323</td>
<td>0.0079</td>
<td>0.6283</td>
</tr>
<tr>
<td>$n = 400$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_p$</td>
<td>0.0012</td>
<td>0.2804</td>
<td>0.2803</td>
<td>0.0112</td>
<td>0.2441</td>
<td>0.2442</td>
<td>0.0013</td>
<td>0.1975</td>
</tr>
<tr>
<td>$\beta_{wp}$</td>
<td>-0.0007</td>
<td>0.4789</td>
<td>0.4786</td>
<td>0.0195</td>
<td>0.4623</td>
<td>0.4625</td>
<td>0.0009</td>
<td>0.3968</td>
</tr>
<tr>
<td>$n = 600$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_p$</td>
<td>-0.0001</td>
<td>0.2226</td>
<td>0.2225</td>
<td>0.0016</td>
<td>0.2035</td>
<td>0.2034</td>
<td>-0.0013</td>
<td>0.1836</td>
</tr>
<tr>
<td>$\beta_{wp}$</td>
<td>-0.0038</td>
<td>0.4295</td>
<td>0.4293</td>
<td>0.0064</td>
<td>0.3975</td>
<td>0.3974</td>
<td>-0.0027</td>
<td>0.3802</td>
</tr>
<tr>
<td>$\beta = 1$</td>
<td>$n = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_p$</td>
<td>0.0032</td>
<td>0.3368</td>
<td>0.3366</td>
<td>-0.0077</td>
<td>0.3764</td>
<td>0.3763</td>
<td>-0.0323</td>
<td>0.3307</td>
</tr>
<tr>
<td>$\beta_{wp}$</td>
<td>0.0068</td>
<td>0.5799</td>
<td>0.5797</td>
<td>-0.0082</td>
<td>0.6155</td>
<td>0.6153</td>
<td>-0.0525</td>
<td>0.5697</td>
</tr>
<tr>
<td>$n = 400$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_p$</td>
<td>0.0049</td>
<td>0.2830</td>
<td>0.2829</td>
<td>0.0159</td>
<td>0.2507</td>
<td>0.2511</td>
<td>0.0083</td>
<td>0.2242</td>
</tr>
<tr>
<td>$\beta_{wp}$</td>
<td>0.0118</td>
<td>0.5334</td>
<td>0.5332</td>
<td>0.0399</td>
<td>0.5815</td>
<td>0.5825</td>
<td>0.0151</td>
<td>0.4396</td>
</tr>
<tr>
<td>$n = 600$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_p$</td>
<td>0.0021</td>
<td>0.2456</td>
<td>0.2454</td>
<td>-0.0122</td>
<td>0.2360</td>
<td>0.2362</td>
<td>-0.0032</td>
<td>0.1643</td>
</tr>
<tr>
<td>$\beta_{wp}$</td>
<td>0.0044</td>
<td>0.4720</td>
<td>0.4718</td>
<td>-0.0252</td>
<td>0.4547</td>
<td>0.4552</td>
<td>-0.0046</td>
<td>0.3503</td>
</tr>
<tr>
<td>$M$</td>
<td>$DGP_1$</td>
<td>$DGP_2$</td>
<td>$DGP_3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\theta = .3$</td>
<td>$\theta = .6$</td>
<td>$\theta = .9$</td>
<td>$\theta = .3$</td>
<td>$\theta = .6$</td>
<td>$\theta = .9$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta = .5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| $n = 200$ | (\hat{n}(x), \beta_p) 0.1939 | 0.1684 | 0.1385 | 0.4040 | 0.3979 | 0.3831 | 0.3913 | 0.3754 | 0.3476  
|  | (\hat{n}(x), \beta_{up}) 0.1980 | 0.1712 | 0.1532 | 0.4016 | 0.4056 | 0.3932 | 0.4054 | 0.3670 | 0.3451  
|  | (\hat{n}(x), \beta_p) 0.1602 | 0.1429 | 0.1223 | 0.2896 | 0.2970 | 0.3027 | 0.1627 | 0.1574 | 0.1488  
|  | (\hat{n}(x), \beta_{up}) 0.1637 | 0.1454 | 0.1280 | 0.2859 | 0.2945 | 0.3016 | 0.1745 | 0.1650 | 0.1542  
| $n = 400$ | (\hat{n}(x), \beta_p) 0.1845 | 0.1291 | 0.1036 | 0.3632 | 0.3567 | 0.3400 | 0.3398 | 0.3321 | 0.3045  
|  | (\hat{n}(x), \beta_{up}) 0.1838 | 0.1296 | 0.1105 | 0.3610 | 0.3668 | 0.3496 | 0.3296 | 0.3252 | 0.2969  
|  | (\hat{n}(x), \beta_p) 0.1212 | 0.1047 | 0.0829 | 0.2688 | 0.2766 | 0.2873 | 0.1434 | 0.1409 | 0.1420  
|  | (\hat{n}(x), \beta_{up}) 0.1221 | 0.1070 | 0.0852 | 0.2646 | 0.2728 | 0.2852 | 0.1449 | 0.1404 | 0.1414  
| $n = 600$ | (\hat{n}(x), \beta_p) 0.1355 | 0.1066 | 0.0799 | 0.3481 | 0.3309 | 0.3281 | 0.3115 | 0.3017 | 0.2890  
|  | (\hat{n}(x), \beta_{up}) 0.1325 | 0.1103 | 0.0826 | 0.3492 | 0.3345 | 0.3348 | 0.2966 | 0.2974 | 0.2794  
|  | (\hat{n}(x), \beta_p) 0.0988 | 0.0825 | 0.0654 | 0.2569 | 0.2677 | 0.2805 | 0.1346 | 0.1370 | 0.1453  
|  | (\hat{n}(x), \beta_{up}) 0.0997 | 0.0834 | 0.0672 | 0.2545 | 0.2649 | 0.2793 | 0.1332 | 0.1367 | 0.1426  
| $\beta = 1$ |  
| $n = 200$ | (\hat{n}(x), \beta_p) 0.2051 | 0.1764 | 0.1325 | 0.4089 | 0.4261 | 0.3837 | 0.4355 | 0.3823 | 0.3762  
|  | (\hat{n}(x), \beta_{up}) 0.2128 | 0.1909 | 0.1411 | 0.4145 | 0.4350 | 0.3944 | 0.4300 | 0.3802 | 0.3848  
|  | (\hat{n}(x), \beta_p) 0.1568 | 0.1429 | 0.1215 | 0.2900 | 0.3043 | 0.3033 | 0.1792 | 0.1676 | 0.1480  
|  | (\hat{n}(x), \beta_{up}) 0.1597 | 0.1467 | 0.1273 | 0.2868 | 0.3001 | 0.3027 | 0.1806 | 0.1712 | 0.1520  
| $n = 400$ | (\hat{n}(x), \beta_p) 0.1547 | 0.1233 | 0.0939 | 0.3637 | 0.3541 | 0.3430 | 0.3637 | 0.3308 | 0.3029  
|  | (\hat{n}(x), \beta_{up}) 0.1520 | 0.1235 | 0.0972 | 0.3711 | 0.3586 | 0.3523 | 0.3440 | 0.3195 | 0.3054  
|  | (\hat{n}(x), \beta_p) 0.1155 | 0.1012 | 0.0823 | 0.2705 | 0.2762 | 0.2860 | 0.1490 | 0.1437 | 0.1438  
|  | (\hat{n}(x), \beta_{up}) 0.1163 | 0.1029 | 0.0833 | 0.2672 | 0.2736 | 0.2835 | 0.1472 | 0.1421 | 0.1445  
| $n = 600$ | (\hat{n}(x), \beta_p) 0.1328 | 0.1062 | 0.0843 | 0.3428 | 0.3309 | 0.3260 | 0.3165 | 0.2932 | 0.2873  
|  | (\hat{n}(x), \beta_{up}) 0.1296 | 0.1065 | 0.0881 | 0.3486 | 0.3356 | 0.3317 | 0.2991 | 0.2876 | 0.2776  
|  | (\hat{n}(x), \beta_p) 0.0992 | 0.0857 | 0.0647 | 0.2561 | 0.2671 | 0.2813 | 0.1330 | 0.1348 | 0.1461  
|  | (\hat{n}(x), \beta_{up}) 0.1005 | 0.0869 | 0.0671 | 0.2536 | 0.2641 | 0.2792 | 0.1316 | 0.1336 | 0.1438  

Table 4: Finite sample mean root mean square error (M) of $\beta$ estimators under all designs.
Figure 1: This is a plot of $m(x)$ and the estimates using $\beta_p$ under the design $DGP_2$, $\beta = 1$, $\theta = .9$ and $n = 200$. 