Identification and Nonparametric Estimation of a Transformed Additively Separable Model[*]

DavidJacho-Chávez[†]
Indiana University

Arthur Lewbel[‡]
Boston College

Oliver Linton[§]
London School of Economics

[†] 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/

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)] and M (x, z) = G(x) + F (z). 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 and test 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

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. As an economic model and to speed convergence rates by reducing the curse of dimensionality as d increases (see, e.g., Stone (1980) and Stone (1986)), 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 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), as well as present evidence of their small sample performance in a some Monte Carlo experiments and 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 (2006), 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 $r(x, z) = \widetilde{H}\left[\widetilde{G}(x), z\right]$, 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, and we attain faster convergence rates.

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 ε, with ε 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 $Y = Y^* 1(Y^* \geq 0)$, or $Y = Y^* | Y^* \geq 0$, or $Y = 1(Y^* \geq 0)$, 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 or a quantile function of ε. 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).

Model (1.1) may arise in a nonparametric regression model with unknown transformation of the dependent variable, $F(z) = G(x) + \varepsilon$, where ε 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 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, ϕ is a continuous strictly decreasing convex function from $[0, 1]$ to $[0, \infty]$ such that $\phi(1) = 0$, and ϕ^{-1} denotes the inverse. A collection of one–parameter families of Archimedean copulas can be found in Nelsen (2006).

We also identify and estimate 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 (2006), are the special case in which $F(z) = z$.

We implement our methodology to estimate generalized homothetic production functions for four industries in the People's Republic of 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 use knowledge of H, and with Horowitz (2001). This section also provides an empirical illustration of our methodology for the estimation of generalized production functions in four industries within the Chinese economy for the years 1995 and 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 the nonparametric 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 $f_W(w) > 0$ for all $w = (x, z) \in \Psi_x \times \Psi_z$. 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, $f(z)$, 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 are not nonparametrically identified if (X, Z) has discrete elements, a restriction which is common in nonparametric models

3

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

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

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

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

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

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 $|H^{-1}[r_0]| < \infty$. 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.

Lewbel and Linton (2002) also use a similar result (2.2) in the nonparametric censored regression setup, $Y = \max[0, M(w) - \varepsilon]$. Their estimator assumes independence between W and ε with $E(\varepsilon) = 0$. For the case where their M is partly separable case, Theorem 2.1 above replicates their Theorem 3 (page 769), but with additional normalizations. In particular, $q(t, z_0) = F_\varepsilon[\mathfrak{F}^{-1}(t)] f(z_0)$, where F_ε is the cumulative distribution function of ε and $\mathfrak{F}(m) = \int\limits_{-\infty}^{m} F_\varepsilon(e) \, de$. As is in our case, their location constant must be known a priori. The assumed additive separability with respect to z also adds an extra normalization on Ψ_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 $f_W(w) > 0$ for all $w = (x, z) \in \Psi_x \times \Psi_z$. 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 and H, G and F are continuous and differentiable with respect to any mixture of their arguments. (ii) F has finite first derivative, $f(z)$, 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 $w = (x, z) \in \Psi_x \times \Psi_z$, where $\Psi_{r(x,z)}$ is the image of the function $r(x, z)$.

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

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

Proof. See the appendix. ∎

If r_l is greater than $r(x, z)$, for any nonnegative constant, r_l, 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 $l = 0, 1$. 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 $(c = c_1 \times c_2)$.

Given the definition of $r(x,z)$, it follows that $H(M(x,z)) = E[r(X,Z) | M(X,Z) = M(x,z)]$, thus the function H may also be identified. 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 H.

We could replace the scale–sign normalization in Assumption (I2)(ii) by the assumption that there is a bounded, non–negative function, ω, such that

$$\int \frac{\omega(z_0)}{f(z_0)} dz_0 = 1,$$

with ω integrating to one over its compact support. For the applied researcher, a normalization restriction such as (I2) is empirically appealing because it entails the selection of a single value rather than a whole function, and it reduces computation time. These restrictions may also 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.

Strict monotonicity of the link function H plays an important role in these results. Because of it, 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 contrasts with Horowitz (2001) and Ekeland, Heckman, and Nesheim (2004), where strict monotonicity is neither assumed nor is it exploited for identification, rather it is the separability of the partial derivatives of $r(x,z)$ that is used instead. It is also worth noting that our identification result does not require stochastic variation in $s(x,z)$ (it could be known or take on random values) once conditioned on r and z.

3 Estimation

In this section, for the case $r(x, z) \equiv E[Y \mid X = x, Z = z]$, 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 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 a local polynomial estimator, which produces estimators for r and s simultaneously. These nonparametric estimators also have better boundary behavior and the ability to adapt to non–uniform designs, among other desirable properties (see Fan and Gijbels (1996)).

For a given random sample $\{Y_i, X_i, Z_i\}_{i=1}^n$, estimators of M, G, F and H in the additive case, can be constructed by following these steps:

1) Obtain a consistent estimator of $\widehat{r}_i = \widehat{r}(X_i, Z_i)$ and $\widehat{s}_i = \widehat{s}(X_i, Z_i)$ 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 $h_1 = h_1(n)$ for $i = 1, \ldots, n$.

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 with corresponding kernel K_2 and bandwidth sequence $h_2 = h_2(n)$ for $i = 1, \ldots, n$. Denote this estimate as $\widehat{q}(t, z_0) = \widehat{E}[\widehat{s} \mid \widehat{r}(X, Z) = t, Z = z_0]$.

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 $\widetilde{c} = (1/2) \left[\int_{\Psi_x} \widehat{\alpha}_{P_1}(x) \, dP_2(x) + \int_{\Psi_z} \widehat{\alpha}_{P_2}(z) \, dP_1(z) \right]$, 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 $\widehat{H}(m)$.

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 $\widetilde{c} = (1/2)\left[\int_{\Psi_x} \widehat{\alpha}_{P_1}(x)\, dP_2(x) + \int_{\Psi_z} \widehat{\alpha}_{P_2}(z)\, dP_1(z)\right]$, 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.

Known Link Function

In many practical situations, especially with binary and survival time data, the conditional distribution of Y given (X, Z) belongs to a known family with a known link function H. For example, the logit and probit link functions are common for binary data, and the logarithm transform for Poisson count data; see McCullagh and Nelder (1989). If the known H is twice continuously differentiable such that $h(M) = H^{(1)}(M) \equiv \partial H(m)/\partial m|_{m=M} \neq 0$ over its entire support, the function $q(t, z_0)$ in Theorem 2.1 and Corollary 2.1 can be replaced by $q_{add}(t) \equiv h[H^{-1}(t)]$ in the additive case, or by $q_{mult}(t) \equiv h[H^{-1}(t)]H^{-1}(t)$ in the multiplicative one, so a scale normalization is then not needed.

Specifically,

$$
\int_{r_0}^{r(x,z)} \frac{dt}{q_{add}(t)} + H^{-1}[r_0] = \int_{r_0}^{r(x,z)} \frac{dt}{h[H^{-1}(t)]} + H^{-1}[r_0] \tag{3.4}
$$

$$
= H^{-1}[r(x,z)] = M(x,z) \equiv G(x) + F(z),
$$

and similarly

$$
\exp\left(\int_{r_1}^{r(x,z)} \frac{dt}{q_{mult}(t)} + \ln\left(H^{-1}[r_1]\right) \right) = \exp\left(\int_{r_1}^{r(x,z)} \frac{dt}{h[H^{-1}(t)]H^{-1}(t)} + H^{-1}[r_1] \right) \tag{3.5}
$$

$$
= H^{-1}[r(x,z)] = M(x,z) \equiv G(x)F(z),
$$

by a change of variables $m = H^{-1}(t)$, such that $dt = h(m)\,dm$. The above equalities hold for any r_l such that $H^{-1}[r_l] < \infty$ for $l = 0, 1$, so it does not require a location normalization as well. Notice that $q(t, z_0) = q_{add}(t)(1/f(z_0))$ and $q(t, z_0) = q_{mult}(t)(F(z_0)/f(z_0))$, in the additive and multiplicative cases respectively.

After replacing the unknown conditional mean function $r(x, z)$, in (3.4) and (3.5), by a local p_1-th order polynomial regression of Y on X and Z with kernel K_1, and bandwidth sequence $h_1 = h_1(n)$, we obtain $\widehat{M}(x, z) = H^{-1}[\widehat{r}(x, z)]$, which corresponds to the estimator proposed by Linton and Härdle (1996) and to that proposed by Linton and Nielsen (1995) for the identity link. In the fully additive case, $G(x) = \sum_{k=1}^{d} G(x_k)$, they also derive the asymptotic properties of $\widehat{M}$, $\widehat{G}_k$ and $\widehat{F}$. Later, in Section 5, we compare the performance of our procedure to that of these two estimators in the special case where they apply, which is when H is known and $d = 1$.

4 Asymptotic Properties

This section gives assumptions under which we present theorems providing 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 (2004), Su and Ullah (2006), and Lewbel and Linton (2006). Once the asymptotic normal distribution of M is established, the asymptotic properties of G and F will follow from ordinary 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 $H_{l\mathbf{j}} = u^{\mathbf{j}} K_l(u)$ 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 $W_i^\top \equiv (X_i^\top, Z_i)$ 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 $\xi > 2$, $E[|\varepsilon_{r,i}|^\xi] < \infty$, $E[|\varepsilon_{q,i}|^\xi] < \infty$, and $E[|\varepsilon_{r,i}\varepsilon_{q,i}|^\xi] < \infty$ where $\varepsilon_{r,i} = Y_i - r(X_i, Z_i)$ and $\varepsilon_{q,i} = S_i - q(r_i, Z_i)$. Also, $E[\varepsilon_{r,i}^2 | X_i = x, Z_i = z] \equiv \sigma_r^2(x, z)$, be such that $\nu_{P_1}(x) \equiv \int p_1^2(z) \sigma_r^2(x, z) f_W^{-1}(x, z) q^{-2}(r, z_0) \, dz < \infty$ and $\nu_{P_2}(z) \equiv \int p_2^2(x) \sigma_r^2(x, z) f_W^{-1}(x, z) q^{-2}(r, z_0) \, dx < \infty$.

(E4) The function $r(\cdot)$ is $(p_1 + 1)$ times partially continuously differentiable and the function $q(\cdot)$ 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) $n h_1^{d+1} h_2^{2(p_2+1)} \to c \in [0, \infty)$,

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

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 $E[\varepsilon_{q,i} | r(X_i, z) = r, Z_i = z] = 0$. However, $E[\varepsilon_{q,i} | X_i = x, Z_i = z] \neq 0$, so we write $\varepsilon_{q,i} = g_q(x, z) + \eta_i$, where $E[\eta_i | X_i = x, Z_i = z] = 0$ 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), $\max_{1 \le j \le n} \| \widehat{r}(W_j) - r(W_j) \| = O_p(\nu_{1n})$, $\max_{1 \le j \le n} \| \widehat{s}(W_j) - s(W_j) \| = O_p(h_1^{-1}\nu_{1n})$ and $\sup_v \| \widehat{q}(v) - q(v) \| = O_p(\nu_{2n})$ 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 $h_2^{-1}\nu_{1n} = o(1)$, 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 bounded continuous function $\mathcal{B}(x, z)$ such that*

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

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. ■

We should mention that there are four sources of biases, defined in the appendix, i.e. $\mathcal{B}(x, z) = h_1^{p_1+1}\mathcal{B}_1(x, z) + h_1^{p_1}h_2\mathcal{B}_2(x, z) + h_2^{p_2+1}\mathcal{B}_3(x, z) + h_1^{p_1+1}\mathcal{B}_4(x, z)$, 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 $\mathcal{B}_2$ are because of the use of generated regressor $\widehat{r}$, and generated response $\widehat{s}$ in constructing $\widehat{q}$ respectively 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 $Var\{\widehat{M}(x, z)\} = O(n^{-1}h_1^{-(d+1)})$, 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*

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

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

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

Proof. Given the 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 the conditional variance $\sigma_r^2(x, z)$ of Y. They 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 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 $\max_{1 \leq j \leq n} \| \widetilde{M}_j - M_j \| = O_p(\nu_{\dagger n})$, where $\nu_{\dagger n} = n^{-1/2}h_1^{-d/2}\sqrt{\ln n} + h_\dagger$.

To obtain the limiting distribution of $\widehat{H}$, 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 $0 \leq j \leq 2p_* + 1$ 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) $nh_*^{2(p_*+1)+1} \to c \in [0, \infty)$, $nh_*h_\dagger^2 \to c \in [0, \infty)$,

(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 $\left(h_*^{-1}\nu_{\dagger n}\right) = o\left(1\right)$. 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 $\sigma_H^2\left(m\right) = E\left[\varepsilon_r^2\middle|\, M\left(X,Z\right) = m\right]$, then we have the following theorem:

Theorem 4.2 *Suppose that Assumption I holds, then, under Assumption E and F, there exists a bounded continuous function* $\mathcal{B}_H\left(\cdot\right)$, *such that*

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

for $m \in \Psi_{M(x,z)}$, *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. ■

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, $\mathcal{B}_H\left(\cdot\right)$ 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 $\widehat{r}\left(X_i, Z_i\right) = \widehat{H}[\widetilde{M}\left(X_i, Z_i\right)]$ 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 $M\left(X_i, Z_i\right)$. 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. In the next section, we compare the performance of our estimator to that of Horowitz (2001).

5   Numerical Results

5.1   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 in two leading scenarios: When the link function is known and the case when it is not. Code for these simulations was written in GAUSS. Our experimental designs are not chosen resemble common economic models, but rather to emphasize performance issues. For simplicity we restrict attention to $d = 1$.

5.1.1 Known Link Function

We first compare the performance of our estimator to those of Linton and Nielsen (1995) and Linton and Härdle (1996). These alternative estimators are not fully efficient, but they do use knowledge of the link function. Hence, they provide a strong benchmarks for comparison with our estimator.

The first set of experiments are as follows: A number, n, of observations (Y, X, Z) are generated from $Y = r(X, Z) + \sigma_r \cdot \varepsilon$, where the distributions of x and z are $U[0, 1]$, ε is chosen independently of X and Z with a standard normal distribution, and $r(x, z) = H[M(x, z)]$ where $M(x, z) = G(x) + F(z)$ with

$$G(x) = (1/2) \sin(2\pi x)$$
$$F(z) = -2z^2 + 2z - 1/3.$$

The curvature and non-monotonicity of G and F provide a test for the estimators described in Section 3. Notice that neither G nor F is homogeneous and both are chosen such that $E[G(X)] = E[F(Z)] = 0$. Also, at $z_0 = 1/4$, we have $f(z_0) = 1$. For each scenario, $\sigma_r^2 = 1$ and $\sigma_r^2 = 2$, two specifications of H are used,

$$H[m] = m \tag{5.1}$$
$$H[m] = \ln\left(m + \sqrt{1 + m^2}\right) + 3. \tag{5.2}$$

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 $h_1 = cc\widehat{s}_W n^{-1/9}$, where cc is a constant term and $\widehat{s}_W$ is the squared 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 ($1.06n^{-1/5}$ times the squared root of the average of the regressors variances). Three different choices of cc are considered: $cc \in \{0.5, 1, 1.5\}$.

Each function is estimated on a 50×50 equally spaced grid in $[0, 1] \times [0, 1]$ when $n = 150$, and at another 60×60 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 on these tables (on the first column from the left under each fitting criteria respectively) the results obtained using the estimators proposed by Linton and Nielsen (1995) for H design (5.1), and Linton and Härdle (1996) for H design (5.2). 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 by roughly 5 to 85 percent 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, while the proposed estimator has a smaller IRMSE and IMAE when estimating F relative to estimates of G. 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 σ_r is increased.

5.1.2 Unknown Link Function

As noted earlier, when $d = 1$, model (1.1) is nested in the class of models Horowitz (2001) considers. We therefore compare our estimator with his in this specific case. We first replicate the Horowitz (2001) original experiment[1] which is as follows: 1000 observations (Y, X, Z) are generated from $Y = 1\left(G\left(X\right) + F\left(Z\right) - \varepsilon > 0\right)$, where $\varepsilon \sim N\left(0, 1\right)$, $X \sim N\left(0, 16\right)$ and $Z \sim N\left(0, 16\right)$, independent of each other. The functions G, F and H are[2]

$$G\left(x\right) = 3 \sin\left(\frac{\pi}{3} x^2\right),$$
$$F\left(z\right) = 3\left[\exp\left(0.35z\right) - 1\right], \text{ and}$$
$$H\left(m\right) = \Phi\left(m\right),$$

where Φ is the standard normal distribution function. This is a binary probit model, where $\Pr\left(Y = 1 | X = x, Z = z\right) = \Phi\left(G\left(x\right) + F\left(z\right)\right) \equiv r\left(x, z\right)$.

Horowitz (2001) (NP2) uses the following fourth and second order kernels to estimate G, F and H:

$$K\left(u\right) = \frac{105}{64}\left(1 - 5u^2 + 7u^4 - 3u^6\right) 1\left(|u| \leq 1\right),$$
$$K_H\left(u\right) = \frac{15}{16}\left(1 - u^2\right)^2 1\left(|u| \leq 1\right).$$

The weight functions used to calculate $\widehat{G}$, $\widehat{F}$ and $\widehat{H}$ are $w_2\left(x\right) = K_H\left(x\right)$, $w_1\left(z\right) = \left(1/2\right) K_H\left(z/2\right)$, and $w_H\left(x, z\right) = w_2\left(x\right) w_1\left(z\right)$ respectively. He also uses bandwidths $h_{11} = 6$, $h_{21} = 5$, and $h_H = 3.25$. He chose these bandwidths through Monte Carlo experimentation to approximately minimize the unweighted integrated mean-square errors of his estimators of G, F and H. The additional

---

[1]The computer code we wrote to implement Horowitz (2001) estimator, was not fast enough to conduct simulations on as large a scale as the previous experiment.

[2]In Horowitz (2001) notation: $F \equiv f_1$, $G \equiv f_2$, and $H \equiv G$, with $x^1 \equiv z$, $x^2 \equiv x$ and $v \equiv m$.

bandwidths his estimator needs are set using his suggested rule–of–thumb $h_{k2} = h_{k1}n^{-1/72}$ for $k = 1, 2$.

We implement our proposed estimator (NP1) for this design, using a second order Gaussian kernel as before, with $p_1 = 0$, $p_2 = 1$, and $p_* = 1$. We find the optimal bandwidths $h_1 = 0.925$, $h_2 = 2.5$ and $h_* = 0.2$ for this design by Monte Carlo experimentation as Horowitz (2001) did.

Figure 1 shows the standardized Q–Q plots of both set of estimators at different points in the interior of the support of each function. These points are sufficiently far from the boundary of the data to avoid boundary effects for both estimators. These plots are based on 300 replications. We observe that the normal approximation of our estimator for G and F are better than Horowitz's at the chosen points. Similar results (not presented) hold for other points in the interior of the support of (X, Z) for G and F. The normal approximation of our estimator for H is similar to Horowitz's for low values of m, while it outperforms Horowitz's for higher values.

Finally, Figure 2 displays a visualization of the resulting output of 5000 replications of a fourth design using only our procedure. Data are generated as before with the same G and F, but now using $H[m] = 1 + (16/7)\,m$, with $\sigma_r^2 = 1$. Other information is set accordingly, in particular $n = 400$, $h_1 = 0.15$, $h_2 = 0.7$, $p_1 = 3$, and $p_2 = 1$. The white plane and dashed lines represent medians of simulations and gray planes and dotted lines represent 90% simulation envelopes.

5.2 Generalized Homothetic Production Function Estimation

Let y be the log output of a firm and $(\widetilde{x}, z)$ be a vector of inputs. Going back at least as far as Shephard (1953) and Shephard (1970), 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 $H(m) = m$, see, e.g., Bairam (1994) and Chung (1994) for parametric models, and Tripathi and Kim (2003) and Tripathi (1998) for fully nonparametric options. Zellner and Ryu (1998) provide empirical comparisons of a large number of different homothetic production functional forms. In the nonparametric framework[3], Lewbel and Linton (2006) 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* Ψ_w *if*

---

[3]Other examples of nonparametric estimators include Varian (1984) and Primont and Primont (1994). Also, Hanoch and Rothschild (1972) provides a test to verify whether a homothetic production function exists that could, without statistical errors, generate a given data set. Although these papers do not assume a parametric form, by assumption they have no statistical errors. Consequently, they also have no associated distribution theory.

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* λ.

Definition 5.2 *A function* $r^* : \Psi_w \subset \Re^{d+1} \to \Re$ *is said to be generalized homothetic on* Ψ_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* Ψ_w.

Homogeneity of degree κ 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

$$r^*(\widetilde{x}, z) = H[M^*(\widetilde{x}, z)] = H\left[M^*(\widetilde{x}/z, 1) g(1/z)^{-1}\right]$$
$$= H[G(x) F(z)] = H[M(x, z)] \equiv r(x, z), \tag{5.3}$$

where $x = \widetilde{x}/z$ and $F(z) = 1/g(1/z)$. When H is assumed known and equal to the identity function, Tripathi and Kim (2003) and Tripathi (1998) use the assumption that $M(x, z)$ is a homogeneous function of degree one, i.e. $F(z) = 1/z$, to identify the model and achieve dimension reduction. Lewbel and Linton (2006) uses the same functional assumption regarding F but with an unknown strictly monotonic link function H. In contrast, our proposed estimator identifies M, G, F and H in models such as (5.3), i.e. $y = r(x, z) + \varepsilon_r$, without imposing any such parametric specification of F, instead exploiting the partial separability of M with respect to z and $f(z) > 0$[4]. Like these other estimators, our estimator reduces the dimensionality by one as described earlier.

We have constructed an `R` package, `JLLprod`, which can be freely downloaded from the author's websites. The package includes access to an Ecuadorian production data set[5], and to the Chinese data set described below. We use this software to estimate generalized homothetic production functions for four industries in mainland China in two time periods, 1995 and 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 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 $r(k, L) \in \mathcal{P}$, which is a set of smooth production functions, and $k = K/L$ as in (5.3). 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 and year. Both regressors are also normalized by their respective median prior to estimation.

---

[4]As $\partial g(\lambda) / \partial \lambda > 0$, and $\lambda = z^{-1}$, it follows that $F(z)$ is strictly increasing, i.e. $f(z) = \partial F(z) / \partial z > 0$ over its entire domain.

[5]This is data on 406 firms in the Petroleum, Chemical and Plastics industries in Ecuador in 2002.

5.2.1 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

$$r_{\psi_{P1}}(k, L) = \theta_0 + \theta_1 \ln(k) + \theta_2 \ln(L + \gamma) + \theta_3 [\ln(k)]^2$$
$$+ \theta_4 \ln(k) \ln(L + \gamma) + \theta_5 [\ln(L + \gamma)]^2, \tag{5.4}$$

and $\psi_{P1} = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \gamma)^\top$. When $\gamma = 0$, (5.4) reduces to the ordinary unrestricted Translog production function. When $2\theta_1\theta_5 - \theta_2\theta_4 = 0$ and $\theta_1^2\theta_5 - \theta_2^2\theta_3 = 0$, (5.4) is equivalent to the following generalized homothetic production function (P2) specification,

$$M(k, L) = k^\alpha (L + \gamma)$$
$$r_{\psi_{P2}}(k, L) = H(M) = \beta_0 + \beta_1 \ln(M) + \beta_2 [\ln(M)]^2, \tag{5.5}$$

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

$$M(k, L) = k^\alpha L$$
$$r_{\psi_{P3}}(k, L) = H(M) = \beta_0 + \beta_1 \ln(M) + \beta_2 [\ln(M)]^2, \tag{5.6}$$

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).

Figure 3 shows isoquants for P2 with $\psi_{P2} = (1/2, 10, 1/2, 1, \gamma)^\top$, where $\gamma = -1, 0, +1$. At any level of output, these isoquants are steeper at high levels of k for negative γ than for positive γ. However, as in the (P3) homothetic case of $\gamma = 0$, the slopes of these level surfaces are constant along rays through the origin. This important property of homothetic models is preserved by the generalized homothetic specification (P2).

Fitting these models by nonlinear least squares in each year yields the parameter estimates reported in Tables 3–5 (Heteroskedasticity robust standard errors are in parentheses).

Specification Test

Two sets of parametric restrictions are tested on model (5.4) for each year and industry. In order to assess whether model (5.4) may be further simplified by (5.5), $H_0 : 2\theta_1\theta_5 - \theta_2\theta_4 = 0; \theta_1^2\theta_5 - \theta_2^2\theta_3 = 0$ is tested by means of a Wald statistic, W_{12} which is distributed under H_0 as $\chi_{(2)}$. The further simplification (5.6) is also tested by a Wald statistic, W_{13}, which under the $H_0 : 2\theta_1\theta_5 - \theta_2\theta_4 = 0; \theta_1^2\theta_5 - \theta_2^2\theta_3 = 0; \gamma = 0$, is distributed as $\chi_{(3)}$. The results of these tests are presented below.

Industry	1995				2001			
	W_{12}	p-value	W_{13}	p-value	W_{12}	p-value	W_{13}	p-value
Chemical	1.280	0.527	2.244	0.523	17.286	0.000	1,095	0.000
Iron	8.834	0.012	14.261	0.003	2.272	0.321	2.343	0.504
Petroleum	1.790	0.409	3.076	0.380	0.791	0.673	0.813	0.846
Transportation	1.735	0.420	1.997	0.573	7.980	0.019	8.252	0.041

Models (5.5) and (5.6) appear to be valid parametric simplifications of the production function (5.4), except for the iron industry in 1995 and the chemical and transportation industries in 2001.

The suitability of the parametric Generalized Homothetic and Homothetic Translog production function fits, $r_{\widehat{\psi}_{P2}}(k, L)$ and $r_{\widehat{\psi}_{P3}}(k, L)$, in these industries may be also judged by the use of a residual based test. For this purpose, we employ the test proposed by Zheng (1996) for the hypothesis $H_0: r \in \overline{\mathcal{P}} \{r \in \mathcal{P} \mid r = r_{\psi_{Pl}} \text{ for some } \psi_{Pl}\}$. For $l = 2, 3$, their test statistics are given by

$$U_{Pl} = \frac{1}{\lambda^2 n^2} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \left(Y_i - r_{\widehat{\psi}_{Pl}}(k_i, L_i)\right) \left(Y_j - r_{\widehat{\psi}_{Pl}}(k_j, L_j)\right) K\left(\frac{k_j - k_i}{\lambda}\right) K\left(\frac{L_j - L_i}{\lambda}\right), \quad (5.7)$$

with kernel $K(\cdot)$, the Gaussian kernel here, and bandwidth λ, set equal to h_1 in all cases. Given standard regularity conditions, under the null hypothesis that the parametric specification is correct,

$$n\lambda U_{Pl} \sim N\left(0, 2 \int K^2(u)\, du \int \left[\sigma_r^2(k, L)\, p(k, L)\right]^2 dk dL\right). \quad (5.8)$$

Replacing integrals by sums and unknown functions by their nonparametric estimates in (5.7) and (5.8), we obtain the following test results:

Industry	1995			2001		
	λ	U_{P2}	p-value	λ	U_{P2}	p-value
Chemical	5.125	-0.7698	0.7793	2.25	-0.4728	0.6818
Iron	4.250	-0.7532	0.7743	4	-0.7367	0.7693
Petroleum	2.750	-0.8117	0.7915	11	-0.7325	0.7681
Transportation	1.750	-0.6519	0.7428	4.37	-0.7124	0.7619

	λ	U_{P3}	p-value	λ	U_{P3}	p-value
Chemical	5.125	-0.7721	0.7800	2.25	-0.4130	0.6602
Iron	4.250	-0.7195	0.7641	4	-0.7417	0.7709
Petroleum	2.750	-0.8088	0.7907	11	-0.7327	0.7681
Transportation	1.750	-0.6503	0.7422	4.37	-0.7065	0.7601

We fail to reject both H_0 for all industries in both years at any level of significance. In all cases, test results are not altered by the choice of smoothing parameter λ. Both sets of tests justify the use of both models as sensible parametric simplifications of the data[6] against which we may compare our more flexible specification. Other kernel–based specification tests that could be employed include Bierens (1990), Härdle and Mammen (1993), Gozalo (1993) and Horowitz and Spokoiny (2001).

5.2.2 Nonparametric Modeling

Figures 4 to 11 show generalized homothetic nonparametric estimates $\widehat{M}(k, L)$, $\widehat{G}(k)$, $\widehat{F}(L)$ and $\widehat{H}(M)$ for both years. For each industry and year, we use local quadratic regression with a Gaussian kernel and bandwidths h_1 given 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	1995			2001		
	n	$\ln L_0$	r_0	n	$\ln L_0$	r_0
Chemical	1560	3.40	7	1637	3.06	7.0
Iron	376	-0.37	7	341	4.06	8.0
Petroleum	93	2.73	7	119	2.27	8.5
Transportation	989	3.44	7	1230	4.04	7.5

The nonparametric fits of the generalized homogeneous component, $\widehat{M}$, shown in Figures 4 and 8, 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[7]. Nonparametric estimates of the functions G and F differ from the parametric Translog model estimates (P3) in Figures 5, 6, 9 and 10[8], but they are roughly similar to the parametric generalized homothetic model (P2) at low levels of L. 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 F in 1995 and G in 2001 for most industries. Comparing the nonparametric estimator of F in Figures 6 and 10 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 chemical and transportation industry in 1995 and petroleum and transportation industries in 2001. In any case, they are all strictly increasing functions in labor, implying a generalized

---

[6]Although the appropriateness of these parametric models may change through time, see Konishi and Nishiyama (2002).

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

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

homogeneous structure for M as conjectured. Figures 7 and 11 show 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. They also show 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 and years, 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 in 2001 for the chemical and iron industries.

Specification Test

We are interested in testing our proposed nonparametric generalized homothetic specification against a general unrestricted nonparametric regression alternative, that is $H_0 : r \in \overline{\mathcal{P}}\{r \in \mathcal{P} \,|\, r = H[G(k) F(L)]$ for some H, G and $F\}$. Given $\widehat{H}$, $\widehat{G}$ and $\widehat{F}$, the implied restricted estimator of the regression surface is $\widehat{r}(k, L) = \widehat{H}[\widehat{G}(k) \widehat{F}(L)]$. As before, we employ a U–statistic based test as suggested in Fan and Li (1996). That is,

$$U_{NP} = \frac{1}{\lambda^2 n^2} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} (Y_i - \widehat{r}(k_i, L_i)) (Y_j - \widehat{r}(k_j, L_j)) K\left(\frac{k_j - k_i}{\lambda}\right) K\left(\frac{L_j - L_i}{\lambda}\right),$$

which under the null hypothesis that the generalized homothetic specification proposed in this paper is correct,

$$n\lambda U_{NP} \sim N\left(0, 2 \int K^2(u)\, du \int \left[\sigma_r^2(k, L)\, p(k, L)\right]^2 dk dL\right).$$

The results are as follows:

Industry	1995			2001		
	λ	U_{NP}	p-value	λ	U_{NP}	p-value
Chemical	5.125	-0.7702	0.7794	2.25	-0.1448	0.5575
Iron	4.250	-0.7578	0.7757	4	-0.2274	0.5899
Petroleum	2.750	-0.8458	0.8012	11	-0.7334	0.7684
Transportation	1.750	-0.6472	0.7413	4.37	-0.7047	0.7595

As in the parametric case, at all levels of significance we fail to reject the hypothesis that our specification is a correct nonparametric simplification of the data for all industries and years.

Substitutability and Returns to Scale

Important properties of production functions are measures of substitutability of inputs and returns to scale. A standard measure of the substitutability of inputs for production is the Technical Rate of Substitution, σ^*, defined as the slope of the isoquants in Figure 3, that is, $\sigma^* = -\left.(dK/dL)\right|_{r(k,L)=r}$, for some constant level of output r. For an arbitrary production function $E[Y|k,L] = r(k,L)$, an equivalent measure of input substitutability that is more convenient for our empirical application is

$$
\begin{aligned}
T(k,L) &\equiv \ln\left(\frac{\sigma^*}{k}\right) \\
&= \ln\left(\frac{\partial r(k,L)}{\partial \ln L}\right) - \ln\left(\frac{\partial r(k,L)}{\partial \ln k}\right).
\end{aligned}
\tag{5.9}
$$

If $r(k,L)$ is generalized homothetic (NP), so $r(k,L) = H[M(k,L)]$, then $T(k,L) = \ln(\partial \ln F(L)/\partial \ln L) - \ln(\partial \ln G(k)/\partial \ln k)$. For the parametric generalized homothetic model (P2), this measure simplifies to $T(k,L) = \ln(L/(L+\gamma)) - \ln(\alpha)$, and for the Homothetic Translog model (P3) $T(k,L) = -\ln(\alpha)$, a constant.

To estimate $T(k,L)$ in the nonparametric model (NP) we use the approximation $\widehat{T}(k,L) = \ln[(\ln \widehat{F}(L_j) - \ln \widehat{F}(L_{j-1}))/(\ln L_j - \ln L_{j-1})] - \ln[(\ln \widehat{G}(k_j) - \ln \widehat{G}(k_{j-1}))/(\ln k_j - \ln k_{j-1})]$ after ordering the estimation grid points j, and approximations for (P2) and (P3) are obtained by replacing unknown quantities with their parametric estimates. Table 6 provides their averages along with standard deviations in parentheses.

Another property of production that is empirically important is economies of scale, defined as $\varepsilon^*(K,L) = \left.(\partial r^*(cK,cL)/\partial \ln c)\right|_{c=1}$, which by (5.3), simplifies to $\varepsilon(k,L) = \partial r(k,L)/\partial \ln L$. If $r(k,L)$ is generalized homothetic, then $\varepsilon(k,L) = RTS(M(k,L),L)$ (Returns to Scale), where

$$
RTS(M,L) = \frac{\partial H(M)}{\partial \ln M}\frac{\partial \ln F(L)}{\partial \ln L}.
\tag{5.10}
$$

For model (P2), $RTS(M,L) = [\beta_1 + 2\beta_2 \ln(M)](L/(L+\gamma))$, and $RTS(M,L) = \beta_1 + 2\beta_2 \ln(M)$ for the Translog model (P3). These were calculated by replacing the unknown parameters with their respective parametric estimates. In the nonparametric model this measure is estimated as $\widehat{RTS}(M,L) = [(\widehat{H}(\widehat{M}_j) - \widehat{H}(\widehat{M}_{j-1}))/(\ln \widehat{M}_j - \ln \widehat{M}_{j-1})] \times [(\ln \widehat{F}(L_j) - \ln \widehat{F}(L_{j-1}))/(\ln L_j - \ln L_{j-1})]$, using the same ordering as before. Table 7 provides summary statistics for all four industries in both years.

Calculating these measures in our data set generates mixed results. The parametric model estimates have similar average $T(k,L)$ in each year, but differ from the nonparametric model estimates, which show a sizeable increase in 2001 relative to their values in 1995 for all sectors but the petroleum industry (which may be caused by the small number of observations for this industry in 1995, $n = 93$). This industry is also the only one for which the three averages coincide in 2001, because of the closeness of the parametric models to the nonparametric fit in Figures 9 and 10. However, all models show a reduction in economies of scales for all industries from 1995

to 2001. Although average increasing returns to scale, $RTS(M, L) > 1$, are predicted for some sectors in 1995, no industries show increasing returns in 2001. The chemical and iron industries seem to have decreasing returns to scale, $RTS(M, L) < 1$, in 2001, while the remaining sectors report approximately constant returns to scale, $RTS(M, L) \simeq 1$ in 2001.

Economically, homotheticity (P3) generalizes the idea that pure economic profit will be zero. Since this situation is descriptive of the long–run equilibrium under perfect competition, a possible explanation for the observed decrease in returns to scale over time may be the substantial ownership reform during this period. Many more firms in the Chinese industrial sectors were state–owned in 1995 than in 2001, resulting in substantial restructuring of these industries and increased exposure to competitive pressures. This would explain the move towards constant returns to scale in 2001, and coupled with the hetereogenity regarding different firms specializing in different products in each industry may can create the appearance of decreasing returns on average for some industries in 2001. The changes in production functions over time may more generally be due to changes in technology, demand, and other aspects of China's increasing economic liberalization and growth over this time period.

On caveat regarding these estimates is that they assume inputs are exogenously determined. Firms with positive productivity shocks may respond by using more inputs, resulting in endogeneity. More generally, when returns to scale are not constant, selection of input levels will depend in part on the profit associated with different output levels, again inducing potential endogeneity. In the next section, we discuss how our estimator could be modified to deal with this potential problem in a more general framework.

6 Conclusion and Extensions

We have provided a general nonparametric estimator for a transformed partly additive or multiplicatively separable model. Its small sample properties were analyzed in some Monte Carlo experiments, and found to compare favorably with 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, and tested its validity. We now conclude by describing some extensions.

Additional Regressors

Consider identification of $G(x)$ and $F(z)$ in the model $r(x, z, w) = H[M(x, z), w] \equiv H[G(x) + F(z), w]$, where H is strictly monotonic on its first element, $(x, z) \in \Re^{d+1}$, and $w \in \Psi_W \subseteq \Re^{d_w}$ is a vector of additional regressors. It is straightforward to extend Theorem 2.1 or Corollary 2.1

in these cases. Specifically, let $s(x, z, w) \equiv \partial r(x, z, w) / \partial z$, and define the function $q(t, z, w)$ by $q(t, z, w) = E[s(X, Z, W) | r(X, Z, W) = t, Z = z, W = w]$. Then, the desired identification is achieved by replacing $q(t, z_0)$ in (2.2) or (2.3) by $q(t, z_0, w)$. For example, in the additive case, $s(x, z, w) = h[M(x, z), w]$, where h now represents the first derivative of H with respect to its first argument, and consequently $q(r, z) = h[H^{-1}(r, w), w] f(z)$. It follows that

$$
\begin{aligned}
\int_{r_{0,w}}^{r(x,z,w)} \frac{dt}{q(t, z_0, w)} &= \int_{r_{0,w}}^{r(x,z,w)} \frac{dt}{h[H^{-1}(t, w), w] f(z_0)} \\
&= \int_{H^{-1}[r_{0,w}, w]}^{H^{-1}[r(x,z,w), w]} \frac{h(m, w) \, dm}{h(m, w) f(z_0)} \\
&= \left(H^{-1}[r(x, z, w), w] - H^{-1}[r_{0,w}, w]\right)(1 / f(z_0)) \\
&= \left(H^{-1}[H[M(x, z), w], w]\right) \equiv M(x, z),
\end{aligned}
$$

where the second equality follows from the change of variables $m = H^{-1}(t, w)$, so $dt = h(m, w) \, dm$, and the last equality follows after assuming that $f(z_0) = 1$ and that $r_0 = H[0, w]$ for all w. This result holds for all $w \in \Psi_W$ and (x, z), so it holds in expectation replacing w with $\widetilde{W}$, thereby yielding

$$
M(x, z) = E\left[\int_{r_{0,\widetilde{W}}}^{r(x,z,\widetilde{W})} \frac{dt}{q(t, z_0, \widetilde{W})}\right].
$$

A consistent estimator of $M(x, z)$ and therefore, by virtue of marginal integration, of $G(x)$ and $F(z)$, is then given by

$$
\widehat{M}(x, z) = \frac{1}{n} \sum_{i=1}^{n} \left[\int_{r_0}^{\widehat{r}(x,z,W_i)} \frac{dt}{\widehat{q}(t, z_0, W_i)}\right], \tag{6.1}
$$

and a consistent estimator of h is then given by a nonparametric regression of $\widehat{r}(x, z, w)$ on $(\widetilde{M}(x, z), w)$, as before. The asymptotic properties of these estimators can be analyzed using similar tools as in Lewbel and Linton (2006).

Endogenous Regressors

Consider estimation of $M(x, z) \equiv G(x) + F(z)$ in the model $y = H^*[M(x, z), \varepsilon]$ where ε is now unobserved and H^* is strictly monotonic in its first argument. If ε is independent of (X, Z), then $r(x, z) = E[Y | X = x, Z = z] = H[M(x, z)]$, and our estimator can be applied. However, when some of the covariates (X, Z) are endogenous, and so correlated with ε, estimation of $M(x, z)$ is still possible, under the following conditions. For an observed vector T of exoge-

nous covariates[9], which may include exogenous elements of (X, Z), define $m_x(t) = E[X|T = t]$, $U_x = X - m_x(t)$, $m_z(t) = E[Z|T = t]$, $U_z = Z - m_z(t)$ and let $U = (U_x, U_z)$. Then by construction $\varepsilon | X, Z, T \sim \varepsilon | U, T$. Define $r(x, z, u) \equiv E[Y|X = x, Z = z, U = u]$ and $H[M(x, z), u] = E[H^*[M(x, z), \varepsilon]|X = x, Z = z, T = t, U = u]$. If we then assume that

$$\varepsilon | U, T \sim \varepsilon | U, \tag{6.2}$$

which is the form of endogeneity analyzed in the control function models of Blundell and Powell (2003)[10], it then follows that $r(x, z, u) = H[M(x, z), u]$. If U were observed, then the estimator proposed in (6.1) could be employed by redefining W as U. Otherwise, U must be estimated, that is, first estimate $\widehat{m}_x(T_i)$ and $\widehat{m}_z(T_i)$ by nonparametric regressions of X and Z on T respectively. Then compute $\widehat{r}$ as a nonparametric regression of Y on $(X, Z, \widehat{U}_i)$, and construct (6.1) by replacing W_i everywhere with $\widehat{U}_i = (\widehat{U}_{x,i}, \widehat{U}_{z,i})$, where $\widehat{U}_{x,i} = X_i - \widehat{m}_x(T_i)$, $\widehat{U}_{z,i} = Z_i - \widehat{m}_z(T_i)$. Consistency of the resulting estimator of the functions M and H will follow from uniform consistency of the nonparametric estimators involved.

Once $M(x, z)$ is estimated in this way, it can be treated as an observed endogenous regressor, and estimation of H^* (or any identifiable functional of H^* that is of applied interest) then reduces to estimation of a nonparametric triangular system. Examples of estimators of such systems are Blundell and Powell (2003), Imbens and Newey (2002) and Chesher (2001).

Further Testing

In production theory, homotheticity of production functions can be assessed by comparing the estimated component $\widehat{F}$ with the parametric model, $F(L) = L$. The assumed separability may also be tested by comparing the unrestricted nonparametric estimator $\widehat{r}(k, L)$ with the implied estimator for r given by the proposed structure, that is, $\widehat{r}(\widehat{M}(k, L))$. Such tests can be performed as in Gozalo and Linton (2001), by using asymptotic critical values or by direct implementation of their bootstrap procedure. Nonetheless, their theoretical justification in our framework would require considerable further work, and so remains a topic of future research.

---

[9]In production theory, they could include investment as in Olley and Pakes (1996), or intermediate inputs as suggested by Levinsohn and Petrin (2003).

[10]Assumption (6.2) also yields a nonparametric triangular system similar to Newey, Powell, and Vella (1999) and Imbens and Newey (2002)

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.

BIERENS, H. J. (1990): "A Consistent Conditional Moment Test of Functional Form," *Econometrica*, 58(6), 1443–1458.

BILLINGSLEY, P. (1986): *Probability and Measure*, Wiley series in probability and mathematical statistics. Wiley, 2 edn.

BLUNDELL, R., AND J. L. POWELL (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econonometrics: Theory and Applications*, ed. by M. Dewatripont, L. Hansen, and S. Turnovsky, vol. II of *Eighth World Congress*. Cambridge University Press, Cambridge.

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

CHESHER, A. (2001): "Quantile Driven Identification of Structural Derivatives," Working Paper CWP08/01, Cemmap.

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.

FAN, Y., AND Q. LI (1996): "Consistent Model Specification Tests: Omitted Variables and Semi-parametric Functional Forms," *Econometrica*, 64(4), 865 – 890.

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.

GOZALO, P. L. (1993): "A Consistent Model Specification Test for Nonparametric Estimation of Regression Function Models," *Econometric Theory*, 9(3), 451–477.

GOZALO, P. L., AND O. B. LINTON (2001): "Testing Additivity in Generalized Nonparametric Regression Models with Estimated Parameters," *Journal of Econometrics*, 104(1), 1–48.

HANOCH, G., AND M. ROTHSCHILD (1972): "Testing the Assumptions of Production Theory: A Nonparametric Approach," *Journal of Political Economy*, 80, 256–275.

HÄRDLE, W. (1990): *Applied Nonparametric Regression*, vol. 9 of *Econometric Society Monographs*. Cambridge University Press, 1 edn.

HÄRDLE, W., AND E. MAMMEN (1993): "Comparing Nonparametric versus Parametric Regression Fits," *Annals of Statistics*, 21(4), 1926–1947.

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.

HOROWITZ, J. L., AND V. G. SPOKOINY (2001): "An Adaptive, Rate-Optimal Test of a Parametric Model Against a Nonparametric Alternative," *Econometrica*, 69(3), 599–631.

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

IMBENS, G. W., AND W. K. NEWEY (2002): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," Working Paper 285, NBER.

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.

KONISHI, Y., AND Y. NISHIYAMA (2002): "Nonparametric Test for Translog Specification of Production Function in Japanese Manufacturing Industry," in *Integrated Assessment and Decision Support, Proceedings of the First Biennial Meeting of the International Environmental Modelling and Software Society*, ed. by A. E. Rizzoli, and A. J. Jakeman, vol. 2, pp. 597–602. iEMSs.

LEVINSOHN, J., AND A. PETRIN (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70(2), 317–341.

LEWBEL, A., AND O. B. LINTON (1999): "Nonparametric Censored and Truncated Regression," Sticerd Working Paper EM/00/389, STICERD.

——— (2002): "Nonparametric Censored and Truncated Regression," *Econometrica*, 70(2), 765–779.

——— (2006): "Matching Estimation and Efficient Estimators of Nonparametric Homothetically Separable Functions," Unpublished Manuscript.

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.

MCCULLAGH, P., AND J. A. NELDER (1989): *Generalized Linear Models*, vol. 37 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London, 2 edn.

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

NEWEY, W. K., J. L. POWELL, AND F. VELLA (1999): "Nonparametric Estimation of Triangular Simultaneous Equation Models," *Econometrica*, 67, 565–603.

OLLEY, S., AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunication Equipment Industry," *Econometrica*, 64(6), 1263–1298.

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

POWELL, J. L., J. H. STOCK, AND T. M. STOKER (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57(6), 1403–1430.

PRIMONT, D., AND D. PRIMONT (1994): "Homothetic Non-parametric Production Models," *Economics Letters*, 45, 191–195.

SHEPHARD, R. W. (1953): *Cost and Production Functions*. Princeton University Press, New Jersey.

——— (1970): *Theory of Cost and Production Functions*. Princeton University Press, New Jersey.

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 (2004): "Local Polynomial Estimation of Nonparametric Simultaneous Equation Models," Unpublished Manuscript.

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

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.

VARIAN, H. R. (1984): "The Nonparametric Approach to Production Analysis," *Econometrica*, 52, 579–597.

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.

ZELLNER, A., AND H. RYU (1998): "Alternative Functional Forms for Production, Cost and Returns to scale Functions," *Journal of Applied Econometrics*, 13, 101–127.

ZHENG, J. X. (1996): "A Consistent Test of Functional Form via Nonparametric Estimation Techniques," *Journal of Econometrics*, 75(2), 263–289.

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

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

where the first element in θ denotes the minimizing intercept of (A-1), θ_0, and

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

We also use the following conventions:

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

$$a^{\mathbf{j}} = a_1^{j_1} \times \ldots \times a_d^{j_d} \times a_{d+1}^{j_{d+1}}$$

$$\sum_{0 \le |\mathbf{j}| \le p_1} = \sum_{k=0}^{p_1} \sum_{\substack{j_1=0 \\ j_1 + \ldots + j_d + j_{d+1} = k}}^k \cdots \sum_{j_d=0}^k \sum_{j_{d+1}=0}^k$$

where $w = \left(x^\top, z\right)^\top$. Let $N_{r,(l)} = \left(l + k - 1\right)! / \left(l! \left(k - 1\right)!\right)$ 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 ϕ_l^{-1} denote this one-to-one map. For each $\mathbf{j}$ with $0 \le |\mathbf{j}| \le 2p_1$, let

$$\mu_{\mathbf{j}}\left(K_1\right) = \int_{\Re^{d+1}} u^{\mathbf{j}} K_1\left(u\right) du,$$

$$\gamma_{\mathbf{j}}\left(K_1\right) = \int_{\Re^{d+1}} u^{\mathbf{j}} K_1^2\left(u\right) du,$$

$$\gamma_{\mathbf{k},\mathbf{l}}^1\left(K_1\right) = \int_{\Re^d} \int_{\Re} \int_{\Re} \left(u_d, u_1\right)^{\mathbf{k}} \left(u_d, \widetilde{u}_1\right)^{\mathbf{l}} K_1\left(u_d, u_1\right) K_1\left(u_d, \widetilde{u}_1\right) du_1 d\widetilde{u}_1, \text{ and}$$

$$\gamma_{\mathbf{k},\mathbf{l}}^2\left(K_1\right) = \int_{\Re} \int_{\Re^d} \int_{\Re^d} \left(u_d, u_1\right)^{\mathbf{k}} \left(\widetilde{u}_d, u_1\right)^{\mathbf{l}} K_1\left(u_d, u_1\right) K_1\left(\widetilde{u}_d, u_1\right) du_d d\widetilde{u}_d,$$

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 $\boldsymbol{\Gamma}_r$, and the $N_r \times N_{r,(p_1+1)}$ matrix $\mathbf{B}_r$ by

$$
\mathbf{M}_r = \left[\begin{array}{cccc}
\mathbf{M}_{r;0,0} & \mathbf{M}_{r;0,1} & \cdots & \mathbf{M}_{r;0,p_1} \\
\mathbf{M}_{r;1,0} & \mathbf{M}_{r;1,1} & \cdots & \mathbf{M}_{r;1,p_1} \\
\vdots & \vdots & & \vdots \\
\mathbf{M}_{r;p_1,0} & \mathbf{M}_{r;p_1,1} & \cdots & \mathbf{M}_{r;p_1,p_1}
\end{array} \right],
$$

$$
\boldsymbol{\Gamma}_r = \left[\begin{array}{cccc}
\boldsymbol{\Gamma}_{r;0,0} & \boldsymbol{\Gamma}_{r;0,1} & \cdots & \boldsymbol{\Gamma}_{r;0,p_1} \\
\boldsymbol{\Gamma}_{r;1,0} & \boldsymbol{\Gamma}_{r;1,1} & \cdots & \boldsymbol{\Gamma}_{r;1,p_1} \\
\vdots & \vdots & & \vdots \\
\boldsymbol{\Gamma}_{r;p_1,0} & \boldsymbol{\Gamma}_{r;p_1,1} & \cdots & \boldsymbol{\Gamma}_{r;p_1,p_1}
\end{array} \right], \mathbf{B}_r = \left[\begin{array}{c}
\mathbf{M}_{r;0,p_1+1} \\
\mathbf{M}_{r;1,p_1+1} \\
\vdots \\
\mathbf{M}_{r;p_1,p_1+1}
\end{array} \right] \tag{A-2}
$$

where $N_r = \sum_{l=0}^{p_1} N_{r,(l)}$, $\mathbf{M}_{r;i,j}$ and $\boldsymbol{\Gamma}_{r;i,j}$ 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. $\boldsymbol{\Gamma}_r^1$ and $\boldsymbol{\Gamma}_r^2$ are defined similarly by the $N_{r,(i)} \times N_{r,(j)}$ matrices $\boldsymbol{\Gamma}_{r;i,j}^1$, $\boldsymbol{\Gamma}_{r;i,j}^2$, whose (l,m) elements are given by $\gamma_{\phi_i(l),\phi_j(m)}^1$ and $\gamma_{\phi_i(l),\phi_j(m)}^2$ 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 ϕ_l^{-1} denote this one-to-one map. For each $\mathbf{j}$ with $0 \le |\mathbf{j}| \le 2p_2$, let

$$
\mu_{\mathbf{j}}(K_2) = \int_{\Re^2} u^{\mathbf{j}} K_2(u)\, du, \text{ and}
$$

$$
\gamma_{\mathbf{j}}(K_2) = \int_{\Re^2} u^{\mathbf{j}} K_2^2(u)\, du.
$$

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

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

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

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

$$
\mathcal{K}_{2,i}(v) = \begin{bmatrix}
\mathcal{K}_{2,i;0}(v) \\
\mathcal{K}_{2,i;1}(v) \\
\vdots \\
\mathcal{K}_{2,i;p_2}(v)
\end{bmatrix}, \mathcal{K}_{2,i}^{(1)}(v) = \begin{bmatrix}
\mathcal{K}_{2,i;0}^{(1)}(v) \\
\mathcal{K}_{2,i;1}^{(1)}(v) \\
\vdots \\
\mathcal{K}_{2,i;p_2}^{(1)}(v)
\end{bmatrix} \tag{A-4}
$$

$$
\mathbf{M}_{q,n}(v) = \begin{bmatrix}
\mathbf{M}_{q,n;0,0}(v) & \mathbf{M}_{q,n;0,1}(v) & \cdots & \mathbf{M}_{q,n;0,p_2}(v) \\
\mathbf{M}_{q,n;1,0}(v) & \mathbf{M}_{q,n;1,1}(v) & \cdots & \mathbf{M}_{q,n;1,p_2}(v) \\
\vdots & \vdots & & \vdots \\
\mathbf{M}_{q,n;p_2,0}(v) & \mathbf{M}_{q,n;p_2,1}(v) & \cdots & \mathbf{M}_{q,n;p_2,p_2}(v)
\end{bmatrix},
$$

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 $\zeta(t, z) = \partial [f_V(t, z) q^2(t, z)]^{-1} / \partial t$, which are well defined given

33

Assumptions (E1) and (E2). Thus, by integration by parts, it follows that

$$\int\limits_{r_0}^{r(x,z)} h_2^{-1} \mathcal{K}_{2,i}^{(1)}(t,z) \left[f_V(t,z) q^2(t,z) \right]^{-1} dt = \left\{ \mathcal{K}_{2,i}(r,z) \left[f_V(r,z) q^2(r,z) \right]^{-1} - \mathcal{K}_{2,i}(r_0,z) \times \right.$$

$$\left. \left[f_V(r_0,z) q^2(r_0,z) \right]^{-1} \right\} - \int\limits_{r_0}^{r(x,z)} \mathcal{K}_{2,i}(t,z) \zeta(t,z) dt$$

$$\equiv \varrho_{i,1}^0 - \varrho_{i,2}^0. \tag{A-5}$$

Similarly, let us define $dQ(t) = 1(r_0 \le t \le r(x,z)) dt$, 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

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

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

$$\widehat{r}(w) - r(w) = \iota_1^\top \left[\mathbf{M}_r f(w) \right]^{-1} \left\{ 1 + o_p(1) \right\}$$

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

$$\widehat{s}(w) - s(w) = h_1^{-1} \iota_1^{*\top} \left[\mathbf{M}_r f(w) \right]^{-1} \left\{ 1 + o_p(1) \right\}$$

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

uniformly in w, where

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

$$\times \int_0^1 \left\{ D^{\mathbf{k}} r(w + \tau(W_i - w)) - D^{\mathbf{k}} r(w) \right\} (1 - \tau)^{p_1} d\tau.$$

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

$$\sup_w |\gamma(w)| = o(h_1^{p_1+1}),$$

$$\sup_w |h_1^{-(p_1+1)} \gamma_n(w) - \gamma(w)| = h_1^{p_1+1} O_p(n^{-1/2} h_1^{-(d+1)/2} \sqrt{\ln n}). \tag{A-8}$$

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

$$\sup_w |h_1^{-(p_1+1)} \beta_n(w) - \beta(w)| = O_p(n^{-1/2} h_1^{-(d+1)/2} \sqrt{\ln n}). \tag{A-9}$$

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

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

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 $h\left[H^{-1}\left(t\right)\right]=h\left(m\right)$ and $dt=h\left(m\right)dm$, we obtain

$$
\int_{r_1}^{r(x,z)}\frac{dt}{q(t,z_0)}=\int_{r_1}^{r(x,z)}\frac{F\left(z_0\right)}{h\left[H^{-1}\left(t\right)\right]H^{-1}\left(t\right)f\left(z_0\right)}dt
$$

$$
=\int_{H^{-1}(r_1)}^{H^{-1}(r(x,z))}\frac{F\left(z_0\right)}{h\left(m\right)mf\left(z_0\right)}h\left(m\right)dm
$$

$$
=\left[\frac{F\left(z_0\right)}{f\left(z_0\right)}\right]\left[\ln\left(H^{-1}\left[r\left(x,z\right)\right]\right)-\ln\left(H^{-1}\left[r_1\right]\right)\right]
$$

$$
=\ln\left(M\left(x,z\right)\right)\equiv\ln\left(G\left(x\right)F\left(z\right)\right).
$$

This proves the result.

Proof of Theorem 4.1

Rearranging terms, we have

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

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

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

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

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

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

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

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

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

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

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

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

$$
\widehat{V}_{q,n}(v) = n^{-1}h^{-2}\sum_{i=1}^{n}\widehat{\mathcal{K}}_{2,i}(v)\,\varepsilon_{q,i},
$$

$$
\widehat{V}_{q,n}^*(v) = n^{-1}h^{-2}\sum_{i=1}^{n}\widehat{\mathcal{K}}_{2,i}(v)\,[\widehat{S}_i - S_i],
$$

$$
\widehat{B}_{q,n}(v) = n^{-1}h^{-2}\sum_{i=1}^{n}\widehat{\mathcal{K}}_{2,i}(v)\,\widehat{\Delta}_{q,i}(v),\text{ and}
$$

$$
\widehat{\Delta}_{q,i}(v) \equiv q(\widehat{V}_i) - \sum_{0\leq|\mathbf{m}|\leq p_2}\frac{1}{\mathbf{m}!}(D^{\mathbf{m}}q)(v)(\widehat{V}_i - v)^{\mathbf{m}}.
$$

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 $dQ(t) = 1\,(r_0 \leq t \leq r(x,z))\,dt$. 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

B-1 to B-4, from which we conclude that

$$
\mathcal{M}_{2n}\left(x,z\right) = h_1^{p_1+1}\iota_2^\top \mathbf{M}_q^{-1}\mathbf{M}_{q,0}^0\iota_1^\top \mathbf{M}_r^{-1}\mathbf{B}_r\left[\frac{E\left[r^{(p_1+1)}(X,Z)g_q\left(X,Z\right)\middle|\,r\left(X,z_0\right)=r,Z=z_0\right]}{q^2\left(r,z_0\right)}\right.
$$

$$
\left.-\frac{E\left[r^{(p_1+1)}(X,Z)g_q\left(X,Z\right)\middle|\,r\left(X,z_0\right)=r_0,Z=z_0\right]}{q^2\left(r_0,z_0\right)}\right]
$$

$$
+ h_1^{p_1}h_2\iota_2^\top \mathbf{M}_q^{-1}\mathbf{M}_{q,0}^1\iota_1^\top \mathbf{M}_r^{-1}\mathbf{B}_r\left[\frac{E\left[r^{(p_1+1)}(X,Z)\middle|\,r\left(X,z_0\right)=r,Z=z_0\right]}{q^2\left(r,z_0\right)}\right.
$$

$$
\left.-\frac{E\left[r^{(p_1+1)}(X,Z)\middle|\,r\left(X,z_0\right)=r_0,Z=z_0\right]}{q^2\left(r_0,z_0\right)}\right]
$$

$$
+ h_2^{p_2+1}\iota_2^\top \mathbf{M}_q^{-1}\mathbf{B}_q\int_{r_0}^{r(x,z)}\frac{q^{(p_2+1)}\left(t,z_0\right)}{q^2\left(t,z_0\right)}dt + o_p(n^{-1/2}h_1^{-(d+1)/2})
$$

$$
= h_1^{p_1+1}\mathcal{B}_1\left(x,z\right) + h_1^{p_1}h_2\mathcal{B}_2\left(x,z\right) + h_2^{p_2+1}\mathcal{B}_3\left(x,z\right) + o_p(n^{-1/2}h_1^{-(d+1)/2}). \tag{A-14}
$$

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

Proof of Theorem 4.2

As before, we can write

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

where

$$\widetilde{V}_{H,n}(m) \equiv \widehat{V}_{H,n}(m) + \widehat{V}_{H,n}^*(m),$$

$$\widehat{V}_{H,n}(m) = n^{-1}h_*^{-1}\sum_{i=1}^{n}\widehat{\mathcal{K}}_{*,i}(m)\varepsilon_{r,i},$$

$$\widehat{V}_{H,n}^*(m) = n^{-1}h_*^{-1}\sum_{i=1}^{n}\widehat{\mathcal{K}}_{*,i}(m)[H(M_i) - H(\widetilde{M_i})], \text{ and}$$

$$\widehat{B}_{H,n}(m) = n^{-1}h_*^{-1}\sum_{i=1}^{n}\widehat{\mathcal{K}}_{*,i}(m)\widehat{\Delta}_{H,i}(m), \text{ with}$$

$$\widehat{\Delta}_{H,i}(m) \equiv H(\widehat{M_i}) - \sum_{0\leq j\leq p_*}\frac{1}{j!}(\partial^j H(m)/\partial m^j)(\widehat{M_i} - m)^j.$$

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

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

$$\mathcal{B}_{H3}(m) \equiv h_*^{p_*+1}\iota_*^\top \mathbf{M}_H^{-1}\mathbf{B}_H H^{(p_*+1)}(m),$$

with $\beta(w) \equiv \int \mathcal{B}(x,z)\,dP_1(z) + \int \mathcal{B}(x,z)\,dP_2(x) + \int\int \mathcal{B}(x,z)\,dP_1(z)\,dP_2(x)$ which is $O(h_\dagger)$ by construction. By defining $\mathcal{B}_H(m) \equiv \mathcal{B}_{H2}(m) + \mathcal{B}_{H3}(m)$, the proof is completed.

Appendix B: Technical Lemmas

Lemma B-1 *Under Assumption E, we have*

$$\sup_{t,z}|T_{q,n,1}(t,z)| = O_p\left(h_2^{-1}\nu_{1n} + n^{-1/2}h_2^{-1}\sqrt{\ln n}\right), \text{ and} \tag{B-1}$$

$$\int_{r_0}^{r(x,z)} T_{q,n,1}(t,z)\,dt = O_p\left(n^{-1/2}h_1^{-(d+1)/2}\right). \tag{B-2}$$

Proof. We may rewrite

$$\widehat{V}_{q,n}(t,z) = nh_2^{-2}\sum_{i=1}^{n}\mathcal{K}_{2,i}(v)\varepsilon_{q,i} + nh_2^{-2}\sum_{i=1}^{n}[\widehat{\mathcal{K}}_{2,i}(v) - \mathcal{K}_{2,i}(v)]\varepsilon_{q,i}.$$

Then, by Theorem 5 (page 593) in Masry (1996a), it follows that the first term is

$$nh_2^{-2}\sum_{i=1}^{n}\mathcal{K}_{2,i}(v)\varepsilon_{q,i} = O_p\left(n^{-1/2}h_2^{-1}\sqrt{\ln n}\right).$$

In the other hand, after a Taylor–series expansion, the second term is bounded by

$$nh_2^{-2}\sum_{i=1}^{n}[\widehat{\mathcal{K}}_{2,i}\left(v\right)-\mathcal{K}_{2,i}\left(v\right)]\varepsilon_{q,i}$$

$$\leq\{n^{-1}h_2^{-3}\sum_{i=1}^{n}|\mathcal{K}_{2,i}^{(1)}(t,z)|\,|\varepsilon_{q,i}|\}\max_{1\leq i\leq n}|\widehat{r}_i-r_i|+O_p\left(h_2^{-2}\nu_{1n}^2\right)$$

$$=O_p(h_2^{-1}\nu_{1n})+o_p(n^{-1/2}h_1^{-(d+1)/2}),$$

by Assumption (E5(ii)). After collecting terms, (B-1) follows.

By using (A-6), we can further write

$$\widehat{V}_{q,n}\left(v\right)-V_{q,n}\left(v\right)=[V_{q,n,b}\left(v\right)+V_{q,n,c}\left(v\right)+V_{q,n,d}\left(v\right)]\left\{1+o_p\left(1\right)\right\}+o_p(n^{-1/2}h_1^{-(d+1)/2}),$$

where

$$V_{q,n,b}\left(v\right)\equiv n^{-2}h_1^{-(d+1)}h_2^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\widetilde{\alpha}_n\left(W_i,W_j;v\right),$$

$$V_{q,n,c}\left(v\right)\equiv n^{-1}h_2^{-2}\sum_{i=1}^{n}h_2^{-1}\mathcal{K}_{2,i}^{(1)}\left(v\right)\varepsilon_{q,i}\widetilde{\beta_n}\left(W_i\right),\text{ and}$$

$$V_{q,n,d}\left(v\right)\equiv n^{-1}h_2^{-2}\sum_{i=1}^{n}h_2^{-1}\mathcal{K}_{2,i}^{(1)}\left(v\right)\varepsilon_{q,i}\widetilde{\gamma}_n\left(W_i\right);$$

where, we have defined

$$\widetilde{\alpha}_n\left(W_i,W_j;v\right)\equiv\mathcal{K}_{2,i}^{(1)}\left(v\right)\iota_1^\top\left[\mathbf{M}_rf\left(W_i\right)\right]^{-1}\mathcal{K}_{1,j}\left(W_i\right)\varepsilon_{q,i}\varepsilon_{r,j},$$

$$\widetilde{\beta}_n\left(w\right)\equiv n^{-1}h_1^{-(d+1)}\iota_1^\top\left[\mathbf{M}_rf\left(w\right)\right]^{-1}\sum_{j=1}^{n}\mathcal{K}_{1,j}\left(w\right)\frac{1}{\mathbf{k}!}\sum_{|\mathbf{k}|=p_1+1}D^{\mathbf{k}}r\left(w\right)\left(W_i-w\right)^{\mathbf{k}},\text{ and}$$

$$\widetilde{\gamma}_n\left(w\right)\equiv n^{-1}h_1^{-(d+1)}\iota_1^\top\left[\mathbf{M}_rf\left(w\right)\right]^{-1}\gamma_n\left(w\right).$$

Thus, we have

$$\int T_{q,n,1}\left(t,z\right)dQ\left(t\right)=\mathcal{T}_{q,n,1a}+\left(\mathcal{T}_{q,n,1b}+\mathcal{T}_{q,n,1c}+\mathcal{T}_{q,n,1d}\right)\left\{1+o_p\left(1\right)\right\}+o_p(n^{-1/2}h_1^{-(d+1)/2}),$$

where

$$\mathcal{T}_{q,n,1a} = \frac{1}{nh_2} \sum_{i=1}^{n} \varepsilon_{q,i} \iota_2^\top \mathbf{M}_q^{-1} \int_{r_0}^{r(x,z)} h_2^{-1} \mathcal{K}_{2,i}(v) \left[f_V(v) q^2(v) \right]^{-1} dt,$$

$$\mathcal{T}_{q,n,1b} = \frac{1}{n^2 h_1^{(d+1)} h_2^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \iota_2^\top \mathbf{M}_q^{-1} \int_{r_0}^{r(x,z)} h_2^{-1} \widetilde{\alpha}_n(W_i, W_j; v) \left[f_V(v) q^2(v) \right]^{-1} dt,$$

$$\mathcal{T}_{q,n,1c} = \frac{1}{nh_2^2} \sum_{i=1}^{n} \varepsilon_{q,i} \iota_2^\top \mathbf{M}_q^{-1} \int_{r_0}^{r(x,z)} h_2^{-1} \mathcal{K}_{2,i}^{(1)}(v) \widetilde{\beta}_n(W_i) \left[f_V(v) q^2(v) \right]^{-1} dt, \text{ and}$$

$$\mathcal{T}_{q,n,1d} = \frac{1}{nh_2^2} \sum_{i=1}^{n} \varepsilon_{q,i} \iota_2^\top \mathbf{M}_q^{-1} \int_{r_0}^{r(x,z)} h_2^{-1} \mathcal{K}_{2,i}^{(1)}(v) \widetilde{\gamma}_n(W_i) \left[f_V(v) q^2(v) \right]^{-1} dt.$$

Firstly, by the law of iterated expectations, notice that $E[\mathcal{T}_{q,n,1a}] = 0$. While using representation (A-5), we are able to rewrite

$$\mathcal{T}_{q,n,1a} = n^{-1} h_2^{-1} \sum_{i=1}^{n} \varepsilon_{q,i} \iota_2^\top \mathbf{M}_q^{-1} \varrho_{i,1}^1 + n^{-1} \sum_{i=1}^{n} \varepsilon_{q,i} \iota_2^\top \mathbf{M}_q^{-1} h_2^{-1} \varrho_{i,2}^1$$
$$\equiv \mathcal{T}_{q,n,1a}^{(I)} + \mathcal{T}_{q,n,1a}^{(II)}.$$

By another change of variable and integration by parts, it is not difficult to see that $\mathcal{T}_{q,n,1a}^{(II)} = O_p\left(n^{-1/2} h_2\right)$ which is clearly $o_p(n^{-1/2} h_1^{-(d+1)/2})$. Moreover, $\mathcal{T}_{q,n,1a}^{(I)}$ satisfies the Linderberg–Feller Central Limit Theorem by virtue of Assumption E (see Härdle (1990)), thus $\mathcal{T}_{q,n,1a}^{(I)} = O_p\left(n^{-1/2}\right)$ and we conclude that $\sqrt{n h_1^{d+1}} \mathcal{T}_{q,n,1a} = o_p(1)$.

Now, under Assumptions (E1) – (E5), it is straightforward to extend the proof of Lemmas 3.1 (page 24) and 3.3 (page 26) in Lewbel and Linton (1999) to show that

$$\mathcal{T}_{q,n,1b} = O_p(n^{-1} h_1^{-(d+1)/2} h_2^{-1}),$$
$$= o_p(n^{-1/2} h_1^{-(d+1)/2}).$$

Let $\widetilde{\beta}(w) = \iota_1^\top \mathbf{M}_r^{-1} \mathbf{B}_r r^{(p_1+1)}(w)$, then

$$\sup_{w} |h_1^{-(p_1+1)} \widetilde{\beta}_n(w) - \widetilde{\beta}(w)| = O_p(n^{-1/2} h_1^{-(d+1)/2} \sqrt{\ln n}),$$

by (A-9). Therefore, we can write (recall $v = (t, z)$)

$$\mathcal{T}_{q,n,1c} \equiv \mathcal{T}_{q,n,1c}^{(I)} + \mathcal{T}_{q,n,1c}^{(II)}, \text{ where}$$

$$\mathcal{T}_{q,n,1c}^{(I)} = h_1^{p_1+1} n^{-1} h_2^{-2} \sum_{i=1}^{n} \varepsilon_{q,i} \iota_2^{\top} \mathbf{M}_q^{-1} \int h_2^{-1} \mathcal{K}_{2,i}^{(1)}(v) \widetilde{\beta}(W_i) \left[f_V(v) q^2(v) \right]^{-1} dQ(t), \text{ and}$$

$$\mathcal{T}_{q,n,1c}^{(II)} = h_1^{p_1+1} n^{-1} h_2^{-2} \sum_{i=1}^{n} \varepsilon_{q,i} \iota_2^{\top} \mathbf{M}_q^{-1}$$

$$\times \int h_2^{-1} \mathcal{K}_{2,i}^{(1)}(v) (h_1^{-(p_1+1)} \widetilde{\beta}_n(W_i) - \widetilde{\beta}(W_i)) [f_V(v) q^2(v)]^{-1} dQ(t).$$

Recall $\varepsilon_{q,i} = g_q(W_i) + \eta_i$ with $E[\eta_i | W_i] = 0$, then we can further write

$$\mathcal{T}_{q,n,1c}^{(I)} \equiv \mathcal{T}_{q,n,1c}^{(I-a)} + \mathcal{T}_{q,n,1c}^{(I-b)}, \text{ where}$$

$$\mathcal{T}_{q,n,1c}^{(I-a)} = h_1^{p_1+1} n^{-1} h_2^{-2} \sum_{i=1}^{n} g_q(W_i) \iota_2^{\top} \mathbf{M}_q^{-1} \varrho_{i,1}^0 \widetilde{\beta}(W_i)$$

$$- h_1^{p_1+1} n^{-1} h_2^{-1} \sum_{i=1}^{n} g_q(W_i) \iota_2^{\top} \mathbf{M}_q^{-1} h_2^{-1} \varrho_{i,2}^0 \widetilde{\beta}(W_i),$$

and $\mathcal{T}_{q,n,1c}^{(I-b)}$ is like $\mathcal{T}_{q,n,1c}^{(I-a)}$, but with η_i replacing $g_q(W_i)$. It follows by Bochner's Lemma that

$$\mathcal{T}_{q,n,1c}^{(I-a)} = h_1^{p_1+1} \iota_2^{\top} \mathbf{M}_q^{-1} \mathbf{M}_{q,0}^0 \left[\frac{E\left[\iota_1^{\top} \mathbf{M}_r^{-1} \mathbf{B}_r r^{(p_1+1)}(X, Z) g_q(X, Z) \middle| r(X, z_0) = r, Z = z_0 \right]}{q^2(r, z_0)} \right.$$

$$\left. - \frac{E\left[\iota_1^{\top} \mathbf{M}_r^{-1} \mathbf{B}_r r^{(p_1+1)}(X, Z) g_q(X, Z) \middle| r(X, z_0) = r_0, Z = z_0 \right]}{q^2(r_0, z_0)} \right] + o_p(n^{-1/2} h_1^{-(d+1)/2})$$

$$= h_1^{p_1+1} \mathcal{B}_1(x, z) + o_p(n^{-1/2} h_1^{-(d+1)/2}), \text{ by Assumption (E5(iii))}.$$

Similarly, by construction $\mathcal{T}_{q,n,1c}^{(I-b)}$ has mean zero and by the Cauchy–Schwarz inequality,

$$|\mathcal{T}_{q,n,1c}^{(I-b)}| = O_p(h_1^{p_1+1} n^{-1/2} h_2^{-1}) + O_p(h_1^{p_1+1} n^{-1/2})$$

$$= o_p(n^{-1/2} h_1^{-(d+1)/2}),$$

by Assumption (E5). With regards to $\mathcal{T}_{q,n,1c}^{(II)}$, this term may be written as

$$\mathcal{T}_{q,n,1c}^{(II)} = \mathcal{T}_{q,n,1c}^{(II-a)} + \mathcal{T}_{q,n,1c}^{(II-b)},$$

which are like $\mathcal{T}_{q,n,1c}^{(I-a)}$ and $\mathcal{T}_{q,n,1c}^{(I-b)}$, but with $h_1^{-(p_1+1)} \widetilde{\beta}_n(W_i) - \widetilde{\beta}(W_i)$ replacing $\widetilde{\beta}(W_i)$. Then by using similar arguments as above, we can show that

$$\mathcal{T}_{q,n,1c}^{(II-a)} = O_p(h_1^{p_1+1}) O_p(n^{-1/2} h_1^{-(d+1)/2} \sqrt{\ln n}) + O_p(h_1^{p_1+1} h_2) O_p(n^{-1/2} h_1^{-(d+1)/2} \sqrt{\ln n})$$

$$= o_p(h_1^{p_1+1} h_2) + o_p(h_1^{p_1+1} h_2^2), \text{ by Assumption (E5(ii))},$$

$$= o_p(n^{-1/2} h_1^{-(d+1)/2}) \text{ by Assumption (E5(iii))}.$$

Define $\widetilde{\gamma}(w) = E\left[\widetilde{\gamma}_n(w)\right]$, then by result (A-8), it follows that

$$\sup_w |\widetilde{\gamma}(w)| = o(h_1^{p_1+1}), \text{ and}$$

$$|h_1^{-(p_1+1)}\widetilde{\gamma}_n(w) - \widetilde{\gamma}(w)| = h_1^{p_1+1}O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n}) \text{ uniformly over } w.$$

Therefore, we can write

$$\mathcal{T}_{q,n,1d} = \mathcal{T}_{q,n,1d}^{(I)} + \mathcal{T}_{q,n,1d}^{(II)},$$

where $\mathcal{T}_{q,n,1d}^{(I)}$ and $\mathcal{T}_{q,n,1d}^{(II)}$ are like $\mathcal{T}_{q,n,1c}^{(I)}$ and $\mathcal{T}_{q,n,1c}^{(II)}$, but with $\widetilde{\gamma}(W_i)$ and $h_1^{-(p_1+1)}\widetilde{\gamma}_n(W_i) - \widetilde{\gamma}(W_i)$ replacing $\widetilde{\beta}(W_i)$ and $h_1^{-(p_1+1)}\widetilde{\beta}_n(W_i) - \widetilde{\beta}(W_i)$ respectively. Then by the Cauchy–Schwarz inequality,

$$|\mathcal{T}_{q,n,1d}^{(I)}| = h_1^{p_1+1}o_p(n^{-1/2}h_1^{-(d+1)/2}) + h_1^{p_1}h_2 o_p(n^{-1/2}h_1^{-(d+1)/2}) \text{ by Assumption (E5(iii))}.$$

Similarly, by Assumption (E5(ii)),

$$\mathcal{T}_{q,n,1d}^{(II)} = o_p(n^{-1/2}h_1^{-(d+1)/2}).$$

Thus, $\sqrt{nh_1^{d+1}}\mathcal{T}_{q,n,1d} = o_p(1)$. ■

Lemma B-2 *Under Assumption E, we have*

$$\sup_{t,z} |T_{q,n,2}(t,z)| = O_p\left(h_1^{-1}\nu_{1n}\right), \text{ and} \tag{B-3}$$

$$\int_{r_0}^{r(x,z)} T_{q,n,2}(t,z)\,dt = O_p\left(n^{-1/2}h_1^{-(d+1)/2}\right). \tag{B-4}$$

Proof. Let $\widehat{S}_i - S_i = |\widehat{D}^{|1,0|}r(X_i, Z_i) - D^{|1,0|}r(X_i, Z_i)|$. Then, by Theorem 6 (page 594) in Masry (1996a),

$$\max_{1\le i\le n} |\widehat{S}_i - S_i| = O_p\left(h_1^{-1}\nu_{1n}\right).$$

We now write

$$\widehat{V}_{q,n}^*(v) = nh_2^{-2}\sum_{i=1}^n \mathcal{K}_{2,i}(v)(\widehat{S}_i - S_i) + nh_2^{-2}\sum_{i=1}^n [\widehat{\mathcal{K}}_{2,i}(v) - \mathcal{K}_{2,i}(v)](\widehat{S}_i - S_i).$$

The first term is clearly

$$nh_2^{-2}\sum_{i=1}^n \mathcal{K}_{2,i}(v)(\widehat{S}_i - S_i) = O_p\left(h_1^{-1}\nu_{1n}\right) \text{ uniformly in } v.$$

The second term, after a Taylor–series expansion, is

$$nh_2^{-2}\sum_{i=1}^n [\widehat{\mathcal{K}}_{2,i}(v) - \mathcal{K}_{2,i}(v)](\widehat{S}_i - S_i)$$

$$\le \{n^{-1}h_2^{-3}\sum_{i=1}^n |\mathcal{K}_{2,i}^{(1)}(t,z)|\}\max_{1\le i\le n}|\widehat{r}_i - r_i|\max_{1\le i\le n}\left|\widehat{S}_i - S_i\right| + O_p\left(h_1^{-1}\nu_{1n}\right)O_p\left(h_2^{-2}\nu_{1n}^2\right),$$

$$= O_p\left(h_2^{-1}\nu_{1n}\right)O_p\left(h_1^{-1}\nu_{1n}\right) + O_p\left(h_1^{-1}\nu_{1n}\right)O_p\left(h_2^{-2}\nu_{1n}^2\right), \text{ by Assumption (E5(ii))},$$

$$= o_p(n^{-1/2}h_1^{-(d+1)/2}).$$

Therefore, (B-3) follows immediately.

We can write

$$\widehat{V}_{q,n}^*(v) = [V_{q,n,a}^*(v) + V_{q,n,b}^*(v) + V_{q,n,c}^*(v)]\{1 + o_p(1)\} + o_p(n^{-1/2}h_1^{-(d+1)/2})$$

by using (A-7), where

$$V_{q,n,a}^*(v) \equiv n^{-2}h_1^{-(d+2)}h_2^{-2} \sum_{i=1}^n \sum_{j=1}^n \widetilde{\alpha}_n^*(W_i, W_j; v, w),$$

$$V_{q,n,b}^*(v) \equiv n^{-1}h_2^{-2} \sum_{i=1}^n \mathcal{K}_{2,i}(v)\widetilde{\beta}_n^*(W_i), \text{ and}$$

$$V_{q,n,c}^*(v) \equiv n^{-1}h_2^{-2} \sum_{i=1}^n \mathcal{K}_{2,i}(v)\widetilde{\gamma}_n^*(W_i)$$

with

$$\widetilde{\alpha}_n^*(W_i, W_j; v) \equiv \mathcal{K}_{2,i}(v)\iota_1^{*\top}[\mathbf{M}_r f_W(W_i)]^{-1}\mathcal{K}_{1,j}(W_i)\varepsilon_{r,j},$$

$$\widetilde{\beta}_n^*(w) \equiv n^{-1}h_1^{-(d+1)}\iota_1^{*\top}[\mathbf{M}_r f_W(w)]^{-1}\sum_{j=1}^n \mathcal{K}_{1,j}(w)$$

$$\times \frac{1}{\mathbf{k}!}\sum_{|\mathbf{k}|=p_1+1}D^{\mathbf{k}}r(w)(W_i - w)^{\mathbf{k}}, \text{ and}$$

$$\widetilde{\gamma}_n^*(w) \equiv n^{-1}h_1^{-(d+1)}\iota_1^{*\top}[\mathbf{M}_r f_W(w)]^{-1}\gamma_n(w).$$

Thus, we have

$$\int T_{q,n,2}(t,z)\,dQ(t) = (\mathcal{T}_{q,n,2a} + \mathcal{T}_{q,n,2b} + \mathcal{T}_{q,n,2c})\{1 + o_p(1)\} + o_p(n^{-1/2}h_1^{-(d+1)/2}),$$

where

$$\mathcal{T}_{q,n,2a} = \frac{h_2}{n^2 h_1^{(d+2)}h_2^2}\sum_{i=1}^n \sum_{j=1}^n \iota_2^\top \mathbf{M}_q^{-1}\int_{r_0}^{r(x,z)} h_2^{-1}\widetilde{\alpha}_n^*(W_i, W_j; v)\left[f_V(v)q^2(v)\right]^{-1}dt,$$

$$\mathcal{T}_{q,n,2b} = \frac{h_2}{nh_1 h_2^2}\sum_{i=1}^n \iota_2^\top \mathbf{M}_q^{-1}\int_{r_0}^{r(x,z)} h_2^{-1}\mathcal{K}_{2,i}(v)\widetilde{\beta}_n^*(W_i)\left[f_V(v)q^2(v)\right]^{-1}dt, \text{ and}$$

$$\mathcal{T}_{q,n,2c} = \frac{h_2}{nh_1 h_2^2}\sum_{i=1}^n \iota_2^\top \mathbf{M}_q^{-1}\int_{r_0}^{r(x,z)} h_2^{-1}\mathcal{K}_{2,i}(v)\widetilde{\gamma}_n^*(W_i)\left[f_V(v)q^2(v)\right]^{-1}dt.$$

Using (A-5),

$$\int h_2^{-1}\widetilde{\alpha}_n^*(W_i, W_j; v)\left[f_V(v)q^2(v)\right]^{-1}dQ(t)$$

$$\equiv \alpha_{n,1}^*(W_i, W_j; r, r_0, z_0) - \alpha_{n,2}^*(W_i, W_j; r, r_0, z_0),$$

44

where

$$\alpha_{n,1}^* \left(W_i, W_j; r, r_0, z_0 \right) \equiv \varrho_{i,1}^1 \iota_1^{*\top} \left[\mathbf{M}_r f_W \left(W_i \right) \right]^{-1} \mathcal{K}_{1,j} \left(W_i \right) \varepsilon_{r_{,j}}, \text{ and}$$

$$\alpha_{n,2}^* \left(W_i, W_j; r, r_0, z_0 \right) \equiv \varrho_{i,2}^1 \iota_1^{*\top} \left[\mathbf{M}_r f_W \left(W_i \right) \right]^{-1} \mathcal{K}_{1,j} \left(W_i \right) \varepsilon_{r_{,j}}.$$

By the law of iterated expectations, $E\left[\alpha_{n,1}^*\right] = E\left[\alpha_{n,2}^*\right] = 0$, and by applying a second order U-statistic theory for random samples (e.g. Powell, Stock, and Stoker (1989)), it is not difficult but lengthy to show that

$$\mathcal{T}_{q,n,2a} = O_p(n^{-1}h_1^{-(d+2)/2}) + O_p(n^{-1}h_1^{-(d+2)/2}h_2).$$

Thus, by Assumption (E5), $\sqrt{nh_1^{d+1}}\mathcal{T}_{q,n,2a} = o_p\left(1\right)$.

By (A-9),

$$\sup_w |h_1^{-(p_1+1)}\widetilde{\beta}_n^* \left(w \right) - \widetilde{\beta}^* \left(w \right)| = O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n}),$$

where $\widetilde{\beta}^* \left(w \right) = \iota_1^{*\top}\mathbf{M}_r^{-1}\mathbf{B}_r r^{(p_1+1)}(w)$. Therefore, we can write

$$\mathcal{T}_{q,n,2b} = \mathcal{T}_{q,n,2b}^{(I)} + \mathcal{T}_{q,n,2b}^{(II)}, \text{ where}$$

$$\mathcal{T}_{q,n,2b}^{(I)} = h_1^{p_1}n^{-1}h_2^{-1}\sum_{i=1}^n \iota_2^\top\mathbf{M}_q^{-1}\int h_2^{-1}\mathcal{K}_{2,i}\left(v\right)\widetilde{\beta}^*\left(W_i\right)\left[f_V\left(v\right)q^2\left(v\right)\right]^{-1}dQ\left(t\right),$$

$$\mathcal{T}_{q,n,2b}^{(II)} = h_1^{p_1}n^{-1}h_2^{-1}\sum_{i=1}^n \iota_2^\top\mathbf{M}_q^{-1}$$

$$\times \int h_2^{-1}\mathcal{K}_{2,i}\left(v\right)\left(h_1^{-(p_1+1)}\widetilde{\beta}_n^*\left(W_i\right) - \widetilde{\beta}^*\left(W_i\right)\right)\left[f_V\left(v\right)q^2\left(v\right)\right]^{-1}dQ\left(t\right).$$

Then, by using representation (A-5), we may further write

$$\mathcal{T}_{q,n,2b}^{(I)} = n^{-1}h_2^{-1}h_1^{p_1}\sum_{i=1}^n \iota_2^\top\mathbf{M}_q^{-1}\varrho_{i,1}^1\widetilde{\beta}^*\left(W_i\right) - n^{-1}h_1^{p_1}\sum_{i=1}^n \iota_2^\top\mathbf{M}_q^{-1}h_2^{-1}\varrho_{i,2}^1\widetilde{\beta}^*\left(W_i\right).$$

The right hand side of the above expression converges in mean squared to (by Bochner's Lemma)

$$\mathcal{T}_{q,n,2b}^{(I)} = h_1^{p_1}h_2\iota_2^\top\mathbf{M}_q^{-1}\mathbf{M}_{q,0}^1 \left[\frac{E\left[\iota_1^\top\mathbf{M}_r^{-1}\mathbf{B}_r r^{(p_1+1)}(X, Z)\,\middle|\, r\left(X, z_0\right) = r, Z = z_0\right]}{q^2\left(r, z_0\right)} \right.$$

$$\left. - \frac{E\left[\iota_1^\top\mathbf{M}_r^{-1}\mathbf{B}_r r^{(p_1+1)}(X, Z)\,\middle|\, r\left(X, z_0\right) = r_0, Z = z_0\right]}{q^2\left(r_0, z_0\right)} \right] + o_p(n^{-1/2}h_1^{-(d+1)/2})$$

$$= h_1^{p_1}h_2\mathcal{B}_2\left(x, z\right) + o_p(n^{-1/2}h_1^{-(d+1)/2}), \text{ by Assumption (E5(iii))}.$$

Furthermore, $\mathcal{T}_{q,n,2b}^{(II)}$ is like $\mathcal{T}_{q,n,2b}^{(I)}$, but with $h_1^{-(p_1+1)}\widetilde{\beta}_n^*\left(w\right) - \widetilde{\beta}^*\left(w\right)$ replacing $\widetilde{\beta}^*\left(W_i\right)$. Then, it follows by Cauchy–Schwarz inequality, that

$$|\mathcal{T}_{q,n,2b}^{(II)}| = O_p(h_1^{p_1}h_2)O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n}) + O_p(h_1^{p_1}h_2^2)O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n}),$$

$$= O_p(h_1^{p_1}h_2)o_p\left(h_2\right) + O_p(h_1^{p_1}h_2^2)o_p\left(h_2\right), \text{ by Assumption (E5(ii))},$$

$$= o_p(n^{-1/2}h_1^{-(d+1)/2}) \text{ by virtue of Assumption (E5(iii))}.$$

Consequently, $\sqrt{nh_1^{d+1}}\mathcal{T}_{q,n,2b} = O_p(1)$.

Define $\widetilde{\gamma}^*(w) = E[\widetilde{\gamma}_n^*(w)]$, then by result (A-8), it follows that

$$\sup_w |\widetilde{\gamma}^*(w)| = o(h_1^{p_1+1}), \text{ and}$$

$$|h_1^{-(p_1+1)}\widetilde{\gamma}_n^*(w) - \widetilde{\gamma}^*(w)| = h_1^{p_1+1}O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n}) \text{ uniformly in } w.$$

Therefore, we can write

$$\mathcal{T}_{q,n,2c} = \mathcal{T}_{q,n,2c}^{(I)} + \mathcal{T}_{q,n,2c}^{(II)},$$

where these two terms are like $\mathcal{T}_{q,n,2b}^{(I)}$ and $\mathcal{T}_{q,n,2b}^{(II)}$, but with $\widetilde{\gamma}^*(W_i)$ and $h_1^{-(p_1+1)}\widetilde{\gamma}_n^*(W_i) - \widetilde{\gamma}^*(W_i)$ replacing $\widetilde{\beta}^*(W_i)$ and $h_1^{-(p_1+1)}\widetilde{\beta}_n^*(W_i) - \widetilde{\beta}^*(W_i)$ respectively. Then by the Cauchy–Schwarz inequality,

$$\begin{aligned}|\mathcal{T}_{q,n,2c}^{(I)}| &= O_p(h_1^{p_1}h_2)o_p(n^{-1/2}h_1^{-(d+1)/2}) \\ &\quad + O_p(h_1^{p_1}h_2^2)o_p(n^{-1/2}h_1^{-(d+1)/2}) \text{ by Assumption (E5(iii))}.\end{aligned}$$

Similarly, by Assumption (E5(ii)) and (E5(iii)),

$$\begin{aligned}|\mathcal{T}_{q,n,2c}^{(II)}| &= O_p(h_1^{p_1}h_2)h_1^{p_1+1}O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n}) \\ &\quad + O_p(h_1^{p_1}h_2^2)h_1^{p_1+1}O_p(n^{-1/2}h_1^{-(d+1)/2}\sqrt{\ln n}) \\ &= o_p(n^{-1/2}h_1^{-(d+1)/2}).\end{aligned}$$

Thus, $\sqrt{nh_1^{d+1}}\mathcal{T}_{q,n,2c} = o_p(1)$. ∎

Lemma B-3 *Under Assumption E, we have*

$$\sup_{t,z}|T_{q,n,3}(t,z)| = O_p\left(h_2^{p_2+1}\right), \text{ and} \tag{B-5}$$

$$\int_{r_0}^{r(x,z)} T_{q,n,3}(t,z)\,dt = O_p\left(n^{-1/2}h_1^{-(d+1)/2}\right). \tag{B-6}$$

Proof. Define

$$\Delta_{q,i}(v) = q(V_i) - \sum_{0\le|\mathbf{k}|\le p_2}\frac{1}{\mathbf{k}!}D^{\mathbf{k}}q(v)(V_i - v)^{\mathbf{k}},$$

and $\widehat{\Delta}_{q,i}(v)$ is like $\Delta_{q,i}(v)$ but with $\widehat{V}_i$ in place of $V_i = (r(X_i,Z_i),Z_i)^\top$. Then by Assumption (E4),

$$\Delta_{q,i}(v) = \sum_{|\mathbf{k}|=p_2+1}\frac{1}{\mathbf{k}!}D^{\mathbf{k}}q(v^*)(V_i - v)^{\mathbf{k}}$$

for some v^* that lies between V_i and v, also

$$\widehat{\Delta}_{q,i}(v) = \sum_{|\mathbf{k}|=p_2+1}\frac{1}{\mathbf{k}!}D^{\mathbf{k}}q(\widehat{v}^*)(V_i - v)^{\mathbf{k}},$$

where $\widehat{v}^*$ lies between $\widehat{V}_i$ and v. It is also clear that $\|\widehat{v}^* - v^*\| = O_p(\nu_{1n})$ and $|\Delta_{q,i}(v)| = O_p(h_2^{p_2+1})$ for $\|V_i - v\| \le ch_2$. These observations along with Assumption (E5(ii)) imply that

$$\widehat{\Delta}_{q,i}(v) - \Delta_{q,i}(v) = \sum_{|\mathbf{k}|=p_2+1} \frac{1}{\mathbf{k}!} D^{\mathbf{k}} q(\widehat{v}^*) [(\widehat{V}_i - v)^{\mathbf{k}} - (V_i - v)^{\mathbf{k}}]$$
$$+ \sum_{|\mathbf{k}|=p_2+1} \frac{1}{\mathbf{k}!} [D^{\mathbf{k}} q(\widehat{v}^*) - D^{\mathbf{k}} q(v^*)] (V_i - v)^{\mathbf{k}},$$
$$= O_p(h_2^{p_2} \nu_{1n}) = o_p(n^{-1/2} h_1^{-(d+1)/2}),$$

uniformly in v and i such that $\|V_i - v\| \le ch_2$. So we conclude that

$$|\widehat{\Delta}_{q,i}(v)| = O_p(h_2^{p_2+1}) + o_p(n^{-1/2} h_1^{-(d+1)/2})$$

uniformly in v and i for $\|V_i - v\| \le ch_2$.

We now write

$$\left| \widehat{\mathbf{B}}_{q,n}(t,z) - \mathbf{B}_{q,n}(t,z) \right| \le \left\{ \frac{1}{nh_2^2} \sum_{i=1}^{n} |\mathcal{K}_{2,i}(t,z)| \right\} \max_{1\le i\le n} \sup_{v} |\widehat{\Delta}_{q,i}(v) - \Delta_{q,i}(v)| \qquad \text{(B-7)}$$
$$+ \left\{ \frac{1}{nh_2^2} \sum_{i=1}^{n} \left| \widehat{\mathcal{K}}_{2,i}(t,z) - \mathcal{K}_{2,i}(t,z) \right| \right\} \max_{1\le i\le n} \sup_{v} |\widehat{\Delta}_{q,i}(v)|. \qquad \text{(B-8)}$$

It is clear that (B-7) is $o_p(n^{-1/2} h_1^{-(d+1)/2})$. The order in probability of (B-8), after a Taylor–series expansion, is given by

$$\{n^{-1} h_2^{-3} \sum_{i=1}^{n} |\mathcal{K}_{2,i}^{(1)}(t,z)|\} \max_{1\le i\le n} |\widehat{r}_i - r_i| \max_{1\le i\le n} \sup_{v} |\widehat{\Delta}_{q,i}(v)| + O_p\left(h_2^{-2} \nu_{1n}^2\right) \max_{1\le i\le n} \sup_{v} |\widehat{\Delta}_{q,i}(v)|$$
$$= O_p(h_2^{-1} \nu_{1n}) O_p(h_2^{p_2+1}) + O_p(h_2^{-2} \nu_{1n}^2) O_p(h_2^{p_2+1})$$
$$= o_p(n^{-1/2} h_1^{-(d+1)/2}), \text{ by Assumption (E5(ii))}$$

Therefore,

$$\sup_{v} |\widehat{\mathbf{B}}_{q,n}(v)| \le \sup_{v} |\mathbf{B}_{q,n}(v)| + \sup_{v} |\widehat{\mathbf{B}}_{q,n}(v) - \mathbf{B}_{q,n}(v)|$$
$$= O_p(h_2^{p_2+1}) + O_p(h_2^{p_2+1} \nu_{1n}),$$

proving (B-5). Furthermore, we can rewrite

$$T_{q,n,3}(t,z_0) = h_2^{p_2+1} \iota_2^\top \mathbf{M}_q^{-1} \mathbf{B}_q q^{(p_2+1)}(t,z_0) q^{-2}(t,z_0)$$
$$+ h_2^{p_2+1} \iota_2^\top \mathbf{M}_q^{-1} [\mathbf{B}_{q,n}(t,z_0) - \mathbf{B}_q f_V(t,z_0)] q^{(p_2+1)}(t,z_0) q^{-2}(t,z_0)$$
$$+ o_p(n^{-1/2} h_1^{-(d+1)/2}).$$

Clearly, the first term of the above equation is $O_p(h_2^{p_2+1})$, and the second is $o_p(h_2^{p_2+1})$ by Corollary 2 (page 580) in Masry (1996a). Then,

$$\int_{r_0}^{r(x,z)} T_{q,n,3}(t,z)\, dt = h_2^{p_2+1} i_2^\top \mathbf{M}_q^{-1} \mathbf{B}_q \int_{r_0}^{r(x,z)} \frac{q^{(p_2+1)}(t,z_0)}{q^2(t,z_0)}\, dt + o_p(n^{-1/2} h_1^{-(d+1)/2})$$

$$= h_2^{p_2+1} \mathcal{B}_3(x,z) + o_p(n^{-1/2} h_1^{-(d+1)/2})$$

follows. ∎

Lemma B-4 *Under Assumption E, we have*

$$\mathcal{R}_{q,n}(x,z) = \int T_{q,n,4}(t,z_0)\, dQ(t) + \int T_{q,n,5}(t,z_0)\, dQ(t) + \int T_{q,n,6}(t,z_0)\, dQ(t)$$

$$= o_p(n^{-1/2} h_1^{-(d+1)/2}).$$

Proof. A typical element of $\widehat{\mathbf{M}}_{q,n}(v) - \mathbf{M}_{q,n}(v)$ is given by

$$[\widehat{\mathbf{M}}_{q,n,j,k}(v)]_{l,l^0} - [\mathbf{M}_{q,n,j,k}(v)]_{l,l^0}$$

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

After a Taylor–series expansion of the last expression at V_i, it is not difficult to show that

$$\sup_{t,z} |[\widehat{\mathbf{M}}_{q,n,j,k}(t,z)]_{l,l^0} - [\mathbf{M}_{q,n,j,k}(t,z)]_{l,l^0}| = O_p\left(h_2^{-1} \nu_{1n}\right).$$

By the triangle inequality, we have

$$\sup_{t,z} |\widehat{\mathbf{M}}_{q,n}(t,z) - f_V(t,z)\mathbf{M}_q| \leq \sup_{t,z} |\widehat{\mathbf{M}}_{q,n}(t,z) - \mathbf{M}_{q,n}(t,z)| + \sup_{t,z} |\mathbf{M}_{q,n}(t,z) - f_V(t,z)\mathbf{M}_q|.$$

The first term of the right hand side of the inequality is $O_p\left(h_2^{-1}\nu_{1n}\right) = o_p(1)$, while the second is, by Corollary 2 (page 580) in Masry (1996a), $O_p(n^{-1/2} h_2^{-1} \sqrt{\ln n} + h_2) = o_p(1)$. Furthermore, by Assumption (E1), $\widehat{\mathbf{M}}_{q,n}^{-1}(v) = O_p(1)$ with probability approaching one. Therefore, results (B-2), (B-4) and (B-6) imply that

$$\int T_{q,n,4}(t,z_0)\, dQ(t) = o_p(1) O_p(n^{-1/2} h_1^{-(d+1)/2}),$$

$$\int T_{q,n,5}(t,z_0)\, dQ(t) = o_p(1) O_p(n^{-1/2} h_1^{-(d+1)/2}), \text{ and}$$

$$\int T_{q,n,6}(t,z_0)\, dQ(t) = o_p(1) O_p(n^{-1/2} h_1^{-(d+1)/2}),$$

respectively. ∎

Lemma B-5 *Under Assumption E, we have*

$$\sup_{t,z} |\widehat{q}(t,z) - q(t,z)| = O_p\left(h_2^{-1}\nu_{1n} + h_1^{-1}\nu_{1n} + \nu_{2n}\right).$$

Proof. This result follows from (B-1), (B-3), (B-5) and Lemma B-4. ∎

Lemma B-6 *Let Assumptions E and F hold, then the estimators $\widehat{\alpha}_{P_1}(x)$ and $\widehat{\alpha}_{P_2}(z)$ satisfies the following asymptotic expansions:*

$$\widehat{\alpha}_{P_1}(x) - \alpha_{P_1}(x) = \iota_1^\top \mathbf{M}_r^{-1}\{1 + o_p(1)\}$$
$$\times \left\{n^{-1}h_1^{-d}\sum_{j=1}^{n}\left[\int h_1^{-1}\mathcal{K}_{1,j}(x,z)\frac{dP_1(z)}{q(r,z_0)f_W(x,z)}\right]\varepsilon_{r,j}\right\}$$
$$+ \int \mathcal{B}(x,z)\,dP_1(z) + R_{P_1,n}(x),$$

$$\widehat{\alpha}_{P_2}(z) - \alpha_{P_2}(z) = \iota_1^\top \mathbf{M}_r^{-1}\{1 + o_p(1)\}$$
$$\times \left\{n^{-1}h_1^{-1}\sum_{j=1}^{n}\left[\int h_1^{-d}\mathcal{K}_{1,j}(x,z)\frac{dP_2(x)}{q(r,z_0)f_W(x,z)}\right]\varepsilon_{r,j}\right\}$$
$$+ \int \mathcal{B}(x,z)\,dP_2(x) + R_{P_2,n}(z),$$

where $o_p(1)$'s are uniformly in x and z, and the remainder terms $R_{P_1,n}(x)$, and $R_{P_2,n}(z)$ satisfy

$$\sup_x |R_{P_1,n}(x)| = o_p(n^{-1/2}h_*^{-1/2}), \text{ and}$$
$$\sup_z |R_{P_2,n}(z)| = o_p(n^{-1/2}h_*^{-1/2}) \text{ respectively.}$$

Proof. This result follows from Lemmas B-1–B-5 and Assumption (F3). ∎

Lemma B-7 *Let Assumptions E and F hold, then*

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

Proof. Let $V_{H,n}(m) = nh_*^{-1}\sum_{i=1}^{n}\mathcal{K}_{*,i}(m)\varepsilon_{r,i}$, then we have

$$\widehat{V}_{H,n}(m) - V_{H,n}(m) = \frac{1}{nh_*}\sum_{i=1}^{n}[\widehat{\mathcal{K}}_{*,i}(m) - \mathcal{K}_{*,i}(m)]\varepsilon_{r,i}$$
$$= \frac{1}{nh_*^2}\sum_{i=1}^{n}\mathcal{K}_{*,i}^{(1)}(m)(\widetilde{M}_i - M_i)\varepsilon_{r,i} + O_p\left(h_*^{-2}\nu_{\dagger n}^2\right)$$
$$= \frac{1}{nh_*^2}\sum_{i=1}^{n}\mathcal{K}_{*,i}^{(1)}(m)\left\{[\widehat{\alpha}_{P_1}(X_i) - \alpha_{P_1}(X_i)]\right.$$
$$\left. + [\widehat{\alpha}_{P_2}(Z_i) - \alpha_{P_2}(Z_i)] - [\widetilde{c}_i - c_i]\right\}\varepsilon_{r,i}$$
$$+ o_p(n^{-1/2}h_*^{(d+1)/2}),$$

where the second equality follows from a Taylor–series expansion and Assumption (F3), and the last by Assumption (F3(ii)). So,

$$\widehat{T}_{H,n}(m) = T_{H,n,a}(m) + T_{H,n,b}(m)\{1 + o(1)\}$$
$$+ T_{H,n,c}(m) + T_{H,n,d}(m) + o_p(n^{-1/2}h_*^{-1/2}),$$

where

$$T_{H,n,a}(m) = \frac{1}{nh_*}\sum_{i=1}^n \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1} \mathcal{K}_{*,i}(m)\varepsilon_{r,i}$$

$$T_{H,n,b}(m) = \frac{1}{n^2 h_*^2}\sum_{i=1}^n\sum_{j=1}^n \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1} \alpha_n(\xi_i,\xi_j;m), \tag{B-9}$$

$$T_{H,n,c}(m) = \frac{1}{nh_*^2}\sum_{i=1}^n \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1} \mathcal{K}_{*,i}^{(1)}(m)\varepsilon_{r,i}\beta(W_i), \tag{B-10}$$

$$T_{H,n,d}(m) = \frac{1}{nh_*^2}\sum_{i=1}^n \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1} \mathcal{K}_{*,i}^{(1)}(m)\varepsilon_{r,i}R_n(W_i). \tag{B-11}$$

We now discuss the properties of each term above.

Firstly, let $\mathcal{F}_W$ and $\mathcal{F}_H$ be the sigma algebras generated by $W^\top = (X^\top, Z)$ and $r(W) = H[M(W)]$ respectively, then by the tower property of conditional expectations, i.e. Theorem (34.3) in Billingsley (1986), we have $E[\varepsilon_{r,i}|r(W_i)] = 0$, which implies that $E[\varsigma_i] = 0$ by the law of iterated expectations, where

$$\varsigma_i = h_*^{-1}\iota_*^\top [f_M(m)\mathbf{M}_H]^{-1} \mathcal{K}_{*,i}(m)\varepsilon_{r,i}.$$

Therefore, by Theorem 4 (page 94) in Masry (1996b), it follows that

$$\sqrt{nh_*}T_{H,n}(m) = \frac{1}{n^{1/2}h_*^{1/2}}\sum_{i=1}^n \varsigma_i$$
$$\xrightarrow{d} N\left(0, \frac{\sigma_M^2(m)}{f_M(m)}\left[\mathbf{M}_H^{-1}\Gamma_H\mathbf{M}_H^{-1}\right]_{0,0}\right),$$

where $\sigma_H^2(b) = E[\varepsilon_{r,i}^2|B(W_i) = b]$. The term α_n in (B-9) may be written as

$$\alpha_n(\xi_i,\xi_j;b) \equiv \alpha_n^I(\xi_i,\xi_j;b) + \alpha_n^{II}(\xi_i,\xi_j;b) - \alpha_n^{III}(\xi_i,\xi_j;b), \text{ with}$$
$$\alpha_n^I(\xi_i,\xi_j;m) \equiv h_1^{-d}\mathcal{K}_{*,i}^{(1)}(m)\varepsilon_{r,i}\int h_1^{-1}\iota_1^\top\mathbf{M}_r^{-1}\mathcal{K}_{1,j}(W_i)\frac{dP_1(Z_i)}{q(r_i,z_0)f_W(W_i)}\varepsilon_{r,j},$$
$$\alpha_n^{II}(\xi_i,\xi_j;m) \equiv h_1^{-1}\mathcal{K}_{*,i}^{(1)}(m)\varepsilon_{r,i}\int h_1^{-d}\iota_1^\top\mathbf{M}_r^{-1}\mathcal{K}_{1,j}(W_i)\frac{dP_2(X_i)}{q(r_i,z_0)f_W(W_i)}\varepsilon_{r,j},$$
$$\alpha_n^{III}(\xi_i,\xi_j;m) \equiv \mathcal{K}_{*,i}^{(1)}(m)\varepsilon_{r,i}\int\int h_1^{-(d+1)}\iota_1^\top\mathbf{M}_r^{-1}\mathcal{K}_{1,j}(W_i)\frac{dP_1(Z_i)\,dP_2(X_i)}{q(r_i,z_0)f_W(W_i)}\varepsilon_{r,j},$$

where $\xi_i \equiv (W_i^\top, \varepsilon_{r,i})^\top$. Thus, by applying a second order U-statistic theory for random samples (e.g. Powell, Stock, and Stoker (1989)), we can show that under Assumptions E and F,

$$T_{H,n,b}(m) = O_p(n^{-1}h_1^{-d/2}h_*^{-3/2}) + O_p(n^{-1}h_1^{-1/2}h_*^{-3/2}) + O_p(n^{-1}h_*^{-3/2}).$$

Consequently, by Assumption (F3(ii)),

$$\begin{aligned}
\sqrt{nh_*}T_{H,n,b}(m) &= \sqrt{nh_*}O_p(n^{-1}h_1^{-d/2}h_*^{-3/2}) \\
&+ \sqrt{nh_*}O_p(n^{-1}h_1^{-1/2}h_*^{-3/2}) + \sqrt{nh_*}O_p(n^{-1}h_*^{-3/2}) \\
&= O_p(n^{-1/2}h_1^{-d/2}h_*^{-1}) + O_p(n^{-1/2}h_1^{-1/2}h_*^{-1}) + O_p(n^{-1/2}h_*^{-1}) = o_p(1).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\beta(w) &\equiv \int \mathcal{B}(x,z)\,dP_1(z) + \int \mathcal{B}(x,z)\,dP_2(x) + \int\int \mathcal{B}(x,z)\,dP_1(z)\,dP_2(x), \\
&= O(h_\dagger) \text{ by construction.}
\end{aligned}$$

Then, by Assumption (F3(ii)), (B-10) is $h_\dagger O_p(n^{-1/2}h_*^{-3/2})$, so

$$\sqrt{nh_*}T_{H,n,c}(m) = O_p\left(h_\dagger h_*^{-1}\right) = o_p(1).$$

Finally, the term (B-11) satisfies

$$\left| n^{-1}h_*^{-2} \sum_{i=1}^n \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1} \mathcal{K}_{*,i}^{(1)}(m)\,\varepsilon_{r,i}R_n(W_i) \right|$$

$$\leq 3 \times \sup_x |R_{P_1,n}(x)| \left\{ n^{-1}h_*^{-2} \sum_{i=1}^n \left| \iota_*^\top [f_M(m)\mathbf{M}_H]^{-1} \mathcal{K}_{*,i}^{(1)}(m)\,\varepsilon_{r,i} \right| \right\},$$

and the right hand side is $o_p(n^{-1/2}h_*^{-1/2})$ by Lemma B-6, completing the proof. ∎

Lemma B-8 *Let Assumptions E and F hold, then*

$$T_{H,n,2}(m) = O_p\left(n^{-1/2}h_*^{-1/2}\right).$$

Proof. Firstly, $\max_{1\leq i\leq n}|H(M_i) - H(\widetilde{M_i})| = O_p(\nu_{\dagger n})$ by Assumption (I2). Let

$$V_{H,n}^*(m) = n^{-1}h_*^{-1} \sum_{i=1}^n \mathcal{K}_{*,i}(m)[H(M_i) - H(\widetilde{M_i})],$$

then after a Taylor–series expansion,

$$\begin{aligned}
\widehat{V}_{H,n}^*(m) - V_{H,n}^*(m) &= n^{-1}h_*^{-1} \sum_{i=1}^n \left\{ \widehat{\mathcal{K}}_{*,i}(m) - \mathcal{K}_{*,i}(m) \right\} [H(M_i) - H(\widetilde{M_i})] \\
&= n^{-1}h_*^{-2} \sum_{i=1}^n \mathcal{K}_{*,i}^{(1)}(m)[\widetilde{M_i} - M_i][H(M_i) - H(\widetilde{M_i})] + O_p(h_*^{-2}\nu_{\dagger n}^3) \\
&= O_p(h_*^{-1}\nu_{\dagger n}^2) + O_p(h_*^{-2}\nu_{\dagger n}^3) = o_p(n^{-1/2}h_*^{-1/2}),
\end{aligned}$$

by Assumption (F3(ii)). Consequently, after a further Taylor–series expansion,

$$T_{H,n,2}(m) = n^{-1}h_*^{-1} \sum_{i=1}^{n} \mathcal{K}_{*,i}(m)[H(M_i) - H(\widetilde{M_i})] + o_p(n^{-1/2}h_*^{-1/2})$$

$$= -n^{-1}h_*^{-1} \sum_{i=1}^{n} \mathcal{K}_{*,i}(m) \frac{\partial H(M_i)}{\partial b} [\widetilde{M_i} - M_i] + O_p(\nu_{\dagger n}^2) + o_p(n^{-1/2}h_*^{-1/2})$$

$$= -n^{-1}h_*^{-1} \sum_{i=1}^{n} \mathcal{K}_{*,i}(m) H^{(1)}(M_i)[\widetilde{M_i} - M_i] + o_p(n^{-1/2}h_*^{-1/2}),$$

where the last equality follows from Assumption (F3(ii)). Therefore, by Lemma B-6, we have

$$T_{H,n,2}(m) = T_{H,n,2a}(m)\{1 + o_p(1)\}$$
$$+ T_{H,n,2b}(m) + T_{H,n,2c}(m) + o_p(n^{-1/2}h_*^{-1/2}),$$

where

$$T_{H,n,2a}(m) = -\frac{1}{n^2 h_*} \sum_{i=1}^{n}\sum_{j=1}^{n} \iota_*^{\top}[f_M(m)\mathbf{M}_H]^{-1}\alpha_n^*(W_i, W_j; m),$$

$$T_{H,n,2b}(m) = -\frac{1}{nh_*} \sum_{i=1}^{n} \iota_*^{\top}[f_M(m)\mathbf{M}_H]^{-1}\mathcal{K}_{*,i}(m)H^{(1)}(M_i)\beta(W_i),$$

$$T_{H,n,2c}(m) = -\frac{1}{nh_*} \sum_{i=1}^{n} \iota_*^{\top}[f_M(m)\mathbf{M}_H]^{-1}\mathcal{K}_{*,i}(m)H^{(1)}(M_i)R_n(W_i),$$

with $\beta(\cdot)$, and $R_n(\cdot)$ defined as in Lemma B-7, and

$$\alpha_n^*(W_i, W_j; m) \equiv \alpha_n^{*I}(W_i, W_j; m) + \alpha_n^{*II}(W_i, W_j; m) - \alpha_n^{*III}(W_i, W_j; m), \text{ with}$$

$$\alpha_n^{*I}(W_i, W_j; m) \equiv h_1^{-d}\mathcal{K}_{*,i}(m)\int h_1^{-1}\iota_1^{\top}\mathbf{M}_r^{-1}\mathcal{K}_{1,j}(W_i)\frac{dP_1(Z_i)}{q(r_i, z_0)f_W(W_i)}\varepsilon_{r,j},$$

$$\alpha_n^{*II}(W_i, W_j; m) \equiv h_1^{-1}\mathcal{K}_{*,i}(m)\int h_1^{-d}\iota_1^{\top}\mathbf{M}_r^{-1}\mathcal{K}_{1,j}(W_i)\frac{dP_2(X_i)}{q(r_i, z_0)f_W(W_i)}\varepsilon_{r,j},$$

$$\alpha_n^{*III}(W_i, W_j; m) \equiv \mathcal{K}_{*,i}(m)\int\int h_1^{-(d+1)}\iota_1^{\top}\mathbf{M}_r^{-1}\mathcal{K}_{1,j}(W_i)\frac{dP_1(Z_i)dP_2(X_i)}{q(r_i, z_0)f_W(W_i)}\varepsilon_{r,j}.$$

Again, by applying a second order U–statistic theory (e.g. Powell, Stock, and Stoker (1989)), we can show that under Assumptions E and F,

$$T_{H,n,2a}(m) = O_p(n^{-1}h_1^{-d/2}h_*^{-1/2}) + O_p(n^{-1}h_1^{-1/2}h_*^{-1/2}) + O_p(n^{-1}h_*^{-1/2})$$
$$= o_p(n^{-1/2}h_*^{-1/2}).$$

By Bochner's Lemma and Lemma B-6, it follows that $T_{H,n,2b}(m)$ converges in mean squared to

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

which is $O(h_\dagger)$. Finally, the last term is bounded by

$$|T_{H,n,2c}(m)| \leq \max_{1\leq i\leq n} |R_n(W_i)| \max_{1\leq i\leq n} \left|H^{(1)}(M_i)\right| \left\{ n^{-1}h_*^{-1} \sum_{i=1}^{n} \left|\iota_*^\top [f_M(m)\mathbf{M}_H]^{-1} \mathcal{K}_{*,i}(m)\right| \right\},$$

where the right hand side is $o_p(n^{-1/2}h_*^{-1/2})$. $\blacksquare$

Lemma B-9 *Let Assumptions E and F hold, then*

$$T_{H,n,3}(m) = O_p\left(n^{-1/2}h_*^{-1/2}\right).$$

Proof. As in Lemma B-3, by Assumption (E4), we can write

$$\widehat{\Delta}_{H,i}(m) - \Delta_{H,i}(m) = \sum_{j=p_*+1} \frac{1}{j!} \left(\partial^j H(\widehat{m}^*)/\partial m^j\right) [(\widehat{M}_i - m)^j - (M_i - m)^j]$$
$$+ \sum_{j=p_*+1} \frac{1}{j!} \left[\partial^j H(\widehat{m}^*)/\partial m^j - \partial^j H(m^*)/\partial m^j\right] (M_i - m)^j, \qquad \text{(B-12)}$$

where $(\widetilde{m}^*, m^*)$ lie between $(\widetilde{M}_i, M_i)$ and m, such that $||\widetilde{m}^* - m^*|| = O_p(\nu_{\dagger n})$. For $\|M_i - m\| \leq ch_*$, $|\Delta_{H,i}(m)| = O_p(h_*^{p_*+1})$. These observations imply that (B-12) is $O_p(h_*^{p_*}\nu_{\dagger n})$, which by Assumption (F3(ii)), is $o_p(n^{-1/2}h_*^{-1/2})$. Therefore, we conclude that

$$|\widehat{\Delta}_{H,i}(m)| = O_p(h_*^{p_*+1}) + o_p(n^{-1/2}h_*^{-1/2})$$

uniformly in m and i for $\|M_i - m\| \leq ch_*$. Now, we write by the triangle inequality

$$\left|\widehat{\mathbf{B}}_{H,n}(m) - \mathbf{B}_{H,n}(m)\right| \leq \left\{\frac{1}{nh_*}\sum_{i=1}^{n} |\mathcal{K}_{*,i}(m)|\right\} \max_{1\leq i\leq n} \sup_b |\widehat{\Delta}_{H,i}(m) - \Delta_{H,i}(m)|$$
$$+ \left\{\frac{1}{nh_*}\sum_{i=1}^{n} \left|\widehat{\mathcal{K}}_{*,i}(m) - \mathcal{K}_{*,i}(m)\right|\right\} \max_{1\leq i\leq n} \sup_b |\widehat{\Delta}_{H,i}(m)|.$$

The sup of the first term is $o_p(n^{-1/2}h_*^{-1/2})$, and by Taylor–series expansion, the second term is

$$\frac{1}{nh_*^2}\sum_{i=1}^{n} \left|\mathcal{K}_{*,i}^{(1)}(m)\right| \left|\widetilde{M}_i - M_i\right| |\widehat{\Delta}_{H,i}(m)| + O_p((h_*^{-1}\nu_{\dagger n})^2) \max_{1\leq i\leq n} \sup_b |\widehat{\Delta}_{H,i}(m)|$$
$$= O_p\left(h_*^{-1}\nu_{\dagger n}\right) O_p(h_*^{p_*+1}) + O_p((h_*^{-1}\nu_{\dagger n})^2) O_p(h_*^{p_*+1})$$
$$= o_p(n^{-1/2}h_*^{-1/2}) \text{ uniformly in } m \text{ by Assumption (F3)}.$$

Therefore, we conclude that

$$\widehat{\mathbf{B}}_{H,n}(m) = \mathbf{B}_{H,n}(m) + o_p(n^{-1/2}h_*^{-1/2})$$

uniformly in m, and by Kolmogorov's Law of Large numbers, it follows that

$$T_{H,n,3}(m) = h_*^{p_*+1} \iota_*^\top \mathbf{M}_H^{-1} \mathbf{B}_H H^{(p_*+1)}(m) + o_p(n^{-1/2}h_*^{-1/2}),$$

as required. $\blacksquare$

Lemma B-10 *Let Assumptions E and F hold, then*

$$\mathcal{R}_{H,n}(m) = T_{H,n,4}(m) + T_{H,n,5}(m) + T_{H,n,6}(m)$$
$$= o_p\left(n^{-1/2}h_*^{-1/2}\right).$$

Proof. For a typical element of $\widehat{\mathbf{M}}_{H,n}(m) - \mathbf{M}_{H,n}(m)$ is given by

$$[\widehat{\mathbf{M}}_{H,n,j,k}(m)]_{l,l^0} - [\mathbf{M}_{H,n,j,k}(m)]_{l,l^0}$$
$$= \frac{1}{nh_*}\sum_{i=1}^{n}\left[\left(\frac{\widetilde{M_i}-m}{h_*}\right)^{\phi_{H;j}(l)+\phi_{H;k}(l^0)} K_2\left(\frac{\widetilde{M_i}-m}{h_*}\right)\right.$$
$$\left. - \left(\frac{M_i-m}{h_*}\right)^{\phi_{H;j}(l)+\phi_{H;k}(l^0)} K_2\left(\frac{M_i-m}{h_*}\right)\right].$$

After expanding the last expression at M_i, it is not difficult to show that

$$\sup_m |[\widehat{\mathbf{M}}_{H,n,j,k}(m)]_{l,l^0} - [\mathbf{M}_{H,n,j,k}(m)]_{l,l^0}| = O_p\left(h_*^{-1}\nu_{\dagger n}\right).$$

By the triangle inequality, we have

$$\sup_b |\widehat{\mathbf{M}}_{H,n}(m) - f_M(m)\mathbf{M}_H| \leq \sup_b |\widehat{\mathbf{M}}_{H,n}(m) - \mathbf{M}_{H,n}(m)|$$
$$+ \sup_b |\mathbf{M}_{H,n}(m) - f_M(m)\mathbf{M}_H|$$
$$= O_p\left(h_*^{-1}\nu_{\dagger n}\right) + O_p(n^{-1/2}h_*^{-1/2}\sqrt{\ln n} + h_*) = o_p(1),$$

where the last equality follows from Assumption (F3(ii)) and Corollary 2 (page 580) in Masry (1996a). Furthermore, by Assumption (F1), $\widehat{\mathbf{M}}_{H,n}^{-1}(m) = O_p(1)$ with probability approaching one. Therefore, Lemmas B-7–B-9 imply

$$T_{H,n,4}(m) = o_p(1)O_p(n^{-1/2}h_*^{-1/2}),$$
$$T_{H,n,5}(m) = o_p(1)O_p(n^{-1/2}h_*^{-1/2}), \text{ and}$$
$$T_{H,n,6}(m) = o_p(1)O_p(n^{-1/2}h_1^{-1/2}),$$

respectively. ∎

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

$\widehat{M}$ vs M		$\sigma_r^2 = 1$				$\sigma_r^2 = 2$			
cc	n	$IRMSE$		$IMAE$		$IRMSE$		$IMAE$	
0.5	150	0.2491	0.4526	0.2027	0.3119	0.3511	0.6744	0.2855	0.4795
	600	0.1299	0.2101	0.1067	0.1669	0.1826	0.3230	0.1493	0.2412
1	150	0.2491	0.4073	0.2028	0.2942	0.3512	0.6486	0.2854	0.4606
	600	0.1299	0.2026	0.1067	0.1592	0.1827	0.3028	0.1494	0.2264
1.5	150	0.2470	0.3888	0.2004	0.2854	0.3471	0.6292	0.2821	0.4405
	600	0.1250	0.1860	0.1024	0.1473	0.1752	0.2631	0.1434	0.2025

$\widehat{G}$ vs G		$\sigma_r^2 = 1$				$\sigma_r^2 = 2$			
cc	n	$IRMSE$		$IMAE$		$IRMSE$		$IMAE$	
0.5	150	0.1600	0.2863	0.1301	0.2420	0.2243	0.4374	0.1833	0.3415
	600	0.0854	0.1546	0.0697	0.1334	0.1192	0.2173	0.0972	0.1849
1	150	0.1600	0.2651	0.1301	0.2267	0.2242	0.3821	0.1832	0.3116
	600	0.0855	0.1467	0.0696	0.1247	0.1191	0.2028	0.0973	0.1731
1.5	150	0.1583	0.2558	0.1291	0.2170	0.2223	0.3988	0.1808	0.3198
	600	0.0820	0.1401	0.0671	0.1201	0.1136	0.1890	0.0926	0.1626

$\widehat{F}$ vs F		$\sigma_r^2 = 1$				$\sigma_r^2 = 2$			
cc	n	$IRMSE$		$IMAE$		$IRMSE$		$IMAE$	
0.5	150	0.1573	0.2359	0.1289	0.1878	0.2220	0.3223	0.1826	0.2532
	600	0.0815	0.1118	0.0661	0.0917	0.1153	0.1744	0.0937	0.1428
1	150	0.1571	0.2113	0.1289	0.1709	0.2221	0.2978	0.1823	0.2314
	600	0.0817	0.1050	0.0662	0.0865	0.1154	0.1624	0.0937	0.1323
1.5	150	0.1554	0.2028	0.1279	0.1608	0.2195	0.2849	0.1801	0.2262
	600	0.0775	0.0944	0.0628	0.0768	0.1094	0.1401	0.0886	0.1124

$\widehat{H}$ vs H		$\sigma_r^2 = 1$				$\sigma_r^2 = 2$			
cc	n	$IRMSE$		$IMAE$		$IRMSE$		$IMAE$	
0.5	150	0.2491	0.2874	0.2027	0.2241	0.3511	0.3939	0.2855	0.3089
	600	0.1299	0.1607	0.1067	0.1263	0.1826	0.2275	0.1493	0.1786
1	150	0.2491	0.2876	0.2028	0.2232	0.3512	0.3958	0.2854	0.3114
	600	0.1299	0.1607	0.1067	0.1262	0.1827	0.2270	0.1494	0.1784
1.5	150	0.2470	0.2816	0.2004	0.2178	0.3471	0.3908	0.2821	0.3044
	600	0.1250	0.1474	0.1024	0.1142	0.1752	0.2082	0.1434	0.1611

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

$\widehat{M}$ vs M		$\sigma_r^2 = 1$				$\sigma_r^2 = 2$			
cc	n	$IRMSE$		$IMAE$		$IRMSE$		$IMAE$	
0.5	150	0.2836	0.5179	0.2297	0.3518	0.4210	0.7313	0.3378	0.5354
	600	0.1421	0.2325	0.1148	0.1800	0.2023	0.3690	0.1643	0.2644
1	150	0.2834	0.4762	0.2297	0.3393	0.4206	0.7035	0.3382	0.5139
	600	0.1422	0.2172	0.1149	0.1710	0.2023	0.3334	0.1640	0.2454
1.5	150	0.2804	0.4545	0.2268	0.3245	0.4155	0.6845	0.3344	0.5064
	600	0.1363	0.1994	0.1109	0.1574	0.1923	0.2877	0.1566	0.2187

$\widehat{G}$ vs G		$\sigma_r^2 = 1$				$\sigma_r^2 = 2$			
cc	n	$IRMSE$		$IMAE$		$IRMSE$		$IMAE$	
0.5	150	0.1801	0.3243	0.1461	0.2702	0.2661	0.4954	0.2141	0.4090
	600	0.0925	0.1696	0.0752	0.1426	0.1301	0.2391	0.1058	0.1999
1	150	0.1800	0.3052	0.1461	0.2559	0.2663	0.4605	0.2141	0.3677
	600	0.0925	0.1588	0.0752	0.1334	0.1302	0.2218	0.1061	0.1842
1.5	150	0.1778	0.2939	0.1442	0.2495	0.2614	0.4745	0.2112	0.3775
	600	0.0887	0.1491	0.0723	0.1260	0.1243	0.2025	0.1014	0.1732

$\widehat{F}$ vs F		$\sigma_r^2 = 1$				$\sigma_r^2 = 2$			
cc	n	$IRMSE$		$IMAE$		$IRMSE$		$IMAE$	
0.5	150	0.1801	0.2553	0.1440	0.2027	0.2679	0.3525	0.2119	0.2812
	600	0.0908	0.1261	0.0741	0.1034	0.1300	0.1926	0.1061	0.1533
1	150	0.1800	0.2371	0.1439	0.1922	0.2678	0.3215	0.2124	0.2601
	600	0.0909	0.1153	0.0742	0.0934	0.1301	0.1734	0.1060	0.1392
1.5	150	0.1774	0.2272	0.1410	0.1822	0.2625	0.3105	0.2086	0.2533
	600	0.0861	0.1038	0.0701	0.0847	0.1229	0.1557	0.0999	0.1262

$\widehat{H}$ vs H		$\sigma_r^2 = 1$				$\sigma_r^2 = 2$			
cc	n	$IRMSE$		$IMAE$		$IRMSE$		$IMAE$	
0.5	150	0.2582	0.2875	0.2098	0.2240	0.3717	0.3910	0.3030	0.3083
	600	0.1321	0.1625	0.1069	0.1272	0.1869	0.2298	0.1525	0.1794
1	150	0.2582	0.2870	0.2099	0.2228	0.3717	0.3948	0.3031	0.3101
	600	0.1323	0.1626	0.1070	0.1273	0.1870	0.2283	0.1526	0.1788
1.5	150	0.2556	0.2818	0.2077	0.2183	0.3673	0.3890	0.2992	0.3038
	600	0.1263	0.1483	0.1032	0.1146	0.1777	0.2087	0.1457	0.1608

Table 3: Parametric General Production Function Estimates (P1)

Industry	θ_0		θ_1		θ_2		θ_3		θ_4		θ_5		γ	
	1995	2001	1995	2001	1995	2001	1995	2001	1995	2001	1995	2001	1995	2001
Chemical	9.622	9.771	0.452	0.498	0.932	0.799	0.064	0.092	0.047	0.001	0.053	0.050	-0.028	-0.022
	(0.043)	(0.036)	(0.041)	(0.032)	(0.056)	(0.028)	(0.046)	(0.022)	(0.046)	(0.025)	(0.027)	(0.015)	(0.038)	(0.001)
Iron	10.148	10.191	0.835	0.434	1.028	1.280	0.053	0.027	-0.162	0.251	0.060	-0.073	-0.032	0.301
	(0.122)	(0.390)	(0.113)	(0.133)	(0.146)	(0.370)	(0.116)	(0.071)	(0.069)	(0.105)	(0.047)	(0.085)	(0.098)	(0.253)
Petroleum	11.090	11.337	0.724	0.809	1.419	1.252	-0.224	0.060	0.272	0.030	-0.120	-0.079	0.061	0.120
	(0.217)	(0.290)	(0.183)	(0.101)	(0.339)	(0.440)	(0.186)	(0.071)	(0.175)	(0.098)	(0.114)	(0.160)	(0.118)	(0.194)
Transportation	8.963	9.630	0.493	0.493	1.538	1.071	-0.005	0.172	0.259	0.116	-0.084	0.023	0.546	0.118
	(0.588)	(0.150)	(0.147)	(0.048)	(0.512)	(0.165)	(0.070)	(0.037)	(0.117)	(0.053)	(0.117)	(0.049)	(0.389)	(0.124)

Table 4: Parametric Generalized Homothetic Estimates (P2)

Industry	α		β_0		β_1		β_2		γ	
	1995	2001	1995	2001	1995	2001	1995	2001	1995	2001
Raw chemical materials	0.500	0.605	9.609	9.736	0.926	0.877	0.053	0.024	-0.025	0.077
(Chemical)	(0.050)	(0.064)	(0.037)	(0.070)	(0.055)	(0.072)	(0.021)	(0.018)	(0.029)	(0.058)
Smelting and processing	0.563	0.653	9.634	10.518	1.570	0.816	-0.066	0.059	0.399	-0.019
of ferrous metals (Iron)	(0.130)	(0.112)	(0.417)	(0.082)	(0.329)	(0.071)	(0.058)	(0.020)	(0.246)	(0.028)
Petroleum processing	0.823	0.783	10.860	11.352	1.099	1.014	0.013	0.033	-0.054	0.004
(Petroleum)	(0.177)	(0.155)	(0.102)	(0.128)	(0.128)	(0.131)	(0.031)	(0.032)	(0.026)	(0.034)
Transportation equipment	0.689	0.601	9.605	9.819	0.923	0.911	0.075	0.073	0.017	0.021
(Transportation)	(0.092)	(0.061)	(0.090)	(0.060)	(0.089)	(0.059)	(0.027)	(0.019)	(0.079)	(0.043)

Table 5: Parametric Translog Estimates (P3)

Industry	α		β_0		β_1		β_2	
	1995	2001	1995	2001	1995	2001	1995	2001
Raw chemical materials	0.479	0.696	9.585	9.815	0.961	0.783	0.045	0.036
(Chemical)	(0.041)	(0.040)	(0.026)	(0.032)	(0.035)	(0.028)	(0.019)	(0.013)
Smelting and processing	0.932	0.621	10.262	10.499	1.025	0.847	0.017	0.054
of ferrous metals (Iron)	(0.095)	(0.098)	(0.064)	(0.079)	(0.055)	(0.051)	(0.019)	(0.018)
Petroleum processing	0.674	0.795	10.810	11.351	1.271	1.001	-0.020	0.036
(Petroleum)	(0.121)	(0.125)	(0.100)	(0.125)	(0.093)	(0.081)	(0.031)	(0.028)
Transportation equipment	0.705	0.626	9.623	9.839	0.905	0.883	0.078	0.079
(Transportation)	(0.062)	(0.045)	(0.038)	(0.038)	(0.046)	(0.031)	(0.024)	(0.016)

Table 6: Average Substitutability, $T(k, L)$.

Industry	P2		P3		NP	
	1995	2001	1995	2001	1995	2001
Raw chemical materials	0.728	0.384	0.738	0.390	0.795	1.328
(Chemical)	(0.042)	(0.135)	–	–	(0.135)	(0.762)
Smelting and processing	0.163	0.474	0.081	0.480	0.383	1.188
of ferrous metals (Iron)	(0.376)	(0.108)	–	–	(0.510)	(0.244)
Petroleum processing	0.286	0.234	0.442	0.228	0.451	0.255
(Petroleum)	(0.165)	(0.023)	–	–	(0.259)	(0.158)
Transportation equipment	0.351	0.473	0.352	0.470	0.147	0.834
(Transportation)	(0.020)	(0.049)	–	–	(0.321)	(0.842)

Table 7: Average Return to Scale, $RTS(M, L)$.

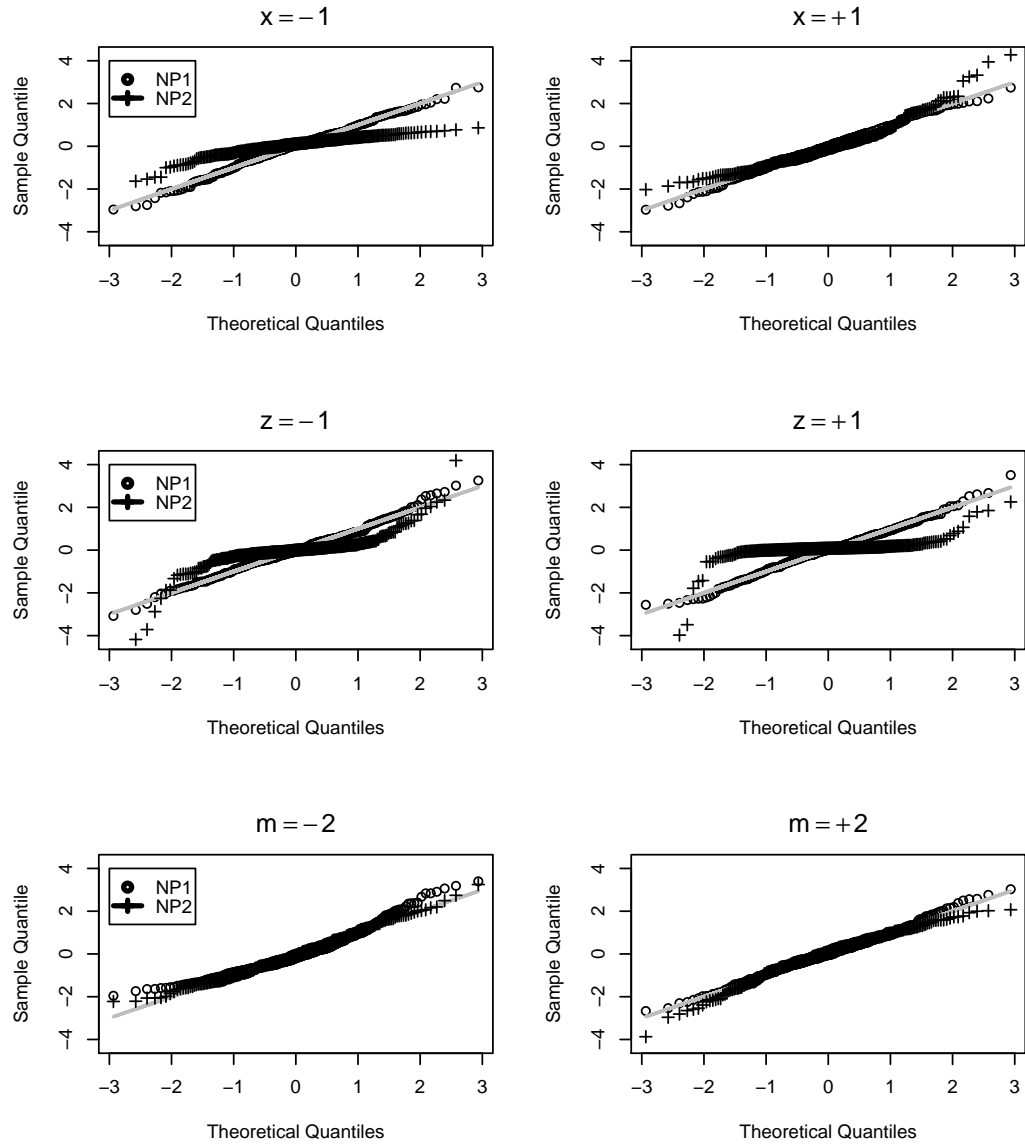| Industry	P2		P3		NP	
	1995	2001	1995	2001	1995	2001
Raw chemical materials	0.962	0.796	0.968	0.799	1.016	0.752
(Chemical)	(0.074)	(0.120)	(0.073)	(0.077)	(0.166)	(0.181)
Smelting and processing	1.186	0.881	1.035	0.881	1.034	0.828
of ferrous metals (Iron)	(0.428)	(0.178)	(0.049)	(0.179)	0.373)	(0.201)
Petroleum processing	1.231	1.020	1.258	1.016	1.162	0.946
(Petroleum)	(0.279)	(0.124)	(0.060)	(0.126)	(0.173)	(0.147)
Transportation equipment	0.934	0.901	0.932	0.896	0.944	0.964
(Transportation)	(0.153)	(0.183)	(0.152)	(0.182)	(0.291)	(0.328)

Figure 1: $Q - Q$ plots for G, F and H.
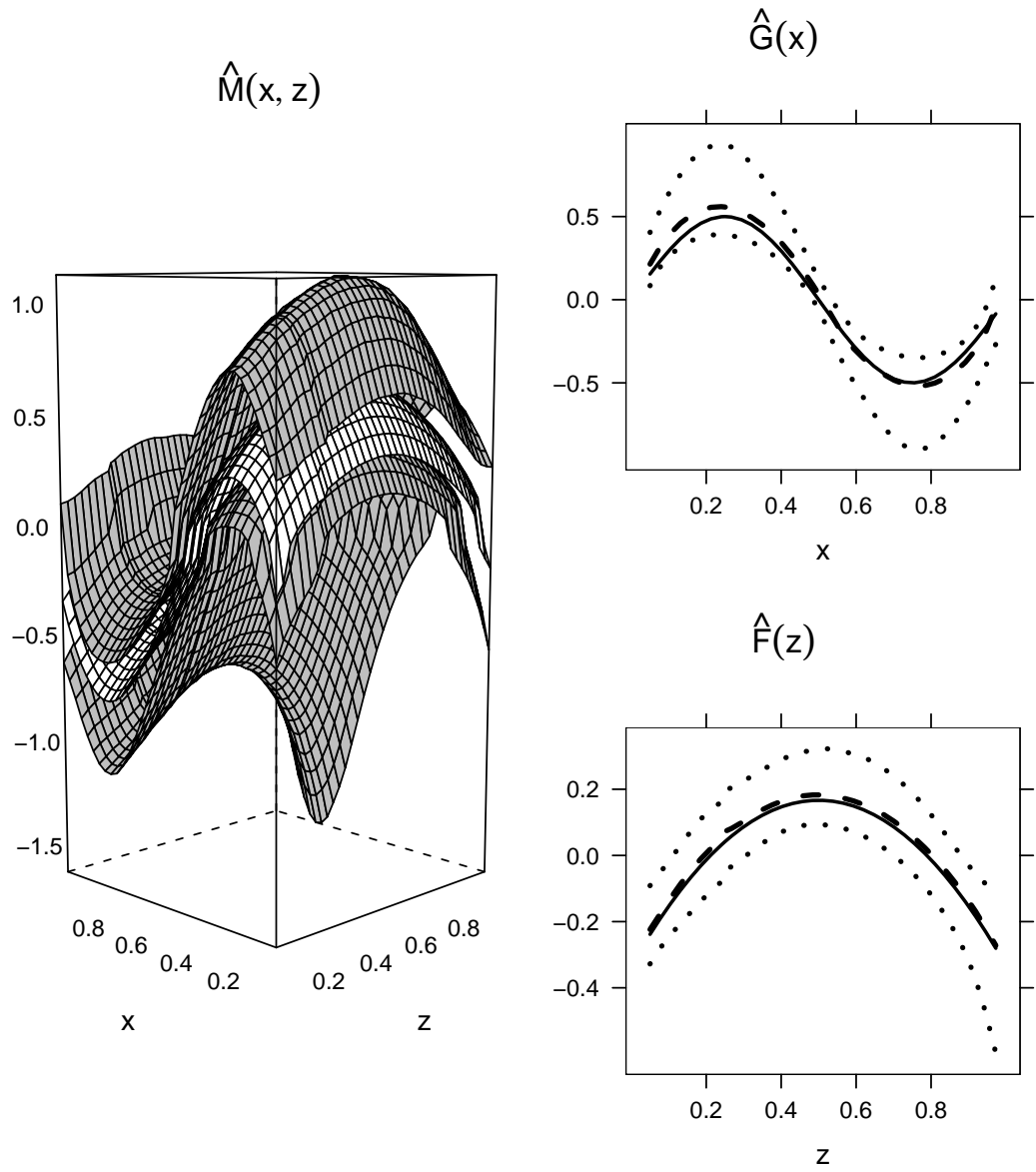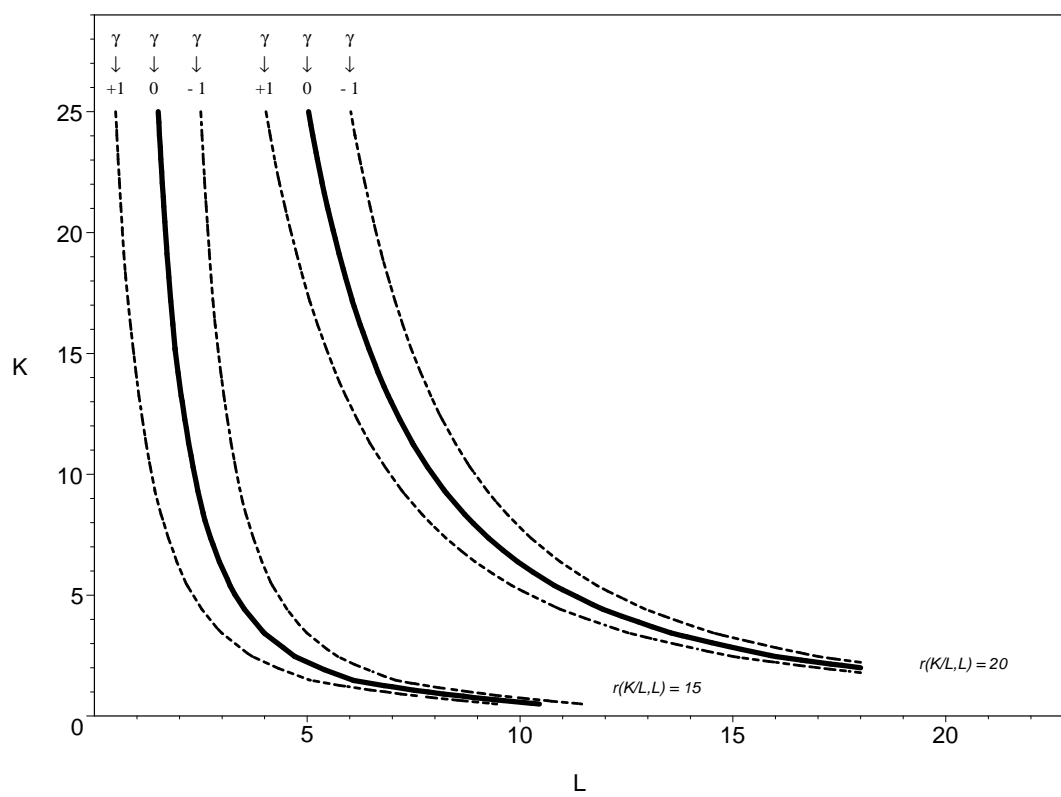
Figure 2: Simulation Envelopes for M, G and F.
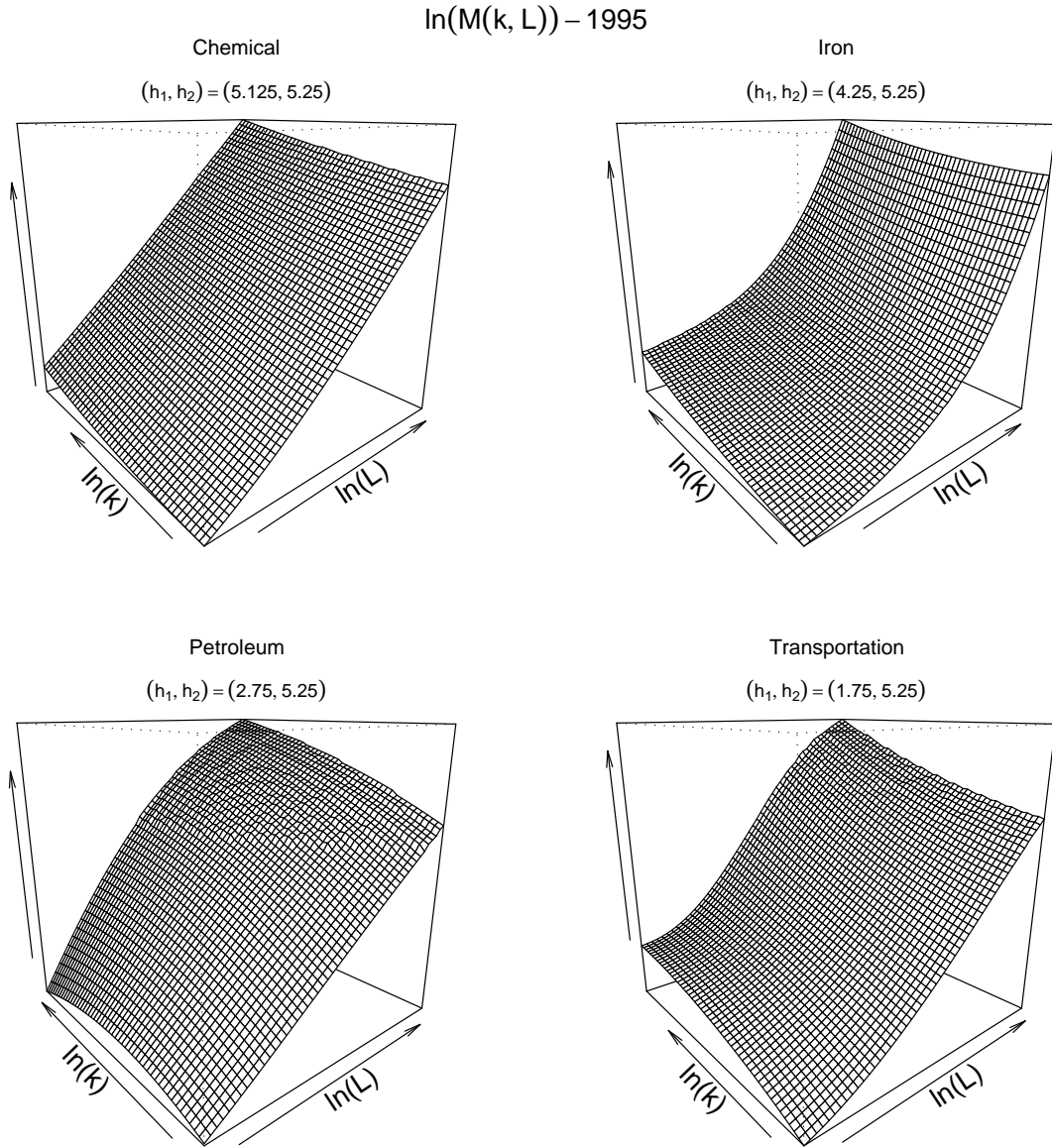
Figure 3: Generalized Homothetic Translog Isoquants.

ln(M(k, L)) − 1995

Chemical

$(h_1, h_2) = (5.125, 5.25)$

ln(k) ln(L)

Iron

$(h_1, h_2) = (4.25, 5.25)$

ln(k) ln(L)

Petroleum

$(h_1, h_2) = (2.75, 5.25)$

ln(k) ln(L)

Transportation

$(h_1, h_2) = (1.75, 5.25)$

ln(k) ln(L)

Figure 4: Generalized Homogeneous M.

64
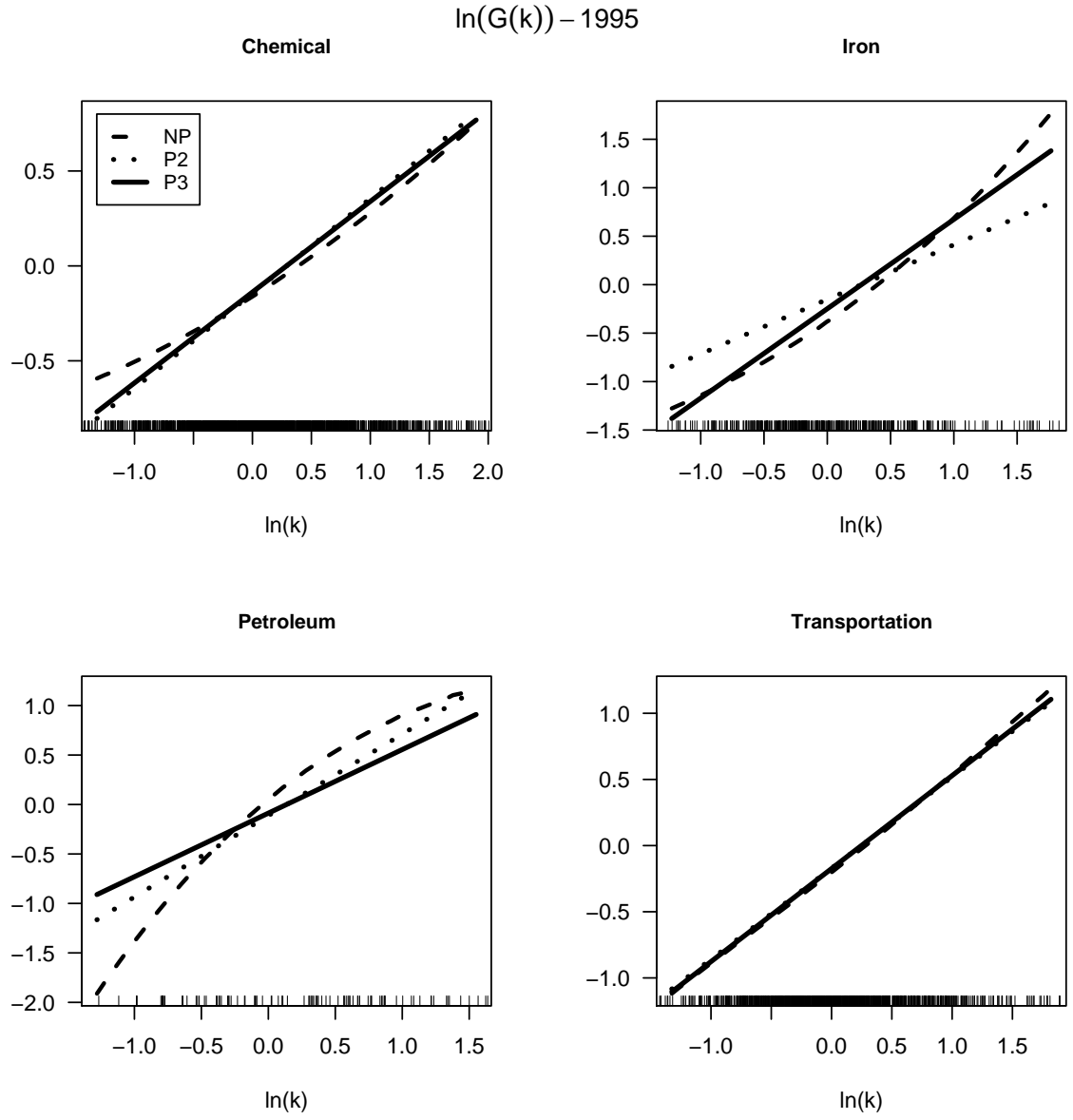
Figure 5: Generalized Homogeneous Component G.
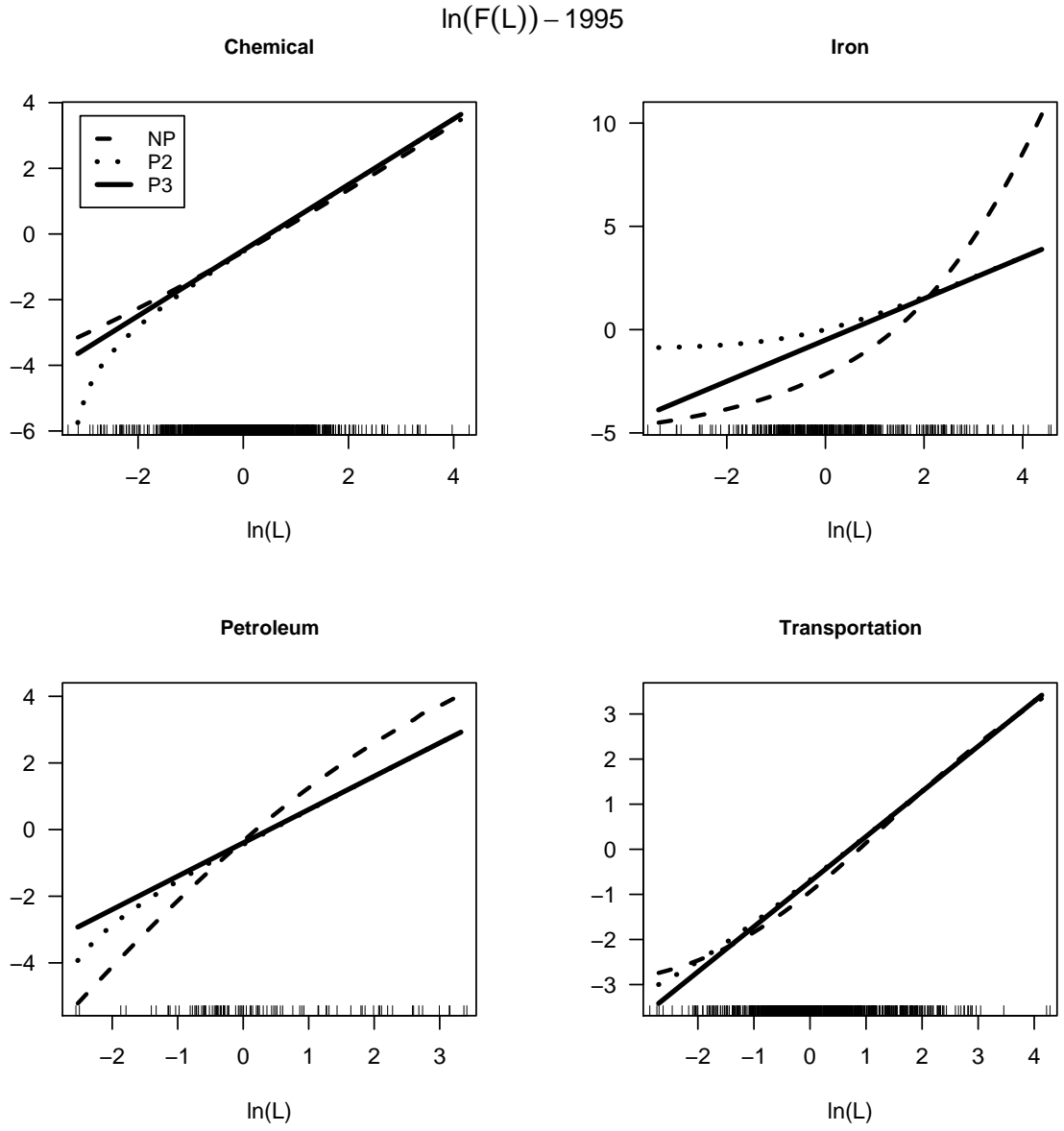
Figure 6: Generalized Homogeneous Component F.
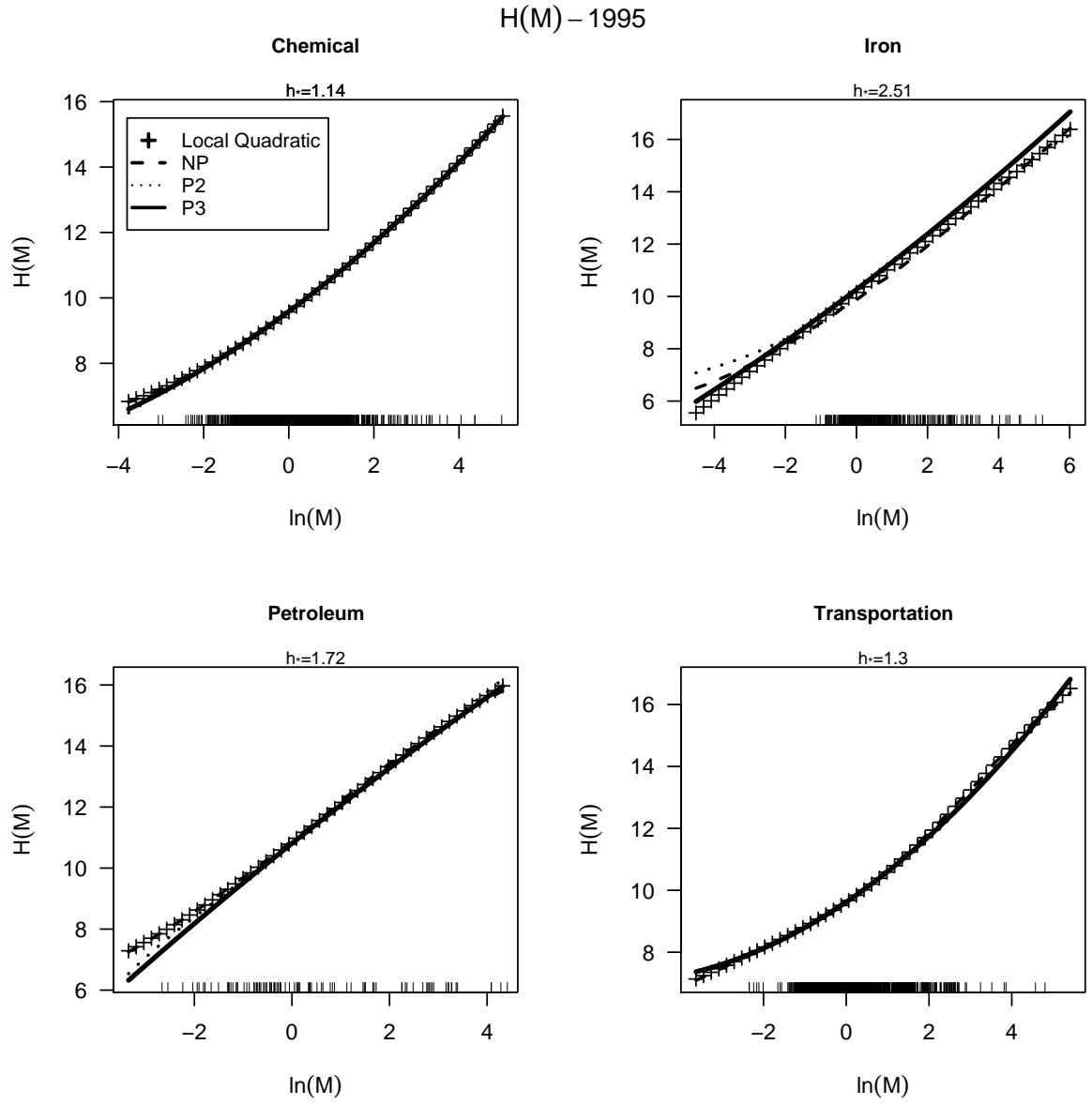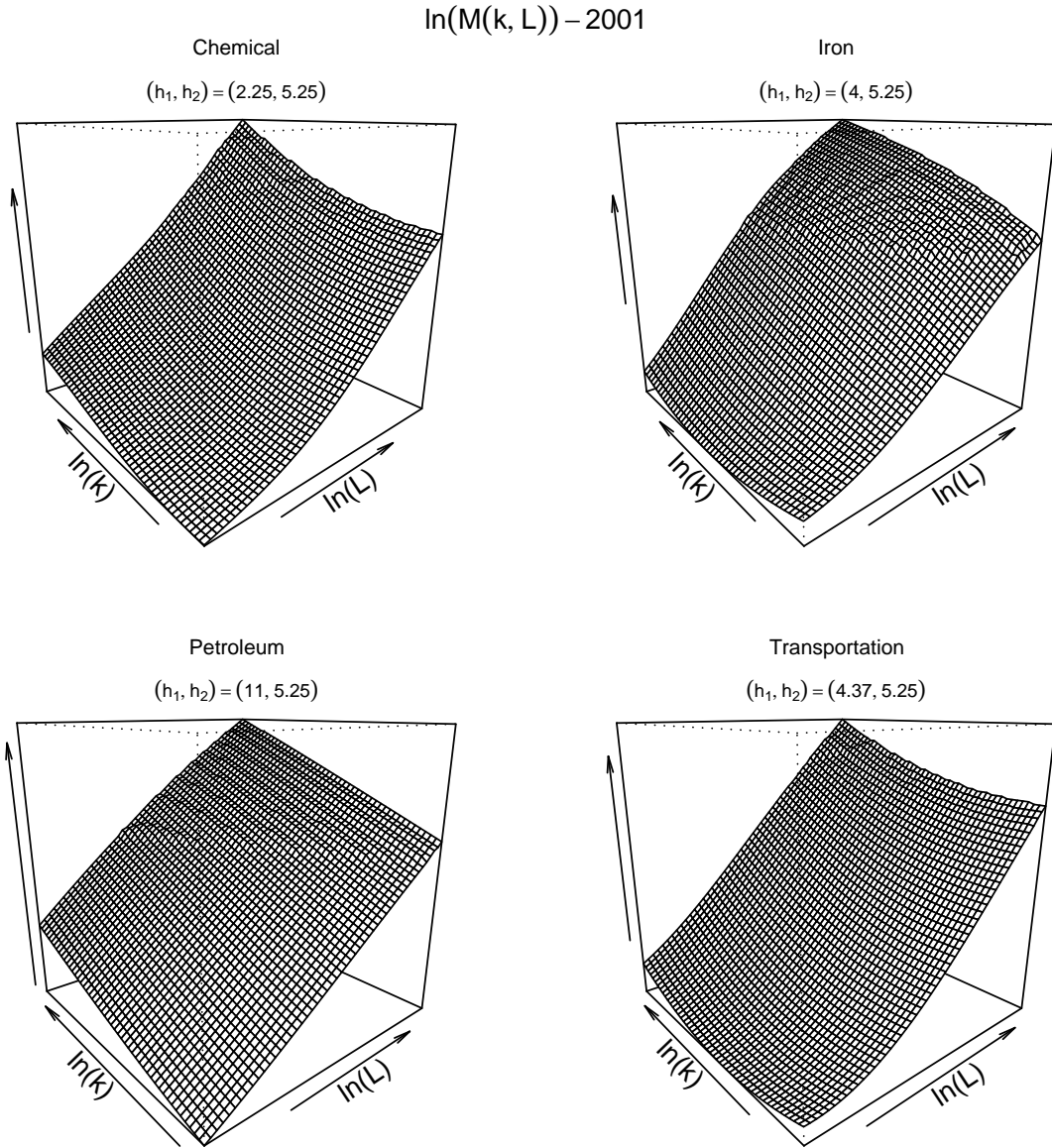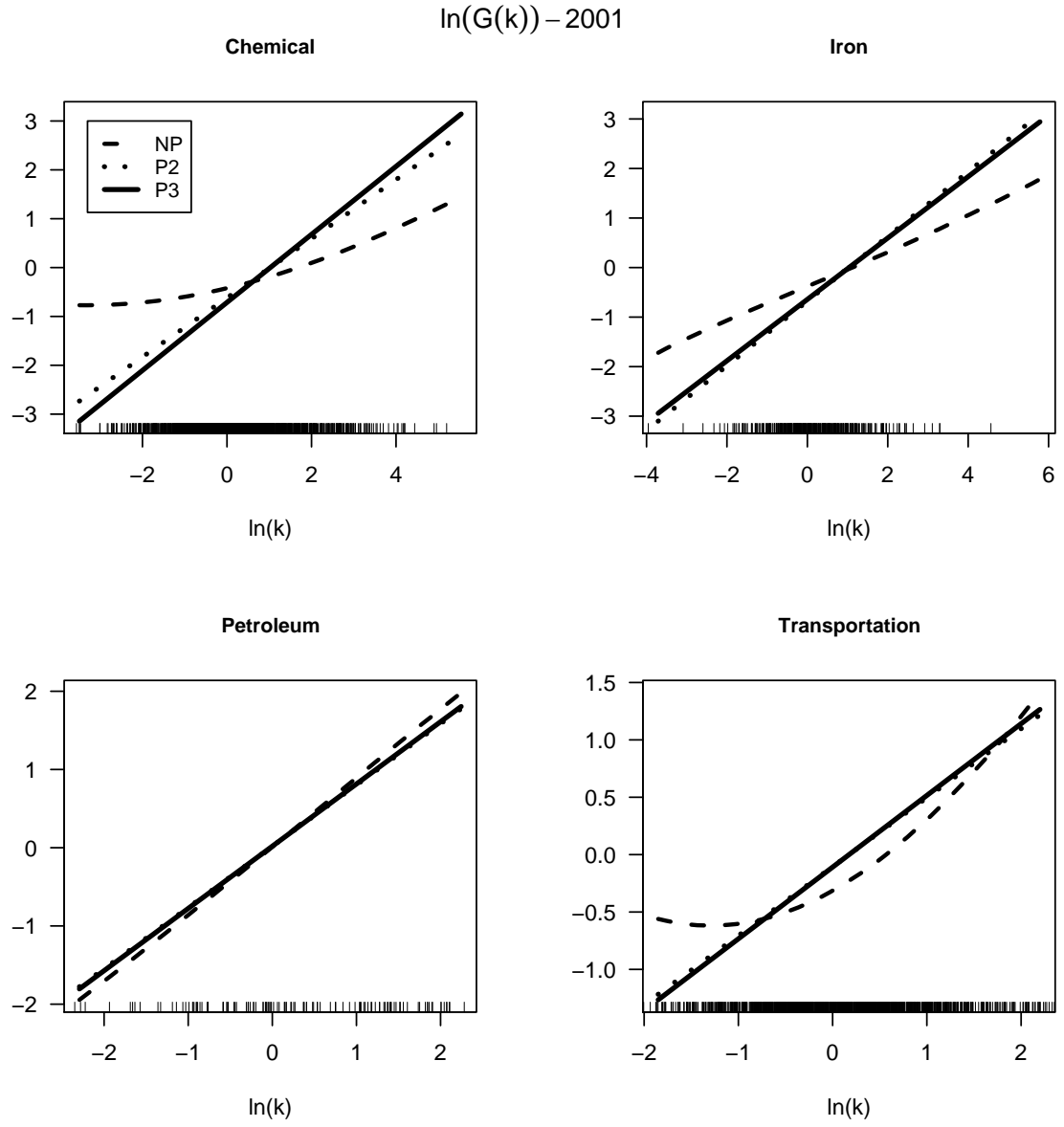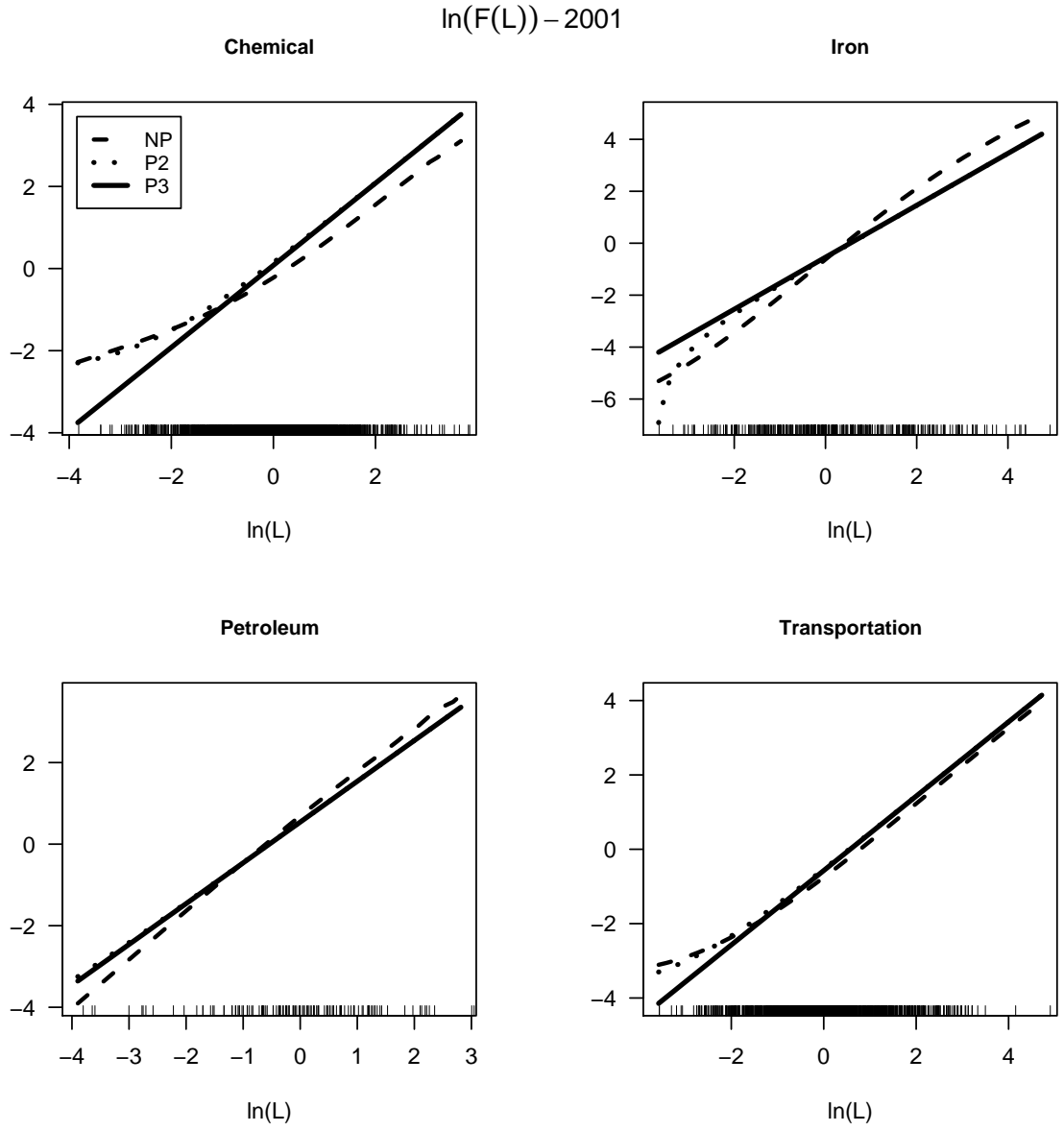
Figure 7: Strictly Monotonic Component H.

ln(M(k, L)) − 2001

Chemical

$(h_1, h_2) = (2.25, 5.25)$

Iron

$(h_1, h_2) = (4, 5.25)$

Petroleum

$(h_1, h_2) = (11, 5.25)$

Transportation

$(h_1, h_2) = (4.37, 5.25)$

Figure 8: Generalized Homogeneous M.

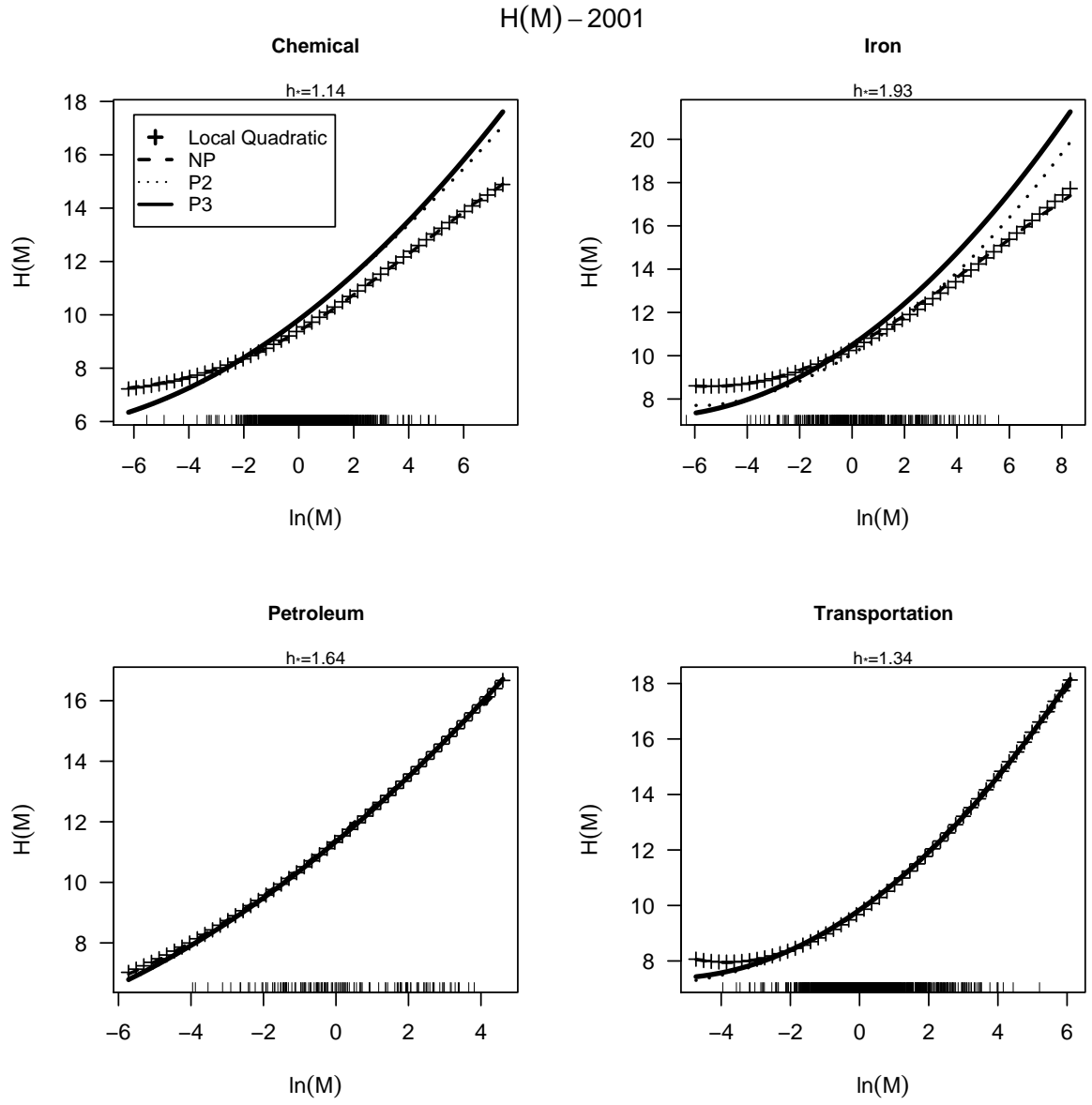Figure 9: Generalized Homogeneous Component G.

Figure 10: Generalized Homogeneous Component F.

Figure 11: Strictly Monotonic Component H