# Identification and Nonparametric Estimation of a Transformed Additively Separable Model*

David Jacho-Chávez† Indiana University

Arthur Lewbel‡ Boston College

Oliver Linton§ London School of Economics

May 5, 2008

**Abstract**

Let $r(x,z)$ be a function that, along with its derivatives, can be consistently estimated nonparametrically. This paper discusses identification and consistent estimation of the unknown functions $H$, $M$, $G$ and $F$, where $r(x,z) = H[M(x,z)]$, $M(x,z) = G(x) + F(z)$, and $H$ is strictly monotonic. An estimation algorithm is proposed for each of the model's unknown components when $r(x,z)$ represents a conditional mean function. The resulting estimators use marginal integration, and are shown to have a limiting Normal distribution with a faster rate of convergence than unrestricted nonparametric alternatives. Their small sample performance is studied in a Monte Carlo experiment. We empirically apply our results to nonparametrically estimate generalized homothetic production functions in four industries within the Chinese economy.

**Keywords:** Partly separable models; Nonparametric regression; Dimension reduction; Generalized homothetic function; Production function.

***JEL classification:*** C13; C14; C21; D24

*We would like to thank the associate editor and two anonymous referees for many helpful and insightful suggestions. We are indebted to Gary Jefferson and Miguel Acosta for providing us with the production data for the Chinese and Ecuadorian economies respectively. We also thank seminar participants at the Singapore Econometrics Study Group, ESRC Econometric Study Group (Bristol), the Central Bank of Ecuador, University College London, Indiana University, Queen Mary University (London), and the Midwest Econometric Group 16th Annual Conference (Cincinnati) for helpful comments and suggestions, that greatly improved the presentation and discussion of this paper.

†Department of Economics, Indiana University, Wylie Hall 251, 100 South Woodlawn Avenue, Bloomington, IN 47405, USA. E-mail: djachoch@indiana.edu. Web Page: http://mypage.iu.edu/~djachoch/

‡Department of Economics, Boston College, 140 Commonwealth Avenue, Chesnut Hill, MA 02467, USA. E-mail: lewbel@bc.edu. Web Page: http://www2.bc.edu/~lewbel/

§Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, UK. E-mail: O.Linton@lse.ac.uk. Web Page: http://personal.lse.ac.uk/lintono/. I would like to thank the ESRC for financial support.

# 1 Introduction

For vector $x \in \Re^d$ and scalar $z$, let $r(x,z)$ be a function that, along with its derivatives, can be consistently estimated nonparametrically. We assume there exist unknown functions $H$, $G$ and $F$ such that

$$r(x,z) = H[M(x,z)] = H[G(x) + F(z)] \tag{1.1}$$

where $M(x,z) \equiv G(x) + F(z)$ and $H$ is strictly monotonic. This structure can arise from an economic model and from the statistical objective of reducing the curse of dimensionality as $d$ increases (see, e.g., Stone (1980) and Stone (1986)). This paper provides new sufficient conditions for identification of $H$, $M$, $G$ and $F$. An estimation algorithm is then proposed when $r(x,z)$ represents a conditional mean function for a given sample $\{Y_i, X_i, Z_i\}_{i=1}^n$. We provide limiting distributions for the resulting nonparametric estimators of each component of (1.1), we present evidence of their small sample performance in some Monte Carlo experiments, and we provide an empirical application.

This framework encompasses a large class of economic models. For example, the function $r(x,z)$ could be a utility or consumer cost function recovered from estimated consumer demand functions via revealed preference theory, or it could be an estimated production or producer cost function. Chiang (1984), Simon and Blume (1994), Bairam (1994), and Chung (1994) review popular parametric examples of (1.1) with $H[m] = m$, the identity function. In demand analysis, Goldman and Uzawa (1964) provide an overview of the variety of separability concepts implicit in such specifications.

Many methods have been developed for the identification and estimation of strongly or additively separable models, where $r(x,z) = \sum_{k=1}^d G_k(x_k) + F(z)$ or its generalized version $r(x,z) = H[\sum_{k=1}^d G_k(x_k) + F(z)]$. Friedman and Stutzle (1981), Breiman and Friedman (1985), Andrews (1991), Tjøstheim and Auestad (1994) and Linton and Nielsen (1995) are examples of the former while Linton and Härdle (1996), and Horowitz and Mammen (2004) provide estimators of the latter for known $H$. Horowitz (2001) uses this strong separability to identify the components of the model when $H$ is unknown, and proposes a kernel–based consistent and asymptotically normal estimator. In contrast with Horowitz, we obtain identification by assuming the link function $H$ is strictly monotonic instead of by assuming that $G$ has the additive form $\sum_{k=1}^d G_k(x_k)$.

A related result is Lewbel and Linton (2007), who identify and estimate models in the special case of (1.1) where $F(z) = z$, or equivalently where $F(z)$ is known. Pinkse (2001) provides a general nonparametric estimator for $\widetilde{G}$ in weakly separable models $r(x,z) = \widetilde{H}[\widetilde{G}(x), z]$. However, in Pinkse's specification, $\widetilde{G}$ is only identified up to an arbitrary monotonic transformation, while our model provides the unique $G$ and $F$ up to sign–scale and location normalizations.

One derivation of our model comes from ordinary partly additive regression models in which the dependent variable is censored, truncated, binary, or otherwise limited. These are models in which $Y^* = G(X) + F(Z) + \varepsilon$ for some unobserved $Y^*$ and $\varepsilon$, where $\varepsilon$ is independent of $(X,Z)$ with an absolutely continuous distribution function, and what is observed is $(Y,X,Z)$, where $Y$ is some function of $Y^*$ such as $Y = 1(Y^* \geq 0)$, or $Y = Y^* | Y^* \geq 0$, or $Y = 1(Y^* \geq 0)Y^*$, in which case $r(x,z) = E[Y|X=x, Z=z]$ or $r(x,z) = \text{med}[Y|X=x, Z=z]$. The function $H$ would then be the distribution function or quantile function of $\varepsilon$. Threshold or selection equations in particular are commonly of this form, having $Y = 1[G(X) + \varepsilon \geq -Z]$, where $-Z$ is some threshold, e.g., a price or a bid, with $G(X) + \varepsilon$ equalling willingness to pay or a reservation price; see, e.g., Lewbel, Linton, and McFadden (2002). Monotonicity of $H$ holds automatically in most of these examples because $H$ either equals or is a monotonic transformation of a distribution function.

Model (1.1) may arise in a nonparametric regression model with unknown transformation of the dependent variable, $F(z) = G(x) + \varepsilon$, where $\varepsilon$ has an absolutely continuous distribution function $H$ which is independent of $x$, $F$ is an unknown monotonic transformation and $G$ is an unknown regression function. It follows that the conditional distribution of $Z$ given $X$, $F_{Z|X}$, has the form $H(F(z) - G(x)) \equiv r(z,x)$, where $F_{Z|X} \equiv r(z,x)$. For this model, Ekeland, Heckman, and Nesheim (2004) provide an identification result that exploits separability between $x$ and $z$, but not the monotonicity of $H$ as we do here. Monotonicity of $H$ again holds in this example because $H$ is a distribution function.

The identification result presented here can also be used for identifying copulas nonparametrically. For example, 'strict' Archimedean copulas can be written as in (1.1), where the joint distribution of $(X,Z)$, $F_{XZ}(x,z)$, is such that $F_{XZ}(x,z) = \phi^{-1}(\phi(F_X(x)) + \phi(F_Z(z)))$, where $F_X(x)$, $F_Z(z)$ represent the marginal distributions of $X$ and $Z$ respectively, $\phi$ is a continuous strictly decreasing convex function from $[0,1]$ to $[0,\infty]$ such that $\phi(1) = 0$, and $\phi^{-1}$ denotes the inverse. A collection of Archimedean copulas can be found in Nelsen (2006).

Our methodology also encompasses models of the transformed multiplicative form $H[M(x,z)] = H[G(x)F(z)]$, which are common in the production function literature. Particularly, if $z \neq 0$ then a function $r(\widetilde{x},z)$ is defined to be "generalized homothetic" if and only if $r(\widetilde{x},z) = H[G(\widetilde{x}/z)F(z)]$ where $H$ is strictly monotonic, so by letting $x = \widetilde{x}/z$ we are providing a nonparametric estimator of generalized homothetic functions. Ordinary homothetic models, as estimated nonparametrically by Lewbel and Linton (2007), are the special case in which $F(z) = z$. Homotheticity and its variants, including the implied monotonicity of $H$, are commonly assumed in utility, production, and cost function contexts; see, e.g., Chiang (1984), Simon and Blume (1994), Chung (1994), and Goldman and Uzawa (1964).

We implement our methodology to estimate generalized homothetic production functions for four industries in China. For this, we have built an `R` package (see Ihaka and Gentleman (1996)), `JLLprod`, containing the functions that implement the techniques proposed here.

In most of the applications listed above the functions $H$, $G$ and $F$ are of direct economic interest, but even when they are not our proposed estimators will still be useful for dimension reduction and for testing whether or not functions have the proposed separability, by comparing $\widehat{r}(x,z)$ with $\widehat{H}[\widehat{G}(x) + \widehat{F}(z)]$, or in the production theory context, to test whether production functions are homothetic, by comparing $F(z) = z$ with $\widehat{F}(z)$. In addition, the more general model $r(x,z,w) = H[M(x,z), w]$ can also be identified using our methods when $M(x,z)$ is additive or multiplicative and $H$ is strictly monotonic with respect to its first argument.

Section 2 sets out the main identification results. Our proposed estimation algorithm is presented in Section 3. Section 4 analyzes the asymptotic properties of the estimators. A Monte Carlo experiment is presented in section 5 comparing our estimators to those proposed by Linton and Nielsen (1995), and Linton and Härdle (1996), both of which assume knowledge of $H$. This section also provides an empirical illustration of our methodology for the estimation of generalized production functions in four industries within the Chinese economy in 2001. Finally, Section 6 concludes and briefly outlines possible extensions.

## 2 Identification

The main identification idea is presented in this section. Observe that (1.1) is unchanged if $G$ and $F$ are replaced by $G + c_G$ and $F + c_F$, respectively, and $H(m)$ is replaced by $\widetilde{H}(m) = H(m - c_G - c_F)$. Similarly, (1.1) remains unchanged if $G$ and $F$ are replaced by $cG$ and $cF$ respectively, for some $c \neq 0$ and $H(m)$ is replaced by $\widetilde{H}(m) = H(m/c)$. Therefore, as is commonly the case in this literature, location and scale normalizations are needed to make identification possible. We will describe and discuss these normalizations below, but first we state the following conditions which are assumed to hold throughout our exposition.

Assumption I:

(I1) Let $W \equiv (X, Z)$ be a $(d+1)$-dimensional random vector with support $\Psi_x \times \Psi_z$, where $\Psi_x \subseteq \Re^d$, and $\Psi_z \subseteq \Re$, for some $d \geq 1$. The distribution of $W$ is absolutely continuous with respect to Lebesgue measure with probability density $f_W$ such that $\inf_{w=(x,z)\in\Psi_x\times\Psi_z} f_W(w) > 0$. There exists functions $r$, $H$, $G$ and $F$ such that $r(x,z) = H[G(x) + F(z)]$ for all $w \equiv (x,z) \in \Psi_x \times \Psi_z$.

(I2) (i) The function $H$ is strictly monotonic and $H$, $G$ and $F$ are continuous and differentiable with respect to any mixture of their arguments. (ii) $F$ has finite first derivative, $f$, over its entire support, and $f(z_0) = 1$ for some $z_0 \in int(\Psi_z)$. (iii) Let $H(0) = r_0$, where $r_0$ is a constant. In addition, (iv) Let $r(x,z) \in \Psi_{r(x,z_0)}$ for all $w \equiv (x,z) \in \Psi_x \times \Psi_z$, where $\Psi_{r(x,z)}$ is the image of the function $r(x,z)$.

Assumption (I1) specifies the model. The functions $M$, $G$ and $F$ may not be nonparametrically identified if $(X, Z)$ has discrete elements, so we rule this out, which is a common restriction in nonparametric models with unknown link function (see Horowitz (2001)). Assumption (I2) defines the location and scale normalizations required for identification. It also requires that the image of $r(x,z)$ over its entire support is replicated once $r$ is evaluated at $z_0$ for all $x$. This assumption implies that $s(x,z) \equiv \partial r(x,z)/\partial z$ is a well defined function for all $w \in \Psi_x \times \Psi_z$. Then, for the random variables $r(X,Z)$ and $s(X,Z)$, define the function $q(t,z)$ by

$$q(t,z) = E[s(X,Z)|\, r(X,Z) = t, Z = z]. \tag{2.1}$$

The assumed strict monotonicity of $H$ ensures that $H^{-1}$, the inverse function of $H$, is well defined over its entire support. Let $h$ be the first derivative of $H$.

**Theorem 2.1** *Let Assumption I hold. Then,*

$$M(x,z) \equiv G(x) + F(z) = \int_{r_0}^{r(x,z)} \frac{dt}{q(t,z_0)}. \tag{2.2}$$

**Proof.** It follows from Assumption (I1) that $s(x,z) = h[M(x,z)]\,f(z)$, so

$$\begin{aligned} E\left[s(X,Z)\right| r(X,Z) = t, Z = z_0] &= E\left[h[M(X,Z)]\,f(Z)\right| r(X,Z) = t, Z = z_0] \\ &= E\left[h\left[H^{-1}(r(X,Z))\right] f(Z)\right| r(X,Z) = t, Z = z_0\big] \\ &= h\left[H^{-1}(t)\right] f(z_0)\,, \text{ and} \end{aligned}$$

$q(t,z_0) = h\left[H^{-1}(t)\right] f(z_0)$. Then using the change of variables $m = H^{-1}(t)$, and noticing that $h\left[H^{-1}(t)\right] = h(m)$ and $dt = h(m)\,dm$, we obtain

$$\begin{aligned} \int_{r_0}^{r(x,z)} \frac{dt}{q(t,z_0)} &= \int_{r_0}^{r(x,z)} \frac{dt}{h\left[H^{-1}(t)\right] f(z_0)} = \int_{H^{-1}[r_0]}^{H^{-1}[r(x,z)]} \frac{h(m)\,dm}{h(m)\,f(z_0)} \\ &= \left(H^{-1}[r(x,z)] - H^{-1}[r_0]\right)(1/f(z_0)) = M(x,z) \equiv G(x) + F(z)\,, \end{aligned}$$

as required. ■

In the special case of an identity link function, i.e. $H(m) = m$, $q$ has a simple form $q(t,z_0) = f(z_0) \equiv q(z_0)$ which is constant over all $t$ and equals 1 by Assumption (I2). It is clear from the proof of this theorem that without knowledge of $z_0$ and $r_0$ in Assumptions (I2)(ii) and (I2)(iii), the function $M(x,z)$ could only be identified up to a sign–scale factor $1/f(z_0)$ and a location constant $H^{-1}(r_0)\,(1/f(z_0))$, provided $|f(z_0)| > 0$ and $\left|H^{-1}[r_0]\right| < \infty$. In addition, (I2)(iv) assumes a range of $(X,Z)$ that is large enough to obtain the function $r(X,Z)$ everywhere in the interval $r_0$ to $r(x,z)$. This ensures that $q$ exists everywhere on $\Psi_{r(x,z)} \times \Psi_z$, making $M(x,z)$ identifiable for all $x$ and $z$.

For the multiplicative model, $M(x,z) = G(x)\,F(z)$, which is a more natural representation of the model in some contexts such as production functions as discussed in the introduction, the following alternative assumption and corollary provides the necessary identification.

Assumption I*:

(I*1) Let $W \equiv (X,Z)$ be a $(d+1)$-dimensional random vector with support $\Psi_x \times \Psi_z$, where $\Psi_x \subseteq \Re^d$, and $\Psi_z \subseteq \Re$, for some $d \geq 1$. The distribution of $W$ is absolutely continuous with respect to Lebesgue measure with probability density $f_W(w)$ such that $\inf_{w=(x,z)\in\Psi_x\times\Psi_z} f_W(w) > 0$. There exists functions $r$, $H$, $G$ and $F$ such that $r(x,z) = H[G(x)\,F(z)]$ for all $w \equiv (x,z) \in \Psi_x \times \Psi_z$.

(I*2) (i) The function $H$ is strictly monotonic, continuous and differentiable. $G$ and $F$ are strictly positive continuous functions and differentiable with respect to any mixture of their arguments. (ii) $F$ has finite first derivative, $f$, such that $F(z_0)/f(z_0) = 1$ for some $z_0 \in int(\Psi_z)$. (iii) Let $H(1) = r_1$, where $r_1$ is a constant. In addition, (iv) Let $r(x,z) \in \Psi_{r(x,z_0)}$ for all $w \equiv (x,z) \in \Psi_x \times \Psi_z$, where $\Psi_{r(x,z)}$ is the image of the function $r(x,z)$.

**Corollary 2.1** *Let Assumption I* hold. Then,*

$$M(x,z) = G(x)\,F(z) = \exp\left(\int_{r_1}^{r(x,z)} \frac{dt}{q(t,z_0)}\right). \tag{2.3}$$

**Proof.** See the appendix. ■

If $r_l$ is greater than $r(x,z)$, for any nonnegative constant, then the integrals of the form $\int_{r_l}^{r(x,z)}$ above are to be interpreted as $-\int_{r_l}^{r(x,z)}$, for $l = 0, 1$. The normalization quantities $z_0$ and $r_1$ can often arise from economic theory. For example, the neoclassical production function of two inputs (say, capital $K$ and labor $L$) with positive, decreasing marginal products with respect to each factor and constant returns to scale, requires positive inputs of both factors for a positive output. If $r(K,L)$ represents such a function, $r_1 = r(0,L) = r(K,0) \equiv \min_{K,L} r(K,L)$ is a natural choice of normalization. Furthermore, if the production function has a multiplicative structure (see Section 5) with $F(L) = L$, then $f(L) = 1$ and any $L_0 > 0$ may be chosen, thereby providing all the normalizations needed for full identification.

Once $M(x,z)$ has been pulled out of the unknown (but strictly monotonic) function $H$ in (2.2) or (2.3), we may recover $G$ and $F$ by standard marginal integration as in Linton and Nielsen (1995). Let $P_1$ and $P_2$ be deterministic discrete or continuous weighting functions with Stieltjes integrals $\int_{\Psi_z} dP_1(z) = 1$ and $\int_{\Psi_x} dP_2(x) = 1$. Let $p_1$ and $p_2$ be the densities of $P_1$ and $P_2$ with respect to Lebesgue measure in $\Re$ and $\Re^d$ respectively. Then

$$\alpha_{P_1}(x) = \int_{\Psi_z} M(x,z)\, dP_1(z)\,, \text{ and } \alpha_{P_2}(z) = \int_{\Psi_x} M(x,z)\, dP_2(x)\,.$$

In the additive model, $\alpha_{P_1}(x) = G(x) + c_1$ and $\alpha_{P_2}(z) = F(z) + c_2$, where $c_1 = \int_{\Psi_z} F(z)\, dP_1(z)$ and $c_2 = \int_{\Psi_x} G(x)\, dP_2(x)$. While in the multiplicative case, $\alpha_{P_1}(x) = c_1 G(x)$ and $\alpha_{P_2}(z) = c_2 F(z)$. Hence, $\alpha_{P_1}(x)$ and $\alpha_{P_2}(z)$ are, up to identification normalizations, the components of $M$ in both the additive ($c = c_1 + c_2$) and multiplicative structures[1] ($c = c_1 \times c_2$).

Given the definition of $r(x,z)$, it follows that $H(M(x,z)) = E[r(X,Z)|\, M(X,Z) = M(x,z)]$, which shows that the function $H$ is identified since we have already identified $M$. If $r(x,z) \equiv E[Y|\, X = x, Z = z]$ for some random $Y$, then the equality $H(M(x,z)) = E[Y|\, M(X,Z) = M(x,z)]$ may also be used to identify[2] $H$.

Strict monotonicity of the link function $H$ plays an important role in these results. Because of this property, the conditional mean of $s(x,z)$ given $r$ and $z$ is a well-defined function, with a known structure which is separable in $z$. This monotonicity may often arise from the theory underlying the model, for example, strict monotonicity of $H$ follows from strict monotonicity of latent error distribution functions in the limited dependent variable examples described in the introduction, and this monotonicity may follow as a consequence of technology or preference axioms in production, utility, or cost function applications

[1]Similarly, $F$ can be recover directly the right-hand side of equations $\partial F(z)/\partial z = s(x,z)/q(r(x,z), z_0)$, and $\partial \ln F(z)/\partial z = s(x,z)/q(r(x,z), z_0)$ in the additive and multiplicative case. We thank an anonymous referee for pointing this out.

[2]Alternatively, $H$ can be identified directly using the reciprocal of $H^{-1}(r)$, where $H^{-1}(r) = \int_{r_0}^{r} dt/q(t, z_0)$ and $H^{-1}(r) = \exp \int_{r_1}^{r} dt/q(t, z_0)$ for the additive and multiplicative cases respectively. We thank an anonymous referee for pointing this out.

## 3 Estimation

In this section, for the case $r(x,z) \equiv E[Y|\, X = x, Z = z]$, we describe estimators of $M$, $G$, $F$ and $H$ based on replacing the unknown functions $r(x,z)$, $s(x,z)$ and $q(t,z)$ in (2.2) by multidimensional regression smoothers. Since an estimator of the partial derivative of the regression surface, $r(x,z)$ with respect to $z$ is needed, a natural choice of smoother will be the local polynomial estimator, which produces estimators for $r$ and $s$ simultaneously. Relative to other nonparametric regression estimators, local polynomials also have better boundary behavior and the ability to adapt to non–uniform designs, among other desirable properties; see e.g. see Fan and Gijbels (1996).

For a given random sample $\{Y_i, X_i, Z_i\}_{i=1}^n$, we propose the following steps to estimate $M$, $G$, $F$ and $H$ in the additive case:

1) Obtain consistent estimators $\widehat{r}_i = \widehat{r}(X_i, Z_i)$ and $\widehat{s}_i = \widehat{s}(X_i, Z_i)$ of $r(X_i, Z_i)$ and $s(X_i, Z_i)$ for $i = 1, \ldots, n$ by local $p_1$-th order polynomial regression of $Y_i$ on $X_i$ and $Z_i$ with corresponding kernel $K_1$, and bandwidth sequence $h_1 = h_1(n)$.

2) Obtain a consistent estimator of $q(t,z)$, given $z_0$ for all $t$, by local $p_2$-th order polynomial regression of $\widehat{s}_i$ on $\widehat{r}_i$ and $Z_i$ for $i = 1, \ldots, n$ with corresponding kernel $K_2$ and bandwidth sequence $h_2 = h_2(n)$. Denote this estimate as $\widehat{q}(t, z_0) = \widehat{E}[\, \widehat{s}(X,Z)|\widehat{r}(X,Z) = t, Z = z_0]$.

3) For a constant $r_0$, define an estimate of $M(x,z) \equiv G(x) + F(z)$ by

$$\widehat{M}(x,z) = \int_{r_0}^{\widehat{r}(x,z)} \frac{dt}{\widehat{q}(t, z_0)}. \tag{3.1}$$

4) Estimate $G(x)$ and $F(z)$ consistently up to an additive constant by marginal integration,

$$\widehat{\alpha}_{P_1}(x) = \int_{\Psi_z} \widehat{M}(x,z)\, dP_1(z), \tag{3.2}$$

$$\widehat{\alpha}_{P_2}(z) = \int_{\Psi_x} \widehat{M}(x,z)\, dP_2(x). \tag{3.3}$$

5) Now for $\widetilde{c} = (1/2)[\int_{\Psi_x} \widehat{\alpha}_{P_1}(x)\, dP_2(x) + \int_{\Psi_z} \widehat{\alpha}_{P_2}(z)\, dP_1(z)]$, define $\widetilde{G}(x) = \widehat{\alpha}_{P_1}(x) - \widetilde{c}$ , $\widetilde{F}(z) = \widehat{\alpha}_{P_2}(z) - \widetilde{c}$ and $\widetilde{M}(X_i, Z_i) \equiv \widetilde{G}(X_i) + \widetilde{F}(Z_i) + \widetilde{c}$, then we can obtain a consistent estimator of $H(m)$ by local $p_*$-th polynomial regression of $Y_i$ or $\widehat{r}(X_i, Z_i)$ on $\widetilde{M}(X_i, Z_i)$ for $i = 1, \ldots, n$ with corresponding kernel $k_*$ and bandwidth sequence $h_* = h_*(n)$. Denote this estimate as $\widehat{H}(m)$.

For estimating the alternative multiplicative $M$ model instead, replace steps 3–5 above by:

3*) For a constant $r_1$, define an estimate of $M(x,z) \equiv G(x)\, F(z)$ by

$$\widehat{M}(x,z) = \exp\left(\int_{r_1}^{\widehat{r}(x,z)} \frac{dt}{\widehat{q}(t, z_0)}\right).$$

4*) Estimate $G(x)$ and $F(z)$ consistently up to a scale factor by marginal integration,

$$\widehat{\alpha}_{P_1}(x) = \int_{\Psi_z} \widehat{M}(x,z)\, dP_1(z),$$
$$\widehat{\alpha}_{P_2}(z) = \int_{\Psi_x} \widehat{M}(x,z)\, dP_2(x).$$

5*) Now for $\widetilde{c} = (1/2)[\int_{\Psi_x} \widehat{\alpha}_{P_1}(x)\, dP_2(x) + \int_{\Psi_z} \widehat{\alpha}_{P_2}(z)\, dP_1(z)]$, define $\widetilde{G}(x) = \widehat{\alpha}_{P_1}(x)/\widetilde{c}$, $\widetilde{F}(z) = \widehat{\alpha}_{P_2}(z)/\widetilde{c}$, and $\widetilde{M}(X_i, Z_i) \equiv \widetilde{G}(X_i)\widetilde{F}(Z_i)\widetilde{c}$, then we can obtain a consistent estimator of $H(m)$ by local $p_*$-th polynomial regression of $Y_i$ or $\widehat{r}(X_i, Z_i)$ on $\widetilde{M}(X_i, Z_i)$ with corresponding kernel $k_*$ and bandwidth sequence $h_* = h_*(n)$ for $i = 1, \ldots, n$. Denote this estimate as $\widehat{H}(m)$.

We can immediately observe how important the joint–unconstrained nonparametric estimation of $r$ and $s$ is in step 1. They are not only used for estimating $q$ in step 2, but $r$ along with the preset $r_0$ ($r_1$) also define the limits of the integral in (3.1) in step 3 (3*). Operationally, because of estimation error in step 1, the function $\widehat{q}(t, z_0)$ is only observed for $t \in$range$(\widehat{r}(X_i, z_0))$, but we continue it beyond this support for step 3 (3*) using linear extrapolation, with slope equal to the derivative of $\widehat{q}$ at the corresponding end of the support (this choice of extrapolation method does not affect the resulting limiting distributions). (3.1) is then easily evaluated using numerical integration. Convenient choices of $P_1(z)$ and $P_2(x)$, in (3.2) and (3.3), are $F_z(z)$ and $F_x(x)$, which are the distribution functions of $Z$ and $X$ respectively. We can replace them by their empirical analogs, $\widehat{F}_z(z)$ and $\widehat{F}_x(x)$, yielding $\widehat{\alpha}_1(x) \equiv n^{-1}\sum_{i=1}^{n} \widehat{M}(x, Z_i)$ and $\widehat{\alpha}_2(z) = n^{-1}\sum_{i=1}^{n} \widehat{M}(X_i, z)$. Finally, notice that $\widehat{H}$ in step 5 (5*) involves a simple univariate nonparametric regression.

# 4 Asymptotic Properties

This section gives assumptions under which our theorems provide the pointwise distribution of our estimators of $M$, $G$, $F$ and $H$ for some $z = z_0$ and $r = r_0$. This is done for the additive case in conditional mean function estimation as described in the previous section. The technical issues involving the distribution of $M$ and $H$ are those of generated regressors, see Ahn (1995), Ahn (1997), Su and Ullah (2006), Lewbel and Linton (2007) and Su and Ullah (2008) for example. The proofs also use techniques to deal with nonparametrically generated dependent variables, which may be of use elsewhere. Once the asymptotic expansion of $\widehat{M}$ is established, the asymptotic properties of $G$ and $F$ will follow from standard marginal integration results.

Assumption E:

(E1) The kernels $K_l$, $l = 1, 2$, satisfy $K_1 = \Pi_{j=1}^{d+1} k_1(w_j)$, $K_2 = \Pi_{j=1}^{2} k_2(v_j)$, and $k_l$, $l = 1, 2$, are bounded, symmetric about zero, with compact support $[-c_l, c_l]$ and satisfy the property that $\int_{\Re} k_l(u)\, du = 1$. For $l = 1$ and 2, the functions $H_{l\mathbf{j}} = w^{\mathbf{j}} K_l(u)$ for all $\mathbf{j}$ with $0 \le |\mathbf{j}| \le 2p_l + 1$ are Lipschitz continuous. The matrices $\mathbf{M}_r$ and $\mathbf{M}_q$, multivariate moments of the kernels $K_1$ and $K_2$ respectively (defined in the appendix), are nonsingular.

(E2) The densities $f_W$ of $W_i$, and $f_V$ of $V_i$ for $W_i^\top \equiv \left(X_i^\top, Z_i\right)$ and $V_i \equiv (r_i, Z_i)$ respectively are uniformly bounded, and they are also bounded away from zero on their compact support.

(E3) For some $\xi > 5/2$, $E[|\varepsilon_{r,i}|^\xi] < \infty$, $E[|\varepsilon_{q,i}|^\xi] < \infty$, and $E[|\varepsilon_{r,i}\varepsilon_{q,i}|^\xi] < \infty$ where $\varepsilon_{r,i} = Y_i - r\left(X_i, Z_i\right)$ and $\varepsilon_{q,i} = S_i - q\left(r_i, Z_i\right)$. Also, $E[\varepsilon_{r,i}^2|X_i = x, Z_i = z] \equiv \sigma_r^2\left(x, z\right)$, be such that $\nu_{P_1}\left(x\right) \equiv \int p_1^2\left(z\right) \sigma_r^2\left(x, z\right) f_W^{-1}\left(x, z\right) q^{-2}\left(r, z_0\right) dz < \infty$ and $\nu_{P_2}\left(z\right) \equiv \int p_2^2\left(x\right) \sigma_r^2\left(x, z\right) f_W^{-1}\left(x, z\right) q^{-2}\left(r, z_0\right) dx < \infty$.

(E4) The function $r\left(\cdot\right)$ is $(p_1 + 1)$ times partially continuously differentiable and the function $q\left(\cdot\right)$ is $(p_2 + 1)$ times partially continuously differentiable. The corresponding $(p_1 + 1)$-th or $(p_2 + 1)$-th order partial derivatives are Lipschitz continuous on their compact support.

(E5) The bandwidth sequences $h_1$, and $h_2$ go to zero as $n \to \infty$, and satisfy the following conditions:

(i) $nh_1^{d+1}h_2^{2(p_2+1)} \to c \in [0, \infty)$,

(ii) $n^{1/2}h_1^{d+1}h_2^2/\ln n \to \infty$, $n^{1/2}h_1^{2(p_1+1)}h_2^{-2} \to 0$,

(iii) $nh_1^{d+1}h_1^{2(p_1+1)} \to c \in [0, \infty)$, and $nh_1^{d+1}h_1^{2p_1}h_2^2 \to c \in [0, \infty)$.

Assumptions (E1)–(E4) provide the regularity conditions needed for the existence of an asymptotic distribution. The estimation error $\varepsilon_{q,i}$, in Assumption (E3), is such that $E[\varepsilon_{q,i}|\, r\left(X_i, z\right) = r, Z_i = z] = 0$. However, $E\left[\varepsilon_{q,i}|\, X_i = x, Z_i = z\right] \neq 0$, so we write $\varepsilon_{q,i} = g_q\left(x, z\right) + \eta_i$, where $E\left[\eta_i|\, X_i = x, Z_i = z\right] = 0$ by construction. Assumption (E4) ensures Taylor–series expansions to appropriate orders.

Let $\nu_{1n} = n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n} + h_1^{p_1+1}$ and $\nu_{2n} = n^{-1/2}h_2^{-1}\sqrt{\ln n} + h_2^{p_2+1}$, then by Theorem 6 (page 593) in Masry (1996a), $\max_{1\leq j\leq n} \parallel \widehat{r}\left(W_j\right) - r\left(W_j\right) \parallel = O_p\left(\nu_{1n}\right)$, $\max_{1\leq j\leq n} \parallel \widehat{s}\left(W_j\right) - s\left(W_j\right) \parallel = O_p\left(h_1^{-1}\nu_{1n}\right)$ and $\sup_v \parallel \widehat{q}\left(v\right) - q\left(v\right) \parallel = O_p\left(\nu_{2n}\right)$ if the unobserved $\{V_1, \ldots, V_n\}$ were used in constructing $\widehat{q}$. Because $\{\widehat{V}_1, \ldots, \widehat{V}_n\}$ were used instead, the approximation error is accounted for in Assumption (E5)(ii), which implies that $(h_2^{-1}\nu_{1n})^2 = o(n^{-1/2}h_2^{-1})$ and so $h_2^{-1}\nu_{1n} = o\left(1\right)$, where the appearance of $h_2^{-1}$ is because of the use of Taylor–series expansions in our proofs. Assumption (E5) permits various choices of bandwidths for given polynomial orders. For example, if $p_1 = p_2 = 3$, we could set $h_1 \propto n^{-1/9}$, and $h_2 = bb \times h_1$ when $d = 1$, for a nonzero scalar $bb$, as in our Monte Carlo experiment in Section 5. More generally, in view of Assumption (E5)(iii), $h_1 \propto n^{-1/[2(p_1+1)+d]}$ and $h_2 \propto n^{-1/[2p_2+3]}$ will work for a variety of combinations of $d$, $p_1$, and $p_2$.

**Theorem 4.1** *Suppose that Assumption I holds. Then, under Assumption E, there exists a sequence of bounded continuous function $\mathcal{B}_n\left(x, z\right)$ with $\mathcal{B}_n\left(x, z\right) \to 0$ such that*

$$\sqrt{nh_1^{d+1}}\left(\widehat{M}\left(x, z\right) - M\left(x, z\right) + \mathcal{B}_n\left(x, z\right)\right) \xrightarrow{d} N\left[0, \frac{\sigma_r^2\left(x, z\right)}{q^2\left(r, z_0\right) f_W\left(x, z\right)}\left[\mathbf{M}_r^{-1}\Gamma_r\mathbf{M}_r^{-1}\right]_{0,0}\right],$$

*where $\left[\mathbf{A}\right]_{0,0}$ means the upper-left element of matrix $\mathbf{A}$.*

**Proof.** The proof of this theorem, along with definitions of each component, is given in the appendix. ■

As defined in the appendix, there are four components to the bias term $\mathcal{B}_n$, specifically, $\mathcal{B}_n(x,z) = h_1^{p_1+1}\mathcal{B}_1(x,z) + h_1^{p_1}h_2\mathcal{B}_2(x,z) + h_2^{p_2+1}\mathcal{B}_3(x,z) - h_1^{p_1+1}\mathcal{B}_4(x,z)$, where $\mathcal{B}_3$ corresponds to the ordinary nonparametric bias of $\widehat{q}$ if the unobserved $r$ and $s$ were used instead in step 2, and $\mathcal{B}_4$ corresponds to the standard nonparametric bias while calculating $\widehat{r}$ in step 1 weighted by $q^{-1}(r,z_0)$. $\mathcal{B}_1$ and $\mathcal{B}_2$ arise from estimation error in the generated regressor $\widehat{r}$ and the generated response $\widehat{s}$ in constructing $\widehat{q}$ in step 2.

Given this result, $E[\widehat{M}(x,z)] - M(x,z) = O(h_1^{p_1+1}) + O(h_1^{p_1}h_2) + O(h_2^{p_2+1})$ and $\text{var}[\widehat{M}(x,z)] = O(n^{-1}h_1^{-(d+1)})$, and these orders of magnitude also hold at boundary points by virtue of using local polynomial regression in each step. By employing generic marginal integration of this preliminary smoother, as described in step 4, we obtain by straightforward calculation the following result:

**Corollary 4.1** *Suppose that Assumption I holds. Then, under Assumption E*

$$\sqrt{nh_1^d}\left(\widehat{\alpha}_{P_1}(x) - \alpha_{P_1}(x) + \int \mathcal{B}_n(x,z)\,dP_1(z)\right) \xrightarrow{d} N\left[0, \nu_{P_1}(x)\left[\mathbf{M}_r^{-1}\Gamma_r^1\mathbf{M}_r^{-1}\right]_{0,0}\right], \quad (4.1)$$

$$\sqrt{nh_1}\left(\widehat{\alpha}_{P_2}(z) - \alpha_{P_2}(z) + \int \mathcal{B}_n(x,z)\,dP_2(x)\right) \xrightarrow{d} N\left[0, \nu_{P_2}(z)\left[\mathbf{M}_r^{-1}\Gamma_r^2\mathbf{M}_r^{-1}\right]_{0,0}\right]. \quad (4.2)$$

*where* $[\mathbf{A}]_{0,0}$ *means the upper-left element of matrix* $\mathbf{A}$.

**Proof.** Given the asymptotic normality of $\widehat{M}$, the proof follows immediately from results in Linton and Nielsen (1995) and Linton and Härdle (1996), and therefore is omitted. ■

Our procedure is similar to many other kernel-based multi-stage nonparametric procedures in that the first estimation step does not contribute to the asymptotic variance of the final stage estimators, see, e.g. Linton (2000) and Xiao, Linton, Carroll, and Mammen (2003). However, the asymptotic variances of $\widehat{M}(x,z)$, $\widehat{\alpha}_{P_1}(x)$ and $\widehat{\alpha}_{P_2}(z)$ reflect the lack of knowledge of the link function $H$ through the appearance of the function $q$ in the denominator, which by Assumption I is bounded away from zero and depends on the scale normalization $z_0$, and on the conditional variance $\sigma_r^2(x,z)$ of $Y$. These quantities can be consistently estimated from the estimates of $r(x,z_0)$, $q(r,z_0)$ in steps 1 and 2, and $\sigma_r^2(x,z)$. For example, if $P_i$, $l = 1,2$, are empirical distribution functions, the standard errors of $\widehat{\alpha}_{P_1}(X_i)$ and $\widehat{\alpha}_{P_2}(Z_i)$ can be computed as

$$\psi^1(k_1)\,\widehat{\sigma}_r^2 n^{-1}\sum_{j=1}^n \left[\widehat{f}_W(X_i,Z_j)\,\widehat{q}^2(r(X_i,Z_j),z_0)\right]^{-1}\widehat{f}_Z(Z_j)\text{, and}$$

$$\psi^2(k_1)\,\widehat{\sigma}_r^2 n^{-1}\sum_{j=1}^n \left[\widehat{f}_W(X_j,Z_i)\,\widehat{q}^2(r(X_j,Z_i),z_0)\right]^{-1}\widehat{f}_X(X_j)$$

respectively, in which $\psi^l(k_1) \equiv \left[\mathbf{M}_r^{-1}\Gamma_r^l\mathbf{M}_r^{-1}\right]_{0,0}$ for $l = 1,2$, $\widehat{f}_W$, $\widehat{f}_X$ and $\widehat{f}_Z$ are the corresponding kernel estimates of $f_W$, $f_X$ and $f_Z$, while $\widehat{\sigma}_r^2 = n^{-1}\sum_{i=1}^n[Y_i - \widehat{r}(X_i,Z_i)]^2$ or $\widehat{\sigma}_r^2 = n^{-1}\sum_{i=1}^n[Y_i - \widehat{H}(\widetilde{M}(X_i,Z_i))]^2$.

Our estimators are based on marginal integration of a function of a preliminary $(d+1)$-dimensional nonparametric estimator, hence the smoothness of $G$ and $F$ we require must increase as the dimension of $X$ increases to achieve the rate $n^{-p_1/(2p_1+1)}$, which is the optimal rate of convergence when $G$ and $F$ have $p_1$ continuous derivatives, see e.g. Stone (1985) and Stone (1986).

Now consider $H$. Define $\Psi_{M(x,z)} = \{m : m = G(x) + F(z), (x,z) \in \Psi_x \times \Psi_z\}$. If $G$ and $F$ were known, $H$ could be estimated consistently by a local $p^*$–polynomial mean regression of $Y$ on $M(X,Z) \equiv G(X) + F(Z)$. Otherwise, $H$ can be estimated with unknown $M$ by replacing $G(X_i)$ and $F(Z_i)$ with estimators in the expression for $M(X_i, Z_i)$. This is a classic generated regressors problem as in Ahn (1995). Denote these by $\widehat{\alpha}_{P_1}(X_i)$ and $\widehat{\alpha}_{P_2}(Z_i)$, with $\widetilde{M}_i \equiv \widehat{\alpha}_{P_1}(X_i) + \widehat{\alpha}_{P_2}(Z_i) - \widetilde{c}$ and $M_i \equiv \alpha_{P_1}(X_i) + \alpha_{P_2}(Z_i) - c$. Let $h_\dagger = \max(h_1^{p_1+1}, h_2^{p_2+1}, h_1^{p_1} h_2)$, then $\max_{1 \le j \le n} \| \widetilde{M}_j - M_j \| = O_p(\nu_{\dagger n})$, where $\nu_{\dagger n} = n^{-1/2} h_1^{-d/2} \sqrt{\ln n} + h_\dagger$.

To obtain the limiting distribution of $\widehat{H}$, we make the following additional assumption:

Assumption F:

(F1) The kernel $k_*$ is bounded, symmetric about zero, with compact support $[-c_*, c_*]$ and satisfies the property that $\int_{\Re} k_*(u)\, du = 1$. The functions $H_{*j} = u^j k_*(u)$ for all $j$ with $0 \le j \le 2p_* + 1$ are Lipschitz continuous. The matrix $\mathbf{M}_H$, defined in the appendix, is nonsingular.

(F2) Let $f_M$ be the density of $M(X,Z)$, which is assumed to exist, to inherit the smoothness properties of $M$ and $f_W$ and to be bounded away from zero on its compact support.

(F3) The bandwidth sequence $h_*$ goes to zero as $n \to \infty$, and satisfies the following conditions:

(i) $n h_*^{2(p_*+1)+1} \to c \in [0, \infty)$, $n h_* h_\dagger^2 \to c \in [0, \infty)$,

(ii) $n^{1/2} h_1^d h_*^{3/2} / \ln n \to \infty$, and $n^{1/2} h_\dagger^2 h_*^{-3/2} \to 0$.

Assumptions (F1) to (F3) are similar to those in Assumption E. As before, Assumption (F3)(ii) implies that $\left(h_*^{-1} \nu_{\dagger n}\right)^2 = o(n^{-1/2} h_*^{-1/2})$ and also that $h_*^{-1} \nu_{\dagger n} = o(1)$. Assumption (F3) imposes restrictions on the rate at which $h_* \to 0$ as $n \to \infty$. They ensure that no contributions to the asymptotic variance of $\widehat{H}$ are made by previous estimation stages. Let $\sigma_H^2(m) = E\left[\varepsilon_r^2 \middle| M(X,Z) = m\right]$, then we have the following theorem:

**Theorem 4.2** *Suppose that Assumption I holds, then, under Assumption E and F, there exists a sequence of bounded continuous functions $\mathcal{B}_{nH}(\cdot)$, such that $\mathcal{B}_{nH}(m) \to 0$ and*

$$\sqrt{n h_*} \left( \widehat{H}(m) - H(m) - \mathcal{B}_{nH}(m) \right) \xrightarrow{d} N\left(0, \frac{\sigma_H^2(m)}{f_M(m)} \left[\mathbf{M}_H^{-1} \Gamma_H \mathbf{M}_H^{-1}\right]_{0,0}\right),$$

*for $m \in \Psi_{M(x,z)}$, where $[\mathbf{A}]_{0,0}$ means the upper-left element of matrix $\mathbf{A}$.*

**Proof.** The proof of this theorem, along with definitions of each component, is given in the appendix. ■

When $p_* = 1$, $h_*$ admits the rate $n^{-1/5}$ when $h_1$ and $h_2$ are chosen as suggested above when $d = 1$, as it is done in the application and simulations in Section 5. In which case, $\mathcal{B}_{nH}(\cdot)$ simplifies to the standard bias from a univariate local linear regression. Standard errors can be easily computed from the formula above. By evaluating $\widehat{H}$ at each data point, the implied estimator of $\widehat{r}(X_i, Z_i) = \widehat{H}[\widetilde{M}(X_i, Z_i)]$ is $O_p(n^{-1/2}h_1^{-(d-1)/2})$, for large $h_1$ and $d$, which can be seen by a straightforward local Taylor–series expansion around $M(X_i, Z_i)$. That is, our proposed methodology has effectively reduced the curse of dimensionality in estimating $r$ by 1 with respect to its fully unrestricted nonparametric counterpart.

# 5 Numerical Results

## 5.1 Generalized Homothetic Production Function Estimation

Let $y$ be the log output of a firm and $(\widetilde{x}, z)$ be a vector of inputs. Many parametric production function models of the form $y = r^*(\widetilde{x}, z) + \varepsilon_{r^*}$ have been estimated that impose either linear homogeneity (constant returns to scale) or homotheticity for the function $r^*$. In the homogenous case, corresponding to known $H(m) = m$, examples include Bairam (1994) and Chung (1994) for parametric models, and Tripathi and Kim (2003) and Tripathi (1998) for fully nonparametric settings. In the nonparametric framework, Lewbel and Linton (2007) presents an estimator for a homothetically separable function $r^*$.

Consider the following generalization of homogeneous and homothetic functions:

**Definition 5.1** *A function $M^* : \Psi_w \subset \Re^{d+1} \to \Re$ is said to be generalized homogeneous on $\Psi_w$ if and only if the equation $M^*(\lambda w) = g(\lambda) M^*(w)$ holds for all $(\lambda, w) \in \Re_{++} \times \Psi_w$ such that $\lambda w \in \Psi_w$. The function $g : \Re_{++} \to \Re_{++}$ is such that $g(1) = 1$ and $\partial g(\lambda)/\partial \lambda > 0$ for all $\lambda$.*

**Definition 5.2** *A function $r^* : \Psi_w \subset \Re^{d+1} \to \Re$ is said to be generalized homothetic on $\Psi_w$ if and only if $r^*(w) = H[M^*(w)]$, where $H : \Re \to \Re$ is a strictly monotonic function and $M^*$ is generalized homogeneous on $\Psi_w$.*

Homogeneity of any degree $\kappa$ and homotheticity are the special cases of definitions 5.1 and 5.2, respectively, in which the function $g$ takes the functional form $g(\lambda) = \lambda^\kappa$. Given a generalized homothetic production function we have

$$\begin{aligned} r^*(\widetilde{x}, z) &= H[M^*(\widetilde{x}, z)] = H\left[M^*(\widetilde{x}/z, 1)\, g(1/z)^{-1}\right] \\ &= H[G(x) F(z)] = H[M(x, z)] \equiv r(x, z), \end{aligned} \tag{5.1}$$

where $x = \widetilde{x}/z$ and $F(z) = 1/g(1/z)$.

We have constructed an `R` package, `JLLprod`, which can be freely downloaded from the first author's websites. The package includes access to an Ecuadorian production data set and to the Chinese data set described below.[3] We use this software to estimate generalized homothetic production functions for four industries in mainland China in 2001. For each firm in each industry we observe the net value of real fixed assets $K$, the number of employees $L$, and $Y$ defined as the log of value-added real output. $K$ and $Y$ are measured in thousands of Yuan converted to the base year 2000 using a general price

[3]This is data on 406 firms in the Petroleum, Chemical and Plastics industries in Ecuador in 2002, see Huynh and Jacho-Chávez (2007) for details.

deflator for the Chinese economy. For details regarding the collection and construction of this data set, see Jefferson, Hu, Guan, and Yu (2003).

We consider both nonparametric and parametric estimates of the production function $r(k, L) \in \mathcal{P}$, where $k = K/L$ and $\mathcal{P}$ is a set of smooth production functions, so in (5.1), $z = L$, $\widetilde{x} = K$, and $x = k$. To eliminate extreme outliers (which in some cases are likely due to gross measurement errors in the data) we sort the data by $k$ and remove the top and bottom 2.5% of observations in each industry. Both regressors are also normalized by their respective median prior to estimation.

### 5.1.1 Parametric Modeling

Consider a general production function (P1) in which log output $Y = r_{\psi_{P1}}(k, L) + \varepsilon_r$, where $r_{\psi_{P1}}$ is an unrestricted quadratic function in $\ln(k)$ and $\ln(L + \gamma)$, so

$$\begin{aligned} r_{\psi_{P1}}(k, L) = \theta_0 + \theta_1 \ln(k) + \theta_2 \ln(L + \gamma) + \theta_3 [\ln(k)]^2 \\ + \theta_4 \ln(k) \ln(L + \gamma) + \theta_5 [\ln(L + \gamma)]^2, \end{aligned} \tag{5.2}$$

and $\psi_{P1} = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \gamma)^\top$. When $\gamma = 0$, (5.2) reduces to the ordinary unrestricted Translog production function. When $2\theta_1\theta_5 - \theta_2\theta_4 = 0$ and $\theta_1^2\theta_5 - \theta_2^2\theta_3 = 0$, (5.2) is equivalent to the following generalized homothetic production function (P2) specification,

$$\begin{aligned} M(k, L) &= k^\alpha (L + \gamma) \\ r_{\psi_{P2}}(k, L) &= H(M) = \beta_0 + \beta_1 \ln(M) + \beta_2 [\ln(M)]^2, \end{aligned} \tag{5.3}$$

where $\psi_{P2} = (\alpha, \beta_0, \beta_1, \beta_2, \gamma)^\top$. If we impose both (P2) and $\gamma = 0$, then the model reduces to

$$\begin{aligned} M(k, L) &= k^\alpha L \\ r_{\psi_{P3}}(k, L) &= H(M) = \beta_0 + \beta_1 \ln(M) + \beta_2 [\ln(M)]^2, \end{aligned} \tag{5.4}$$

where $\psi_{P3} = (\alpha, \beta_0, \beta_1, \beta_2)^\top$, which is the homothetic Translog production function proposed by Christensen, Jorgenson, and Lau (1973).

Models (P1), (P2), and (P3) are fitted to the data using nonlinear least squares estimation. The implied parametric estimates of $G$, $F$ and $H$ are shown in Figures 1 and 2. Various specification tests justify the use of these models as sensible parametric simplifications for our data, see Jacho-Chávez, Lewbel, and Linton (2006) for details.

### 5.1.2 Nonparametric Modeling

Figures 1 and 2 also display generalized homothetic nonparametric estimates $\widehat{M}(k, L)$, $\widehat{G}(k)$, $\widehat{F}(L)$ and $\widehat{H}(M)$. For each industry, we use local quadratic regression with a Gaussian kernel and bandwidths $h_1$ obtained by a standard unrestricted leave–one–out cross validation method for regression functions. In the second stage, we set bandwidth $h_2$ to be the same in local linear regressions across industries and time. We also choose the location and scale normalizations to obtain estimated surfaces $\widehat{M}$ with approximately the same range, yielding the following normalizations:

| Industry | 2001 | | |
|---|---|---|---|
| | $n$ | $\ln L_0$ | $r_0$ |
| Chemical | 1637 | 3.06 | 7.0 |
| Iron | 341 | 4.06 | 8.0 |
| Petroleum | 119 | 2.27 | 8.5 |
| Transportation | 1230 | 4.04 | 7.5 |

The nonparametric fits of the generalized homogeneous component, $\widehat{M}$, shown in Figure 1, are quite similar. They are both increasing in $k$ and $L$ with ranges varying more with labor than with respect to capital to labor ratios, as we would expect.[4] Nonparametric estimates of the functions $G$ and $F$ differ from the parametric Translog model estimates (P3) in Figures 2, but they are roughly similar to the parametric generalized homothetic model (P2) at low levels of $L$.[5] The nonparametric estimates are all strictly increasing in their arguments, but show quite a bit more curvature, departing most markedly from the parametric models for $G$ in most industries. Comparing the nonparametric estimator of $F$ in 2 with the parametric estimates also provides a quick check for the presence of homotheticity in the data set. If homotheticity were present, i.e. $F(L) = L$, all curves would be close to each other, as happens for the petroleum and transportation industries. In any case, they are all strictly increasing functions in labor, implying a generalized homogeneous structure for $M$ as conjectured. Figure 1 shows parametric and nonparametric fits of the unknown link function $H$, obtained by a local linear regression of $\widehat{r}$ on $\widehat{M}$ with a normal kernel and bandwidth $h_*$ given by Silverman's rule of thumb. It also shows fits from the unconstrained estimator of the function $r$ used in the construction of our estimator in the first stage for each $(k, L)$ for which $\widehat{M}$ was calculated. The nonparametric fits of $r$ and those of $H$ are quite similar in all industries, indicating that the imposition of generalized homotheticity is reasonable for these industries. The parametric fits are also broadly similar to the nonparametric ones, but showing more curvature for the chemical and iron industries. However, these similarities do not always translate into comparable measures of substitutability and returns to scale, see Jacho-Chávez, Lewbel, and Linton (2006) for details.

## 5.2 Simulations

In this section, we describe Monte Carlo experiments to study the finite sample properties of the proposed estimator, and compare its performance with that of direct competitors when the link function is known. Code for these simulations was written in `GAUSS`. Our experimental designs are based on the parametric production function models in Section 5.1. In particular, $n$ observations $\{Y_i, X_i, Z_i\}_{i=1}^n$ are generated from $Y = r_{\psi_{P2}}(X, Z) + \sigma_r \cdot \varepsilon_r$, where the distributions of $X$ and $Z$ are $U[1, 2]$, $\varepsilon_r$ is chosen independently of $X$ and $Z$ with a standard normal distribution, and $r_{\psi_{P2}}(X, Z)$ is given by (5.3). We

[4]It was a similar observation by Cobb and Douglas (1928) that motivated the use of homogeneous functions in production theory, see Douglas (1967).

[5]The means of the observed ranges were subtracted from both sets of curves before plotting.

consider two designs:

$$\text{Design 1:} \qquad \psi_{P2} = (3/2, 0, 1, 0, 0)^{\top},$$
$$\text{Design 2:} \qquad \psi_{P2} = (3/2, 3, 3/4, 0, -1/2)^{\top},$$

and two possible scenarios, $\sigma_r^2 = 1$ and $\sigma_r^2 = 2$. Notice that $z_0 = 3/2$, and $r_0 = 0$, and $r_0 = 3$ provide the required normalizations in designs 1 and 2 respectively.

In constructing our estimators $\widehat{M}$, $\widehat{G}$, $\widehat{F}$ and $\widehat{H}$, we use the second order Gaussian kernel $k_i(u) = \left(1/\sqrt{2\pi}\right)\exp\left(-u^2/2\right)$, $i = 1, 2, *$. The integral in $\widehat{M}$ in step 2 of section 3, is evaluated numerically using the trapezoid method. We also fix $p_1 = 3$, $p_2 = 1$ and $p_* = 1$. We use the bandwidth $h_1 = \widetilde{k}\widehat{s}_W n^{-1/9}$, where $\widetilde{k}$ is a constant term and $\widehat{s}_W$ is the square root of the average of the sample variances of $X_i$ and $Z_i$. This bandwidth is proportional to the optimal rate for 3rd order local polynomial estimation in the first stage, and for simplicity $h_2$ is fixed as $3h_1$. The bandwidth $h_*$ is set to follow Silverman's rule of thumb ($1.06n^{-1/5}$ times the squared root of the average of the regressors variances). Three different choices of $\widetilde{k}$ are considered: $\widetilde{k} \in \{0.5, 1, 1.5\}$.

We compare the performance of our estimator to those of Linton and Nielsen (1995) in Design 1, and Linton and Härdle (1996) in Design 2. These alternative estimators may not be fully efficient, but they assume the link function $H$ is known, and so they provide strong benchmarks for comparison with our estimator where $H$ is unknown.

Each function is estimated on a $50 \times 50$ equally spaced grid in $[1, 2] \times [1, 2]$ when $n = 150$, and at another $60 \times 60$ uniform grid on the same domain when $n = 600$. Two criteria summarizing goodness of fit, the Integrated Root Mean Squared Error (IRMSE) and Integrated Mean Absolute Error (IMAE), are calculated at all grid points and then averaged. Tables 1 and 2 report the median of these averages over 2000 replications for each design, scenario and bandwidth. For comparison, we also report the results obtained using the estimators proposed by Linton and Nielsen (1995), and Linton and Härdle (1996) in column (1) in Tables 1, and 2 respectively. These were constructed using the same unrestricted first stage nonparametric regression $r$ employed by our estimator.

As seen in the Tables, for either sample size, lack of knowledge of the link function increases the fitting error of our estimator relative to estimates using that knowledge. For each scenario, the IRMSE and IMAE decline as the sample size is quadrupled for both sets of estimators, at somewhat similar, less than $\sqrt{n}$-rates. Larger bandwidths produce superior estimates for all functional components in all designs. In the estimation of the additive components, $G$ and $F$, the fitting criteria of Linton and Nielsen (1995) and Linton and Härdle (1996) are of approximately the same magnitude. There does not seem to be a dramatic difference in estimates of $H$ between estimators in both designs. All sets of estimates deteriorate as expected when $\sigma_r$ is increased.

# 6 Conclusion

We provide a general nonparametric estimator for a transformed partly additive or multiplicatively separable model of a regression function. Its small sample properties were analyzed in some Monte Carlo experiments, and found to compare favorably with respect to other estimators. We have shown that many popular empirical models implied by economic theory share this partly separable structure.

We empirically applied our model to estimate generalized homothetic production functions. Possible extensions include the identification and estimation of 1.1 with additional regressors; the possibility of endogenous regressors; and a test for homotheticity, see Jacho-Chávez, Lewbel, and Linton (2006) for more details.

## References


Ahn, H. (1995): "Nonparametric Two-stage Estimation of Conditional Choice Probabilities in a Binary Choice Model under Uncertainty," *Journal of Econometrics*, 67, 337–378.

——— (1997): "Semiparametric Estimation of a Single-index Model with Nonparametric Generated Regressors," *Econometric Theory*, 13(1), 3–31.

Andrews, D. W. (1991): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," *Econometrica*, 59, 307–345.

Bairam, E. I. (1994): *Homogeneous and Nonhomogeneous Production Functions*. Avebury, Vermont, 1 edn.

Breiman, L., and J. H. Friedman (1985): "Estimating Optimal Transformations for Multiple Regression and Correlation," *Journal of the American Statistical Association*, 80, 580–598.

Chiang, A. (1984): *Fundamental Methods of Mathematical Economics*. McGraw-Hill, New York, 3 edn.

Christensen, L. R., D. W. Jorgenson, and L. J. Lau (1973): "Transcendental Logarithmic Production Frontiers," *The Review of Economics and Statistics*, 55, 28–45.

Chung, J. W. (1994): *Utility and Production Functions*. Blackwell Publishers, Massachusetts, 1 edn.

Cobb, C. W., and P. H. Douglas (1928): "Theory of Production," *American Economic Review*, 18, 139–165.

Douglas, P. H. (1967): "Comments on the Cobb-Douglas Production Function," in *The Theory and Empirical Analysis of Production*, ed. by M. Brown, pp. 15–22. Columbia University Press, New York.

Ekeland, I., J. J. Heckman, and L. Nesheim (2004): "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112(1), S60–S109.

Fan, J., and I. Gijbels (1996): *Local Polynomial Modelling and Its Applications*, vol. 66 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, 1 edn.

Friedman, J. H., and W. Stutzle (1981): "Projection Pursuit Regression," *Journal of The American Statistical Association*, 76, 817–823.

Goldman, S. M., and H. Uzawa (1964): "A Note on Separability in Demand Analysis," *Econometrica*, 32(3), 387–398.

Horowitz, J. L. (2001): "Nonparametric Estimation of a Generalized Additive Model with an Unknown Link Function," *Econometrica*, 69, 499–513.

Horowitz, J. L., and E. Mammen (2004): "Nonparametric Estimation of an additive Model with a Link Function," *Annals of Statistics*, 36(2), 2412–2443.

Huynh, K., and D. Jacho-Chávez (2007): "Conditional density estimation: an application to the Ecuadorian manufacturing sector," *Economics Bulletin*, 3(62), 1–6.

Ihaka, R., and R. Gentleman (1996): "R: A Language for Data Analysis and Graphics," *Journal of Computational and Graphical Statistics*, 5(3), 299–314.

Jacho-Chávez, D. T., A. Lewbel, and O. B. Linton (2006): "Identification and Nonparametric Estimation of a Transformed Additively Separable Model," Sticerd Working Paper EM/2006/505, STICERD.

Jefferson, G., A. G. Z. Hu, X. Guan, and X. Yu (2003): "Ownership, Performance, and Innovation in China's Large- and Medium-size Industrial Enterprise Sector," *China Economic Review*, 14(1), 89–113.

Lewbel, A., and O. B. Linton (2007): "Nonparametric Matching and Efficient Estimators of Homothetically Separable Functions," *Econometrica*, 75(4), 1209–1227.

Lewbel, A., O. B. Linton, and D. McFadden (2002): "Estimating Features of a Distribution from Binomial Data," Unpublished Manuscript.

Linton, O. B. (2000): "Efficient Estimation of Generalized Additive Nonparametric Regression Models," *Econometric Theory*, 16(4), 502–523.

Linton, O. B., and W. Härdle (1996): "Estimating Additive Regression with Known Links," *Biometrika*, 83, 529–540.

Linton, O. B., and J. P. Nielsen (1995): "A Kernel Model of Estimating Structured Nonparametric Regression Based on Marginal Integration," *Biometrika*, 82, 93–100.

Masry, E. (1996a): "Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates," *Journal of Time Series Analysis*, 17(6), 571–599.

——— (1996b): "Multivariate Regression Estimation Local Polynomial Fitting for Time Series," *Stochastic Processes and their Application*, 65, 81–101.

Nelsen, R. B. (2006): *An Introduction to Copulas*, Springer Series in Statistics. Springer-Verlag New York, Inc, New York, 2 edn.

Pinkse, J. (2001): "Nonparametric Regression Estimation using Weak Separability," Unpublished manuscript.

Simon, C. P., and L. E. Blume (1994): *Mathematics for Economists*. W.W. Norton and Company Ltd, USA, 1 edn.

Stone, C. J. (1980): "Optimal Rates of Convergence for Nonparametric Estimators," *The Annals of Statistics*, 8(8), 1348–1360.

——— (1985): "Additive Regression and other Nonparametric Models," *Annals of Statistics*, 13(2), 689–705.

——— (1986): "The Dimensionality Reduction Principle for Generalized Additive Models," *Annals of Statistics*, 14(2), 590–606.

Su, L., and A. Ullah (2006): "More Efficient Estimation in Nonparametric Regression with Nonparametric Autocorrelated Errors," *Econometric Theory*, 22(1), 98–126.

——— (2008): "Local Polynomial Estimation of Nonparametric Simultaneous Equation Models," forthcoming in *Journal of Econometrics*.

Tjøstheim, D., and B. H. Auestad (1994): "Nonparametric Identification of Nonlinear Time Series: Projections," *Journal of The American Statistical Association*, 89, 1398–1409.

Tripathi, G. (1998): "Nonparametric Estimation and Testing of Homogeneous Functional Forms," Unpublished Manuscript.

Tripathi, G., and W. Kim (2003): "Nonparametric Estimation of Homogeneous Functions," *Econometric Theory*, 19, 640–663.

Xiao, Z., O. B. Linton, R. J. Carroll, and E. Mammen (2003): "More Efficient Local Polynomial Estimation in Nonparametric Regression With Autocorrelated Errors," *Journal of the American Statistical Association*, 98(464), 980–992.

# Appendix A: Main Proofs

### Preliminaries

We use the notation as well as the general approach introduced by Masry (1996b). For the sample $\{Y_i, X_i, Z_i\}_{i=1}^n$, let $W_i = \left(X_i^\top, Z_i\right)^\top$ so we obtained the $p_1$–th order local polynomial regression of $Y_i$ on $W_i$ by minimizing

$$Q_{r,n}(\theta) = n^{-1} h_1^{-(d+1)} \sum_{i=1}^{n} K_1\left(\frac{W_i - w}{h_1}\right)\left[Y_i - \sum_{0 \le |\mathbf{j}| \le p_1} \theta_{\mathbf{j}} (W_i - w)^{\mathbf{j}}\right]^2, \tag{A-1}$$

where the first element in $\theta$ denotes the minimizing intercept of (A-1), $\theta_0$, and

$$\theta_{\mathbf{j}} = \frac{1}{\mathbf{j}!} \frac{\partial^{|\mathbf{j}|} r(w)}{\partial^{j_1} w_1 \cdots \partial^{j_d} w_d \partial^{j_{d+1}} w_{d+1}}.$$

We also use the following conventions:

$$\mathbf{j} = (j_1, \ldots, j_d, j_{d+1})^\top, \ \mathbf{j}! = j_1! \times \ldots \times j_d \times j_{d+1}!, \ |\mathbf{j}| = \sum_{k=1}^{d+1} j_k$$

$$a^{\mathbf{j}} = a_1^{j_1} \times \ldots \times a_d^{j_d} \times a_{d+1}^{j_{d+1}}, \quad \sum_{0 \le |\mathbf{j}| \le p_1} = \sum_{k=0}^{p_1} \sum_{j_1=0}^{k} \cdots \sum_{j_d=0}^{k} \sum_{\substack{j_{d+1}=0 \\ j_1+\ldots+j_d+j_{d+1}=k}}^{k}$$

where $w = (x^\top, z)^\top$. Let $N_{r,(l)} = (l + k - 1)! / (l! (k-1)!)$ be the number of distinct $k-$tuples $\mathbf{j}$ with $|\mathbf{j}| = l$, where $k = d + 1$. After arranging them in the corresponding lexicographical order, we let $\phi_l^{-1}$ denote this one-to-one map. For each $\mathbf{j}$ with $0 \le |\mathbf{j}| \le 2p_1$, let

$$\begin{aligned}
\mu_{\mathbf{j}}(K_1) &= \int_{\Re^{d+1}} w^{\mathbf{j}} K_1(u)\, du, \ \gamma_{\mathbf{j}}(K_1) = \int_{\Re^{d+1}} w^{\mathbf{j}} K_1^2(u)\, du, \\
\gamma^1_{\mathbf{k},\mathbf{l}}(K_1) &= \int_{\Re^d} \int_{\Re} \int_{\Re} (u_d, u_1)^{\mathbf{k}} (u_d, \widetilde{u}_1)^{\mathbf{l}} K_1(u_d, u_1) K_1(u_d, \widetilde{u}_1)\, du_1 d\widetilde{u}_1, \text{ and} \\
\gamma^2_{\mathbf{k},\mathbf{l}}(K_1) &= \int_{\Re} \int_{\Re^d} \int_{\Re^d} (u_d, u_1)^{\mathbf{k}} (\widetilde{u}_d, u_1)^{\mathbf{l}} K_1(u_d, u_1) K_1(\widetilde{u}_d, u_1)\, du_d d\widetilde{u}_d,
\end{aligned}$$

where $u_d$ and $u_1$ represent the first $d$ and last element of the $d+1$ vector $u$ respectively. Define the $N_r \times N_r$ dimensional matrices $\mathbf{M}_r$ and $\mathbf{\Gamma}_r$, and the $N_r \times N_{r,(p_1+1)}$ matrix $\mathbf{B}_r$ by

$$\mathbf{M}_r = \begin{bmatrix} \mathbf{M}_{r;0,0} & \mathbf{M}_{r;0,1} & \ldots & \mathbf{M}_{r;0,p_1} \\ \mathbf{M}_{r;1,0} & \mathbf{M}_{r;1,1} & \ldots & \mathbf{M}_{r;1,p_1} \\ \vdots & \vdots & & \vdots \\ \mathbf{M}_{r;p_1,0} & \mathbf{M}_{r;p_1,1} & \ldots & \mathbf{M}_{r;p_1,p_1} \end{bmatrix},$$

$$\mathbf{\Gamma}_r = \begin{bmatrix} \mathbf{\Gamma}_{r;0,0} & \mathbf{\Gamma}_{r;0,1} & \ldots & \mathbf{\Gamma}_{r;0,p_1} \\ \mathbf{\Gamma}_{r;1,0} & \mathbf{\Gamma}_{r;1,1} & \ldots & \mathbf{\Gamma}_{r;1,p_1} \\ \vdots & \vdots & & \vdots \\ \mathbf{\Gamma}_{r;p_1,0} & \mathbf{\Gamma}_{r;p_1,1} & \ldots & \mathbf{\Gamma}_{r;p_1,p_1} \end{bmatrix}, \mathbf{B}_r = \begin{bmatrix} \mathbf{M}_{r;0,p_1+1} \\ \mathbf{M}_{r;1,p_1+1} \\ \vdots \\ \mathbf{M}_{r;p_1,p_1+1} \end{bmatrix} \tag{A-2}$$

where $N_r = \sum_{l=0}^{p_1} N_{r,(l)}$, $\mathbf{M}_{r;i,j}$ and $\mathbf{\Gamma}_{r;i,j}$ are $N_{r,(i)} \times N_{r,(j)}$ dimensional matrices whose $(l,m)$ elements are $\mu_{\phi_i(l)+\phi_j(m)}$ and $\gamma_{\phi_i(l),\phi_j(m)}$ respectively. $\mathbf{\Gamma}^1_r$ and $\mathbf{\Gamma}^2_r$ are defined similarly by the $N_{r,(i)} \times N_{r,(j)}$ matrices $\mathbf{\Gamma}^1_{r;i,j}$, $\mathbf{\Gamma}^2_{r;i,j}$, whose $(l,m)$ elements are given by $\gamma^1_{\phi_i(l),\phi_j(m)}$ and $\gamma^2_{\phi_i(l),\phi_j(m)}$ respectively. The elements of $\mathbf{M}_r = \mathbf{M}_r(K_1, p_1)$ and $\mathbf{B}_r = \mathbf{B}_r(K_1, p_1)$ are simply multivariate moments of the kernel $K_1$.

Similarly, for the generated sub-sample set $\{\widehat{s}(X_i, Z_i), \widehat{r}(X_i, Z_i), Z_i\}_{i=1}^n$, an estimator of the function $q$, defined as $q(t,z) = E[S| r(X,Z) = t, Z = z]$, is obtained by the intercept of the following minimizing problem,

$$Q_{q,n}(\theta) = n^{-1} h_2^{-2} \sum_{i=1}^{n} K_2\left(\frac{\widehat{V}_i - v}{h_2}\right)\left[\widehat{S}_i - \sum_{0 \le |\mathbf{j}| \le p_2} \theta_{\mathbf{j}}\left(\widehat{V}_i - v\right)^{\mathbf{j}}\right]^2,$$

where $\widehat{V}_i = (\widehat{r}_i, Z_i)^\top$ and $v = (t, z)^\top$, define $V_i = (r_i, Z_i)^\top$ accordingly. Let $N_{q,(l)} = (l + k - 1)!/(l! \times (k-1)!)$ be the number of distinct $k-$tuples $\mathbf{j}$ with $|\mathbf{j}| = l$, where $k = 2$. After arranging them in the corresponding lexicographical order, we let $\phi_l^{-1}$ denote this one-to-one map. For each $\mathbf{j}$ with $0 \le |\mathbf{j}| \le 2p_2$, let $\mu_{\mathbf{j}}(K_2) = \int_{\Re^2} w^{\mathbf{j}} K_2(u)\, du$, and $\gamma_{\mathbf{j}}(K_2) = \int_{\Re^2} w^{\mathbf{j}} K_2^2(u)\, du$. Define the $N_q \times N_q$ dimensional matrices $\mathbf{M}_q$ and $\mathbf{\Gamma}_q$, and the $N_q \times N_{q,(p_2+1)}$ matrix $\mathbf{B}_q$ by

$$\mathbf{M}_q = \begin{bmatrix} \mathbf{M}_{q;0,0} & \mathbf{M}_{q;0,1} & \ldots & \mathbf{M}_{q;0,p_2} \\ \mathbf{M}_{q;1,0} & \mathbf{M}_{q;1,1} & \ldots & \mathbf{M}_{q;1,p_2} \\ \vdots & \vdots & & \vdots \\ \mathbf{M}_{q;p_2,0} & \mathbf{M}_{q;p_2,1} & \ldots & \mathbf{M}_{q;p_2,p_2} \end{bmatrix},$$

$$\mathbf{\Gamma}_q = \begin{bmatrix} \mathbf{\Gamma}_{q;0,0} & \mathbf{\Gamma}_{q;0,1} & \ldots & \mathbf{\Gamma}_{q;0,p_2} \\ \mathbf{\Gamma}_{q;1,0} & \mathbf{\Gamma}_{q;1,1} & \ldots & \mathbf{\Gamma}_{q;1,p_2} \\ \vdots & \vdots & & \vdots \\ \mathbf{\Gamma}_{q;p_2,0} & \mathbf{\Gamma}_{q;p_2,1} & \ldots & \mathbf{\Gamma}_{q;p_2,p_2} \end{bmatrix}, \mathbf{B}_q = \begin{bmatrix} \mathbf{M}_{q;0,p_2+1} \\ \mathbf{M}_{q;1,p_2+1} \\ \vdots \\ \mathbf{M}_{q;p_2,p_2+1} \end{bmatrix}, \tag{A-3}$$

where $N_q = \sum_{l=0}^{p_2} N_{q,(l)}$, $\mathbf{M}_{q;j,k}$ and $\mathbf{\Gamma}_{q;j,k}$ are $N_{q,(j)} \times N_{q,(k)}$ dimensional matrices whose $(l,m)$ elements are $\mu_{\phi_{q;j}(l)+\phi_{q;k}(m)}$ and $\gamma_{\phi_{q;j}(l),\phi_{q;k}(m)}$ respectively. The elements of $\mathbf{M}_q = \mathbf{M}_q(K_2, p_2)$ and $\mathbf{B}_q = \mathbf{B}_q(K_2, p_2)$ are simply multivariate moments of the kernel $K_2$. To facilitate the proof, let $\mathcal{K}_{2,i}(v)$ be a $N_q \times 1$ vector, $\mathcal{K}_{2,i}^{(1)}(v)$ be a $N_q \times 2$ matrix, and $\mathbf{M}_{q,n}(v)$ be a symmetric $N_q \times N_q$ matrix such that

$$\mathbf{M}_{q,n}(v) = \begin{bmatrix} \mathbf{M}_{q,n;0,0}(v) & \mathbf{M}_{q,n;0,1}(v) & \ldots & \mathbf{M}_{q,n;0,p_2}(v) \\ \mathbf{M}_{q,n;1,0}(v) & \mathbf{M}_{q,n;1,1}(v) & \ldots & \mathbf{M}_{q,n;1,p_2}(v) \\ \vdots & \vdots & & \vdots \\ \mathbf{M}_{q,n;p_2,0}(v) & \mathbf{M}_{q,n;p_2,1}(v) & \ldots & \mathbf{M}_{q,n;p_2,p_2}(v) \end{bmatrix}, \tag{A-4}$$

$$\mathcal{K}_{2,i}(v) = \begin{bmatrix} \mathcal{K}_{2,i;0}(v) \\ \mathcal{K}_{2,i;1}(v) \\ \vdots \\ \mathcal{K}_{2,i;p_2}(v) \end{bmatrix}, \mathcal{K}_{2,i}^{(1)}(v) = \begin{bmatrix} \mathcal{K}_{2,i;0}^{(1)}(v) \\ \mathcal{K}_{2,i;1}^{(1)}(v) \\ \vdots \\ \mathcal{K}_{2,i;p_2}^{(1)}(v) \end{bmatrix},$$

where $\mathcal{K}_{2,i;l}(v)$ is a $N_{q,(l)} \times 1$ dimensional subvector whose $l^0-$th element is given by $[\mathcal{K}_{2,i;l}(v)]_{l^0} = ((V_i - v)/h_2)^{\phi_{q;l}(l^0)}\ K_2((V_i - v)/h_2)$. The $N_{q,(l)} \times 1$ matrix $\mathcal{K}_{2,i;l}^{(1)}(v)$ has $l^0$ element being the partial derivative of $[\mathcal{K}_{2,i;l}(t,z)]_{l^0}$ with respect to $r$, and $\mathbf{M}_{q,n;j,k}(v)$ is a $N_{q,(j)} \times N_{q,(k)}$ dimensional submatrix with the $(l, l^0)$ element given by

$$[\mathbf{M}_{q,n;j,k}(v)]_{l,l^0} = \frac{1}{nh_2^2} \sum_{i=1}^{n} \left(\frac{V_i - v}{h_2}\right)^{\phi_{q;j}(l)+\phi_{q;k}(l^0)} K_2\left(\frac{V_i - v}{h_2}\right).$$

$\widehat{\mathcal{K}}_{2,i}(v)$ and $\widehat{\mathbf{M}}_{q,n}(v)$ are defined similarly as $\mathcal{K}_{2,i}(v)$ and $\mathbf{M}_{q,n}(v)$ respectively, but with the generated regressors $\{\widehat{r}_i\}_{i=1}^n$ in place of the unobserved variables $\{r_i\}_{i=1}^n$. Let us define the functions $\widetilde{\mathcal{K}}_{2,i}(z) = \int h_2^{-1}\mathcal{K}_{2,i}(t,z)\,dt$ and $\zeta(t,z) = \partial\left[f_V(t,z)\,q^2(t,z)\right]^{-1}/\partial t$, which are well defined given Assumptions (E1) and (E2). Thus, by integration by parts, it follows that

$$\int_{r_0}^{r(x,z)} h_2^{-1}\mathcal{K}_{2,i}^{(1)}(t,z)\left[f_V(t,z)\,q^2(t,z)\right]^{-1}dt = \Big\{\mathcal{K}_{2,i}(r,z)\left[f_V(r,z)\,q^2(r,z)\right]^{-1} - \mathcal{K}_{2,i}(r_0,z)\times \left[f_V(r_0,z)\,q^2(r_0,z)\right]^{-1}\Big\} - \int_{r_0}^{r(x,z)}\mathcal{K}_{2,i}(t,z)\,\zeta(t,z)\,dt \equiv \varrho_{i,1}^0 - \varrho_{i,2}^0. \tag{A-5}$$

Similarly, let us define $dQ(t) = 1(r_0 \leq t \leq r(x,z))\,dt$, so we can write

$$\int h_2^{-1}\mathcal{K}_{2,i}(t,z)\left[f_V(t,z)\,q^2(t,z)\right]^{-1}dQ(t) \equiv \varrho_{i,1}^1 - \varrho_{i,2}^1,$$

where $\varrho_{i,1}^1$ and $\varrho_{i,2}^1$ are like $\varrho_{i,1}^0$ and $\varrho_{i,2}^0$ in (A-5), but with $\mathcal{K}_{2,i}^1(r,z)$ replacing $\mathcal{K}_{2,i}(r,z)$, where $\mathcal{K}_{2,i}^1(r,z) = \int_{-\infty}^r \mathcal{K}_{2,i}(s,z)\,ds$, a $N_q\times 1$ vector with well-defined functions as elements by virtue of Assumption (E1). Furthermore, $n^{-1}h_2^2\sum_{i=1}^n\mathcal{K}_{2,i}^1(r,z)$ converges to $\mathbf{M}_{q,0}^1 f_V(r,z)$ in mean squared, where $\mathbf{M}_{q,0}^1$ is a $N_q\times 1$ vector with $l_0$ element given by $\int u^{\phi_{q:l}(l^0)}K_2^1(u)\,du$, and $K_2^1(u) = \int_{-\infty}^u K_2(v)\,dv$. Similarly, $n^{-1}h_2^2\sum_{i=1}^n\mathcal{K}_{2,i}(r,z)$ converges in mean squared to $\mathbf{M}_{q,0}^0 f_V(r,z)$.

Let also arrange the $N_{r,(m)}$ and $N_{q,(m)}$ elements of the derivatives

$$D^{\mathbf{m}}r(w) \equiv \frac{\partial^m r(w)}{\partial^{m_1}w_1,\ldots,\partial^{m_k}w_k},\ D^{\mathbf{m}}q(v) \equiv \frac{\partial^m q(v)}{\partial^{m_1}v_1,\ldots,\partial^{m_k}v_k},\ \text{for } |\mathbf{m}| = m$$

as the $N_{r,(m)}\times 1$ and $N_{q,(m)}\times 1$ column vectors $r^{(m)}(w)$ and $q^{(m)}(v)$ in the lexicographical order mentioned above.

Let $\iota_1 = (1,0,\ldots,0)^\top \in \Re^{N_r}$ and $\iota_1^* = (0,1,0,\ldots,0)^\top \in \Re^{N_r}$, then by equation (2.13) (page 574) and Corollary 2(ii) (page 580) in Masry (1996a), we can write

$$\widehat{r}(w) - r(w) = \iota_1^\top\left[\mathbf{M}_r f(w)\right]^{-1}\{1+o_p(1)\} \times\left\{n^{-1}h_1^{-(d+1)}\sum_{j=1}^n\mathcal{K}_{1,j}(w)\left[\varepsilon_{r,j} + \sum_{|\mathbf{k}|=p_1+1}\frac{1}{\mathbf{k}!}D^{\mathbf{k}}r(w)\,(W_i-w)^{\mathbf{k}}\right] + \gamma_n(w)\right\}, \tag{A-6}$$

$$\widehat{s}(w) - s(w) = h_1^{-1}\iota_1^{*\top}\left[\mathbf{M}_r f(w)\right]^{-1}\{1+o_p(1)\} \times\left\{n^{-1}h_1^{-(d+1)}\sum_{j=1}^n\mathcal{K}_{1,j}(w)\left[\varepsilon_{r,j} + \sum_{|\mathbf{k}|=p_1+1}\frac{1}{\mathbf{k}!}D^{\mathbf{k}}r(w)\,(W_i-w)^{\mathbf{k}}\right] + \gamma_n(w)\right\} \tag{A-7}$$

uniformly in $w$, where

$$\gamma_n(w) \equiv (p_1+1)\,n^{-1}h_1^{-(d+1)}\frac{1}{\mathbf{k}!}\sum_{|\mathbf{k}|=p_1+1}\mathcal{K}_{1,j}(w)\,(W_j-w)^{\mathbf{k}} \times\int_0^1\{D^{\mathbf{k}}r(w+\tau(W_i-w)) - D^{\mathbf{k}}r(w)\}\,(1-\tau)^{p_1}\,d\tau.$$

As before $\mathcal{K}_{1,i}(w)$, a $N_r \times 1$ dimensional vector, is defined analogously as $\mathcal{K}_{2,i}(v)$ in (A-4), with a $N_{r,(l)} \times 1$ dimensional subvector with $l^0$–th element given by $[\mathcal{K}_{1,i;l}(w)]_{l^0} = ((W_i - w)/h_1)^{\phi_{r;l}(l^0)} K_1((W_i - w)/h_1)$, such that $n^{-1}h_1^{-(d+1)} \sum_{j=1}^{n} \mathcal{K}_{1,j}(w)$ converges in mean squared to $\mathbf{M}_{r,0} f_W(w)$. Define $\gamma(w) = E[\gamma_n(w)]$, then by Proposition 2 (page 581) and by Theorem 4 (page 582) in Masry (1996a), it follows that

$$\sup_{w=(x,z)\in\Psi_x\times\Psi_z} |\gamma(w)| = o(h_1^{p_1+1}),$$

$$\sup_{w=(x,z)\in\Psi_x\times\Psi_z} |h_1^{-(p_1+1)}\gamma_n(w) - \gamma(w)| = h_1^{p_1+1} O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n}). \tag{A-8}$$

Let

$$\beta_n(w) \equiv n^{-1}h_1^{-(d+1)} \sum_{j=1}^{n} \mathcal{K}_{1,j}(w) \frac{1}{\mathbf{k}!} \sum_{|\mathbf{k}|=p_1+1} D^{\mathbf{k}} r(w) (W_i - w)^{\mathbf{k}}, \text{ and}$$

$$\beta(w) = \mathbf{B}_r r^{(p_1+1)}(w) f_W(w),$$

then by Theorem 2 (page 579) in Masry (1996a), it follows that

$$\sup_{w=(x,z)\in\Psi_x\times\Psi_z} |h_1^{-(p_1+1)}\beta_n(w) - \beta(w)| = O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n}). \tag{A-9}$$

For the set $\{Y_i, \widetilde{M}_i\}_{i=1}^{n}$, as discussed in the main text, an estimator of the function $H$ is obtained by the intercept of the following minimizing problem

$$Q_{H,n}(\theta) = n^{-1}h_*^{-1} \sum_{i=1}^{n} k_* \left(\frac{\widetilde{M}_i - m}{h_*}\right) \left[Y_i - \sum_{0\le j\le p_*} \theta_j \left(\widetilde{M}_i - m\right)^j\right]^2.$$

Because this is a simple univariate nonparametric regression, its associated matrices $\mathbf{M}_H$, $\mathbf{M}_{H,0}^0$, $\mathbf{\Gamma}_H$, $\mathbf{B}_H$, $\mathbf{M}_{H,n}(m)$, $\widehat{\mathbf{M}}_{H,n}(m)$, and vector $\mathcal{K}_{*,i;l}(m)$ have simpler forms. They are as those previously described but replacing the responses by $Y_i$ and the conditioning variables by $M_i$ or $\widetilde{M}_i$ accordingly.

## Proof of Corollary 2.1

As before, given Assumption I*, it follows that $s(x,z) = h[G(x)F(z)]G(x)f(z)$, consequently $q(t,z_0) = h[H^{-1}(t)]H^{-1}(t)[f(z_0)/F(z_0)]$, and using the change of variables $m = H^{-1}(t)$, after noticing that $h[H^{-1}(t)] = h(m)$ and $dt = h(m)\,dm$, we obtain

$$\begin{aligned}
\int_{r_1}^{r(x,z)} \frac{dt}{q(t,z_0)} &= \int_{r_1}^{r(x,z)} \frac{F(z_0)}{h[H^{-1}(t)]H^{-1}(t)f(z_0)} dt \\
&= \int_{H^{-1}(r_1)}^{H^{-1}(r(x,z))} \frac{F(z_0)}{h(m)\,m f(z_0)} h(m)\,dm \\
&= \left[\frac{F(z_0)}{f(z_0)}\right] \left[\ln\left(H^{-1}[r(x,z)]\right) - \ln\left(H^{-1}[r_1]\right)\right] \\
&= \ln(M(x,z)) \equiv \ln(G(x)F(z)).
\end{aligned}$$

This proves the result. ■

## Proof of Theorem 4.1

Rearranging terms, we have

$$\begin{aligned}\widehat{M}(x,z)-M(x,z)&=\int_{r_0}^{\widehat{r}(x,z)}\frac{dt}{\widehat{q}(t,z_0)}-\int_{r_0}^{r(x,z)}\frac{dt}{q(t,z_0)}\\&=\left(\int_{r_0}^{\widehat{r}(x,z)}-\int_{r_0}^{r(x,z)}\right)\frac{dt}{q(t,z_0)}-\int_{r_0}^{\widehat{r}(x,z)}\left(\frac{\widehat{q}(t,z_0)-q(t,z_0)}{\widehat{q}(t,z_0)\,q(t,z_0)}\right)dt.\end{aligned}$$

By mean value expansions of the first term, in the last equality above, and after some manipulation we obtain,

$$\widehat{M}(x,z)-M(x,z)=\frac{1}{q(r,z_0)}\left(\widehat{r}(x,z)-r(x,z)\right)-\int_{r_0}^{r(x,z)}\frac{\widehat{q}(t,z_0)-q(t,z_0)}{q^2(t,z_0)}dt \tag{A-10}$$

$$-\int_{r(x,z)}^{\widehat{r}(x,z)}\frac{\widehat{q}(t,z_0)-q(t,z_0)}{q^2(t,z_0)}dt+\int_{r_0}^{\widehat{r}(x,z)}\frac{\left(\widehat{q}(t,z_0)-q(t,z_0)\right)^2}{\widehat{q}(t,z_0)\,q^2(t,z_0)}dt \tag{A-11}$$

$$=\mathcal{M}_{1,n}(x,z)-\mathcal{M}_{2,n}(x,z)-\mathcal{R}_{M,n}(x,z). \tag{A-12}$$

The terms in (A-10), $\mathcal{M}_{1,n}(x,z)$ and $\mathcal{M}_{2,n}(x,z)$, are linear in the estimation error from the two nonparametric regressions, while the remaining terms in (A-11), $\mathcal{R}_{Mn}(x,z)$, are both quadratic in such errors, and thus they will be shown to be of smaller order. $\mathcal{M}_{1,n}(x,z)$ is just a constant times the estimation error of $\widehat{r}(x,z)$, the unconstrained first–stage nonparametric estimator of $r(x,z)$, and under Assumption E, it can be analyzed directly using Theorem 4 (page 94) in Masry (1996b), given that $q(r(x,z),z)>0$ over $\Psi_x\times\Psi_z$. That is,

$$\sqrt{nh_1^{d+1}}\left(\mathcal{M}_{1,n}(x,z)-h_1^{p_1+1}\mathcal{B}_4(x,z)\right)\xrightarrow{d}N\left[0,\frac{\sigma_r^2(x,z)}{q^2(r,z_0)\,f_W(x,z)}\left[\mathbf{M}_r^{-1}\Gamma_r\mathbf{M}_r^{-1}\right]_{0,0}\right],$$

$$\mathcal{B}_4(x,z)=\left[\mathbf{M}_r^{-1}\mathbf{B}_r r^{(p_1+1)}(x,z)\right]_{0,0}q^{-1}(r,z_0) \tag{A-13}$$

where $[A]_{0,0}$ is the upper-left element of matrix $A$. In order to analyze the second term, $\mathcal{M}_{2,n}(x,z)$, we first notice that for any two symmetric nonsingular matrices $A_1$ and $A_2$, we have that $A_1^{-1}-A_2^{-1}=A_2^{-1}(A_2-A_1)A_1^{-1}$, which implies

$$
\begin{aligned}
\frac{\widehat{q}(t,z)-q(t,z)}{q^2(t,z)} &= \iota_2^\top \widehat{\mathbf{M}}_{q,n}^{-1}(v)\left[q^2(v)\right]^{-1}\widetilde{V}_{q,n}(v)+\iota_2^\top \widehat{\mathbf{M}}_{q,n}^{-1}(v)\left[q^2(v)\right]^{-1}\widehat{B}_{q,n}(v)\\
&= \iota_2^\top \widehat{\mathbf{M}}_{q,n}^{-1}(v)\left[q^2(v)\right]^{-1}\widehat{V}_{q,n}(v)+\iota_2^\top \widehat{\mathbf{M}}_{q,n}^{-1}(v)\left[q^2(v)\right]^{-1}\widehat{V}_{q,n}^{*}(v)\\
&+ \iota_2^\top \widehat{\mathbf{M}}_{q,n}^{-1}(v)\left[q^2(v)\right]^{-1}\widehat{B}_{q,n}(v)\\
&= \iota_2^\top\left[f_V(v)\,q^2(v)\,\mathbf{M}_q\right]^{-1}\widehat{V}_{q,n}(v)+\iota_2^\top\left[f_V(v)\,q^2(v)\,\mathbf{M}_q\right]^{-1}\widehat{V}_{q,n}^{*}(v)\\
&+ \iota_2^\top\left[f_V(v)\,q^2(v)\,\mathbf{M}_q\right]^{-1}\widehat{B}_{q,n}(v)\\
&- \iota_2^\top\left[f_V(v)\,q^2(v)\,\mathbf{M}_q\right]^{-1}\left[\widehat{\mathbf{M}}_{q,n}(v)-f_V(v)\,\mathbf{M}_q\right]\widehat{\mathbf{M}}_{q,n}^{-1}(v)\,\widehat{V}_{q,n}(v)\\
&- \iota_2^\top\left[f_V(v)\,q^2(v)\,\mathbf{M}_q\right]^{-1}\left[\widehat{\mathbf{M}}_{q,n}(v)-f_V(v)\,\mathbf{M}_q\right]\widehat{\mathbf{M}}_{q,n}^{-1}(v)\,\widehat{V}_{q,n}^{*}(v)\\
&- \iota_2^\top\left[f_V(v)\,q^2(v)\,\mathbf{M}_q\right]^{-1}\left[\widehat{\mathbf{M}}_{q,n}(v)-f_V(v)\,\mathbf{M}_q\right]\widehat{\mathbf{M}}_{q,n}^{-1}(v)\,\widehat{B}_{q,n}(v)\\
&\equiv T_{q,n,1}(v)+T_{q,n,2}(v)+T_{q,n,3}(v)-T_{q,n,4}(v)-T_{q,n,5}(v)-T_{q,n,6}(v)
\end{aligned}
$$

where $\mathbf{M}_q$ is defined in (A-3). We have also defined $\widetilde{V}_{q,n}(v)=\widehat{V}_{q,n}(v)+\widehat{V}_{q,n}^{*}(v)$, where the $N_q\times 1$ vectors $\widehat{V}_{q,n}(v)$, $\widehat{V}_{q,n}^{*}(v)$, and $\widehat{B}_{q,n}(v)$ are

$$
\begin{aligned}
\widehat{V}_{q,n}(v) &= n^{-1}h^{-2}\sum_{i=1}^{n}\widehat{\mathcal{K}}_{2,i}(v)\,\varepsilon_{q,i},\\
\widehat{V}_{q,n}^{*}(v) &= n^{-1}h^{-2}\sum_{i=1}^{n}\widehat{\mathcal{K}}_{2,i}(v)\,[\widehat{S}_i-S_i],\\
\widehat{B}_{q,n}(v) &= n^{-1}h^{-2}\sum_{i=1}^{n}\widehat{\mathcal{K}}_{2,i}(v)\,\widehat{\Delta}_{q,i}(v)\,,\text{ and}\\
\widehat{\Delta}_{q,i}(v) &\equiv q(\widehat{V}_i)-\sum_{0\le|\mathbf{m}|\le p_2}\frac{1}{\mathbf{m}!}\,(D^{\mathbf{m}}q)\,(v)\,(\widehat{V}_i-v)^{\mathbf{m}}.
\end{aligned}
$$

Consequently,

$$
\mathcal{M}_{2n}(x,z)=\mathcal{T}_{q,n,1}(x,z)+\mathcal{T}_{q,n,2}(x,z)+\mathcal{T}_{q,n,3}(x,z)+\mathcal{R}_{q,n}(x,z)\,,
$$

where $\mathcal{T}_{q,n,l}(x,z)=\int T_{q,n,l}(t,z_0)\,dQ(t)$ for $l=1,2,3$ and $dQ(t)=1\,(r_0\le t\le r(x,z))\,dt$. These terms, along with the remainder $\mathcal{R}_{q,n}(x,z)=\sum_{l=4}^{6}\int T_{q,n,l}(t,z_0)\,dQ(t)$ are dealt with in Lemmas B1 to B-4 in Jacho-Chávez, Lewbel, and Linton (2006), from which we conclude that

$$\begin{aligned}\mathcal{M}_{2n}(x,z) &= h_1^{p_1+1}\iota_2^\top \mathbf{M}_q^{-1}\mathbf{M}_{q,0}^0\iota_1^\top \mathbf{M}_r^{-1}\mathbf{B}_r\left[\frac{E\left[r^{(p_1+1)}(X,Z)g_q(X,Z)\right| r(X,z_0)=r, Z=z_0\right]}{q^2(r,z_0)}\right.\\ &\qquad\left.-\frac{E\left[r^{(p_1+1)}(X,Z)g_q(X,Z)\right| r(X,z_0)=r_0, Z=z_0\right]}{q^2(r_0,z_0)}\right]\\ &+ h_1^{p_1}h_2\iota_2^\top \mathbf{M}_q^{-1}\mathbf{M}_{q,0}^1\iota_1^\top \mathbf{M}_r^{-1}\mathbf{B}_r\left[\frac{E\left[r^{(p_1+1)}(X,Z)\right| r(X,z_0)=r, Z=z_0\right]}{q^2(r,z_0)}\right.\\ &\qquad\left.-\frac{E\left[r^{(p_1+1)}(X,Z)\right| r(X,z_0)=r_0, Z=z_0\right]}{q^2(r_0,z_0)}\right]\\ &+ h_2^{p_2+1}\iota_2^\top \mathbf{M}_q^{-1}\mathbf{B}_q\int_{r_0}^{r(x,z)}\frac{q^{(p_2+1)}(t,z_0)}{q^2(t,z_0)}dt + o_p(n^{-1/2}h_1^{-(d+1)/2})\\ &= h_1^{p_1+1}\mathcal{B}_1(x,z) + h_1^{p_1}h_2\mathcal{B}_2(x,z) + h_2^{p_2+1}\mathcal{B}_3(x,z) + o_p(n^{-1/2}h_1^{-(d+1)/2}). \qquad \text{(A-14)}\end{aligned}$$

Finally, the last term in (A-12), $\mathcal{R}_{M,n}(x,z) = O_p(\nu_{1n})O_p(h_2^{-1}\nu_{1n} + h_1^{-1}\nu_{1n} + \nu_{2n}) + O_p((h_2^{-1}\nu_{1n} + h_1^{-1}\nu_{1n} + \nu_{2n})^2)$, by Theorem 6 (page 594) in Masry (1996a) and Lemma B-5 in Jacho-Chávez, Lewbel, and Linton (2006). Therefore, it follows from Assumption (E5) that $\mathcal{R}_{M,n}(x,z) = o_p(n^{-1/2}h_1^{(d+1)/2})$. By grouping terms, $\mathcal{B}_n(x,z) \equiv h_1^{p_1+1}\mathcal{B}_1(x,z) + h_1^{p_1}h_2\mathcal{B}_2(x,z) + h_2^{p_2+1}\mathcal{B}_3(x,z) - h_1^{p_1+1}\mathcal{B}_4(x,z)$, we conclude the proof of the theorem. ■

## Proof of Theorem 4.2

As before, we can write

$$\begin{aligned}\widehat{H}(m) - H(m) &= \iota_*^\top \widehat{\mathbf{M}}_{H,n}^{-1}(m)\widetilde{V}_{H,n}(m) + \iota_*^\top \widehat{\mathbf{M}}_{H,n}^{-1}(m)\widehat{B}_{H,n}(m)\\ &= \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1}\widetilde{V}_{H,n}(m) + \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1}\widehat{B}_{H,n}(m)\\ &- \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1}\left[\widehat{\mathbf{M}}_{H,n}(m) - f_M(m)\mathbf{M}_H\right]\widehat{\mathbf{M}}_{H,n}^{-1}(m)\widetilde{V}_{H,n}(m)\\ &- \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1}\left[\widehat{\mathbf{M}}_{H,n}(m) - f_M(m)\mathbf{M}_H\right]\widehat{\mathbf{M}}_{H,n}^{-1}(m)\widehat{B}_{H,n}(m)\\ &= \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1}\widehat{V}_{H,n}(m) + \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1}\widehat{V}_{H,n}^*(m)\\ &+ \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1}\widehat{B}_{H,n}(m)\\ &- \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1}\left[\widehat{\mathbf{M}}_{H,n}(m) - f_M(m)\mathbf{M}_H\right]\widehat{\mathbf{M}}_{H,n}^{-1}(b)\widehat{V}_{H,n}(b)\\ &- \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1}\left[\widehat{\mathbf{M}}_{H,n}(m) - f_M(m)\mathbf{M}_H\right]\widehat{\mathbf{M}}_{H,n}^{-1}(m)\widehat{V}_{H,n}^*(m)\\ &- \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1}\left[\widehat{\mathbf{M}}_{H,n}(m) - f_M(m)\mathbf{M}_H\right]\widehat{\mathbf{M}}_{H,n}^{-1}(m)\widehat{B}_{H,n}(m)\\ &\equiv T_{H,n,1}(m) + T_{H,n,2}(m) + T_{H,n,3}(m) - T_{H,n,4}(m) - T_{H,n,5}(m) - T_{H,n,6}(m),\end{aligned}$$

where

$$\widetilde{V}_{H,n}(m) \equiv \widehat{V}_{H,n}(m) + \widehat{V}^{*}_{H,n}(m),$$
$$\widehat{V}_{H,n}(m) = n^{-1}h_*^{-1}\sum_{i=1}^{n}\widehat{\mathcal{K}}_{*,i}(m)\varepsilon_{r,i},$$
$$\widehat{V}^{*}_{H,n}(m) = n^{-1}h_*^{-1}\sum_{i=1}^{n}\widehat{\mathcal{K}}_{*,i}(m)[H(M_i) - H(\widetilde{M}_i)], \text{ and}$$
$$\widehat{B}_{H,n}(m) = n^{-1}h_*^{-1}\sum_{i=1}^{n}\widehat{\mathcal{K}}_{*,i}(m)\widehat{\Delta}_{H,i}(m), \text{ with}$$
$$\widehat{\Delta}_{H,i}(m) \equiv H(\widehat{M}_i) - \sum_{0\le j\le p_*}\frac{1}{j!}(\partial^j H\left(m\right)/\partial m^j)(\widehat{M}_i - m)^j.$$

We analyze the properties of $T_{H,n,l}(b)$, $l = 1,\ldots,6$ in Lemmas B-7 to B-10 in Jacho-Chávez, Lewbel, and Linton (2006), which show that $T_{H,n,1}(m) = O_p(n^{-1/2}h_*^{-1/2})$ and that $T_{H,n,2}(m) \xrightarrow{p} \mathcal{B}_{H2}\left(m\right)$, $T_{H,n,3}(m) \xrightarrow{p} \mathcal{B}_{H3}\left(m\right)$, where

$$\mathcal{B}_{H2}\left(m\right) \equiv -\iota_*^{\top}\mathbf{M}_H^{-1}\mathbf{M}_{H,0}^{0}E\left[H^{(1)}\left(M\left(X,Z\right)\right)\theta\left(X,Z\right)\middle| H\left(M\left(X,Z\right)\right)=m\right],$$
$$\mathcal{B}_{H3}\left(m\right) \equiv h_*^{p_*+1}\iota_*^{\top}\mathbf{M}_H^{-1}\mathbf{B}_H H^{(p_*+1)}\left(m\right),$$

with $-\theta\left(w\right) \equiv \int\mathcal{B}_n\left(x,z\right)dP_1\left(z\right) + \int\mathcal{B}_n\left(x,z\right)dP_2\left(x\right) - \int\int\mathcal{B}_n\left(x,z\right)dP_1\left(z\right)dP_2\left(x\right)$ which is $O\left(h_\dagger\right)$ by construction. By defining $\mathcal{B}_{nH}\left(m\right) \equiv \mathcal{B}_{H2}\left(m\right) + \mathcal{B}_{H3}\left(m\right)$, the proof is completed. ■

Table 1: Median of Monte Carlo fit criteria over grid for Design 1.

| | | | $\sigma^2 = 1$ | | | | $\sigma^2 = 2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | (1) | | (2) | | (1) | | (2) | |
| | $cc$ | $n$ | IRMSE | IMAE | IRMSE | IMAE | IRMSE | IMAE | IRMSE | IMAE |
| $G(x)$ | 0.5 | 150 | 0.223 | 0.175 | 0.621 | 0.467 | 0.314 | 0.247 | 0.575 | 0.438 |
| | | 600 | 0.115 | 0.090 | 0.456 | 0.336 | 0.160 | 0.126 | 0.551 | 0.409 |
| | 1 | 150 | 0.154 | 0.123 | 0.230 | 0.178 | 0.219 | 0.176 | 0.380 | 0.281 |
| | | 600 | 0.082 | 0.066 | 0.137 | 0.107 | 0.115 | 0.091 | 0.212 | 0.161 |
| | 1.5 | 150 | 0.130 | 0.105 | 0.136 | 0.111 | 0.187 | 0.151 | 0.200 | 0.163 |
| | | 600 | 0.071 | 0.057 | 0.080 | 0.064 | 0.098 | 0.079 | 0.116 | 0.091 |
| $F(z)$ | 0.5 | 150 | 0.225 | 0.175 | 0.632 | 0.481 | 0.314 | 0.245 | 0.580 | 0.446 |
| | | 600 | 0.113 | 0.089 | 0.453 | 0.339 | 0.161 | 0.126 | 0.554 | 0.425 |
| | 1 | 150 | 0.156 | 0.124 | 0.246 | 0.191 | 0.216 | 0.175 | 0.394 | 0.309 |
| | | 600 | 0.083 | 0.067 | 0.148 | 0.117 | 0.117 | 0.092 | 0.218 | 0.174 |
| | 1.5 | 150 | 0.135 | 0.110 | 0.173 | 0.141 | 0.185 | 0.152 | 0.256 | 0.209 |
| | | 600 | 0.073 | 0.059 | 0.098 | 0.078 | 0.102 | 0.082 | 0.137 | 0.112 |
| $M(x,z)$ | 0.5 | 150 | 0.343 | 0.271 | 1.047 | 0.826 | 0.485 | 0.379 | 0.936 | 0.743 |
| | | 600 | 0.174 | 0.136 | 0.737 | 0.592 | 0.242 | 0.191 | 0.914 | 0.704 |
| | 1 | 150 | 0.247 | 0.197 | 0.386 | 0.311 | 0.345 | 0.277 | 0.641 | 0.501 |
| | | 600 | 0.130 | 0.104 | 0.234 | 0.189 | 0.181 | 0.145 | 0.357 | 0.292 |
| | 1.5 | 150 | 0.215 | 0.174 | 0.260 | 0.215 | 0.303 | 0.247 | 0.377 | 0.317 |
| | | 600 | 0.115 | 0.092 | 0.147 | 0.119 | 0.162 | 0.130 | 0.209 | 0.169 |
| $H(m)$ | 0.5 | 150 | 0.343 | 0.271 | 0.326 | 0.258 | 0.485 | 0.379 | 0.401 | 0.315 |
| | | 600 | 0.174 | 0.136 | 0.208 | 0.165 | 0.242 | 0.191 | 0.270 | 0.212 |
| | 1 | 150 | 0.247 | 0.197 | 0.228 | 0.180 | 0.345 | 0.277 | 0.311 | 0.246 |
| | | 600 | 0.130 | 0.104 | 0.133 | 0.104 | 0.181 | 0.145 | 0.183 | 0.144 |
| | 1.5 | 150 | 0.215 | 0.174 | 0.184 | 0.147 | 0.303 | 0.247 | 0.255 | 0.206 |
| | | 600 | 0.115 | 0.092 | 0.107 | 0.083 | 0.162 | 0.130 | 0.146 | 0.115 |

Note: Results for Linton and Nielsen (1995) are in column (1), and column (2) corresponds to the new estimator.

Table 2: Median of Monte Carlo fit criteria over grid for Design 2.

| | | | $\sigma^2 = 1$ | | | | $\sigma^2 = 2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | (1) | | (2) | | (1) | | (2) | |
| | $cc$ | $n$ | IRMSE | IMAE | IRMSE | IMAE | IRMSE | IMAE | IRMSE | IMAE |
| $G(x)$ | 0.5 | 150 | 0.298 | 0.234 | 0.632 | 0.472 | 0.419 | 0.328 | 0.591 | 0.461 |
| | | 600 | 0.153 | 0.119 | 0.494 | 0.375 | 0.212 | 0.167 | 0.542 | 0.413 |
| | 1 | 150 | 0.206 | 0.163 | 0.323 | 0.249 | 0.292 | 0.235 | 0.507 | 0.382 |
| | | 600 | 0.109 | 0.087 | 0.189 | 0.145 | 0.152 | 0.121 | 0.301 | 0.225 |
| | 1.5 | 150 | 0.173 | 0.139 | 0.191 | 0.154 | 0.250 | 0.201 | 0.276 | 0.226 |
| | | 600 | 0.094 | 0.076 | 0.108 | 0.086 | 0.131 | 0.105 | 0.159 | 0.125 |
| $F(z)$ | 0.5 | 150 | 0.299 | 0.233 | 0.624 | 0.475 | 0.419 | 0.326 | 0.608 | 0.471 |
| | | 600 | 0.151 | 0.118 | 0.493 | 0.382 | 0.214 | 0.167 | 0.564 | 0.425 |
| | 1 | 150 | 0.206 | 0.165 | 0.337 | 0.257 | 0.287 | 0.233 | 0.549 | 0.421 |
| | | 600 | 0.110 | 0.087 | 0.204 | 0.162 | 0.154 | 0.122 | 0.319 | 0.240 |
| | 1.5 | 150 | 0.178 | 0.145 | 0.236 | 0.191 | 0.246 | 0.201 | 0.348 | 0.282 |
| | | 600 | 0.095 | 0.076 | 0.130 | 0.105 | 0.134 | 0.107 | 0.186 | 0.151 |
| $M(x,z)$ | 0.5 | 150 | 0.457 | 0.360 | 1.027 | 0.801 | 0.647 | 0.504 | 0.995 | 0.761 |
| | | 600 | 0.230 | 0.180 | 0.818 | 0.653 | 0.322 | 0.255 | 0.883 | 0.690 |
| | 1 | 150 | 0.327 | 0.262 | 0.535 | 0.427 | 0.458 | 0.368 | 0.881 | 0.688 |
| | | 600 | 0.171 | 0.137 | 0.323 | 0.261 | 0.240 | 0.192 | 0.522 | 0.413 |
| | 1.5 | 150 | 0.282 | 0.230 | 0.355 | 0.295 | 0.402 | 0.327 | 0.522 | 0.434 |
| | | 600 | 0.150 | 0.121 | 0.196 | 0.160 | 0.213 | 0.171 | 0.285 | 0.235 |
| $H(m)$ | 0.5 | 150 | 0.329 | 0.261 | 0.291 | 0.229 | 0.463 | 0.367 | 0.349 | 0.273 |
| | | 600 | 0.165 | 0.130 | 0.194 | 0.153 | 0.231 | 0.184 | 0.244 | 0.192 |
| | 1 | 150 | 0.238 | 0.192 | 0.223 | 0.176 | 0.334 | 0.270 | 0.287 | 0.228 |
| | | 600 | 0.125 | 0.100 | 0.132 | 0.104 | 0.175 | 0.141 | 0.181 | 0.143 |
| | 1.5 | 150 | 0.208 | 0.170 | 0.183 | 0.146 | 0.294 | 0.242 | 0.250 | 0.201 |
| | | 600 | 0.110 | 0.089 | 0.104 | 0.082 | 0.157 | 0.126 | 0.144 | 0.113 |

Note: Results for Linton and Härdle (1996) are in column (1), and column (2) corresponds to the new estimator.

Figure 1: Generalized Homogeneous $M$, and Strictly Monotonic $H$.

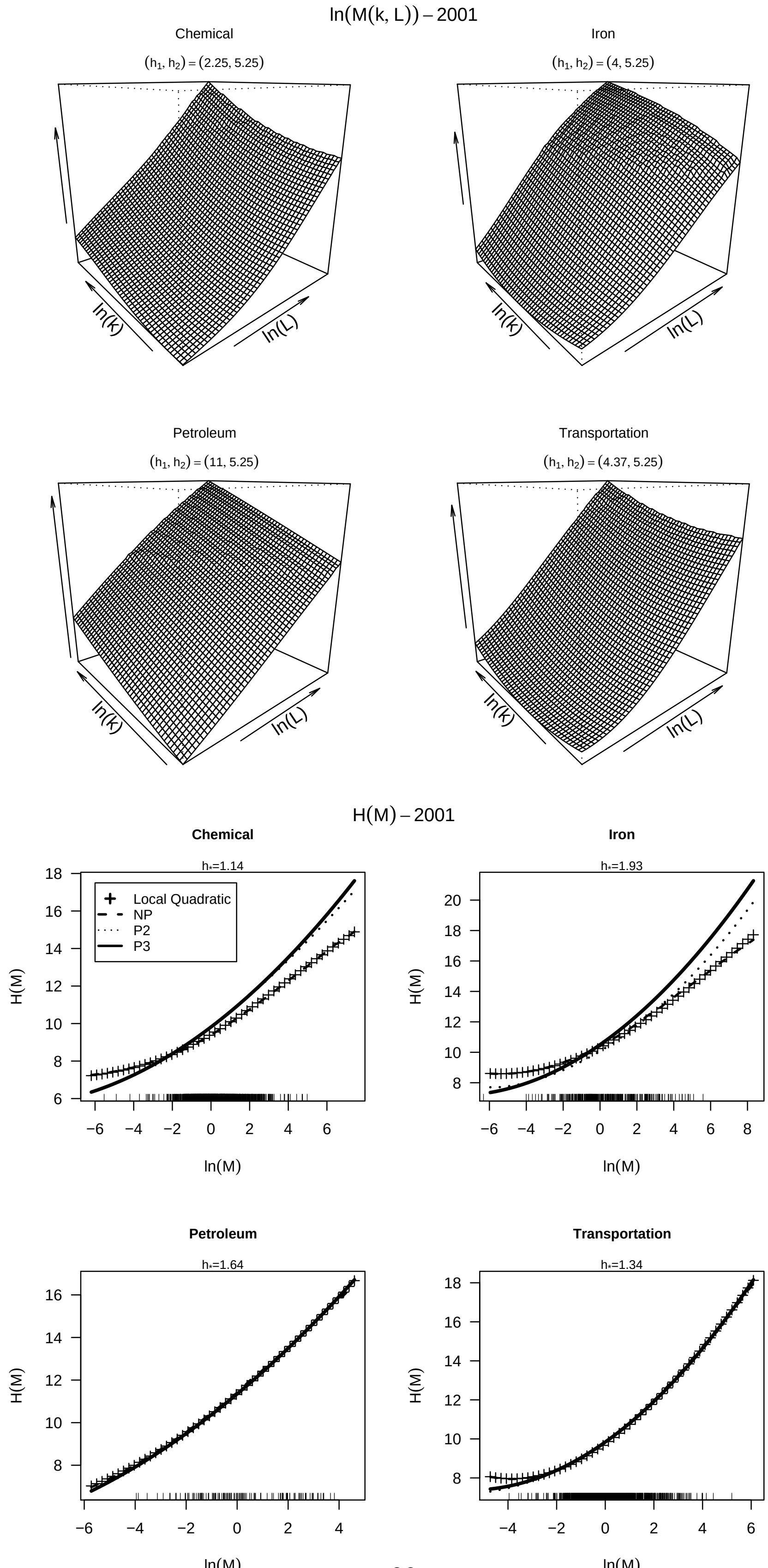

Figure 2: Generalized Homogeneous Components $G$ and $F$.