Smooth Coefficient Models With Endogenous Environmental Variables

Michael S. Delgado[*]
Department of Agricultural Economics
Purdue University

Deniz Ozabaci[†]
Department of Economics
University of Connecticut

Yiguo Sun[‡]
Department of Economics and Finance
University of Guelph

Subal C. Kumbhakar[§]
Department of Economics
Binghamton University

March 9, 2015

Abstract

We propose a three-step estimator for a structural semiparametric smooth coefficient model with endogenous variables in the *nonparametric* component. We use a control function approach to avoid the ill-posed inverse problem, and use a combination of series and kernel estimators to obtain consistent and asymptotically normal estimators of the functionals and their partial derivatives. Further, the third step estimator is oracle efficient. A Monte Carlo simulation illustrates the finite sample performance of our estimator, and we provide a complete empirical application to a model of household gasoline demand.

Keywords: Control function; Endogeneity; Gasoline demand; Instrumental variables; Kernel estimation; Series estimation; Smooth coefficient models.

JEL Codes: C14; C26

[*]Michael S. Delgado, Department of Agricultural Economics, Purdue University, West Lafayette, IN 47907-2056. Phone: 765-494-4211, Fax: 765-494-9176, E-mail: delgado2@purdue.edu.

[†]Corresponding author: Deniz Ozabaci, Department of Economics, University of Connecticut, 365 Fairfield Way, U-1063 Storrs, CT, 06269, E-mail:deniz.ozabaci@uconn.edu.

[‡]Yiguo Sun, Department of Economics and Finance, University of Guelph, Guelph, ON, N1G 2W1, E-mail:yisun@uoguelph.ca.

[§]Subal C. Kumbhakar, Department of Economics, Binghamton University, PO Box 6000, Binghamton, NY 13902-6000. Phone: 607-777-4762. Fax: 607-777-2681. E-mail: kkar@binghamton.edu.

1 Introduction

We develop a three-step estimator for a structural semiparametric smooth coefficient model in which the endogenous variables appear in the nonparametric part of the regression. We contribute to the growing literature on structural estimation of both nonparametric and semiparametric regression models. Seminal papers discussing structural estimation of fully nonparametric models include Roehrig (1988), Newey, Powell & Vella (1999), Pinkse (2000), Ai & Chen (2003), Newey & Powell (2003), Hall & Horowitz (2005), Su & Ullah (2008), Darolles, Fan, Florens & Renault (2011), Horowitz (2011), and Chen & Pouzo (2012). Estimators for semiparametric structural regression models have been developed for the partial linear structure, assuming the endogenous variables appear in either the parametric or nonparametric part of the regression (Yao & Martins-Filho 2012, Delgado & Parmeter 2014); the additive structure (Ozabaci, Henderson & Su 2014); and the smooth coefficient structure (Das 2005, Cai, Das, Xiong & Wu 2006, Cai & Li 2008, Cai & Xiong 2012, Su, Murtazashvili & Ullah 2013). In the latter case, however, contributions thus far have only led to estimators suited for structural estimation in which the endogenous variables appear only in the parametric part of the regression. Our contribution, therefore, is to develop an estimator that allows for endogenous variables in the nonparametric part of a smooth coefficient regression model.

While it may appear that our structure is a straightforward extension of the structural smooth coefficient models with endogenous variables only in the parametric part of the model, we illustrate that the estimation procedure under our assumed structure with endogenous variables in the nonparametric part of the model is substantially more complex. We consider a smooth coefficient model given by

$$\begin{cases} Y_t = m(Z_t, X_t) + \epsilon_t \\ m(Z_t, X_t) = \beta_0(Z_t) + \sum_{j=1}^{d_x} \beta_j(Z_t) X_{j,i} \equiv \mathcal{X}_t^T \beta(Z_t),\ \mathcal{X}_t \equiv (1, X_t)^T \\ Z_t = g\left(W_t, X_t\right) = \mu_z + \sum_{j=1}^{d_w} g_{w,j}(W_{j,t}) + \sum_{j=1}^{d_x} g_{x,j}(X_{j,t}) + U_t \\ E\left(U_t | W_t, X_t\right) = 0 \\ E\left(\epsilon_t | W_t, X_t, U_t\right) = E\left(\epsilon_t | U_t\right),\ t = 1, 2, \ldots, T, \end{cases} \tag{1.1}$$

where Y_t is a scalar outcome, $X_t = (X_{1,t}, X_{2,t}, \ldots, X_{d_x,t})^T$ is a $d_x \times 1$ vector of exogenous regressors, $\beta(Z_t)$ is a $(d_x + 1)$-dimensional vector of unspecified nonparametric structural coefficient functions, and ϵ_t is a zero mean error term. We define Z_t to be an endogenous variable, and assume a triangular system of equations in which Z_t is additively separable in the nonparametric functions of W_t and X_t, in which W_t is a $d_w \geq 1$ dimensional vector of instrumental variables.[1] We follow Newey et al. (1999) and assume that (W_t, X_t) is

[1]In general, one might assume that Z_t is a $d_z \geq 1$ dimensioned vector of endogenous variables. Such a setup would imply that U_t is a d_z-dimensioned vector of errors, and $d_w \geq d_z$ would be necessary for identification. However, the implication of this setup in the context of Equation (1.1) would be that the

orthogonal to U_t. Clearly, the endogeneity of Z_t precludes direct estimation of $\beta(Z_t)$ using standard nonparametric methods (e.g., Li, Huang, Li & Fu 2002).

Given the conditional moment assumptions in (1.1), identification follows from Newey et al. (1999):

$$\begin{aligned} E\left(Y_t|Z_t, X_t, W_t, U_t\right) &= m(Z_t, X_t) + E\left(\epsilon_t|Z_t, X_t, W_t, U_t\right) \\ &= m(Z_t, X_t) + E\left(\epsilon_t|X_t, W_t, U_t\right) \\ &= m(Z_t, X_t) + E\left(\epsilon_t|U_t\right) \end{aligned} \tag{1.2}$$

which, after setting $E\left(\epsilon_t|U_t\right) \equiv \gamma(U_t)$ yields

$$m_0(Z_t, X_t, U_t) \equiv E\left(Y_t|Z_t, X_t, U_t\right) = \beta_0(Z_t) + \sum_{j=1}^{d_x} \beta_j(Z_t) X_{j,t} + \gamma(U_t). \tag{1.3}$$

By Theorem 2.1 in Newey et al. (1999), $\beta_0(Z_t)$ and $\gamma(U_t)$ are identified up to an additive constant as long as there does not exist an additive functional relationship between Z_t and U_t. If one is further willing to assume that $E\left(\epsilon_t\right) = 0$, $\beta_0(Z_t)$ and $\gamma(U_t)$ can be exactly identified.

Our structure is clearly more complicated than the structural smooth coefficient models in which the endogenous variables only appear in the parametric part of the regression, because under those assumptions estimators simply manage endogeneity by substituting fitted values for endogenous regressors or use a GMM type approach. Further, estimation of Equation (1.3) is complicated because of the appearance of two additive nonparametric functions, $\beta_0(Z_t)$ and $\gamma(U_t)$, because the slope coefficients $\beta(Z_t)$ are smooth functions instead of constant parameters, and because U_t is not generally observed. In the fully nonparametric models of Newey et al. (1999) and Su & Ullah (2008), their control function equations analogous to our Equation (1.3) were standard *additive* nonparametric functions, which are easily estimable by marginal integration, or related tools. In a partial linear structure, i.e. in which $\beta(Z_t) = \beta$, estimation would proceed following a straightforward conditional mean transformation (Robinson 1988, Christopeit & Hoderlein 2006, Delgado & Parmeter 2014). In Ozabaci et al. (2014), the functions are all by assumption additively separable; hence no interaction is possible unless an additional partial linear structure is imposed. We might also allow for additional partial linear interactions as in Ozabaci et al. (2014), but given the flexible interactions provided by the smooth coefficient structure, these additional parametric interactions seem less practical. Hence, compared to other common structures typically assumed, our smooth coefficient structure adds some flexibility but at the same time creates additional complications for

last term in the Z_t equation would be $\sum_{j=1}^{d_z} U_{j,t}$. Estimation of this slightly generalized setup would simply be repetition of our proposed control function approach for each variable in Z_t one at a time. To simplify notation slightly, our derivation assumes Z_t is a scalar.

estimating the structural functions.

To estimate the functionals and their first-order partial derivatives given our smooth coefficient structure, we elect to combine both series and kernel estimators in order to ensure that our third step estimators are oracle efficient (Horowitz & Mammen 2004, Ozabaci et al. 2014). As is evident from Equation (1.3), we deploy a control function approach for controlling for endogeneity (among many others, Newey et al. 1999, Su & Ullah 2008). We prove that our estimator is consistent, asymptotically normal, and oracle efficient. Moreover, in our theoretical developments, we allow for both independent and weakly dependent data. Hence, the estimator builds on the current literature by providing an estimator for a semiparametric structure that has not previously been considered, as well as further developing semiparametric regression estimators that achieve a high level of efficiency for both independent and dependent data. We support the theoretical developments with a Monte Carlo simulation that illustrates the effectiveness and practicality of our estimator in finite samples. We conduct our simulations under both data independence and weak dependence, and show that our estimator performs well by different performance measures.

To provide practical motivation for our assumed structure, we recognize that practitioners are often motivated to consider semiparametric structures given well-known dimensionality issues often present in fully nonparametric regression models. Among the commonly deployed semiparametric structures, the smooth coefficient model is particularly suited for studying empirical economic phenomenon, as the structure provides a straightforward generalization of the workhorse linear parametric model by adding flexibility in the coefficients that amounts to allowing for interactions between X and arbitrary forms of Z. In many contexts, these interactions may have important implications for economic policy (see, for instance, Delgado, McCloud & Kumbhakar 2014, Liu 2014). It is straightforward then, to econometrically test such generalizations against the standard linear model using specification tests (Cai, Fan & Yao 2000, Li et al. 2002), which may further support the development of robust economic policies. Recently, the smooth coefficient model has been applied in many areas of economics, including economic growth (Durlauf, Kourtellos & Minkin 2001, Mamuneas, Savvides & Stengos 2006, Delgado et al. 2014), microfinance (Hartarska, Parmeter & Nadolynak 2011, Delgado, Parmeter, Hartarska & Nadolynak 2015), environmental economics (Wadud, Noland & Graham 2010, Delgado 2013, Liu 2014), and firm productivity (Zhang, Sun, Delgado & Kumbhakar 2012).

To specifically illustrate the applicability of the proposed estimator, we develop a structural smooth coefficient model of household gasoline demand. We build on recent studies that have identified significant heterogeneity in gasoline consumption via semiparametric smooth coefficient regression (Wadud et al. 2010, Liu 2014) by exploring endogeneity within the structural coefficients. We find that household income is an important source of heterogeneity in the gasoline demand relationship, and that our estimator is able to correct for endogeneity that leads to a bias in several of our coefficient functions if estimated using a standard non-instrumental variables estimator.

This rest of the paper is outlined as follows. In Section 2 we develop our proposed estimator. In Section 3 we establish the asymptotic properties of our estimator and Section 4 contains the setup and results from our Monte Carlo study. Section 5 develops an empirical application to household gasoline demand, and Section 6 provides concluding remarks. We provide proofs for our asymptotic results in the Appendix.

2 Smooth Coefficient IV Estimation

We start with the structure in (1.1). Our main objective is to get estimators of $\beta(Z_t)$ and its first order partial derivative, $\dot{\beta}(Z_t)$. In doing so we follow these steps:

1. Calculate $\hat{U}_t$ from the reduced-form equation

$$Z_t = \mu_z + \sum_{j=1}^{d_w} g_{w,j}(W_{j,t}) + \sum_{j=1}^{d_x} g_{x,j}(X_{j,t}) + U_t.$$

2. Use sieve estimation to model $Y_t = \mathcal{X}_t^T \beta(Z_t) + \gamma(\hat{U}_t) + e_t$ to obtain the estimates $\hat{\beta}(Z_t)$ and $\hat{\gamma}(\hat{U}_t)$ for $t = 1, 2, \ldots, T$. Then, calculate the partial residuals $Y_{bx,t} = Y_t - \hat{\gamma}(\hat{U}_t)$.

3. Apply a local linear least squares kernel estimator to model $Y_{bx,t} = \mathcal{X}_t^T \beta(Z_t) + v_{1,t}$ to obtain $\widetilde{\beta}(Z_t)$ and $\widetilde{\dot{\beta}}(Z_t)$, in which $\widetilde{\dot{\beta}}(Z_t)$ is our final step estimator of the first order partial derivative of $\beta(Z_t)$.

Let $\{p_1(\cdot), p_2(\cdot), \ldots\}$ be a sequence of twice continuously differentiable orthonormal basis functions on $L_2[0,1]$. Also, for some integer $\kappa > 0$, we denote $\mathbf{p}^\kappa(v) = [p_1(v), \ldots, p_\kappa(v)]^T$, $P^\kappa(w,x) = \left[1, \mathbf{p}^\kappa(w_1)^T, \ldots, \mathbf{p}^\kappa(w_{d_w})^T, \mathbf{p}^\kappa(x_1)^T, \ldots, \mathbf{p}^\kappa(x_{d_x})^T\right]^T$, and $Q^\kappa(z,x,u) = \left[\mathbf{p}^\kappa(z)^T, \mathbf{p}^\kappa(z)^T x_1 \ldots, \mathbf{p}^\kappa(z)^T x_{d_x}, \mathbf{p}^\kappa(u)^T\right]^T$.

In Step 1, we approximate respectively $g_{w,j}(w_j)$ and $g_{x,j'}(x_{j'})$ by

$$g^*_{w,j}(w_j) = \mathbf{p}^{\kappa_{1,T}}(w_j)^T \alpha_j \text{ and } g^*_{x,j'}(x_j) = \mathbf{p}^{\kappa_{1,T}}(x_{j'})^T \vartheta_{j'},\ j = 1, \ldots, d_w, j' = 1, \ldots, d_x,$$

where α_j and $\vartheta_{j'}$ are $\kappa_{1,T} \times 1$ vectors, and we denote $\alpha = \left[\alpha_1^T, \ldots, \alpha_{d_w}^T\right]^T$ and $\vartheta = \left[\vartheta_1^T, \ldots, \vartheta_{d_x}^T\right]^T$. The sieve estimator of $g\left(w,x\right)$ is given by

$$
\begin{aligned}
& \hat{g}\left(w,x\right) \\
= \; & P^{\kappa_{1,T}}\left(w,x\right)^T \left[\hat{\mu}_z, \hat{\alpha}^T, \hat{\vartheta}^T\right]^T \\
= \; & P^{\kappa_{1,T}}\left(w,x\right)^T \left[T^{-1}\sum_{t=1}^{T} P^{\kappa_{1,T}}\left(W_t,X_t\right) P^{\kappa_{1,T}}\left(W_t,X_t\right)^T\right]^{+} T^{-1}\sum_{t=1}^{T} P^{\kappa_{1,T}}\left(W_t,X_t\right) Z_t \\
\equiv \; & \hat{\mu}_z + \sum_{j=1}^{d_w} \hat{g}_{w,j}(W_{j,t}) + \sum_{j=1}^{d_x} \hat{g}_{x,j}(X_{j,t})
\end{aligned}
$$

where A^+ denotes the Moore-Penrose generalized inverse of A, $\hat{g}_{w,j}(W_{j,t}) \equiv \mathbf{p}^{\kappa_{1,T}}\left(W_{j,t}\right)^T \hat{\alpha}_j$, $\hat{g}_{x,j}(X_{j,t}) \equiv \mathbf{p}^{\kappa_{1,T}}\left(X_{j,t}\right)^T \hat{\vartheta}_j$, and

$$
\left[\hat{\mu}_z, \hat{\alpha}^T, \hat{\vartheta}^T\right]^T = \arg\min_{(\mu_z,\alpha,\vartheta)} \sum_{t=1}^{T} \left\{Z_t - P^{\kappa_{1,T}}\left(W_t,X_t\right)^T \left[\mu_z, \alpha^T, \vartheta^T\right]^T\right\}^2.
$$

In Step 2, we approximate respectively $\gamma(u)$ and $\beta_j(z)$ by

$$
\gamma^*(u) = \mathbf{p}^{\kappa_2,T}\left(u\right)^T \rho \text{ and } \beta_j^*(z) = \mathbf{p}^{\kappa_{2,T}}\left(z\right)^T \psi_j,\; j = 0, 1, \ldots, d_x
$$

where ρ and ψ_l's are $\kappa_{2,T} \times 1$ vectors, and we denote $\psi = \left[\psi_0^T, \psi_1^T, \ldots, \psi_{d_x}^T\right]^T$. The sieve estimator of $m_0(z,x,u)$ is given by

$$
\begin{aligned}
\hat{m}_0(z,x,u) \;&\equiv\; Q^{\kappa_{2,T}}\left(z,x,u\right)^T \left[\hat{\psi}^T, \hat{\rho}^T\right]^T \\
&=\; Q^{\kappa_{2,T}}\left(z,x,u\right)^T \left[T^{-1}\sum_{t=1}^{T} Q^{\kappa_{2,T}}\left(Z_t,X_t,\hat{U}_t\right) Q^{\kappa_{2,T}}\left(Z_t,X_t,\hat{U}_t\right)^T\right]^{+} \\
&\quad \times T^{-1}\sum_{t=1}^{T} Q^{\kappa_{2,T}}\left(Z_t,X_t,\hat{U}_t\right) Y_t \\
&\equiv\; \hat{\beta}_0(z) + \sum_{j=1}^{d_x} \hat{\beta}_j(z) X_{j,t} + \hat{\gamma}(u),
\end{aligned}
$$

where $\hat{\beta}_j(z) \equiv \mathbf{p}^{\kappa_{2,T}}\left(z\right)^T \hat{\psi}_j$ for $j = 0, \ldots, d_x$, $\hat{\gamma}(u) = \mathbf{p}^{\kappa_{2,T}}\left(u\right)^T \hat{\rho}$,

$$
\left[\hat{\psi}^T, \hat{\rho}^T\right]^T = \arg\min_{(\psi,\rho)} \sum_{t=1}^{T} \left\{Y_t - \mathbf{p}^{\kappa_{2,T}}\left(Z_t\right)^T \psi_0 - \sum_{j=1}^{d_x} X_{j,t} \mathbf{p}^{\kappa_{2,T}}\left(Z_t\right)^T \psi_j - \mathbf{p}^{\kappa_{2,T}}\left(\hat{U}_t\right)^T \rho\right\}^2,
$$

and $\hat{U}_t = Z_t - \hat{g}\left(W_t, X_t\right)$ is calculated from Step 1.

In Step 3, denoting $\mathcal{Z}_t = [1, (Z_t - z\ )/h]^T$, $\mathcal{W}_t = \mathcal{Z}_t \otimes \mathcal{X}_t$, $K_h(v) = K(v/h)$, and a diagonal matrix $H = \text{diag}(1,h) \otimes I_{2(d_x+1)}$, we estimate $\beta(z)$ and $\dot{\beta}(z)$ by the local linear regression approach

$$\begin{pmatrix} \tilde{\beta}(z) \\ \tilde{\dot{\beta}}(z) \end{pmatrix} = H^{-1} \left[\sum_{t=1}^{T} K_h(Z_t - z) \mathcal{W}_t \mathcal{W}_t^T \right]^{-1} \sum_{t=1}^{T} K_h(Z_t - z) \mathcal{W}_t Y_{bx,t}, \tag{2.1}$$

where

$$Y_{bx,t} \equiv Y_t - \hat{\gamma}\left(\hat{U}_t\right) = \mathcal{X}_t^T \beta(Z_t) + \gamma(U_t) - \hat{\gamma}\left(\hat{U}_t\right) + \eta_t, \tag{2.2}$$

$$\eta_t \equiv \epsilon_t - \gamma(U_t) = Y_t - \beta_0(Z_t) - \sum_{j=1}^{d_x} \beta_j(Z_t) X_{j,t} - \gamma(U_t). \tag{2.3}$$

We note that if one were also interested in a third-step estimate of $\gamma(U_t)$, we can add to Step 2 computation of $Y_{\gamma,t} = Y_t - \mathcal{X}_t^T \hat{\beta}(Z_t)$ which will allow for application of local linear least squares to the model $Y_{\gamma,t} = \gamma(\hat{U}_t) + v_{2,t}$ in Step 3 to obtain $\widetilde{\gamma}(U_t)$. We assume throughout, however, that interest is centered on $\beta(Z_t)$ and $\dot{\beta}(Z_t)$.

It is worth acknowledging that while we prefer a local linear least squares approach for the kernel estimators, local constant or local polynomial (of degree greater than one) regression may also be applicable. For the series estimators, we prefer smoothing splines with B-splines given their faster convergence rates and low multicollinearity, however power series may be adopted as well. In practice, we advocate using cross validation techniques for both the series approximation terms and for the kernel regressions bandwidths to better benefit from the semiparametric structure and because it is well-known that data-driven bandwidths are more desirable in an empirical setting.

We also acknowledge our choice of functional form for the first and third step estimation. In the first step, we assume an additive nonparametric functional form. While this is helpful for our theoretical results, this assumption may be replaced with any semiparametric or parametric specification; we chose an additive nonparametric structure believing that it is empirically efficient. For the final step, on the other hand, the efficiency of the estimator may be further increased if one chooses to take partial residuals per each additive smooth coefficient component (that is, calculate $d_x + 2$ partial residuals including the one for U). Below, we prove that it is possible to estimate the smooth coefficient model free from the errors of the first two steps of estimation, i.e. our proposed estimator is asymptotically equivalent to the estimator calculated from a standard smooth coefficient model. Our estimator is oracle efficient in the sense that the smooth coefficient part of the model can be estimated with the asymptotic accuracy achieved if the smooth nonparametric function of $\hat{U}$ is known.

3 Asymptotic Properties

We now turn to the asymptotic behavior of our estimators. We provide two theorems and focus on estimation of $\beta(Z_t)$ and its gradients $\dot{\beta}(Z_t)$. The first two theorems are for the second step estimator $\hat{\beta}(Z_t)$ and the third step estimators of $\tilde{\beta}(Z_t)$ and $\widetilde{\dot{\beta}}(Z_t)$, respectively. In the first theorem we show that the asymptotic bias expression for $\hat{\beta}(Z_t)$ contains the terms from first step estimation. In the second theorem, we show that the terms from first two steps vanish and the estimates are asymptotically equivalent to those from a standard smooth coefficient model. The second theorem also shows that $\tilde{\beta}(Z_t)$ and $\widetilde{\dot{\beta}}(Z_t)$ are asymptotically normal. We note that it is also straightforward to establish uniform consistency.

3.1 Assumptions

Let $\mathcal{S}_x$, $\mathcal{S}_z$ and $\mathcal{S}_w$ be the support of variables X, Z, and W. We assume that all three sets are compact. Also, we denote $\|\mathbf{p}^\kappa\|_r = \max_{0\leq s\leq r} \|\partial^s \mathbf{p}^\kappa(v)/\partial v^s\|$ for a given positive integer r, and $\|\cdot\|$ refers to the Euclidean norm.

Assumption 1. (i) Let $\left\{\left(Y_t, X_t^T, Z_t, W_t^T, U_t, \varepsilon_t\right)\right\}$ be a strictly stationary α-mixing process with mixing coefficients $\{\alpha_t\}$ satisfying $\alpha_t = O\left(t^{-\varrho}\right)$ for some $\varrho > 2(\delta+2)/\delta$ and $\delta > 0$.

(ii) Z_t has a Lebesgue density function $f_z(z) > 0$ which is twice continuously differentiable and is bounded away from zero over its support, $\mathcal{S}_z$.

(iii) $f(z|x) \leq M < \infty$ and $\sup_{l\geq 1} f(z_0, z_l|x_0, x_l; l) \leq M < \infty$, where $f(z|x)$ and $f(z_0, z_l|x_0, x_l; l)$ are the conditional density of Z given $X = x$ and of (Z_0, Z_l) given $(X_0, X_l) = (x_0, x_l)$, respectively.

(iv) For some $\delta^* > \delta > 0$, $E\left(\|X_t\|^{2(2+\delta^*)}\right) < M < \infty$, $E\left(|Y_t|^{2+\delta^*}|Z_t = z, X_t = x\right) < M < \infty$ for all $x \in \mathcal{S}_x$ in the neighborhood of z, and $E\left(Y_0^2 + Y_t^2|Z_0 = z_0, X_0 = x_0, Z_t = z, X_t = x_t\right)$ $< M < \infty$ for all $(z_0, x_0, x_t) \in \mathcal{S}_z \times \mathcal{S}_x \times \mathcal{S}_x$ in the neighborhood of z.

(v) $\beta(z)$, $E(X_t|Z_t = z)$, and $E\left(\mathcal{X}_t\mathcal{X}_t^T|z\right)$ are all twice continuously differentiable in z.

(vi) $E\left(\mathcal{X}_t\mathcal{X}_t^T|z\right)$ is a $(d_x+1)\times(d_x+1)$ non-singular matrix.

Assumption 2. There exists a large T_0 and $\underline{c}_1$, $\bar{c}_1$, $\underline{c}_2$ and $\bar{c}_2$ such that for any $T \geq T_0$:

(i) $0 < \underline{c}_1 \leq \lambda_{\min}(\Sigma_{PP}) \leq \lambda_{\max}(\Sigma_{PP}) \leq \bar{c}_1 < \infty$, $\lambda_{\max}(\Sigma_{PP,U}) \leq \bar{c}_1 < \infty$, $\lambda_{\max}(\Sigma_{PP,\eta})$ $\leq \bar{c}_1 < \infty$ and $\lambda_{\max}\left(E\left[P_t^{\kappa_1,T}\left(P_t^{\kappa_1,T}\right)^T|Z_t = z\right]\right) \leq \bar{c}_1 < \infty$, where we denote $P_t^{\kappa_1,T} \equiv P^{\kappa_1,T}(W_t, X_t)$, $\Sigma_{PP} = E\left[P_t^{\kappa_1,T}\left(P_t^{\kappa_1,T}\right)^T\right]$, $\Sigma_{PP,U} = E\left[U_t^2 P_t^{\kappa_1,T}\left(P_t^{\kappa_1,T}\right)^T\right]$, and $\Sigma_{PP,\eta} = E\left[\eta_t^2 P_t^{\kappa_1,T}\left(P_t^{\kappa_1,T}\right)^T\right]$.

(ii) $0 < \underline{c}_2 \leq \lambda_{\min}(\Sigma_{QQ}) \leq \lambda_{\max}(\Sigma_{QQ}) \leq \bar{c}_2 < \infty$, and $\lambda_{\max}(\Sigma_{QQ,\eta}) \leq \bar{c}_2 < \infty$, where we denote $Q_t^{\kappa_2,T} \equiv Q^{\kappa_2,T}(Z_t, X_t, U_t)$, $\Sigma_{QQ} = E\left[Q_t^{\kappa_2,T}\left(Q_t^{\kappa_2,T}\right)^T\right]$ and $\Sigma_{QQ,\eta} =$

$E\left[\eta_t^2 Q_t^{\kappa_{2,T}}\left(Q_t^{\kappa_{2,T}}\right)^T\right]$.

Assumption 3. (i) $E(\varepsilon_t) = E[\gamma(U_t)] = 0$.

(ii) The unknown functions $\beta(\cdot)$, $g_w(\cdot)$, $g_x(\cdot)$, and $\gamma(\cdot)$ are all ξ-smooth functions with $\xi \geq 2$, where a ξ-smooth function is $[\xi]$-times continuously differentiable over its support and its $[\xi]$th derivative satisfies a Hölder condition with exponent $\xi - [\xi]$, and $[\xi]$ = the integer part of ξ.

(iii) There exist α_j, $\vartheta_{j'}$, ρ and $\psi_{j'}$ such that

$$\sup_{w_j \in \mathcal{S}_{w,j}} \left| g_{w,j}(w_j) - \mathbf{p}^{\kappa_{1,T}}(w_j)^T \alpha_j \right| \leq M\kappa_{1,T}^{-\xi}, \sup_{x_{j'} \in \mathcal{S}_{x,j'}} \left| g_{x,j'}(x_{j'}) - \mathbf{p}^{\kappa_{1,T}}(x_{j'})^T \vartheta_{j'} \right| \leq M\kappa_{1,T}^{-\xi}$$
$$\sup_{u \in \mathcal{S}_u} \left| \gamma(u) - \mathbf{p}^{\kappa_{2,T}}(u)^T \rho \right| \leq M\kappa_{2,T}^{-\xi}, \sup_{z \in \mathcal{S}_z} \left| \beta_{j'}(z) - \mathbf{p}^{\kappa_{2,T}}(z)^T \psi_{j'} \right| \leq M\kappa_{2,T}^{-\xi}$$

as $\kappa_{1,T} \to \infty$ and $\kappa_{2,T} \to \infty$, for $j = 1, \ldots, d_w$ and $j' = 1, \ldots, d_x$.

(iv) The set of basis functions, $\{p_j(\cdot), j = 1, 2, \ldots\}$, are twice continuously differentiable everywhere on its support for all j. Also, for $\sum_{l=1}^{\kappa_T}\left[E\left(\varpi_{l,t}^{2(2+\delta)}\right)\right]^{1/(2+\delta)} = O(\varsigma_{\kappa_T,\delta})$ for $\omega_{l,t} \equiv p_l(Z_t)$, $p_l(U_t)$, $p_l(W_{j,t})$, and $p_l(X_{j',t})$, and $\sum_{l=1}^{\kappa_{2,T}}\left\{E\left[(\eta_t \dot{p}_l(U_t))^{2(2+\delta)}\right]\right\}^{1/(2+\delta)} = O\left(\dot{\varsigma}_{\kappa_{2,T},\delta}\right)$ for all $j = 1, \ldots, d_w$ and $j' = 1, \ldots, d_x$.

Assumption 4. (i) $\sigma_\eta^2(x, z) = E(\eta_t^2 | Z_t = z, X_t = x)$ is positive and uniformly bounded over its support, $E\left(\eta_t^{2(2+\delta)}\right) < M < \infty$, and $E\left(\|X_t \eta_t\|^{2(2+\delta)}\right) < M < \infty$.

(ii) $E(U_t | W_t, X_t) = 0$, $E\left(U_t^{2(2+\delta)}\right) < M < \infty$, and $\sigma_u^2(w, x) = E(U_t^2 | W_t = w, X_t = x)$ $< M < \infty$ holds uniformly over its support $\mathcal{S}_w \times \mathcal{S}_x$.

(iii) There exists a large T_0 and $\bar{c}_3 > 0$ such that for any $T \geq T_0$, $\lambda_{\max}\left\{E\left(\varpi_t \mathbf{p}^{\kappa_{1,T}}(X_{j,t}) \mathbf{p}^{\kappa_{1,T}}(X_{j,t})^T\right)\right\} \leq \bar{c}_3 < \infty$, where $\varpi_t = |\eta_t|$, $X_{1,t}^2$, …,$X_{d_x,t}^2$.

(v) (U_t, Z_t) and $(U_t, e_{j,t}, Z_t)$ have joint Lebesgue probability density function, $f(u, z)$ and $f(u, e_j, z)$, respectively, where $e_{j,t} = X_{j,t}$ or $W_{j,t}$ for $j = 1, \ldots, d_x$ (or d_w). $\gamma(u)$, $f(u, z)$, $\partial f(u, z)/\partial z$, $\partial f(u, e_j, z)/\partial u$, and $\partial^2 f(u, e_j, z)/\partial u \partial z$ are all square-integrable functions of $u \in \mathcal{S}_u$.

Assumption 5. The kernel function $K(\cdot)$ is a symmetric probability density function over a compact support $[-1, 1]$, where we define $\omega_{i,j} = \int K^i(v) v^j dv$.

Assumption 6. Let $\pi_{j,T} = \kappa_{j,T}^{-\xi} + \sqrt{\kappa_{j,T}/T}\left(1 + \sqrt{\varsigma_{\kappa_{j,T},\delta}/\kappa_{j,T}}\right)$ for $j = 1, 2$.

(i) As $T \to \infty$, $\kappa_{j,T} \to \infty$, $\varsigma_{\kappa_{j,T},\delta}/\sqrt{T} \to 0$, $\varsigma_{\kappa_{1,T},\delta}\dot{\varsigma}_{\kappa_{2,T},\delta}/T \to 0$, $\kappa_{2,T}^{-\xi}\|\mathbf{p}^{\kappa_{2,T}}\|_1 \to 0$, $\|\mathbf{p}^{\kappa_{j,T}}\|_0^2 \kappa_{j,T}/T \to 0$, and $\pi_{1,T}\left(\|\mathbf{p}^{\kappa_{2,T}}\|_0 + \|\mathbf{p}^{\kappa_{2,T}}\|_1 + \|\mathbf{p}^{\kappa_{2,T}}\|_2\right) \to 0$.

(ii) As $T \to \infty$, $h \to 0$, $Th \to \infty$, and $Th^5 \to c \neq 0$.

(iii)$\sqrt{Th}\left(\|\mathbf{p}^{\kappa_{2,T}}\|_1 \kappa_{1,T}^{-\xi} + \pi_{2,T}\right) \to 0$.

Strict stationarity is assumed to simplify the assumptions of moment conditions appearing in the paper. Assumptions 1(i)-(iv), 5, and 6 (ii) are equivalent to conditions A.1

and A.2 in Cai et al. (2000, p. 952). They impose the sampling conditions and ensure that the mixing coefficient follow a certain decay rate. Assumption 1(ii) assumes that Z_t is continuously valued. An extension would be to allow for discrete endogenous variables, but it is not pursued in this paper. Assumption 2 ensures that we have nonsingular covariance matrices in both steps; we do not assume uniform bounds here, which makes the assumption less restrictive. Assumption 3 ensures that the unknown functions are ξ-smooth (such as Chen 2007, Ozabaci et al. 2014). Assumption 3 (iii)-(iv) quantifies the approximation error of the basis functions, given the smoothness assumption. Assumption 3 (iv) contains additional requirements due to weak dependence of time series data, where the term $\varsigma_{\kappa_{j,T},\delta} \geq M\kappa_{j,T}$ and $\dot{\varsigma}_{\kappa_{2,T},\delta}$vanish for independent data. If B-splines series are used to construct the basis functions, it is well known that $\|\mathbf{p}^{\kappa}\|_r = \kappa^{2^{-1}+r}$; e.g., Newey (1997, p. 160). Assumptions 4 and 5 are needed for the asymptotic results of Theorems 1 and 2, where Assumption 4(v) is required by the use of Parseval's identity. Assumption 6, on the other hand, defines the number of basis functions and bandwidths, and ensures that they follow our theoretical results. It is also worth mentioning that for our theoretical work, we assume that all exogenous regressors are continuously valued. However, this assumption can be relaxed to include additional discrete regressors, as well. This extension is straightforward, hence we leave it for now to keep our theoretical presentation clean.

3.2 Limiting results

Theorem 1 *Under Assumptions 1-4 and 6(i), we have*

$$\begin{aligned}
&\left[\left(\hat{\psi}-\psi\right)^T, (\hat{\rho}-\rho)^T\right]^T \\
= \;& \Sigma_{QQ}^{-1}\left[T^{-1}\sum_{t=1}^{T} Q_t^{\kappa_{2,T}}\eta_t + T^{-1}\sum_{t=1}^{T} Q_t^{\kappa_{2,T}}\left\{\mathcal{X}_t^T\left[\beta\left(Z_t\right)-\beta^*\left(Z_t\right)\right]+\gamma\left(U_t\right)-\gamma^*\left(U_t\right)\right\}\right] \\
& + T^{-1}\sum_{t=1}^{T} Q_t^{\kappa_{2,T}}\dot{\gamma}\left(U_t\right)\left(U_t-\hat{U}_t\right)+o_p\left(\pi_{1,T}\right)
\end{aligned}$$

and

$$\left[\left(\hat{\psi}-\psi\right)^T, (\hat{\rho}-\rho)^T\right]^T = O_p\left(\pi_{2,T}+\pi_{1,T}\right). \tag{3.1}$$

Theorem 2 *Under Assumptions 1-6, we have*

$$\begin{bmatrix} \sqrt{nh}I_{d_x+1} & 0_{d_x+1} \\ 0_{d_x+1} & \sqrt{nh^3}I_{d_x+1} \end{bmatrix}\begin{pmatrix} \tilde{\beta}(z)-\beta(z)-\omega_{1,2}h^2\ddot{\beta}\left(z\right)/2 \\ \tilde{\dot{\beta}}(z)-\dot{\beta}(z) \end{pmatrix} \xrightarrow{d} N\left(0_{2(d_x+1)}, \Omega\left(z\right)\right),$$

where

$$\Omega(z) = \frac{1}{f_z(z)} \begin{bmatrix} \omega_{2,0} & 0 \\ 0 & \frac{\omega_{2,2}}{\omega_{1,2}^2} \end{bmatrix} \otimes \left\{ \left[E\left(\mathcal{X}_t \mathcal{X}_t^T | z \right) \right]^{-1} E\left[\mathcal{X}_t \mathcal{X}_t^T \sigma_\eta^2 (X_t, Z_t) | z \right] \left[E\left(\mathcal{X}_t \mathcal{X}_t^T | z \right) \right]^{-1} \right\}.$$

In contrast to Theorem 1, Theorem 2 indicates that the third step kernel estimator of $\left[\beta(z)^T, \dot{\beta}(z)^T\right]^T$ is oracle efficient in the sense that the limiting distribution is the same as the limiting distribution when $\{\gamma(U_t) : 1 \leq t \leq T\}$ is known. That is, sampling errors present in the estimators of our first two steps do not asymptotically affect our third stage estimator.

4 Monte Carlo Analysis

So far, we have proposed an estimator for a smooth coefficient model with endogenous variables in the nonparametric function, and have proved that our estimator is consistent, asymptotically normal, and in the third step achieves oracle efficiency. In this section, we explore the finite sample performance of our estimator via Monte Carlo simulations. In particular, we consider a data generating process under assumptions of both independence and weak-dependence as our theoretical assumptions include the weak-dependence case. Our focus in these experiments is on the root mean squared error ($RMSE$) and variance of the conditional mean estimator $\widetilde{m}(X_t, Z_t)$ and the smooth coefficient estimators $\widetilde{\beta}(Z_t)$.

We first consider the following five-dimensional data generating process:

<u>DGP-1:</u>

$$\begin{aligned} Y_t &= \sin(Z_t) + \sin(Z_t) X_{1,t} + \sin(Z_t) X_{2,t} + \sin(Z_t) X_{3,t} + \sin(Z_t) X_{4,t} + \varepsilon_{1,t} + \varepsilon_{2,t} \\ Z_t &= \sin(X_{1,t}) + \sin(W_{1,t}) + \sin(W_{2,t}) + \sin(W_{3,t}) + \sin(W_{4,t}) + U_t \end{aligned} \quad (4.1)$$

where we draw both ε_1 and U independently from a standard normal distribution, $N(0,1)$, and draw each of the regressors in (X, Z, W) independently from each other from a uniform distribution over the unit interval, $U[0,1]$. ε_2, on the other hand, corresponds to the bias component, where $E[\varepsilon_2|U] \neq 0$; we model it as $E[\varepsilon_2|U] = \sin(U)$. Hence, this design corresponds to the standard cross-sectional setting.

We have assumed a nonparametric additive structure for Z to reflect our belief that the nonparametric additive structure may be desirable from an empirical perspective because of its flexibility as well as ability to avoid dimensionality issues common in fully nonparametric specifications. Similar exercises could be done under alternative structures that may include either nonparametric or parametric assumptions. Further, we have elected to use the $\sin(\cdot)$ function in our DGP because of its differentiability and nonlinearity. As with our assumed additive structure, different functional forms can easily be tested as well. Notice that in our DGP, we allow for only one included and four excluded regressors; our estimator will work, however, as long as the dimension of excluded regressors is greater than zero since the DGP assumes only a single endogenous regressor. Further, while our DGP is designed to be five-dimensional, once we add the control function to the structural equation the dimension will increase to six because of the additional smooth function. We note that we chose a high dimensional setting in order to provide results for a more general case that may correspond more directly to an empirical setting; the relatively low computational demand of our estimator allows us to perform these computations without high time costs. Of course, low dimensional simulations may also be considered.

To explore the performance of our estimator under weak-dependence, our second and third designs modify (4.1) by adding lags of some of the right-hand side variables. The two additional DGP's that we consider are:

DGP-2:

$$\begin{aligned} Y_t &= \sin(Z_t) + \sin(Z_t) X_{1,t} + \sin(Z_t) X_{2,t} + \sin(Z_t) X_{2,t-1} + \sin(Z_t) X_{2,t-2} + \varepsilon_{1,t} + \varepsilon_{2,t} \\ Z_t &= \sin(X_{1,t}) + \sin(W_{1,t}) + \sin(W_{2,t}) + \sin(W_{3,t}) + \sin(W_{4,t}) + U_t \end{aligned} \tag{4.2}$$

DGP-3:

$$\begin{aligned} Y_t &= \sin(Z_t) + \sin(Z_t) X_{1,t} + \sin(Z_t) X_{1,t-1} + \sin(Z_t) X_{2,t} + \sin(Z_t) X_{2,t-1} + \varepsilon_{1,t} + \varepsilon_{2,t} \\ Z_t &= \sin(X_{1,t}) + 0.1 W_{1,t} + 0.2 W_{1,t-1} + \sin(W_{2,t}) + \sin(W_{3,t}) + U_t \end{aligned} \tag{4.3}$$

in which all variables and errors are generated according to the structure in DGP-1. However, to model weak dependence, in DGP-2, we include lags only in the Y_t equation, and consider $X_{2,t-1}$ and $X_{2,t-2}$ as the one and two period lags of $X_{2,t}$. We extend the design in DGP-3 to include both $X_{1,t-1}$ and $X_{2,t-1}$ as one period lags of $X_{1,t}$ and $X_{2,t}$ in the Y_t equation, as well as $W_{1,t-1}$ as a one period lag of $W_{1,t}$ in the Z_t equation. Also, in DGP-3, we allow $W_{1,t}$ and $W_{1,t-1}$ to enter linearly instead of nonlinearly via the $\sin(\cdot)$ function as in DGP-1 and DGP-2. We make this modification to provide simulation results that relate more directly to alternative empirical structures and to illustrate that the empirical design of the first stage regression may include some parametric structure. Simulations reverting back to the $\sin(\cdot)$ function for these regressors also illustrate consistency of our estimator and are available upon request.

We use cubic B-splines for the first two steps in our estimation procedure, with a fixed number of knots given by $\lfloor 2T^{1/5} \rfloor$ where $\lfloor \cdot \rfloor$ is the floor function and T is the sample size. The third stage regression uses a local linear least squares kernel estimator to estimate the smooth coefficients and their first order partial derivatives. In both our simulations and empirical application we deploy a Gaussian kernel function, $K(v) = \exp(-0.5v^2)/\sqrt{2\pi}$ with a Silverman rule-of-thumb bandwidth, $h = 1.06\sigma_Z T^{-1/5}$ in which σ_Z is the sample standard deviation of Z. It is well known that in empirical settings a data-driven approach to selecting the knots or kernel bandwidths is desirable. In practice we advocate the same, however in our simulations we prefer to use the fixed number of knots and rule-of-thumb bandwidth to avoid unnecessarily long computation time. For each design, we consider simulations of size $T = \{200, 400, 1600\}$ over 999 Monte Carlo replications.

We report the results from our simulations in Table 1; in particular, for the slope coefficients we report the $RMSE$ and variance for the intercept and first slope coefficient. In all cases, the performance of our estimator is as anticipated. Within a particular DGP setup, as the sample size increases both the $RMSE$ and variance of our estimators of both the conditional mean and coefficient functions decrease. Further, the table shows that the downward movement in $RMSE$ as the sample size increases is faster for the smooth coefficients compared to the conditional mean. It is interesting to note that both the $RMSE$ and variance is higher for the intercept coefficient relative to the slope coefficient for any particular DGP structure and sample size. That has important implications for practical applications because it indicates that, all else equal, the endogeneity bias will manifest more strongly in the intercept coefficient compared to the slopes. Our simulation exercises also indicate that the consistency of our estimator holds across all structures of data independence as well as weak dependence, but that the $RMSE$ and variance are generally highest for DGP-3 compared to both DGP-2 and DGP-1.

5 Application to Household Gasoline Demand

5.1 Preliminary Discussion

This section develops an empirical application of our estimator to household gasoline demand. Transition to an empirical setting with observational data usually implies certain complexities not typically discussed in theoretical research or included in construction of computer generated samples. For example, economic and social data is often discrete, which requires special treatment. Fortunately, recent tools (e.g., Ma, Racine & Yang 2013, Ma, Racine & Yang 2014) make this relatively straightforward by introducing discrete-kernel smoothing for the discrete regressors. In this application, we couple our estimator with recently developed discrete-data smoothing methods for spline regression (Ma et al. 2013, Ma et al. 2014), using cross-validation to select the bandwidths in the discrete kernel functions as well as assisting with specification of the spline setup. We, however, use a rule of thumb bandwidth given by $2.12\hat{\sigma}_x T^{-1/5}$ for the final smooth coefficient estimates in order to provide a comparison between our estimator and a standard non-instrumental variables smooth coefficient estimator without introducing variability across estimators due to differences in bandwidths. Note that we use a slightly larger bandwidth than the traditional Silverman 1.06 rule-of-thumb in order to reduce the variability in our estimates of the coefficient functions; in other words, our reported results are qualitatively unchanged if we deploy the 1.06 rule-of-thumb, except that in that case there is increased variability in our estimated coefficient functions.

5.2 Motivation for Semiparametric Model of Gasoline Demand

Gasoline consumption is a critical component of global carbon emissions, and as a result, national and international policies have taken keen interest in understanding consumer gasoline demand (e.g., Davis & Killian 2011). Of particular interest is the price elasticity of demand for gasoline, as this measures the responsiveness of consumers to economic policies that may influence the price of gasoline, such as a tax, in hopes of encouraging households to drive less. In particular, several notable studies have deployed semiparametric estimators to capture heterogeneity and nonlinearities in consumer demand (Schmalensee & Stoker 1999, Yatchew & No 2001, Wadud et al. 2010, Liu 2014), with the latter two studies assuming a smooth coefficient structure. These studies have produced convincing evidence that there is heterogeneity in consumer gasoline demand, and that semiparametric methods - in particular, the smooth coefficient setup - is well suited for modeling gasoline demand elasticities and providing results that support a rich analysis of heterogeneous response to policy levers (e.g., a tax).

One source of heterogeneity in the price elasticity of demand comes from differences in incomes across households; Liu (2014), for instance, assumes a parametric interaction between the price of gasoline and income to model this heterogeneity. High- versus low-income consumers likely have many differences that affect gasoline consumption. For instance, high-income consumers typically have the luxury to take more vacations, or can substitute to alternative forms/frequencies of transportation in response to rising gasoline prices. Low-income consumers typically do not have the luxury of taking frequent vacations, or do not have the flexibility of alternative forms/frequencies of transportation given rising prices of gasoline. Alternatively, gasoline constitutes a larger share of household income for low-income households, which might lead us to suspect that low-income consumers will be more responsive to gasoline price changes relative to high-income consumers.

Yet, the price elasticity of demand is only one way in which consumer income may induce heterogeneity into consumer gasoline demand. The relationship between many gasoline demand drivers and gasoline consumption may differ for households with different incomes as high- and low-income consumers may be fundamentally different in many ways. For instance, the way in which moving from a one-car household to a two-car household impacts gasoline consumption may differ across income groups; or, similarly, an increase in family size (persons per household) may have a different effect on gasoline consumption for consumers with different incomes. In other words, the importance of consumer income as a driver of consumer demand - potentially through multiple channels - provides adequate motivation for us to consider specifying our semiparametric model as a smooth coefficient model in which the coefficients vary with respect to income. Income may also be correlated with education, and may play an important role in the impact of education on gasoline consumption. Of course, we might also think of income having a direct effect on gasoline consumption in our model through the intercept function.

Yet, consumer income is likely to be correlated with unobservable characteristics that also affect gasoline demand. Consumer preferences for positive environmental amenities also influence gasoline consumption. Research in environmental economics has emphasized consumer preferences for improved environmental quality in general (e.g., Kotchen 2005, Kotchen 2006, Kotchen & Moore 2007, Kotchen 2009, Jacobsen, Kotchen & Vandenbergh 2012); Clark, Kotchen & Moore (2003) show that higher income consumers are more likely to demand green electricity. This suggests that income is likely to be correlated with unobservable preferences that also influence gasoline consumption. Other empirical evidence of the relationship between income and environmental quality comes from the much-discussed environmental Kuznets hypothesis that environmental quality eventually improves as income rises, which might imply that consumers with higher income are more likely to value environmental amenities. These preferences, then, in the context of our gasoline demand model implies that consumer income may be endogenous. So, while other important contributions have emphasized the importance of accounting for heterogeneity in the household gasoline demand model, none have considered the use of instrumental variables to assuage concerns that household income is endogenous. To manage this endogeneity, we deploy our proposed estimation strategy so that we can model heterogeneity in the price elasticity of gasoline demand as a function of income while still obtaining consistent estimates. To make our estimator operational we use the age of the head of the household and the race of the household head as instrumental variables since both are correlated with income, but are not likely to affect gasoline consumption directly.

5.3 Model Specification and Estimation

We propose the empirical design

$$\ln G_t = \beta_1(\ln I_t) \ln P_t + X_{1t}\beta(\ln I_t) + \epsilon_t \tag{5.1}$$

in which G_t is the quantity of gasoline consumed by the tth household, I_t is household income, P_t is the price of gasoline, X_{1t} is a vector of control variables that includes a column of ones to allow for an intercept, $(\beta_1(\cdot), \beta(\cdot)^T)$ are coefficient functions we wish to estimate, and ϵ_t is the error such that $E[\epsilon_t|I_t] \neq 0$ but $E[\epsilon_t|P_t, X_{1t}] = 0$. To complete the specification, we model income in the first stage as

$$\ln I_t = g(W_t, X_{2t}) + U_t \tag{5.2}$$

in which W_t is a vector of instrumental variables that are excluded from Equation (5.1) and X_{2t} are other control variables that may appear in X_{1t}. We estimate (5.2) using the tools from Ma et al. (2013, 2014) to allow for categorical regressors in X_{2t}. Note that, following our theoretical setup in (1.1), one might assume $g(\cdot)$ has an additive structure; alternatively - as we elect to do here - one can follow Ma et al. (2013, 2014) and use least squares cross-validation to select the appropriate spline orders, number of interior knots, and bandwidths for the discrete variables. This approach is appealing in practical applications since it alleviates the burden of a prior specifying a particular spline regression structure. Further, if the true structure is additive, the cross-validation procedure will converge to this setup.

Following generalized spline estimation of $\hat{U}_t$, we return to our proposed estimation procedure of using sieve methods and then local linear least squares kernel regression to recover consistent and oracle efficient estimates of $(\beta_1(\cdot), \beta(\cdot)^T)$.

5.4 Data

The data we use for this application comes from Wadud et al. (2010). The data comes from the United States Consumer Expenditure Surveys from the years 1997-2002, and contains 30,000 quarterly measured, household level observations. In addition to household demographics (age, number of wage earners, race, education of household head), the survey contains total family quarterly expenditure, expenditure on gasoline, the number of vehicles in the household, average household vehicle fuel efficiency, and an indicator for rurality. Table 2 reports descriptive statistics for these data.

In our application, X_{1t} includes the family size, number of alternative vehicles available to the household, fuel economy of the household's car, the education level of the household head, and the indicator for rurality. In addition to price, we believe that households with more members will have higher gasoline consumption, as will households with a greater number of vehicles. Households with a higher average fuel efficiency rating on their vehicles will likely consume less gasoline, as will households in which the head has a higher education. We assume that rural households also consume more gasoline as average distances traveled is likely higher because these households are likely farther from local amenities. We see in Table 2 that the average family in our data has 2 or 3 members (mean is 2.5), and the average household has about 2 vehicles that average around 21 MPG. Education of the household head is categorical, ranging from 1 to 4 which indicates the highest level of education attained being less than high school, high school, some college, or college degree. On average, the head of the household has either high school or some college education (mean is 2.811), and nearly 11 percent of these households are in a rural area.

W_t and X_{2t} include, respectively, the age of the household head and the race of the household head defined as an indicator that is equal to unity if the household head is non-white, and the number of wage earners in the household and the education level of the household head. It is common in empirical research to find that age is nonlinearly related to income, with income peaking somewhere around forty-five years of age. Further, it is well-known that being non-white is correlated with lower income. Clearly, household income is likely higher in households with a greater number of wage earners, and is likely higher in households in which the head has a higher level of education; however, these two variables are not likely to be uncorrelated with gasoline consumption so they are included in both equations. In Table 2 we see that the average age of the head of the household is just under 52 years of age, with a minimum age of 17 and maximum age of 94 years. Further, just over 14 percent of households have a head that is non-white.

The dependent variable in (5.1) is quarterly consumption of gasoline in gallons, and the key regressors include total quarterly expenditure as a proxy for income, and the price of gasoline measured as the average state-level price in cents per gallon for each household over the quarter in which the household was surveyed in order to measure the price the household faced at time of interview. In our data, the average quarterly consumption of gasoline per household is 222 gallons, with a minimum and maximum of 2.5 and over 5,000 gallons. Average quarterly income (expenditure) is \$8,380, and the average price of gasoline faced by the household is 152 cents per gallon (\$1.52 per gallon). It is important to emphasize that in our model, prices are not endogenous to gasoline consumption because gasoline consumption is individual household gasoline consumption and prices are quarterly-state-level averages. It is not likely that individual gasoline consumption from a household can lead to any changes in the local price of gasoline. The continuous variables in the model are in logs to allow $\beta_1(\cdot)$ to represent the price elasticity of demand. In our preferred model, we do not log age, however we point out that empirical results are qualitatively unchanged if the first stage regression is modeled using the log of age. For more specific details regarding data construction and sources, please see Wadud et al. (2010).

5.5 Sieve Estimation Results

We use the tools proposed by Ma et al. (2013, 2014) to estimate (5.2). The number of wage earners, race, and education level are treated as categorical regressors, and age is treated as continuous. The cross-validation procedure indicates that the optimal number of spline segments for the age variable is 3, and each of our categorical variables are statistically relevant predictors of income. That is, the cross-validated bandwidths on these variables are 0.078, 0.001, and 0.012, respectively for wage earners, race, and education. The R^2 for this regression is 0.370.

We plot a partial regression relationship between log income and age, number of earners, and education in Figure 1. While number of earners and education appear in our regression as discrete variables, we plot these relationships as a line because it is easier to visually see the relationship between these variables and income. As can be seen from the figure, the relationship between income and age is nonlinear, and peaks around 40 years of age. Our estimated peak in the income-age relationship is consistent with our expectation for this relationship, and as can be seen from the figure, this relationship is statistically significant. The effect of a larger number of wage earners or a higher education on income is monotonically increasing, and both are statistically significant as well.

While not shown graphically, the mean relationship between being white and log income is 8.80 with upper and lower confidence bounds being (8.81, 8.79). The mean effect of being non-white on log income is 8.55, with upper and lower bounds (8.56, 8.54). Hence, the relationship between log income and race is consistent with our expectations, and statistically significant both from each other and from zero. In other words, Figure 1 and our estimates of the income-race relationship constitutes evidence that our first stage regression is well specified, and our instrumental variables are highly correlated with income and are relevant.

5.6 Kernel Estimation Results

5.6.1 Function Estimates

Following our proposed estimation strategy, we obtain $\hat{U}_t$ from the first step calculation, and then plug this estimate into our second and third step regressions. We report our final step estimates of $(\beta_1(\cdot), \beta(\cdot)^T)$ in Figure 2. In Figure 2, we compare our final step semiparametric estimates to a standard non-IV smooth coefficient estimator; we show the instrumental variables estimates with the (black) circles, and the non-IV estimates with the (red) plus symbols. Interpretation of the smooth coefficient functions is that, at a given point (log income), the function is the marginal effect of the parametric (i.e., price and X_{1t}) variables on gasoline consumption.

Consider, first, the estimated price elasticity of gasoline demand, $\hat{\beta}_1(\cdot)$, shown in the upper right corner of the figure. As anticipated, the price elasticity of demand is negative - indicating a downward sloping demand curve for gasoline - and nonlinear. This nonlinearity is evidence that income induces heterogeneity in the gasoline demand relationship. At the lowest levels of income, the demand elasticity slightly dips downwards (becomes more elastic), but then gradually rises with income (becomes less elastic). Hence, higher income consumers are less responsive to increases in the price of gasoline, likely because gasoline is a smaller fraction of their overall budget and they have less incentive to alter their driving habits on account of rising gasoline prices. However, when comparing our IV and non-IV estimates, we see that there is not a significant difference in the price elasticity of demand. While not shown to keep the graph visually clear, this is confirmed by placing confidence bounds over the function estimates - it is not statistically possible to distinguish the semiparametric and semiparametric-IV estimates of this elasticity. If the price of gasoline is exogenous to household gasoline consumption, then it is possible that any endogeneity in income does not lead to noticeable bias in our estimated elasticity given the nonlinear interaction between income and price through $\beta_1(\ln I_t) \ln P_t$.

Our estimate of the intercept function might be interpreted as the direct effect of income on gasoline consumption. For this estimated coefficient function, it is clear that there is a substantial divergence between our IV and non-IV estimates, with our IV-estimated function lying substantially below our non-IV estimate. Further, the shape of the coefficient function differs slightly between the two models: in the IV model, the function gradually increases with income, while in the non-IV model, the function is generally flat (or at least slightly rises and then falls). Figure 3 plots this relationship with bootstrap confidence bounds around the semiparametric-IV estimate to show that the standard semiparametric estimate is significantly different. It is clear that the endogeneity of income leads to a larger bias in the intercept than in some of the slope coefficients, a fact that is shown clearly in our Monte Carlo exercises as well. Hence, while there is not a noticeable bias in our estimate of the price elasticity of gasoline, we do see a strong bias in the intercept coefficient function, and that our new IV estimator controls for this bias. Further, interpreting the intercept coefficient as a direct effect of income on gasoline consumption, the endogeneity of income leads to an over-estimate of this effect.

Turning to our other estimated coefficient functions, it is clear that income is a source of heterogeneity for the effect of most of the X_{1t} variables on gasoline consumption. It is clear that the effect of the number of vehicles, fuel efficiency, and rurality on gasoline consumption are all decreasing with income. Yet, for these three variables, the estimated coefficient function does not differ significantly across the IV and non-IV estimators. However, the effect of family size and education does differ across these two estimators, and is confirmed by our confidence bounds shown in Figure 3. It is clear that the family size coefficient is also biased upwards in the semiparametric model, since the function estimate from the semiparametric-IV model lies significantly below. The non-IV semiparametric education coefficient is approximately zero, but the semiparametric-IV estimate is clearly downward sloping and is becoming significantly negative as education increases. This indicates that, without controlling for the endogeneity of income, one might conclude that the relationship between gasoline consumption, education, and income is not statistically significant; yet, it is clear that after controlling for income endogeneity, we see that households with higher income and education have significantly lower consumption of gasoline. We believe this result is more in line with intuition than the semiparametric non-IV estimates.

5.6.2 Point Estimates

To provide more specifics, we report a summary of point estimates and standard errors from our semiparametric IV estimator in Table 3. Specifically, we report the mean estimate across all observations, as well as the estimate at the 25th, 50th, and 75th percentiles of the distribution of estimates. It is clear from Table 3 that all of the point estimates we report are highly statistically significant.

In particular, we find that the price elasticity of gasoline demand ranges from -0.698 to -0.494, with a mean of -0.587. We further note that each of these point estimates lie outside of two standard deviations, which indicates that these estimates are significantly heterogeneous. Therefore, as we hypothesized, the price elasticity of gasoline demand is significantly negative, and varies significantly across households with different levels of income.

Our other point estimates confirm that income leads to significant differences in the magnitude of our point estimates. We find, for instance, that on average, gasoline consumption increases by about 7 percent for households with an additional member, or by nearly 12 percent on average for households with an additional vehicle. It is interesting to note that households with higher miles per gallon ratings on their vehicles consume significantly less gasoline, which may reflect both a mechanical difference in gasoline requirements as well as another channel through which environmentally conscious consumers reduce gasoline consumption. Finally, as expected we see that households with higher education (of the head) consume nearly 10 percent less gasoline, and rural households on average consume about 10 percent more gasoline. The latter effect is likely because of a farther distance from these households to amenities.

6 Conclusion

We develop an estimator for a semiparametric smooth coefficient structure with endogenous variables in the nonparametric part of the model. Estimators have been proposed for managing endogeneity in the parametric part of a smooth coefficient model, but so far none have considered endogeneity in the nonparametric functions. We propose to manage the endogeneity problem in the nonparametric functions via a control function approach which yields a hybrid semiparametric structure in which there are additive nonparametric functions as well as the smooth coefficients, and is therefore more complicated that a traditional smooth coefficient setup.

We propose a three-step estimator that uses both series and kernel estimators, and establish that our estimators are consistent and asymptotically normal. We further show that our third stage estimator is oracle efficient, so that the asymptotic performance of our final step estimator is identical to that of an estimator with a known control function. We develop the theory for our estimators under both independent and weakly dependent data structures, and use Monte Carlo simulations to demonstrate that our estimator performs well in finite samples for cases involving both independent and weakly dependent data.

To develop the practicality of our estimator and illustrate its applicability to economic data in policy-relevant settings, we develop a semiparametric model of household gasoline demand in which the smooth coefficients vary with respect to household income. We find that income is an important source of heterogeneity in our estimated coefficient functions, as well as evidence that our estimator is able to correct for the endogeneity of income. In particular, consistent with our Monte Carlo simulations, we find that the endogeneity of income has a more pronounced effect on our intercept coefficient. We also find that this endogeneity impacts our estimates of the effect of education on gasoline consumption; in a non-IV model, the effect of education as a function of income on gasoline consumption is not significantly different from zero, but after controlling for the endogeneity of income we find that there is a significantly negative effect of education on gasoline consumption.

Acknowledgements

We thank Zia Wadud for kindly providing the data used for the empirical application, and Raymond Florax, Daniel Henderson, Weiwei Liu, and seminar participants at Purdue University and the University of Connecticut for helpful comments. This research was supported in part by computational resources provided by Information Technology at Purdue - Rosen Center for Advanced Computing, Purdue University, West Lafayette, Indiana.

References

Ai, C. & Chen, X. (2003), 'Efficient estimation of models with conditional moment restrictions containing unknown functions', *Econometrica* **71**, 1795–1843.

Cai, Z., Das, M., Xiong, H. & Wu, X. (2006), 'Functional coefficient instrumental variables models', *Journal of Econometrics* **133**, 207–241.

Cai, Z., Fan, J. & Yao, Q. (2000), 'Functional coefficient regression models for nonlinear time series', *Journal of the American Statistical Association* **95**, 941–956.

Cai, Z. & Li, Q. (2008), 'Nonparametric estimation of varying coefficient dynamic panel data models', *Econometric Theory* **24**, 1321–1342.

Cai, Z. & Xiong, H. (2012), 'Partially varying coefficient instrumental variables models', *Statistica Neerlandica* **66**, 85–110.

Chen, X. (2007), Large sample sieve estimation of semi-nonparametric models, *in* J. J. Heckman & E. Leamer, eds, 'Handbook of Econometrics', New York: Elsevier Science, chapter 76, pp. 5549–5632.

Chen, X. & Pouzo, D. (2012), 'Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals', *Econometrica* **80**, 277–321.

Christopeit, N. & Hoderlein, S. G. N. (2006), 'Local partitioned regression', *Econometrica* **74**, 787–817.

Clark, C. F., Kotchen, M. J. & Moore, M. R. (2003), 'Internal and external influences on pro-environmental behavior: Participation in a green electricity program', *Journal of Environmental Psychology* **23**, 237–246.

Darolles, S., Fan, Y., Florens, J. P. & Renault, E. (2011), 'Nonparametric instrumental regression', *Econometrica* **79**, 1541–1565.

Das, M. (2005), 'Instrumental variables estimators for nonparametric models with discrete endogenous variables', *Journal of Econometrics* **124**, 335–361.

Davis, L. W. & Killian, L. (2011), 'Estimating the effect of a gasoline tax on carbon emissions', *Journal of Applied Econometrics* **26**, 1187–1214.

Delgado, M. S. (2013), 'A smooth coefficient model of carbon emissions', *Empirical Economics* **45**, 1049–1071.

Delgado, M. S., McCloud, N. & Kumbhakar, S. C. (2014), 'A generalized empirical model of corruption, foreign direct investment, and growth', *Journal of Macroeconomics* **42**, 298–316.

Delgado, M. S. & Parmeter, C. F. (2014), 'A simple estimator for partial linear regression with endogenous nonparametric variables', *Economics Letters* **124**, 100–103.

Delgado, M. S., Parmeter, C. F., Hartarska, V. & Nadolynak, D. (2015), 'Should all microfinance institutions mobilize microsavings? evidence from economies of scope', *Empirical Economics* **48**, 193–225.

Durlauf, S. N., Kourtellos, A. & Minkin, A. (2001), 'The local Solow growth model', *European Economic Review* **45**, 928–940.

Hall, P. & Horowitz, J. L. (2005), 'Nonparametric methods for inference in the presence of instrumental variables', *Annals of Statistics* **33**, 2904–2929.

Hartarska, V., Parmeter, C. F. & Nadolynak, D. (2011), 'Economies of scope of lending and mobilizing deposits in microfinance institutions: A semiparametric analysis', *American Journal of Agricultural Economics* **93**, 389–398.

Horowitz, J. L. (2011), 'Applied nonparametric instrumental variables estimation', *Econometrica* **79**, 347–394.

Horowitz, J. L. & Mammen, E. (2004), 'Nonparametric estimation of an additive model with a link function', *Annals of Statistics* **32**, 2412–2443.

Jacobsen, G. D., Kotchen, M. J. & Vandenbergh, M. P. (2012), 'The behavioral response to voluntary provision of an environmental public good: Evidence from residential electricity demand', *European Economic Review* **56**, 946–960.

Kotchen, M. J. (2005), 'Impure public goods and the comparative statics of environmentally friendly consumption', *Journal of Environmental Economics and Management* **49**, 281–300.

Kotchen, M. J. (2006), 'Green markets and private provision of public goods', *Journal of Political Economy* **114**, 816–834.

Kotchen, M. J. (2009), 'Voluntary provision of public goods for bads: A theory of environmental offsets', *Economic Journal* **119**, 883–899.

Kotchen, M. J. & Moore, M. R. (2007), 'Private provision of environmental public goods: Household participation in green-electricity programs', *Journal of Environmental Economics and Management* **53**, 1–16.

Li, Q., Huang, C. J., Li, D. & Fu, T. (2002), 'Semiparametric smooth coefficient models', *Journal of Business and Economic Statistics* **20**, 412–422.

Liu, W. (2014), 'Modeling gasoline demand in the United States: A flexible semiparametric approach', *Energy Economics* **45**, 244–253.

Ma, S., Racine, J. S. & Yang, L. (2013), 'Additive regresison splines with irrelevant regressors', *Statistica Sinica* **23**, 515–541.

Ma, S., Racine, J. S. & Yang, L. (2014), 'Spline regression in the presence of categorical predictors', *Journal of Applied Econometrics* **forthcoming**.

Mamuneas, T. P., Savvides, A. & Stengos, T. (2006), 'Economic development and the return to human capital: A smooth coefficient semiparametric approach', *Journal of Applied Econometrics* **21**, 111–132.

Newey, W. K. (1997), 'Convergence rates and asymptotic normality for series estimators', *Journal of Econometrics* **79**, 147–168.

Newey, W. K. & Powell, J. L. (2003), 'Instrumental variables estimation of nonparametric models', *Econometrica* **71**, 1565–1578.

Newey, W. K., Powell, J. L. & Vella, F. (1999), 'Nonparametric estimation of triangular simultaneous equations models', *Econometrica* **67**, 565–603.

Ozabaci, D., Henderson, D. J. & Su, L. (2014), 'Additive nonparametric regression in the presence of endogenous regressors', *Journal of Business and Economic Statistics* **32**, 555–575.

Pinkse, J. (2000), 'Nonparametric two-step regression estimation when regressors and error are dependent', *Canadian Journal of Statistics* **28**, 289–300.

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

Roehrig, C. S. (1988), 'Conditions for identification in nonparametric and parametric models', *Econometrica* **56**, 433–447.

Schmalensee, R. & Stoker, T. M. (1999), 'Household gasoline demand in the United States', *Econometrica* **67**, 645–662.

Seber, G. A. F. (2008), A matrix handbook for statistics, *in* 'Wiley Series in Probability and Statistics', John Wiley & Sons.

Su, L., Murtazashvili, I. & Ullah, A. (2013), 'Local linear GMM estimation of functional coefficient IV models with application to the estimation of rate of return to schooling', *Journal of Business and Economic Statistics* **31**, 184–207.

Su, L. & Ullah, A. (2008), 'Local polynomial estimation of nonparametric simultaneous equations models', *Journal of Econometrics* **144**, 193–218.

Wadud, Z., Noland, R. B. & Graham, D. J. (2010), 'A semiparametric model of household gasoline demand', *Energy Economics* **32**, 93–101.

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

Yatchew, A. & No, J. (2001), 'Household gasoline demand in Canada', *Econometrica* **69**, 1679–1709.

Zhang, R., Sun, K., Delgado, M. S. & Kumbhakar, S. C. (2012), 'Productivity in China's high technology industry: Regional heterogeneity and R&D', *Technological Forecasting and Social Change* **79**, 127–141.

7 Appendix: Brief Mathematical Proofs

Proof of Theorem 1 : Denoting $Q_t^{\kappa_{2,T}} \equiv Q^{\kappa_{2,T}}(Z_t, X_t, U_t)$, $\hat{Q}_t^{\kappa_{2,T}} \equiv Q^{\kappa_{2,T}}\left(Z_t, X_t, \hat{U}_t\right)$, $\Sigma_{T,QQ} = T^{-1}\sum_{t=1}^T Q_t^{\kappa_{2,T}}\left(Q_t^{\kappa_{2,T}}\right)^T$, and $\Sigma_{T,\hat{Q}\hat{Q}} = T^{-1}\sum_{t=1}^T \hat{Q}_t^{\kappa_{2,T}}\left(\hat{Q}_t^{\kappa_{2,T}}\right)^T$, from Step 2, we have

$$
\begin{aligned}
&\left[\left(\hat{\psi}-\psi\right)^T, (\hat{\rho}-\rho)^T\right]^T \\
= \;& \Sigma_{T,\hat{Q}\hat{Q}}^{+} T^{-1}\sum_{t=1}^T \hat{Q}_t^{\kappa_{2,T}}\eta_t + \Sigma_{T,\hat{Q}\hat{Q}}^{+} T^{-1}\sum_{t=1}^T \hat{Q}_t^{\kappa_{2,T}}\left\{\mathcal{X}_t^T\left[\beta(Z_t)-\beta^*(Z_t)\right]+\gamma(U_t)-\gamma^*\left(\hat{U}_t\right)\right\} \\
= \;& \Sigma_{T,\hat{Q}\hat{Q}}^{+}\Bigg[T^{-1}\sum_{t=1}^T Q_t^{\kappa_{2,T}}\eta_t + T^{-1}\sum_{t=1}^T Q_t^{\kappa_{2,T}}\left\{\mathcal{X}_t^T\left[\beta(Z_t)-\beta^*(Z_t)\right]+\gamma(U_t)-\gamma^*(U_t)\right\} \\
&+T^{-1}\sum_{t=1}^T Q_t^{\kappa_{2,T}}\left[\gamma^*(U_t)-\gamma^*\left(\hat{U}_t\right)\right] + T^{-1}\sum_{t=1}^T\left(\hat{Q}_t^{\kappa_{2,T}}-Q_t^{\kappa_{2,T}}\right)\eta_t \\
&+T^{-1}\sum_{t=1}^T\left(\hat{Q}_t^{\kappa_{2,T}}-Q_t^{\kappa_{2,T}}\right)\left\{\mathcal{X}_t^T\left[\beta(Z_t)-\beta^*(Z_t)\right]+\gamma(U_t)-\gamma^*(U_t)\right\} \\
&+T^{-1}\sum_{t=1}^T\left(\hat{Q}_t^{\kappa_{2,T}}-Q_t^{\kappa_{2,T}}\right)\left[\gamma^*(U_t)-\gamma^*\left(\hat{U}_t\right)\right] \\
\equiv \;& \Sigma_{T,\hat{Q}\hat{Q}}^{+}\left(A_{T1}+A_{T2}+A_{T3}+A_{T4}+A_{T5}+A_{T6}\right),
\end{aligned}
$$

where the definitions of A_{Tj} (j=1,2,3,4) will be clear from the following context.

First, we consider $A_{T1} \equiv T^{-1}\sum_{t=1}^T Q_t^{\kappa_{2,T}}\eta_t$. As $E(\eta_t|Z_t, X_t, W_t, U_t) = E(\epsilon_t|Z_t, X_t, W_t, U_t)$ $-E(\epsilon_t|U_t) = 0$ by (1.2), we have $E(A_{T1}) = 0$. As all the variables are strictly stationary we have

$$
E\left(\|A_{T1}\|^2\right) = T^{-1}E\left[\eta_t^2\left(Q_t^{\kappa_{2,T}}\right)^T Q_t^{\kappa_{2,T}}\right] + \frac{2}{T}\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)E\left[\eta_1\eta_t\left(Q_1^{\kappa_{2,T}}\right)^T Q_t^{\kappa_{2,T}}\right],
$$

where by Assumption 2(ii), we have

$$
E\left[\eta_t^2\left(Q_t^{\kappa_{2,T}}\right)^T Q_t^{\kappa_{2,T}}\right] = tr\left(\Sigma_{QQ,\eta}\right) \leq \kappa_{2,T}(d_x+1)\bar{c}_2 = O(\kappa_{2,T}). \tag{A.1}
$$

Under Assumption 1(i), using Davoydov's inequality and Hölder's inequality, we obtain

$$
\begin{aligned}
E(\eta_1\eta_t\varpi_1\varpi_t) &\leq M\alpha_{t-1}^{\delta/(2+\delta)}\left[E\left(|\tilde{\eta}_1|^{2+\delta}|\varpi_1|^{2+\delta}\right)\right]^{1/(2+\delta)}\left[E\left(|\tilde{\eta}_t|^{2+\delta}|\varpi_t|^{2+\delta}\right)\right]^{1/(2+\delta)} \\
&\leq M\alpha_{t-1}^{\delta/(2+\delta)}\left[E\left(\tilde{\eta}_1^{2(2+\delta)}\right)E\left(\varpi_1^{2(2+\delta)}\right)\right]^{1/(2+\delta)}
\end{aligned}
$$

where $\varpi_t = p_l(Z_t)$ or $p_l(U_t)$ and $\tilde{\eta}_t = \eta_t$ or $\eta_t X_{j,t}$ for $l = 1, \ldots, \kappa_{2,T}$ and $j = 1, \ldots, d_x$.

Hence, under Assumptions 1(i), 3(iv), and 4(i), applying , we have

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)E\left[\eta_1\eta_t\left(Q_1^{\kappa_{2,T}}\right)^T Q_t^{\kappa_{2,T}}\right]\\
=&\ \frac{1}{T}\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)E\left\{\eta_1\eta_t\left[\sum_{l=1}^{\kappa_{2,T}}\left[p_l\left(Z_1\right)p_l\left(Z_t\right)+p_l\left(U_1\right)p_l\left(U_t\right)\right]\right.\right.\\
&\left.\left.+\sum_{l=1}^{\kappa_{2,T}}\sum_{j=1}^{d_x}p_l\left(Z_1\right)p_l\left(Z_t\right)X_{j,1}X_{j,t}\right]\right\}\\
=&\ O\left(\varsigma_{\kappa_{2,T},\delta}\right)\frac{1}{T}\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)\alpha_{t-1}^{\delta/(2+\delta)} = O\left(\varsigma_{\kappa_{2,T},\delta}/T\right),
\end{aligned}$$

where $\sum_{t=1}^{T-1}\left(1-t/T\right)\alpha_{t-1}^{\delta/(2+\delta)} < M < \infty$ under Assumption 1(i). Hence, we obtain $A_{T1}=O_p\left(T^{-1/2}\left(\kappa_{2,T}^{1/2}+\varsigma_{\kappa_{2,T},\delta}^{1/2}\right)\right)$.

Second, we consider $A_{T2}\equiv T^{-1}\sum_{t=1}^{T}Q_t^{\kappa_{2,T}}\left\{\mathcal{X}_t^T\left[\beta\left(Z_t\right)-\beta^*\left(Z_t\right)\right]+\gamma\left(U_t\right)-\gamma^*\left(U_t\right)\right\}$. Under Assumptions 2(ii) and 3(iii), we have $\|A_{T2}\|=O_p\left(\kappa_{2,T}^{-\xi}\right)$, closely following the proof of Lemma B.2 (ii) in Ozabaci et al. (2014).

Third, we consider $A_{T3}\equiv T^{-1}\sum_{t=1}^{T}Q_t^{\kappa_{2,T}}\left[\gamma^*\left(U_t\right)-\gamma^*\left(\hat{U}_t\right)\right]=T^{-1}\sum_{t=1}^{T}Q_t^{\kappa_{2,T}}\dot{\gamma}^*\left(\bar{U}_t\right)\left(U_t-\hat{U}_t\right)$,

where $\bar{U}_t$ lies between U_t and $\hat{U}_t$, and $\dot{\gamma}^*\left(\bar{U}_t\right)=\dot{\gamma}^*\left(U_t\right)+\left[\dot{\gamma}\left(U_t\right)-\dot{\gamma}^*\left(U_t\right)\right]+\left[\dot{\gamma}\left(\bar{U}_t\right)-\dot{\gamma}\left(U_t\right)\right]$ $+\left[\dot{\gamma}^*\left(\bar{U}_t\right)-\dot{\gamma}\left(\bar{U}_t\right)\right]$. As $\left|\dot{\gamma}\left(u'\right)-\dot{\gamma}\left(u\right)\right|\le M\left|u'-u\right|$ and $\left|\dot{\gamma}^*\left(u\right)-\dot{\gamma}\left(u\right)\right|\le M\kappa_{2,T}^{1-\xi}$ hold uniformly for all u, $u'\in\mathcal{S}_u$ under Assumption 3 (ii)-(iii), we have

$$\begin{aligned}
&\left\|T^{-1}\sum_{t=1}^{T}Q_t^{\kappa_{2,T}}\left[\dot{\gamma}\left(\bar{U}_t\right)-\dot{\gamma}\left(U_t\right)+\dot{\gamma}\left(U_t\right)-\dot{\gamma}^*\left(U_t\right)+\dot{\gamma}^*\left(\bar{U}_t\right)-\dot{\gamma}\left(\bar{U}_t\right)\right]\left(U_t-\hat{U}_t\right)\right\|\\
\le&\ MT^{-1}\sum_{t=1}^{T}\left\|Q_t^{\kappa_{2,T}}\right\|\left(U_t-\hat{U}_t\right)^2+M\kappa_{2,T}^{1-\xi}T^{-1}\sum_{t=1}^{T}\left\|Q_t^{\kappa_{2,T}}\right\|\left|U_t-\hat{U}_t\right|\\
=&\ O_p\left(\pi_{1,T}^2\left\|\mathbf{p}^{\kappa_{2,T}}\right\|_0\right)+M\kappa_{2,T}^{1-\xi}\sqrt{T^{-1}\sum_{t=1}^{T}\left\|Q_t^{\kappa_{2,T}}\right\|^2}\sqrt{T^{-1}\sum_{t=1}^{T}\left(U_t-\hat{U}_t\right)^2}\\
=&\ O_p\left(\pi_{1,T}^2\left\|\mathbf{p}^{\kappa_{2,T}}\right\|_0+\kappa_{2,T}^{-\xi+3/2}\pi_{1,T}\right)
\end{aligned}$$

by (B.9) and $T^{-1}\sum_{t=1}^{T}\left\|Q_t^{\kappa_{2,T}}\right\|^2=E\left(\left\|Q_1^{\kappa_{2,T}}\right\|^2\right)+o_p\left(1\right)=O\left(\kappa_{2,T}\right)$, following the proof of (B.6), so that

$$\begin{aligned}
A_{T3} &= T^{-1}\sum_{t=1}^{T}Q_t^{\kappa_{2,T}}\dot{\gamma}^*\left(U_t\right)\left(U_t-\hat{U}_t\right)+O_p\left(\pi_{1,T}^2\left\|\mathbf{p}^{\kappa_{2,T}}\right\|_0+\kappa_{2,T}^{-\xi+3/2}\pi_{1,T}\right)\\
&= O_p\left(\pi_{1,T}\right)+O_p\left(\pi_{1,T}^2\left\|\mathbf{p}^{\kappa_{2,T}}\right\|_0+\kappa_{2,T}^{-\xi+3/2}\pi_{1,T}\right), \qquad \text{(A.2)}
\end{aligned}$$

where $T^{-1}\sum_{t=1}^{T} Q_t^{\kappa_{2,T}} \dot{\gamma}^*(U_t)\left(U_t - \hat{U}_t\right) = O_p(\pi_{1,T})$ can be shown by closely following the proof of Lemma B.5 (iii) in Ozabaci et al. (2014).

For $A_{T6} \equiv T^{-1}\sum_{t=1}^{T}\left(\hat{Q}_t^{\kappa_{2,T}} - Q_t^{\kappa_{2,T}}\right)\left[\gamma^*(U_t) - \gamma^*\left(\hat{U}_t\right)\right]$. We have

$$\begin{aligned} A_{T6} &= T^{-1}\sum_{t=1}^{T}\left[\mathbf{p}^{\kappa_{2,T}}\left(\hat{U}_t\right) - \mathbf{p}^{\kappa_{2,T}}(U_t)\right]\left[\gamma^*(U_t) - \gamma^*\left(\hat{U}_t\right)\right] \\ &= T^{-1}\sum_{t=1}^{T}\dot{\mathbf{p}}^{\kappa_{2,T}}\left(\bar{U}_t\right)\gamma^*\left(\vec{U}_t\right)\left(\hat{U}_t - U_t\right)^2 \\ &= O_p\left(\pi_{1,T}^2\left\|\mathbf{p}^{\kappa_{2,T}}\right\|_1\right) \end{aligned}$$

by (B.9) where $\bar{U}_t$ and $\vec{U}_t$ both lie between U_t and $\hat{U}_t$.

Fourth, we consider $A_{T4} \equiv T^{-1}\sum_{t=1}^{T}\left(\hat{Q}_t^{\kappa_{2,T}} - Q_t^{\kappa_{2,T}}\right)\eta_t$. Applying a Taylor expansion to $\mathbf{p}^{\kappa_{2,T}}\left(\hat{U}_t\right)$ and by the fact that

$$\begin{aligned} & \hat{U}_t - U_t \\ =\ & Z_t - \hat{g}(W_t, X_t) - U_t = g(W_t, X_t) - \hat{g}(W_t, X_t) \\ =\ & g(W_t, X_t) - g^*(W_t, X_t) - P^{\kappa_{1,T}}(W_t, X_t)^T\left[\hat{\mu}_z - \mu_z, (\hat{\alpha} - \alpha)^T, \left(\hat{\vartheta} - \vartheta\right)^T\right] \end{aligned} \tag{A.3}$$

we decompose A_{T4} as

$$\begin{aligned} A_{T4} &\equiv T^{-1}\sum_{t=1}^{T}\left(\hat{Q}_t^{\kappa_{2,T}} - Q_t^{\kappa_{2,T}}\right)\eta_t = T^{-1}\sum_{t=1}^{T}\left[\mathbf{p}^{\kappa_{2,T}}\left(\hat{U}_t\right) - \mathbf{p}^{\kappa_{2,T}}(U_t)\right]\eta_t \\ &= T^{-1}\sum_{t=1}^{T}\eta_t\dot{\mathbf{p}}^{\kappa_{2,T}}(U_t)\left(\hat{U}_t - U_t\right) + (2T)^{-1}\sum_{t=1}^{T}\eta_t\ddot{\mathbf{p}}^{\kappa_{2,T}}\left(\bar{U}_t\right)\left(\hat{U}_t - U_t\right)^2 \\ &= -T^{-1}\sum_{t=1}^{T}\eta_t\dot{\mathbf{p}}^{\kappa_{2,T}}(U_t)P^{\kappa_{1,T}}(W_t, X_t)^T\Sigma_{T,PP}^{+}\left[T^{-1}\sum_{t=1}^{T}P^{\kappa_{1,T}}(W_t, X_t)U_t\right] \\ &\quad -T^{-1}\sum_{t=1}^{T}\eta_t\dot{\mathbf{p}}^{\kappa_{2,T}}(U_t)P^{\kappa_{1,T}}(W_t, X_t)^T\Sigma_{T,PP}^{+}(\chi_{T2} + \chi_{T3}) \\ &\quad +T^{-1}\sum_{t=1}^{T}\eta_t\dot{\mathbf{p}}^{\kappa_{2,T}}(U_t)\left[g(W_t, X_t) - g^*(W_t, X_t)\right] \\ &\quad +(2T)^{-1}\sum_{t=1}^{T}\eta_t\ddot{\mathbf{p}}^{\kappa_{2,T}}\left(\bar{U}_t\right)\left(\hat{U}_t - U_t\right)^2 \\ &\equiv A_{T4,1} + A_{T4,2} + A_{T4,3} + 2A_{T4,4}, \end{aligned}$$

where χ_{T2} and χ_{T3} are defined in Lemma 1 below, and $\|A_{T4,4}\| \leq T^{-1}\sum_{t=1}^{T}|\eta_t|\left\|\ddot{\mathbf{p}}^{\kappa_{2,T}}\left(\bar{U}_t\right)\right\|\left(\hat{U}_t - U_t\right)^2$ $= O_p\left(\pi_{1,T}^2\left\|\mathbf{p}^{\kappa_{2,T}}\right\|_2\right)$ by (B.9).

Let $B_T \equiv T^{-1} \sum_{t=1}^{T} \eta_t \dot{\mathbf{p}}^{\kappa_{2,T}}(U_t) P^{\kappa_{1,T}}(W_t, X_t)^T$. We have $E(B_T) = 0$ and

$$
\begin{aligned}
& E\left(\|B_T\|^2\right) \\
= \;& T^{-1} E\left[\eta_t^2 P^{\kappa_{1,T}}(W_t, X_t)^T P^{\kappa_{1,T}}(W_t, X_t) \dot{\mathbf{p}}^{\kappa_{2,T}}(U_t)^T \dot{\mathbf{p}}^{\kappa_{2,T}}(U_t)\right] \\
& +2T^{-2} \sum_{t=1}^{T} \sum_{s \neq t} E\left[\eta_t \eta_s P^{\kappa_{1,T}}(W_t, X_t)^T P^{\kappa_{1,T}}(W_s, X_s) \dot{\mathbf{p}}^{\kappa_{2,T}}(U_s)^T \dot{\mathbf{p}}^{\kappa_{2,T}}(U_t)\right] \\
= \;& O\left(T^{-1} \|\mathbf{p}^{\kappa_{2,T}}\|_1^2\right) E\left[\eta_t^2 P^{\kappa_{1,T}}(W_t, X_t)^T P^{\kappa_{1,T}}(W_t, X_t)\right] \\
& +2T^{-2} \sum_{l=1}^{\kappa_{2,T}} \sum_{t=1}^{T} \sum_{s \neq t} E\left\{\left[1 + \sum_{l'=1}^{\kappa_{1,T}} \left(\sum_{j=1}^{d_w} p_{l'}(W_{j,t}) p_{l'}(W_{j,s}) + \sum_{j=1}^{d_x} p_{l'}(X_{j,t}) p_{l'}(X_{j,s})\right)\right]\right. \\
& \left.\eta_t \eta_s \dot{p}_l(U_t) \dot{p}_l(U_s)\right\} \\
= \;& O\left(T^{-1} \kappa_{1,T} \|\mathbf{p}^{\kappa_{2,T}}\|_1^2\right) + O\left(T^{-1} \varsigma_{\kappa_{1,T},\delta} \dot{\varsigma}_{\kappa_{2,T},\delta}\right)
\end{aligned}
$$

by Assumptions 1(i), 2 (i), and 3(iv), where we apply Davoydov's inequality and Hölder's inequality to the second term

$$
\begin{aligned}
E(\eta_1 \eta_t \varpi_1 \varpi_t) \;&\leq\; M \alpha_{t-1}^{\delta/(2+\delta)} \left[E\left(|\tilde{\eta}_1|^{2+\delta} |\varpi_1|^{2+\delta}\right)\right]^{1/(2+\delta)} \left[E\left(|\tilde{\eta}_t|^{2+\delta} |\varpi_t|^{2+\delta}\right)\right]^{1/(2+\delta)} \\
&\leq\; M \alpha_{t-1}^{\delta/(2+\delta)} \left[E\left(\tilde{\eta}_1^{2(2+\delta)}\right) E\left(\varpi_1^{2(2+\delta)}\right)\right]^{1/(2+\delta)}
\end{aligned}
$$

for $\varpi_t = 1$, $p_{l'}(W_{j,t})$ or $p_{l'}(X_{j,t})$, and $\tilde{\eta}_t = \eta_t \dot{p}_l(U_t)$, $l = 1, \ldots, \kappa_{2,T}$, $l' = 1, \ldots, \kappa_{1,T}$ and $j = 1, \ldots, d_x$ (or d_w). By Markov's inequality we obtain $B_T = O_p\left(T^{-1/2}\left(\kappa_{1,T}^{1/2} \|\mathbf{p}^{\kappa_{2,T}}\|_1 + \varsigma_{\kappa_{1,T},\delta}^{1/2} \dot{\varsigma}_{\kappa_{2,T},\delta}^{1/2}\right)\right)$. Therefore, combining with Lemma 1, we have

$$
\begin{aligned}
\|A_{T4,1}\| \;&=\; \|A_{T4,1}\|_\infty \leq \|B_T\|_\infty \left\|\Sigma_{T,PP}^{+}\right\|_\infty \left\|T^{-1} \sum_{t=1}^{T} P^{\kappa_{1,T}}(W_t, X_t) U_t\right\| \\
&=\; O_p\left(T^{-1}\left(\kappa_{1,T}^{1/2} + \varsigma_{\kappa_{1,T},\delta}^{1/2}\right)\left(\kappa_{1,T}^{1/2} \|\mathbf{p}^{\kappa_{2,T}}\|_1 + \varsigma_{\kappa_{1,T},\delta}^{1/2} \dot{\varsigma}_{\kappa_{2,T},\delta}^{1/2}\right)\right) \qquad \text{(A.4)}
\end{aligned}
$$

and $A_{T4,2} = O_p\left(T^{-1/2} \kappa_{1,T}^{-\xi}\left(\kappa_{1,T}^{1/2} \|\mathbf{p}^{\kappa_{2,T}}\|_2 + \varsigma_{\kappa_{1,T},\delta}^{1/2} \dot{\varsigma}_{\kappa_{2,T},\delta}^{1/2}\right)\right)$. Similarly, we can show $A_{T4,3} = O_p\left(T^{-1/2} \kappa_{1,T}^{-\xi} \dot{\varsigma}_{\kappa_{2,T},\delta}^{1/2}\right)$. Taking these results together gives

$$
A_{T4} = O_p\left(\pi_{1,T}^2 \|\mathbf{p}^{\kappa_{2,T}}\|_2\right) + O_p\left(T^{-1/2} \pi_{1,T}\left(\kappa_{1,T}^{1/2} \|\mathbf{p}^{\kappa_{2,T}}\|_1 + \varsigma_{\kappa_{1,T},\delta}^{1/2} \dot{\varsigma}_{\kappa_{2,T},\delta}^{1/2}\right)\right).
$$

Fifth, we consider

$$\begin{aligned}
A_{T5} &\equiv T^{-1}\sum_{t=1}^{T}\left(\hat{Q}_t^{\kappa_{2,T}}-Q_t^{\kappa_{2,T}}\right)\left\{\mathcal{X}_t^T\left[\beta\left(Z_t\right)-\beta^*\left(Z_t\right)\right]+\gamma\left(U_t\right)-\gamma^*\left(U_t\right)\right\} \\
&= T^{-1}\sum_{t=1}^{T}\left(\hat{Q}_t^{\kappa_{2,T}}-Q_t^{\kappa_{2,T}}\right)\left\{\left[\beta_0\left(Z_t\right)-\beta_0^*\left(Z_t\right)\right]+\gamma\left(U_t\right)-\gamma^*\left(U_t\right)\right\} \\
&\quad +T^{-1}\sum_{j=1}^{d_x}\sum_{t=1}^{T}\left(\hat{Q}_t^{\kappa_{2,T}}-Q_t^{\kappa_{2,T}}\right)X_{j,t}\left[\beta_j\left(Z_t\right)-\beta_j^*\left(Z_t\right)\right] \\
&\equiv A_{T5,0}+\sum_{j=1}^{d_x}A_{T5,j},
\end{aligned}$$

where by Assumption 3 (ii), the triangular inequality, and $\left(\sum_{i=1}^{T}a_i\right)^2\leq T\sum_{i=1}^{T}a_i^2$, we have

$$\begin{aligned}
\left\|A_{T5,0}\right\| &\leq M\kappa_{2,T}^{-\xi}T^{-1}\sum_{t=1}^{T}\left\|\hat{Q}_t^{\kappa_{2,T}}-Q_t^{\kappa_{2,T}}\right\| \\
&\leq M\kappa_{2,T}^{-\xi}\left(T^{-1}\sum_{t=1}^{T}\left\|\hat{Q}_t^{\kappa_{2,T}}-Q_t^{\kappa_{2,T}}\right\|^2\right)^{1/2}=O_p\left(\pi_{1,T}\kappa_{2,T}^{-\xi}\left\|\mathbf{p}^{\kappa_{2,T}}\right\|_1\right)
\end{aligned}$$

and for any $j=1,\ldots,d_x$

$$\left\|A_{T5,j}\right\|\leq M\kappa_{2,T}^{-\xi}\left(T^{-1}\sum_{t=1}^{T}\left\|\hat{Q}_t^{\kappa_{2,T}}-Q_t^{\kappa_{2,T}}\right\|^2X_{j,t}^2\right)^{1/2}=O_p\left(\pi_{1,T}\kappa_{2,T}^{-\xi}\left\|\mathbf{p}^{\kappa_{2,T}}\right\|_1\right)$$

by (B.9) and (B.10) in Lemma 2.

Next, following closely the proof of (B.1) in Lemma 1, we can show

$$\Sigma_{T,\hat{Q}\hat{Q}}^{+}=\Sigma_{QQ}^{-1}+O_p\left(\left(\varsigma_{\kappa_{2,T},\delta}\kappa_{2,T}^{-1/2}+\left\|\mathbf{p}^{\kappa_{2,T}}\right\|_0\right)\sqrt{\kappa_{2,T}/T}\right)=\Sigma_{QQ}^{-1}+o_p\left(1\right)$$

under Assumption 6(i). Therefore, we obtain $\left[\left(\hat{\psi}-\psi\right)^T,\left(\hat{\rho}-\rho\right)^T\right]^T=O_p\left(\pi_{2,T}+\pi_{1,T}\right)$ $+O_p\left(\pi_{1,T}^2\left(\left\|\mathbf{p}^{\kappa_{2,T}}\right\|_1+\left\|\mathbf{p}^{\kappa_{2,T}}\right\|_2\right)\right)+O_p\left(T^{-1/2}\pi_{1,T}\left(\kappa_{1,T}^{1/2}\left\|\mathbf{p}^{\kappa_{2,T}}\right\|_1+\varsigma_{\kappa_{1,T},\delta}^{1/2}\dot{\varsigma}_{\kappa_{2,T},\delta}^{1/2}\right)\right)$ $+O_p\left(\pi_{1,T}\kappa_{2,T}^{-\xi}\left\|\mathbf{p}^{\kappa_{2,T}}\right\|_1\right)$. This completes the proof of this theorem.

Proof of Theorem 2 : Applying a Taylor expansion to $\beta\left(Z_t\right)$ at an interior point z, we have

$$\beta\left(Z_t\right)=\beta\left(z\right)+\dot{\beta}\left(z\right)\left(Z_t-z\right)+\ddot{\beta}\left(\tilde{Z}_t\right)\left(Z_t-z\right)^2/2,$$

where $\tilde{Z}_t$ lies between Z_t and z. Then we have

$$\begin{pmatrix} \tilde{\beta}(z) \\ \tilde{\dot{\beta}}(z) \end{pmatrix} - \begin{pmatrix} \beta(z) \\ \dot{\beta}(z) \end{pmatrix} = H^{-1} A_{T1}^{-1} \left(A_{T2}/2 + A_{T3} + A_{T4} \right),$$

where

$$\begin{aligned} A_{T1} &= (Th)^{-1} \sum_{t=1}^{T} K_h (Z_t - z) \mathcal{W}_t \mathcal{W}_t^T \\ A_{T2} &= (Th)^{-1} \sum_{t=1}^{T} K_h (Z_t - z) \mathcal{W}_t \mathcal{X}_t^T \ddot{\beta} \left(\tilde{Z}_t \right) (Z_t - z)^2, \\ A_{T3} &= (Th)^{-1} \sum_{t=1}^{T} K_h (Z_t - z) \mathcal{W}_t \left[\gamma (U_t) - \hat{\gamma} \left(\hat{U}_t \right) \right] \\ A_{T4} &= (Th)^{-1} \sum_{t=1}^{T} K_h (Z_t - z) \mathcal{W}_t \eta_t. \end{aligned}$$

Letting $\phi(z) = E\left(\mathcal{X}_t \mathcal{X}_t^T | Z_t = z \right) \ddot{\beta}(z) f_z(z)$ and applying Theorems 1 and 2 in Cai et al. (2000), we have

$$\begin{aligned} A_{T1} &\xrightarrow{p} \begin{bmatrix} 1 & 0 \\ 0 & \omega_{1,2} \end{bmatrix} \otimes E\left(\mathcal{X}_t \mathcal{X}_t^T | z \right) f_z(z) \\ A_{T2} &= \begin{bmatrix} h^2 \omega_{1,2} \phi(z) \\ 0 \end{bmatrix} + O_p\left(h^4 \right) \\ \sqrt{Th} A_{T4} &\xrightarrow{d} N(0, \Sigma(z)) \end{aligned}$$

where

$$\Sigma(z) = f_z(z) \begin{bmatrix} \omega_{2,0} & 0 \\ 0 & \omega_{2,2} \end{bmatrix} \otimes E\left[\mathcal{X}_t \mathcal{X}_t^T \sigma_\eta^2 (X_t, Z_t) | z \right].$$

Next, we need to verify $\sqrt{Th} A_{T3} = o_p(1)$. Decomposing A_{T3} into three terms yields $A_{T3} \equiv A_{T3,1} + A_{T3,2} + A_{T3,3}$, where under Assumptions 1(vi), 3(iii)-(vi), 5, and 6, we have, by Lemma 3,

$$A_{T3,1} \equiv \frac{1}{Th} \sum_{t=1}^{T} K_h (Z_t - z) \mathcal{W}_t \left[\gamma (U_t) - \gamma^* (U_t) \right] = O_p\left(\kappa_{2,T}^{-\xi} \right),$$

$$
\begin{aligned}
A_{T3,2} &\equiv \frac{1}{Th}\sum_{t=1}^{T} K_h\left(Z_t - z\right)\mathcal{W}_t\left[\gamma^*\left(U_t\right) - \gamma^*\left(\hat{U}_t\right)\right] \\
&= \frac{1}{Th}\sum_{t=1}^{T} K_h\left(Z_t - z\right)\mathcal{W}_t\left(U_t - \hat{U}_t\right)\dot{\mathbf{p}}^{\kappa_{2,T}}\left(\bar{U}_t\right)^T\rho \\
&= O_p\left(\left\|\mathbf{p}^{\kappa_{2,T}}\right\|_1\kappa_{1,T}^{-\xi} + \pi_{1,T}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
& A_{T3,3} \\
\equiv\ & \frac{1}{Th}\sum_{t=1}^{T} K_h\left(Z_t - z\right)\mathcal{W}_t\left[\gamma^*\left(\hat{U}_t\right) - \hat{\gamma}\left(\hat{U}_t\right)\right] \\
=\ & \frac{1}{Th}\sum_{t=1}^{T} K_h\left(Z_t - z\right)\mathcal{W}_t\mathbf{p}^{\kappa_{2,T}}\left(\hat{U}_t\right)^T\left(\rho - \hat{\rho}\right) \\
=\ & \frac{1}{Th}\sum_{t=1}^{T} K_h\left(Z_t - z\right)\mathcal{W}_t\mathbf{p}^{\kappa_{2,T}}\left(U_t\right)^T\left(\rho - \hat{\rho}\right) \\
& + \frac{1}{Th}\sum_{t=1}^{T} K_h\left(Z_t - z\right)\mathcal{W}_t\left(\hat{U}_t - U_t\right)\dot{\mathbf{p}}^{\kappa_{2,T}}\left(U_t\right)^T\left(\rho - \hat{\rho}\right) \\
& + \frac{1}{2Th}\sum_{t=1}^{T} K_h\left(Z_t - z\right)\mathcal{W}_t\left(\hat{U}_t - U_t\right)^2\ddot{\mathbf{p}}^{\kappa_{2,T}}\left(\vec{U}_t\right)^T\left(\rho - \hat{\rho}\right) \\
=\ & O_p\left(1 + \left\|\mathbf{p}^{\kappa_{2,T}}\right\|_1\kappa_{1,T}^{-\xi} + \pi_{1,T} + \left\|\mathbf{p}^{\kappa_{2,T}}\right\|_2\pi_{1,T}^2\right)O_p\left(\pi_{1,T} + \pi_{2,T}\right),
\end{aligned}
$$

where $\vec{U}_t$ lies between U_t and $\hat{U}_t$, as $\rho - \hat{\rho} = O_p\left(\pi_{1,T} + \pi_{2,T}\right)$ by Theorem 1. Taking all the results above gives $A_{T3} = O_p\left(\left\|\mathbf{p}^{\kappa_{2,T}}\right\|_1\kappa_{1,T}^{-\xi} + \pi_{2,T}\right)$. This completes the proof of this theorem.

8 Lemmas

Lemma 1 *Under Assumptions 1–4, we have*

$$\left\|\Sigma_{T,PP}^{+}-\Sigma_{PP}^{-1}\right\|=O_p\left(\left(\varsigma_{\kappa_{1,T},\delta}\kappa_{1,T}^{-1/2}+\|\mathbf{p}^{\kappa_{1,T}}\|_0\right)\sqrt{\kappa_{1,T}/T}\right), \tag{B.1}$$

and

$$\begin{aligned}&\left[\hat{\mu}_z-\mu_z,(\hat{\alpha}-\alpha)^T,\left(\hat{\vartheta}-\vartheta\right)^T\right]^T\\ =\;&\Sigma_{PP}^{-1}\left(\chi_{T1}+\chi_{T2}+\chi_{T3}\right)+O_p\left(\pi_{1,T}\left(\varsigma_{\kappa_{1,T},\delta}\kappa_{1,T}^{-1/2}+\|\mathbf{p}^{\kappa_{1,T}}\|_0\right)\sqrt{\kappa_{1,T}/T}\right),\end{aligned} \tag{B.2}$$

where we denote

$$\begin{aligned}\chi_{T1}&\equiv T^{-1}\sum_{t=1}^{T}P^{\kappa_{1,T}}\left(W_t,X_t\right)U_t=O_p\left(T^{-1/2}\left(\kappa_{1,T}^{1/2}+\varsigma_{\kappa_{1,T},\delta}^{1/2}\right)\right)\\ \chi_{T2}&\equiv T^{-1}\sum_{j=1}^{d_w}\sum_{t=1}^{T}\mathbf{p}^{\kappa_{1,T}}\left(W_{j,t}\right)\left[g_{w,j}(W_{j,t})-g_{w,j}^{*}(W_{j,t})\right]=O_p\left(\kappa_{1,T}^{-\xi}\right)\\ \chi_{T3}&\equiv T^{-1}\sum_{j=1}^{d_x}\sum_{t=1}^{T}\mathbf{p}^{\kappa_{1,T}}\left(X_{j,t}\right)\left[g_{x,j}(X_{j,t})-g_{x,j}^{*}(X_{j,t})\right]=O_p\left(\kappa_{1,T}^{-\xi}\right),\end{aligned}$$

and $\pi_{1,T}=\kappa_{1,T}^{-\xi}+\sqrt{\kappa_{1,T}/T}\left(1+\sqrt{\varsigma_{\kappa_{1,T},\delta}/\kappa_{1,T}}\right)$.

Proof: First, we verify (B.1). Let $\|A\|_\infty=\max_{\|\omega\|=1,\omega\neq0}\|A\omega\|=\lambda_{\max}^{1/2}\left(A^TA\right)$, where $\omega\in R^n$, A is an $n\times n$ matrix, and $\|\cdot\|$ refers to the Euclidean norm.[2] As $\|\cdot\|_\infty$ is submultiplicative, we have

$$\begin{aligned}\left\|\Sigma_{T,PP}^{-1}-\Sigma_{PP}^{-1}\right\|_\infty&=\left\|\Sigma_{T,PP}^{-1}\left(\Sigma_{n,PP}-\Sigma_{PP}\right)\Sigma_{PP}^{-1}\right\|_\infty\\ &\le\left\|\Sigma_{T,PP}^{-1}\right\|_\infty\left\|\Sigma_{T,PP}-\Sigma_{PP}\right\|_\infty\left\|\Sigma_{PP}^{-1}\right\|_\infty\\ &=O_p(1)\left\|\Sigma_{T,PP}^{-1}\right\|_\infty\left\|\Sigma_{T,PP}-\Sigma_{PP}\right\|\end{aligned} \tag{B.3}$$

where $\left\|\Sigma_{PP}^{-1}\right\|_\infty=O_p(1)$ by Assumption 2 (i) and $\|\Sigma_{T,PP}-\Sigma_{PP}\|_\infty\le\|\Sigma_{T,PP}-\Sigma_{PP}\|$. And, we have

$$\left\|\Sigma_{T,PP}^{-1}\right\|_\infty=\lambda_{\min}^{-1}\left(\Sigma_{T,PP}\right)=\lambda_{\min}^{-1}\left(\Sigma_{PP}\right)+O\left(\|\Sigma_{T,PP}-\Sigma_{PP}\|\right), \tag{B.4}$$

[2] The matrix norm $\|A\|_\infty$ has following useful properties: (i) $\|A\|_\infty\le\|A\|$; (ii) $\|AB\|_\infty\le\|A\|_\infty\|B\|_\infty$ l (iii) $\|B\|_\infty\le\|A\|_\infty$ if B is a submatrix of A; see, Seber (2008, p. 72).

where we apply Weyl's theorem (Seber 2008, p. 117) to obtain

$$
\begin{aligned}
\lambda_{\min}\left(\Sigma_{T,PP}\right) &\leq \lambda_{\min}\left(\Sigma_{T,PP}\right) + \lambda_{\max}\left(\Sigma_{T,PP} - \Sigma_{PP}\right) \leq \lambda_{\min}\left(\Sigma_{T,PP}\right) + \left\|\Sigma_{T,PP} - \Sigma_{PP}\right\| \\
\lambda_{\min}\left(\Sigma_{T,PP}\right) &\geq \lambda_{\min}\left(\Sigma_{T,PP}\right) + \lambda_{\min}\left(\Sigma_{T,PP} - \Sigma_{PP}\right) \geq \lambda_{\min}\left(\Sigma_{T,PP}\right) - \left\|\Sigma_{T,PP} - \Sigma_{PP}\right\|.
\end{aligned}
$$

Hence, to obtain (B.1), we need to obtain the stochastic order of $\left\|\Sigma_{T,PP} - \Sigma_{PP}\right\|$.

Denoting the element of $P^{\kappa_{1,T}}\left(W_t, X_t\right)$ by $\phi_{l,t}$ for $l = 1, \ldots, v_T$ ($v_T = 1 + \left(d_x + d_w\right)\kappa_{1,T}$), we have

$$
\begin{aligned}
&E\left(\left\|\Sigma_{T,PP} - \Sigma_{PP}\right\|^2\right) = \sum_{l=1}^{v_T}\sum_{j=1}^{v_T} E\left[T^{-1}\sum_{t=1}^{T}\phi_{l,t}\phi_{j,t} - E\left(\phi_{l,t}\phi_{j,t}\right)\right]^2 \\
=\;& T^{-1}\sum_{l=1}^{v_T}\sum_{j=1}^{v_T} E\left[\phi_{l,t}\phi_{j,t} - E\left(\phi_{l,t}\phi_{j,t}\right)\right]^2 \\
&+T^{-2}\sum_{l=1}^{v_T}\sum_{j=1}^{v_T}\sum_{t=1}^{T}\sum_{s\neq t} E\left[\phi_{l,t}\phi_{j,t} - E\left(\phi_{l,t}\phi_{j,t}\right)\right]\left[\phi_{l,s}\phi_{j,s} - E\left(\phi_{l,s}\phi_{j,s}\right)\right] \\
\leq\;& T^{-1}E\left(\sum_{l=1}^{v_T}\phi_{l,t}^2\sum_{j=1}^{v_T}\phi_{j,t}^2\right) + T^{-2}\sum_{l=1}^{v_T}\sum_{j=1}^{v_T}\sum_{t=1}^{T}\sum_{s\neq t}\alpha_{|t-s|}^{\delta/(2+\delta)}\left[E\left(\left|\phi_{l,t}\phi_{j,t}\right|^{2+\delta}\right)E\left|\phi_{l,s}\phi_{j,s}\right|^{2+\delta}\right]^{1/(2+\delta)} \\
\equiv\;& a_1 + a_2,
\end{aligned}
$$

where we apply Davoydov's inequality to obtain the second term in the first inequality. As $\sum_{l=1}^{v_T}\phi_{l,t}^2 \leq \left(d_x + d_w + 1\right)\left\|\mathbf{p}^{\kappa_{1,T}}\right\|_0^2$ under Assumption 3 (iv), and $E\left(\sum_{j=1}^{v_T}\phi_{j,t}^2\right) = tr\left(\Sigma_{PP}\right)$ $= \sum_{j=1}^{v_T}\lambda_j\left(\Sigma_{PP}\right) = O\left(\kappa_{1,T}\right)$ under Assumption 2(i), we have $a_1 = O\left(\left\|\mathbf{p}^{\kappa_{1,T}}\right\|_0^2\kappa_{1,T}/T\right)$.[3] Applying Hölder's inequality gives

$$
\sum_{l=1}^{v_T}\sum_{j=1}^{v_T}\left[E\left(\left|\phi_{l,t}\phi_{j,t}\right|^{2+\delta}\right)\right]^{2/(2+\delta)} \leq \sum_{l=1}^{v_T}\left[E\left(\phi_{l,t}^{2(2+\delta)}\right)\right]^{1/(2+\delta)}\sum_{j=1}^{v_T}\left[E\left(\phi_{j,t}^{2(2+\delta)}\right)\right]^{1/(2+\delta)} \leq M\varsigma_{\kappa_{1,T},\delta}^2 \tag{B.5}
$$

by Assumptions 1(i) and 3 (iv), where $\varsigma_{\kappa_{1,T},\delta} \geq M\kappa_{1,T}$ as $\sum_{l=1}^{v_T}\left[E\left(\phi_{l,t}^{2(2+\delta)}\right)\right]^{1/(2+\delta)} \geq$ $\sum_{l=1}^{v_T}E\left(\phi_{l,t}^2\right) \geq M\kappa_{1,T}$ by Jensen's inequality under Assumption 2 (i). As $T^{-1}\sum_{t=1}^{T}\sum_{s\neq t}\alpha_{|t-s|}^{\delta/(2+\delta)}$ $= O(1)$ under Assumption 1(i), by Markov's inequality we obtain

$$
\left\|\Sigma_{T,PP} - \Sigma_{PP}\right\| = O_p\left(\left(\varsigma_{\kappa_{1,T},\delta}\kappa_{1,T}^{-1/2} + \left\|\mathbf{p}^{\kappa_{1,T}}\right\|_0\right)\sqrt{\kappa_{1,T}/T}\right). \tag{B.6}
$$

Combining this result with (B.3) and (B.4) gives (B.1).

[3] Lemma A.1 (i) in Ozabaci et al. (2014) may be wrong; e.g., Newey (1997, p. 162).

Next, from Step 1, we have

$$\begin{aligned}
& \left[\hat{\mu}_z-\mu_z,(\hat{\alpha}-\alpha)^T,\left(\hat{\vartheta}-\vartheta\right)^T\right]^T \\
= \; & \Sigma_{T,PP}^{+}\left(\chi_{T1}+\chi_{T2}+\chi_{T3}\right) \\
= \; & \Sigma_{PP}^{-1}\left(\chi_{T1}+\chi_{T2}+\chi_{T3}\right)+\left(\Sigma_{T,PP}^{+}-\Sigma_{PP}^{-1}\right)\left(\chi_{T1}+\chi_{T2}+\chi_{T3}\right),
\end{aligned}$$

where closely following the proof of Lemma A.2 (ii) in Ozabaci et al. (2014) we obtain $\chi_{T2}=O_p\left(\kappa_{1,T}^{-\xi}\right)$ and $\chi_{T3}=O_p\left(\kappa_{1,T}^{-\xi}\right)$. Under Assumptions 1(i), 2(i), 3(iv), and 4(ii), we have $E\left(\chi_{T1}\right)=0$ and

$$\begin{aligned}
& E\left(\|\chi_{T1}\|^2\right) \\
= \; & T^{-1}E\left[P^{\kappa_{1,T}}\left(W_t,X_t\right)^T P^{\kappa_{1,T}}\left(W_t,X_t\right)U_t^2\right] \\
& +T^{-2}\sum_{t=1}^{T}\sum_{s\neq t}E\left[P^{\kappa_{1,T}}\left(W_s,X_s\right)^T P^{\kappa_{1,T}}\left(W_t,X_t\right)U_tU_s\right] \\
= \; & T^{-1}E\left(U_t^2\right)+T^{-1}\sum_{j=1}^{d_w}\sum_{i=1}^{\kappa_{1,T}}E\left[p_i^2\left(W_{j,t}\right)U_t^2\right]+T^{-1}\sum_{l=1}^{d_x}\sum_{i=1}^{\kappa_{1,T}}E\left[p_i^2\left(X_{l,t}\right)U_t^2\right] \\
& +T^{-2}\sum_{t=1}^{T}\sum_{s\neq t}E\left(U_tU_s\right)+T^{-2}\sum_{j=1}^{d_w}\sum_{i=1}^{\kappa_{1,T}}\sum_{t=1}^{T}\sum_{s\neq t}E\left[p_i\left(W_{j,t}\right)p_i\left(W_{j,s}\right)U_tU_s\right] \\
& +T^{-2}\sum_{l=1}^{d_x}\sum_{i=1}^{\kappa_{1,T}}\sum_{t=1}^{T}\sum_{s\neq t}E\left[p_i\left(X_{l,t}\right)p_i\left(X_{l,s}\right)U_tU_s\right] \\
\leq \; & M\kappa_{1,T}T^{-1}+O\left(T^{-1}\varsigma_{\kappa_{1,T},\delta}\right),
\end{aligned}$$

where we again obtain the second term by applying Davoydov's and Hölder's inequality. Hence, we have

$$\chi_{T1}=O_p\left(T^{-1/2}\left(\kappa_{1,T}^{1/2}+\varsigma_{\kappa_{1,T},\delta}^{1/2}\right)\right). \tag{B.7}$$

by Markov's inequality. Therefore, we have for j=1,2,3

$$\begin{aligned}
& \left\|\left(\Sigma_{T,PP}^{+}-\Sigma_{PP}^{-1}\right)\chi_{Tj}\right\| \\
= \; & \sqrt{\chi_{Tj}^T\left(\Sigma_{T,PP}^{+}-\Sigma_{PP}^{-1}\right)\left(\Sigma_{T,PP}^{+}-\Sigma_{PP}^{-1}\right)\chi_{Tj}} \\
\leq \; & \|\chi_{Tj}\|\,\lambda_{\max}^{1/2}\left\{\left(\Sigma_{T,PP}^{+}-\Sigma_{PP}^{-1}\right)^2\right\}\leq\|\chi_{Tj}\|\left\|\Sigma_{T,PP}-\Sigma_{PP}\right\|.
\end{aligned} \tag{B.8}$$

Combining (B.6), (B.7), and (B.8) gives (B.2) . This completes the proof of this lemma.

Lemma 2 *Under Assumptions 1-4, we have*

$$\Delta_{T,1} \equiv T^{-1}\sum_{t=1}^{T}\left(\hat{U}_t - U_t\right)^2 \varpi_t = O_p\left(\pi_{1,T}^2\right) \tag{B.9}$$

$$\Delta_{T,2} \equiv T^{-1}\sum_{t=1}^{T}\left\|\hat{Q}_t^{\kappa_{2,T}} - Q_t^{\kappa_{2,T}}\right\|^2 \varpi_t = O_p\left(\pi_{1,T}^2 \left\|\mathbf{p}^{\kappa_{2,T}}\right\|_1^2\right) \tag{B.10}$$

where $\varpi_t = 1,\ |\eta_t|,\ X_{1,t}^2,\ \ldots,\ X_{d_x,t}^2$.

Proof: We first verify (B.9). From Step 1, we have $\hat{U}_t - U_t = Z_t - \hat{g}\left(W_t, X_t\right) - U_t$ $= g\left(W_t, X_t\right) - \hat{g}\left(W_t, X_t\right)$, where

$$\begin{aligned}
& \left[\hat{g}\left(W_t, X_t\right) - g\left(W_t, X_t\right)\right]^2 \\
\leq\ & 2\left[\hat{g}\left(W_t, X_t\right) - g^*\left(W_t, X_t\right)\right]^2 + 2\left[g^*\left(W_t, X_t\right) - g\left(W_t, X_t\right)\right]^2 \\
\leq\ & 2\left\{P^{\kappa_{1,T}}\left(W_t, X_t\right)^T\left[\hat{\mu}_z - \mu_z, \left(\hat{\alpha} - \alpha\right)^T, \left(\hat{\vartheta} - \vartheta\right)^T\right]^T\right\}^2 + M\kappa_{1,T}^{-2\xi} \\
\leq\ & 6\left(\hat{\mu}_z - \mu_z\right)^2 + 6d_x\sum_{j=1}^{d_x}\left[\mathbf{p}^{\kappa_{1,T}}\left(X_{j,t}\right)^T\left(\hat{\alpha}_j - \alpha_j\right)\right]^2 \\
& +6d_w\sum_{j=1}^{d_w}\left[\mathbf{p}^{\kappa_{1,T}}\left(W_{j,t}\right)^T\left(\hat{\vartheta}_j - \vartheta_j\right)\right]^2 + M\kappa_{1,T}^{-2\xi},
\end{aligned} \tag{B.11}$$

by Assumption 3 (iii). So, we have

$$\begin{aligned}
\Delta_{T,1} \equiv\ & T^{-1}\sum_{t=1}^{T}\left(\hat{U}_t - U_t\right)^2 \varpi_t \\
\leq\ & 6\left(\hat{\mu}_z - \mu_z\right)^2 T^{-1}\sum_{t=1}^{T}\varpi_t + 6d_x\sum_{j=1}^{d_x}\left(\hat{\alpha}_j - \alpha_j\right)^T\sum_{t=1}^{T}\varpi_t\mathbf{p}^{\kappa_{1,T}}\left(X_{j,t}\right)^T\mathbf{p}^{\kappa_{1,T}}\left(X_{j,t}\right)\left(\hat{\alpha}_j - \alpha_j\right) \\
& +6d_w\sum_{j=1}^{d_w}\left(\hat{\vartheta}_j - \vartheta_j\right)^T\sum_{t=1}^{T}\varpi_t\mathbf{p}^{\kappa_{1,T}}\left(W_{j,t}\right)^T\mathbf{p}^{\kappa_{1,T}}\left(W_{j,t}\right)\left(\hat{\vartheta}_j - \vartheta_j\right) + M\kappa_{1,T}^{-2\xi},
\end{aligned}$$

where we have

$$\begin{aligned}
& a_j \equiv \left(\hat{\alpha}_j - \alpha_j\right)^T T^{-1}\sum_{t=1}^{T}\varpi_t\mathbf{p}^{\kappa_{1,T}}\left(X_{j,t}\right)\mathbf{p}^{\kappa_{1,T}}\left(X_{j,t}\right)^T\left(\hat{\alpha}_j - \alpha_j\right) \\
\leq\ & \lambda_{\max}\left(T^{-1}\sum_{t=1}^{T}\varpi_t\mathbf{p}^{\kappa_{1,T}}\left(X_{j,t}\right)\mathbf{p}^{\kappa_{1,T}}\left(X_{j,t}\right)^T\right)\left\|\hat{\alpha}_j - \alpha_j\right\|^2.
\end{aligned} \tag{B.12}$$

Following the proof of Lemma 1 we have $\lambda_{\max}\left(T^{-1}\sum_{t=1}^{T}\varpi_t\mathbf{p}^{\kappa_{1,T}}\left(X_{j,t}\right)\mathbf{p}^{\kappa_{1,T}}\left(X_{j,t}\right)^T\right) =$ $\lambda_{\max}\left\{E\left(\varpi_t\mathbf{p}^{\kappa_{1,T}}\left(X_{j,t}\right)\mathbf{p}^{\kappa_{1,T}}\left(X_{j,t}\right)^T\right)\right\}$ $+o_p\left(1\right) = O_p\left(1\right)$ by Assumption 4(iii). Hence,

$a_j \equiv O_p\left(\|\hat{\alpha}_j - \alpha_j\|^2\right)$. Similarly, we obtain $\left(\hat{\vartheta}_j - \vartheta_j\right)^T T^{-1}\sum_{t=1}^T \varpi_t \mathbf{p}^{\kappa_{1,T}}(W_{j,t})\,\mathbf{p}^{\kappa_{1,T}}(W_{j,t})^T\left(\hat{\vartheta}_j - \vartheta_j\right)$ $= O_p\left(\left\|\hat{\vartheta}_j - \vartheta_j\right\|^2\right)$. Applying Lemma 1 gives (B.9).

It then follows $\Delta_{T,2} \equiv T^{-1}\sum_{t=1}^T \left\|\hat{Q}_t^{\kappa_{2,T}} - Q_t^{\kappa_{2,T}}\right\|^2 \varpi_t = T^{-1}\sum_{t=1}^T \varpi_t \left\|\mathbf{p}^{\kappa_{2,T}}(U_t) - \mathbf{p}^{\kappa_{2,T}}\left(\hat{U}_t\right)\right\|^2$ $\leq M\|\mathbf{p}^{\kappa_{2,T}}\|_1^2 T^{-1}\sum_{t=1}^T \varpi_t\left(\hat{U}_t - U_t\right)^2 = O_p\left(\pi_{1,T}^2\|\mathbf{p}^{\kappa_{2,T}}\|_1^2\right)$. This completes the proof of this lemma.

Lemma 3 *Under Assumptions 1-6, we have*

$$
\begin{aligned}
\Delta_{T1} &\equiv \frac{1}{Th}\sum_{t=1}^T K_h(Z_t - z)\,\mathcal{W}_t \mathbf{p}^{\kappa_{2,T}}(U_t)^T \rho = O_p(1)\ , \\
\Delta_{T2} &\equiv \frac{1}{Th}\sum_{t=1}^T K_h(Z_t - z)\,\|\mathcal{W}_t\|\left(\hat{U}_t - U_t\right)^2 = O_p\left(\pi_{1,T}^2\right), \\
\Delta_{T3} &\equiv \frac{1}{Th}\sum_{t=1}^T K_h(Z_t - z)\,\mathcal{W}_t\left(\hat{U}_t - U_t\right)\dot{\mathbf{p}}^{\kappa_{2,T}}(U_t)^T \rho = O_p\left(\|\mathbf{p}^{\kappa_{2,T}}\|_1 \kappa_{1,T}^{-\xi} + \pi_{1,T}\right).
\end{aligned}
$$

Proof: First, we consider Δ_{T1}. We have for j=0,1,

$$
\begin{aligned}
&\frac{1}{Th}\sum_{t=1}^T E\left[K_h(Z_t - z)\left(\frac{Z_t - z}{h}\right)^j \mathbf{p}^{\kappa_{2,T}}(U_t)^T \rho\right] \\
= &\sum_{l=1}^{\kappa_{2,T}} \rho_l \int\int K(v)\, v^j p_l(u)\, f(u, hv + z)\, dv du \\
= &\left(\int K(v)\, v^j dv\right)\sum_{l=1}^{\kappa_{2,T}} \rho_l \int p_l(u)\, f(u, z)\, du\left[1 + O\left(h^2\right)\right].
\end{aligned}
$$

By Parseval's identity we have $\sum_{l=1}^\infty \rho_l^2 = \int \gamma^2(u)\, du$ and $\sum_{l=1}^\infty\left[\int p_l(u)\, f(u,z)\right]^2 = \int f^2(u,z)\, du$. Under Assumption 4(v) and applying Hölder's inequality, we show that the above equation is bounded by a constant, which implies $E(\Delta_{T1}) \leq M < \infty$. Similarly, we can show that $E\left(\Delta_{T1}^2\right) = O(1)$. Therefore, we have $\Delta_{T1} = O_p(1)$.

Next, we consider Δ_{T2}, which can be decomposed as

$$
\begin{aligned}
\Delta_{T2} &\leq \frac{2}{Th}(\hat{\chi} - \chi)^T \sum_{t=1}^T K_h(Z_t - z)\,\|\mathcal{W}_t\|\, P^{\kappa_{1,T}}(W_t, X_t)\, P^{\kappa_{1,T}}(W_t, X_t)^T (\hat{\chi} - \chi) \\
&\quad + \frac{2}{Th}\sum_{t=1}^T K_h(Z_t - z)\,\|\mathcal{W}_t\|\left[g(W_t, X_t) - g^*(W_t, X_t)\right]^2 \\
&= 2\left(\Delta_{T2,1} + \Delta_{T2,2}\right),
\end{aligned}
$$

where $\Delta_{T2,2} \equiv (Th)^{-1}\sum_{t=1}^T K_h(Z_t - z)\,\|\mathcal{W}_t\|\left[g(W_t, X_t) - g^*(W_t, X_t)\right]^2 = O_p\left(\kappa_{1,T}^{-2\xi}\right)$ under Assumptions 3 (ii)-(iii). Let $B_T \equiv (Th)^{-1}\sum_{t=1}^T K_h(Z_t - z)\,\|\mathcal{W}_t\|\, P^{\kappa_{1,T}}(W_t, X_t)\, P^{\kappa_{1,T}}(W_t, X_t)^T$.

Evidently, B_T is a p.d.f. matrix, and it is readily seen that $B_T = O_p(1)$ under Assumption 2(i). As $\Delta_{T2,1} = (\hat{\chi} - \chi)^T B_T (\hat{\chi} - \chi) \leq \lambda_{\max}(B_T) \|\hat{\chi} - \chi\|^2$, we have $\Delta_{T2,1} = O_p(\pi_{1,T}^2)$ by Lemma 1.

Finally, we consider Δ_{T3}. By (A.3) and letting $\hat{\chi} - \chi = \left[\hat{\mu}_z - \mu_z, (\hat{\alpha} - \alpha)^T, \left(\hat{\vartheta} - \vartheta\right)^T\right]$, we have

$$\begin{aligned} \Delta_{T3} &= -\frac{1}{Th}\sum_{t=1}^{T} K_h(Z_t - z)\mathcal{W}_t \dot{\mathbf{p}}^{\kappa_{2,T}}(U_t)^T \rho P^{\kappa_{1,T}}(W_t, X_t)^T (\hat{\chi} - \chi) \\ &\quad + \frac{1}{Th}\sum_{t=1}^{T} K_h(Z_t - z)\mathcal{W}_t \dot{\mathbf{p}}^{\kappa_{2,T}}(U_t)^T \rho \left[g(W_t, X_t) - g^*(W_t, X_t)\right] \\ &= -\Delta_{T3,1} + \Delta_{T3,2}, \end{aligned}$$

where $\Delta_{T3,2} \equiv (Th)^{-1}\sum_{t=1}^{T} K_h(Z_t - z)\mathcal{W}_t \dot{\mathbf{p}}^{\kappa_{2,T}}(U_t)^T \rho \left[g(W_t, X_t) - g^*(W_t, X_t)\right] = O_p\left(\|\mathbf{p}^{\kappa_{2,T}}\|_1 \kappa_{1,T}^{-\xi}\right)$ under Assumptions 3(ii)-(iii) and Assumption 4(v). Following the proof of Δ_{T1} we can show that

$$\frac{1}{Th}\sum_{t=1}^{T} K_h(Z_t - z)\mathcal{W}_t \dot{\mathbf{p}}^{\kappa_{2,T}}(U_t)^T \rho \mathbf{p}^{\kappa_{2,T}}(e_{j,t}) = O_p(1)$$

where $e_{j,t} = X_{j,t}$ or $W_{j,t}$ for $j = 1, \ldots, d_x$ (or d_w), as $\int f(u, e_j, z)\dot{p}_l^{\kappa_{2,T}}(u)\,du = -\int p_l^{\kappa_{2,T}}(u)\,\partial f(u, e_j, z)/\partial u du$. Combining this result with Lemma 1, we obtain $\Delta_{T3} = O_p\left(\|\mathbf{p}^{\kappa_{2,T}}\|_1 \kappa_{1,T}^{-\xi} + \pi_{1,T}\right)$. This completes the proof of the lemma.

Table 1: Results from the Monte Carlo simulations for both the conditional mean and smooth coefficients

Sample size	T = 200			T = 400			T = 1600		
	$\widehat{m}(X,Z)$	$\widehat{\beta}_0(Z)$	$\widehat{\beta}_1(Z)$	$\widehat{m}(X,Z)$	$\widehat{\beta}_0(Z)$	$\widehat{\beta}_1(Z)$	$\widehat{m}(X,Z)$	$\widehat{\beta}_0(Z)$	$\widehat{\beta}_1(Z)$
$RMSE$									
DGP_1	0.4799	1.8860	0.3889	0.3797	1.7970	0.2802	0.2332	1.3540	0.1572
DGP_2	0.4889	2.4430	0.8044	0.3831	2.0960	0.5444	0.2336	1.5890	0.2966
DGP_3	0.5209	2.5500	0.8275	0.4304	2.2710	0.5510	0.3138	1.8160	0.2985
$Variance$									
DGP_1	0.2210	0.4215	0.1287	0.1380	0.2333	0.0685	0.0515	0.0775	0.0221
DGP_2	0.2336	1.5840	1.0550	0.1419	0.3207	0.5366	0.0522	0.1026	0.1192
DGP_3	0.2649	1.0000	1.2580	0.1804	0.4943	0.3376	0.0956	0.1331	0.1007

Reported results are the mean values of the root mean squared error and variance of our estimator across simulations.

Table 2: Descriptive statistics

Variable	Mean	Std. Dev.	Min	Max
Dependent Variable				
Gasoline Consumption	222.326	183.501	2.518	5062.414
Environmental Variable				
Income	8380.284	6170.567	522.000	250256.797
Independent Variables				
Price	151.950	20.272	98.999	214.985
Family Size	2.508	1.413	1	14
Number of Vehicles	2.016	1.255	1	20
MPG Rating	21.286	3.589	11.811	27.500
Education	2.811	0.923	1	4
Rural	0.108	-	0	1
Instrumental Variables				
Age	51.929	16.855	17	94
Race Indicator	0.143	-	0	1

Sample size is 30,000. Table reports continuous variables in levels.

Table 3: Final stage point estimates from the proposed semiparametric IV estimator

Variable	Mean	25th	50th	75th
Intercept	5.780***	5.643***	5.880***	6.098***
	0.332	0.277	0.254	0.236
Price	-0.587***	-0.698***	-0.609***	-0.494***
	0.060	0.047	0.045	0.051
Family Size	0.072***	0.067***	0.069***	0.076***
	0.007	0.007	0.004	0.006
Number of Vehicles	0.116***	0.090***	0.112***	0.131***
	0.009	0.007	0.006	0.007
MPG	-0.493***	-0.597***	-0.496***	-0.451***
	0.052	0.038	0.038	0.040
Education	-0.109***	-0.133***	-0.113***	-0.087***
	0.011	0.008	0.007	0.008
Rural	0.094***	0.066**	0.087***	0.116***
	0.030	0.023	0.021	0.018

Sample size is 30,000. Statistically significant point estimates at the 1, 5, or 10 percent level are denoted by ***,**,*.

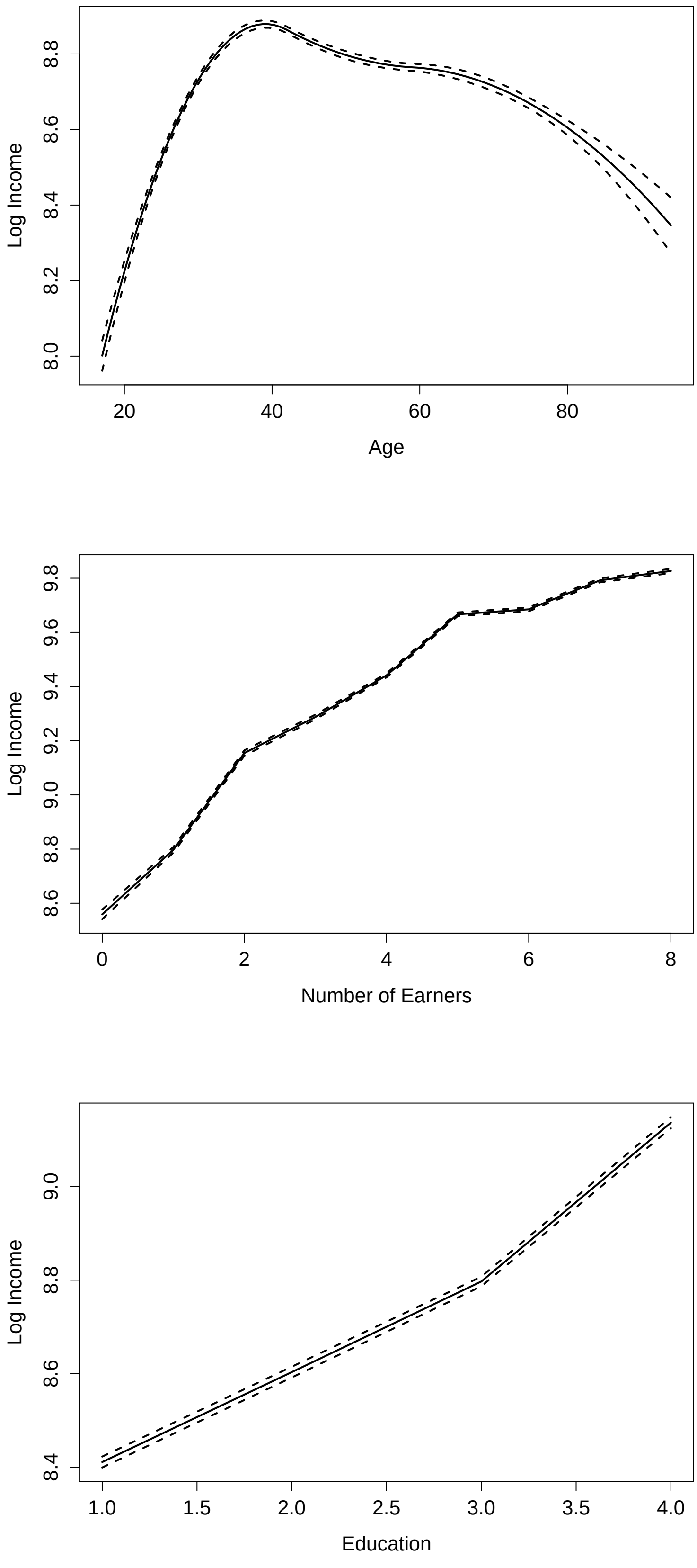

Figure 1: Estimated first stage relationship between log income and age, number of wage earners, and education.

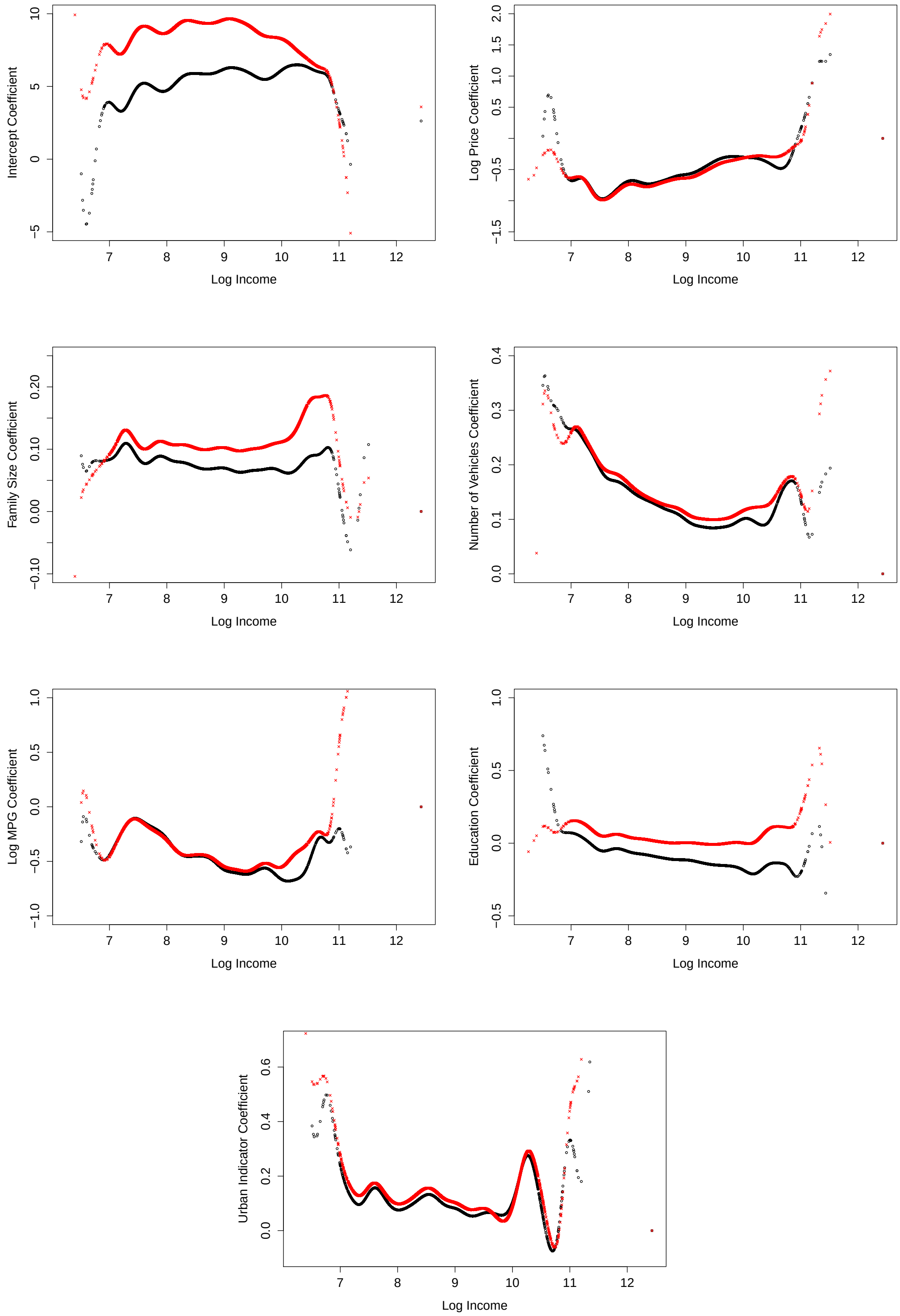

Figure 2: Estimated smooth coefficient functions of log income. Black circles denote estimates from the proposed IV estimator, and the red plus symbols denote the non-IV smooth coefficient estimates.

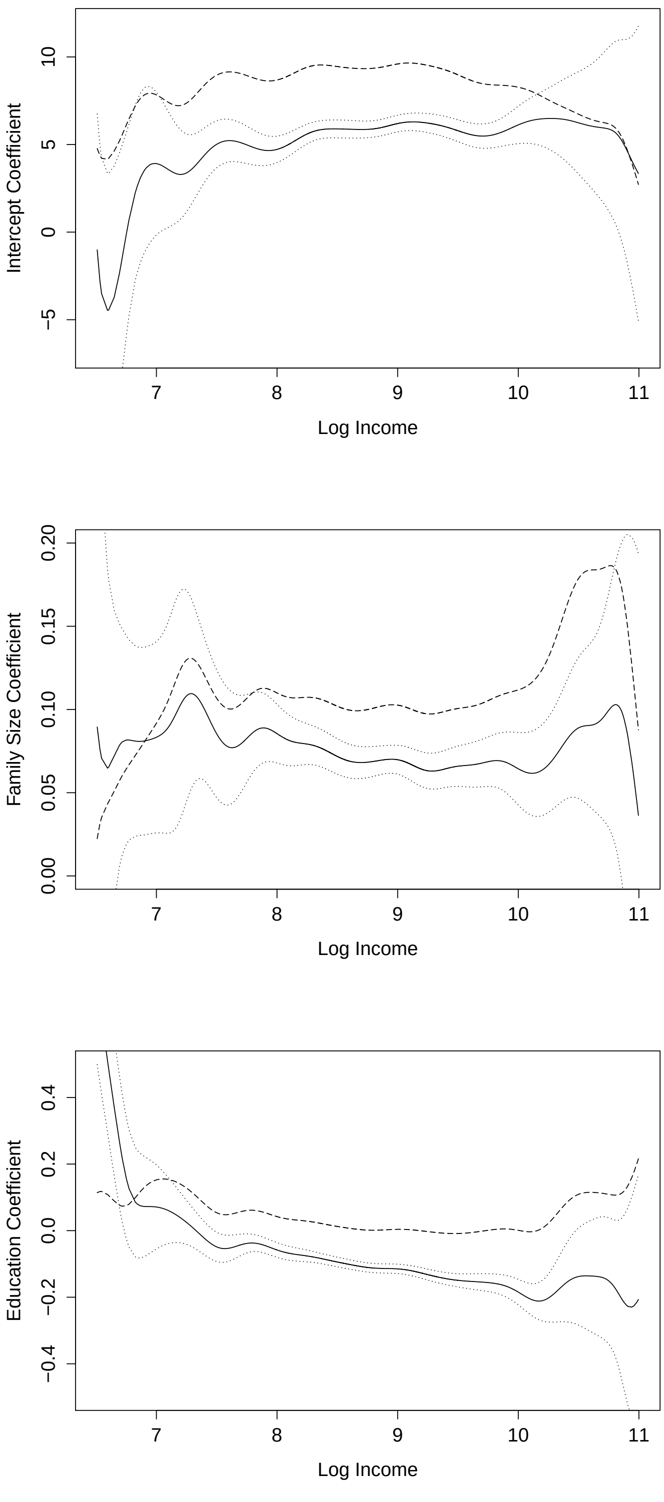

Figure 3: Estimated smooth coefficient functions of log income for the intercept, family size, and education variables. The solid line represents the proposed IV estimate with bootstrapped 95 percent confidence bounds placed above and below. The dashed line is the function estimate from a standard non-IV local linear estimator.