EFFICIENT ESTIMATION OF A SEMIPARAMETRIC PARTIALLY LINEAR SMOOTH COEFFICIENT MODEL

A Dissertation

by

SITTISAK LEELAHANON

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2002

Major Subject: Economics

UMI Number: 3072483

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

EFFICIENT ESTIMATION OF A SEMIPARAMETRIC PARTIALLY LINEAR SMOOTH COEFFICIENT MODEL

A Dissertation

by

SITTISAK LEELAHANON

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Qi Li
(Chair of Committee)

James M. Griffin
(Member)

Rajiv Sarin
(Member)

Emanuel Parzen
(Member)

Leonardo Auernheimer
(Head of Department)

December 2002

Major Subject: Economics

ABSTRACT

Efficient Estimation of a Semiparametric

Partially Linear Smooth Coefficient Model. (December 2002)

Sittisak Leelahanon, B.Eng., Chulalongkorn University;

B.B.A., Ramkhamhaeng University

Chair of Advisory Committee: Dr. Qi Li

In this dissertation I propose a general series method to estimate the semiparametric partially linear smooth coefficient model. The consistency and $\sqrt{n}$-normality property of the estimated parameters of the partially linear part are established and furthermore it attains the semiparametric efficiency bound when the error is conditional homoskedastic. The convergence rates of the estimators of the smooth coefficient functions are also derived and a simulation is conducted to make the results more concrete.

The application of this model is illustrated by the case of inflation rate forecasting using the unemployment rate and the industry capacity utilization rate. Forecasting efficiency is compared using the simple autoregressive model, the smooth coefficient model, and a semiparametric partially linear smooth coefficient model. Specification tests are also performed.

Another application in this dissertation is to show that one can better forecast inflation using the past information of money growth by allowing for potentially complicated nonlinearities in the relationship between money growth and inflation. Many nonparametric and semiparametric models have been used to compare the forecasting efficiency with the parametric VAR approach.

To My Mother

ACKNOWLEDGMENTS

I thank Dr. Qi Li for his valuable advice and for encouraging me to finish this dissertation as well as Lance J. Bachmeier for his help in completing chapter III. I also thank Dr. Emanuel Parzen for his valuable suggestions, Dr. James M. Griffin for his comments in doing empirical research and every committee member for all the help.

TABLE OF CONTENTS

Page

LIST OF TABLES

TABLE Page

CHAPTER I

INTRODUCTION

Semiparametric and nonparametric estimation techniques have attracted much attention among econometricians and statisticians. One popular semiparametic specification is the partially linear model of Robinson (1988). It can be genrealized to the semiparametric smooth coefficient model. Li et al. (2002) has shown that the semiparametric smooth coefficient model can be estimated consistently by using the kernel method. However, it can be shown that their estimator is not efficient; and it is hard to establish the asymptotic distribution of an estimator of the smooth coefficient model with partially linear component included.

The partially linear smooth coefficient model can be used in many economics applications such as demand analysis, In chapter II, I propose to estimate the semiparametric partially linear smooth coefficient model using the general series method. The advantage of using the series method is that one can estimate the parameters and the unknown function simultaneously. Furthermore the series estimators have well-defined meanings even when the model is misspecified. However, using general series estimation methods, it is difficult to establish the asymptotic normality result for the nonparametric components under primitive conditions. Using the results of Newey (1997), one can derive their convergence rates.

One of the central propositions in economics is that there is a relationship between the inflation rate and the growth rate of the money supply. This relationship has been the basis of a number of policies that have been proposed for the US economy, including the use of money growth as an indicator variable for inflation. Friedman

---

The journal model is Journal of Econometrics.

(1988) summarized the evidence on the use of money growth as an indicator variable for inflation, and concluded that there is no stable empirical relationship between the two variables. Friedman and Kuttner (1992) presented extensive evidence confirming that conclusion. Chapter III asks whether the dismissal of money growth as an indicator for inflation is warranted. Specifically, empirical research has focused on forecasts made using vector autoregressive (VAR) models. If the goal is to produce out-of-sample forecasts, however, VAR models may not be the best choice because they are often overparametrized. If the system of interest is nonlinear, there may be gains from using a nonlinear forecasting model rather than a linear approximation.

Chapter III allows for potentially complicated nonlinearities in the relationship between money growth and inflation. The approach is to ask whether a nonparametric model, that includes money growth, forecasts inflation better out-of-sample than an autoregressive model of inflation.

Over the last few decades, the semiparametric methods have been proposed and widely used in many real world applications. Both econometricians and statisticians have developed the theory specifically for the semiparametric model. One of the most popular model is the Robinson's (1988) semiparametric partially linear model. However, the literature assumes that the bandwidth h is prespecified or simply given by some ad-hoc method.

It is well known that the bandwidth is of crucial importance in nonparametric and semiparametric estimations. The optimal bandwidth is needed to obtain a good estimate. Many data-driven bandwidth methods have been proposed and the commonly used ones are the cross-validation (CV) method and generalized cross-validation (GCV) method. Härdle, Hall and Marron (1988) defined the theoretically optimal bandwidth and show that the bandwidth obtained by the cross-validation method converges in probability to the theoretically optimal value.

Chapter IV will consider the convergence behavior of the cross-validation bandwidth under the semiparametric partially linear model setting. I propose that the finite dimensional parameter estimates and the cross-validation bandwidth can be obtained simultaneously by minimizing a sum of squared errors. This means the bandwidth need not be prespecified. I will show that the parameter estimates are still $\sqrt{n}$-consistent and asymptotically normally distributed and the cross-validation bandwidth estimates also converge in probability uniformly over some shrinking compact sets to the theoretically optimal bandwidth.

CHAPTER II

EFFICIENT ESTIMATION OF A SEMIPARAMETRIC PARTIALLY LINEAR SMOOTH COEFFICIENT MODEL

A. Introduction

Semiparametric and nonparametric estimation techniques have attracted much attention among econometricians and statisticians. One popular semiparametic specification is the partially linear model of Robinson (1988),

$$Y_i = X_i'\beta + g(Z_i) + u_i \qquad i = 1, \ldots, n \tag{2.1}$$

where $X_i'\beta$ is the parametric component and $g(Z_i)$ is the nonparametric component which is the unknown function. This model can be genrealized to the semiparametric smooth coefficient model

$$Y_i = X_i'\beta(Z_i) + u_i \qquad i = 1, \ldots, n \tag{2.2}$$

where $\beta(z)$ is a vector of unknown smooth functions of Z_i. The smooth coefficient model is an appropriate setting, for example, in the framework of a cross-sectional production function where $X_i = (Labor_i, Capital_i)$ and $Z_i = R\&D_i$. The smooth coefficient model suggests that the labor and capital input coefficients may vary directly with firm's R&D input, so the marginal productivity of labor and capital depend on firm's R&D values. The partially linear model suggests that the R&D variable has a neutral effect on the production function, i.e., only shift the production frontier. In contrast, the smooth coefficient model allows the R&D variable to have a non-neutral effect on the production function.

The semiparametric smooth coefficient model has the advantage that it allows more flexibility in functional form than a parametric linear model or a semiparametric partilly linear model. However, if the dimension of Z_i is greater than one, the smooth coefficient model still suffers form the curse of dimensionality problem. Li et al. (2002) has shown that the semiparametric smooth coefficient model can be estimated consistently by using the kernel method. However, it can be shown that their estimator is not efficient; and it is hard to establish the asymptotic distribution of an estimator of the smooth coefficient model with partially linear component included. So, in this chapter I will consider the semiparametric partially linear smooth coefficient model

$$Y_i = W_i'\gamma + X_i'\beta\,(Z_i) + u_i \qquad i = 1, \ldots, n \tag{2.3}$$

The partially linear smooth coefficient model can be used in many economics applications such as demand analysis, where in the model, $Y_i = Quantity\ Demand_i$, $X_i = (1, Price_i)$, $Z_i = Price\ of\ Substitution\ good_i$ and $W_i = Income_i$. Bresnahan(1982) has shown that in some situations one needs to allow the effect of substitution good price not only shift the demand curve but also change the slope of the demand curve. In this chapter I propose to estimate the semiparametric partially linear smooth coefficient model using the general series method. The advantage of using the series method is that one can estimate the parameters and the unknown function simultaneously.[1] Furthermore the series estimators have well-defined meanings even when the model is misspecified. However, using general series estimation methods, it is difficult to establish the asymptotic normality result for the nonparametric components under primitive conditions. Using the results of Newey (1997), one can derive their convergence rates.

This chapter is organized as follows. Section B discusses the estimation method

[1]See Li (2000).

and states the main theorem. Section C shows the simulation results. Section D applies the partially linear model to inflation rate prediction using the unemployment rate and the industry capacity utilization rate. Section E concludes and the last section contains the proofs.

B. Estimation

Consider the partially linear smooth coefficient regression model:

$$Y_i = W_i'\gamma + X_i'\beta(Z_i) + u_i \qquad i = 1, \ldots, n \tag{2.4}$$

where the prime denotes transpose, $W_i = (W_{i1}, \ldots, W_{iq})'$ is a $q \times 1$ vector of random variables, $\gamma = (\gamma_1, \ldots \gamma_q)'$ is a $q \times 1$ vector of unknown parameters, $X_i = (X_{i1}, \ldots, X_{id})'$ is a $d \times 1$ vector of random variables, $Z_i = (Z_{i1}, \ldots, Z_{ir})'$ is an $r \times 1$ vector of random variables, and $\beta(\cdot) = (\beta_1(\cdot), \ldots, \beta_d(\cdot))'$ is a $d \times 1$ vector of unknown smooth functions. u_i is an error term with $E(u_i|W_i, X_i, Z_i) = 0$.

Using series estimator method, I approximate $\beta_l(z)$ by a linear combination of k_l basis functions, i.e. $p_l^{k_l}(z)'\alpha_l^{k_l}$,where $p_l^{k_l}(z) = [p_{l1}(z), \ldots, p_{lk_l}(z)]'$ is a $k_l \times 1$ vector of basis functions and $\alpha_l^{k_l} = (\alpha_{l1}, \ldots \alpha_{lk_l})'$ is a $k_l \times 1$ vector of unknown constants.The approximation functions $p_l^{k_l}(z)$ have the property that, as k_l grows, there is a linear combination of $p_l^{k_l}(z)$ that can approximate any smooth function β_l arbitrarily well in the mean squared error sense.

Define the $K \times 1$ matrices $p_i^K(X_i, Z_i) = \left(X_{i1}p_1^{k_1}(Z_i)', \ldots, X_{id}p_d^{k_d}(Z_i)'\right)'$ and $\alpha = \left(\alpha_1^{k_1\prime}, \ldots, \alpha_d^{k_d\prime}\right)'$, where $K = \sum_{l=1}^{d} k_l$, hence, I use a linear combination of K functions, $p_i^K(X_i, Z_i)'\alpha$, to approximate $X_i'\beta(Z_i)$. Hence I can rewrite (2.4) as:

$$Y_i = W_i'\gamma + p_i^K(X_i, Z_i)'\alpha + \left(X_i'\beta(Z_i) - p_i^K(X_i, Z_i)'\alpha\right) + u_i \tag{2.5}$$

In matrix notation, let $Y = (Y_1, \ldots, Y_n)'$, $u = (u_1, \ldots, u_n)'$, $W = (W_1, \ldots, W_n)'$, $G = (X_1'\beta(Z_1), \ldots, X_n'\beta(Z_n))'$, and $P = \left(p_1^K(X_1, Z_1), \ldots, p_n^K(X_n, Z_n)\right)'$. Hence, the model (2.5) can be written in matrix notation as:

$$Y = W\gamma + P\alpha + (G - P\alpha) + u \tag{2.6}$$

Define $M = P(P'P)^{-1}P'$ and $\tilde{A} = MA$. Premultiply (2.6) by M leads to:

$$\tilde{Y} = \tilde{W}\gamma + P\alpha + \left(\tilde{G} - P\alpha\right) + \tilde{u} \tag{2.7}$$

Subtracting (2.7) to (2.6) gives:

$$Y - \tilde{Y} = \left(W - \tilde{W}\right)\gamma + \left(G - \tilde{G}\right) + u - \tilde{u} \tag{2.8}$$

The proposed estimator of γ is the least squares regression of $Y - \tilde{Y}$ on $W - \tilde{W}$, i.e.,

$$\hat{\gamma} = \left[\left(W - \tilde{W}\right)'\left(W - \tilde{W}\right)\right]^{-1}\left(W - \tilde{W}\right)'\left(Y - \tilde{Y}\right) \tag{2.9}$$

and the estimator of $\beta(z)$ is $\hat{\beta}(z) = \left(\hat{\beta}_1(z), \ldots, \hat{\beta}_d(z)\right)$, where $\hat{\beta}_l(z) = p_l^{k_l}(z)\,\hat{\alpha}_l^{k_l}$,and $\hat{\alpha} = \left(\hat{\alpha}_1^{k_1\prime}, \ldots, \hat{\alpha}_d^{k_d\prime}\right)'$ can be calculated by:

$$\hat{\alpha} = (P'P)^{-1}P'(Y - W\hat{\gamma}) \tag{2.10}$$

Under the assumptions given below, both $\left[\left(W - \tilde{W}\right)'\left(W - \tilde{W}\right)\right]^{-1}$ and $(P'P)^{-1}$ are asymptotically nonsingular, hence, $\hat{\gamma}$ and $\hat{\alpha}$ given in (2.9) and (2.10) are numerically identical to the least squares estimator of regressing Y on (W, P).

First I introduce the definition of the class of smooth coefficient functions.

Definition II.1 *A function $g(x, z)$ belongs to a smooth coefficient class of functions $\mathcal{G}$ if (i) $g(x, z) = \sum_{l=1}^{d} x_l h_l(z)$ for some $h_l(z)$ and $g(x, z)$ is continuous in its support S_l, where S_l is a compact subset of $\mathbb{R}^{r+d}$, and (ii) $E\left[g(X, Z)^2\right] = \sum_{l=1}^{d} E\left[X_l^2 h_l(Z)^2\right] <$*

∞.

Note that, by the second condition in the definition, the function g is an element in L_2 and by the independence of X and Z, which will be assumed later, this implies that h_l is also an element in L_2.

For any function $f(x,z)$, let $E_C[f(x,z)]$ denote the projection of $f(x,z)$ onto the smooth coefficient space $\mathcal{G}$ (under the L_2-norm). That is, $E_C[f(x,z)]$ is an element that belongs to $\mathcal{G}$ and it is the closest function to $f(x,z)$ among all the functions in $\mathcal{G}$. More specifically,

$$\begin{aligned} & E\left[\left(f(X_i,Z_i) - E_C[f(X_i,Z_i)]\right)\left(f(X_i,Z_i) - E_C[f(X_i,Z_i)]\right)'\right] \qquad (2.11) \\ = & \inf_{h_l \in L_2} E\left[\left(f(X_i,Z_i) - \sum_{l=1}^{d} X_{il}h_l(Z_i)\right)\left(f(X_i,Z_i) - \sum_{l=1}^{d} X_{il}h_l(Z_i)\right)'\right] \end{aligned}$$

where the infimum of (2.11) is in the sense that, for all $g(x,z) = \sum_{l=1}^{d} x_l h_l(z) \in \mathcal{G}$ (or equivalently for all $h_l \in L_2$),

$$\begin{aligned} & E\left[\left(f(X_i,Z_i) - E_C[f(X_i,Z_i)]\right)\left(f(X_i,Z_i) - E_C[f(X_i,Z_i)]\right)'\right] \qquad (2.12) \\ \leq & E\left[\left(f(X_i,Z_i) - \sum_{l=1}^{d} X_{il}h_l(Z_i)\right)\left(f(X_i,Z_i) - \sum_{l=1}^{d} X_{il}h_l(Z_i)\right)'\right] \end{aligned}$$

where for square matrices A and B, $A \leq B$ means that $A - B$ is negative semidefinite.

Proposition II.1 *Define $\theta(X,Z) = E[W|X,Z]$. There exists $m(X,Z) = E_C[\theta(X,Z)]$ which is the solution of the minimization problem:*

$$\begin{aligned} & E\left[\left(\theta(X_i,Z_i) - m(X_i,Z_i)\right)\left(\theta(X_i,Z_i) - m(X_i,Z_i)\right)'\right] \qquad (2.13) \\ = & \inf_{h_l \in L_2} E\left[\left(\theta(X_i,Z_i) - \sum_{l=1}^{d} X_{il}h_l(Z_i)\right)\left(\theta(X_i,Z_i) - \sum_{l=1}^{d} X_{il}h_l(Z_i)\right)'\right] \end{aligned}$$

Proof. The proof is given in the last section. ■

Since $W_i = \theta(X_i, Z_i) + v_i$, where $E[v_i|X_i, Z_i] = 0$. Hence, one can find the projection of W_i on $\mathcal{G}$ as the following:

$$\begin{aligned}
& \inf_{g\in\mathcal{G}} E\left[(W_i - g(X_i, Z_i))(W_i - g(X_i, Z_i))'\right] \\
= & \inf_{g\in\mathcal{G}} E\left[(\theta(X_i, Z_i) - g(X_i, Z_i))(\theta(X_i, Z_i) - g(X_i, Z_i))'\right] + E[v_i v_i'] \\
= & E\left[(\theta(X_i, Z_i) - m(X_i, Z_i))(\theta(X_i, Z_i) - m(X_i, Z_i))'\right] + E[v_i v_i'] \\
= & E\left[(W_i - m(X_i, Z_i))(W_i - m(X_i, Z_i))'\right] \qquad (2.14)
\end{aligned}$$

Since $v_i \perp \mathcal{G}$, therefore, $m(X_i, Z_i)$ is also the projection of W_i on $\mathcal{G}$. Note that $m(X_i, Z_i)$ has dimension $q \times 1$. Let $m_j(X_i, Z_i)$ be the jth component of $m(X_i, Z_i)$, i.e. $m(X_i, Z_i) = (m_1(X_i, Z_i), \ldots, m_q(X_i, Z_i))'$.

The following assumptions are needed to establish the asymptotic distribution of $\hat{\gamma}$ and the convergence rates of $\hat{\beta}(z)$.

Assumption II.1 *(i) $(Y_i, W_i, X_i, Z_i)_{i=1}^n$ are independent and identically distributed as (Y, W, X, Z) and the support of (W, X, Z) is a compact subset of $\mathbb{R}^{q+d+r}$; (ii) both $\theta(X_i, Z_i)$ and $var[Y|W = w, X = x, Z = z]$ are bounded functions on the support of (W, X, Z).*

Assumption II.2 *(i) For every K there is a nonsingular matrix B such that for $P^K(x, z) = Bp^K(x, z)$; the smallest eigenvalue of $E\left[P^K(X_i, Z_i) P^K(X_i, Z_i)'\right]$ is bounded away from zero uniformly in K; (ii) there is a sequence of constants $\zeta_0(K)$ satisfying $\sup_{(x,z)\in S} \left\|P^K(x, z)\right\| \leq \zeta_0(K)$ and $K = K_n$ such that $(\zeta_0(K))^2 K/n \to 0$ as $n \to \infty$, where S is the support of (X, Z).*

Assumption II.3 *(i) For $f(x, z) = g(x, z) = \sum_{l=1}^d x^l \beta(z)$ or $f(x, z) = m_j(x, z)$ $(j = 1, \ldots, q)$, there exists some $\delta_l > 0$ $(l = 1, \ldots, d)$, $\alpha_f = \alpha_{fK} = \left(\alpha_1^{k_1 \prime}, \ldots, \alpha_d^{k_d \prime}\right)'$, $\sup_{(x,z)\in S} \left|f(x, z) - P^K(x, z)' \alpha_f\right| = O\left(\sum_{l=1}^d k_l^{-\delta_l}\right)$ as $\min\{k_1, \ldots, k_d\} \to \infty$; (ii) $\sqrt{n}\left(\sum_{l=1}^d k_l^{-\delta_l}\right) \to 0$ as $n \to \infty$.*

Assumption II.1 is a standard assumption of the semiparametric regression model. Assumption II.2 imposes an orthonormalization on the approximating functions and ensures that $P'P$ is asymptotically nonsingular. Assumption II.3 says that there exists some $\delta_l > 0$ $(l = 1, \ldots, d)$ such that the uniform approximation error to the function shrinks at rate $\sum_{l=1}^{d} k_l^{-\delta_l}$. The assumption II.2 and assumption II.3 are not the primitive conditions but it is known that many series functions satisfy these conditions, e.g., power series and spline.

Under the above assumptions, I can state the main theorem.

Theorem II.1 *Define* $\varepsilon_i = W_i - m(X_i, Z_i)$, *where* $m(X_i, Z_i)$*is defined by* (2.13), *and assume that* $\Phi \equiv E[\varepsilon_i \varepsilon_i']$ *is positive definite, then under assumption II.1 to II.3, I have*

(i)$\sqrt{n}(\hat{\gamma} - \gamma) \rightarrow N(0, \Sigma)$ *in distribution, where* $\Sigma = \Phi^{-1}\Omega\Phi^{-1}$,

where $\Omega = E[\sigma_u^2(W_i, X_i, Z_i)\varepsilon_i\varepsilon_i']$ *and* $\sigma_u^2(W_i, X_i, Z_i) = E[u_i^2|W_i = w, X_i = x, Z_i = z]$.

(ii) A consistent estimator of Σ *is given by* $\hat{\Sigma} = \hat{\Phi}^{-1}\hat{\Omega}\hat{\Phi}^{-1}$,

where $\hat{\Phi} = \frac{1}{n}\sum_{i=1}^{n}\left(W_i - \tilde{W}_i\right)\left(W_i - \tilde{W}_i\right)', \hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2\left(W_i - \tilde{W}_i\right)\left(W_i - \tilde{W}_i\right)'$, $\tilde{W}_i$ *is the ith row of* $\tilde{W}$ *and* $\hat{u}_i = Y_i - W_i'\hat{\gamma} - P_i^K(X_i, Z_i)'\hat{\alpha}$.

Proof. The proof is given in the last section. ∎

Under the conditional homoskedastic error assumption, $E[u_i^2|W_i, X_i, Z_i] = \sigma_u^2$, The estimator $\hat{\gamma}$ is semiparametric efficient in the sense that the inverse of the asymptotic variance of $\sqrt{n}(\hat{\gamma} - \gamma)$ equals the semiparametric efficiency bound. From the result of Chamberlain(1992), the semiparemetric efficiency bound for the inverse of the asymptotic variance of an estimator of γ is

$$J_0 = \inf_{g \in \mathcal{G}} E\left[(W_i - g(X_i, Z_i))(var[u_i|W_i, X_i, Z_i])^{-1}(W_i - g(X_i, Z_i))'\right] \quad (2.15)$$

Under the conditional homoskedastic error assumption, $var[u_i|W_i, X_i, Z_i] = \sigma_u^2$, then (2.15) can be rewritten as

$$\begin{aligned} J_0 &= \frac{1}{\sigma_u^2}\inf_{g\in\mathcal{G}} E\left[\left(W_i - g\left(X_i, Z_i\right)\right)\left(W_i - g\left(X_i, Z_i\right)\right)'\right] \\ &= \frac{1}{\sigma_u^2} E\left[\left(W_i - m\left(X_i, Z_i\right)\right)\left(W_i - m\left(X_i, Z_i\right)\right)'\right] \\ &= \frac{1}{\sigma_u^2} E\left[\varepsilon_i \varepsilon_i'\right] = \frac{\Phi}{\sigma_u^2} \end{aligned} \tag{2.16}$$

Also, under the conditional homoskedastic error assumption, $\Omega = \sigma_u^2 \Phi$, and thus the variance of $\hat{\gamma}$ equals to $\sigma_u^2 \Phi^{-1}$. Hence, $\Sigma^{-1} = J_0$ which implies that $\hat{\gamma}$ attains the semiparemetric efficiency bound under the conditional homoskedastic error assumption.

However, under the conditional hetroskedastic error, $E\left[u_i^2 | W_i, X_i, Z_i\right] = \sigma_u^2\left(W_i, X_i, Z_i\right)$, the estimator $\hat{\gamma}$ is not semiparametric efficient. One can improve the semiparametric efficiency of the estimator by using GLS instead of OLS. First assume for a moment that $\sigma_u^2\left(w, x, z\right)$ is known, let $\sigma_i = \sqrt{\sigma_u^2\left(W_i, X_i, Z_i\right)}$ and assume that σ_i is bounded away from zero. Intuitively, the estimator $\hat{\gamma}_{GLS}$ obtained from regressing Y_i/σ_i on W_i/σ_i and $p_i^K\left(X_i, Z_i\right)/\sigma_i$ is semiparametric efficient because u_i/σ_i is now conditional homoskedastic.

In practice, $\sigma_u^2\left(w, x, z\right)$ is unknown. One can use the OLS method to estimate $\hat{\gamma}$ and $\hat{\alpha}$ by using (2.9) and (2.10) and then estimate u_i by $\hat{u}_i = Y_i - W_i'\hat{\gamma} - P_i^K\left(X_i, Z_i\right)'\hat{\alpha}$. Based on $\left(\hat{u}_i^2, W_i, X_i, Z_i\right)_{i=1}^n$, one can get the estimator $\hat{\sigma}_i = \sqrt{\hat{\sigma}_u^2\left(W_i, X_i, Z_i\right)}$ where $\hat{\sigma}_u^2\left(w, x, z\right)$ can be obtained by using some nonparametric methods. Then regressing $Y_i/\hat{\sigma}_i$ on $W_i/\hat{\sigma}_i$ and $p_i^K\left(X_i, Z_i\right)/\hat{\sigma}_i$ will result in a semiparametric efficient estimator of γ provided that $\hat{\sigma}_u^2\left(w, x, z\right)$ converges to $\sigma_u^2\left(w, x, z\right)$ uniformly for all (w, x, z) in the compact support of (W, X, Z) with some certain rates and perhaps with some other extra regularity conditions.

The next theorem gives the convergence rates of $\hat{g}\left(x, z\right) = p^K\left(x, z\right)'\hat{\alpha}$ to $g\left(x, z\right) =$

$x'\beta(z)$.

Theorem II.2 *Under assumption II.1 to II.3, let $\mathcal{S}$ denote the support of (X, Z), then*

(i) $\sup_{(x,z)\in\mathcal{S}} |\hat{g}(x,z) - g(x,z)| = O_p\left(\zeta_0(K)\left(\sqrt{K}/\sqrt{n} + \sum_{l=1}^{d} k_l^{-\delta_l}\right)\right)$.

(ii) $\frac{1}{n}\sum_{i=1}^{n}(\hat{g}(X_i, Z_i) - g(X_i, Z_i))^2 = O_p\left(K/n + \sum_{l=1}^{d} k_l^{-2\delta_l}\right)$.

(iii) $\int_{\mathcal{S}}(\hat{g}(x,z) - g(x,z))^2 dF(x,z) = O_p\left(K/n + \sum_{l=1}^{d} k_l^{-2\delta_l}\right)$, *where F is the joint cummulative distribution function of X and Z.*

Proof. The proof is given in the last section. ∎

Note that the convergence rate of $\hat{g}(x,z)$ to $g(x,z)$ using semiparametric estimator $\hat{\gamma}$ is the same as the convergence rate of nonparametric series estimators. This is to be expected because the convergence rate of $\hat{\gamma}$ to γ is faster.

The next theorem states the convergence rate of $\hat{\beta}_l(z) = p_l^{k_l}(z)\hat{\alpha}_l^{k_l}$ to $\beta_l(z)$ for $l = 1, \ldots, d$.

Theorem II.3 *Under assumption II.1 to II.3, let $\mathcal{S}_z$ denote the support of Z, then, for $l = 1, \ldots, d$,*

(i) $\sup_{z\in\mathcal{S}_z}\left|\hat{\beta}_l(z) - \beta_l(z)\right| = O_p\left(\zeta_0(K)\left(\sqrt{K}/\sqrt{n} + k_l^{-\delta_l}\right)\right)$.

(ii) $\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_l(z) - \beta_l(z)\right)^2 = O_p\left(K/n + k_l^{-2\delta_l}\right)$.

(iii) $\int_{\mathcal{S}_z}\left(\hat{\beta}_l(z) - \beta_l(z)\right)^2 dF_z(z) = O_p\left(K/n + k_l^{-2\delta_l}\right)$, *where F_z is the cummulative distribution function of Z.*

Proof. The proof is very similar to the proof of theorem II.2, so it will be omitted here. ∎

The remaining of this chapter is to give the primitive conditions for power series and B-splines such that the assumption II.1 to II.3 hold. Newey (1997) gave such conditions and they are restated here for conveniences.

Assumption II.4 *(i) The support of (X, Z) is a Cartesian product of compact connected intervals on which (X, Z) has an absolutely continuous probability density function that is bounded above by a positive constant and bounded away from zero; (ii) for $l = 1, \ldots, d$, $f_l(x, z)$ is continuously differentiable of order c_l on the support S, where $f_l(x, z) = x\beta_l(z)$ or $f_l = m_l$.*

Assumption II.5 *The support of (X, Z) is $[-1, 1]^{d+r}$.*

Newey(1997) showed that assumption II.4 implies that assumptions II.2 and II.3 hold for the power series with $\zeta_0(K) = O(K)$ and $\delta_l = c_l/(r+1)$, $l = 1, \ldots, d$. Also Newey(1997) showed that Assumption II.4 and II.5 imply that assumption II.2 and II.3 hold for B-splines with $\zeta_0(K) = O\left(\sqrt{K}\right)$. Hence, the results of theorem II.1 to II.3 still hold with $\zeta_0(K)$ replaced by K for the power series and $\zeta_0(K)$ replaced by $\sqrt{K}$ for B-splines.

C. Simulation

To concrete the results, I conduct a Monte Carlo simulation. The first data generating process (DGP1) based on the true regression,

$$Y_i = 4 + 0.5W_i + X_i + X_i \cdot 24Z_i^3 \exp(-24Z_i) + u_i \qquad i = 1, \ldots, n \tag{2.17}$$

where $\{u_i\}_{i=1}^n$ is i.i.d. vector of normal random variables with mean 0 and variance 0.25, $\{Z_i\}_{i=1}^n$ is generated by the i.i.d. uniform[0,2] vector of random variables, $W = V_1 + 2V_3$ and $X = V_2 + V_3$, where V_j, $j = 1, 2, 3$, are i.i.d. uniform[0,2] vector of random variables. Let $g_i = 4 + 0.5W_i + X_i + X_i \cdot 24Z_i^3 \exp(-24Z_i)$ be the undisturbed value of the function. [2] For each sample size $n = 100$ and $n = 200$, I repeat the

[2] This function is picked from an example in Hart (1997).

simulation for 2000 times for each estimation method and compare the average mean squared error (AMSE) of $\hat{\gamma}$ and $\hat{g}$ defined by $AMSE(\hat{\gamma}) = \frac{1}{2000}\sum_{i=1}^{2000}(\hat{\gamma}_i - 0.5)^2$ and $AMSE(\hat{g}) = \frac{1}{2000}\sum_{i=1}^{2000} ASE(\hat{g}_i)$ where $\hat{\gamma}_i$ and $ASE(\hat{g}_i) = \frac{1}{n}\sum_{j=1}^{n}(\hat{g}_{ij} - g_{ij})^2$ are the estimates of $\gamma = 0.5$ and the average squared error (ASE) of $\hat{g}$ for the i^{th} iteration respectively. There are three estimation methods conducted in this simulation, i.e., B-Spline, Power Series, and Kernel method. For a B-spline method, I use the univariate B-spline basis function with order $r = 4$ defined by

$$B_r(z|t_0, \ldots, t_r) = \frac{h^{1-r}}{(r-1)!}\sum_{i=0}^{r}(-1)^i\binom{r}{i}[\max(0, z - t_i)]^{r-1} \tag{2.18}$$

where $t_0, \ldots, t_r$ are the evenly spaced designed knots on the support of Z, and h is the distance between knots. The estimates from the kernel method is obtained by first estimate $\tilde{\gamma}$ by

$$\tilde{\gamma} = \left[(W - \tilde{W})'(W - \tilde{W})\right]^{-1}(W - \tilde{W})'(Y - \tilde{Y}) \tag{2.19}$$

where $\tilde{W}$ and $\tilde{Y}$ are the kernel estimation of $E[W|X, Z]$ and $E[Y|X, Z]$ using the plug-in bandwidths $\tilde{h}_x = x_{s.d.}n^{-1/5}$ and $\tilde{h}_z = z_{s.d.}n^{-1/5}$, where $x_{s.d.}$ and $z_{s.d.}$ are the standard deviations of X and Z respectively. Then I estimated $\tilde{\beta}(z)$ by using the method of Li et al. (2002) replacing Y_i in their formulae with $Y_i - W_i\tilde{\gamma}$ and included the constant term into the matrix X, i.e. $\mathcal{X} = (1, X)$. So the estimate $\hat{\gamma}$ is now obtained by regressing $Y_i - \mathcal{X}_i\tilde{\beta}(Z_i)$ on W_i and the estimate $\hat{g}_i = W_i'\hat{\gamma} + \mathcal{X}_i'\hat{\beta}(Z_i)$ where $\hat{\beta}(Z_i)$ is estimated by using the method of Li et al. (2002) replacing Y_i in their formulae with $Y_i - W_i\hat{\gamma}$ and X_i with $\mathcal{X}_i$.

The second data generating process (DGP2) based on the true regression,

$$Y_i = 4 + 0.5W_i + X_{i1} + X_{i1}\cdot 24Z_i^3\exp(-24Z_i) + X_{i2}Z_i + X_{i2}\sin(Z_i) + u_i \tag{2.20}$$

where $\{u_i\}_{i=1}^{n}$ is i.i.d. vector of normal random variables with mean 0 and variance

0.25, $\{Z_i\}_{i=1}^n$ is generated by the i.i.d. uniform[0,2] vector of random variables, $W = V_1 + 2V_3$ and $X_1 = V_2 + V_3$, and $X_2 = V_4 + 0.5V_3$, where V_j, $j = 1, 2, 3, 4$, are i.i.d. uniform[0,2] vector of random variables. Now let $g_i = 4 + 0.5W_i + X_{i1} + X_{i1} \cdot 24Z_i^3 \exp(-24Z_i) + X_{i2}Z_i + X_{i2}\sin(Z_i)$ be the undisturbed value of the function. The estimation method, the sample size and the number of iteration are the same as I did in DGP1 but replacing X with (X_1, X_2) and $\mathcal{X}$=$(1, X_1, X_2)$ in the kernel method. The simulation results are presented in Table I.

Table I. Simulation Results

		DGP1		DGP2	
		$n = 100$	$n = 200$	$n = 100$	$n = 200$
B-Spline	$AMSE(\hat{\gamma})$	0.0027831	0.0013348	0.0035712	0.0015268
	$AMSE(\hat{g})$	0.0339976	0.0208716	0.0449357	0.0268447
Power Series	$AMSE(\hat{\gamma})$	0.0028169	0.0013550	0.0036243	0.0015515
	$AMSE(\hat{g})$	0.0357869	0.0243554	0.0456148	0.0295840
Kernel	$AMSE(\hat{\gamma})$	0.0046368	0.0020657	0.0071594	0.0030235
	$AMSE(\hat{g})$	0.0884829	0.0644006	0.1152154	0.0857579

From Table I, one can see that the B-Spline method gives the smallest AMSE of both $\hat{\gamma}$ and $\hat{g}$ for every sample size and for both DGPs. Also the power seires method has smaller AMSE of both $\hat{\gamma}$ and $\hat{g}$ for every sample size and for both DGPs than the kernel method. Hence, this simulation shows that one can get the more effcient estimation using series estimation and the simulation results fit with the theory.

D. Application

The application of this model is illustrated by the case of inflation rate forecasting using the unemployment rate and the industry capacity utilization rate. In the macroeconomic literature, the inflation rate and the unemployment rate have a nonlinear relationship known as the Phillips curve. This section will propose the method to forecast the future inflation rate from the past unemployment rate and these two indicators. The data series run from January 1967 to April 2004 and were downloaded from the St. Louis Federal Reserve website.[3]

The forecasting efficiency will be compared using many different models such as the simple autoregressive model (AR model), the smooth coefficient model (SC model), and a semiparametric partially linear smooth coefficient model (PLSC model). For the parametric model analysis, I use the SIC to choose the number of lags in the AR model, and calculate out-of-sample mean squared prediction error (MSPE) for this model (for 100 forecasts). Since the SIC choose 2 lags in parametric model, hence I will also use 2 lags in the semiparametric models. Therefore, consider the following models:

$$\begin{aligned}
\text{AR model:} \quad \pi_t &= \alpha_0 + \alpha_1 \pi_{t-s} + \alpha_2 \pi_{t-s-1} + \alpha_3 u_{t-s} + \alpha_4 z_{t-s} + \epsilon_t \\
\text{PLSC model:} \quad \pi_t &= \alpha_0(z_{t-s}) + \alpha_1 \pi_{t-s} + \alpha_2 \pi_{t-s-1} + \beta(z_{t-s}) u_{t-s} + \epsilon_t \\
\text{SC model:} \quad \pi_t &= \alpha_0(z_{t-s}) + \alpha_1(z_{t-s}) \pi_{t-s} + \alpha_2(z_{t-s}) \pi_{t-s-1} + \beta(z_{t-s}) u_{t-s} + \epsilon_t
\end{aligned}$$

where $s = 1, 6, 12, 24$, π_t is the inflation rate at time t, u_t is the unemployment rate at time t, z_t is the industry capacity utilization rate at time t, and the ϵ_t is the error term at time t satisfies all the assumptions in section B.

---

[3]http://www.research.stlouisfed.org/fred/

Each of these models are estimated and calculated the mean squared prediction errors from the formulae $MSPE = \frac{1}{100}\sum_{t=349}^{448}(\pi_t - \hat{\pi}_t)^2$, where the $\hat{\pi}_t$ is the inflation prediction using the data up to time $t-1$. The results show that the MSPEs of the semiparametric models are smaller than the MSPEs of the parametric models. To identify the model, I also do the following in-sample specification tests for $s = 1, 6, 12, 24$.

Test 1:

H_0: $\pi_t = \alpha_0 + \alpha_1\pi_{t-s} + \alpha_2\pi_{t-s-1} + \alpha_3 u_{t-s} + \alpha_4 z_{t-s} + \epsilon_t$ almost surely

H_1: $\pi_t = \alpha_0(z_{t-s}) + \alpha_1(z_{t-s})\pi_{t-s} + \alpha_2(z_{t-s})\pi_{t-s-1} + \beta(z_{t-s})u_{t-s} + \epsilon_t$, $\alpha_0(z_{t-s}) \neq \alpha_0, \alpha_1(z_{t-s}) \neq \alpha_1$ and $\alpha_2(z_{t-s}) \neq \alpha_2$ almost surely

Test 2:

H_0: $\pi_t = \alpha_0(z_{t-s}) + \alpha_1\pi_{t-s} + \alpha_2\pi_{t-s-1} + \beta(z_{t-s})u_{t-s} + \epsilon_t$ almost surely

H_1: $\pi_t = \alpha_0(z_{t-s}) + \alpha_1(z_{t-s})\pi_{t-s} + \alpha_2(z_{t-s})\pi_{t-s-1} + \beta(z_{t-s})u_{t-s} + \epsilon_t$, $\alpha_1(z_{t-s}) \neq \alpha_1$ and $\alpha_2(z_{t-s}) \neq \alpha_2$ almost surely

The test statistic [4] is

$$C_n = \frac{n^{-1}\sum_{t=1}^{n}(\hat{\pi}_t^{H_0} - \hat{\pi}_t^{H_1})^2}{\hat{\sigma}^2} \tag{2.21}$$

where $\hat{\sigma}^2$ is the estimator of the error variance under H_1. The 95% confidence interval of this test statistic is generated by wild bootstrap method. The test results are shown in table II.

From Table II, one can conclude from Test 1 that, with 95% confidence, the AR model is misspecified and, from Test 2, one cannot reject the semiparametric partially linear smooth coefficient model. These results show that the semiparametric partially linear smooth coefficient model is appropriate to use in inflation rate prediction using

[4]See Hart (1997) for details of this test.

Table II. The Specification Tests between Autoregressive Model, Smooth Coefficient Model and Partially Linear Smooth Coefficient Model

Horizon	1 Month	6 Months	12 Months	24 Months
Test 1	Reject H_0	Reject H_0	Reject H_0	Reject H_0
Test 2	Not Reject H_0	Not Reject H_0	Not Reject H_0	Not Reject H_0

the unemployment rate and the industry capacity utilization rate.

E. Conclusion

In this chapter I propose using a general series method to estimate the semiparametric partially linear smooth coefficient model. I show that the estimator $\hat{\gamma}$ has the $\sqrt{n}$-normality property and it attains the semiparametric efficiency bound when the error is homoskedastic. I also show that it is easy to modify the method to the hetroskedastic error case and that the estimator is still efficient under the modification. The convergence rate of the smooth coefficient function is proven and the primitive conditions of the power series and B-spline are restated here as examples.

Where the Monte Carlo simulation is also conducted, three different methods of estimation are used. I find that for every sample size and every data generating process used in this simulation, the series estimations perform better than the kernel method. In particular, the B-spline method performs best and the kernel method is the last.

Finally the application of the semiparametic partially linear smooth coefficient model is illustrated by applying it to forecasting the inflation rate using the unemployment rate and the industry capacity utilization rate. The specification test results show that the semiparametric partially linear smooth coefficient model is more appropriate than the full smooth coefficient model and the parametric autoregressive model.

F. Proofs

Throughout this section, let C denote a generic constant that may be different in different uses and $\sum_i = \sum_{i=1}^n$. The norm $\|\cdot\|$ for a matrix A is defined by $\|A\| = [tr(A'A)]^{1/2}$.

Proof of Proposition II.1. First consider the case that $\theta(X_i, Z_i)$ is scalar. If f and g are functions from $\mathcal{S}$ to $\mathbb{R}$ such that $E[f^2]$ and $E[g^2]$ are both finite, define the inner product $\langle f, g\rangle = E[fg]$. Hence the class $\mathcal{G}$ defined in Definition 1 is a Hilbert space. I need to show that there exists $m(x,z)$ that satisfies

$$E\left[(\theta(X_i, Z_i) - m(X_i, Z_i))^2\right] = \inf_{h \in L_2} E\left[\left(\theta(X_i, Z_i) - \sum_{l=1}^{d} X_{il} h_l(Z_i)\right)^2\right] \tag{2.22}$$

If $\theta \in \mathcal{G}$ then the solution $m = \theta$ is obvious, so consider when $\theta \notin \mathcal{G}$. For simplicity, I will prove only $d = 2$, the proof for $d > 2$ follows the same arguments.

Let $\{a_i(z)\}_{i=1}^\infty$ be a complete base functions that can expand any $h_1(z) \in L_2$ and also let $\{b_i(z)\}_{i=1}^\infty$ be a complete base functions that can expand any $h_2(z) \in L_2$. Define, for $i \in \mathbb{N}$, $\varphi_{2i-1} = x_1 a_i(z)$ and $\varphi_{2i} = x_2 b_i(z)$. Hence, $\{\varphi_j(x,z)\}_{j=1}^\infty$ is a complete base funtions that can expand any $g(x,z) = x_1 h_1(z) + x_2 h_2(z) \in \mathcal{G}$. Using Gram-Schmidt orthonormalization procedure to orthonormalize $\{\varphi_j(x,z)\}_{j=1}^\infty$, so I will get the orthonormal basis functions $\{\phi_j(x,z)\}_{j=1}^\infty$. Now define

$$m(x,z) = \sum_{j=1}^{\infty} \phi_j(x,z)\beta_j \tag{2.23}$$

where $\beta_j = \langle \theta(X_i, Z_i), \phi_j(X_i, Z_i)\rangle = E[\theta(X_i, Z_i)\phi_j(X_i, Z_i)]$. Now let $\eta(x,z) =$

$\theta(x,z) - m(x,z)$, substitute into (2.23) I get

$$\theta(x,z) = \sum_{j=1}^{\infty} \phi_j(x,z)\beta_j + \eta(x,z) \tag{2.24}$$

Multiply (2.24) both sides by $\phi_j(x,z)$, take expectation and using the orthonormaility of $\{\phi_j(x,z)\}_{j=1}^{\infty}$, I get

$$E[\eta(X_i,Z_i)\phi_j(X_i,Z_i)] = 0 \qquad \forall j \in \mathbb{N} \tag{2.25}$$

This implies that η is orthogonal to the basis functions $\{\phi_j(x,z)\}_{j=1}^{\infty}$ and hence it is also orthogonal to any function $f \in \mathcal{G}$, i.e.,

$$E[\eta(X_i,Z_i) f(X_i,Z_i)] = 0 \qquad \forall f \in \mathcal{G} \tag{2.26}$$

Now square both sides of (2.24) then take expectation; using the orthonormaility of $\{\phi_j(x,z)\}_{j=1}^{\infty}$ and the fact that $\eta \perp \{\phi_j(x,z)\}_{j=1}^{\infty}$, I get

$$E\left[\theta(X_i,Z_i)^2\right] = \sum_{j=1}^{\infty} \beta_j^2 + E\left[\eta(X_i,Z_i)^2\right] \tag{2.27}$$

Since $E\left[\theta(X_i,Z_i)^2\right] < \infty$, therefore $\sum_{j=1}^{\infty} \beta_j^2$ is also finite. This implies that $\sum_{j=1}^{\infty} \phi_j(x,z)\beta_j$ will converge to the well-defined function in $\mathcal{G}$ and because $\theta = m+\eta$, $m \in \mathcal{G}$ and $\eta \perp \mathcal{G}$, $m(x,z)$ reaches the infimum of (2.22).

Next suppose that $\theta(X_i,Z_i)$ is a $q \times 1$ vector. Let $\theta(X_i,Z_i) = (\theta_1(X_i,Z_i),\ldots,\theta_q(X_i,Z_i))'$ and $m(X_i,Z_i) = (m_1(X_i,Z_i),\ldots,m_q(X_i,Z_i))'$ where, for $s = 1,\ldots,q$, $m_s(x,z)$ is defined as (2.23) with β_j replaced by $\beta_{sj} = E[\theta_s(X_i,Z_i)\phi_j(X_i,Z_i)]$. Define $\eta = \theta - m$ as before, then follows the same arguments as above, it can be easily shown that $(\eta_1, \ldots, \eta_q) = \eta \perp \mathcal{G}$; hence, for any $g \in \mathcal{G}$,

$$\begin{aligned} E\left[(\theta - g)(\theta - g)'\right] &= E\left[(\theta - m + m - g)(\theta - m + m - g)'\right] \\ &= E\left[(\theta - m)(\theta - m)'\right] + E\left[(m - g)(m - g)'\right] \\ &\quad +0 \end{aligned} \tag{2.28}$$

where the 0 comes from the facts that $\eta = \theta - m \perp \mathcal{G}$ and $m - g \in \mathcal{G}$. Since $E\left[(m - g)(m - g)'\right]$ is positive definite and $m \in \mathcal{G}$, (2.28) implies that

$$E\left[(\theta - m)(\theta - m)'\right] = \inf_{g \in \mathcal{G}} E\left[(\theta - g)(\theta - g)'\right]$$

■

Proof of theorem II.1. Recall that $\theta(X_i, Z_i) = E[W_i | X_i, Z_i]$, $v_i = W_i - \theta(X_i, Z_i)$, $\varepsilon_i = W_i - m(X_i, Z_i)$ and $\eta_i = \theta(X_i, Z_i) - m(X_i, Z_i)$. I will use the following shorthand notations: $\theta_i = \theta(X_i, Z_i)$, $g_i = X_i'\beta(Z_i)$, and $m_i = m(X_i, Z_i)$. Hence, $v_i = W_i - \theta_i$, $\varepsilon_i = \theta_i + v_i - m_i$, $\eta_i = \theta_i - m_i$. Finally, the variables without the subscript represent the matrix, e.g. $\theta = (\theta_1, \ldots, \theta_n)'$.

Also recall that for any matrix A with n rows, I define $\tilde{A} = P(P'P)^{-1} P'A$. Apply this definition to θ, m, η, u, v, I get $\tilde{\theta}, \tilde{m}, \tilde{\eta}, \tilde{u}, \tilde{v}$.

Since $W_i = \theta_i + v_i$ and $\theta_i = m_i + \eta_i$, I get $W_i = \eta_i + v_i + m_i$ and $\tilde{W}_i = \tilde{\eta}_i + \tilde{v}_i + \tilde{m}_i$. In matrix notation, $W = \eta + v + m$ and $\tilde{W} = \tilde{\eta} + \tilde{v} + \tilde{m}$. Therefore, I have

$$W - \tilde{W} = \eta + v + (m - \tilde{m}) - \tilde{v} - \tilde{\eta} \tag{2.29}$$

For scalars or column vectors A_i and B_i, I define $S_{A,B} = \frac{1}{n}\sum_i A_i B_i'$ and $S_A = S_{A,A}$. Note that if $S_{W-\tilde{W}}^{-1}$ exists, then from (2.8) and (2.9), I get

$$\sqrt{n}(\hat{\gamma} - \gamma) = S_{W-\tilde{W}}^{-1} \sqrt{n} S_{W-\tilde{W}, G-\tilde{G}+u-\tilde{u}} \tag{2.30}$$

For the first part of the theorem, I will proof the followings: (i) $S_{W-\tilde{W}} = \Phi + o_p(1)$, (ii) $S_{W-\tilde{W},G-\tilde{G}} = o_p(n^{-1/2})$, (iii) $S_{W-\tilde{W},\tilde{u}} = o_p(n^{-1/2})$ and (iv) $\sqrt{n} S_{W-\tilde{W},u} \to N(0,\Omega)$ in distribution.

Proof of (i): Using equation (2.29), I have

$$S_{W-\tilde{W}} = S_{\eta+v+(m-\tilde{m})-\tilde{v}-\tilde{\eta}} = S_{\eta+v} + S_{(m-\tilde{m})-\tilde{v}-\tilde{\eta}} + 2S_{\eta+v,(m-\tilde{m})-\tilde{v}-\tilde{\eta}} \tag{2.31}$$

The first term, $S_{\eta+v} = \frac{1}{n}\sum_i (\eta_i + v_i)(\eta_i + v_i)' = \frac{1}{n}\sum_i \varepsilon_i \varepsilon_i' = \Phi + o_p(1)$ by law of large numbers.

The second term, $S_{(m-\tilde{m})-\tilde{v}-\tilde{\eta}} \leq 3\left(S_{(m-\tilde{m})} + S_{\tilde{v}} + S_{\tilde{\eta}}\right) = o_p(1)$ by lemma II.3, lemma II.4(i), and lemma II.4(iii).

The last term, $S_{\eta+v,(m-\tilde{m})-\tilde{v}-\tilde{\eta}} \leq \left(S_{\eta+v} S_{(m-\tilde{m})-\tilde{v}-\tilde{\eta}}\right)^{1/2} = \left(O_p(1)\, o_p(1)\right)^{1/2} = o_p(1)$ by the preceeding results.

Proof of (ii): Using equation (2.29), I have

$$\begin{aligned} S_{W-\tilde{W},G-\tilde{G}} &= S_{\eta+v+(m-\tilde{m})-\tilde{v}-\tilde{\eta},G-\tilde{G}} \\ &= S_{\eta+v,G-\tilde{G}} + S_{m-\tilde{m},G-\tilde{G}} - S_{\tilde{v},G-\tilde{G}} - S_{\tilde{\eta},G-\tilde{G}} \end{aligned} \tag{2.32}$$

The first term, $S_{\eta+v,G-\tilde{G}} \leq \left(S_{\eta+v} S_{G-\tilde{G}}\right)^{1/2} = O_p\left(\sum_{l=1}^d k_l^{-\delta_l}\right)$ by lemma II.3.

The second term, $S_{m-\tilde{m},G-\tilde{G}} \leq \left(S_{m-\tilde{m}} S_{G-\tilde{G}}\right)^{1/2} = O_p\left(\sum_{l=1}^d k_l^{-2\delta_l}\right)$ by lemma II.3.

The third term, $S_{\tilde{v},G-\tilde{G}} \leq \left(S_{\tilde{v}} S_{G-\tilde{G}}\right)^{1/2} = o_p(1)\, O_p\left(\sum_{l=1}^d k_l^{-\delta_l}\right)$ by lemma II.3 and lemma II.4(i).

The last term, $S_{\tilde{\eta},G-\tilde{G}} \leq \left(S_{\tilde{\eta}} S_{G-\tilde{G}}\right)^{1/2} o_p(1)\, O_p\left(\sum_{l=1}^d k_l^{-\delta_l}\right)$ by lemma II.3 and lemma II.4(iii).

Combine all four terms, I get $S_{W-\tilde{W},G-\tilde{G}} = o_p(n^{-1/2})$ by assumption II.3.

Proof of (iii): Using equation (2.29), I have

$$S_{W-\tilde{W},\tilde{u}} = S_{\eta+v+(m-\tilde{m})-\tilde{v}-\tilde{\eta},\tilde{u}} = S_{\eta+v,\tilde{u}} + S_{m-\tilde{m},\tilde{u}} - S_{\tilde{v},\tilde{u}} - S_{\tilde{\eta},\tilde{u}} \qquad (2.33)$$

The first term, $S_{\eta+v,\tilde{u}} \leq (S_{\eta+v}S_{\tilde{u}})^{1/2} = O_p\left(\sqrt{K}/\sqrt{n}\right)$ by lemma II.4(ii).

The second term, $S_{m-\tilde{m},\tilde{u}} \leq (S_{m-\tilde{m}}S_{\tilde{u}})^{1/2} = O_p\left(\sum_{l=1}^{d} k_l^{-\delta_l}\right) O_p\left(\sqrt{K}/\sqrt{n}\right)$ by lemma II.3 and lemma II.4(ii).

The third term, $S_{\tilde{v},\tilde{u}} \leq (S_{\tilde{v}}S_{\tilde{u}})^{1/2} = O_p(K/n)$ by lemma II.4(i) and lemma II.4(ii).

The last term, $S_{\tilde{\eta},\tilde{u}} \leq (S_{\tilde{\eta}}S_{\tilde{u}})^{1/2} = O_p(K/n)$ by lemma II.4(ii) and lemma II.4(iii).

Combine all four terms, I get $S_{W-\tilde{W},\tilde{u}} = o_p\left(n^{-1/2}\right)$ by assumption II.3.

Proof of (iv): Using equation (2.29), I have

$$\begin{aligned} \sqrt{n}S_{W-\tilde{W},u} &= \sqrt{n}S_{\eta+v+(m-\tilde{m})-\tilde{v}-\tilde{\eta},u} \\ &= \sqrt{n}S_{\eta+v,u} + \sqrt{n}\left(S_{m-\tilde{m},u} - S_{\tilde{v},u} - S_{\tilde{\eta},u}\right) \qquad (2.34) \end{aligned}$$

The first term, $\sqrt{n}S_{\eta+v,u} = \sqrt{n}\sum_{i=1}^{n}(\eta_i + v_i)u_i = \sqrt{n}\sum_{i=1}^{n}\varepsilon_i u_i \to N(0,\Omega)$ in distribution by the central limit theorem.

The second term, $E\left[S_{m-\tilde{m},u}^2|X,Z\right] = \frac{1}{n^2}tr\left((m-\tilde{m})(m-\tilde{m})' E\left[uu'|X,Z\right]\right) \leq \frac{C}{n}tr\left((m-\tilde{m})'(m-\tilde{m})/n\right) = \frac{C}{n}S_{m-\tilde{m}} = o_p(n^{-1})$ by lemma II.3. Hence, $S_{m-\tilde{m},u} = o_p\left(n^{-1/2}\right)$.

The third term, $E\left[S_{\tilde{v},u}^2|X,Z\right] = \frac{1}{n^2}tr\left(P(P'P)^{-1}P'vv'P(P'P)^{-1}P'E\left[uu'|X,Z\right]\right) \leq \frac{C}{n^2}tr\left(P(P'P)^{-1}P'vv'P(P'P)^{-1}P'\right) = \frac{C}{n}tr(\tilde{v}\tilde{v}'/n) = \frac{C}{n}S_{\tilde{v}} = o_p(n^{-1})$ by lemma II.4(i). Hence, $S_{\tilde{v},u} = o_p\left(n^{-1/2}\right)$.

The last term, similar to the proof of $S_{\tilde{v},u}$ but instead use lemma 4(iii). Hence $S_{\tilde{\eta},u} = o_p\left(n^{-1/2}\right)$.

Combine the proof of (i), (ii), (iii) and (iv) with (2.30), I can conclude that $\sqrt{n}(\hat{\gamma} - \gamma) \to N(0, \Phi^{-1}\Omega\Phi^{-1})$ in distribution.

For the second part of the proof, I need to show that $\hat{\Sigma} = \Sigma + o_p(1)$, where $\hat{\Sigma} = \hat{\Phi}^{-1}\hat{\Omega}\hat{\Phi}^{-1}$. Since $\hat{\Phi} = S_{W-\tilde{W}}$, so by the proof in theorem II.1, $\hat{\Phi} = \Phi + o_p(1)$. The remaining of the proof is to show that $\hat{\Omega} = \Omega + o_p(1)$.

By the proof of the first part of theorem II.1, I have $\hat{\gamma} = \gamma + O_p\left(n^{-1/2}\right)$; and from the proof of theorem II.2 next, I will have $\hat{g}_i = g_i + o_p(1)$. Hence, $\hat{u}_i = Y_i - W_i'\hat{\gamma} - \hat{g}_i = (Y_i - W_i'\gamma - g_i) + o_p(1) = u_i + o_p(1)$. By the lemma II.3, lemma II.4(i) and lemma II.4(iii), I have $m_i - \tilde{m}_i = o_p(1)$, $\tilde{v}_i = o_p(1)$ and $\tilde{\eta}_i = o_p(1)$. Using equation (2.29), I have

$$W_i - \tilde{W}_i = \eta_i + v_i + (m_i - \tilde{m}_i) - \tilde{v}_i - \tilde{\eta}_i = \varepsilon_i + o_p(1) \tag{2.35}$$

Therefore, $\hat{\Omega} = \frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 \left(W_i - \tilde{W}_i\right)\left(W_i - \tilde{W}_i\right)' = \frac{1}{n}\sum_{i=1}^{n} u_i^2 \varepsilon_i \varepsilon_i' + o_p(1) = \Omega + o_p(1)$ by the law of large numbers. ∎

Proof of theorem II.2. The proof of this theorem is similar to the proof of Theorem 1 in Newey(1997). First find the convergence rate of $\mathbf{1}_n \|\hat{\alpha} - \alpha\|$, by (2.10) and (2.6), I get

$$\begin{aligned}
\hat{\alpha} &= (P'P)^{-1} P'(Y - W\hat{\gamma}) \\
&= (P'P)^{-1} P'(Y - W\gamma - W(\hat{\gamma} - \gamma)) \\
&= (P'P)^{-1} P'(P\alpha + (G - P\alpha) + u - W(\hat{\gamma} - \gamma)) \\
&= \alpha + (P'P/n)^{-1} P'(G - P\alpha)/n + (P'P/n)^{-1} P'u/n \\
&\quad - (P'P/n)^{-1} P'W(\hat{\gamma} - \gamma)/n
\end{aligned} \tag{2.36}$$

Hence,

$$\begin{aligned}
\mathbf{1}_n \|\hat{\alpha} - \alpha\| \leq\ & \mathbf{1}_n \left\|(P'P/n)^{-1} P'(G - P\alpha)/n\right\| \\
& + \mathbf{1}_n \left\|(P'P/n)^{-1} P'u/n\right\| \\
& + \mathbf{1}_n \left\|(P'P/n)^{-1} P'W(\hat{\gamma} - \gamma)/n\right\|
\end{aligned} \tag{2.37}$$

The first term, $\mathbf{1}_n \left\|(P'P/n)^{-1} P'(G-P\alpha)/n\right\| = O_p\left(\sum_{l=1}^{d} k_l^{-\delta_l}\right)$ by lemma II.2.

The second term, $E\left[\mathbf{1}_n \left\|(P'P/n)^{-1} P'u/n\right\| |X,Z\right]$

$= \mathbf{1}_n E\left[\left((u'P/n)(P'P/n)^{-1}(P'P/n)^{-1}(P'u/n)\right)^{1/2} |X,Z\right]$

$\leq O_p(1)\,\mathbf{1}_n tr\left(P(P'P)^{-1}P'E[uu'|X,Z]/n\right)^{1/2} \leq O_p(1)\,\mathbf{1}_n C\sqrt{K}/\sqrt{n}$ by lemma II.1 and assumption II.1. Hence, $\mathbf{1}_n \left\|(P'P/n)^{-1} P'u/n\right\| = O_p\left(\sqrt{K}/\sqrt{n}\right)$.

The last term, note that $W = \eta + v + m = \varepsilon + m$ and $\hat{\gamma} - \gamma = O_p\left(n^{-1/2}\right)$ by theorem II.1, so

$E\left[\mathbf{1}_n \left\|(P'P/n)^{-1} P'W/n\right\| |X,Z\right] = \mathbf{1}_n E\left[\left\|(P'P/n)^{-1} P'(\varepsilon+m)/n\right\| |X,Z\right]$

$\leq \mathbf{1}_n E\left[\left\|(P'P/n)^{-1} P'\varepsilon/n\right\| |X,Z\right] + \mathbf{1}_n E\left[\left\|(P'P/n)^{-1} P'm/n\right\| |X,Z\right]$.

Since, $\mathbf{1}_n E\left[\left\|(P'P/n)^{-1} P'\varepsilon/n\right\| |X,Z\right]$

$= \mathbf{1}_n E\left[\left\|(\varepsilon'P/n)(P'P/n)^{-1}(P'P/n)^{-1}(P'\varepsilon/n)\right\| |X,Z\right]$

$\leq O_p(1)\,\mathbf{1}_n tr\left(P(P'P)^{-1}P'E[\varepsilon\varepsilon'|X,Z]/n\right)^{1/2} \leq O_p(1)\,\mathbf{1}_n C\sqrt{K}/\sqrt{n}$ by lemma II.1 and the proof in theorem II.1, hence, $\mathbf{1}_n \left\|(P'P/n)^{-1} P'\varepsilon/n\right\| = O_p\left(\sqrt{K}/\sqrt{n}\right) = o_p(1)$.

And, $\mathbf{1}_n \left\|(P'P)^{-1} P'm\right\| = \mathbf{1}_n \|\tilde{\alpha}_m\| = \mathbf{1}_n \|\tilde{\alpha}_m - \alpha_m + \alpha_m\|$

$\leq \mathbf{1}_n \|\tilde{\alpha}_m - \alpha_m\| + \mathbf{1}_n \|\alpha_m\| = O_p\left(\sum_{l=1}^{d} k_l^{-\delta_l}\right) + O_p(1) = o_p\left(n^{-1/2}\right) + O_p(1) = O_p(1)$, by lemma II.2.

Combine all the terms, I have

$$\mathbf{1}_n \|\hat{\alpha} - \alpha\| = O_p\left(\sum_{l=1}^{d} k_l^{-\delta_l} + \sqrt{K}/\sqrt{n}\right) \tag{2.38}$$

To prove part (i), using (2.38) and assumption II.3, I have

$$\begin{aligned}
&\sup_{(x,z)\in S} |\hat{g}(x,z) - g(x,z)| \\
\leq\; &\sup_{(x,z)\in S} \left|p^K(x,z)'(\hat{\alpha}-\alpha)\right| + \left|p^K(x,z)'\alpha - g(x,z)\right| \\
\leq\; &\zeta_0(K)\|\hat{\alpha}-\alpha\| + O\left(\sum_{l=1}^{d} k_l^{-\delta_l}\right) \\
=\; &O_p\left(\zeta_0(K)\left(\sum_{l=1}^{d} k_l^{-\delta_l} + \sqrt{K}/\sqrt{n}\right)\right)
\end{aligned}$$

Since the proof of (ii) are similar to the proof of (iii), so I will prove only part (iii). Using (2.38) and assumption II.3, I have

$$\begin{aligned}
&\int_S (\hat{g}(x,z) - g(x,z))^2 dF(x,z) \\
=\; &\int_S \left(p^K(x,z)'(\hat{\alpha}-\alpha) + p^K(x,z)'\alpha - g(x,z)\right)^2 dF(x,z) \\
\leq\; &\|\hat{\alpha}-\alpha\|^2 + \int_S \left(p^K(x,z)'\alpha - g(x,z)\right)^2 dF(x,z) \\
=\; &O_p\left(K/n + \sum_{l=1}^{d} k_l^{-2\delta_l}\right) + O_p\left(\sum_{l=1}^{d} k_l^{-2\delta_l}\right) \\
=\; &O_p\left(K/n + \sum_{l=1}^{d} k_l^{-2\delta_l}\right)
\end{aligned}$$

■

The followings are the proofs of the lemmas that are used in the proofs of the theorems. Following the arguments in Newey (1997), I can assume without loss of generality that $B = I$, hence $P^K(X,Z) = p^K(X,Z)$, and $Q = E\left[p^K(X_i,Z_i)\,p^K(X_i,Z_i)'\right] = I$, see Newey (1997) for the reasons and more discussions on these issues. Recall that $p^K(X,Z)$ is a $K \times 1$ matrix, rewrite each component of this matrix as $p^K(X,Z) = (p_{1K}(X,Z), \ldots, p_{KK}(X,Z))'$.

Lemma II.1 $\left\|\hat{Q} - I\right\| = O_p\left(\zeta_0(K)\sqrt{K}/\sqrt{n}\right) = o_p(1)$, *where* $\hat{Q} = P'P/n$.

Proof. This is the proof of Theorem 1 in Newey (1997). The proof is copied here for the reference.

$$\begin{aligned}
E\left[\left\|\hat{Q}-I\right\|^2\right] &= \sum_{k=1}^{K}\sum_{j=1}^{K} E\left[\left(\sum_{i=1}^{n} p_{kK}(X_i,Z_i)\, p_{jK}(X_i,Z_i)/n - I_{jk}\right)^2\right] \\
&\leq \sum_{k=1}^{K}\sum_{j=1}^{K} E\left[p_{kK}(X_i,Z_i)^2\, p_{jK}(X_i,Z_i)^2\right]/n \\
&= E\left[\sum_{k=1}^{K} p_{kK}(X_i,Z_i)^2 \sum_{j=1}^{K} p_{jK}(X_i,Z_i)^2\right]/n \\
&\leq \zeta_0(K)^2\, tr(I)/n \\
&\leq \zeta_0(K)^2\, K/n \to 0
\end{aligned}$$

Hence,

$$\left\|\hat{Q}-I\right\| = O_p\left(\zeta_0(K)\sqrt{K}/\sqrt{n}\right) = o_p(1)$$

■

Let $\mathbf{1}_n$ be an indicator function which has value 1 if $P'P$ is invertible and 0 otherwise. Note that by Lemma II.1, $P(\mathbf{1}_n = 1) \to 1$.

Lemma II.2 $\|\tilde{\alpha}_f - \alpha_f\| = O_p\left(\sum_{l=1}^{d} k_l^{-\delta_l}\right)$, *where* $\tilde{\alpha}_f = (P'P)^{-1} P'f$, α_f *satisfies assumption II.3,* $f = G$ *or* $f = m$.

Proof. By lemma II.1, assumption II.3 and the fact that $P(P'P)^{-1}P'$ is idempotent,

$$\begin{aligned}
\mathbf{1}_n \|\tilde{\alpha}_f - \alpha_f\| &= \mathbf{1}_n \left\|(P'P)^{-1} P'(f - P\alpha_f)\right\| \\
&= \mathbf{1}_n \left\|(f-P\alpha_f)' P (P'P)^{-1} \hat{Q} P' (f - P\alpha_f)/n\right\|^{1/2} \\
&\leq \mathbf{1}_n O_p(1) \left\|(f-P\alpha_f)' P (P'P)^{-1} P' (f - P\alpha_f)/n\right\|^{1/2} \\
&\leq O_p(1) \left\|(f-P\alpha_f)'(f-P\alpha_f)/n\right\|^{1/2} = O_p\left(\sum_{l=1}^{d} k_l^{-\delta_l}\right)
\end{aligned}$$

Since, $P(\mathbf{1}_n = 1) \to 1$, $\|\tilde{\alpha}_f - \alpha_f\| = O_p\left(\sum_{l=1}^d k_l^{-\delta_l}\right)$. ■

Lemma II.3 *$S_{f-\tilde{f}} = O_p\left(\sum_{l=1}^d k_l^{-2\delta_l}\right)$, where $f = G$ or $f = m$.*

Proof. Note that $\tilde{f} = P\tilde{\alpha}_f$, by assumption II.3, lemma II.1 and lemma II.2,

$$\begin{aligned} S_{f-\tilde{f}} &= \frac{1}{n}\left\|f - \tilde{f}\right\|^2 \le \frac{1}{n}\left(\|f - P\alpha_f\|^2 + \|P(\alpha_f - \tilde{\alpha}_f)\|^2\right) \\ &= O\left(\sum_{l=1}^d k_l^{-2\delta_l}\right) + (\alpha_f - \tilde{\alpha}_f)'(P'P/n)(\alpha_f - \tilde{\alpha}_f) \\ &\le O\left(\sum_{l=1}^d k_l^{-2\delta_l}\right) + O_p(1)\|\alpha_f - \tilde{\alpha}_f\|^2 = O_p\left(\sum_{l=1}^d k_l^{-2\delta_l}\right) \end{aligned}$$

■

Lemma II.4 *(i) $S_{\tilde{v}} = O_p(K/n)$, (ii) $S_{\tilde{u}} = O_p(K/n)$ and (iii) $S_{\tilde{\eta}} = O_p(K/n)$.*

Proof. (i) This proof is similar to the proof of Theorem 1 of Newey (1997),

$$\begin{aligned} E[S_{\tilde{v}}|X, Z] &= \frac{1}{n}E\left[v'P(P'P)^{-1}P'v|X, Z\right] \\ &= \frac{1}{n}E\left[tr\left(P(P'P)^{-1}P'E[vv'|X, Z]\right)\right] \\ &\le \frac{C}{n}tr\left(P(P'P)^{-1}P'\right) = C(K/n) \end{aligned}$$

Hence, $S_{\tilde{v}} = O_p(K/n)$.

(ii) Follow the same proof as in the proof of lemma II.4(i).

(iii) Note that since $E[\theta(X,Z)\theta(X,Z)'] < \infty$, $\theta(x,z) = m(x,z) + \eta(x,z)$, $m \in \mathcal{G}$ and $\eta \perp \mathcal{G}$, these imply that $E[\eta\eta'|X,Z] < \infty$. So I can also apply the proof of lemma II.4(i) to this case and get the same result. ■

CHAPTER III

THE RELATIONSHIP BETWEEN MONEY GROWTH AND INFLATION IN THE UNITED STATES

A. Introduction

One of the central propositions in economics is that there is a relationship between the inflation rate and the growth rate of the money supply. This relationship has been the basis of a number of policies that have been proposed for the US economy, including the use of money growth as an indicator variable for inflation.[1] Friedman (1988) summarized the evidence on the use of money growth as an indicator variable for inflation, and concluded that there is no stable empirical relationship between the two variables. Friedman and Kuttner (1992) presented extensive evidence confirming that conclusion.[2] Svensson and Woodford (2002) have summarized the existing empirical literature, "Under normal circumstances, the information content of money growth for inflation forecasts in the short and medium term seems to be quite low. Only in the long run does a high correlation between money growth and inflation result."

This chapter asks whether the dismissal of money growth as an indicator for inflation is warranted. Specifically, empirical research has focused on forecasts made using vector autoregressive (VAR) models:

$$X_t = \alpha + \beta(L)X_{t-1} + \epsilon_t \tag{3.1}$$

[1]See, among many others, Bernanke and Mishkin (1997), Rudebusch and Svensson (1999) and Svensson and Woodford (2002) for discussion of policies making use of indicator variables. Money growth targeting is another example of using the relationship between money growth and inflation in a policy rule. See e.g. Friedman (1988) and Friedman and Kuttner (1996) for further discussion.

[2]See e.g. Cecchetti (1995), Friedman and Kuttner (1996), Estrella and Mishkin (1997) and Stock and Watson (1999) for additional evidence that money growth does not have marginal predictive content for inflation.

where $X_t = (\pi_t, z_t)'$, π_t is the inflation rate at time t, and z_t is either the growth rate of a monetary aggregate or a "price gap" measure such as velocity (see e.g. Hallman, Porter and Small (1991) and Gerlach and Svensson (2001)). The popularity of VAR modeling arises from the fact that it is an atheoretical approach, requiring very few assumptions. Based on the Wold Decomposition Theorem, most time series can be assumed to have a representation that can be approximated as an autoregressive process (see e.g. Sargent (1987)). If the goal is to produce out-of-sample forecasts, however, VAR models may not be the best choice because they are often overparametrized. If the system of interest is nonlinear, there may be gains from using a nonlinear forecasting model rather than a linear approximation.

This chapter allows for potentially complicated nonlinearities in the relationship between money growth and inflation. The approach is to ask whether a nonparametric model, that includes money growth, forecasts inflation better out-of-sample than an autoregressive model of inflation. In principle, a nonparametric approach should be preferred to parametric approaches, because it is more general. As few other macroeconomic papers have attempted to exploit the gains from nonparametric modeling[3], it is worth discussing some reasons why this may be so. First, the curse of dimensionality requires that only very parsimonious models be considered. The computational burden of implementing a nonparametric approach is non-trivial, requiring infeasible amounts of computing time for models with more than four or five independent variables. On the other hand, it is possible to estimate even the most complicated linear models in just a few seconds. A second disadvantage is that nonparametric models are less efficient when the data generating process can be approximated well with a linear model. The assumption of linearity is prevalent in empirical macroeconomics,

[3]See e.g. Diebold and Nason (1990).

presumably because of a lack of evidence to the contrary.[4] Evidence that a nonparametric inflation forecasting model outperforms an autoregressive model provides a strong case for policymakers to place weight on money growth when setting policy.

This chapter proceeds as follows. Section B describes the data. Section C discusses the findings on the information content of monetary aggregates and velocity for inflation. Section D is the conclusion. Finally, section E discusses the computational details.

B. Data

This section describes the variables that are used in the empirical analysis. All of the data series run from January 1959 to June 2002 , and were downloaded from the St. Louis Federal Reserve website.[5] The monetary aggregate data are simple sum M2, and M3, and corresponding M2, and M3 Divisia monetary services index data (see e.g. Belongia and Chalfant (1989), Belongia (1996) and the collection of papers in Barnett and Serletis (2000) for discussion and empirical evidence on the advantages and disadvantages of using Divisia monetary aggregates).

The velocity data are computed as $V = PQ/M$, where P is the seasonally adjusted consumer price index for all urban consumers, Q is the index of total industrial production, and M is one of the monetary aggregates. The consumer price index and industrial production series are used in place of other possible measures of the price level and output, because these variables are available at a monthly frequency. Use of data available at a quarterly frequency would prohibit the use of nonparametric methods, as the sample size would be too small to allow for an informative post-sample

[4]For exceptions, see Michael, Nobay and Peel (1997), Taylor (2001) and Hamilton (2001).

[5]http://www.research.stlouisfed.org/fred/

analysis.

C. Results

1. Parametric Forecasting Models

For the parametric model analysis, I use the SIC to choose the number of lags in the AR model, and calculate out-of-sample mean squared prediction error (MSPE) for this model (for 100 forecasts). Then I need to use the SIC to choose the number of lags in the bivariate VAR model of inflation and money growth, and make forecasts using that model.

All the forecasting relative MSPE reported in this chapter are relative to the MSPE of a simple linear AR(2) model given below (SIC selects 2 lags):

$$\pi_t = \alpha_0 + \alpha_1 \pi_{t-s} + \alpha_2 \pi_{t-s-1} + u_t, \tag{3.2}$$

where $s = 1, 6, 12, 24$.

The data is from January 1959 to June 2002. I make one-step-ahead forecasting for the last 100 periods. Let $\hat{\pi}_t$ denote the one-step-ahead forecasted value of π_t ($t = 1, \ldots, 100$, corresponds to the last 100 periods), the MSPE is computed as $\frac{1}{100}\sum_{t=1}^{100}(\hat{\pi}_t - \pi_t)^2$. Based on model (3.2), the 1,6,12 and 24 month ahead forcasting MSPEs are: 3.52, 3.69, 4.12 and 4.48, respectively. Thus based on an AR(2) model, the longer the forecasting horizon, the worse the forecasting.

Applying SIC to the VAR model also picks 2-lag, therefore, the VAR model for inflation is as follows:

$$\pi_t = \alpha_0 + \alpha_1 \pi_{t-s} + \alpha_2 \pi_{t-s-1} + \alpha_3 z_{t-s} + \alpha_4 z_{t-s-1} + u_t. \tag{3.3}$$

The relative forecasting MSPE of the parametric linear VAR models are the ratio of MSPEs of the VAR models to MSPEs of the simple AR(2) model of (3.2). The results are given in Table III below.

Table III. Relative MSPE of Linear Models

Horizon	M2	M3	M2D	M3D	V2	V3	V2D	V3D
1 Month	.99	.99	1.00	1.01	1.09	1.11	1.01	1.00
6 Months	1.02	1.10	.98	1.03	.93	1.05	.90	.90
12 Months	.93	1.04	.97	1.01	.85	.85	.84	.83
24 Months	.79	1.06	.89	.96	.71	.67	.63	.62

From Table III one observes that, for 1-month-ahead forecasting, out of 8 different money and velocity measures used, only M2 and V3D have slightly smaller forecasting errors than the simple AR model. Moreover, compared with the simple AR(2) model, the best parametric models with money or velocity (M2 and M3) have no more than 1% reduction in MSPE. This clearly shows that, in a linear regression model framework, adding money or velocity as additional regressors do not help forecasting 1-month-ahead inflation. For 6-month-ahead forecasting, half of the models with money or velocity preform better than the AR(2) model. For the longer horizon of 12-month-ahead and 24-month-ahead forecastings, 6 of 8 models with money or velocity give smaller MSPE than the AR(2) model. Therefore, the estimation results based on the linear models suggest that money may help forecasting inflation only in the long run (1 year or more), but it does not help forecasting inflation in the short-term (1 or 6 months).

The above results are consistent with the previous literature that various economic indicators may help predicting inflation only in the long run. There is a strong correlation between inflation and money growth in the long run but not in the short run (Cecchetti(1995)). I also find that when the horizon is longer, the R-squared is smaller. This result shows that there is no relationship between the ranking of in-sample goodness-of-fit and the ranking of out-of-sample forecasting, this reinforces the finding of Cecchetti (1995).

2. Nonlinearity in the Money Growth-Inflation Relationship

In this subsection I use a nonparametric approach to model the relationship between inflation and money growth rate. I compute the relative MSPE from the nonparametric model to the simple linear AR(2) model. Comparison of the linear AR(2) model MSPEs to the fully nonparametric model MSPEs. The linear AR(2) model is given in (3.2). The nonparametric model has the following form ($s = 1, 6, 12, 24$):

$$\pi_t = g(\pi_{t-s}, \pi_{t-s-1}, z_{t-s}) + u_t, \tag{3.4}$$

where $g(.)$ has an unknown functional form, $z_t = m_t$ or $z_t = v_t$, where $m_t = ln(M_t/M_{t-1})$, $v_t = P_tQ_t/M_t$, M_t can be any one of the four monetary measures: $M2_t$, $M3_t$, $M2D_t$, and $M3D_t$.[6]

Table IV reports the relative MSPE of the nonparametric estimation results to those of the simple linear AR(2) model.

From Table IV one observes that for the short-run 1-month and 6-month ahead forecasting, the nonparametric models give much smaller MSPEs than the simple linear AR(2) model for all cases. First note that, compared with the linear AR(2) model, the nonparametric AR(2) model has 13% and 11% reductions in MSPEs for

[6]I have also estimated a nonparametric model with two lags in z_t. The MSPE results are similar to that of model (3.4) with only one lag in z_t. These results are not reported here to save space.

Table IV. Relative MSPE of Nonparametric Regression Models

Horizon	AR	M2	M3	M2D	M3D	V2	V3	V2D	V3D
1 Month	.87	.80	.84	.86	.86	.74	.83	.74	.73
6 Months	.89	.85	.89	.86	.89	.74	.81	.72	.71
12 Months	.90	.80	.89	.88	.89	.62	.69	.61	.61
24 Months	1.06	.71	1.04	.93	.98	.58	.69	.58	.57

one-month and six-month ahead forecastings, respectively. Secondly note that all the models with money or velocity perform better than the nonparametric AR(2) model. The best performers are V2, V2D and V3D, the MSPE reduction, compared with the nonparametric AR(2) model, is around 15% for 1-month-ahead forecasting, and 20% for 6-month-ahead forecasting. Hence, even for the 1-month or 6-month short-run, using money or velocity help improve forecasting inflation.

For the 12 and 24 months ahead forecasting, all the models with money or velocity perform better than the nonparametric AR(2) model. Thus, the nonparametric estimation result shows that using money or velocity can improve forecasting inflation even in the short-run. This suggests that money and velocity affect inflation in a nonlinear way, and the misspecified linear models fail to detect the correlation between money (velocity) and inflation.

3. Identifying the Source of the Nonlinearity

In this subsection I try to identify the source of nonlinearity. I try a parametric model allowing nonlinear interaction terms as well as two semiparametric models to see whether the nonlinear effect of money on inflation can be summarized in a simple way.

a. Linear interaction model

First I consider a linear model with interactions ($s = 1, 6, 12, 24$):

$$\pi_t = \alpha_0 + \alpha_1 \pi_{t-s} + \alpha_2 \pi_{t-s-1} + \alpha_3 z_{t-s} + \alpha_4 \pi_{t-s} z_{t-s} + \alpha_5 \pi_{t-s-1} z_{t-s} + \alpha_6 \pi_{t-s} \pi_{t-s-1} + u_t. \tag{3.5}$$

The out-of-sample MSPE based on model (3.5) is given in Table V. Comparing the result of Table V to that of Table IV, one observes that the relative MSPE of the parametric model (3.5) is larger than those based on the nonparametric model (3.4). This suggests that the simple quadratic interaction terms do not catch the nonlinear money effect on inflation.

Table V. Relative MSPE of Parametric Linear Models with Interactions

Horizon	AR	M2	M3	M2D	M3D	V2	V3	V2D	V3D
1 Month	.90	.95	.89	.93	.92	1.00	.96	.89	.89
6 Months	.98	1.04	1.08	.99	1.00	.89	1.03	.86	.84
12 Months	.99	1.01	1.06	.98	1.00	.68	.99	.68	.66
24 Months	1.03	.83	1.08	.96	.99	.65	.69	.75	.70

b. Partially linear model

Next, I consider a semiparametric partially linear model as follows ($s = 1, 6, 12, 24$).

$$\pi_t = \alpha_1 \pi_{t-s} + \alpha_2 \pi_{t-s-1} + g(z_{t-s}) + u_t, \tag{3.6}$$

where $g(.)$ is of unknown functional form. Although the partially linear model allows nonlinearity in z_{t-s} because $g(.)$ has flexible functional form. It does not allow nonlinearity in π_{t-s} and π_{t-s-1}, nor does it allow interaction among z_{t-s}, π_{t-s} and π_{t-s-1}. The relative MSPE (relative to a linear AR(2) model) for the partially linear model is given in Table VI below.

Table VI. Relative MSPE of Partially Linear Regression Models

Horizon	M2	M3	M2D	M3D	V2	V3	V2D	V3D
1 Month	.99	1.01	1.00	.99	.96	.98	.90	.91
6 Months	.96	1.02	.95	.97	.86	.91	.80	.81
12 Months	.94	1.01	.98	.99	.82	.83	.72	.74
24 Months	.71	.94	.87	.92	.65	.72	.60	.60

From Table VI one observes that the performance of a partially linear model is similar to that of the parametric model (3.5) with interactions. Its MSPEs are in general larger than those from the nonparametric model (3.4).

c. Smooth coefficient model

Finally I consider a more general semiparametric model: the smooth coefficient model which is given by ($s = 1, 6, 12, 24$)

$$\pi_t = \beta_0(z_{t-s}) + \beta_1(z_{t-s})\pi_{t-s} + \beta_2(z_{t-s})\pi_{t-s-1} + u_t, \qquad (3.7)$$

where the functional form of $\beta_0(.)$, $\beta_1(.)$ and $\beta_2(.)$ are not specified. Note that the smooth coefficient model (3.7) contains the partially linear model as a special case. When $\beta_1(z)$ and $\beta_2(z)$ are constant functions, the smooth coefficient model reduces to a partially linear model.

The relative MSPE of the smooth coefficient model is given in Table VII. From Table VII one sees that the smooth coefficient model performs much better than the simple linear AR(2) model. It also performs much better than both the parametric model with interactions, and the partially linear model. In fact there are a few cases where the smooth coefficient model even out-perform the nonparametric model. Compared with the nonparametric model (3.4), the smooth coefficient model has the advantage that it only contains a univariate nonparametric component. Thus, it does not suffer from the 'curse of dimensionality' problem. The smooth coefficient model suggests that the money growth rate affects inflation nonlinerly and that the interaction between lagged inflation and lagged money growth rate are important determinants of future inflation.

Table VII. Relative MSPE of Smooth Coefficient Models

Horizon	M2	M3	M2D	M3D	V2	V3	V2D	V3D
1 Month	1.00	.99	.86	1.00	1.00	1.00	.75	.73
6 Months	.96	.98	.81	.89	.94	.94	.76	.77
12 Months	.93	1.00	.67	.79	.97	.99	.60	.59
24 Months	.71	.90	.60	.67	.88	.92	.60	.58

D. Conclusion

This chapter has made several notable contributions. By not relying on the straight-jacket of a linear VAR forecasting model, I show strong and robust evidence that money growth has been useful as an indicator of future inflation in the US. The finding of a strong and sufficiently stable relationship between money growth and inflation is the opposite of that found in most recent work. In particular, there is a near consensus that the only relationship between money growth and inflation is at long horizons. I have presented evidence that money growth has marginal predictive content for inflation at horizons as short as one month. More generally, the results indicate that nonparametric methods might be useful for the monetary policy process. Nonlinearities in the relationships of interest to central bankers appear to be strong. When modeling inflation, the benefits of relaxing the assumptions on functional form substantially outweigh any disadvantages of nonparametric estimation.

E. Computational Details

Discussion of how each of the nonparametric models was estimated is written for someone who is familiar with all of the technical details of nonparametric estimation.

The general nonparametric models are estimated by using the Nadaraya-Watson kernel type estimator. The bandwidth is choosen by the least squares cross-validation method.

The patially linear model is estimated by using the method purposed by Robinson (1988). First I estimated the parameters in the model by regressing $\pi_t - \hat{E}(\pi_t|z_{t-1})$ on $\pi_{t-s} - \hat{E}(\pi_{t-s}|z_{t-1})$ and $\pi_{t-s-1} - \hat{E}(\pi_{t-s-1}|z_{t-1})$ where the $\hat{E}(\cdot|z_{t-1})$ is the estimated conditional expectation given z_{t-1}. The bandwidth of the estimated conditional expectation is chosen from the values of C in the plug-in rule that minimizes the leave-one-out sum of squared errors. The nonparametric part can be estimated again using the kernel method by smoothing $\pi_t - \hat{\alpha}_1\pi_{t-s} - \hat{\alpha}_2\pi_{t-s-1}$ over z_{t-1}. The bandwidth is chosen by the plug-in rule and by choosing the differnt values of C around 1 that gives the smallest MSPE. The MSPE is calculated by $MSPE = \frac{1}{100}\sum_{t=420}^{519}(\pi_t - \hat{\pi}_t)^2$

where $\hat{\pi}_t = \hat{\alpha}_1 \pi_{t-s} + \hat{\alpha}_2 \pi_{t-s-1} + \hat{g}(z_{t-1})$ is the prediction using the data from the first period to period $t-1$.

The smooth coefficient model is estimated by the method of Li et al. (2002). The bandwidth is chosen from the plug-in rule with different values of C around 1 that gives the smallest MSPE. The MSPE is calculated by $MSPE = \frac{1}{100} \sum_{t=420}^{519} (\pi_t - \hat{\pi}_t)^2$ where $\hat{\pi}_t = \hat{\beta}_0(z_{t-s}) + \hat{\beta}_1(z_{t-s})\pi_{t-s} + \hat{\beta}_2(z_{t-s})\pi_{t-s-1}$ is the prediction using the data from the first period to period $t-1$.

CHAPTER IV

OPTIMAL SMOOTHING OF THE PARTIALLY LINEAR MODEL

A. Introduction

Over the last few decades, the semiparametric methods have been proposed and widely used in many real world applications. Both econometricians and statisticians have developed the theory specifically for the semiparametric model. One of the most popular model is the Robinson's (1988) semiparametric partially linear model. Robinson (1988) has shown that the parametric estimates are $\sqrt{n}$-consistent and asymptotically normally distributed with finite asymptotic variance. Li (1996) has purposed an alternative proof of Robinson's (1988) result and obtains a faster convergence rate for some of the average nonparametric kernel estimators, and relaxed some regularity conditions proposed by Robinson (1988). However, both of these literatures assume that the bandwidth h is prespecified or simply given by some ad-hoc method.

It is well known that the bandwidth is of crucial importance in nonparametric and semiparametric estimations. The optimal bandwidth is needed to obtain a good estimate. Many data-driven bandwidth methods have been proposed and the commonly used ones are the cross-validation (CV) method and generalized cross-validation (GCV) method. Härdle, Hall and Marron (1988) defined the theoretically optimal bandwidth and show that the bandwidth obtained by the cross-validation method converges in probability to the theoretically optimal value. This result has been used in many nonparametric and semiparametric literatures. Härdle, Hall and Ichimura (1993) used this result to show that the cross-validation bandwidth converges in probability uniformly over some shrinking compact sets to the theoretically optimal value in the single-index model setting.

This chapter will consider the convergence behavior of the cross-validation bandwidth under the semiparametric partially linear model setting. I propose that the finite dimensional parameter estimates and the cross-validation bandwidth can be obtained simultaneously by minimizing a sum of squared errors. This means the bandwidth need not be prespecified. I will show that the parameter estimates are still $\sqrt{n}$-consistent and asymptotically normally distributed and the cross-validation bandwidth estimates also converge in probability uniformly over some shrinking compact sets to the theoretically optimal bandwidth. The chapter is organized as follows: section B presents the idea and the methodology of how to obtain the estimators, section C gives the conclusion and the last section is dedicated to the proof of the main theorem.

B. Methodology

Consider Robinson's (1988) semiparametric partially linear model:

$$Y_i = X_i'\beta_0 + \theta(Z_i) + u_i \quad i = 1, ..., n \tag{4.1}$$

where Y_i is a scalar value dependent random variable, Z_i is a $p \times 1$ vector of random variables, X_i is a $d \times 1$ vector of random variables, β_0 is a $d \times 1$ vector of parameters, θ is an unknown scalar value function of Z and $\{u_i\}_{i=1}^n$ are the independent and identically distributed (i.i.d.) random variables with $E[u_i|X_i, Z_i] = 0$ and $E[u_i^2|X_i, Z_i] = \sigma^2 < \infty$. For $A = Y$ or X, define

$$\hat{A}_i = \hat{E}[A_i|Z_i] = \frac{1}{nh^p\hat{f}(Z_i)} \sum_{\substack{j=1 \\ j \neq i}}^{n} A_j K\left(\frac{Z_i - Z_j}{h}\right), \tag{4.2}$$

which is the leave-one-out nonparametric estimate of conditional expectation where h

is a bandwidth, K is a fixed kernel function and $\hat{f}(Z_i) = (nh^p)^{-1}\sum_{j=1,j\neq i}^{n} K\left((Z_i - Z_j)/h\right)$ is the nonparametric density estimation of Z at Z_i. Also define $\tilde{A}_i = A_i - \hat{A}_i$. The objective is to show that the bandwidth obtained from the next minimization problem converges in probability and does so uniformly over some shrinking compact sets to the theoretically optimal value. So I consider the following minimization problem:

$$(\hat{h}, \hat{\beta}) = \arg\min_{(h,\beta)} \frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \tilde{X}_i'\beta)^2 \tag{4.3}$$

The first order condition with respect to β in matrix notation is

$$\hat{\beta}_h = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y} \tag{4.4}$$

Substitute (4.4) to (4.3) gives

$$\hat{h} = \arg\min_{h} \frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \tilde{X}_i'\hat{\beta}_h)^2. \tag{4.5}$$

First I will decompose the objective function to

$$\frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \tilde{X}_i'\hat{\beta}_h)^2 = T(h) + R_1(h) + R_2(h) + R_3(h) + \text{terms do not depend on } h \tag{4.6}$$

Then I will show that the uniform convergence rate of $R_1(h)$, $R_2(h)$ and $R_3(h)$ is smaller than the uniform convergence rate of $T(h)$. So minimizing the objective function will lead to minimizing $T(h)$. The theoretically optimal value of h is defined to be the value h_0 that minimizes the mean squared error, i.e.,

$$h_0 = \arg\min_{h} E(\tilde{Y} - \tilde{X}'\beta_0)^2 \tag{4.7}$$

Härdle, Hall and Marron (1988) have shown that $\hat{h} = h_0 + o_p(n^{-1/(4+p)})$, where $\hat{h}$ is the minimizer of $T(h)$. Therefore, under the following assumptions, I can state the main theorem.

Assumption IV.1 *$(Y_i, X_i, Z_i)_{i=1}^n$ are independent and identically distributed as (Y, X, Z) and $\mathcal{Z}$, the support of Z, is a compact subset of $\mathbb{R}^p$.*

Assumption IV.2 *The density function f is bounded away from zero, i.e., $f(z) \geq b > 0$ for all $z \in \mathcal{Z}$.*

Assumption IV.3 *The kernel function is of second order with compact support, symmetric around zero, integrate to one and Lipschitz continuous, i.e., there exists a constant $D \geq 0$ such that $|K(x) - K(y)| \leq D|x - y|$ for all x, y in the support of the kernel function.*

Assumption IV.4 *(i) The functions $\phi(z) = E[Y|Z = z]$, $\xi(z) = E[X|Z = z]$ and $\theta(z)$ have two bounded, continuous derivatives on $\mathcal{Z}$; (ii) the matrix $\Phi \equiv E[\{X - E[X|Z]\}\{X - E[X|Z]\}']$ is positive definite.*

Assumption IV.5 *(i) U is independent of X and Z. For all $i \leq n$, $E[u_i|X_i, Z_i] = 0$, $E[u_i^2|X_i, Z_i] = \sigma^2 < \infty$, furthermore, $E|u_i|^l$, $E|X_i|^l$, $E|\xi(Z_i)|^l$ and $E|\theta(Z_i)|^l$ are all finite for all $l \in \mathbb{N}$.*

The assumptions IV.1 and IV.4 are the standard assumptions in the semiparametric and nonparametric literature. That the density function is bounded away from zero in assumption IV.2 can be relaxed and use the truncation technique in the proofs. Assumption IV.3 is the standard assumption but note that for higher dimension $p > 5$ the higher order kernel function is needed to be used to obtain $\sqrt{n}$-consistency of $\hat{\beta}_h$. The boundedness of all moments in assumption IV.5 can be relaxed to be bounded for some sufficiently large moments.

Now I can state the main theorem.

Theorem IV.1 *Under assumption IV.1-IV.5, define a compact set $\mathcal{H}_n = [\underline{c}n^{-1/(4+p)}, \bar{c}n^{-1/(4+p)}]$. For any $\zeta > 0$ and $p \leq 5$,* *(i)* $\sup_{h\in\mathcal{H}_n} T(h) = \sigma^2 + o_p(n^{-4/(4+p)+\varsigma})$

(ii) $\sup_{h\in\mathcal{H}_n} |R_2(h)| = o_p(n^{-(8+p)/(4+p)+\varsigma})$

(iii) $\sup_{h\in\mathcal{H}_n} |R_3(h)| = o_p(n^{-(12+p)/(8+2p)+\varsigma})$

Proof. The proof will be given in the last section. ∎

The theorem shows that the uniform convergence rate of $T(h)$ is slower than all other remaining terms, therefore, minimizing $T(h)$ will give $\hat{h}$ such that $\hat{h}/h_0 \to 1$ in probability by the result of Härdle, Hall and Marron (1988). Since for a fixed h, Robinson (1988) and Li(1996) showed that $\sqrt{n}(\hat{\beta}_h - \beta_0) \to N(0, \sigma^2\Phi^{-1})$ in distribution, and since $\hat{h}/h_0 \to 1$ in probability; hence one can show shat $\sqrt{n}(\hat{\beta}_{\hat{h}} - \beta_0) \to N(0, \sigma^2\Phi^{-1})$ in distribution.

C. Conclusion

In this chapter the uniform convergence rate of the cross-validation bandwidth is proved. I propose that, to get the parametric estimates, the bandwidth need not be prespecified; and the parameter estimates and the cross-validation bandwidth can be obtained simultaneously by minimizing a sum of squared errors. I show that the parameter estimates are still $\sqrt{n}$-consistent and asymptotically normally distributed and the cross-validation bandwidth estimates also converge in probability not only pointwise but also uniformly over some shrinking compact sets to the theoretically optimal bandwidth. This result is stronger than the pointwise convergence in Li (1996) and Robinson (1988). However, the boundednesses of higher moments of all the functions in the model are needed to obtain the uniform convergence property.

D. Proofs

Defind the short notation $\sum_i = \sum_{i=1}^n$, $\sum_{j\neq i} = \sum_{j=1,j\neq i}^n$, $\theta_i = \theta(Z_i)$, $\theta = (\theta_1, \theta_2, \ldots \theta_n)'$, $K_{ij} = K_{ijh} = K((Z_i - Z_j)/h)$. Since I am interested in the rate of convergence, I will write only the leading term, e.g., for $n(n-1)$ I will write n^2 for short, and the smaller order term will be abbreviated as $(s.o.)$. Define C with any subscripts to be the generic constants which may be different at different places. Also let

$$S_{A\hat{f},B\hat{f}} = \frac{1}{n}\sum_i A_i B_i' \hat{f}_i^2 \tag{4.8}$$

for any vector A_i and $B_i$and $S_A = S_{A,A}$. Since,

$$\frac{1}{\hat{f}} = \frac{1}{f} + \left(\frac{f-\hat{f}}{f\hat{f}}\right) \tag{4.9}$$

by using (4.8) and (4.9), one can show that

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n (\tilde{Y}_i - \tilde{X}_i'\beta)^2 &= \frac{1}{n}\sum_{i=1}^n (\tilde{Y}_i - \tilde{X}_i'\beta)^2 \frac{\hat{f}_i^2}{f_i^2} + \frac{1}{n}\sum_{i=1}^n (\tilde{Y}_i - \tilde{X}_i'\beta)^2 \hat{f}_i^2 \left(\frac{1}{\hat{f}_i^2} - \frac{1}{f_i^2}\right) \\ &= A(h) + R_1(h) \end{aligned} \tag{4.10}$$

since $|R_1(h)|$ has a smaller order than $A(h)$ by the uniform convergence of the non-parametric density estimation. Hence,

$$\begin{aligned} A(h) &= \frac{1}{n}\sum_{i=1}^n [\tilde{Y}_i - \tilde{X}_i'\beta_0 - \tilde{X}_i'(\hat{\beta}_h - \beta_0)]^2 \frac{\hat{f}_i^2}{f_i^2} \\ &= \frac{1}{n}\sum_{i=1}^n [\tilde{Y}_i - \tilde{X}_i'\beta_0]^2 \frac{\hat{f}_i^2}{f_i^2} + \frac{1}{n}\sum_{i=1}^n (\hat{\beta}_h - \beta_0)'\tilde{X}_i\tilde{X}_i'(\hat{\beta}_h - \beta_0)\frac{\hat{f}_i^2}{f_i^2} \\ &\quad - \frac{2}{n}\sum_{i=1}^n (\tilde{Y}_i - \tilde{X}_i'\beta_0)\tilde{X}_i'(\hat{\beta}_h - \beta_0)\frac{\hat{f}_i^2}{f_i^2} \\ &= T(h) + R_2(h) - 2R_3(h) \end{aligned} \tag{4.11}$$

For a fix value of h, I will show the pointwise convergence as follows:

Proposition IV.1 $T(h) = \sigma^2 + O_p(h^4 + \frac{1}{nh^p} + h^6 + \frac{1}{nh^{p/2}} + \frac{h^2}{\sqrt{n}h^{p/2}}) + o_p(1)$

Proof.

$$\begin{aligned} T(h) &= \frac{1}{n}\sum_{i=1}^{n}[\tilde{Y}_i - \tilde{X}_i'\beta_0]^2\frac{\hat{f}_i^2}{f_i^2} = \frac{1}{n}\sum_{i=1}^{n}[\theta_i - \hat{\theta}_i + u_i - \hat{u}_i]^2\frac{\hat{f}_i^2}{f_i^2} \\ &= S_{(\theta-\hat{\theta})\hat{f}/f} + S_{u\hat{f}/f} + S_{\hat{u}\hat{f}/f} - 2S_{u\hat{f}/f,\hat{u}\hat{f}/f} + 2S_{(\theta-\hat{\theta})\hat{f}/f,u\hat{f}/f} - 2S_{(\theta-\hat{\theta})\hat{f}/f,\hat{u}\hat{f}/f} \end{aligned} \tag{4.12}$$

Now showing the rate of each term, and by assumption IV.1 to IV.5,

$$\begin{aligned} E[S_{(\theta-\hat{\theta})\hat{f}/f}] &= E[(\theta_1 - \hat{\theta}_1)\hat{f}_1/f_1]^2 = E\left[\frac{1}{nh^p f_1}\sum_{j\neq 1}(\theta_1 - \theta_j)K_{1j}\right]^2 \\ &\leq \frac{C}{n^2h^{2p}}E\left[\sum_{j\neq 1}(\theta_1 - \theta_j)^2K_{1j}^2\right] + \frac{C}{n^2h^{2p}}E\left[\sum_{j\neq 1}\sum_{j\neq k,1}(\theta_1 - \theta_j)(\theta_1 - \theta_k)K_{1j}K_{1k}\right] \\ &= L_1 + L_2 \end{aligned}$$

One can show that

$$L_1 = \frac{Ch^2}{nh^p}[C_1 + h^2C_2] + (s.o.) \text{ and } L_2 = C_1h^4 + C_2h^6 + (s.o.)$$

So,

$$S_{(\theta-\hat{\theta})\hat{f}/f} = O_p(h^4 + \frac{h^2}{nh^p} + h^6 + \frac{h^4}{nh^p}) \tag{4.13}$$

Next, by using the law of large numbers,

$$
\begin{aligned}
S_{u\hat{f}/f} &= \frac{1}{n}\sum_i u_i^2 \hat{f}_i^2/f_i^2 = \frac{1}{n}\sum_i u_i^2 \frac{(\hat{f}_i - f_i + f_i)^2}{f_i^2} \\
&= \frac{1}{n}\sum_i u_i^2 \left[1 + \frac{(\hat{f}_i - f_i)^2}{f_i^2} + 2\frac{(\hat{f}_i - f_i)}{f_i}\right] \\
&= (\sigma^2 + o_p(1))\left(1 + O_p\left(h^4 + \frac{1}{nh^p}\right) + O_p\left(h^2 + \frac{1}{\sqrt{nh^p}}\right)\right) \\
&= \sigma^2 + o_p(1) \qquad (4.14)
\end{aligned}
$$

Next,

$$
\begin{aligned}
E[S_{\hat{u}\hat{f}/f}] &= E\left[\frac{1}{nh^p f_1}\sum_{j\neq 1} u_j K_{1j}\right]^2 \\
&\leq \frac{C}{n^2h^{2p}} E\left[\sum_{j\neq 1} u_j^2 K_{1j}^2 + \sum_{j\neq 1}\sum_{j\neq k,1} u_j u_k K_{1j} K_{1k}\right] \\
&= \frac{C\sigma^2}{n^2h^{2p}}(C_0 + C_1 h^2)
\end{aligned}
$$

Hence,

$$
S_{\hat{u}\hat{f}/f} = O_p\left(\frac{1}{nh^p} + \frac{h^2}{nh^p}\right) \qquad (4.15)
$$

Next,

$$
S^2_{u\hat{f}/f,\hat{u}\hat{f}/f} = \frac{1}{n^2}\left(\sum_i u_i^2 \hat{u}_i^2 \frac{\hat{f}_i^4}{f_i^4} + \sum_i\sum_{j\neq i} u_i u_j \hat{u}_i \hat{u}_j \frac{\hat{f}_i^2}{f_i^2}\frac{\hat{f}_j^2}{f_j^2}\right)
$$

$$
\begin{aligned}
E[S^2_{u\hat{f}/f,\hat{u}\hat{f}/f}] &= \frac{1}{n}E[u_1^2\hat{u}_1^2\hat{f}_1^4/f_1^4] + E[u_1u_2\hat{u}_1\hat{u}_2\hat{f}_1^2\hat{f}_2^2/(f_1f_2)^2] \\
&\leq \frac{C}{n^3h^{2p}}E\left[u_1^2\left(\sum_{i\neq 1}u_iK_{1i}\right)^2\hat{f}_1^2/f_1^2\right] \\
&+ \frac{C}{n^2h^{2p}}E\left[u_1u_2\left(\sum_{i\neq 1}u_iK_{1i}\right)\left(\sum_{j\neq 2}u_jK_{2j}\right)\hat{f}_1\hat{f}_2/(f_1f_2)\right] \\
&= \frac{C}{n^3h^{2p}}E\left[u_1^2\sum_{i\neq 1}u_i^2K_{1i}^2\hat{f}_1^2/f_1^2\right] + \frac{C}{n^2h^{2p}}E\left[u_1^2u_2^2K_{12}^2\hat{f}_1\hat{f}_2/(f_1f_2)\right] \\
&= \frac{C\sigma^4}{n^2h^p}O(1+h^2) + \frac{C\sigma^4}{n^2h^p}O(1+h^2) = O(\frac{1}{n^2h^p} + \frac{h^2}{n^2h^p})
\end{aligned}
$$

Hence,

$$
S_{u\hat{f}/f,\hat{u}\hat{f}/f} = O_p(\frac{1}{nh^{p/2}} + \frac{h}{nh^{p/2}}) \tag{4.16}
$$

Next,

$$
S^2_{(\theta-\hat{\theta})\hat{f}/f,u\hat{f}/f} = \frac{1}{n^2}\left(\sum_i(\theta_i-\hat{\theta}_i)^2u_i^2\frac{\hat{f}_i^4}{f_i^4} + \sum_i\sum_{j\neq i}(\theta_i-\hat{\theta}_i)(\theta_j-\hat{\theta}_j)u_iu_j\frac{\hat{f}_i^2}{f_i^2}\frac{\hat{f}_j^2}{f_j^2}\right)
$$

$$
\begin{aligned}
E[S^2_{(\theta-\hat{\theta})\hat{f}/f,u\hat{f}/f}] &= \frac{\sigma^2}{n^2}E\left[\sum_i(\theta_i-\hat{\theta}_i)^2\frac{\hat{f}_i^4}{f_i^4}\right] \\
&\leq \frac{\sigma^2C}{n^2}E\left[\sum_i(\theta_i-\hat{\theta}_i)^2\frac{\hat{f}^2}{f_i^2}\right] + (s.o.) \\
&= \frac{\sigma^2C}{n}O(h^4 + h^6 + \frac{h^2}{nh^p} + \frac{h^4}{nh^p})
\end{aligned}
$$

Hence,

$$
S_{(\theta-\hat{\theta})\hat{f}/f,u\hat{f}/f} = O_p(\frac{h^2}{\sqrt{n}} + \frac{h^3}{\sqrt{n}} + \frac{h}{nh^{p/2}} + \frac{h^2}{nh^{p/2}}) \tag{4.17}
$$

Last term,

$$S^2_{(\theta-\hat{\theta})\hat{f}/f,\hat{u}\hat{f}/f} = \frac{1}{n^2}\left(\sum_i (\theta_i - \hat{\theta}_i)^2 \hat{u}_i^2 \frac{\hat{f}_i^4}{f_i^4} + \sum_i \sum_{j\neq i} (\theta_i - \hat{\theta}_i)(\theta_j - \hat{\theta}_j)\hat{u}_i \hat{u}_j \frac{\hat{f}_i^2}{f_i^2}\frac{\hat{f}_j^2}{f_j^2}\right)$$

$$= L_3 + L_4$$

$$\begin{aligned}
E[L_3] &= \frac{1}{n}E[(\theta_1 - \hat{\theta}_1)^2 \hat{u}_1^2 \hat{f}_i^4/f_i^4] \leq \frac{C}{n^3h^{2p}}E\left[(\theta_1 - \hat{\theta}_1)^2 \hat{f}_1^2/f_1^2 \left(\sum_{j\neq 1} u_j K_{1j}\right)^2\right] \\
&= \frac{C}{n^3h^{2p}}E\left[(\theta_1 - \hat{\theta}_1)^2 \hat{f}_1^2/f_1^2 \sum_{i\neq 1} u_i^2 K_{1i}^2 + (\theta_1 - \hat{\theta}_1)^2 \hat{f}_1^2/f_1^2 \sum_{i\neq 1}\sum_{j\neq i,1} u_i u_j K_{1i} K_{1j}\right] \\
&= \frac{C}{n^3h^{2p}}E\left[(\theta_1 - \hat{\theta}_1)^2 \hat{f}_1^2/f_1^2 \sum_{i\neq 1} u_i^2 K_{1i}^2\right] \\
&\leq \frac{C^2}{n^5h^{4p}}E\left[\left(\sum_{j\neq 1}(\theta_1 - \theta_j)K_{1j}\right)^2 \left(\sum_{i\neq 1} u_i^2 K_{1i}^2\right)\right] \\
&= \frac{C^2}{n^5h^{4p}}E\left[\left(\sum_{i\neq 1}(\theta_1 - \theta_i)^2 K_{1i}^2\right)\left(\sum_{i\neq 1} u_i^2 K_{1i}^2\right)\right. \\
&\quad \left. + \left(\sum_{i\neq 1}\sum_{j\neq i,1}(\theta_1 - \theta_i)(\theta_1 - \theta_j)K_{1i}K_{1j}\right)\left(\sum_{i\neq 1} u_i^2 K_{1i}^2\right)\right] \\
&= \frac{\sigma^2 C^2}{n^5h^{4p}}E\left[\left(\sum_{i\neq 1}(\theta_1 - \theta_i)^2 K_{1i}^2\right)\left(\sum_{i\neq 1} K_{1i}^2\right)\right] \\
&\quad + \frac{\sigma^2 C^2}{n^5h^{4p}}E\left[\left(\sum_{i\neq 1}\sum_{j\neq i,1}(\theta_1 - \theta_i)(\theta_1 - \theta_j)K_{1i}K_{1j}\right)\left(\sum_{i\neq 1} K_{1i}^2\right)\right] \\
&= L_5 + L_6
\end{aligned}$$

$$\begin{aligned}
E[L_5] &= \frac{\sigma^2 C^2}{n^5h^{4p}}E\left[\sum_{i\neq 1}(\theta_1 - \theta_i)^2 K_{1i}^4 + \sum_{i\neq 1}\sum_{j\neq i,1}(\theta_1 - \theta_i)^2 K_{1i}^2 K_{1j}^2\right] \\
&= \frac{\sigma^2 C^2}{n^4h^{3p}}O(h^2 + h^4) + \frac{\sigma^2 C^2}{n^3h^{2p}}O(h^4 + h^6) \\
&= O\left(\frac{h^2}{n^4h^{3p}} + \frac{h^4}{n^4h^{3p}} + \frac{h^4}{n^3h^{2p}} + \frac{h^6}{n^3h^{2p}}\right)
\end{aligned}$$

$$\begin{aligned}
E[L_6] &= \frac{\sigma^2C^2}{n^5h^{4p}}E\left[\sum_{i\neq 1}\sum_{j\neq i,1}(\theta_1-\theta_i)(\theta_1-\theta_j)K_{1i}^3K_{1j}\right] \\
&+ \frac{\sigma^2C^2}{n^5h^{4p}}E\left[\sum_{i\neq 1}\sum_{j\neq i,1}\sum_{k\neq i,j,1}(\theta_1-\theta_i)(\theta_1-\theta_j)K_{1i}K_{1j}K_{1k}^2\right] \\
&= \frac{\sigma^2C^2}{n^3h^{2p}}O(h^4+h^6)+\frac{\sigma^2C^2}{n^2h^p}O(h^4+h^6) \\
&= O(\frac{h^4}{n^2h^p}+\frac{h^6}{n^2h^p})
\end{aligned}$$

$$\begin{aligned}
E[L_4] &= E[(\theta_1-\hat\theta_1)(\theta_2-\hat\theta_2)\hat u_1\hat u_2\hat f_1^2\hat f_2^2/(f_1f_2)^2] \\
&= E\left[(\theta_1-\hat\theta_1)(\theta_2-\hat\theta_2)\frac{\hat f_1\hat f_2}{f_1f_2}\left(\frac{1}{nh^pf_1}\sum_{i\neq 1}u_iK_{1i}\right)\left(\frac{1}{nh^pf_2}\sum_{j\neq 1}u_jK_{2j}\right)\right] \\
&\leq \frac{C}{n^2h^{2p}}E\left[(\theta_1-\hat\theta_1)(\theta_2-\hat\theta_2)\frac{\hat f_1\hat f_2}{f_1f_2}\left(\sum_{i\neq 1,2}u_i^2K_{1i}K_{2i}+\sum_{i\neq 1,2}\sum_{j\neq i,1,2}u_iu_jK_{1i}K_{2j}\right)\right] \\
&= \frac{C}{n^2h^{2p}}E\left[(\theta_1-\hat\theta_1)(\theta_2-\hat\theta_2)\frac{\hat f_1\hat f_2}{f_1f_2}\left(\sum_{i\neq 1,2}u_i^2K_{1i}K_{2i}\right)\right] \\
&\leq \frac{\sigma^2C^2}{n^4h^{4p}}E\left[\left(\sum_{j\neq 1}(\theta_1-\theta_j)K_{1j}\right)\left(\sum_{k\neq 2}(\theta_2-\theta_k)K_{2k}\right)\left(\sum_{i\neq 1,2}K_{1i}K_{2i}\right)\right] \\
&= \frac{\sigma^2C^2}{n^4h^{4p}}E\left[\sum_{i\neq 1,2}(\theta_1-\theta_i)(\theta_2-\theta_i)K_{1i}^2K_{2i}^2\right] \\
&+ \frac{\sigma^2C^2}{n^4h^{4p}}E\left[\sum_{j\neq 1}\sum_{k\neq j,1,2}(\theta_1-\theta_j)(\theta_2-\theta_k)K_{1j}^2K_{2j}K_{2k}\right] \\
&+ \frac{\sigma^2C^2}{n^4h^{4p}}E\left[\sum_{j\neq 2}\sum_{k\neq j,1,2}(\theta_1-\theta_k)(\theta_2-\theta_j)K_{1j}K_{1k}K_{2k}^2\right] \\
&+ \frac{\sigma^2C^2}{n^4h^{4p}}E\left[\sum_{i\neq 1,2}\sum_{j\neq 1,2}(\theta_1-\theta_j)(\vartheta_2-\theta_j)K_{1j}K_{2j}K_{1i}K_{2i}\right] \\
&+ \frac{\sigma^2C^2}{n^4h^{4p}}E\left[\sum_{j\neq 1}\sum_{k\neq j,2}\sum_{k\neq j,1,2}(\theta_1-\theta_j)(\theta_2-\theta_k)K_{1i}K_{2i}K_{1j}K_{2k}\right]
\end{aligned}$$

$$= \frac{\sigma^2 C^2}{n^3 h^{2p}} O(h^4 + h^6) + \frac{\sigma^2 C^2}{n^2 h^{2p}} O(h^4 + h^6) + \frac{\sigma^2 C^2}{n^2 h^{2p}} O(h^4 + h^6)$$
$$+ \frac{\sigma^2 C^2}{n^2 h^{2p}} O(h^4 + h^6) + \frac{\sigma^2 C^2}{n h^p} O(h^4 + h^6)$$
$$= O(\frac{h^4}{nh^p} + \frac{h^6}{nh^p})$$

Hence,

$$S_{\theta-\hat{\theta},\hat{u}} = O_p(\frac{h^2}{nh^{p/2}} + \frac{h^3}{nh^{p/2}} + \frac{h^2}{\sqrt{n}h^{p/2}} + \frac{h^3}{\sqrt{n}h^{p/2}}) \tag{4.18}$$

Combining all the terms I get,

$$T(h) = \sigma^2 + O_p(h^4 + \frac{1}{nh^p} + h^6 + \frac{1}{nh^{p/2}} + \frac{h^2}{\sqrt{n}h^{p/2}}) + o_p(1) \tag{4.19}$$

■

Proposition IV.2 $R_2(h) = O_p(\frac{h^4}{n} + \frac{1}{n^2 h^p} + \frac{h^6}{n} + \frac{1}{n^2 h^{p/2}} + \frac{h^2}{n^{3/2} h^{p/2}})$+*terms do not depend on h*

Proof. Since for fixed h, $\left\|\hat{\beta}_h - \beta_0\right\| = O_p(n^{-1/2})$, where $\|\cdot\|$ is the Euclidean norm, then

$$R_2(h) \leq \left\|\hat{\beta}_h - \beta_0\right\|^2 \left\|\frac{1}{n}\sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i' \frac{\hat{f}_i^2}{f_i^2}\right\| = O_p(n^{-1}) \left\|S_{\tilde{x}\hat{f}/f}\right\| \tag{4.20}$$

Let $X_i = E[X_i|Z_i] + v_i = \xi_i + v_i$, then $\hat{X}_i = \hat{\xi}_i + \hat{v}_i$ and hence, $\tilde{X}_i = \xi_i - \hat{\xi}_i + v_i - \hat{v}_i$. So,

$$S_{\tilde{X}\hat{f}/f} = S_{(\xi-\hat{\xi})\hat{f}/f} + S_{v\hat{f}/f} + S_{\hat{v}\hat{f}/f} + 2S_{(\xi-\hat{\xi})\hat{f}/f, v\hat{f}/f} - 2S_{(\xi-\hat{\xi})\hat{f}/f, \hat{v}\hat{f}/f} - 2S_{v\hat{f}/f, \hat{v}\hat{f}/f} \tag{4.21}$$

One can show that, $S_{(\xi-\hat{\xi})\hat{f}/f} = O_p(h^4 + h^6 + \frac{h^2}{nh^p} + \frac{h^4}{nh^p})$, $S_{v\hat{f}/f} = \Phi + o_p(1)$, $S_{\hat{v}\hat{f}/f} = O_p(\frac{1}{nh^p} + \frac{h^2}{nh^p})$, $S_{(\xi-\hat{\xi})\hat{f}/f, v\hat{f}/f} = O_p(\frac{h^2}{\sqrt{n}} + \frac{h^3}{\sqrt{n}} + \frac{h}{nh^{p/2}} + \frac{h^2}{nh^{p/2}})$, $S_{(\xi-\hat{\xi})\hat{f}/f, \hat{v}\hat{f}/f} = O_p(\frac{h^2}{nh^{p/2}} + \frac{h^3}{nh^{p/2}} + \frac{h^2}{\sqrt{n}h^{p/2}} + \frac{h^3}{\sqrt{n}h^{p/2}})$ and $S_{v\hat{f}/f, \hat{v}\hat{f}/f} = O_p(\frac{1}{nh^{p/2}} + \frac{h}{nh^{p/2}})$.

Therefore,

$$
\begin{aligned}
R_2(h) &= O_p(n^{-1})O_p(h^4 + \frac{1}{nh^p} + h^6 + \frac{1}{nh^{p/2}} + \frac{h^2}{\sqrt{n}h^{p/2}}) + \text{terms do not depend on } h \\
&= O_p(\frac{h^4}{n} + \frac{1}{n^2h^p} + \frac{h^6}{n} + \frac{1}{n^2h^{p/2}} + \frac{h^2}{n^{3/2}h^{p/2}}) + \text{terms do not depend on } h
\end{aligned}
\tag{4.22}
$$

■

Proposition IV.3 $R_3(h) = O_p(\frac{h^4}{\sqrt{n}} + \frac{1}{n^{3/2}h^p})$+*terms do not depend on h*

Proof.

$$
\begin{aligned}
|R_3(h)| &\leq \left\| \frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \tilde{X}_i'\beta_0)\tilde{X}_i'\frac{\hat{f}_i^2}{f_i^2} \right\| \left\| (\hat{\beta}_h - \beta_0) \right\| \\
&= O_p(n^{-1/2}) \left\| S_{(\theta-\hat{\theta}+u-\hat{u})\hat{f}/f,(\xi-\hat{\xi}+v-\hat{v})\hat{f}/f} \right\| \\
&= O_p(n^{-1/2}) \left\| S_{(\theta-\hat{\theta})\hat{f}/f,(\xi-\hat{\xi})\hat{f}/f} + S_{(\theta-\hat{\theta})\hat{f}/f,v\hat{f}/f} - S_{(\theta-\hat{\theta})\hat{f}/f,\hat{v}\hat{f}/f} + S_{u\hat{f}/f,(\xi-\hat{\xi})\hat{f}/f} \right. \\
&\quad \left. + S_{u\hat{f}/f,v\hat{f}/f} - S_{u\hat{f}/f,\hat{v}\hat{f}/f} - S_{\hat{u}\hat{f}/f,(\xi-\hat{\xi})\hat{f}/f} - S_{\hat{u}\hat{f}/f,v\hat{f}/f} + S_{\hat{u}\hat{f}/f,\hat{v}\hat{f}/f} \right\|
\end{aligned}
\tag{4.23}
$$

One can show that, $S_{(\theta-\hat{\theta})\hat{f}/f,(\xi-\hat{\xi})\hat{f}/f} = O_p(h^4 + h^6 + \frac{h^2}{nh^p} + \frac{h^4}{nh^p})$, $S_{(\theta-\hat{\theta})\hat{f}/f,v\hat{f}/f} = O_p(\frac{h^2}{\sqrt{n}} + \frac{h^3}{\sqrt{n}} + \frac{h}{nh^{p/2}} + \frac{h^2}{nh^{p/2}})$, $S_{(\theta-\hat{\theta})\hat{f}/f,\hat{v}\hat{f}/f} = O_p(\frac{h^2}{nh^{p/2}} + \frac{h^3}{nh^{p/2}} + \frac{h^2}{\sqrt{n}h^{p/2}} + \frac{h^3}{\sqrt{n}h^{p/2}})$, $S_{u\hat{f}/f,(\xi-\hat{\xi})\hat{f}/f} = O_p(\frac{h^2}{\sqrt{n}} + \frac{h^3}{\sqrt{n}} + \frac{h}{nh^{p/2}} + \frac{h^2}{nh^{p/2}})$, $S_{u\hat{f}/f,v\hat{f}/f} = \sigma_{uv} + o_p(1)$ does not depend on h, $S_{u\hat{f}/f,\hat{v}\hat{f}/f} = O_p(\frac{1}{nh^{p/2}} + \frac{h}{nh^{p/2}})$, $S_{\hat{u}\hat{f}/f,(\xi-\hat{\xi})\hat{f}/f} = O_p(\frac{h^2}{nh^{p/2}} + \frac{h^3}{nh^{p/2}} + \frac{h^2}{\sqrt{n}h^{p/2}} + \frac{h^3}{\sqrt{n}h^{p/2}})$, $S_{\hat{u}\hat{f}/f,v\hat{f}/f} = O_p(\frac{1}{nh^{p/2}} + \frac{h}{nh^{p/2}})$ and $S_{\hat{u}\hat{f}/f,\hat{v}\hat{f}/f} = O_p(\frac{1}{nh^{p/2}} + \frac{h}{nh^{p/2}})$.

Therefore,

$$
\begin{aligned}
R_3(h) &= O_p(n^{-1/2})O_p(h^4 + \frac{1}{nh^p}) + \text{terms do not depend on } h \\
&= O_p(\frac{h^4}{\sqrt{n}} + \frac{1}{n^{3/2}h^p}) + \text{terms do not depend on } h
\end{aligned}
\tag{4.24}
$$

■

To show the uniform convergence I simply prove it when $p = 1$. For $p > 1$ the proof will be more tedious but similar. for $p > 5$, the higher order kernel is used.

Proposition IV.4 *For* $p = 1$ *and* $\mathcal{H}_n = [\underline{c}n^{-1/(4+p)}, \bar{c}n^{-1/(4+p)}] = [\underline{c}n^{-1/5}, \bar{c}n^{-1/5}]$,

(i) $\sup_{h\in\mathcal{H}_n} |R_2(h)| = o_p(n^{-9/5+\varsigma})$

(ii) $\sup_{h\in\mathcal{H}_n} |R_3(h)| = o_p(n^{-13/10+\varsigma})$

Proof. Since $R_2(h) = \frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}_h-\beta_0)'\tilde{X}_i\tilde{X}_i'(\hat{\beta}_h-\beta_0)\hat{f}_i^2/f_i^2$, hence $\sup_{h\in\mathcal{H}_n}|R_2(h)| \leq \sup_{h\in\mathcal{H}_n}\left\|\hat{\beta}_h - \beta_0\right\|^2 \sup_{h\in\mathcal{H}_n}\left\|S_{\tilde{x}\hat{f}/f}\right\|$. And since $\hat{\beta}_h - \beta_0 = S^{-1}_{\tilde{x}\hat{f}/f}S_{\tilde{x}\hat{f}/f,(\tilde{\theta}+\tilde{u})\hat{f}/f}$, $\tilde{X} = (\xi - \hat{\xi}) + v - \hat{v}$, $\tilde{\theta} = \theta - \hat{\theta}$ and $\tilde{u} = u - \hat{u}$, so

$$S_{\tilde{x}\hat{f}/f} = S_{(\xi-\hat{\xi})\hat{f}/f} + S_{v\hat{f}/f} + S_{\hat{v}\hat{f}/f} + 2S_{(\xi-\hat{\xi})\hat{f}/f,v\hat{f}/f} + 2S_{(\xi-\hat{\xi})\hat{f}/f,\hat{v}\hat{f}/f} - 2S_{v\hat{f}/f,\hat{v}\hat{f}/f} \quad (4.25)$$

Hence,

$$\begin{aligned}\hat{\beta}_h - \beta_0 = &\left(S_{(\xi-\hat{\xi})\hat{f}/f} + S_{v\hat{f}/f} + S_{\hat{v}\hat{f}/f} + 2S_{(\xi-\hat{\xi})\hat{f}/f,v\hat{f}/f} + 2S_{(\xi-\hat{\xi})\hat{f}/f,\hat{v}\hat{f}/f} - 2S_{v\hat{f}/f,\hat{v}\hat{f}/f}\right)^{-1}\\ &\times\left(S_{(\xi-\hat{\xi})\hat{f}/f,(\theta-\hat{\theta})\hat{f}/f} + S_{(\xi-\hat{\xi})\hat{f}/f,u\hat{f}/f} + S_{(\xi-\hat{\xi})\hat{f}/f,\hat{u}\hat{f}/f} + S_{v\hat{f}/f,(\theta-\hat{\theta})\hat{f}/f} + S_{v\hat{f}/f,u\hat{f}/f}\right.\\ &\left.+S_{v\hat{f}/f,\hat{u}\hat{f}/f} + S_{\hat{v}\hat{f}/f,(\theta-\hat{\theta})\hat{f}/f} + S_{\hat{v}\hat{f}/f,u\hat{f}/f} + S_{\hat{v}\hat{f}/f,\hat{u}\hat{f}/f}\right) \quad (4.26)\end{aligned}$$

Similarly, $R_3(h) = \frac{1}{n}\sum_{i=1}^{n}(\tilde{Y}_i - \tilde{X}_i'\beta_0)\tilde{X}_i'(\hat{\beta}_h - \beta_0)\hat{f}_i^2/f_i^2$, so $\sup_{h\in\mathcal{H}_n}|R_3(h)| \leq \sup_{h\in\mathcal{H}_n}\left\|(\hat{\beta}_h - \beta_0)\right\| \sup_{h\in\mathcal{H}_n}\left\|S_{(\theta-\hat{\theta}+u-\hat{u})\hat{f}/f,(\xi-\hat{\xi}+v-\hat{v})\hat{f}/f}\right\|$. I will show the uniform convergence of $R_2(h)$ and $R_3(h)$ by proving the uniform convergence for each term on the right hand side of (4.26) in the following lemmas and use the fact from the central limit theorem that $\sqrt{n}S_{u\hat{f}/f,v\hat{f}/f} \rightarrow N(0,\sigma^2\Phi)$ in distribution to obtain the result. First note that, to show the uniform convergence, since $\mathcal{H}_n$ is compact, so for any $\delta > 0$, there exist a finite number N_δ and a finite set $\mathcal{H}_n' = \{h_1, h_2, ..., h_{N_\delta}\} \subset \mathcal{H}_n$ such that for all $h \in \mathcal{H}_n$, there exists $h' \in \mathcal{H}_n'$ such that $|h - h'| < \delta$. In this context, I let $\delta = (\bar{c} - \underline{c})n^{-1/5-c}$ for some constants c which will be selected for each cases differently later. Note that there are at most n^c elements in $\mathcal{H}_n'$. Hence, for any function $\varphi(h)$, I decompose it to

$$\sup_{h\in\mathcal{H}_n}|\varphi(h)| \le \sup_{h'\in\mathcal{H}'_n}|\varphi(h')| + \sup_{\substack{h\in\mathcal{H}_n,h'\in\mathcal{H}'_n\\|h-h'|<\delta}}|\varphi(h)-\varphi(h')| \tag{4.27}$$

So if I can show the convergence rate for each term on the right hand side of (4.27), then the uniform convergence rate of $\varphi(h)$ is proved. ■

These following lemmas will be used to prove the proposition $IV.4$.

Lemma IV.1 $\sup_{\substack{h\in\mathcal{H}_n,h'\in\mathcal{H}'_n\\|h-h'|<\delta}}|S_{(m-\hat{m}_h)\hat{f}_h/f}-S_{(m-\hat{m}_{h'})\hat{f}_{h'}/f}| = o_p(n^{-4/5+\zeta})$ *for* $\zeta>0, \delta = (\bar{c}-\underline{c})n^{-8/5}, m=\theta$ *or* $m=\xi$

Proof. I have to show that for any $\varepsilon>0$, $M>0$ there exists an n_0 such that $\forall n>n_0$

$$P\left(n^{4/5-\zeta}\sup_{\substack{h\in\mathcal{H}_n,h'\in\mathcal{H}'_n\\|h-h'|<n^{-8/5}}}|S_{(m-\hat{m}_h)\hat{f}_h/f}-S_{(m-\hat{m}_{h'})\hat{f}_{h'}/f}|>M\right)<\varepsilon \tag{4.28}$$

decompose $|S_{(m-\hat{m}_h)\hat{f}_h/f}-S_{(m-\hat{m}_{h'})\hat{f}_{h'}/f}|$ as

$$\begin{aligned}
&|S_{(m-\hat{m}_h)\hat{f}_h/f}-S_{(m-\hat{m}_{h'})\hat{f}_{h'}/f}|\\
&=\left|\frac{1}{n}\sum_i\left((m_i-\hat{m}_{ih})^2\hat{f}_{ih}^2/f_i^2-(m_i-\hat{m}_{ih'})^2\hat{f}_{ih'}^2/f_i^2\right)\right|\\
&=\left|\frac{1}{n}\sum_i\left\{\left(\frac{1}{nhf_i}\sum_{j\neq i}(m_i-m_j)K_{ijh}\right)^2-\left(\frac{1}{nh'f_i}\sum_{j\neq i}(m_i-m_j)K_{ijh'}\right)^2\right\}\right|\\
&\le\left|\frac{C}{n^3}\sum_i\left\{\frac{1}{h^2}\sum_{j\neq i}(m_i-m_j)^2K_{ijh}^2+\frac{1}{h^2}\sum_{j\neq i}\sum_{k\neq j,i}(m_i-m_j)(m_i-m_k)K_{ijh}K_{ikh}\right.\right.\\
&\left.\left.-\frac{1}{h'^2}\sum_{j\neq i}(m_i-m_j)^2K_{ijh'}^2-\frac{1}{h'^2}\sum_{j\neq i}\sum_{k\neq j,i}(m_i-m_j)(m_i-m_k)K_{ijh'}K_{ikh'}\right\}\right|
\end{aligned}$$

$$= \left| \frac{C}{n^3} \sum_i \left\{ \frac{1}{h^2} \sum_{j \neq i} (m_i - m_j)^2 K_{ijh}^2 - \frac{1}{h'^2} \sum_{j \neq i} (m_i - m_j)^2 K_{ijh'}^2 \right\} \right.$$

$$+ \frac{C}{n^3} \sum_i \left\{ \frac{1}{h^2} \sum_{j \neq i} \sum_{k \neq j,i} (m_i - m_j)(m_i - m_k) K_{ijh} K_{ikh} \right.$$

$$\left. \left. - \frac{1}{h'^2} \sum_{j \neq i} \sum_{k \neq j,i} (m_i - m_j)(m_i - m_k) K_{ijh'} K_{ikh'} \right\} \right|$$

$$= |A_1 + A_2|$$

$$|A_1| = \left| \frac{C}{n^3} \sum_i \left\{ \frac{1}{h^2} \sum_{j \neq i} (m_i - m_j)^2 K_{ijh}^2 - \frac{1}{h'^2} \sum_{j \neq i} (m_i - m_j)^2 K_{ijh}^2 \right. \right.$$

$$\left. \left. + \frac{1}{h'^2} \sum_{j \neq i} (m_i - m_j)^2 K_{ijh}^2 - \frac{1}{h'^2} \sum_{j \neq i} (m_i - m_j)^2 K_{ijh'}^2 \right\} \right|$$

$$\leq \left| \frac{C}{n^3} \sum_i \sum_{j \neq i} (m_i - m_j)^2 K_{ijh}^2 \left(\frac{1}{h^2} - \frac{1}{h'^2} \right) \right| + \left| \frac{C}{n^3} \sum_i \frac{1}{h'^2} \sum_{j \neq i} (m_i - m_j)^2 (K_{ijh}^2 - K_{ijh'}^2) \right|$$

$$\leq \frac{C}{n^3} \sum_i \sum_{j \neq i} (m_i - m_j)^2 K_{ijh}^2 \frac{(h + h')|h - h'|}{h^2 h'^2}$$

$$+ \frac{C}{n^3 h'^2} \sum_i \sum_{j \neq i} (m_i - m_j)^2 (K_{ijh} + K_{ijh'}) D |Z_i - Z_j| \frac{|h - h'|}{h h'}$$

$$|A_2| = \left| \frac{C}{n^3} \sum_i \left\{ \frac{1}{h^2} \sum_{j \neq i} \sum_{k \neq j,i} (m_i - m_j)(m_i - m_k) K_{ijh} K_{ikh} \right. \right.$$

$$- \frac{1}{h'^2} \sum_{j \neq i} \sum_{k \neq j,i} (m_i - m_j)(m_i - m_k) K_{ijh} K_{ikh}$$

$$+ \frac{1}{h'^2} \sum_{j \neq i} \sum_{k \neq j,i} (m_i - m_j)(m_i - m_k) K_{ijh} K_{ikh}$$

$$\left. \left. - \frac{1}{h'^2} \sum_{j \neq i} \sum_{k \neq j,i} (m_i - m_j)(m_i - m_k) K_{ijh'} K_{ikh'} \right\} \right|$$

$$\leq \frac{C}{n^3}\sum_i \sum_{j\neq i} \sum_{k\neq j,i} |(m_i - m_j)(m_i - m_k)K_{ijh}K_{ikh}| \frac{(h+h')|h-h'|}{h^2h'^2}$$

$$+ \frac{C}{n^3h'^2}\sum_i \left| \sum_{j\neq i} \sum_{k\neq j,i} (m_i - m_j)(m_i - m_k)K_{ijh}K_{ikh} \right.$$

$$- \sum_{j\neq i} \sum_{k\neq j,i} (m_i - m_j)(m_i - m_k)K_{ijh'}K_{ikh}$$

$$+ \sum_{j\neq i} \sum_{k\neq j,i} (m_i - m_j)(m_i - m_k)K_{ijh'}K_{ikh}$$

$$\left. - \sum_{j\neq i} \sum_{k\neq j,i} (m_i - m_j)(m_i - m_k)K_{ijh'}K_{ikh'} \right|$$

$$\leq \frac{C}{n^3}\sum_i \sum_{j\neq i} \sum_{k\neq j,i} |(m_i - m_j)(m_i - m_k)K_{ijh}K_{ikh}| \frac{(h+h')|h-h'|}{h^2h'^2}$$

$$+ \frac{C}{n^3h'^2}\sum_i \sum_{j\neq i} \sum_{k\neq j,i} |(m_i - m_j)(m_i - m_k)K_{ikh}||K_{ijh} - K_{ijh'}|$$

$$+ \frac{C}{n^3h'^2}\sum_i \sum_{j\neq i} \sum_{k\neq j,i} |(m_i - m_j)(m_i - m_k)K_{ijh'}||K_{ikh} - K_{ikh'}|$$

$$\leq \frac{C}{n^3}\sum_i \sum_{j\neq i} \sum_{k\neq j,i} |(m_i - m_j)(m_i - m_k)K_{ijh}K_{ikh}| \frac{(h+h')|h-h'|}{h^2h'^2}$$

$$+ \frac{C}{n^3h'^2}\sum_i \sum_{j\neq i} \sum_{k\neq j,i} |(m_i - m_j)(m_i - m_k)K_{ikh}|D|Z_i - Z_j| \frac{|h-h'|}{hh'}$$

$$+ \frac{C}{n^3h'^2}\sum_i \sum_{j\neq i} \sum_{k\neq j,i} |(m_i - m_j)(m_i - m_k)K_{ijh'}|D|Z_i - Z_k| \frac{|h-h'|}{hh'}$$

Hence,

$$
\begin{aligned}
&\sup_{\substack{h\in\mathcal{H}_n,h'\in\mathcal{H}'_n\\|h-h'|<\delta}} |S_{(m-\hat{m}_h)\hat{f}_h/f} - S_{(m-\hat{m}_{h'})\hat{f}_{h'}/f}| \\
&\leq \frac{C}{n^3}\sum_i\sum_{j\neq i}(m_i-m_j)^2K_{ijh}^2\frac{(h+h')|h-h'|}{h^2h'^2} \\
&+\frac{C}{n^3h'^2}\sum_i\sum_{j\neq i}(m_i-m_j)^2(K_{ijh}+K_{ijh'})D|Z_i-Z_j|\frac{|h-h'|}{hh'} \\
&+\frac{C}{n^3}\sum_i\sum_{j\neq i}\sum_{k\neq j,i}|(m_i-m_j)(m_i-m_k)K_{ijh}K_{ikh}|\frac{(h+h')|h-h'|}{h^2h'^2} \\
&+\frac{C}{n^3h'^2}\sum_i\sum_{j\neq i}\sum_{k\neq j,i}|(m_i-m_j)(m_i-m_k)K_{ikh}|D|Z_i-Z_j|\frac{|h-h'|}{hh'} \\
&+\frac{C}{n^3h'^2}\sum_i\sum_{j\neq i}\sum_{k\neq j,i}|(m_i-m_j)(m_i-m_k)K_{ijh'}|D|Z_i-Z_k|\frac{|h-h'|}{hh'} \\
&\leq \frac{C_1n^{-1/5}\delta}{nn^{-4/5}}\left(\frac{1}{n^2}\sum_i\sum_{j\neq i}(m_i-m_j)^2\right)+\frac{C_2\delta}{nn^{-2/5}n^{-2/5}}\left(\frac{1}{n^2}\sum_i\sum_{j\neq i}(m_i-m_j)^2|Z_i-Z_j|\right) \\
&+\frac{C_3n^{-1/5}\delta}{n^{-4/5}}\left(\frac{1}{n^3}\sum_i\sum_{j\neq i}\sum_{k\neq j,i}|(m_i-m_j)(m_i-m_k)|\right) \\
&+\frac{C_4\delta}{n^{-2/5}n^{-2/5}}\left(\frac{1}{n^3}\sum_i\sum_{j\neq i}\sum_{k\neq j,i}|(m_i-m_j)(m_i-m_k)||Z_i-Z_j|\right) \\
&+\frac{C_4\delta}{n^{-2/5}n^{-2/5}}\left(\frac{1}{n^3}\sum_i\sum_{j\neq i}\sum_{k\neq j,i}|(m_i-m_j)(m_i-m_k)||Z_i-Z_k|\right) \\
&= A_3
\end{aligned}
$$

Using Markov's inequality, I find that

$$
\begin{aligned}
P\left(n^{4/5-\varsigma}\sup_{\substack{h\in\mathcal{H}_n,h'\in\mathcal{H}'_n\\|h-h'|<(\bar{c}-\underline{c})n^{-8/5}}}|S_{(m-\hat{m}_h)\hat{f}_h/f}-S_{(m-\hat{m}_{h'})\hat{f}_{h'}/f}|>M\right) &\leq P(n^{4/5-\varsigma}A_3>M) \\
&\leq E[n^{4/5-\varsigma}A_3]/M
\end{aligned}
$$

and

$$E[n^{4/5-\zeta}A_3] = C_1'n^{-6/5-\zeta}E[m_1 - m_2]^2 + C_2'n^{-1-\zeta}E[(m_1 - m_2)^2|Z_1 - Z_2|]$$
$$+ C_3'n^{-1/5-\zeta}E[|(m_1 - m_2)(m_1 - m_3)|]$$
$$+ C_4'n^{-\zeta}E[|(m_1 - m_2)(m_1 - m_3)||Z_1 - Z_2|]$$
$$+ C_4'n^{-\zeta}E[|(m_1 - m_2)(m_1 - m_3)||Z_1 - Z_3|]$$

Since all the expectations are finite, so it is obvious that such n_0 exists. ∎

Lemma IV.2 $\sup_{h'\in\mathcal{H}_n'} |S_{(m-\hat{m}_{h'})\hat{f}_{h'}/f}| = o_p(n^{-4/5+\zeta})$ *for* $\zeta > 0, m = \theta$ *or* $m = \xi$

Proof. First note that there are at most $n^{7/5}$ elements in $\mathcal{H}_n'$, by Markov's inequality,

$$P\left(n^{4/5-\zeta}\sup_{h'\in\mathcal{H}_n'} |S_{(m-\hat{m}_{h'})\hat{f}_{h'}/f}| > M\right) \leq 2n^{7/5}\sup_{h'\in\mathcal{H}_n'} P\left(n^{4/5-\zeta}|S_{(m-\hat{m}_{h'})\hat{f}_{h'}/f}| > M\right)$$
$$\leq 2n^{7/5}(n^{-\zeta}/M)^{2l}\sup_{h'\in\mathcal{H}_n'} E[n^{4/5}S_{(m-\hat{m}_{h'})\hat{f}_{h'}/f}]^{2l}$$

provided that a positive integer $l > \frac{7}{10\zeta}$.

If I can show that $\sup_{h'\in\mathcal{H}_n'} E[n^{4/5}S_{(m-\hat{m}_{h'})\hat{f}_{h'}/f}]^{2l} = O(1)$ then the proof is done. Since $E[S_{(m-\hat{m}_{h'})\hat{f}_{h'}/f}]^{2l} = E[(m_1 - \hat{m}_1)\hat{f}_1/f_1]^{2l}$, expand the right hand side first and then take the expectation, then the leading term will be

$$L = \frac{1}{n^{2l}h'^{2lp}f_1^{2l}}\sum_{i_1\neq 1}\sum_{i_2\neq i_1,1}\cdots\sum_{i_{2l}\neq i_{2l-1},i_{2l-2},\ldots,i_1,1}\prod_{j=1}^{2l}(m_1 - m_{i_j})K_{1i_j}$$

Given that the $2l^{\text{th}}$ moment of m is finite, it can be easily shown that $E[L] = O_p\left(h'^{4l}\right)$. For $h' \in \mathcal{H}_n'$, $E[L] = O_p(n^{-4l/5})$, hence $\sup_{h'\in\mathcal{H}_n'} E[n^{4/5}S_{(m-\hat{m}_{h'})\hat{f}_{h'}/f}]^{2l} = O(1)$. ∎

The next lemma will be used to prove uniform convergence. The proof of the lemma is similar to the lemma *IV*.1 and lemma *IV*.2, so I will omit the proof.

Lemma IV.3 *for any* $\zeta > 0$,

(i) $\sup_{\substack{h\in\mathcal{H}_n,h'\in\mathcal{H}'_n\\ |h-h'|<\delta}} |S_{\hat{t}_h\hat{f}_h/f} - S_{\hat{t}_{h'}\hat{f}_{h'}/f}| = o_p(n^{-4/5+\zeta})$ *for* $\delta = (\bar{c}-\underline{c})n^{-8/5}, t = u$ *or* $t = v$

(ii) $\sup_{\substack{h\in\mathcal{H}_n,h'\in\mathcal{H}'_n\\ |h-h'|<\delta}} |S_{(m-\hat{m}_h)\hat{f}_h/f,\hat{t}_h\hat{f}_h/f} - S_{(m-\hat{m}_{h'})\hat{f}_{h'}/f,\hat{t}_{h'}\hat{f}_{h'}/f}| = o_p(n^{-4/5+\zeta})$
for $\delta = (\bar{c}-\underline{c})n^{-8/5}, m = \theta$ *or* $m = \xi$, *and* $t = u$ *or* $t = v$

(iii) $\sup_{\substack{h\in\mathcal{H}_n,h'\in\mathcal{H}'_n\\ |h-h'|<\delta}} |S_{(m-\hat{m}_h)\hat{f}_h/f,t\hat{f}_h/f} - S_{(m-\hat{m}_{h'})\hat{f}_{h'}/f,t\hat{f}_{h'}/f}| = o_p(n^{-9/10+\zeta})$
for $\delta = (\bar{c}-\underline{c})n^{-3/2}, m = \theta$ *or* $m = \xi$, *and* $t = u$ *or* $t = v$

(iv) $\sup_{\substack{h\in\mathcal{H}_n,h'\in\mathcal{H}'_n\\ |h-h'|<\delta}} |S_{\hat{m}_h\hat{f}_h/f,t\hat{f}_h/f} - S_{\hat{m}_{h'}\hat{f}_{h'}/f,t\hat{f}_{h'}/f}| = o_p(n^{-9/10+\zeta})$ *for* $\delta = (\bar{c}-\underline{c})n^{-3/2}$,
$m = u, t = v$ *or* $m = v$, $t = u$

(v) $\sup_{\substack{h\in\mathcal{H}_n,h'\in\mathcal{H}'_n\\ |h-h'|<\delta}} |S_{\hat{m}_h\hat{f}_h/f,\hat{t}_h\hat{f}_h/f} - S_{\hat{m}_{h'}\hat{f}_{h'}/f,\hat{t}_{h'}\hat{f}_{h'}/f}| = o_p(n^{-9/10+\zeta})$ *for* $\delta = (\bar{c}-\underline{c})n^{-3/2}$,
$m = u, t = v$ *or* $m = v$, $t = u$

(vi) $\sup_{\substack{h\in\mathcal{H}_n,h'\in\mathcal{H}'_n\\ |h-h'|<\delta}} |S_{(m-\hat{m}_h)\hat{f}_h/f,(t-\hat{t}_h)\hat{f}_h/f} - S_{(m-\hat{m}_{h'})\hat{f}_{h'}/f,(t-\hat{t}_{h'})\hat{f}_{h'}/f}| = o_p(n^{-4/5+\zeta})$ *for*
$\delta = (\bar{c}-\underline{c})n^{-8/5}, m = \theta$ *or* $m = \xi$, *and* $t = \theta$ *or* $t = \xi$

(vii) $\sup_{h'\in\mathcal{H}'_n} |S_{\hat{t}_{h'}\hat{f}_{h'}/f}| = o_p(n^{-4/5+\zeta})$ *for* $t = u$ *or* $t = v$

(viii) $\sup_{h'\in\mathcal{H}'_n} |S_{(m-\hat{m}_{h'})\hat{f}_{h'}/f,\hat{t}_{h'}\hat{f}_{h'}/f}| = o_p(n^{-4/5+\zeta})$ *for* $m = \theta$ *or* $m = \xi$, *and* $t = u$ *or*
$t = v$

(ix) $\sup_{h'\in\mathcal{H}'_n} |S_{(m-\hat{m}_{h'})\hat{f}_{h'}/f,t\hat{f}_{h'}/f}| = o_p(n^{-9/10+\zeta})$ *for* $m = \theta$ *or* $m = \xi$, *and* $t = u$ *or*
$t = v$

(x) $\sup_{h'\in\mathcal{H}'_n} |S_{\hat{m}_{h'}\hat{f}_{h'}/f,t\hat{f}_{h'}/f}| = o_p(n^{-9/10+\zeta})$ *for* $m = u, t = v$ *or* $m = v$, $t = u$

(xi) $\sup_{h'\in\mathcal{H}'_n} |S_{\hat{m}_{h'}\hat{f}_{h'}/f,\hat{t}_{h'}\hat{f}_{h'}/f}| = o_p(n^{-9/10+\zeta})$ *for* $m = u, t = v$ *or* $m = v$, $t = u$

(xii) $\sup_{h'\in\mathcal{H}'_n} |S_{(m-\hat{m}_{h'})\hat{f}_{h'}/f,(t-\hat{t}_{h'})\hat{f}_{h'}/f}| = o_p(n^{-4/5+\zeta})$ *for* $m = \theta$ *or* $m = \xi$, *and* $t = \theta$
or $t = \xi$

CHAPTER V

CONCLUSION

In chapter II, I propose using a general series method to estimate the semiparametric partially linear smooth coefficient model. I show that the estimator $\hat{\gamma}$ has the $\sqrt{n}$-normality property and it attains the semiparametric efficiency bound when the error is homoskedastic. I also show that it is easy to modify the method to the hetroskedastic error case and that the estimator is still efficient under the modification. The convergence rate of the smooth coefficient function is proven and the primitive conditions of the power series and B-spline are restated as examples.

Where the Monte Carlo simulation is also conducted, three different methods of estimation are used. I find that for every sample size and every data generating process used in this simulation, the series estimations perform better than the kernel method. In particular, the B-spline method performs best and the kernel method is the last.

The application of the semiparametic partially linear smooth coefficient model is illustrated by applying it to forecasting the inflation rate using the unemployment rate and the industry capacity utilization rate. The specification test results show that the semiparametric partially linear smooth coefficient model is more appropriate than the full smooth coefficient model and the parametric autoregressive model.

Chapter III has made several notable contributions. By not relying on the straightjacket of a linear VAR forecasting model, I show strong and robust evidence that money growth has been useful as an indicator of future inflation in the US. The finding of a strong and sufficiently stable relationship between money growth and inflation is the opposite of that found in most recent work. In particular, there is a near consensus that the only relationship between money growth and inflation is at long horizons. I have presented evidence that money growth has marginal predictive content for inflation at horizons as short as one month. More generally, the results indicate that nonparametric methods might be useful for the monetary policy process. Nonlinearities in the relationships of interest to central bankers appear to be strong. When modeling inflation, the benefits of relaxing the assumptions on functional form substantially outweigh any disadvantages of nonparametric estimation.

In chapter IV, the uniform convergence rate of the cross-validation bandwidth is proved. I propose that, to get the parametric estimates, the bandwidth need not be prespecified; and the parameter estimates and the cross-validation bandwidth can be obtained simultaneously by minimizing a sum of squared errors. I show that the parameter estimates are still $\sqrt{n}$-consistent and asymptotically normally distributed and the cross-validation bandwidth estimates also converge in probability not only pointwise but also uniformly over some shrinking compact sets to the theoretically optimal bandwidth. This result is stronger than the pointwise convergence in Li (1996) and Robinson (1988). However, the boundednesses of higher moments of all the functions in the model are needed to obtain the uniform convergence property.

REFERENCES

Barnett, W.A., Serletis, A., 2000. The Theory of Monetary Aggregation. North-Holland, New York.

Belongia, M.T., 1996. Measurement Matters: Recent Results from Monetary Economics Reexamined. The Journal of Political Economy 104, 1065-1083.

Belongia, M.T., Chalfant, J.A., 1989. The Changing Empirical Definition of Money: Some Estimates from a Model of the Demand for Money Substitutes. The Journal of Political Economy 97, 387–397.

Bernanke, B.S., Mishkin, F.S., 1997. Inflation Targeting: A New Framework for Monetary Policy?. The Journal of Economic Perspectives 11, 97–116.

Bresnahan, T.F., 1982. The Oligopoly Solution Concept is Identified. Economics Letters 10, 87–92.

Cecchetti, S., 1995. Inflation Indicators and Inflation Policy. In: Ben S. Bernanke and Julio Rotemberg (Eds.), National Bureau of Economic Research Macroeconomics Annual 1995. MIT Press, Cambridge, pp. 189–219.

Chamberlain, G., 1992. Efficiency Bounds for Semiparametric Regression. Econometrica 60, 567–596.

Diebold, F., Nason, J., 1990. Nonparametric Exchange Rate Prediction?.

Journal of International Economics 28, 315–332.

Estrella, A., Mishkin, F.S., 1997. Is there a role for monetary aggregates in the conduct of monetary policy?. Journal of Monetary Economics 40, 279–304.

Friedman, B.M., 1988. Lessons on Monetary Policy from the 1980s. The Journal of Economic Perspectives 2, 51–72.

Friedman, B.M., Kuttner, K.N., 1992. Money, Income, Prices, and Interest Rates. The American Economic Review 82, 472–492.

Friedman, B.M., Kuttner, K.N., 1996. A Price Target for U.S. Monetary Policy? Lessons from the Experience with Money Growth Targets. Brookings Papers on Economic Activity 1, 77–125.

Gerlach, S., Svensson, L., 2001. Money and Inflation in the Euro Area: A Case for Monetary Indicators?. Working Paper, Princeton University.

Hallman, J.J., Porter, R.D., Small, D.H., 1991. Is the Price Level Tied to the M2 Monetary Aggregate in the Long Run?. The American Economic Review 81, 841–858.

Hamilton, J.D., 2001. A Parametric Approach to Flexible Nonlinear Inference. Econometrica 69. 537–573.

Härdle, W., Hall, P., Ichimura, H., 1993. Optimal Smoothing in Sigle-

Index Models. Annals of Statistics 21, 157–178.

Härdle, W., Hall, P., Marron, J.S., 1988. How Far Are Automatically Chosen Regression Smoothing Parameters From Their Optimum?. Journal of the American Statistical Association 83, 86–95.

Hart, J.D., 1997. Nonparametric Smoothing and Lack-of-Fit Tests. Springer-Verlag, New York.

Li, Q., 1996. On the Root-N-Consistent Semiparametric Estimation of Partially Linear Models. Economics Letters 51, 277–285.

Li, Q., 2000. Efficient Estimation of Additive Partially Linear Models. International Economic Review 41, 1073–1092.

Li, Q., Huang, C.J., Li, D., Fu, T.T., 2002. Semiparametric Smooth Coefficient Models. Journal of Business and Economic Statistics 20, 412–422.

Michael, P., Nobay, A.R., Peel, D.A., 1997. Transactions Costs and Nonlinear Adjustment in Real Exchange Rates: An Empirical Investigation. The Journal of Political Economy 105, 862–879.

Newey, W.K., 1997. Convergence Rates and Asymptotic Normality for Series Estimators. Journal of Econometrics 79, 147–168.

Robinson, P.M., 1988. Root-N-consistent Semiparametric Regression. Econometrica 56, 931–954.

Rudebusch, G., Svensson, L., 2000. Policy Rules for Inflation Targeting. In: John B. Taylor (Ed.), Monetary Policy Rules. University of Chicago Press, Chicago.

Stock, J.H., Watson, M.W., 1999. Forecasting Inflation. Journal of Monetary Economics 44, 293–335.

Svensson, L., 1997. Inflation Forecast Targeting: Implementing and Monitoring Inflation Targets. European Economic Review 41, 1111–1146.

Svensson, L., Woodford, M., 2002. Indicator variables for optimal policy. Working Paper, Princeton University.

Taylor, A.M., 2001. Potential Pitfalls for the Purchasing-Power Parity Puzzle? Sampling and Specification Biases in Mean-Reversion Tests of the Law of One Price. Econometrica 69, 473–498.

VITA

Sittisak Leelahanon Email: sittisak@lycos.com

64/33 Latprakao 6 Latprakao Rd. Latprao Bangkok 10230 Thailand

---

Education

B.Eng. (Electrical Engineering), Chulalongkorn University (May 1995)

B.B.A. (General Management), Ramkhamhaeng University (June 1996)

Ph.D. Economics, Texas A&M University (Expected December 2002)

Fields of Specialization

Primary: Econometrics and Game Theory

Secondary: Financial Mathematics

Professional Experience

Instructor:

Principles of Microeconomics (Spring 2001)

Teaching Assistant:

(Undergraduate) Principles of Microeconomics (Spring 1999)

(Undergraduate) Principles of Macroeconomics (Fall 1999)

(Graduate) Econometrics I (Fall 2000)

(Undergraduate) Antitrust Economics (Fall 2001, Spring 2002)

(Graduate) Econometrics III (Fall 2002)

Research Assistant:

Research assistant of the Private Enterprise Research Center (Summer 2002)

Referee for Journal of Nonparametric Statistics (Fall 2002)

Reference

Prof. Qi Li Email: qi@econ.tamu.edu Phone: (979) 845-9954

Department of Economics Texas A&M University College Station, TX 77843-4228