The Dissertation Committee for Shih-Chang Lin
certifies that this is the approved version of the following dissertation:

Three Essays in Econometrics

Committee:

Stephen G. Donald, Supervisor

Maxwell B. Stinchcombe

Paul W. Wilson

Thomas S. Shively

Li Gan

Three Essays in Econometrics

by

Shih-Chang Lin, B.A., M.A.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2004

Dedicated to my family.

Acknowledgments

I would like to thank my supervisor, Dr. Stephen Donald for his constant encouragement and wonderful guidance. He has been infinitely patient, caring, understanding and motivating figure during my development of this dissertation. I would also like to thank the other members of my dissertation committee, Dr. Max Stinchcombe, Dr. Paul Wilson, Dr. Li Gan, as well as Dr. Tom Shively, who provided invaluable insights and suggests. I also want to thank Dr. Chien-Fu Lin, Dr. Hui-Lin Lin and Professor Jen-Tsong Chen in National Taiwan University for being my mentors and role models for a long time.

I am grateful to the amazing friends I had met while in the graduate school, Anders, Hamid, Kai-Lun, Wei-Yueh, Shu-Yi, Shu-Keng, Tsu-Yu, and Wee-Keng, who made my study in Economics department free and relaxed. And the members in badminton team, Heng-Rue, Hung-Ming, Hsiu-Chen, Wan-Hsuan, Rong, Shu-Hua, Yang-Yu, Jun-Nan and the Kuo family, Keng family, made my life in Austin infinitely better.

I would like to express my deepest gratitude to my dear Mom and Dad. They have been offering a stimulative and caring environment to me since my childhood. And they provide much spiritual and material support to help me pursuing my dream. I love them. My sister, Shu-Ya, and brother, Shih-Feng always provide unconditional support. I am also grateful to the encouragement from my parents in-law and sisters in-law, Gina and Sarah.

Finally, I would like to thank my two sons, 3.5-year-old Max and 13-month-old Thomas. They are the sources of joy and comfort when I got stuck in research. Most of all, I would like to express my thanks to my wife, Chia-Hsing for her inspiration, patience, and support over the years. She plays the wonderful roles of my wife and mother of my sons as well as a Ph.D student in Education department of the University. I really appreciate her understanding during the stressful time. This dissertation would not be completed without her love and sacrifice.

Three Essays in Econometrics

Publication No. ____________

Shih-Chang Lin, Ph.D.
The University of Texas at Austin, 2004

Supervisor: Stephen G. Donald

In this dissertation, I would like to consider the efficient estimation of various models in the presence of heteroskedasticity of unknown form. The first essay focuses on mean sqaure errors comparison of linear regression model of hetetroskedasticity with unknown form. I compare higher order properties of the efficient estimators which include the GMM-type Cragg estimator, FGLS based on series and kernel estimations. The comparison is to calculate the approximate mean square errors of estimators using the Nagar type stochastic expansion.

In the second essay, I consider the efficient estimation of partial linear regression model under heteroskedastictiy with unknown form. I propose an efficient estimator and prove it achieves Chamberlain's (1992) semi-parametric efficiency bound. The new estimator I propose has the same first order asymptotic properties as Li's (2000) estimator. My estimator has the potential advantage of analyzing the higher order asymptotics.

The third essay considers the two-step series estimation method for generated regressors problem in context of semiparametric regression model under heteroskedastictiy of unknown form. I establish the root-n consistency and asymptotic normality results of the two-step series estimators. Compared to the double kernel estimator introduced by Stengos and Yan (2001), my estimator has some computational advantage and is more accurate in the sense of the asymptotic variance. Simulation results show that the two-step series estimator outperforms the double kernel estimator in terms of mean absolute bias and mean square error.

The estimators considered in three essay involve the problem of choosing smoothing parameters. Therefore, I also demonstrate how to pick optimal smoothing parameters in each essay.

Table of Contents

List of Tables

Chapter 1

Introduction

In this dissertation, I would like to consider the efficient estimation of various models in the presence of heteroskedasticity of unknown form. In Chapter 2, I consider the linear regression model. As we know, a common problem in regression with cross-sectional data is the presence of heteroskedasticity in the residuals. Even though we could get consistently estimated standard errors to conduct hypothesis testing OLS is asymptotically less efficient than other alternative estimators such as the feasible Generalized Least Squares (FGLS) estimator and an efficient GMM estimator. One problem with these estimators is that they typically involve the choice of a smoothing parameter in order to estimate the unknown variance function or to determine the number of moment conditions to be used in GMM. Although there are rules for the smoothing parameters that guarantee efficiency less is known about the impact of the choice of smoothing parameter on the finite sample distribution of the estimators. In Chapter 2, I compare higher order properties of efficient estimators under heteroskedasticity of unknown form and in particular consider how the smoothing parameter affects the finite sample properties of the estimator. The estimators considered include the GMM-type Cragg estimator, FGLS based on series and kernel estimations. The comparison is to calculate the mean square errors (MSE) of estimators using the Nagar type stochastic expansion. It is natural to pick the optimal smoothing parameter by minimizing the approximate MSE.

Chapter 3 discusses efficient estimator of partially linear regression model with heteroskedastictiy of unknown form. I extend the Generalized Least Square estimator considered in Li (2000) to the feasible GLS estimator. I also propose an efficient estimator of partial linear regression model and prove it achieves Chamberlain's (1992) semi-parametric efficiency bound. The new estimator I propose has the same first order asymptotic properties as Li's (2000). Both estimators involve the choice of two smoothing parameters. The advantage of my estimator is that its form makes it more amenable to asymptotic expansions that could potentially be used to pick the two smoothing parameters. In the context of a Monte Carlo experiment, I compare my estimator with Li's (2000) and kernel estimators in terms of absolute mean bias and mean squared error. The simulation results show that my estimator behaves similar to Li's (2000) estimator. Also the performance of the series type estimator seems more robust to the setting of the unknown function than the kernel estimator. To overcome the problem of picking smoothing parameters in series estimation, I propose the bootstrapping approximate mean square error to choose the smoothing parameters. Using the true MSE as the benchmark, the bootstrapping method works very well and provides us a useful criterion for choosing two smoothing parameters simultaneously.

In Chapter 4, I consider the two-step series estimation method for generated regressors problem in context of semiparametric regression model under heteroskedastictiy of unknown form. I establish the $\sqrt{n}$ consistency and asymptotic normality results of the two-step series estimators. These estimators are much simpler to compute than the kernel methods proposed in the literature, such as the double kernel estimator of Stengos and Yan (2001). In addition I have allowed more general pocesses for the residual than considered in that paper. The asymptotic variance of the two-step series estimator is composed of two sources of error – one is the sampling error term and the other is from the fact that the series approximation may not necessarily equal the true function. We consider methods of inference that are robust to heteroskedasticity in the residuals. Simulation results show that my two-step series estimator outperforms the double kernel estimator in terms of mean absolute bias and mean square error. Finally we consider the use of bootstrapping the MSE for determining the two smoothing parameters and find that the method works well in practice.

Chapter 2

Higher Order MSE Comparison of Linear Regression Model Under Heteroskedasticity with Unknown Form

2.1 Introduction

A common problem in regression with cross-sectional data is the presence of heteroskedasticity in the residuals - the variance in many cases varies with the regressors. For example, in the study of family income and expenditures, it seems reasonable to expect that lower income families would spend at a rather steady rate, while the spending patterns of higher income families would be more volatile. In general the ordinary least squares (OLS) estimator of linear regression models is unbiased, consistent and asymptotically normal under heteroskdasticity. However, inference based on the variance of OLS estimator is likely to be misleading due to the use of wrong variance covariance matrix. i.e. The estimated standard errors are inconsistent. In addition, even if one were able to consistently estimate the standard errors for OLS, tests will tend to have lower than optimal power due to the relative inefficiency of the OLS estimator.

There is a large literature in econometrics that addresses the problem of obtaining valid inference for the OLS estimator. The pioneering work of

Eicker (1963) in statistics and White (1980) (referred to as a heteroskedasticity consistent covariance matrix – HCCME) in econometrics suggested a simple method for consistently estimating the OLS standard error. A nice feature of their estimator is the fact that one does not need to know the form of heteroskedasticity, and in addition there is no need for smoothing parameters as is usually required for nonparametric estimation. This has led to the wide use of their method for estimating standard errors. Although there method is popular MacKinnon and White (1985) has shown in Monte Carlo simulations and Chesher and Jewitt (1987) have shown with direct bias approximations that the HCCME tends to give standard errors that are too small in finite samples. Various authors have suggested some alternatives to HCCME that have better finite sample properties. These include some of the alternatives in MacKinnon and White (1985), including the Jackknife and various bootstrap procedures, including the Wild Bootstrap, considered in Wu (1986). Other more recent papers that address the issue of reliable inference for the OLS estimator include Chesher (1989), Chesher and Austin (1991), Cribari-Neto, Ferrari and Cordeiro (2000) and Cribari-Neto and Galvão (2003), Cribari-Neto (2003).

There is also a large theoretical literature aimed at efficient estimation with heteroskedasticity of unknown form. There are two basic approaches for achieving efficient estimation of the linear model. The first is Generalized Least Squares (GLS) with nonparametric estimation of the skedastic function. The papers by Carrol (1982) and Robinson (1987) proposed efficient estimators which assume heteroskedasticity of unknown form but have the same first order asymptotic distribution as the true GLS estimator. Carrol used kernel estimators of the variance function which involved a nonparametric regression of the squared OLS residual on the regressors. His result was based on i.i.d. observations with compactly supported regressors. Robinson used k-nearest neighbor (k-NN) method to estimate the variance function. He kept the i.i.d. observation assumption but discarded the compact support assumption. White and Stinchcombe (1991) studied nonparametric FGLS estimation by allowing the data to be dependent and heterogeneously distributed. Newey (1994) took into account of the series based FGLS estimator and gave the $\sqrt{n}$ consistency and asymptotic efficiency results. The other main approach is to use GMM with an increasing number of moment conditions. This leads to an estimator that has the form suggested in Cragg (1983). Newey (1993) showed that the Cragg (1983) estimator could become efficient if the right set of moment conditions were used in estimation. Both these papers showed in small sampling experiments that there could be efficiency gains even in small samples. Donald (1987) suggested an adjustment to the Cragg estimator that involved a bias correction to the squared residual that gives rise to even greater efficiency gains than found for the Cragg estimator.

Despite the large literature on efficient estimation the methods have not proved popular in applied work. The reasons are twofold. First, one must choose a smoothing parameter or a number of moment conditions. To date there are no justified methods for doing this. The second is that it is not clear how the various estimators compare in terms of their finite samples and how the performance varies with the choice of smoothing parameter. In this chapter we aim to address these deficiencies. In particular we investigate three estimators including the GMM-type Cragg estimator, FGLS estimator based on series and kernel estimations, which are all efficient estimator in the sense that their asymptotic variances get to the Chamberlain's (1992) semiparametric efficiency bound. To understand the finite sample properties for those estimators, we will compare the higher order approximate MSE of the estimators with a focus on terms in the MSE that depend on the smoothing parameter. This approximation is based on arguments that are similar to that of Nagar (1959). We obtain a stochastic expansion for the estimators and find the MSE of the largest terms in the stochastic expansion. Under regularity conditions the MSE of the leading terms is identical to the same expansion as those of Edgeworth approximation (see Rothenberg, 1984). The expansions allow us to compare the different estimators. In addition the MSE can be used as a criterion for choosing the smoothing parameter in much the same way as in Donald and Newey (2001) in the context of instrumental variables estimation. The idea is to estimate the approximate MSE using preliminary estimates and to use this estimated MSE as a criterion.

In section 2 we describe the various estimators. Section 3 presents and compares the approximate MSE for the three estimators we consider. We also propose the criteria to choose smoothing parameters. Section 4 discusses the extension to vector case. Section 5 is a small Monte-Carlo experiment. Section

6 applies the criteria empirically to the estimation of a wage equation. Section 7 concludes this chapter.

2.2 The Model and Estimators

The model we consider in this chapter is the linear regression model.

$$y_i = x_i'\beta + \epsilon_i, \tag{2.1}$$

for $i = 1, 2, ..., n$, where y_i is a scalar, x_i is a vector of exogenous variables. We have the usual assumption

$$E\left[\epsilon_i | x_i\right] = 0 \tag{2.2}$$

and

$$E\left[\epsilon_i^2 | x_i\right] = \sigma^2\left(x_i\right),$$

which allows the variance of the error term ϵ_i to be heteroskedastic. The matrix form of the model is

$$y = X\beta + \epsilon,$$

where $E\left[\epsilon\epsilon' | X\right] = \Sigma = \mathrm{diag}(\sigma^2\left(x_1\right), ..., \sigma^2\left(x_n\right))$. Let's first consider the ordinary least squares estimator (OLS) and its variance.

$$\begin{aligned} \widehat{\beta}_{OLS} &= \left(X'X\right)^{-1} X'y \\ Var\left(\widehat{\beta}_{OLS}\right) &= \left(X'X\right)^{-1} X'\Sigma X \left(X'X\right)^{-1} \end{aligned}$$

We know that OLS estimator is unbiased, consistent and asymptotically normal but not efficient (relative to generalized least squares estimator, GLS) under heteroskedasticity. In addition, as is well known the variance of OLS estimator is no longer $\sigma^2 (X'X)^{-1}$ so that statistical inference based on $\widehat{\sigma}^2 (X'X)^{-1}$ will be invalid.

The most common approach to dealing with heteroskedasticity is the one popularised in White (1980). This involves estimating $Var\left(\widehat{\beta}_{OLS}\right)$ by replacing Σ with $\widehat{\Sigma}$ =diag(e_i^2) , where e_i^2 is the square of the ith least squares residual. This gives rise to,

$$V^W_{\widehat{\beta}_{OLS}} = (X'X)^{-1} X'\widehat{\Sigma}X (X'X)^{-1},$$

which is known as HC0. However, lots of simulation studies reveal the downward bias of HC0, see MacKinnon and White (1985). There are several corrections based on White's setup and they turned out to be more efficient than OLS estimator using White's estimator. Hinkley (1977) did the degree of freedom correction to $V^W_{\widehat{\beta}_{OLS}}$ as below.

$$V^H_{\widehat{\beta}_{OLS}} = \left[\frac{n}{n-d}\right] (X'X)^{-1} X'\widehat{\Sigma}X (X'X)^{-1},$$

which is known as HC1. d is the dimension of covariates X. HC1 inflates the residual by factor $\sqrt{n/n-d}$. Horn, Horn, and Duncan (1975) used an "almost unbiased" estimator[1] for $\sigma^2 (x_i)$.

$$V^{HD}_{\widehat{\beta}_{OLS}} = (X'X)^{-1} X'\text{diag}\left(\frac{e_i^2}{1-h_{ii}}\right) X (X'X)^{-1},$$

[1]In the homoskedastic error case, we can show

$$E\left[e_i^2\right] = \sigma^2 [1-h_{ii}] \neq \sigma^2.$$

Even though $e_i^2/[1-h_{ii}]$ is not unbiased estimator of σ_i^2 under heteroskedasitcity, it is a less biased estimator than e_i^2.

where h_{ii} being the ith diagonal element of $X\left(X'X\right)^{-1}X'$. $V_{\widehat{\beta}_{OLS}}^{HD}$ is known as HC2, which inflates the magnitude of HC0 by factor $(1-h_{ii})$. MacKinnon and White (1985) proposed the HC3 as follows.

$$V_{\widehat{\beta}_{OLS}}^{MW}=\frac{n-1}{n}\left(X'X\right)^{-1}X'\left[\operatorname{diag}\left(e_i^{*2}\right)-\frac{1}{n}e^*e^{*\prime}\right]X\left(X'X\right)^{-1},$$

where $e_i^*=e_i/\left(1-h_{ii}\right)$ and e^* is a $n\times 1$ column vecter of e_i^*. The Monte-Carlo results of MacKinnon and White showed that OLS variance covariance estimator based on ordinary jackknife method[2] will be more efficient that based on other correction methods. Lots of simulation work show that HC3 outperforms other variants of HC0. Long and Ervin (2000) strongly suggest statistical software developers to add HC3 to their programs. Chesher and Austin (1991) found the impact of leverage points on the finite sample behavior. Monte Carlo by Cribari-Neto and Zarkos (2001) showed the presence of high leverage points in the design matrix plays an important role in the small sample properties of the various HCCMEs. And the influence of these high leverage point is more decisive than heteroskedasticity itself. Based on this fact, Cribari-Neto (2003) proposed a modified estimator called HC4.

$$V_{\widehat{\beta}_{OLS}}^{CN}=\left(X'X\right)^{-1}X'\operatorname{diag}\left(\frac{e_i^2}{\left(1-h_{ii}\right)^{\delta_i}}\right)X\left(X'X\right)^{-1},$$

where $\delta_i=\min\left\{4,nh_{ii}/\sum_{i=1}^{n}h_{ii}\right\}$.

[2]Actually the three corrections based on White (1980) considered in MacKinnon and White (1985) could be all thought of as jackknife based estimator. White (1980) is called *infinitesimal jackknife method*. Hinkley (1977) is called *weighted jackknife estimator*. MacKinnon and White (1985) is the direct applicatin of the jackknife covariance estimator from the idea of Efron (1982).

The other problem with the OLS is its inefficiency, which impacts the power of statistical tests and the accuracy of confidence intervals. As is well known the efficient estimator is the GLS or Aitken estimator which has the form,

$$\widehat{\beta}_{Aitken} = \left(X'\Sigma^{-1}X\right)^{-1} X'\Sigma^{-1}y$$

with variance given by $\left(X'\Sigma^{-1}X\right)^{-1}$. The problem with the estimator is that Σ is typically unknown. One may try to parameterize this heteroskedasticity and obtain feasible GLS. However, if the assumed functional form of heteroskedasticity is incorrect then one is not necessarily any better off than one would have been with the OLS estimator – indeed it is possible that one could be even more inefficient than the OLS estimator in terms of variance. To overcome the specification of the functional form of heteroskedasticity, the remedy is to adopt nonparametric methods (series, kernel, local linear, k-NN,...etc) to estimate the unknown variance covariance matrix Σ in the GLS estimator. Carrol (1982) and Robinson (1987) are two leading examples of this type of work. In this paper we consider two methods for this nonparametric part. The first is based on series estimation of the variance, denoted $\widehat{\beta}_{Series}$ and the second is the kernel based FGLS estimator ($\widehat{\beta}_{Kernel}$). The former is based on power series, splines or Fourier series. The latter could be on the basis of Nadaraya-Watson, k-NN or local polynomials. Let's denote two estimators as follows.

$$\begin{aligned} \widehat{\beta}_{Series} &= \left(X'\widehat{\Sigma}_S^{-1}X\right)^{-1} X'\widehat{\Sigma}_S^{-1}y \\ \widehat{\beta}_{Kernel} &= \left(X'\widehat{\Sigma}_K^{-1}X\right)^{-1} X'\widehat{\Sigma}_K^{-1}y. \end{aligned}$$

Their corresponding variance-covariance matrices will be

$$
\begin{aligned}
Var_{\widehat{\beta}_{Series}} &= \left(X'\widehat{\Sigma}_S^{-1}X\right)^{-1} \\
Var_{\widehat{\beta}_{Kernel}} &= \left(X'\widehat{\Sigma}_S^{-1}X\right)^{-1}.
\end{aligned}
$$

Both approaches are sometimes called semiparametric methods in the sense that they do the nonparametric estimation in the first step and the parametric procedure in the second step

An alternative route to efficient estimation is using an estimator due to Cragg (1983). This estimator is based on using a GMM type of approach[3] based on the moment condition (2.2). Specicially if we let q_i be a vector of functions of x_i (including x_i) then the condition (2.2) implies the following set of unconditional moment restrictions,

$$
E(q_i\,(y_i - x_i\beta)) = 0
$$

Then the optimal weighted 2-step GMM procedure (with OLS used in the first step) based on this condition would involve solving the following problem,

$$
\begin{aligned}
\widehat{\beta}_{GMM} &= \arg\min_{\beta}\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\beta)\,q(x_i)\right)'\left(\frac{1}{n}\sum_{i=1}^{n}e_i^2 q(x_i)\,q(x_i)'\right)^{-1} \\
&\quad\times\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\beta)\,q(x_i)\right),
\end{aligned}
$$

where e_i is the OLS residual. This results in the estimator that is,

$$
\widehat{\beta}_{Cragg} = \left(X'Q\left(Q'\hat{\Sigma}Q\right)^{-1}Q'X\right)^{-1}X'Q\left(Q'\hat{\Sigma}Q\right)^{-1}Q'y.
$$

[3]The GMM interpretation of the estimator was not in the original Cragg (1983) article. The GMM interpretation was noted in Newey (1993).

with variance given by,

$$Var_{\widehat{\beta}_{Cragg}} = \left(X'Q\left(Q'\hat{\Sigma}Q\right)^{-1}Q'X\right)^{-1},$$

where $Q = (X : \Psi)$ is the matrix whose rows are q_i and Ψ being a $n \times G$ matrix of auxilliary variables (or instruments) which consists of moments of the variables in X, except for those moments already contained in X. Using similar logic to that in White (1980) this estimator behaves asymptotically like the infeasible estimator given by,

$$\widehat{\beta}_{GMM} = \left(X'Q\left(Q'\Sigma Q\right)^{-1}Q'X\right)^{-1}X'Q\left(Q'\Sigma Q\right)^{-1}Q'y$$

Newey (1993) noted that this estimator can be efficient with appropriate choice of auxilliary functions – in particular functions that have good approximation properties. To see this we note that from Chamberlain (1987) the optimal unconditional moment restriction based on the conditional moment restriction (2.2) is,

$$\begin{aligned} E\left(\psi^*\left(x_i\right)\left(y_i - x_i\beta\right)\right) &= 0. \\ \psi^*\left(x_i\right) &= \frac{x_i}{\sigma_i^2} \end{aligned}$$

Performing GMM with instrument $\psi^*(x_i)$ would lead directly to the GLS estimator. The FGLS estimators described earlier attempt to estimate this by plugging in an estimate of the variance function directly. The GMM estimator of Cragg (1983) and Newey (1993) implicitly uses a moment condition that is an estimated version of,

$$\begin{aligned} E\left(\psi\left(x_i\right)\left(y_i - x_i\beta\right)\right) &= 0. \\ \psi\left(x_i\right) &= q_i'\left(Q'\Sigma Q\right)^{-1} Q'X. \end{aligned}$$

The reason that this can be as efficient as the direct FGLS method is that $\psi\left(x_i\right)$ can eventually approximate $\psi^*\left(x_i\right)$. To see this note that the fitted value in the regression

$$\frac{x_i}{\sigma_i} = \sigma_i q_i'\pi + error$$

is

$$\sigma_i q_i'\left(Q'\Sigma Q\right)^{-1} Q'X = \sigma_i \psi\left(x_i\right).$$

Assuming that q_i is sufficiently in terms of its approximation properties then,

$$\begin{aligned} \min_{\pi} \sup_{i} \left| \frac{x_i}{\sigma_i} - \sigma_i q_i'\pi \right| &\leq \min_{\pi} \sup_{i} \sigma_i \left| \frac{x_i}{\sigma_i^2} - q_i'\pi \right| \\ &\leq C \min_{\pi} \sup_{i} \left| \frac{x_i}{\sigma_i^2} - q_i'\pi \right| \\ &= O\left(K^{-\alpha}\right), \end{aligned}$$

where α is the smoothness index of x_i/σ_i^2. Then since it is the case that $\sigma_i q_i'\left(Q'\Sigma Q\right)^{-1} Q'X$ will eventually approximate x_i/σ_i it should be the case that $q_i'\left(Q'\Sigma Q\right)^{-1} Q'X$ will eventually approximate x_i/σ_i^2. Thus one can attain the efficiency bound by doing standard GMM using an increasing number of moment conditions.

A problem with all of the efficient methods of estimation is that they all require the choice of some smoothing parameter – a bandwidth, or a number of polynomial terms or a number of moment conditions. In general the estimators will have finite sample properties that will depend on the smoothing parameter and it is not clear if one estimator may be more sensitive to the choice of smoothing parameter than another. Through the use of higher order MSE approximations this chapter explores the extent to which the estimators depend on the smoothing parameter so that some comparisons across different methods can be made. In addition we examine the use of the MSE and its estimate as a means for selecting a smoothing parameter in practice.

2.3 The Mean Square Error

The way we calculate approximate MSE is similar to Nagar (1959). We compute the approximate MSE for each estimator we considered. That is to calculate

$$MSE_j = E\left[n\left(\widehat{\beta}_j - \beta\right)\left(\widehat{\beta}_j - \beta\right)'\right],$$

where $j = \{Cragg, Series, Kernel\}$.

2.3.1 Regularity Conditions

Some regularity conditions have to be specified to obtain the results. Let $\|A\| = tr\left(A'A\right)^{1/2}$ denote the usual Euclidean norm for a matrix A.

Assumption 2.3.1. *$\{y_i, x_i\}$ are i.i.d., $E\left[\epsilon_i|x_i\right] = 0$, $E\left[\epsilon_i^2|x_i\right] = \sigma^2\left(x_i\right) = \sigma_i^2 < \infty$ and $E\left[\epsilon_i^3|x_i\right] = 0$.*

This assumption puts the bounded second moment condition and zero third moments condition, which is important in our derivation. Assumption 1 also allows for heteroskedasticity.

Assumption 2.3.2. *For every K there is a nonsingular constant matrix B such that for $q^K(x_i) = Bq_i$; (i) the smallest eigenvalue of $E\left[q^K(x_i)' q^K(x_i)\right]$ is bounded away from zero uniformly in K and; (ii) there is a sequence of constants $\zeta(K)$ satisfying $\sup_{x \in X} \left\|q^K(x)\right\| \leq \zeta(K)$ and $K = K(n)$ such that $\zeta(K)^2 K/n \to 0$ as $n \to \infty$.*

Assumption 2.3.2 is usually imposed on series estimator. See Newey (1997). This assumption normalizes approximating function. Part (i) bounds the second moment matrix away from singularity. Part (ii) controls the convergence rate of the series estimator.

Assumption 2.3.3. *For an integer $d \geq 0$ there exits α and β_K such that $\left|g_0 - q^{K\prime}\beta_K\right|_d = O\left(K^{-\alpha}\right)$ as $K \to \infty$, where $g_0 = E[y|x]$ denotes the true conditional expectation and g denotes some function of x. If $d = 0$, the integer $\alpha = s/r$ is the smoothness index of g_0, where s is the number of continuous derivatives of g_0 and r is the dimension of x.*

Assumption 2.3.3 specifies the bound of the approximation error. Since we do not need the derivative of approximation error, what we need is the case of $d = 0$. In our notation, we will have $1/n \sum_{i=1}^{n} \left(x_i/\sigma_i - q^K(x_i)/\sigma_i)' \pi\right)^2 = O\left(K^{-2\alpha}\right) = O\left(\|f\|^2\right).$

Assumption 2.3.4. *The constant scalar $\zeta(K)$ in Assumption 2.3.2 satisfies (i) $\zeta(K) K/\sqrt{n} \to 0$ and; (ii) $\zeta(K)\sqrt{K}/\sqrt{n} \to 0$.*

Assumption 2.3.4 is imposed to ensure the small order terms converge to zero sufficiently fast. Actually, part (i) implies part (ii).

2.3.2 MSE Formulae

We put all the proofs of the propositions in the Appendix. First, we give the approximate MSE for Cragg estimator.

Proposition 2.3.1. *If Assumptions 2.3.1-2.3.3 are satisfied, then MSE for Cragg estimator is*

$$\begin{aligned}
& \Omega^{*-1}\left[MSE_{Cragg}\right]\Omega^{*-1} \\
= \ & \Omega^{*-1}E\left[n\left(\widehat{\beta}_{Cragg}-\beta\right)\left(\widehat{\beta}_{Cragg}-\beta\right)'\right]\Omega^{*-1} \\
= \ & \left[\frac{X^{*\prime}X^{*}}{n}+\frac{X^{*\prime}\left(I-P^{*}\right)X^{*}}{n}-\frac{1}{n}tr\left(\Xi_1\right)+\frac{5}{n}tr\left(\Xi_2\right)\right]+o\left(\frac{K}{n}\right),
\end{aligned}$$

where

$$\begin{aligned}
\Omega^{*} &= \left(\frac{X^{*\prime}X^{*}}{n}\right)^{-1} \\
\Xi_1 &= \left(Q^{*\prime}Q^{*}\right)^{-1}\sum_i \kappa_i^{*}x_i^{*2}Q_i^{*}Q_i^{*\prime} \\
\Xi_2 &= \left(Q^{*\prime}Q^{*}\right)^{-1}\sum_i x_i^{*2}Q_i^{*}Q_i^{*\prime} \\
\kappa_i^{*} &= E\left[\epsilon_i^{*4}|x_i^{*}\right] \\
x_i^{*} &= \frac{x_i}{\sigma\left(x_i\right)} \\
\epsilon_i^{*} &= \frac{\epsilon_i}{\sigma\left(x_i\right)} \\
Q_i^{*} &= \frac{q^{K}(x_i)}{\sigma\left(x_i\right)}
\end{aligned}$$

Next result is for series based FGLS estimator.

Proposition 2.3.2. *If Assumptions 2.3.1-2.3.4 are satisfied, then MSE for series based FGLS estimator is*

$$\begin{aligned}
& \Omega^{*-1}\left[MSE_{Series}\right]\Omega^{*-1} \\
= \; & \Omega^{*-1}E\left[n\left(\widehat{\beta}_{Series}-\beta\right)\left(\widehat{\beta}_{Series}-\beta\right)'\right]\Omega^{*-1} \\
= \; & \frac{X^{*\prime}X^{*}}{n}+\frac{1}{n}\sum_{i}\frac{x_i^{*2}}{\sigma_i^4}\left[\overline{\sigma}_i^2-\sigma_i^2\right]^2-2\frac{1}{n}\sum_{i}\frac{x_i^{*2}}{\sigma_i^4}P_{ii}\kappa_i \\
& +\frac{1}{n}\sum_{i\neq j}\frac{x_i^{*2}}{\sigma_i^4}P_{ij}^2\kappa_j+6\frac{1}{n}\sum_{i}x_i^{*2}P_{ii}-\frac{1}{n}\sum_{i\neq j}\frac{\sigma_j^4}{\sigma_i^4}x_i^{*2}P_{ij}^2+o\left(\frac{K}{n}\right),
\end{aligned}$$

where

$$\begin{aligned}
P_{ij} &= Q_i\left(Q'Q\right)^{-1}Q_j \\
\kappa_i &= E\left[\epsilon_i^4|x_i\right] \\
\overline{\sigma}_i^2 &= q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\sigma\mathbf{2} \\
\epsilon_i^* &= \frac{\epsilon_i}{\sigma\left(x_i\right)}.
\end{aligned}$$

Now we present the approximate MSE for kernel based FGLS estimator. The method of estimating σ_i^2 is using local polynomial regression[4] considered in Linton (1996).

Proposition 2.3.3. *If all the Assumptions of nonparametric estimators in*

[4]There several advantages of adopting local polynomial regression method such as no boundary effects and design adaptation.

Linton (1996) are satisfied, then MSE for kernel based FGLS estimator is

$$
\begin{aligned}
& \Omega^{*-1}\left[MSE_{Kernel}\right]\Omega^{*-1} \\
= & \Omega^{*-1}E\left[n\left(\widehat{\beta}_{Kernel}-\beta\right)\left(\widehat{\beta}_{Kernel}-\beta\right)'\right]\Omega^{*-1} \\
= & \frac{X^{*\prime}X^{*}}{n}+h^{4}\mathcal{B}+n^{-1}h^{-d}\mathcal{V}+o_{p}\left(n^{-2\mu}\right) \\
= & \frac{X^{*\prime}X^{*}}{n}+h^{4}\left[\Gamma_{2}-\Gamma_{1}\Omega^{*-1}\Gamma_{1}\right]+n^{-1}h^{-d}\left[\left(\kappa_{3}^{2}+\kappa_{4}-1\right)\Omega_{n}^{*}\right]+o_{p}\left(n^{-2\mu}\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\Gamma_{1} &= n^{-1}\sum_{i=1}^{n}x_{i}x_{i}'\sigma_{i}^{-4}\widetilde{B}_{i} \\
\Gamma_{2} &= n^{-1}\sum_{i=1}^{n}x_{i}x_{i}'\sigma_{i}^{-6}\widetilde{B}_{i}^{2} \\
\widetilde{B}_{i} &= h^{-2}\sum_{j\neq i}\left(\sigma_{j}^{2}-\sigma_{i}^{2}\right)w_{ij} \\
\Omega_{n}^{*} &= n^{-1}\sum_{i=1}^{n}x_{i}x_{i}'\sigma_{i}^{-2}\left(nh^{d}\sum_{j\neq i}w_{ij}^{2}\right) \\
\mu &= 2/\left(d+4\right) \\
E\left[\epsilon_{i}^{3}\right] &= \kappa_{3}\sigma_{i}^{3} \\
E\left[\epsilon_{i}^{4}\right] &= \kappa_{4}\sigma_{i}^{4}.
\end{aligned}
$$

First of all, notice that the leading term in the approximate MSE is $X^{*\prime}X^{*}/n$, which is common to all the of our approximate MSEs. It is natural to omit this term in choosing the smoothing parameter since $X^{*\prime}X^{*}/n$ does not involve K or h. The additional terms are the largest (in order) of those that increase or decrease with the smoothing parameters – K in the case of

Cragg and series based FGLS estimator and h in the case of kernel based FGLS. Note that matrix Q plays the role of approximating functions both in Cragg estimator and series based FGLS estimator. This offers the basis for our comparison. The difference is that for Cragg estimator we use series estimators to form unconditional moment restriction. While for series based estimator, series estimators provides nonparametric estimation of unknown parameter σ_i^2.

We observe that there is common structure of approximate MSE we list above. The approximate MSEs are all composed of a leading term which is nothing to do with smoothing parameter, a term decreasing in smoothing parameter, and terms increasing in smoothing parameter. For Cragg estimator, $X^{*\prime}X^*/n$ is the common leading term, the term $X^{*\prime}\left(I-P^*\right)X^*/n$ is decreasing with K and the term $5tr\left(\Xi_2\right)/n-tr\left(\Xi_1\right)/n$ is increasing with K provided the kurtosis is not too high.[5] For series based FGLS estimator, $X^{*\prime}X^*/n$ is the common leading term, the term $\sum_i\left(x_i^{*2}/\sigma_i^4\right)\left[\overline{\sigma}_i^2-\sigma_i^2\right]^2/n$ is decreasing with K and the rest of terms is increasing with K provided the kurtosis is not too high. For kernel based FGLS estimator, $X^{*\prime}X^*/n$ is still the common leading term, the term $n^{-1}h^{-d}\left[\left(\kappa_3^2+\kappa_4-1\right)\Omega_n^*\right]$ is decreasing with h and the term $h^4\left[\Gamma_2-\Gamma_1\Omega^{*-1}\Gamma_1\right]$ is increasing with h.

Looking at the second term of order $O\left(K^{-\alpha_1}\right)$ in Proposition 2.3.1, it

[5] We don't want the kurtosis of the error terms is too high. In this case we will tend to pick the numbers of instruments as many as possible. Please see the simulation results in later section. Also note that our result is along the line with the asymptotic expansion with respect to GMM minimum distance estimation in Konenker et al (1994).

could be rewritten as

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{x_i}{\sigma_i}-\sigma_i q_i'\overline{\pi}_{1K}\right]^2,$$

where $\overline{\pi}_{1K}=\left(Q'\Sigma Q\right)^{-1}Q'x$. It is just the residuals sum of square of regressing x_i/σ_i on $\sigma_i q_i'$. If we pick the approximating function q properly, the residuals sum of square will shrink to zero at rate $O\left(K^{-\alpha_1}\right)$. As we illustrated in previous section. $q_i'\overline{\pi}_{1K}$ will estimate the optimal instrument x_i/σ_i^2 quite well. The second term of order $O\left(K^{-\alpha_2}\right)$ in Proposition 2.3.2 could be expressed as

$$\frac{1}{n}\sum_{i=1}^{n}\frac{x_i^2}{\sigma_i^6}\left[\sigma_i^2-q_i'\overline{\pi}_{2K}\right]^2,$$

where $\overline{\pi}_{2K}=\left(Q'\Sigma Q\right)^{-1}Q'\sigma\mathbf{2}$. The regression implication is the weighted residuals sum of square of regression σ_i^2 on Q_i'. If we pick the approximating function q properly, the weighted residuals sum of square will shrink to zero at rate $O\left(K^{-\alpha_2}\right)$. That is $q_i'\overline{\pi}_{2K}$ could estimate σ_i^2 very well.

In general, we could not determine the relative size of the approximate MSE among three estimators without further assumptions. It is quite interesting to compare the size of approximate MSE of Cragg estimator and series based FGLS estimator due to the similar form of the terms in Proposition 2.3.1 and 2.3.2. We conduct a simulation study to evaluate the three major terms in both estimators. See section in simulation result. We consider different distributions of regressors, errors and forms of heteroskedasticity and find that the numerical evaluation of the approximate MSE of Cragg estimator is quite similar to that of series based FGLS estimator. i.e. We cannot uniformly rank these two estimators.

Proposition 2.3.2 is to come up with data based methods for selecting smoothing parameter K. For series based FGLS we look at a method based on the MSE of the resulting FGLS estimator,

$$E\left[\left(\widehat{\beta}_{Series} - \beta\right)^2\right].$$

An alternative is to minimize MSE in the estimation of the variance function (using cross validation). The regression is (assume one knows the residual)

$$\begin{aligned} \epsilon_i^2 &= \sigma_i^2 + v_i \qquad (2.3) \\ &= Q_i'\overline{\pi}_{2K} + \left(\sigma_i^2 - Q_i'\overline{\pi}_{2K}\right) + v_i. \end{aligned}$$

The series estimator for $\sigma 2$ is

$$\begin{aligned} \widehat{\sigma 2} &= Q\left(Q'Q\right)^{-1}\epsilon 2 \\ &= \sigma 2 - (I - P)\,\sigma 2 + Pv, \end{aligned}$$

so the sample MSE is

$$\begin{aligned} &\frac{1}{n}E\left[\left(\widehat{\sigma 2} - \sigma 2\right)'\left(\widehat{\sigma 2} - \sigma 2\right)\right] \\ =\ &\frac{1}{n}\sigma 2'\left(I - P\right)\sigma 2 + \frac{1}{n}E\left[v'Pv|x\right] \\ =\ &\frac{1}{n}\sum_{i=1}^{n}\left[\sigma_i^2 - Q_i'\overline{\pi}_{2K}\right]^2 + \frac{1}{n}\sum_{i=1}^{n}\left(\kappa_4 - 1\right)\sigma_i^4 P_{ii}. \end{aligned}$$

However, (2.3) is not the actual regression done for FGLS since ϵ_i^2 is unknown. What we actually did is to regress e_i^2 on Q_i',

$$\widetilde{\sigma 2} = Q\left(Q'Q\right)^{-1}e2.$$

The following proposition shows that the MSE of $\widehat{\sigma 2}$ is essentially asymptotically equivalent to that of $\widetilde{\sigma 2}$ in the sense that the difference is of order that is smaller than the order of the MSE.

Proposition 2.3.4.

$$MSE\left(\widetilde{\sigma 2}\right) = MSE\left(\widehat{\sigma 2}\right) + o(\frac{1}{n}\sigma 2' \left(I - P\right) \sigma 2 + \frac{1}{n}\left(\kappa_4 - 1\right) \sum_{i=1}^{n} \sigma_i^4 P_{ii}).$$

One can see from this result that the optimal smoothing parameter from the point of view of estimating the variance function will generally be different from the one that is optimal from the point of view of estimating the regression parameter.

The next proposition, under an assumption of homoskedasticity $\sigma^2\left(x_i\right) = \sigma_i^2 = \sigma^2$, gives an equivalence result between the Cragg estimator and the series based FGLS estimator.

Proposition 2.3.5. *Under homoskedasticity, we have*

$$\begin{aligned} & \Omega^{*-1}\left[MSE_{Cragg}\right]\Omega^{*-1} \\ = \; & \Omega^{*-1}\left[MSE_{Series}\right]\Omega^{*-1} \\ = \; & \frac{1}{n\sigma^2}\sum_i x_i^2 + \frac{5}{n\sigma^2}\sum_i x_i^2 P_{ii} - \frac{1}{n\sigma^6}\sum_i x_i^2 P_{ii}\kappa_i. \end{aligned}$$

Supprisingly, under homoskedasticity the higher order MSE of Cragg estimator is the same as that of series based FGLS estimator. Our simulation, which computes the MSE according to Proposition 2.3.1 and 2.3.2 respectively confirms this finding. Unlike the case of heteroskedasticity, the approximate MSE under homoskedasticity does not depend upon the term which is decreasing with K. The term of increasing with K could be factored out by $5\sigma^4 - \kappa_i = \sigma^4 (5 - \kappa_i/\sigma^4)$. Ine the case of conditional normality of ϵ_i $\kappa_i/\sigma^4 = 3$ so that the terms in approximate MSE (other than the leading term) are strictly increasing with K. This fact is also confirmed by our simulation. However, if the disturbances are much thicker tailed than normal (say $\kappa_i/\sigma^4 > 5$), we may expect the approximate MSE will decrease with K. We can further combine the second and third terms in Proposition 2.3.5 in the case where the conditional fourth moment is constant (ie $\kappa_i = \kappa_4$),

$$\frac{1}{n\sigma^2} (5 - \kappa_4) \sum_i x_i^2 P_{ii}. \tag{2.4}$$

This expression shows exactly how the degree of kurtosis affects the way in which the smoothing parameter impacts the higher order MSE. Note that the term $\sum_i x_i^2 P_{ii} = O(K)$.

Notice that the MSE of Cragg and series based FGLS estimator will involve the projection matrix $P = Q' (Q'Q)^{-1} Q$. However, the MSE of kernel based estimator involves the kernel weight w_{ij}. It arises the difficulty comparing the approximate MSEs between various estimators. To make comparison easily under homoskedasticity, we assume third moment condition holds for kernel based FGLS estimator. Here is the result.

Proposition 2.3.6. *Under homoskedasticity and $E[\epsilon_i^3 | x_i] = 0$, we have*

$$\Omega^{*-1} [MSE_{Kernel}] \Omega^{*-1} = \frac{1}{n\sigma^2} \sum_{i=1}^{n} x_i^2 + \frac{1}{n\sigma^6} \sum_{i \neq j} x_i^2 w_{ij}^2 \kappa_j - \frac{1}{n\sigma^2} \sum_{i \neq j} x_i^2 w_{ij}^2.$$

It is still difficult to compare the approximate MSE of kernel based FGLS estimator with that of other two estimators even though we've imposed the zero third moments condition and assumed homoskedasticity. However, we could combine the second and third terms in Proposition 2.3.6 as followed.

$$\frac{1}{n\sigma^2}\left(\kappa_4 - 1\right)\sum_{i\neq j} x_i^2 w_{ij}^2. \tag{2.5}$$

Now the term in (2.5) looks like term in (2.4).

2.3.3 Choosing the Smoothing Parameters

Since we've established the approximate MSEs for all three estimators, the next question is how to pick the smoothing parameter K or h to minimize approximate MSE. As we mentioned in previous section, the term which does not grow with smoothing parameter has been omitted. The selection criteria for Cragg estimator is

$$S_{Cragg}\left(K\right) = \frac{1}{n}\widehat{X}^{*\prime}\left(I - \widehat{P}^*\right)\widehat{X}^* - \frac{1}{n}\sum_i \widehat{\kappa}_i^* \widehat{x}_i^{*2}\widehat{P}_{ii}^* + \frac{5}{n}\sum_i \widehat{x}_i^{*2}\widehat{P}_{ii}^*,$$

where the "hat" means the estimated variable. The unknown parameters in $S_{Cragg}\left(K\right)$ are σ_i^2 and forth moment κ_i^*. The criteria for series based FGLS estimator is

$$\begin{aligned} S_{FGLS-Series}\left(K\right) &= \frac{1}{n}\sum_i \frac{\widehat{x}_i^{*2}}{\widehat{\sigma}_i^4}\left[\widehat{\widehat{\sigma}}_i^2 - \widehat{\sigma}_i^2\right]^2 - 2\frac{1}{n}\sum_i \frac{\widehat{x}_i^{*2}}{\widehat{\sigma}_i^4}P_{ii}\widehat{\kappa}_i \\ &+ \frac{1}{n}\sum_{i\neq j}\frac{\widehat{x}_i^{*2}}{\widehat{\sigma}_i^4}P_{ij}^2\widehat{\kappa}_j + 6\frac{1}{n}\sum_i \widehat{x}_i^{*2}P_{ii} - \frac{1}{n}\sum_{i\neq j}\frac{\widehat{\sigma}_j^4}{\widehat{\sigma}_i^4}\widehat{x}_i^{*2}P_{ij}^2. \end{aligned}$$

As we discussed in previous subsection, we could pick the smoothing parameter K by minimizing the estimated MSE in variance function. The criteria is as follows.

$$S_{Variance}(K) = \frac{1}{n}\widehat{\sigma 2}'(I-P)\widehat{\sigma 2} + \frac{1}{n}(\widehat{\kappa}_4 - 1)\sum_{i=1}^{n}\widehat{\sigma}_i^4 P_{ii}$$

Another method for choosing K in the variance function is to calculate the cross validation (CV) criteria of the fit. In this case, the CV is

$$\frac{1}{n}\sum_{i=1}^{n}\left[e_i^2 - \widehat{e}_{-i}^2\right]^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{1-P_{ii}}\widehat{u}_i\right]^2,$$

where e_i^2 is the first stage OLS squared residuals, $\widehat{e}_{-i}^2$ is the delete-one version of the fitted value and $\widehat{u}_i$ is corresponding to the redidual of fitted value $\widehat{e}_i^2$.[6] The criteria for kernel based FGLS estimator is

$$S_{FGLS-Kernel}(h) = h^4\widehat{\mathcal{B}} + n^{-1}h^{-d}\widehat{\mathcal{V}}.$$

The second order effect could be minimized by setting h so that $h^4 = \gamma n^{-1}h^{-d}$ or $h = \gamma n^{-1/(4+d)}$. Minimizing h with respect to $S_{FGLS-Kernel}(h)$ gives us the optimal value of γ being

$$\widehat{\gamma} = \left[\frac{d\widehat{\mathcal{V}}}{4\widehat{\mathcal{B}}}\right]^{\frac{1}{(4+d)}}.$$

Of course we have to replace the unknown σ_i^2 and the fourth moment κ_4 in $\widehat{\mathcal{B}}$ and $\widehat{\mathcal{V}}$ by the estimated counterparts. Now the optimal bandwidth is

$$\widehat{h} = \widehat{\gamma}n^{-1/(4+d)} = \left[\frac{d\widehat{\mathcal{V}}}{4\widehat{\mathcal{B}}}\right]^{\frac{1}{(4+d)}} n^{-1/(4+d)}.$$

[6] Note that

$$\begin{aligned}\widehat{e}_{-i}^2 &= q_i'\left(q_{-i}'q_{-i}\right)^{-1}e2\\ \widehat{e}_i^2 &= q_i'\left(q'q\right)^{-1}e2\\ \widehat{u}_i &= e_i^2 - \widehat{e}_i^2.\end{aligned}$$

2.4 Extension to Vector Case

Our calculation for Proposition 2.3.1 to 2.3.3 are based on scalar case. It is quite natural to extend our results to general vector case. Let's consider the scalar standardized quantities:

$$T = \sqrt{n}\lambda' \left(\widehat{\beta} - \beta\right), \tag{2.6}$$

where λ is any $d \times 1$ vector. $\lambda = (0, ..., 0, 1, 0, ...0)'$ is an important special case. The MSE of (2.6) which is the linear combination of coefficients will be

$$MSE(T) = n\lambda' E\left[\left(\widehat{\beta} - \beta\right)\left(\widehat{\beta} - \beta\right)\right]\lambda.$$

The Proposition 2.3.1 to 2.3.3 could be summarized by the following corollaries. The proofs are staightforward.

Corollary 2.4.1. *If Assumptions 2.3.1-2.3.3 are satisfied, then MSE for Cragg estimator is*

$$\begin{aligned}
& \Omega^{*-1}\left[MSE_{Cragg}\right]\Omega^{*-1} \\
= \; & \Omega^{*-1} E\left[n\left(\widehat{\beta}_{Cragg} - \beta\right)\left(\widehat{\beta}_{Cragg} - \beta\right)'\right]\Omega^{*-1} \\
= \; & \left[\frac{\lambda' X^{*\prime} X^{*} \lambda}{n} + \frac{\lambda' X^{*\prime}\left(I - P^{*}\right) X^{*}\lambda}{n} - \frac{1}{n} tr\left(\Xi_1\right) + \frac{5}{n} tr\left(\Xi_2\right)\right] + o\left(\frac{K}{n}\right).
\end{aligned}$$

Corollary 2.4.2. *If Assumptions 2.3.1-2.3.4 are satisfied, then MSE for series* *based FGLS estimator is*

$$
\begin{aligned}
& \Omega^{*-1}\left[MSE_{Series}\right]\Omega^{*-1} \\
= \; & \Omega^{*-1}E\left[n\left(\widehat{\beta}_{Series}-\beta\right)\left(\widehat{\beta}_{Series}-\beta\right)'\right]\Omega^{*-1} \\
= \; & \frac{\lambda' X^{*\prime} X^{*}\lambda}{n}+\frac{1}{n}\sum_{i}\frac{\lambda' x_i^{*} x_i^{*\prime}\lambda}{\sigma_i^4}\left[\overline{\sigma}_i^2-\sigma_i^2\right]^2-2\frac{1}{n}\sum_{i}\frac{\lambda' x_i^{*} x_i^{*\prime}\lambda}{\sigma_i^4}P_{ii}\kappa_i \\
& +\frac{1}{n}\sum_{i\neq j}\frac{\lambda' x_i^{*} x_i^{*\prime}\lambda}{\sigma_i^4}P_{ij}^2\kappa_j+6\frac{1}{n}\sum_{i}\lambda' x_i^{*} x_i^{*\prime}\lambda P_{ii}-\frac{1}{n}\sum_{i\neq j}\frac{\sigma_j^4}{\sigma_i^4}\lambda' x_i^{*} x_i^{*\prime}\lambda P_{ij}^2+o\left(\frac{K}{n}\right).
\end{aligned}
$$

Corollary 2.4.3. *If all the Assumptions of nonparametric estimators in Linton (1996) are satisfied, then MSE for kernel based FGLS estimator is*

$$
\begin{aligned}
& \Omega^{*-1}\left[MSE_{Kernel}\right]\Omega^{*-1} \\
= \; & \Omega^{*-1}E\left[n\left(\widehat{\beta}_{Kernel}-\beta\right)\left(\widehat{\beta}_{Kernel}-\beta\right)'\right]\Omega^{*-1} \\
= \; & \frac{\lambda' X^{*\prime} X^{*}\lambda}{n}+h^4\mathcal{B}+n^{-1}h^{-d}\mathcal{V}+o_p\left(n^{-2\mu}\right) \\
= \; & \frac{\lambda' X^{*\prime} X^{*}\lambda}{n}+h^4\lambda'\left[\Gamma_2-\Gamma_1\Omega^{*-1}\Gamma_1\right]\lambda+n^{-1}h^{-d}\left[\left(\kappa_3^2+\kappa_4-1\right)\lambda'\Omega_n^{*}\lambda\right]+o_p\left(n^{-2\mu}\right).
\end{aligned}
$$

2.5 Monte Carlo Experiment

The first part of the simulation we compare the performance of variants of White estimators, Cragg estimator and semiparametric FGLS estimators under heteroskedasticity of unknown form. The criteria of comparing different HCCME estimators are the accuracy of the inference and MSEs of the estimators. To compare the accuracy of the inference, we list the empirical size of a test and see how far it is from the nominal size. The estimators we consider in the simulation include the OLS, FGLS, and Cragg estimators for β. Among the three types of the estimators, we use different estimators for the unknown parameters in the corresponding variance covariances.

The second part of the simulation we compare the approximate MSE for Cragg, series based FGLS estimators and series based variance estimator in variance function numerically.

2.5.1 Simulation Design

In the simulation design we adapt the model in Cragg (1983) and consider the case of only one regressor.

$$y_i = \beta x_i + \epsilon_i, \tag{2.7}$$

where x_i is a log-normal random variable and ϵ_i is a normally distributed random variable with mean zero and variance given by

$$\sigma_i^2 = \gamma_1 + \gamma_2 x_i + \gamma_3 x_i^2, \tag{2.8}$$

where the vector $\gamma = (\gamma_1, \gamma_2, \gamma_3)'$ could represent the magnitude of heteroskedasticity.

The estimator of β could be OLS ($\widehat{\beta}_{OLS}$), FGLS ($\widehat{\beta}_{FGLS}$) and Cragg's estimator ($\widehat{\beta}_{Cragg}$). To get the accuracy of a test, we compare the nominal and actual size of t-statistics. For each estimator we calculate the t-statistics as follows.

$$t = \frac{\widehat{\beta} - \beta}{s.e.\left(\widehat{\beta}\right)},$$

where $\widehat{\beta}$ is one of three major types of estimators of β and $s.e.\left(\widehat{\beta}\right)$ is computed from potential variance-covariance estimators of β. For OLS estimator, the corresponding variance covariance estimators are based on the true Σ, White's estimator $\widehat{\Sigma}_W$ and MacKinnon-White's estimator $\widehat{\Sigma}_{MW}$. For FGLS estimator, the corresponding variance covariance estimators are based on the true Σ, White's estimator $\widehat{\Sigma}_W$, MacKinnon-White's estimator $\widehat{\Sigma}_{MW}$, series estimator $\widehat{\Sigma}_S$, and kernel estimator $\widehat{\Sigma}_K$. For Cragg estimator, the corresponding variance covariance estimators are based on the true Σ, White's estimator $\widehat{\Sigma}_W$ and MacKinnon-White's estimator $\widehat{\Sigma}_{MW}$. Note that the OLS estimator of β is nothing to do with the estimation of Σ. However, FGLS and Cragg estimators of β and the variance covariance matrices both depend upon the estimation of Σ.

To evaluate the finite sample properties of different estimators, the sample size T is set to $(50, 100, 150)$. The number of replications is 1000. We also assume the true value $\beta = 1$. The parameters related to heteroskedasticity are specified as $\gamma = (0, .2, .5)'$ which corresponds the case of severe heteroskedasticity noted by Cragg.

To evaluate the approximate MSE for Cragg and series based FGLS estimators[7], we generate covariate from uniform distribution and consider three types of error structure with different kurtosis– normal, uniform and logistic.

[7] Here we evaluate Cragg and series based FGLS estimator and exclude kernel based FGLS estimator because the former two estimators both involve the approximating function in terms of the number K.

Also the sample size will increase to 200.

2.5.2 Simulation Results

2.5.2.1 Finite Sample Properties

The basic results are summarized in the following tables. Table D.1 states the size performance of 4 versions of OLS estimators. It is obvious that if we adopt the true variance matrix in t-test, the empirical size is very close to the nominal size. The column $\widehat{\beta}_{OLS}^{B}$ deviates from the nominal size significantly due to the use of the wrong variance matrix in t-test. The use of the White correction improves things to a degree but there are still large size distortions with rejection rates in some cases above 20% when the nominal size is 5%. MacKinnon and White's estimator[8] provides more accurate inference, although even for that approach there are some size distortions. As expected, things improve for the White and MW approaches as the sample size grows while the size distortion grows with the sample size when using the OLS standard error. In this example the the OLS standard errors are too small – this need not always be the case as we see when we consider an empirical example in

[8]There are two verisons of MacKinnon and White estimators. They are

$$
\begin{aligned}
V_1^{MW} &= \frac{n-1}{n}\left(X'X\right)^{-1}X'\left[\operatorname{diag}\left(u_i^{*2}\right)-\frac{1}{n}u^*u^{*\prime}\right]X\left(X'X\right)^{-1} \\
V_2^{MW} &= \left(X'X\right)^{-1}X'\left[\operatorname{diag}\left(u_i^{*2}\right)\right]X\left(X'X\right)^{-1}.
\end{aligned}
$$

Andrew (1991) pointed out that V_2^{MW} is analogue of the leave-one-out cross-validation estimator of the covariance matrix. Our simualtion shows tiny difference between the two versions although Andrew (1991) asserted that V_2^{MW} outperforms V_1^{MW} in terms of size distortion.

the next section.

Table D.2 reports the size distortion of different FGLS estimators. As in the case of OLS estimator, using the true variance matrix gives us almost accurate size. Series based FGLS estimator performs well compared with kernel based FGLS estimator[9]. The size distortion of series based FGLS estimator tends to decrease as sample size increases. In general, semiparametric FGLS estimators provide more accurate inference than White's correction.

The size comparison of Cragg estimator is summarized by Table D.3. True variance matrix performs well as usual. The Cragg estimator using MW performs markedly better than it does when using White. Both methods perform better than the FGLS approach. Indeed among the various approaches to inference including OLS with corrected standard errors using the Cragg estimator with MW provides the best approach in the sense of having size close to nominal size.

The MSEs comparison of FGLS estimators is listed in Table D.4. First of all, we can see the MSEs are decreasing with the sample size in all estimators we consider. By theory, we know that GLS estimator is the most efficient estimator due to correcting heteroskedasticity. The sample MSEs verify this fact. Note that the case of $K = 1$ in FGLS estimator corresponds the OLS estimator which has the highest MSE within expectation. The Cragg estima-

[9]For the kernel based FGLS estimator ,we report the local linear estimation method. We also implement the case of Nadaray-Watson kernel estimation method, which turns out to be worse than local linear estimation method.

tor using true variance matrix has higher MSE than that of GLS estimator because we know that GLS is the most efficient estimator. See Table D.5. When we increase the number of instruments ($K = 5$), the MSE of Cragg estimator (.1207) approaches to GLS estimator (.1202) quickly. This confirms the theoretical argument that Cragg estimator will get the efficiency with an increase in the number of moment conditions. Also note that the feasible Cragg estimators have smaller MSEs than those of FGLS estimators.

2.5.2.2 Evaluation Simulation

In this subsection we conduct the simulation to evaluate the size of the terms in approximate MSE of Cragg and series based FGLS estimator. The first result is to evaluate the terms in Proposition 2.3.1 and 2.3.2 numerically under homoskedasticity. See Table D.6. This result numerically verifies our Proposition 2.3.4 which says the MSE of Cragg and series based FGLS estimator should be the same under homoskedasticity.

For the numerical evaluation of Cragg estimator, we consider three errors structures with different kurtosis. See Table D.7. The kurtosis for normal, uniform and logistic distributions are 3, 1.8 and 12.6 respectively. We can see that optimal number of instrument is 2 in Model 1 and 2. In model 3 we have excess kurtosis and tend to pick as many instruments as possible. This fact could be seen from Proposition 2.3.1.

Table D.8 lists the evaluation result of series based FGLS estimator. We observe that the magnitude of approximate MSE of series based FGLS estimator is very close to that of Cragg estimator for three different data generating process. The optimal instrument is 3 in Model 1 and 2 except using criteria to pick K in model 2. It turns out to be 1 under uniform error assumption. For model 3, excess kurtosis makes approximate MSE grow with K, which is consistent with Proposition 2.3.2.

2.6 Empirical Application

We adopt the empirical example of estimating wage equation from Wooldridge (2000). For the population of people in the work force in 1976, the data set includes wages for 526 individuals, ages, work experience, education level,...,etc. We use log-wages of individuals as the dependent variable and work education level, experience and experience^2 as independent variables. The simple wage regression is

$$\log wage = \beta_0 + \beta_1 educ + \beta_2 \exp er + \beta_3 \exp er^2 + u.$$

Simply looking at the residuals plot or plot of log-wage against work experience may not tell the magnitude of heteroskedasticity.[10] We perform the general White test and Breusch-Pagan tests and find that we could not reject the assumption of homoskedastic errors at 5% (even lower) significance level. See Table D.10.

[10]Heteroskedasticity could be introduced by any independent variables in arbitrary form. We conduct Breusch-Pagan tests using various covariates and find that work experience is relevant to the variance of the error term.

To implement the estimation method we proposed in this paper, we need to pick the smoothing parameters. For Cragg estimator, series based FGLS and variance function, we report the corresponding optimal K using the estimated criteria proposed in section 3. From the Table D.11, we can see that the optimal number of moments for the Cragg estimator will be $K = 7$. For series based FGLS we find that $K = 4$ is optimal whereas using cross validation method on the estimation of the variance function leads to an optimal K that is 6.

After determining the smoothing parameters, one can now do the estimation. The estimation results using different estimation methods considered in this paper are summarized in Table D.12. Although the differences are not large the Cragg and FGLS estimators are more precise with smaller standard errors than the OLS estimator using the more reliable MW approach to estimating the variance of OLS. It is interesting to note that the OLS standard error is smaller than the MW standard error for education while the opposite is true for the experience variables.

2.7 Concluding Remarks

In this chapter we compare the higher order approximate MSE of GMM-type Cragg estimator, series based FGLS estimator and kernel based FGLS estimator through Nagar type stochastic expansion. According to our calculation, it is hard to uniformly rank the three estimators although they have interesting implications under homoskedasticity. Instead, we could numerically compare the sieve based estimators which depend on the same smoothing parameter. Accroding to the numerical evaluation, the Cragg estimator generally has quite similar approximate MSE to that of the series based FGLS estimator. We also derive the criteria for selecting the smoothing parameters. However, the result of estimated version of approximate MSE is mixed. We take empirical example of wage equation to illustrate the estimation procedure we propose in this chapter.

Chapter 3

Efficient Estimation of Partial Linear Model Under Heteroskedasticity with Unknown Form

3.1 Introduction

Nonparametric methods have become quite popular in economics in recent decades. While nonparametric regression is flexible in recovering the true shape of the regression curve without specifying a parametric family for the data, it has some disadvantages. The most fundamental problem is the *curse of dimensionality.* To overcome this problem, a useful approach is to remain nonparametric about certain variables but take a parametric form about the variables of interest. A popular method for doing this is to specify the regression model as,

$$y_i = x_i'\beta + g\left(z_i\right) + u_i, \tag{3.1}$$

where $g\left(\cdot\right)$ is an unknown nonparametric function and is usually highly dimensional. This model is called *partial linear* or *semilinear* regression model. Engle, Granger, Rice and Weiss (1986) applied this model to study the effect of weather on electricity demand. The partial linear specification also appears in various sample selection models such as Newey, Powell and Walker (1990), and Lee, Rosenzweig and Pitt (1992).

Previous work of estimation on partial linear model includes papers by Engle, Granger, Rice and Weiss (1986), Wahba (1984), Heckman (1986), Rice (1985), Chen (1988), Speckman (1988), Robinson (1988), Linton (1995), Donald and Newey (1994), Hong and Cheng (1999), and Li (2000). Engle, Granger, Rice and Weiss (1986), Wahba (1984) proposed the partial spline smoothing approach. The method was further studied by Rice (1985) and Heckman (1986). Rice (1985) obtained the asymptotic bias of a partial spline smoothing estimator of β and showed that this approach can not attain the Berry-Esseen rate $\sqrt{n}$ for the estimator of β unless x and z are uncorrelated or the unknown nonparametric component $g(\cdot)$ is undersmoothed.[1] Chen (1988) proposed a kind of piecewise polynomial approximation to $g(\cdot)$ and the convergence rate of $\widehat{\beta}$ is shown to be $\sqrt{n}$ consistent with the smallest possible variance even when x and z are dependent. Speckman (1988) considered kernel smoothing and proved that the parametric rate of $\widehat{\beta}$ is attainable for the usual "optimal" bandwidth choice under which the optimal nonparametric convergence rate for the estimation of $g(\cdot)$. Robinson (1988) constructed a feasible least squares estimator of β using Nadaraya-Watson kernel estimators of $E[y|z]$ and $E[x|z]$. He proved that $\widehat{\beta}$ is $\sqrt{n}$ consistent and asymptotically normal. Linton (1995) proposed the local polynomial regression method to estimate $E[y|z]$ and $E[x|z]$.[2] He established the $\sqrt{n}$ consistent estimator of

[1]It means that the $\sqrt{n}$ parametric rate for estimation of β and optimal nonparametric rate for estimation of $g(\cdot)$ could not be attained simultaneously in the partial spline smoothing approach.

[2]Linton (1995) adopted local polynomial regression estimator instead of Robinson's (1988) Nadaraya-Watson kernel estimator due to the nice properties of local polynomial

β and found that it is second order optimal using a second order approximation of $\sqrt{n}\left(\widehat{\beta}-\beta\right)$. Donald and Newey (1994) used series approximation to the the unknown function $g(\cdot)$. They showed that the estimator was $\sqrt{n}$ consistent estimator and asymptotic normality under weak conditions.[3] Hong and Cheng (1999) revisited the kernel smoothing method and showed that normal approximation rate of β is achieved only when bandwidth h is chosen with rate $n^{-1/4}$ instead of the usual "optimal" bandwidth rate $n^{-1/5}$.[4] Li (2000) considered the additive partial linear model using series estimation method and proved the estimator of finite dimensional parameter β reaches the semiparametric efficiency bound under homoskedasticity.[5] An alternative approach to partial linear model is to avoid the nonparametric estimation procedure. Yatchew (1997) proposed a differencing estimator to remove the effect of unknown function $g(\cdot)$. The differencing estimator is in general not efficient. Yatchew proposed the generalized method of differencing to achieve the same asymptotic efficiency bound of Robinson (1988).

In this paper we add to the literature on partially linear models by considering a new form of estimator that is efficient when there is heteroskedasticity of unknown form. This involves not only dealing with the unknown

regression estimator which is design adaptive and is able to correct the boundary bias problem.

[3]The condition is weaker than previous studies in that the modulus of continuity of $g(z)$ and $E[x|z]$ be higher than 1/4 the dimension of z and that the number of terms be chosen appropriately. Also the covariates z could be not only multidimensional but also be discrete.

[4]The faster convergence rate of tending to zero of bandwidth than optimal rate is callded "undersmoothed".

[5]Li's result was based on homoskedastic errors. However, Chamberlain's (1992) semiparametric efficiency bound can allow for conditional heteroskedasticity.

function g but also an unknown variance function which is allowed to depend on all of the regressors. The paper by Li (2000) proposed a feasible GLS type estimator that is efficient in the case of heteroskedasticity. His estimator relies on the use of the series method and involves a weighted least squares estimation of a regression model which involves the linear component $x'\beta$ and an approximation to the nonparameteric component g. Although the estimator is efficient its form makes higher order expansions difficult and hence a common method for picking a smoothing parameter may be difficult to use in practice. The alternative estimator proposed here has a simpler form that makes a higher order expansion possible and also does not rely on the use of the series based approximation to the function g. In our case we do weighting of the regression of $y - E[y|z]$ on $x - E[x|z]$. The method could potentially be implemented using any method for estimating conditional expectations and the form of the estimator will potentially make higher order expansion possible and help in the problem of coming up with a method for picking the two smoothing parameters.

The remainder of this paper is organized as follows. In section 2 we describe the model and estimation technique in this paper. The first order asymptotic results for different efficient estimators are provided in Section 3. In Section 4 we conduct a small scale Monte Carlo experiment. Section 5 concludes this paper.

3.2 The Model

Consider a partial linear regression model in (3.1)

$$y_i = x_i'\beta + g(z_i) + u_i, \tag{3.2}$$

where the covariates x_i and z_i are of dimension r and q respectively, $\beta = (\beta_1, ..., \beta_r)'$ is a $r \times 1$ vector of unknown parameter, and $g(\cdot)$ is an unknown function. Of course, we could extend (3.2) to the additive partially linear regression model by setting

$$g(z_i) = g_1(z_{1i}) + g_2(z_{2i}) + ... + g_L(z_{Li}),$$

where $g_l(z_{li})$ is scalar, z_{li} is of dimension q_l ($q_l \geq 1, l = 1, 2, ..., L$). For simplicity, we assume $L = 1, q_l = q$. In matrix form, we could write (3.2) as

$$y = x\beta + g(z) + u. \tag{3.3}$$

The identification condition for β in (3.3) is stated below.

Assumption 3.2.1. *(Identification) To identify the partial linear regression model in (3.3), we need* $E\left[(x - E[x|z])'(x - E[x|z])\right]$ *to be positive definite.*

Literally, we need that random variable x is not fully contained in z. To understand the identification condition, taking expectation conditional on z with respect to (3.3) gives

$$E[y|z] = E[x|z]\beta + g(z) + E[u|z]. \tag{3.4}$$

Substracting (3.4) from (3.3) gives

$$y - E[y|z] = [x - E[x|z]]\beta + u - E[u|z]. \tag{3.5}$$

It is obvious that identification of β requires the full rank of $x - E[x|z]$. In the context of a sample selection model where z would represent the variables that affect selection we can have a situation where z is a linear function of some variable that appear in x provided that there is also a variable that predicts selection but does not appear in x, see Newey, Powell and Walker (1990). In other instances where z just represents some other variables we require that x and z not overlap.

Before providing regularity conditions we discuss the estimation methods to be used in this paper. The estimation strategy of model (3.3) recommended in Robinson (1988) is to estimate $E[y|z]$ and $E[x|z]$ nonparametrically (Nadaraya-Watson type kernel method) and regress $y - E[y|z]$ on $x - E[x|z]$ to get estimate of β. However, when using a series estimator and not using weighted least squares this is equivalent to regression y on x and the series basis functions,.

$$p_K(z) = (p_{1K}(z), p_{2K}(z), ..., p_{KK}(z))',$$

Here we let p_K be the $n \times K$ matrix with ith row $p_{Ki} = p_K(z_i)$. The projection matrix is $Q = p_K(p_K' p_K)^{-1} p_K'$. Then the partialled out series based method of estimating the parameter β, as first suggested in Donald and Newey (1994) is given by,

$$\begin{aligned}\widehat{\beta} &= \left[(x-Qx)'(x-Qx)\right]^{-1}(x-Qx)(y-Qy) \\ &= \left[x'(I-Q)x\right]^{-1}x'(I-Q)y.\end{aligned}$$

So, what we do here is different from Robinson (1988) in that we employ series method to estimate $E[y|z]$ and $E[x|z]$ instead of kernel method. Now, the unknown function $g(\cdot)$ could be estimated by $\widehat{g} = p_K(z)'\widehat{\pi}$, where $\widehat{\pi}$ is given by

$$\widehat{\pi} = \left(p_K' p_K\right)^{-1} p_K' \left(y - x\widehat{\beta}\right).$$

Li (2000) verified that under homoskedasticity assumption $\widehat{\beta}$ will be semiparametric efficient in the sense that the inverse of the asymptotic variance of $\sqrt{n}\left(\widehat{\beta}-\beta\right)$ achieves Chamberlain's (1992) semiparametric efficiency bound. He also established $\sqrt{n}$-consistency of $\widehat{\beta}$ under conditional heteroskedasticity. However, if the disturbances are heteroskedastic, $\widehat{\beta}$ will in general not be semiparametric efficient.[6] Therefore, Li (2000) suggested a GLS type estimator by regressing y_i/σ_i on $\left[x_i/\sigma_i, p_K(z_i)'/\sigma_i\right]$, where $\sigma_i^2 = E(u_i^2|x_i, z_i) = \sigma^2(z_i)$. We let $\widehat{\beta}_{GLS}$ denote the corresponding estimator of β and note that it has the form,

$$\widehat{\beta}_{GLS} = \left[x^{*\prime}(I-Q^*)x^*\right]^{-1}\left[x^{*\prime}(I-Q^*)y^*\right], \tag{3.6}$$

[6]The estimator $\widehat{\beta}$ is said to be "local efficient" according to Li (2000) in that its efficiency is attained when some restrictions are satisfied. Here, it means that assumption of homoskedasticity is satisfied.

where

$$\begin{aligned} x^* &= (x_1/\sigma_1, x_2/\sigma_2, ..., x_n/\sigma_n)' \\ Q^* &= p_K^* \left(p_K^{*\prime} p_K^*\right)^{-1} p_K^{*\prime} \\ p_K^* &= (p_{K1}/\sigma_1, p_{K2}/\sigma_2, ..., p_{Kn}/\sigma_n)'. \end{aligned}$$

Without providing any proof, he asserted the method will produce a semi-parametric efficient estimator of β. We will prove this fact in Theorem 3.3.1. To implement this one would use an estimate of the variance function σ_i^2 which would then be plugged into the matrix Q^*. This could be done by using a preliminary consistent estimator of the model, such as $\widehat{\beta}$, then regressing the squared OLS residuals onto x and z using some nonparametric regression method. In this case we would regress the square of,

$$\widehat{u}_i = y_i - x_i'\widehat{\beta} - \widehat{g}(z_i)$$

on x and z. Robinson (1987) suggested a k-nearest neighbor method for doing this in a linear model, but in principle one could also use a series based method. A difficulty with this approach is that one would like to know how the estimator depends on the smoothing parameters and in this case there will be two. One smoothing parameter relates to the number of functions used to approximate g (say K) as well as the number of functions used to approximate the variance function (say H). On other hand, one could consider to partial out the variable $P^K(z)$ in the first stage and then do the weighted least square in the second stage.

The estimator proposed in this paper differs in that we do a weighted regression of $y - E(y|x)$ on $x - E(x|z)$ using weights that are the inverse of the variance. All conditional expectations are estimated via series regression so that if the variances were known we would have the following GLS estimator,

$$\widetilde{\beta}_{GLS} = \left[x'(I-Q)\Sigma^{-1}(I-Q)x\right]^{-1}\left[x'(I-Q)\Sigma^{-1}(I-Q)y\right] \tag{3.7}$$

Although the two estimators look different, we could prove that our estimator has the same first order result as Li's (2000) estimator. The reason that weighting after removing the mean works is that essentially we are estimating the model,

$$y_i - E\left[y_i|z_i\right] = \left[x_i - E\left[x_i|z_i\right]\right]\beta + u_i$$

by weighted least squares with weights that are the inverse of the variances. This then is equivalent to doing the regression,

$$\frac{y_i - E\left[y_i|z_i\right]}{\sigma_i} = \frac{\left[x_i - E\left[x_i|z_i\right]\right]}{\sigma_i}\beta + \frac{u_i}{\sigma_i}$$

and the residual in this regression is conditionally homoskedastic. It means our estimator inherits the semiparametric efficiency. At this moment, we are considering the infeasible GLS procedure. It is natural to extend our results to feasible GLS by incorporating the estimated version of σ_i^2 using the H series functions of the variable z_i. The feasible versions of Li (2000) and our estimators using these estimated variances will be denoted $\tilde{\beta}_{FGLS}$ and $\hat{\beta}_{FGLS}$.

3.3 First Order Asymptotics

The following assumptions are needed to establish our results.

Assumption 3.3.1. *(i)* $(y_i, x_i, z_i), i = 1, ..., n$ *are i.i.d. (independent and identically distributed); the support of* (x, z) *is a compact subset of* $\mathcal{R}^{q+r}$*; (ii)* $E[u_i|x_i, z_i] = 0$*,* $E[u_i^2|x_i, z_i] = \sigma^2(z_i) = \sigma_i^2$ *and* u_i *has bounded fourth moments; (iii) Let* $x_i = E(x_i|z_i) + \epsilon_i = h(z_i) + \epsilon_i$*,* $E(\epsilon_i|z_i) = 0$*, and* $E(\epsilon_i^2|z_i)$ *is bounded away from* ∞*; (iv) All of* $h(z_i)$ *and* $\sigma^2(z_i)$ *are bounded functions on the support of* (x, z) *.*

Assumption 3.3.1 (i) is quite standard in regression model. Assumption 3.3.1 (ii) allows for conditional heteroskedasticity. Assumption 3.3.1 (iii) assumes that x_i is function of z_i plus a random element that has a finite variance. These conditions plus smoothness, discussed below in Assumption 3.3.3 will make it possible to estimate the various unknown functions.

Note that Assumption 3.3.1 (ii) imposes the restriction that the conditional varianc $E[u_i^2|x_i, z_i]$ depends only on z_i.but not x_i. The reason is to get semiparametric efficient estimator in the sense that the variance attains Chamberlain's (1992) semiparametric efficiency bound. If we assume $E[u_i^2|x_i, z_i] = \sigma^2(x_i, z_i)$, the semiparametric efficiency bound could be expressed as

$$SPEB = \inf_{\xi} E\left\{[x_i - \xi(z_i)][x_i - \xi(z_i)]' / \sigma^2(x_i, z_i)\right\}. \tag{3.8}$$

On the other hand, the asymptotic variance of our estimator is

$$E\left\{[x_i - h(z_i)][x_i - h(z_i)]' / \sigma^2(x_i, z_i)\right\}. \tag{3.9}$$

In general the minimizer ξ in (3.8) will not be equal to $h(z_i)$. However, imposing the assumption of $E[u_i^2|x_i, z_i] = \sigma^2(z_i)$, our estimator will be semiparametric efficient. i.e. The SPEB in (3.8) will reduce to (3.9).

Assumption 3.3.2. *For every K there is a nonsigular constant matrix B such that for $P^K(z) = Bp^K(z)$; (i) the smallest eigenvalue of $E\left[p^K(z_i)\, p^K(z_i)'\right]$ is bounded away from zero uniformly in K and; (ii) there is a sequence of constants $\zeta(K)$ satisfying $\sup_{z\in\mathcal{Z}} \left\|p^K(z)\right\| \leq \zeta(K)$ and $K = K(n)$ such that $\zeta(K)^2 K/n \to 0$ as $n \to \infty$, where $\mathcal{Z}$ is the support of z.*

Assumption 3.3.2 is usually imposed on series estimators. See Newey (1997). This assumption normalizes approximating function. Part (i) bounds the second moment matrix away from singularity. Part (ii) controls the convergence rate of the series estimator.

Assumption 3.3.3. *(i) For $f = g$ or $f = h$, there exits some π_f and $\alpha_f\,(>0)$ such that $\sup_{z\in\mathcal{Z}} \left|f(z) - P^K(z)'\,\pi_f\right| = O\left(K^{-\alpha_f}\right)$ as $K \to \infty$; also, $\sqrt{n}K^{-\alpha} \to 0$ as $n \to \infty$. (ii) For σ^2, there exits some π_{σ^2} and $\alpha\,(>0)$ such that $\sup_{(x,z)\in\mathcal{X}\times\mathcal{Z}} \left|\sigma^2(x,z) - P^H(x,z)'\,\pi_{\sigma^2}\right| = O\left(H^{-\alpha_{\sigma^2}}\right)$ as $H \to \infty$; also, $\sqrt{n}H^{-\alpha_{\sigma^2}} \to 0$ as $n \to \infty$.*

Assumption 3.3.3 specifies the bound of the approximation error when we approximate unknown function g or h as well as the variance function as will be required in order to implement a FGLS estimator of the partially linear model. Note that there are two smoothing parameters K and H that are required for estimation. The following theorem gives the first order asymptotic distribution of $\widehat{\beta}_{GLS}$, which is the infeasible GLS of Li's (2000) estimator in (3.6). All the proofs in this section are put in the Appendix.

Theorem 3.3.1. *Define* $x_i = h(z_i) + \epsilon_i$ *and assume that* $E[\epsilon_i \epsilon_i' / \sigma_i^2]$ *is positive definite, then under Assumptions 3.2.1-3.3.3, we have*

$$\sqrt{n}\left(\widehat{\beta}_{GLS} - \beta\right) \xrightarrow{d} N\left(0, J_0^{-1}\right),$$

where $J_0 = E\{[x_i - h(z_i)][Var(u_i|x_i, z_i)][x_i - h(z_i)]\} = E[\epsilon_i \epsilon_i' / \sigma_i^2]$ *is Chamberlain's semi-parametric efficiency bound.*

Proof. The proof is given in Appendix. □

The next theorem states that our estimator of infeasible version. $\widetilde{\beta}_{GLS}$ in (3.7) is semiparametric efficient.

Theorem 3.3.2. *Define* $x_i = h(z_i) + \epsilon_i$ *and assume that* $E[\epsilon_i \epsilon_i' / \sigma_i^2]$ *is positive definite, then under Assumptions 3.2.1-3.3.3, we have*

$$\sqrt{n}\left(\widetilde{\beta}_{GLS} - \beta\right) \xrightarrow{d} N\left(0, J_0^{-1}\right),$$

where $J_0 = E\{[x_i - h(z_i)][Var(u_i|x_i, z_i)][x_i - h(z_i)]\} = E[\epsilon_i \epsilon_i' / \sigma_i^2]$ *is Chamberlain's semi-parametric efficiency bound.*

Proof. The proof is given in Appendix. □

The results for the feasible GLS estimators are below and show that one can also achieve the efficiency bound even without knowing the variance function provided that the assumptions stated above are satisfied. Note that since we are estimating the variance function the conditions pertaining to this are now required.

Theorem 3.3.3. *Define* $x_i = h(z_i) + \epsilon_i$ *and assume that* $E[\epsilon_i \epsilon_i' / \sigma_i^2]$ *is positive definite, then under Assumptions 3.2.1-3.3.3, we have*

$$\sqrt{n}\left(\widehat{\beta}_{FGLS} - \beta\right) \xrightarrow{d} N\left(0, J_0^{-1}\right),$$

where $J_0 = E\{[x_i - h(z_i)][Var(u_i|x_i, z_i)][x_i - h(z_i)]\} = E[\epsilon_i \epsilon_i' / \sigma_i^2]$ *is Chamberlain's semi-parametric efficiency bound.*

Proof. The proof is given in Appendix. □

Theorem 3.3.4 proves that our feasible estimator is semiparametric efficient.

Theorem 3.3.4. *Define* $x_i = h(z_i) + \epsilon_i$ *and assume that* $E[\epsilon_i \epsilon_i' / \sigma_i^2]$ *is positive definite, then under Assumptions 3.2.1-3.3.3, we have*

$$\sqrt{n}\left(\widetilde{\beta}_{FGLS} - \beta\right) \xrightarrow{d} N\left(0, J_0^{-1}\right),$$

where $J_0 = E\{[x_i - h(z_i)][Var(u_i|x_i, z_i)][x_i - h(z_i)]\} = E[\epsilon_i \epsilon_i' / \sigma_i^2]$ *is Chamberlain's semi-parametric efficiency bound.*

Proof. The proof is given in Appendix. □

If one would be interested in comparing finite sample properties of competing estimators of partial linear model, higher order expansion will be needed. Although our estimator is the same as Li's (2000) estimator in first order sense, our estimator has advantage in doing asymptotic expansion. To see this, let us consider Li's estimator first.

$$\sqrt{n}\left(\widehat{\beta}_{FGLS}-\beta\right)=\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\left[x_i-\widetilde{x}_i^{*}\right]^2}{\widehat{\sigma}_i^2}\right]^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\left[x_i-\widetilde{x}_i^{*}\right]\left[g_i-\widetilde{g}_i^{*}+u_i-\widetilde{u}_i^{*}\right]}{\widehat{\sigma}_i^2}\right], \tag{3.10}$$

where the "*" means the projection matrix in forming $\widetilde{x}$ involves the normalized version of the approximating functons. As we expand equation (3.10), one has to expand not only the term $\widehat{\sigma}_i^2$ in the general denominator in (3.10) but also the implicit $\widehat{\sigma}_i^2$ in normalized approximating function – this will complicate the expansions. As for our estimator,

$$\sqrt{n}\left(\widetilde{\beta}_{FGLS}-\beta\right)=\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\left[x_i-\widetilde{x}_i\right]^2}{\widehat{\sigma}_i^2}\right]^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\left[x_i-\widetilde{x}_i\right]\left[g_i-\widetilde{g}_i+u_i-\widetilde{u}_i\right]}{\widehat{\sigma}_i^2}\right]. \tag{3.11}$$

The projection matrix in forming $\widetilde{x}$ does not involve the normalized random factor, $\widehat{\sigma}_i^2$ and should be more amenable to higher order expansion.

Although the series estimator we propose in this paper is quite easy to implement, we still need to pick the number of approximating functions. We need to pick a smoothing parameter K for the approximation of $E(y|z)$ and $E(x|z)$ and H for the approximation of the variance function $\sigma^2(z)$.[7] The

[7] We have implicitly assumed that the same number of functions is used to approximate $E(y|z)$ and all elements of $E(x|z)$. In principle they could all be different.

natural problem is how to choose the optimal smoothing parameter according to some higher order approximate MSE. Although we do not present results in this direction the discussion above suggests that it may be easier to do for our estimator. The method we consider below is to use, without formal justification, the bootstrap to approximate the MSE and to pick smoothing parameters to minimize the estimated MSE. Using the bootstrap-based procedure for selecting the moment condition has been discussed in Inoue (2003).

3.4 Monte Carlo Experiment

3.4.1 Simulation Design

We briefly state the simulation design in this section.

$$
\begin{aligned}
y_i &= \beta \cdot x_i + \exp(z_i) + u_i \cdot \sigma^2 (x_i, z_i) \qquad (3.12)\\
x_i &= d \cdot z_i + v_i,\\
z_i &= \frac{i}{n}, i = 1, ...n\\
u_i &\sim N(0,1), \; v_i \sim N(0,1)\\
d &= 10, \; \beta = 1
\end{aligned}
$$

Here is the setting for heteroskedasticity.

$$\sigma^2 (x_i, z_i) = .3x_i^2 + 0 \cdot z_i^2$$

We consider 1000 replications for sample sizes of $n = 100$, 200 and 400. The Mean Absolute Bias (BIAS) and Mean Square Error (MSE) are computed for four possible estimators, which include preliminary estimator[8], our estimator, Li's estimator and kernel estimator. Those estimators will be described below.

3.4.2 Estimators of β

3.4.2.1 Li's estimator

1. Regress y_i on $p_K(z_i)$ and x_i on $p_K(z_i)$. Obtain residuals $y_i - \widetilde{y}_i$ and $x_i - \widetilde{x}_i$, where

$$\begin{aligned} y_i - \widetilde{y}_i &= y_i - p_K(z_i)\left(p_K(z)' p_K(z_i)\right)^{-1} p_K(z)' y \\ x_i - \widetilde{x}_i &= x_i - p_K(z_i)\left(p_K(z)' p_K(z_i)\right)^{-1} p_K(z)' x. \end{aligned}$$

2. Regress $y_i - \widetilde{y}_i$ on $x_i - \widetilde{x}_i$. Obtain preliminary estimator of β, b_0.

3. Estimate g by $\widehat{g}(z_i) = p_K(z_i)\left(p_K(z)' p_K(z_i)\right)^{-1} p_K(z)' [[y - xb_0]$.

4. Estimate u_i by $\widehat{u}_i = y_i - x_i b_0 - \widehat{g}(z_i)$.

5. Estimate $\sigma_i^2(z_i)$ by

$$\widehat{\sigma}_i^2(z_i) = p_H(z_i)\left(p_H'(z)\, p_H(z)\right)^{-1} p_H(z)' \widehat{u2},$$

   where $\widehat{u2}$ is a column vector of $\widehat{u}_i^2$.

6. Regess $y_i/\widehat{\sigma}_i$ on $x_i/\widehat{\sigma}_i$, $p_K(z_i)/\widehat{\sigma}_i$ to get Li's (2000) semiparametric efficient estimator of β.

[8] Preliminary estimator represents the series estimator of partial linear model without considering heteroskedasticity.

3.4.2.2 New estimator

1. Regress y_i on $p_{K1}(z_i)$ and x_i on $p_{K1}(z_i)$. Obtain residuals $y_i - \widetilde{y}_i$ and $x_i - \widetilde{x}_i$, where

$$\begin{aligned} y_i - \widetilde{y}_i &= y_i - p_K(z_i)\left(p_K(z)' p_K(z_i)\right)^{-1} p_K(z)' y \\ x_i - \widetilde{x}_i &= x_i - p_K(z_i)\left(p_K(z)' p_K(z_i)\right)^{-1} p_K(z)' x. \end{aligned}$$

2. Regress $y_i - \widetilde{y}_i$ on $x_i - \widetilde{x}_i$. Obtain preliminary estimator of β, b_0.

3. Estimate g by $\widehat{g}(z_i) = p_K(z_i)\left(p_K(z)' p_K(z_i)\right)^{-1} p_K(z)' [[y - xb_0]$.

4. Estimate u_i by $\widehat{u}_i = y_i - x_i b_0 - \widehat{g}(z_i)$.

5. Estimate $\sigma_i^2(z_i)$ by

$$\widehat{\sigma}_i^2(z_i) = p_H(z_i)\left(p_H'(z) p_H(z)\right)^{-1} p_H(z)' \widehat{u2},$$

6. Regress $(y_i - \widetilde{y}_i)/\widehat{\sigma}_i$ on $(x_i - \widetilde{x}_i)/\widehat{\sigma}_i$ to get our semiparametric efficient estimator of β.

3.4.2.3 Kernel estimator

1. Compute the kernel (or local linear) estimators of $E(y|z)$ and $E(x|z)$ by

$$\begin{aligned} \widehat{y}_i &= \frac{\overline{y}_i}{\widehat{f}_i} = \frac{\frac{1}{Nh}\sum_{j=1}^N y_j K\left(\frac{Z_i - Z_j}{h}\right)}{\frac{1}{Nh}\sum_{j=1}^N K\left(\frac{Z_i - Z_j}{h}\right)} \\ \widehat{x}_i &= \frac{\overline{x}_i}{\widehat{f}_i} = \frac{\frac{1}{Nh}\sum_{j=1}^N x_j K\left(\frac{Z_i - Z_j}{h}\right)}{\frac{1}{Nh}\sum_{j=1}^N K\left(\frac{Z_i - Z_j}{h}\right)}, \end{aligned}$$

where $K(\cdot)$ is the kernel function and h is the bandwidth.[9]

2. Obtain initial $b_0 = \left[(x - \widehat{x})' (x - \widehat{x})\right]^{-1} \left[(x - \widehat{x}) (y - \widehat{y})\right]$.

3. Obtain the residuals by $\widehat{u} = (y - \widehat{y}) - (x - \widehat{x}) b_0$.

4. Form the kernel regression (or local linear) of $\widehat{\sigma}_i^2 (z_i)$ using $\widehat{u}_i^2$ with explanatory variable z_i.[10]

5. Regress $(y_i - \widehat{y}_i) / \widehat{\sigma}_i (x_i, z_i)$ on $(x_i - \widehat{x}_i) / \widehat{\sigma}_i (z_i)$ to get efficient estimator of β.

3.4.3 Simulation Results

We know that the series estimator of the variance function is not necessarily positive. One should do trimming to guarantee the positive variance. However, the choice of trimming parameter is not the issue we would explore in this paper. To report the simulation result, we arbitrarily set three possible trimming points (TP), $TP = .1$, .01 and .001.

The simulation result is summarized in Table E.1. Under heteroskedasticity and exponential g function, we know that the preliminary estimator ignoring heteroskedasticity performs worst. Li's (2000) estimator has the min-

[9]Here we utilize Gaussian kernel and pick the bandwidth by

$$h = cn^{-(4+p)},$$

where $c = 1$, $p = 1$ and n is the sample size.

[10]Here we still utilize Gaussian kernel and pick the bandwidth along the same way as step 1.

imum BIAS and MSE in almost all cases. Our estimator only dominates in the case of $n = 400$. However, one could find that our new estimator performs pretty much similar to Li's (2000) estimator. As the sample size increases to 400, the behaviour of the two estimators is just about equivalent. On the other hand, kernel estimator is dominated by Li's (2000) and our estimators in this case.

To see the impact of different setup of unknown g function, we change the setting to be $g(z) = (1 + z)^3$. The result is listed in Table E.2. One could observe that in this setting Li's (2000) estimator still dominates although our estimator is quite close to Li's. Note that the kernel estimator performs poorly in this particular setting.

Note that throughout the simulation we arbitrarily pick the approximating function for $g(z)$ and $\sigma^2(z)$ being $(1, z, z^2)$ and $(1, z, z^2)$ respectively. The issue of how to pick the optimal smoothing parameters will be explored in next subsection.

3.4.4 Picking Smoothing Parameters

We would like to consider picking optimal smoothing parameters by minimizing the mean square error. Since MSE involves the true value of the estimator, in practice we need to expand our series estimator using higher order asymptotic expansion. Then the approximate MSE will not depend on the true value of the estimator. Therefore, one could minimize the approximate MSE to obtain the optimal smoothing parameter. However, it is not trivial to derive the approximate MSE in our context. An alternative approach is to estimate the approximate MSE through bootstrapping. Of course, the bootstrapping method we suggest in this section could be easily applied for Li's (2000) estimator. In this experiment, we consider three possible sets which serve as the functions for approximating $g(z)$ and $\sigma^2(z)$. The DGP follows the same setup in (3.12). The numbers of Monte Carlo and bootstrapping replications are set to 1000 and 399 for all cases. The resampling scheme is to bootstrap the (x, z, y) pairs. The potential instrument sets for approximating $g(z)$ and $\sigma^2(z)$ are

$$\begin{aligned} z^1 &= (1, z) \\ z^2 &= \left(1, z, z^2\right) \\ z^3 &= \left(1, z, z^2, z^3\right). \end{aligned}$$

Conducting the series estimation ends up with 9 combinations of instruments. We use the following notation to record each combination.

$$\begin{array}{lll} K_{11} = (z^1, z^1), & K_{21} = (z^2, z^1), & K_{31} = (z^3, z^1), \\ K_{12} = (z^1, z^2), & K_{22} = (z^2, z^2), & K_{32} = (z^3, z^2), \\ K_{13} = (z^1, z^3), & K_{23} = (z^2, z^3), & K_{33} = (z^3, z^3), \end{array}$$

where K_{32} stands for using $(1, z, z^2, z^3)$ and $(1, z, z^2)$ as the instruments of approximating $g(z)$ and $\sigma^2(z)$ respectively. It also means that we employ 4 and 3 instruments in forming the approximating functions. Note that we restrict our attention as the case of $TP = .001$.

The result is shown in Table E.3. One can observe that it is not the best strategy to choose as many as functions such as picking K_{33}. Most of the situations (such as K_{21} or K_{31}) tend to choose more (say, three or four) instruments for $g(z)$ and just two instruments for $\sigma^2(x,z)$. Because we know the DGP, the true MSE could be actually calculated for different combination of instruments. We compare the true MSE criteria with boostrapping method and find that the optimal smoothing parameters chosen by the two methods are quite similar. For instance, as $n = 100$, true MSE and bootstrapping MSE pick K_{21} and K_{31} respectively.

3.5 Concluding Remarks

In this paper we extend the feasible Generalized Least Square estimator considered in Li (2000) to allow for heteroskedastictiy of unknown form. We also propose an efficient estimator of partial linear regression model and prove it achieves Chamberlain's (1992) semi-parametric efficiency bound. The new estimator we proposed has the same first order asymptotic properties as Li's (2000) but is likely to be an easier estimator when it comes to studying higher order properties and bandwidth selection. In addition our estimator can be implemented using any nonparametric method whereas the method of Li (2000) is only implementable using a series method. The first order asymptotics of the feasible version of Li's (2000) and our estimator are provided.

In the Monte Carlo experiment, we compare our estimatoar with Li's (2000) and kernel estimators in terms of absolute mean bias and mean squared error. The simulation results show that our estimator behaves similar to Li's (2000) estimator. Also the performance of the series type esitmator seems more robust to the setting of the unknown g function than the kernel estimator. One needs to determine two smoothing parameters in estimating unknown $g(z)$ and $\sigma^2(z)$. The ususal way is to derive the approximate MSE. It is not trivial to construct the higher order MSE expansion for our series estimator. This will be left for the future research. To overcome the problem of picking smoothing parameters in series estimation, we propose the bootstrapping approximate mean square error to choose the smoothing parameters. Using the true MSE as the benchmark, the bootstrapping method works very well and provides us the criteria to choose two smoothing parameters simultaneously.

Chapter 4

Two-Step Series Estimation of Semiparametric Model with Generated Regressors

4.1 Introduction

Generated regressors occur in models where conditional expectations enter a regression model. An example in applied microeconomics is the estimation of simultaneous equation models with endogenous dummy variables. In that case, the conditional expectation term may involve a discrete variable which appears in the second step regression. See Amemiya (1985). The problem also occurs in the sample selection model with a nonparametric selection equation. The difference is that the regressor is a nonlinear transformation of a nonparametric estimate in a sample selection model. Macroeconomic models with rational expectations could have unknown conditional mean function on the right hand side, see Barro (1977). In labor economics, the unobservable expectations variables in a wage equation could be the expected job tenure which is a function of marital status, education, and other demographic variables.[1] In the area of international trade, if one would like to test the proposition in the influential paper by Grossman and Helpman (1994), the absolute own

[1]Generated regressor is also a technique to reduce dimensionality in the setting of nonparametric regression model with many regressors. See the discussion in Rilstone (1996).

price elasticity of import demand could be viewed as generated regressor.[2] See the application in Gawande (1997).

The econometric issues in presence of generated regressors have been extensively discussed by Pagan (1984, 1986). Basically, if we ignore the generated regressor problem, the estimation will be inefficient and statistical inference will be invalid. The two-step estimators are generally consistent and efficient but don't provide valid inference. Pagan (1984, 1986) only discussed the parametric setting in both the main regression model and auxiliary regression model. However, if either the functional form of main or auxiliary regression model is misspecified, it will result in incorrect inference. Andrews (1991, 1994) and Newey (1994a) consider the nonparametric setting in auxiliary regression. They estimate the generated regressor nonparametrically in the first step but keep the parametric setting in the second stage regression. One will suffer the inconsistent estimates or invalid inference owing to the misspecification of the regression equation of interest. Ahn (1995), Rilstone (1996) and Stengos and Yan (2001) try to avoid the strong parametric assumption in both first and second stage regression. They estimate both the regression of interest and auxiliary regression nonparametrically. Ahn (1997) established

[2] The regression model of testable prediction of Grossman and Helpman (1994) is

$$\frac{t_i}{1+t_i} = \beta\frac{z_i}{e_i} + X\gamma + \epsilon,$$

where t_i is the ad valorem trade taxes or subsidies for good; z_i is the ratio of domestic output to imports; e_i is the absolute own price elasticity of import demand and X is other control variable. The Grossman and Helpman hypothesis is that $\beta > 0$.

the $\sqrt{n}$ consistency and asymptotic normality results for the two-step estimators. Stengos and Yan (2001) proposed the double kernel estimators and build the $\sqrt{n}$ consistency and asymptotic normality results as the generated regressor enters the interest of semiparametric model in linear form.

Series estimation methods have been proposed to estimate the nonparametric regression model since it is conceptually simple and easily applicable. In addition it is straightforward to impose additive structure on unknown functions using series based approach, see Newey (1994b, 1997) for detailed discussion. Series methods work owing to the Stone-Weierstrass theorem that any continuous function could be approximated by a linear combination of known approximating functions including spline, power series and Fourier series. In many instances one could use relatively low order series to obtain high accuracy. Donald (1992) exploited some asymptotic restults for averages of series based nonparametric estimates. He applied the series based method to handle the generated regressor problem.[3]

In this paper we consider the two steps series estimation method for generated regressors problem in context of semiparametric regression model. We also establish the $\sqrt{n}$ consistency and asymptotic normality results of two-step series estimators. The asymptotic variance of the two-step series estimator is composed of two sources of error – one is the sampling error term and the other is from the fact that the series approximation may not

[3]The generated regressor problem in Donald (1992) is a parametric regression model containing an expectation term.

necessarily equal the true function. Our two step series method incoporates not only the conditional heteroskedastictiy but also the correlation of error terms between the regression of interest and auxiliary regression. This feature is not simultaneously investigated in Stengos and Yan (2001) and Donald (1992).

Section 2 starts with simple parametric model with generated regressors. Section 3 presents the asymptotic distribution of two-step series estimator allowing for more general semiparametric setting. The issue of choosing the smoothing parameters will be addressed in Section 4. Section 5 consists the performance of the estimator we propose through small Monte Carlo simulation. It also demonstrates the procedure of picking smoothing parameters using bootstrap method. Section 6 concludes.

4.2 Parametric Model

Let us consider the parametric model with generated regressor which enters the model linearly.

$$\begin{aligned} y &= x'\beta + E\left[s|z\right] \cdot \alpha + u \\ &= x'\beta + g\left(z\right) \cdot \alpha + u, \end{aligned} \tag{4.1}$$

where we define $E\left[s|z\right] = g\left(z\right)$ and $s = E\left[s|z\right] + \epsilon$. Let $E\left[x|z\right] = e\left(z\right)$ and $x = E\left[x|z\right] + v$. Here we could imagine that $E\left[s|z\right]$ is the expected job tenure. s represents the length of time at the present job. z may include age, marital status, number of children and other demographic characteristics which are assumed to be exogenous. Therefore, we allow the elements of x and z to overlap. More specifically, consider three variables (x_1, x_2, x_3). One could think x as (x_1, x_2) and z as (x_2, x_3). The overlapping variable between x and z is x_2. We also assume that error term is heteroskedastic. To adopt the series method in estimation we need the following assumptions and definitions. For the function g we let $S(g)$ denote the smoothness index. So that for instance a smoothness index of 2.5 means that the function is twice continuously differentiable and the following Lipschitz condition holds,

$$\left\| \frac{\partial^2}{\partial z^2} g(z_1) - \frac{\partial^2}{\partial z^2} g(z_0) \right\| \leq C \left\| z_1 - z_0 \right\|^{1/2}$$

As is well known from the work of Andrews (1991) and Newey (1994a) when the function has smoothness index $S(g)$ then one can approximate the function uinformly well in the sense that one can find coefficients π such that,

$$\sup_z \left\| g(z) - \sum_{j=1}^{K} w_j(z) \pi_j \right\| = O(K^{-S(g)/\dim(z)})$$

where $\dim(z)$ is the dimension of z and where $w_j(z)$ is the jth function of z such as a polynomial or spline type function. Thus the approximation error will go to zero faster if the smoothness index of nonparametric function g is larger. We make two assumptions that also aid in bounding the variance terms.

Assumption 4.2.1. *For the regression model (4.1) the observations are independent and*

$$E\left[u_i | x_i, z_i\right] = 0, \quad E\left[u_i^2 | x_i, z_i\right] = \sigma_i^2,$$

where

$$0 < \inf_i \sigma_i^2 < \sup_i \sigma_i^2 = \overline{\sigma}_u^2 < \infty.$$

Also, $E\left[|u_i|^3 |x_i, z_i\right] < \Delta < \infty$ *almost surely.*

Assumption 4.2.2. *For the regression model (4.1) the observations are independent and*

$$x_i = E[x_i|z_i] + v_i = e(z_i) + v_i,$$

where

$$E[v_i|z_i] = 0, \quad E\left[v_i^2|z_i\right] = \sigma_{vi}^2,$$

and

$$0 < \inf_i \sigma_{vi}^2 < \sup_i \sigma_{vi}^2 = \overline{\sigma}_v^2 < \infty.$$

Assumption 4.2.1 imposes the conditional mean zero, bounded second and third moments on the error term conditioning on x_i and z_i. Assumption 4.2.2 limits the behaviour of x_i conditioning on z_i. Assumption 4.2.1 and 4.2.2 also allow for conditional heteroskedasticity.

4.2.1 The Estimator

We could estimate $E[s|z] = g(z)$ by a series based regression method. Denote the estimator of the vector of g by $\widetilde{g} = P_z s$ where P_z is the projection matrix formed using the K functions $w_j(z_i)$. The model (4.1) becomes

$$y_i = x_i'\beta + \widetilde{g}_i\alpha + (g_i - \widetilde{g}_i)\alpha + u_i. \tag{4.2}$$

Denote $(x_i', \widetilde{g}_i) = w_i^\dagger$, $(x_i', g_i) = w_i$, and $(\beta', \alpha)' = \gamma$. Rewrite (4.2) as

$$y_i = w_i^\dagger \gamma + (g_i - \widetilde{g}_i)\, \alpha + u_i.$$

We could estimate γ by regressing y on $w^\dagger$.

$$\begin{aligned} \widehat{\gamma} &= \left(w^{\dagger\prime} w^\dagger\right)^{-1} w^{\dagger\prime} y \qquad (4.3) \\ &= \gamma + \left(w^{\dagger\prime} w^\dagger\right)^{-1} w^{\dagger\prime} \left[(g - \widetilde{g})\, \alpha + u\right]. \end{aligned}$$

Note that in this case, we only utilize series method once due to the parametric setting. The two step series method will be applied on more general nonparamteric setting in the next section.

4.2.2 First Order Results

We now state the first order asymptotic result by the following proposition.

Proposition 4.2.1. *Given the following assumptions*

(i) $\frac{1}{n} \sum_i w_i^\dagger w_i^{\dagger\prime} = A + o_p(1)$, *where* A $\left(= \frac{1}{n} \sum_i w_i w_i'\right)$ *is positive definite.*

(ii) u_i *and* ϵ_i *satisfy the conditions in Assumption 4.2.1.*

(iii) $\widetilde{g} = P_z s$ *and* $S(g) > \dim(z) + 1,$ $K(n) = O\left(n^{1/2-\xi}\right)$ *with* $0 < \xi < 1/2 - \dim(z)/2S(g)$.

(iv) $S(e) > 0,$ *with* $x_i = e(z_i) + v_i$ *satisfying Assumption 4.2.2.*

(v) The elements of x_i have bounded 4th moments. Then

$$\sqrt{n}\left(\widehat{\gamma}-\gamma\right)\rightarrow N\left(0,A^{-1}\left(S_1+S_2-2S_{12}\right)A^{-1}\right),$$

where

$$S_1=p\lim\frac{1}{n}\sum_i w_i'w_iE\left[u_i^2|x_i,z_i\right]$$

$$S_2=p\lim\frac{\alpha^2}{n}\sum_i\left(e\left(z_i\right)'\ g_i\right)'\left(e\left(z_i\right)'\ g_i\right)E\left[\epsilon_i^2|z_i\right]$$

and

$$S_{12}=p\lim\frac{\alpha}{n}\sum_i E\left[\epsilon_iu_iw_i'\left(e\left(z_i\right)'\ g_i\right)\right]$$

Proof. The proof of Proposition 4.2.1 is given in the Appendix. □

One could observe that the asymptotic variance of γ involves the term both from $\left(w^{\dagger\prime}w^{\dagger}\right)^{-1}w^{\dagger\prime}u$ (sampling errors) and $\left(w^{\dagger\prime}w^{\dagger}\right)^{-1}w^{\dagger\prime}\left(g-\widetilde{g}\right)\alpha$ (approximation errors) as well as the covariance term between these two sources of errors. Compare our estimator under the parametric setting with Stengos and Yan's (2001) double kernel method. We can find the result is basically similar to theirs. Employing series method we could easily establish the $\sqrt{n}$ consistency and asymptotic normaliy properties. The difference is that we relax the assumption of conditional homoskedasticity. Donald (1992) considers the similar model using series method to estimate γ assuming that there is no correlation between the error terms u and ϵ. Our series estimator contains Stengos and Yan's (2001) double kernel method and Donald's (1992) series method as special cases in the sense that we incoporate conditional heteroskedasticity (not considered in Stengos and Yan) and correlation of errors between equation of interest and auxiliary regression (not considered in Donald).

4.2.3 Variance Estimation

Now we briefly discuss the variance estimation in Proposition 4.2.1. Basically, one could replace the population value with the sample counterpart. First of all, the term S_1 could be esitmated as follows by White's (1980) method.

$$\overset{\vee}{S}_1 = \frac{1}{n}\overset{\vee}{w}'\overset{\vee}{\Psi}\overset{\vee}{w},$$

where $\overset{\vee}{w} = (x\ \widetilde{g})$ and $\overset{\vee}{\Psi} = diag\left(\overset{\vee}{u}_i^2\right)$. Similarly, we can estimate S_2 by $\overset{\vee}{S}_2$.

$$\overset{\vee}{S}_2 = \frac{\widehat{\alpha}^2}{n}\left(\overset{\vee}{e}'\ \widetilde{g}\right)'\overset{\vee}{\Xi}\left(\overset{\vee}{e}'\ \widetilde{g}\right),$$

where $\overset{\vee}{e}'$ is the estimate of $E\left[x|z\right]$ and $\overset{\vee}{\Xi} = diag\left(\overset{\vee}{\epsilon}_i^2\right)$. Of course, $\overset{\vee}{e}'$ could be estimated nonparametrically by series, local linear or k-NN estimators. For example, a standard kernel estimator (say, Nadaraya-Watson estimator) of $E\left[x|z\right]$ is given by

$$\overset{\vee}{e}\left(z\right) = \frac{\sum_{l=1}^{n} K\left(\frac{z_l - z}{h}\right)x_l}{\sum_{l=1}^{n} K\left(\frac{z_l - z}{h}\right)}.$$

A series estimator of $E\left[x|z\right]$ is given by

$$\overset{\vee}{e}\left(z\right) = P_z x,$$

where P_z is the projection matrix formed by approximating function q_z. Now deal with the estimate of the convariance term S_{12}.

$$\overset{\vee}{S}_{12} = \frac{\widehat{\alpha}}{n} \sum_i \left[\overset{\vee}{\epsilon}_i \overset{\vee}{u}_i' \overset{\vee}{w}_i' \left(\overset{\vee}{e}_i' \, \widetilde{g} \right) \right].$$

The next proposition says that applying the sample couterpart in variance estimation will end up with consistency result.

Proposition 4.2.2. *$\overset{\vee}{S}_1$, $\overset{\vee}{S}_2$ and $\overset{\vee}{S}_{12}$ are consistent estimators of S_1, S_2 and S_{12} respectively.*

Proof. We know that $\widetilde{g}$ is consistent estimator of g. $\overset{\vee}{w}$ will be the consistent estimator of w. Applying White's (1980) result on heteroskedasticity consistent covariance matrix estimator gives the consistency of $\overset{\vee}{S}_1$. Similarly, one could find a consistent estimator of e, $\overset{\vee}{e}$. $\overset{\vee}{S}_2$ is the consistent estimator of S_2. We already have the consistency result of $\widehat{\alpha}$ in Proposition 4.2.1. $\overset{\vee}{S}_{12}$ will be a consistent estimator of S_{12} as well. □

4.2.4 Efficiency Bound

In this section we discuss the efficiency bound for estimating the parameter β. To simplify the discussion of the semiparametric efficiency bounds, we assume that x and z are scalars. We also need the normality assumption for the distribution of u and e. Homoskedasticity of u and ϵ makes our work easier. A series method is adpoted in this paper to approximate the unknown

g function. The so called parametric submodel K is written as

$$y = \beta x + \alpha \sum_{j=1}^{K} \gamma_j w_j (z) + u \tag{4.4}$$
$$s = \sum_{j=1}^{K} \gamma_j w_j (z) + \epsilon.$$

Now we have parameters of interests (β, α) and nuisance parameters $\gamma^K = (\gamma_1, \gamma_2, ..., \gamma_K)'$. Let $\psi^K = \left(\beta, \alpha, \gamma^K\right)$. The log-likelihood function of the parametric submodel K in (4.4) is

$$\mathcal{L}\left(\beta, \alpha, \gamma^K\right) = \text{constant} - \frac{1}{2\sigma_u^2}\left[y - \beta x - \alpha \sum_{j=1}^{K} \gamma_j w_j (z)\right]^2 - \frac{1}{2\sigma_\epsilon^2}\left[s - \sum_{j=1}^{K} \gamma_j w_j (z)\right]^2.$$

The score function as evaluated at the true model is the $(K+2)$ vector

$$S\left(\psi_0^K\right) = \begin{pmatrix} S_\theta \\ S_{\gamma_K} \end{pmatrix} = \begin{pmatrix} \left(1/\sigma_u^2\right) ux \\ \left(1/\sigma_u^2\right) u g_0 \\ \left[\left(1/\sigma_u^2\right) \alpha u + \left(1/\sigma_\epsilon^2\right) \epsilon\right] w^K \end{pmatrix} = \frac{1}{\sigma_u^2} \begin{pmatrix} ux \\ u g_0 \\ \tau_0 w^K \end{pmatrix},$$

where $S_\theta = (S_\beta, S_\alpha)'$, $\tau_0 = \alpha u + \left(\sigma_u^2/\sigma_\epsilon^2\right)\epsilon$ and $w^K = (w_1, w_2, ..., w_K)'$. To proceed the calculation of semiparametric lower bounds, we now define the tangent set, $\mathcal{T}$, for the nuisance function as the mean square closure of all $K \times 1$ linear combinations of the score functions S_{γ_K}.

$$\mathcal{T} = \left\{ t \in \mathcal{R}^2 : E\left\|t\right\|^2 < \infty, \exists M_j S_{\gamma_K} : \lim_{j\to\infty} \left(E\left\|t - M_j S_{\gamma_K}\right\|^2\right)^{1/2} \to 0 \text{ as } j \to \infty \right\}.$$

The residual vector from the projection of S_θ on the tangent set is defined as the efficient score S^*, which could be expressed as

$$S_\theta^* = S_\theta - \mathcal{P}\left[S_\theta | \mathcal{T}\right].$$

The semiparametric efficiency bound is

$$V_\theta = \left[E\left(S_\theta^* S_\theta^{*\prime}\right)\right]^{-1}.$$

To apply above result to our context, we need the Lemma 1 in Rilstone (1993) by setting $h = \sigma_u^2 S_\theta$ and $B = \tau_0$. Now the projection of the score S_θ on the tangent set will be

$$\begin{aligned}
\mathcal{P}\left[S_\theta|\mathcal{T}\right] &= \frac{E\left[\sigma_u^2 S_\theta|z\right]}{E\left[\tau_0^2|z\right]}\tau_0 \\
&= \frac{1}{\sigma_u^2}\left[\phi\epsilon + \frac{\phi\sigma_u^2}{\alpha\sigma_\epsilon^2}u\right]\begin{pmatrix} E\left[x|z\right] \\ g_0\left(z\right)\end{pmatrix},
\end{aligned}$$

where $\phi = \frac{\alpha^2\sigma_\epsilon^2}{\alpha^2\sigma_\epsilon^2+\sigma_u^2} \in (0,1)$. The efficient score S_θ^* is given by

$$S_\theta^* = \frac{1}{\sigma_u^2}\begin{pmatrix} ux \\ ug_0\left(z\right)\end{pmatrix} - \frac{1}{\sigma_u^2}\left[\phi\epsilon + \frac{\phi\sigma_u^2}{\alpha\sigma_\epsilon^2}u\right]\begin{pmatrix} E\left[x|z\right] \\ g_0\left(z\right)\end{pmatrix}.$$

A little algebra gives

$$E\left(S_\theta^* S_\theta^{*\prime}\right) = \frac{1}{\sigma_u^2}\begin{pmatrix} E\left[x^2\right] & E\left[g_0 x\right] \\ E\left[g_0 x\right] & E\left[g_0^2\right]\end{pmatrix} - \frac{\phi}{\sigma_u^2}\begin{pmatrix} E\left[\left(E\left[x|z\right]\right)^2\right] & E\left[g_0 x\right] \\ E\left[g_0 x\right] & E\left[g_0^2\right]\end{pmatrix}.$$

Therefore,

$$V_\theta = \begin{pmatrix} V_{\beta\beta} & V_{\beta\alpha} \\ V_{\alpha\beta} & V_{\alpha\alpha}\end{pmatrix} = \left[E\left(S_\theta^* S_\theta^{*\prime}\right)\right]^{-1} = \sigma_u^2\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22}\end{pmatrix}^{-1},$$

where

$$\begin{aligned}
B_{11} &= E\left[x^2\right] - \phi E\left[\left(E\left[x|z\right]\right)^2\right] \\
B_{12} &= E\left[g_0 x\right] - \phi E\left[g_0 x\right] = B_{21} \\
B_{22} &= E\left[g_0^2\right] - \phi E\left[g_0^2\right].
\end{aligned}$$

Using the partition inverse formula, we have

$$V_{\beta\beta} = B_{11}^{-1} + B_{11}^{-1} B_{12} F_2 B_{12}' B_{11}^{-1},$$

where $F_2 = \left(B_{22} - B_{12}' B_{11}^{-1} B_{12}\right)^{-1}$.

4.3 General Model

If the model considered in (4.1) is misspecified, we will end up with inconsistent estimation and invalid inference. Therefore, we should avoid specifying the parametric setting like that in (4.1). Now consider the more general semiparametric model (or partial linear regression model) with generated regressor in the linear part. The parametric part $x'\beta$ has been replaced by nonparametric unknown function $\theta(x)$.

$$\begin{aligned} y &= \theta(x) + E[s|z] \cdot \alpha + u \\ &= \theta(x) + g(z) \cdot \alpha + u. \end{aligned} \tag{4.5}$$

The identification of this model is stated below.

Assumption 4.3.1. *(Identification) To identify the semiparametric model in (4.5), we need* $E\left[(g - E[g|x])'(g - E[g|x])\right]$ *is positive definite.*

Literally, we need that random variable z is not fully contained in x. To understand the identification condition 4.3.1, taking expectation conditional on x with respect to (4.5) gives

$$E[y|x] = \theta(x) + E[g(z)|x] + E[u|x]. \tag{4.6}$$

Substracting (4.6) from (4.5) gives

$$y - E[y|x] = [g(z) - E[g(z)|x]]\alpha + u - E[u|x].$$

It is obvious that identification of α requires the full rank of $g(z) - E[g(z)|x]$. Now let's consider the following examples.

Example 4.3.1. *Let $x = (x_1, x_2, x_3)$, $z = (x_3, x_4)$ and $g(z) = x_3^2 + x_4$. In this case, the identification condition in Assumption 4.3.1 holds, α is identified. In fact,*

$$g(z) - E[g(z)|x] = x_4 - E[x_4|x_1, x_2, x_3] \neq 0.$$

Example 4.3.2. *Let $x = (x_1, x_2, x_3)$, $z = x_3$ and $g(z) = x_3^2$. In this case, the identification condition in Assumption 4.3.1 does not hold, α is not identified since every element of z is contained in x. In fact,*

$$g(z) - E[g(z)|x] = x_3^2 - E\left[x_3^2|x_1, x_2, x_3\right] = x_3^2 - x_3^2 = 0.$$

Example 4.3.3. *Let $x = (x_1, x_2)$, $z = (x_1, x_2, x_3, x_4)$ and $g(z) = x_1^2 x_2 + x_3^2 x_4^5$. In this case, the identification condition in Assumption 4.3.1 holds since $X \subset Z$. α is identified. In fact,*

$$g(z) - E[g(z)|x] = x_3^2 x_4^5 - E\left[x_3^2 x_4^5|x_1, x_2\right] \neq 0.$$

4.3.1 The Estimator

The first step for the two step estimator of α is to estimate $E[s|z] = g(z)$ by $\widetilde{g}_i = P_z s$. The model becomes

$$y_i = \theta_i + \widetilde{g}_i \alpha + (g_i - \widetilde{g}_i)\alpha + u_i. \tag{4.7}$$

Premultiplying equation (4.7) by P_x the projection matrix formed using series basis functions of x gives

$$\widehat{y}_i = \widehat{\theta}_i + \widehat{\widetilde{g}}_i \alpha + \left(\widehat{g}_i - \widehat{\widetilde{g}}_i\right) \alpha + \widehat{u}_i, \tag{4.8}$$

where $\widehat{A} = P_x A$. Substracting (4.8) from (4.7) gives

$$y_i - \widehat{y}_i = \theta_i - \widehat{\theta}_i + \left(\widetilde{g}_i - \widehat{\widetilde{g}}_i\right) \alpha + \left(g_i - \widetilde{g}_i\right) \alpha - \left(\widehat{g}_i - \widehat{\widetilde{g}}_i\right) \alpha + u_i - \widehat{u}_i. \tag{4.9}$$

The second step is to estimate α by regressing $y - \widehat{y}$ on $\left(\widetilde{g} - \widehat{\widetilde{g}}\right)$ or using the partialling out formula.

$$\begin{aligned} \widehat{\alpha} &= \left[\left(\widetilde{g} - \widehat{\widetilde{g}}\right)' \left(\widetilde{g} - \widehat{\widetilde{g}}\right)\right]^{-1} \left(\widetilde{g} - \widehat{\widetilde{g}}\right)' \left(y - \widehat{y}\right) \\ &= \left[\widetilde{g}' \left(I - P_x\right) \widetilde{g}\right]^{-1} \widetilde{g}' \left(I - P_x\right) y. \end{aligned} \tag{4.10}$$

4.3.2 First Order Results

Using the fact that $y = \theta(x) + g(z) \cdot \alpha + u$, equation (4.10) could be rewritten as

$$\begin{aligned} \widehat{\alpha} &= \left[\widetilde{g}' \left(I - P_x\right) \widetilde{g}\right]^{-1} \widetilde{g}' \left(I - P_x\right) \left(\theta + g\alpha + u\right) \\ &= \left[\widetilde{g}' \left(I - P_x\right) \widetilde{g}\right]^{-1} \widetilde{g}' \left(I - P_x\right) \left[\widetilde{g}\alpha + \widehat{\theta} + \left(\theta - \widehat{\theta}\right) + \left(g - \widetilde{g}\right) \alpha + u\right] \\ &= \alpha + \left[\widetilde{g}' \left(I - P_x\right) \widetilde{g}\right]^{-1} \widetilde{g}' \left(I - P_x\right) \left[\theta - \left(\widetilde{g} - g\right) \alpha + u\right]. \end{aligned}$$

The idea of obtaining the consistency result is to verify that the term

$$\left[\widetilde{g}' \left(I - P_x\right) \widetilde{g}\right]^{-1} \widetilde{g}' \left(I - P_x\right) \left[\theta - \left(\widetilde{g} - g\right) \alpha + u\right]$$

will approach to zero in probability limit. On the other hand, the term

$$\left[\widetilde{g}' \left(I - P_x\right) \widetilde{g}\right]^{-1} \widetilde{g}' \left(I - P_x\right) \left[\left(\widetilde{g} - g\right) \alpha + u\right]$$

allows us to establish the asymptotic normality. The asymptotic distribution of $\widehat{\alpha}$ is given in Proposition 4.3.1.

Proposition 4.3.1. *Given the following assumptions*

(i) $\frac{1}{n}\widetilde{g}'(I-P_x)\widetilde{g} = E\left[(g-h)'(g-h)\right] + o_p(1) = D + o_p(1)$, *where* $g = E[s|z], h = E[g|x]$ *and* D *is positive definite.*

(ii) u_i *and* ϵ_i *satisfy the conditions in Assumption 4.2.1.*

(iii) $\widetilde{g} = P_z s$ *and* $S(g) > \dim(z) + 1,\ K(n) = O\left(n^{1/2-\xi}\right)$ *with* $0 < \xi < 1/2 - \dim(z)/2S(g)$.

(iv) $S(e) > 0$, *with* $x_i = e(z_i) + v_i$ *satisfying Assumption 4.2.2.*

(v) *The elements of* x_i *have bounded 4th moments. Define* $g = h(x) + \eta$ *and* $\eta = \psi(z) + \omega$. *Then*

$$\sqrt{n}(\widehat{\alpha} - \alpha) \to N\left(0, D^{-1}(T_1 + T_2 - 2T_{12})D^{-1}\right),$$

where

$$\begin{aligned} T_1 &= p\lim \frac{1}{n}\sum_i g_i(I-P_x)g_i E\left[u_i^2|x_i,z_i\right] \\ &= E\left[(g-h)' E[uu'|x,z](g-h)\right] \end{aligned}$$

$$\begin{aligned} T_2 &= p\lim \frac{\alpha^2}{n}\sum_i \psi_i E\left[\epsilon_i^2|z_i\right]\psi_i \\ &= \alpha^2 \psi' E[\epsilon\epsilon'|z]\psi \end{aligned}$$

and

$$\begin{aligned} T_{12} &= p\lim \frac{\alpha}{n}\sum_{i} E\left[\left(g_i - h_i\right)\psi_i \epsilon_i u_i\right] \\ &= \alpha E\left[\left(g-h\right)'\psi' E\left[\epsilon u'|x,z\right]\psi\left(g-h\right)\right]. \end{aligned}$$

Proof. The proof of Proposition 4.3.1 is given in the Appendix. □

Our result under more general setting is different from Stangos and Yan (2001) in two aspects. First, the asymptotic variance of $\widehat{\alpha}$ in Stangos and Yan contains only the aymptotic variance of

$$\frac{\left[\widetilde{g}'\left(I-P_x\right)\widetilde{g}\right]^{-1}\widetilde{g}'\left(I-P_x\right)u}{\sqrt{n}},$$

which is from sampling error. Denote it as T_1. As for our series estimator, there are two extra terms. One is the asymptotic variance of

$$\frac{\left[\widetilde{g}'\left(I-P_x\right)\widetilde{g}\right]^{-1}\widetilde{g}'\left(I-P_x\right)\left(\widetilde{g}-g\right)\alpha}{\sqrt{n}},$$

which is from the approximation error. Denote it as T_2. The other is a covariance term between the sampling and approximatin errors. It is T_{12} defined in Proposition 4.3.1. Stangos and Yan proved that T_2 will eventually disappear in double kernel estimation context. However, our result shows that the asymptotic variance from approximating unknown function g should be taken into account.

We think that our result is more reasonable in that the asymptotic variance of $\widehat{\alpha}$ contains the term from approximation error. Take a look at the first order asymptotic result of Stengos and Yan (2001), Donald (1992) and our series estimator under parametric setting. One can find out the asymptotic variances of the three methods all involves the term from approximation error. Moreover, the parametric case is the special case of the more general case of semiparametric model. One could imagine that in the more general setting, the error of approximating unknown g function should matter as well.

The second major difference between Stangos and Yan (2001) and ours is that our estimator allows conditional heteroskedasticity. While Stangos and Yan merely consider the conditional homoskedastic case. In empirical application, the asymptotic result of our estimator shows that it is quite straightforward to estimate the conditional heteroskedasticity of unknown form using White type method.

4.3.3 Variance Estimation

In this subsection we propose methods to estimate the asymptotic variance in Proposition 4.3.1. Following the similar strategy in previous section, one could replace the population value with the sample counterpart. The term T_1 could be esitmated by

$$\overset{\vee}{T}_1 = \frac{1}{n}\left(\widetilde{g} - \overset{\vee}{h}\right)' \overset{\vee}{\Psi} \left(\widetilde{g} - \overset{\vee}{h}\right),$$

where $\overset{\vee}{\Psi} = diag\left(\overset{\vee}{u}_i^2\right)$ and $\overset{\vee}{h}$, which could be estimated through any nonparametric method described in previous section. Now we would estimate T_2 by

$\overset{\vee}{T}_2$.

$$\overset{\vee}{T}_2 = \frac{\widehat{\alpha}^2}{n} \overset{\vee}{\psi}' \overset{\vee}{\Xi} \overset{\vee}{\psi},$$

where $\overset{\vee}{\Xi} = diag\left(\overset{\vee}{\epsilon}_i^2\right)$ and $\overset{\vee}{\psi}'$, where $\overset{\vee}{\psi}'$ could be estimated using series method as follows.

$$\overset{\vee}{\psi}' = \widetilde{g}' \left(P_z - P_x P_z\right),$$

where P_x and P_z are projection matrix formed by appropriate approximating function[4]. The last term T_{12} could be estimated as follows.

$$\overset{\vee}{T}_{12} = \frac{\widehat{\alpha}}{n} \sum_i \left[\left(\widetilde{g}_i - \overset{\vee}{h}_i\right) \overset{\vee}{\psi}_i \overset{\vee}{e}_i \overset{\vee}{u}_i\right].$$

The next proposition says that applying the sample couterpart in variance estimation will end up with consistency result.

Proposition 4.3.2. *$\overset{\vee}{T}_1$, $\overset{\vee}{T}_2$ and $\overset{\vee}{T}_{12}$ are consistent estimators of T_1, T_2 and T_{12} respectively.*

---

[4]From the Appendix-C, we have

$$\psi(z) = g(z) - E\left[E\left[g(z)|x\right]|z\right].$$

Under the idenfication of α, the series estimator of $\psi(z)$ is

$$\overset{\vee}{\psi} = P_z s - P_z P_x P_z s.$$

However, if $X \subset Z$, α is still identified but the estimator of $\psi(z)$ becomes

$$\overset{\vee}{\psi} = P_z s - P_x P_z s,$$

since in this case

$$E\left[E\left[g(z)|x\right]|z\right] = E\left[g(z)|x\right] = h(x).$$

Using the notation in the Appendix-C, the error term ω will be 0.

Proof. By the consistency result of $\widetilde{g}$ and $\overset{\vee}{h}$. The sandwich form of $\overset{\vee}{T}_1$ is the consistent estimator of T_1 by the White's argument. $\widehat{\alpha}$ is a consistent estimator of α from Proposition 4.3.1. The nonparametric estimator $\overset{\vee}{\psi}$ is also consistent. Consistency of $\overset{\vee}{T}_2$ is easy to establish. Based on the consistency results of $\overset{\vee}{T}_1$ and $\overset{\vee}{T}_2$, consistency of $\overset{\vee}{T}_{12}$ is not hard to verify. □

4.4 Choosing the Smoothing Parameters

Although the 2-step series estimator we propose in this paper is quite easy to implement, we still need to pick the number of column in approximating function (or the smoothing parameter, K). In the parametric regression model, one smoothing parameter will be needed since there is only one unknown function g to be approximated. In the more general case, we have to pick two smoothing parameters due to the approximation of $\theta(x)$ and $g(z)$. In this case, what we need is to choose two smoothing parameters. For instance, let us just simply assume $x = (x_1, x_2)$ and $z = (z_1, z_2)$. For the double kernel estimator, they need to specify four kernel functions for x_1, x_2, z_1 and z_2. Further, they have to pick the 4 bandwidths for the corresponding kernel functions. As the number of (x, z) variables increase, double kernel methods will get into trouble in picking too many smoothing parameters. In this sense, our 2-step series estimator may be easier to use in practice than the double kernel estimator. Our method may also be easier from a computational standpoint and can be implemented with most regression packages.

The natural problem is how to choose the optimal smoothing parameter. One approach to answer this question is to emply higher order asymptotics. It may not be trivial in this context. Alternatively, we may utilize the bootstrap method to get approximate mean square error. Using the bootstrap-based procedure for selecting the moment condition has been discussed in Inoue (2003). Minimizing the bootstrapping approximate MSE will obtain the optimal K. The detail will be discussed in the next section.

4.5 Monte Carlo Experiment

4.5.1 Simulation Design

In the section, we conduct a small Monte Carlo simulation to see the performance of ours and double kernel estimator proposed by Stengos and Yan (2001). The simulation design follows Stengos and Yan (2001) in order to do comparison. The design is as follows[5].

$$
\begin{aligned}
y &= \theta(x) + E[s|z]\alpha + u \qquad (4.11)\\
\theta(x) &= (\beta_1 x_1 + \beta_2 x_2)^2 \\
s &= \left[(z_1+z_2)^2 \;\; z_2\right]\gamma + e \\
z_i &= x_i\delta_i + v_i, \;\; i = 1, 2.
\end{aligned}
$$

[5]Allowing for more general setting of g function could verify the 2 step series estimator is robust to the functional form of the unkown function. We also implement the experiment by setting g function as $\exp\left[(z_1+z_2)^2 + z_2\right]$, which is obviously not a polynomial form employed by Stengos and Yan (2001). See the estimation result in next subsection. Furthermore, since Stangos and Yan (2001) did not take into account conditional heteroskedasticity, we will not explore this issue here. Of course, our two-step estimator could deal with conditional heteroskedasticity well.

where x_i $(i = 1, 2)$ are generated from uniform distribution on [1,2] and $u \sim N(0, \sigma_u^2)$, $v \sim N(0, \sigma_v^2)$, $e \sim N(0, \sigma_e^2)$, $\rho_i = cov(x_i, z_i)/\sigma_{x_i}^2$. We pick the coefficients of δ_i by $\delta_i = \sqrt{12}\sigma\rho/\sqrt{1-\rho^2}$, where ρ is the correlation coefficient between x and z. σ is set to 1. In addition, we introduce the correlation between u and e by setting $\sigma_{ue} = 0.5$, which is not specified in Stengos and Yan (2001). The related parameters are summarized as follows.

$$\begin{array}{lll} \beta_1 = \beta_2 = \alpha = 1, & \gamma = (1,1)', & \sigma_u^2 = \sigma_v^2 = \sigma_e^2 = 1, \\ \delta = .35 \text{ if } \rho = .1, & \delta = 2 \text{ if } \rho = .5, & \delta = 7.15 \text{ if } \rho = .9. \end{array}$$

We consider 4000, 2000 and 1000 replications for sample sizes of $n = 100$, 200 and 400 respectively. We also report three possible correlation between x and z as in Stengos and Yan (2001). In the simulation, we compute the Mean Absolute Bias (BIAS) and Mean Square Error (MSE) for seven estimators rather than three estimators in Stengos and Yan (2001). The seven estimators are described below.

1. True estimator: We assume that the unknown functions g and θ are both known to us. Of course, it's the unattainable goal. However, it provides the benchmark for various estimators.

2. True_series estimator: It means that we directly take the true value of g $(= E[s|z])$ in the auxilliary regression instead of estimating it. Then we use this true value to estimate coefficient α using the series method.

3. True_Kernel estimator: It is computed in Stengos and Yan (2001) for the benchmark. Here we take the true value of g and use the Kernel

method[6] to estimate α.

4. Linear_series estimator: We treat the unknown g function as linear one and get the estimate of g. Then we use the misspecified estimate of g to estimate α using series method.

5. Linear_Kernel estimator: It has the same form as the Linear_series estimator except that we adopt Kernel method in the second stage estimation.

6. 2-Step series estimator: That is the estimator we propose in this paper. We estimate $E[s|z]$ by series method in the first stage and estimate $E[E[s|z]|x]$ by series method again in the second stage. Then we get the estimate of coefficient α.

7. Double Kernel estimator: That is the estimator proposed by Stengos and Yan (2001). The feature is that they applied the Kernel method in both stages of estimation.

4.5.2 Simulation Results

The simulation result is summarized in Table F.1. The preliminary simulation result is encouraging in this particular design. First, under the as-

[6] To implement the Kernel regression, we need the choices of kernel function and bandwidth. Here we follow Stengos and Yan's strategy by selecting Gauss kernel and setting bandwidth as

$$h = cn^{-(4+p)},$$

where $c = 1$, $p = 2$ and n is the sample size.

sumption of knowing the true unknown g function, our series method performs better than Kernel method especially when the correlation between x and z is higher. Second, even when we assume the unknown g function is linear, the performance of Linear_series estimator is better than the Double Kernel method. That means using the series method in the second stage will be better than the Double Kernel method in this case. Finally, simulation evidence shows that our 2-step series estimator performs well. It uniformly outperforms the Double Kernel estimator in all cases. In some cases, our estimator has even smaller BIAS and MSE than those of True_Kernel estimator.

Some may argue that the original specification of the nonparametric g function in Stengos and Yan (2001), $\left[g(z) = (z_1 + z_2)^2 + z_2\right]$, would be in favor of our series estimator since it is kind of polynomial form. Therefore we change the DGP by setting $g(z) = \exp\left[(z_1 + z_2)^2 + z_2\right]$. The estimation result is shown in Table F.2. Even though we modify the DGP of g function, our two-step series estimator still performs very well.

Note that throughout the simulation we arbitrarily pick the approximating function for $g(z)$ and $\theta(x)$ being

$$\left(1, z_1, z_2, z_1^2, z_2^2, z_1 z_2\right),$$
$$\left(1, x_1, x_2, x_1^2, x_2^2, x_1 x_2\right)$$

respectively. The issue of how to pick the optimal smoothing parameters will be explored in next subsection.

4.5.3 Picking Smoothing Parameters

We would consider to pick optimal smoothing parameters by minimizing the mean square error. Since MSE involves the true value of the estimator, in practice we need to expand our two-step series estimator using higher order asymptotic expansion. The approximate MSE will not depend on the true value of the estimator. Therefore, one could minimize the approximate MSE to obtain the optimal smoothing parameter. However, it is not trivial to derive the approximate MSE in our context. An alternative approach is to estimate the apprxoimate MSE through bootstrapping. In this experiment, we consider three possible sets which serve as the approximating functions for $g(z)$ and $\theta(x)$. The DGP follows the same setup in (4.11). The numbers of Monte Carlo and bootstrapping replications are set to 1000 and 399 for all cases[7]. The resampling scheme is to bootstrap the (x, z, y, s) pairs. The instruments for estimating $\theta(x)$ are

$$\begin{aligned} x^1 &= (1, x_1, x_2) \\ x^2 &= \left(1, x_1, x_2, x_1^2, x_2^2, x_1 x_2\right) \\ x^3 &= \left(1, x_1, x_2, x_1^2, x_2^2, x_1 x_2, x_1^3, x_2^3, x_1^2 x_2, x_1 x_2^2\right). \end{aligned}$$

[7] We use 1000 Monte Carlo replications instead of 2000 or 4000 to save the computing time.

The instrument sets for approximating $g(z)$ are

$$\begin{aligned} z^1 &= (1, z_1, z_2) \\ z^2 &= \left(1, z_1, z_2, z_1^2, z_2^2, z_1 z_2\right) \\ z^3 &= \left(1, z_1, z_2, z_1^2, z_2^2, z_1 z_2, z_1^3, z_2^3, z_1^2 z_2, z_1 z_2^2\right). \end{aligned}$$

Thus there are a total of nine combinations of approximating functions. We use the following notation to record each combination.

$$\begin{array}{lll} K_{11} = (x^1, z^1), & K_{12} = (x^1, z^2), & K_{13} = (x^1, z^3), \\ K_{21} = (x^2, z^1), & K_{22} = (x^2, z^2), & K_{23} = (x^2, z^3), \\ K_{31} = (x^3, z^1), & K_{32} = (x^3, z^2), & K_{33} = (x^3, z^3), \end{array}$$

where K_{23} stands for using $(1,\ x_1,\ x_2,\ x_1^2,\ x_2^2,\ x_1 x_2)$ and $(1,\ z_1,\ z_2,\ z_1^2,\ z_2^2,\ z_1 z_2,\ z_1^3,\ z_2^3,\ z_1^2 z_2,\ z_1 z_2^2)$ as the instruments of approximating $\theta(x)$ and $g(z)$ respectively. It also means that we employ 6 and 10 instruments in forming the approximating function.

The result is shown in Table F.3. One can observe that it is not the best strategy to choose as many functions as possible, K_{33}. Most of the situations tend to choose medium numbers of instruments such as K_{22} or K_{23}. When the sample size is 100, fewer functions of x variables (say, 3) are needed in estimating $\theta(x)$. Increasing the sample size ($n = 100$ or 200), results in more functions becoming preferable (say, 6). On the other hand, it was usually optimal to pick a larger number of functions of z variables (say 6 or 10). Because we know the DGP, the true MSE could be actually calculated for different combination of instruments. We compare the true MSE criteria with boostrapping method and find that the optimal instruments chosen by the two methods are quite similar. For instance, as $n = 100$, true MSE and bootstrapping MSE pick K_{12} and K_{13} respectively. For $n = 200$ and 400, K_{22} and K_{23} are usually the best choices.

4.6 Concluding Remarks

In this paper, we propose the two-step series estimation method to estimate a semiparametric regression model with generated regressors. We start with simple parametric model with generated regressors and then consider more general semiparametric setting. We establish the $\sqrt{n}$ consistency and asymptotic normality for the two-step estimator. Our two step series method incoporates not only conditional heteroskedastictiy but also the correlation of error terms between the regression of interest and auxiliary regression. This feature is not simultaneously investigated in Stengos and Yan (2001) and Donald (1992). Also in more general setting, the asymptotic variance of our estimator seems more plausiable than that of Stangos and Yan (2001).

According to our simulation result, our two step series estimator outperforms other competing estimators in terms of mean absolute bias and mean square error. Even if we change the functional form of unknown g function, the performance of our two-step series estimator is overwhelming. Compared to the double kernel method, our estimator has some computational advantage in the sense that running ordinary least squares twice is required in our appraoch. However, one needs to determine two smoothing parameters in estimating unknown $\theta(x)$ and $g(z)$. The ususal way is to derive the approximate MSE. It is not trivial to construct the higher order MSE expansion for our two-step series estimator. We propose the bootstrapping method to approximate MSE. Using the true MSE as the benchmark, the bootstrapping method works very well and provides us the criteria to choose two smoothing parameters simultaneously.

Appendices

Appendix A

Appendix-Chapter 2

We will use the following useful expansion for a matrix repeatedly.

$$\widehat{A}^{-1} = A^{-1} - A^{-1}\left(\widehat{A} - A\right)A^{-1} + A^{-1}\left(\widehat{A} - A\right)A^{-1}\left(\widehat{A} - A\right)A^{-1} + ... \quad \text{(A.1)}$$

A.1 Proof of Proposition 2.3.1

We have that

$$\widehat{\beta}_{Cragg} - \beta = \left[X'Q\left(Q'\widehat{\Sigma}Q\right)^{-1}Q'X\right]^{-1}X'Q\left(Q'\widehat{\Sigma}Q\right)^{-1}Q'\epsilon = \widehat{H}^{-1}\widehat{h}.$$

To simplify the calculation, we adopt White's (1980) method of estimating $\widehat{\Sigma}$ =diag(u_i^2) $\equiv S$. Remember u_i^2 is the square of the OLS residual. We now expand $(Q'SQ)^{-1}$ using (A.1).

$$(Q'SQ)^{-1} = (Q'\Sigma Q)^{-1} - (Q'\Sigma Q)^{-1}Q'(S - \Sigma)Q(Q'\Sigma Q)^{-1}.$$

It follows that

$$\widehat{h} = X'Q(Q'\Sigma Q)^{-1}Q'\epsilon - X'Q(Q'\Sigma Q)^{-1}Q'(S - \Sigma)Q(Q'\Sigma Q)^{-1}Q'\epsilon.$$

and

$$
\begin{aligned}
& \widehat{H}^{-1} \\
= \; & \left(X'Q\left(Q'\Sigma Q\right)^{-1}Q'X\right)^{-1} + \left(X'Q\left(Q'\Sigma Q\right)^{-1}Q'X\right)^{-1} \times \\
& \left(X'Q\left(Q'\Sigma Q\right)^{-1}Q'\left(S-\Sigma\right)Q\left(Q'\Sigma Q\right)^{-1}Q'X\right)\left(X'Q\left(Q'\Sigma Q\right)^{-1}Q'X\right)^{-1}.
\end{aligned}
$$

Define $X^* = \Sigma^{-1/2}X$, $Q^* = \Sigma^{1/2}Q$, $S^* = \Sigma^{-1/2}S\Sigma^{-1/2}$, $\epsilon^* = \Sigma^{-1/2}\epsilon$ and $P^* = Q^*\left(Q^{*\prime}Q^*\right)^{-1}Q^{*\prime}$. After change of variables, we obtain the following.

$$
\begin{aligned}
& \widehat{H}^{-1} \\
= \; & \left(X^{*\prime}P^*X^*\right)^{-1} + \left(X^{*\prime}P^*X^*\right)^{-1}\left(X^{*\prime}P^*\left(S^*-I\right)P^*X^*\right)\left(X^{*\prime}P^*X^*\right)^{-1} \\
& -\left(X^{*\prime}P^*X^*\right)^{-1}\left(X^{*\prime}P^*\left(S^*-I\right)P^*\left(S^*-I\right)P^*X^*\right)\left(X^{*\prime}P^*X^*\right)^{-1}.
\end{aligned}
$$

$$
\widehat{h} = X^{*\prime}P^*\epsilon^* - X^{*\prime}P^*\left(S^*-I\right)P^*\epsilon^* + X^{*\prime}P^*\left(S^*-I\right)P^*\left(S^*-I\right)P^*\epsilon^*.
$$

Now we have

$$
\begin{aligned}
& \widehat{H}^{-1}\widehat{h} \\
= \; & \underbrace{\left(X^{*\prime}PX^{*}\right)^{-1}X^{*\prime}P\epsilon^{*}}_{T1} + \underbrace{\left(X^{*\prime}PX^{*}\right)^{-1}\left(X^{*\prime}P\left(S^{*}-I\right)PX^{*}\right)\left(X^{*\prime}PX^{*}\right)^{-1}X^{*\prime}P\epsilon^{*}}_{T2} \\
& -\underbrace{\left(X^{*\prime}PX^{*}\right)^{-1}\left(X^{*\prime}P\left(S^{*}-I\right)P\left(S^{*}-I\right)PX^{*}\right)\left(X^{*\prime}PX^{*}\right)^{-1}X^{*\prime}P\epsilon^{*}}_{T3} \\
& -\underbrace{\left(X^{*\prime}PX^{*}\right)^{-1}X^{*\prime}P\left(S^{*}-I\right)P\epsilon^{*}}_{T4} \\
& -\underbrace{\left(X^{*\prime}PX^{*}\right)^{-1}\left(X^{*\prime}P\left(S^{*}-I\right)PX^{*}\right)\left(X^{*\prime}PX^{*}\right)^{-1}X^{*\prime}P\left(S^{*}-I\right)P\epsilon^{*}}_{T5} \\
& +\underbrace{\left(X^{*\prime}PX^{*}\right)^{-1}\left(X^{*\prime}P\left(S^{*}-I\right)P\left(S^{*}-I\right)PX^{*}\right)\left(X^{*\prime}PX^{*}\right)^{-1}X^{*\prime}P\left(S^{*}-I\right)P\epsilon^{*}}_{T6} \\
& +\underbrace{\left(X^{*\prime}PX^{*}\right)^{-1}X^{*\prime}P\left(S^{*}-I\right)P\left(S^{*}-I\right)P\epsilon^{*}}_{T7} \\
& +\underbrace{\left(X^{*\prime}PX^{*}\right)^{-1}\left(X^{*\prime}P\left(S^{*}-I\right)PX^{*}\right)\left(X^{*\prime}PX^{*}\right)^{-1}X^{*\prime}P\left(S^{*}-I\right)P\left(S^{*}-I\right)P\epsilon^{*}}_{T8} \\
& -\underbrace{\left(X^{*\prime}PX^{*}\right)^{-1}\left(X^{*\prime}P\left(S^{*}-I\right)P\left(S^{*}-I\right)PX^{*}\right)}_{T9} \\
& \times\underbrace{\left(X^{*\prime}PX^{*}\right)^{-1}X^{*\prime}P\left(S^{*}-I\right)P\left(S^{*}-I\right)P\epsilon^{*}}_{T9}.
\end{aligned}
$$

Furthermore, it is useful to simplify terms T1 to T9 by the following expansion.

$$
\begin{aligned}
\left(X^{*\prime}PX^{*}\right)^{-1} &= \left(X^{*\prime}X^{*}\right)^{-1} + \left(X^{*\prime}X^{*}\right)^{-1}X^{*\prime}\left(I-P\right)X^{*}\left(X^{*\prime}X^{*}\right)^{-1} \\
X^{*\prime}P\epsilon^{*} &= X^{*\prime}\epsilon^{*} - X^{*\prime}\left(I-P\right)\epsilon^{*}.
\end{aligned}
$$

Before calculating the orders of T1 to T9, let's present some results of related orders by the following lemmata.

Lemma A.1.1. $\left(X^{*\prime}X^{*}/n\right)^{-1} = O_p\left(1\right)$ *and* $\left(X^{*\prime}\epsilon^{*}/\sqrt{n}\right) = O_p\left(1\right).$

Proof. By the boundedness of X^* and Central Limit Theorem. $\square$

Lemma A.1.2. $\left[X^{*\prime}\left(I-P^*\right)X^*/n\right]=O_p\left(\|f\|^2\right).$

Proof. By the traditional assmption of series estimation, we have

$$\left[\sum_{i=1}^{n}\left(f\left(x_i^*\right)-q^K\left(x_i^*\right)'\pi\right)^2/n\right]^{1/2}=O\left(K^{-\alpha}\right)=O\left(\|f\|\right).$$

$\square$

Lemma A.1.3. $\left[X^{*\prime}\left(I-P\right)\epsilon^*/\sqrt{n}\right]=O\left(\|f\|\right).$

Proof. It is mean zero, so that

$$Var\left[\frac{X^{*\prime}\left(I-P\right)\epsilon^*}{\sqrt{n}}\right]=\frac{X^{*\prime}\left(I-P\right)X^*}{n}=O_p\left(\|f\|^2\right).$$

By M, the result follows. $\square$

Lemma A.1.4. $\left[X^{*\prime}P\left(S^*-I\right)PX^*/n\right]=\frac{1}{n}X^{*\prime}PS_1PX^*+O\left(1/n\right).$

Proof. We define the following notations.

$$\begin{aligned} S_1 &= \operatorname{diag}\left(\epsilon_i^{*2}-1\right) \\ S_2 &= \operatorname{diag}\left(x_i^*\epsilon_i^*\right) \\ S_3 &= \operatorname{diag}\left(x_i^{*2}\right). \end{aligned}$$

Note that $u_i = \left(\beta - \widehat{\beta}\right) x_i + \epsilon_i$. We have[1]

$$\begin{aligned} u_i^2 &= \left[\left(\beta - \widehat{\beta}\right) x_i + \epsilon_i\right]^2 \\ &= \epsilon_i^2 + \left(\beta - \widehat{\beta}\right)^2 x_i^2 + 2\left(\beta - \widehat{\beta}\right) x_i \epsilon_i. \end{aligned}$$

[1]If we adopt MacKinnon and White (1985) jackknife estimate of covariance matrix, $\mathrm{diag}\left(u^2(x_i)/(1-h_{ii})^2\right)$ will replace $\mathrm{diag}\left(u^2(x_i)\right)$. Note that

$$\begin{aligned} (1-h_{ii})^{-1} &= 1 + h_{ii} + h_{ii}^2 + ... = 1 + h_{ii} \\ (1-h_{ii})^{-2} &= 1 + 2h_{ii} + 3h_{ii}^2 + ... = 1 + 2h_{ii}. \end{aligned}$$

Now we will have

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n} \overline{x}_i^{*2}\left[\left(\epsilon_i^{*2} + \left(\widehat{\beta}-\beta\right)^2 x_i^{*2} - 2\left(\widehat{\beta}-\beta\right) x_i^* \epsilon_i^*\right)(1+2h_{ii}) - 1\right] \\ = & \text{ [terms as in White]} + \frac{2}{n}\sum_{i=1}^{n} \overline{x}_i^{*2}\epsilon_i^{*2}h_{ii} \\ & + \frac{2}{n}\left(\widehat{\beta}-\beta\right)^2 \sum_{i=1}^{n} \overline{x}_i^{*4}h_{ii} + \frac{4}{n}\left(\widehat{\beta}-\beta\right)\sum_{i=1}^{n}\overline{x}_i^{*2}x_i^*\epsilon_i^* h_{ii} \\ = & \text{ [terms as in White]} + \frac{2}{n}\sum_{i=1}^{n}\overline{x}_i^{*2}\left[\epsilon_i^{*2}-1\right]h_{ii} + \frac{2}{n}\sum_{i=1}^{n}\overline{x}_i^{*2}h_{ii} \\ & + \frac{2}{n}\left(\widehat{\beta}-\beta\right)^2 \sum_{i=1}^{n} \overline{x}_i^{*4}h_{ii} + \frac{4}{n}\left(\widehat{\beta}-\beta\right)\sum_{i=1}^{n}\overline{x}_i^{*2}x_i^*\epsilon_i^* h_{ii} \end{aligned}$$

Here are the orders.

$$\frac{2}{n}\sum_{i=1}^{n}\overline{x}_i^{*2}\epsilon_i^{*2}h_{ii} = O_p\left(\frac{d}{n}\right)$$

$$Var\left[\frac{1}{n}\sum_{i=1}^{n}\overline{x}_i^{*2}x_i^*\epsilon_i^*h_{ii}\right] = \frac{1}{n^2}\sum_{i=1}^{n}\overline{x}_i^{*4}x_i^{*2}h_{ii}^2 = O_p\left(\frac{\zeta(K)^2 d}{n^3}\right).$$

Applying the expansion of u_i^2 gives

$$\begin{aligned} & \frac{1}{n} X^{*\prime} P \left(S^* - I\right) P X^* \\ = & \frac{1}{n} X^{*\prime} P \left(\sum_{j=1}^{3} \widetilde{S}_j \right) P X^*, \end{aligned}$$

where $\widetilde{S}_1 = S_1, \widetilde{S}_2 = 2\left(\beta - \widehat{\beta}\right) S_2$, and $\widetilde{S}_3 = \left(\beta - \widehat{\beta}\right)^2 S_3$. By the fact that

$$\frac{1}{n} X^{*\prime} P S_1 P X^* = O\left(\frac{1}{\sqrt{n}}\right)$$

and

$$\frac{1}{n} X^{*\prime} P \left(\sum_{j=1}^{2} \widetilde{S}_j \right) P X^* = O\left(\frac{1}{n}\right),$$

the result follows. □

Lemma A.1.5. $\left[X^{*\prime} P \left(S^* - I\right) P \left(S^* - I\right) P X^* / n\right] = \frac{1}{n} X^{*\prime} P S_1 P S_1 P X^* = O\left(K/n\right)$

Proof. It could be written as

$$\begin{aligned} & \frac{1}{n} X^{*\prime} P \left(\sum_{j=1}^{3} \widetilde{S}_j \right) P \left(\sum_{j=1}^{3} \widetilde{S}_j \right) P X^* \\ = & \frac{1}{n} X^{*\prime} P S_1 P S_1 P X^* + O\left(\frac{1}{n}\right) \\ = & O\left(\frac{K}{n}\right). \end{aligned}$$

□

Lemma A.1.6. $\left[X^{*\prime} P \left(S^* - I\right) P \epsilon^* / \sqrt{n}\right] = \frac{1}{\sqrt{n}} X^{*\prime} P S_1 P \epsilon^* - \frac{2}{\sqrt{n}} \left(\hat{\beta} - \beta\right) X^{*\prime} P S_2 P \epsilon^*$

$+ o\left(\frac{K}{n}\right)$

Proof. Expansion gives

$$\frac{1}{\sqrt{n}}X^{*\prime}P\left(\sum_{j=1}^{3}\widetilde{S}_j\right)P\epsilon^* = \frac{1}{\sqrt{n}}X^{*\prime}PS_1P\epsilon^* - \frac{2}{\sqrt{n}}\left(\hat{\beta}-\beta\right)X^{*\prime}PS_2P\epsilon^* + o\left(\frac{K}{n}\right).$$

We can show

$$\begin{aligned} \frac{1}{\sqrt{n}}X^{*\prime}PS_1P\epsilon^* &= O\left(\frac{\sqrt{K}}{\sqrt{n}}\right) \\ \frac{2}{\sqrt{n}}\left(\beta-\hat{\beta}\right)X^{*\prime}PS_2P\epsilon^* &= O\left(\frac{K}{\sqrt{n}}\right). \end{aligned}$$

The proof ends. □

Lemma A.1.7. $\left[X^{*\prime}P\left(S^*-I\right)P\left(S^*-I\right)P\epsilon^*/\sqrt{n}\right] = \frac{1}{\sqrt{n}}X^{*\prime}PS_1PS_1P\epsilon^* = O\left(\frac{\zeta(K)\sqrt{K}}{n}\right)$

Proof.

$$\begin{aligned} & \frac{1}{\sqrt{n}}X^{*\prime}P\left(\sum_{j=1}^{3}\widetilde{S}_j\right)P\left(\sum_{j=1}^{3}\widetilde{S}_j\right)P\epsilon^* \\ = & \frac{1}{\sqrt{n}}X^{*\prime}PS_1PS_1P\epsilon^* + o\left(\frac{K}{n}\right) \\ = & O\left(\frac{\zeta\left(K\right)K}{n}\right). \end{aligned}$$

□

Now deal with terms $\sqrt{n}$T1 to $\sqrt{n}$T9. To simplify the calculation, we premultiply all of the terms by $\Omega^{*-1} = X^{*\prime}X^*/n$. Furthermore, by Lemma A.1.1-A.1.7 we could drop out lots of terms with small orders.

$$\begin{aligned} \Omega^{*-1}\sqrt{n}T1 &= \left(\frac{X^{*\prime}\epsilon^*}{\sqrt{n}}\right) + \left(\frac{X^{*\prime}\left(I-P\right)X^*}{n}\right)\left(\frac{X^{*\prime}X^*}{n}\right)^{-1}\left(\frac{X^{*\prime}\epsilon^*}{\sqrt{n}}\right) \\ & - \left(\frac{X^{*\prime}\left(I-P\right)\epsilon^*}{\sqrt{n}}\right) + O\left(\frac{1}{n}\right). \end{aligned}$$

$$\Omega^{*-1}\sqrt{n}T2 = \left(\frac{1}{n}X^{*\prime}PS_1PX^*\right)\left(\frac{X^{*\prime}X^*}{n}\right)^{-1}\left(\frac{X^{*\prime}\epsilon^*}{\sqrt{n}}\right) - \left(\frac{1}{n}X^{*\prime}PS_1PX^*\right)\left(\frac{X^{*\prime}X^*}{n}\right)^{-1}\left(\frac{X^{*\prime}\left(I-P\right)\epsilon^*}{\sqrt{n}}\right) + O\left(\frac{1}{n}\right).$$

$$\Omega^{*-1}\sqrt{n}T3 = \left(\frac{1}{n}X^{*\prime}PS_1PS_1PX^*\right)\left(\frac{X^{*\prime}X^*}{n}\right)^{-1}\left(\frac{X^{*\prime}\epsilon^*}{\sqrt{n}}\right) + O\left(\frac{1}{n}\right).$$

$$\Omega^{*-1}\sqrt{n}T4 = \frac{1}{\sqrt{n}}X^{*\prime}PS_1P\epsilon^* - \frac{2}{\sqrt{n}}\left(\hat{\beta}-\beta\right)X^{*\prime}PS_2P\epsilon^* + O\left(\frac{1}{n}\right).$$

$$\Omega^{*-1}\sqrt{n}T5 = \left(\frac{X^{*\prime}PS_1PX^*}{n}\right)\left(\frac{X^{*\prime}X^*}{n}\right)^{-1}\left(\frac{X^{*\prime}PS_1P\epsilon^*}{\sqrt{n}}\right)+O\left(\frac{1}{n}\right) = O\left(\frac{\sqrt{K}}{n}\right).$$

$$\Omega^{*-1}\sqrt{n}T6 = \left(\frac{X^{*\prime}PS_1PS_1PX^*}{n}\right)\left(\frac{X^{*\prime}X^*}{n}\right)^{-1}\left(\frac{X^{*\prime}PS_1P\epsilon^*}{\sqrt{n}}\right)+O\left(\frac{1}{n}\right) = O\left(\frac{1}{n}\right).$$

$$\Omega^{*-1}\sqrt{n}T7 = \frac{X^{*\prime}PS_1PS_1P\epsilon^*}{\sqrt{n}} = O\left(\frac{\zeta\left(K\right)K}{n}\right).$$

$$\Omega^{*-1}\sqrt{n}T8 = \frac{1}{\sqrt{n}}\left(X^{*\prime}PS_1PX^*\right)\left(X^{*\prime}X^*\right)^{-1}\left(X^{*\prime}PS_1PS_1P\epsilon^*\right) = \text{tiny}.$$

$$\Omega^{*-1}\sqrt{n}T9 = \frac{1}{\sqrt{n}}\left(X^{*\prime}PS_1PS_1PX^*\right)\left(X^{*\prime}X^*\right)^{-1}\left(X^{*\prime}PS_1PS_1P\epsilon^*\right) = \text{tiny}.$$

Combining $\Omega^{*-1}\sqrt{n}T1, ..., \Omega^{*-1}\sqrt{n}T9$ gives us the expansion of $\Omega^{*-1}\sqrt{n}\left(\widehat{\beta}_{Cragg} - \beta\right)$.

$$
\begin{aligned}
& \Omega^{*-1}\sqrt{n}\widehat{H}^{-1}\widehat{h} \\
= \; & \Omega^{*-1}\sqrt{n}\sum_{i=1}^{9} Ti \\
= \; & \left(\frac{X^{*\prime}\epsilon^*}{\sqrt{n}}\right) + \left(\frac{X^{*\prime}\left(I-P\right)X^*}{n}\right)\left(\frac{X^{*\prime}X^*}{n}\right)^{-1}\left(\frac{X^{*\prime}\epsilon^*}{\sqrt{n}}\right) \\
& - \left(\frac{X^{*\prime}\left(I-P\right)\epsilon^*}{\sqrt{n}}\right) + \left(\frac{1}{n}X^{*\prime}PS_1PX^*\right)\left(\frac{X^{*\prime}X^*}{n}\right)^{-1}\left(\frac{X^{*\prime}\epsilon^*}{\sqrt{n}}\right) \\
& - \left(\frac{1}{n}X^{*\prime}PS_1PX^*\right)\left(\frac{X^{*\prime}X^*}{n}\right)^{-1}\left(\frac{X^{*\prime}\left(I-P\right)\epsilon^*}{\sqrt{n}}\right) \\
& - \left(\frac{1}{n}X^{*\prime}PS_1PS_1PX^*\right)\left(\frac{X^{*\prime}X^*}{n}\right)^{-1}\left(\frac{X^{*\prime}\epsilon^*}{\sqrt{n}}\right) \\
& - \frac{1}{\sqrt{n}}X^{*\prime}PS_1P\epsilon^* + \frac{2}{\sqrt{n}}\left(\hat{\beta} - \beta\right)X^{*\prime}PS_2P\epsilon^* \\
& + \frac{1}{\sqrt{n}}X^{*\prime}PS_1PS_1P\epsilon^* + O\left(\frac{1}{\sqrt{n}}\right) \\
= \; & \sum_{j=1}^{9} U_j + O\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

Note that the definitions and orders of U1 to U9 are listed below.

$$
\begin{aligned}
U_1 &= \frac{X^{*\prime}\epsilon^*}{\sqrt{n}} = O(1) \\
U_2 &= \left(\frac{X^{*\prime}(I-P)X^*}{n}\right)\left(\frac{X^{*\prime}X^*}{n}\right)^{-1}\left(\frac{X^{*\prime}\epsilon^*}{\sqrt{n}}\right) = O\left(\|f\|^2\right) \\
U_3 &= -\left(\frac{X^{*\prime}(I-P)\epsilon^*}{\sqrt{n}}\right) = O(\|f\|) \\
U_4 &= \left(\frac{1}{n}X^{*\prime}PS_1PX^*\right)\left(\frac{X^{*\prime}X^*}{n}\right)^{-1}\left(\frac{X^{*\prime}\epsilon^*}{\sqrt{n}}\right) = O\left(\frac{1}{\sqrt{n}}\right) \\
U_5 &= -\left(\frac{1}{n}X^{*\prime}PS_1PX^*\right)\left(\frac{X^{*\prime}X^*}{n}\right)^{-1}\left(\frac{X^{*\prime}(I-P)\epsilon^*}{\sqrt{n}}\right) = O\left(\frac{\|f\|}{\sqrt{n}}\right) \\
U_6 &= -\left(\frac{1}{n}X^{*\prime}PS_1PS_1PX^*\right)\left(\frac{X^{*\prime}X^*}{n}\right)^{-1}\left(\frac{X^{*\prime}\epsilon^*}{\sqrt{n}}\right) = O\left(\frac{K}{n}\right) \\
U_7 &= -\frac{1}{\sqrt{n}}X^{*\prime}PS_1P\epsilon^* = O\left(\frac{\sqrt{K}}{\sqrt{n}}\right) \\
U_8 &= \frac{2}{\sqrt{n}}\left(\hat{\beta}-\beta\right)X^{*\prime}PS_2P\epsilon^* = O\left(\frac{K}{\sqrt{n}}\right) \\
U_9 &= \frac{1}{\sqrt{n}}X^{*\prime}PS_1PS_1P\epsilon^* = O\left(\frac{\zeta(K)K}{n}\right)
\end{aligned}
$$

Now define

$$
\begin{aligned}
\Xi_1 &= (Q'Q)^{-1}\sum_i \kappa_i x_i^{*2}Q_iQ_i \\
\Xi_2 &= (Q'Q)^{-1}\sum_i x_i^{*2}Q_iQ_i \\
\kappa_i &= E\left[\epsilon_i^{*4}|x_i^*\right].
\end{aligned}
$$

$$
E[U_1U_1'] = E\left[\frac{X^{*\prime}\epsilon^*}{\sqrt{n}}\frac{\epsilon^{*\prime}X^*}{\sqrt{n}}\right] = \frac{X^{*\prime}X^*}{n} = O(1)
$$

$$
\begin{aligned}
E\left[U_3U_3'\right] &= E\left[\frac{X^{*\prime}\left(I-P\right)\epsilon^*}{\sqrt{n}}\frac{\epsilon^{*\prime}\left(I-P\right)X^*}{\sqrt{n}}\right] \\
&= \frac{X^{*\prime}\left(I-P\right)X^*}{n} \\
&= O\left(\|f\|^2\right)
\end{aligned}
$$

$$
\begin{aligned}
E\left[U_7U_7'\right] &= E\left[\frac{X^{*\prime}S_1P\epsilon^*}{\sqrt{n}}\frac{\epsilon^{*\prime}PS_1X^*}{\sqrt{n}}\right] \\
&= \frac{1}{n}\sum_i x_i^{*2}P_{ii}^2E\left[\left(\epsilon_i^{*2}-1\right)^2\right]\epsilon_i^{*2} \\
&= O\left(\frac{K}{n}\right) \\
&= \frac{1}{n}tr\left(\Xi_1\right)-\frac{1}{n}tr\left(\Xi_2\right)+o\left(\frac{K}{n}\right)
\end{aligned}
$$

$$
\begin{aligned}
2E\left[U_2U_1'\right] &= 2E\left[\left(\frac{X^{*\prime}\left(I-P\right)X^*}{n}\right)\left(\frac{X^{*\prime}X^*}{n}\right)^{-1}\left(\frac{X^{*\prime}\epsilon^*}{\sqrt{n}}\right)\left(\frac{\epsilon^{*\prime}X^*}{\sqrt{n}}\right)\right] \\
&= 2\frac{X^{*\prime}\left(I-P\right)X^*}{n} \\
&= O\left(\|f\|^2\right)
\end{aligned}
$$

$$
\begin{aligned}
2E\left[U_1U_3'\right] &= -2E\left[\frac{X^{*\prime}\epsilon^*}{\sqrt{n}}\frac{\epsilon^{*\prime}\left(I-P\right)X^*}{\sqrt{n}}\right] \\
&= -2\frac{X^{*\prime}\left(I-P\right)X^*}{n} \\
&= O\left(\|f\|^2\right)
\end{aligned}
$$

$$
\begin{aligned}
2E\left[U_4U_1'\right] &= 2E\left[\left(\frac{1}{n}X^{*\prime}PS_1PX^*\right)\left(\frac{X^{*\prime}X^*}{n}\right)^{-1}\left(\frac{X^{*\prime}\epsilon^*}{\sqrt{n}}\right)\frac{\epsilon^{*\prime}X^*}{\sqrt{n}}\right] \\
&= \frac{2}{n^2}E\left[X^{*\prime}PS_1PX^*\Omega^{*-1}X^{*\prime}\epsilon^*\epsilon^{*\prime}X^*\right] \\
&= \frac{2}{n^2}E\left[\sum_{ijk}\overline{X^*}_i^2\left(\epsilon_i^{*2}-1\right)\Omega^{*-1}X_j^*\epsilon_j^*X_k^*\epsilon_k^*\right] \\
&= \frac{2}{n^2}\sum_i\overline{X^*}_i^2\Omega^{*-1}X_i^{*2}E\left[\left(\epsilon_i^{*2}-1\right)\epsilon_i^{*2}\right] \\
&= O\left(\frac{1}{n}\right)
\end{aligned}
$$

$$
\begin{aligned}
2E\left[U_5U_1'\right] &= -2E\left[\left(\frac{1}{n}X^{*\prime}PS_1PX^*\right)\left(\frac{X^{*\prime}X^*}{n}\right)^{-1}\left(\frac{X^{*\prime}\left(I-P\right)\epsilon^*}{\sqrt{n}}\right)\frac{\epsilon^{*\prime}X^*}{\sqrt{n}}\right] \\
&= -\frac{2}{n^2}E\left[X^{*\prime}PS_1PX^*\Omega^{*-1}X^{*\prime}\left(I-P\right)\epsilon^*\epsilon^*X^*\right] \\
&= -\frac{2}{n^2}E\left[\sum_{ijk}\overline{X^*}_i^2\left(\epsilon_i^{*2}-1\right)\Omega^{*-1}\left(X_j^*-\overline{X^*}_j\right)\epsilon_j^*X_k^*\epsilon_k^*\right] \\
&= -\frac{2}{n^2}\sum_i\overline{X^*}_i^2\Omega^{*-1}\left(X_i^*-\overline{X^*}_i\right)E\left[\left(\epsilon_i^{*2}-1\right)\epsilon_i^{*2}\right] \\
&= o\left(\frac{\sqrt{K}}{n}\right)
\end{aligned}
$$

$$
\begin{aligned}
2E\left[U_6U_1'\right] &= -2E\left[\left(\frac{1}{n}X^{*\prime}PS_1PS_1PX^*\right)\left(\frac{X^{*\prime}X^*}{n}\right)^{-1}\left(\frac{X^{*\prime}\epsilon^*}{\sqrt{n}}\right)\frac{\epsilon^{*\prime}X^*}{\sqrt{n}}\right] \\
&= -\frac{2}{n^2}E\left[X^{*\prime}PS_1PS_1PX^*\Omega^{*-1}X^{*\prime}\epsilon^*\epsilon^*X^*\right] \\
&= -\frac{2}{n^2}E\left[\sum_{ijkl}\overline{X^*}_i\overline{X^*}_j\left(\epsilon_i^{*2}-1\right)\left(\epsilon_j^{*2}-1\right)P_{ij}\Omega^{*-1}X_l^*\epsilon_l^*X_k^*\epsilon_k^*\right] \\
&= -\frac{2}{n}\left[tr\left(\Xi_1\right)-tr\left(\Xi_2\right)\right]+o\left(\frac{K}{n}\right)
\end{aligned}
$$

$$
\begin{aligned}
2E\left[U_7U_1'\right] &= -2E\left[\left(\frac{1}{n}X^{*\prime}PS_1P\epsilon^*\right)\frac{\epsilon^{*\prime}X^*}{\sqrt{n}}\right] \\
&= -\frac{2}{n^2}E\left[X^{*\prime}PS_1P\epsilon^*\epsilon^{*\prime}X^*\right] \\
&= -\frac{2}{n^2}E\left[\sum_{ijk}\overline{X^*}_i\left(\epsilon_i^{*2}-1\right)P_{ij}\epsilon_j^*X_k^*\epsilon_k^*\right] \\
&= -\frac{2}{n^2}\sum_i\overline{X^*}_i^2X_i^*P_{ii}E\left[\left(\epsilon_i^{*2}-1\right)\epsilon_i^{*2}\right] \\
&= -\frac{2}{n}\left[tr\left(\Xi_1\right)-tr\left(\Xi_2\right)\right]+o\left(\frac{K}{n}\right)
\end{aligned}
$$

$$
\begin{aligned}
2E\left[U_8U_1'\right] &= 2E\left[\frac{2}{\sqrt{n}}\left(\hat{\beta}-\beta\right)X^{*\prime}PS_2P\epsilon^*\frac{\epsilon^{*\prime}X^*}{\sqrt{n}}\right] \\
&= \frac{4}{n^2}E\left[\sum_{ijkl}\phi_i\epsilon_i^*\overline{X^*}_j\epsilon_j^*X_j^*P_{jk}\epsilon_k^*\epsilon_l^*X_l^*\right] \\
&= \frac{4}{n^2}\left[\sum_i\phi_i\overline{X^*}_iX_i^{*2}P_{ii}E\left(\epsilon_i^{*4}\right)\right]+\left[\frac{4}{n^2}\sum_{i\neq j}\phi_iX_i^*\overline{X^*}_jX_j^*P_{jj}\right] \\
&\quad +\frac{4}{n^2}\left[\sum_{i\neq j}\phi_i\overline{X^*}_jX_j^{*2}P_{ji}\right]+\frac{4}{n^2}\left[\sum_{i\neq j}\phi_i\overline{X^*}_iX_i^*P_{ij}\right] \\
&= \frac{4}{n}tr\left(\Xi_2\right)+o\left(\frac{K}{n}\right)
\end{aligned}
$$

$$
\begin{aligned}
2E\left[U_9U_1'\right] &= 2E\left[\left(\frac{1}{\sqrt{n}}X^{*\prime}PS_1PS_1P\epsilon^*\right)\frac{\epsilon^{*\prime}X^*}{\sqrt{n}}\right] \\
&= \frac{2}{n}\left[tr\left(\Xi_1\right)-tr\left(\Xi_2\right)\right]+o\left(\frac{K}{n}\right)
\end{aligned}
$$

Note that all other terms of tiny orders are neglected. The mean of $\Omega^{*-1}\sqrt{n}\left(\widehat{\beta}_{Cragg}-\beta\right)$ is zero. Hence the MSE of $\Omega^{*-1}\sqrt{n}\left(\widehat{\beta}_{Cragg}-\beta\right)$ will be

$$
\frac{1}{n}X^{*\prime}X^*+\frac{1}{n}X^{*\prime}\left(I-P\right)X^*-\frac{1}{n}tr\left(\Xi_1\right)+\frac{5}{n}tr\left(\Xi_2\right)+o\left(\frac{K}{n}\right).
$$

Q.E.D.

A.2 Proof of Proposition 2.3.2

The series based FGLS estimator $\widehat{\beta}_{Series}$ could be written as

$$\sqrt{N}\left(\widehat{\beta}_{Series}-\beta\right)=\left[\frac{1}{n}\sum_i\frac{x_i^2}{\widehat{\sigma}_i^2}\right]^{-1}\left[\frac{1}{\sqrt{n}}\sum_i\frac{x_i\epsilon_i}{\widehat{\sigma}_i^2}\right]=\widehat{H}^{-1}\widehat{h}.$$

We could expand $\widehat{H}$ and $\widehat{h}$ respectively as

$$\widehat{H}=\frac{1}{n}\sum_i\frac{x_i^2}{\sigma_i^2}-\frac{1}{n}\sum_i\frac{x_i^2}{\sigma_i^4}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)+\frac{1}{n}\sum_i\frac{x_i^2}{\sigma_i^6}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)^2$$

and

$$\widehat{h}=\frac{1}{\sqrt{n}}\sum_i\frac{x_i\epsilon_i}{\sigma_i^2}-\frac{1}{\sqrt{n}}\sum_i\frac{x_i\epsilon_i}{\sigma_i^4}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)+\frac{1}{\sqrt{n}}\sum_i\frac{x_i\epsilon_i}{\sigma_i^6}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)^2.$$

In matrix form, they are

$$\begin{aligned}\widehat{H} &= \frac{1}{n}X^{*\prime}X^{*}-\frac{1}{n}\left[X^{*\prime}\left(\widehat{\Sigma}^{*}-I\right)X^{*}\right]+\frac{1}{n}\left[X^{*\prime}\left(\widehat{\Sigma}^{*}-I\right)\left(\widehat{\Sigma}^{*}-I\right)X^{*}\right]\\ &= H+T^{H}+Z^{H}\end{aligned}$$

and

$$\begin{aligned}\widehat{h} &= \frac{1}{\sqrt{n}}X^{*\prime}\epsilon^{*}-\frac{1}{\sqrt{n}}\left[X^{*\prime}\left(\widehat{\Sigma}^{*}-I\right)\epsilon^{*}\right]+\frac{1}{\sqrt{n}}\left[X^{*\prime}\left(\widehat{\Sigma}^{*}-I\right)\left(\widehat{\Sigma}^{*}-I\right)\epsilon^{*}\right]\\ &= h+T^{h}+Z^{h}.\end{aligned}$$

To simplify the calculation of MSE for series based FGLS estimator, we will verify the Lemma 1 of Donald and Newey (2001) and then compute $E\left[\widehat{A}(K)\right]$ in Lemma 1. Before the calulation, we present Lemmata of leading terms in expansion. The following Lemmata hold for series based FGLS estimator.

Lemma A.2.1. $\left[X^{*\prime}\left(\widehat{\Sigma}^{*}-I\right)X^{*}/n\right] = \frac{1}{n}\sum_i \frac{x_i^{*2}}{\sigma_i^2}\left(\overline{\sigma}_i^2-\sigma_i^2\right)+\frac{1}{n}\sum_i \frac{x_i^{*2}}{\sigma_i^2}v_i + O_p\left(\frac{1}{n}\right).$

Proof. We could rewrite $X^{*\prime}\left(\widehat{\Sigma}^{*}-I\right)X^{*}/n$ as

$$\frac{1}{n}X^{*\prime}\left(\widehat{\Sigma}^{*}-I\right)X^{*}=\frac{1}{n}\sum\left[x_i^{*2}\left(\frac{1}{\sigma_i^2}\widehat{\sigma}_i^2-1\right)\right]. \tag{A.2}$$

The series estimator for the diagonal element of Σ is

$$\widehat{\sigma}_i^2=q^K\left(x_i\right)'\left(Q'Q\right)^{-1}Q'\mathbf{e2},$$

where $q^K\left(x_i\right)$ is a $k\times 1$ vector and Q is a $n\times k$ approximating function. $\mathbf{e2}$ is a $n\times 1$ vector of squared residuals from OLS estimation.[2] The matrix form of the OLS squared residuals expansion is

$$\mathbf{e2}=\epsilon\mathbf{2}+2\left(\beta-\hat{\beta}\right)\mathbf{x}\epsilon+\left(\beta-\hat{\beta}\right)^2\mathbf{x2},$$

where

$$\begin{aligned}\epsilon\mathbf{2} &= \left(\epsilon_1^2,\epsilon_2^2,...,\epsilon_n^2\right)' \\ 2\left(\beta-\hat{\beta}\right)\mathbf{x}\epsilon &= 2\left(\beta-\hat{\beta}\right)\left(x_1\epsilon_1,x_2\epsilon_2,...,x_n\epsilon_n\right)' \\ \left(\beta-\hat{\beta}\right)^2\mathbf{x2} &= \left(\beta-\hat{\beta}\right)^2\left(x_1^2,x_2^2,...,x_n^2\right)'.\end{aligned}$$

Now $\widehat{\sigma}_i^2$ could be expanded as

$$\widehat{\sigma}_i^2=q^K\left(x_i\right)'\left(Q'Q\right)^{-1}Q'\left[\epsilon\mathbf{2}+2\left(\beta-\hat{\beta}\right)\mathbf{x}\epsilon+\left(\beta-\hat{\beta}\right)^2\mathbf{x2}\right]. \tag{A.3}$$

[2]More specifically, $Q=\begin{bmatrix}q^K\left(x_1\right)'\\ q^K\left(x_2\right)'\\ \vdots\\ q^K\left(x_n\right)'\end{bmatrix}$ and $\mathbf{e2}=\begin{bmatrix}e_1^2\\ e_2^2\\ \vdots\\ e_n^2\end{bmatrix}$.

Plugging equation $(A.3)$ into equation $(A.2)$ gives us

$$
\begin{aligned}
& \frac{1}{n}\sum\left[x_i^{*2}\left(\frac{1}{\sigma^2}\widehat{\sigma}_i^2-1\right)\right] \\
= & \frac{1}{n}\sum\left[x_i^{*2}\left[\frac{1}{\sigma_i^2}q^{K\prime}(x_i)(Q'Q)^{-1}Q'\left[\epsilon\mathbf{2}+2\left(\beta-\hat{\beta}\right)\mathbf{x}\epsilon+\left(\beta-\hat{\beta}\right)^2\mathbf{x2}\right]-1\right]\right] \\
= & \frac{1}{n}\left[\mathbf{xs2}'Q(Q'Q)^{-1}Q'\epsilon\mathbf{2}-\mathbf{xs}'\iota\right] \\
& +2\left(\beta-\hat{\beta}\right)\cdot\frac{1}{n}\left[\mathbf{xs2}'Q(Q'Q)^{-1}Q'\mathbf{x}\epsilon\right] \\
& +\left(\beta-\hat{\beta}\right)^2\cdot\frac{1}{n}\left[\mathbf{xs2}'Q(Q'Q)^{-1}Q'\mathbf{x2}\right] \\
= & \frac{1}{n}\left[\mathbf{xs2}'P\epsilon\mathbf{2}-\mathbf{xs}'\iota\right] \qquad (A.4) \\
& +2\left(\beta-\hat{\beta}\right)\cdot\frac{1}{n}\left[\mathbf{xs2}'P\mathbf{x}\epsilon\right] \qquad (A.5) \\
& +\left(\beta-\hat{\beta}\right)^2\cdot\frac{1}{n}\left[\mathbf{xs2}'P\mathbf{x2}\right], \qquad (A.6)
\end{aligned}
$$

where the projection matrix P is $Q(Q'Q)^{-1}Q'$, and

$$
\begin{aligned}
\mathbf{x2} &= \left(x_1^2, x_2^2, ..., x_n^2\right)' \\
\mathbf{xs2} &= \left(x_1^2/\sigma^4(x_1), x_2^2/\sigma^4(x_2), ..., x_n^2/\sigma^4(x_n)\right)' \\
\mathbf{xs} &= \left(x_1^2/\sigma^2(x_1), x_2^2/\sigma^2(x_2), ..., x_n^2/\sigma^2(x_n)\right)' \\
\iota &= (1, ..., 1)'.
\end{aligned}
$$

Let's denote

$$
\begin{aligned}
\epsilon\mathbf{2} &= E(\epsilon\mathbf{2}|\mathbf{x})+\mathbf{v} \\
&= \sigma\mathbf{2}+\mathbf{v},
\end{aligned}
$$

where $E(\mathbf{v}|\mathbf{x}) = 0$ and $\sigma\mathbf{2} = \left(\sigma^2(x_1), ..., \sigma^2(x_n)\right)'$. Also $\overline{v}_i = \sum_j P_{ij}v_j$.

Rewrite the term in (A.4) as follows.

$$\begin{aligned}
& \frac{1}{n}\left[\mathbf{xs2}'P\epsilon\mathbf{2}-\mathbf{xs}'\iota\right] \\
= & \frac{1}{n}\left[\mathbf{xs2}'P\left(\sigma\mathbf{2}+\mathbf{v}\right)-\mathbf{xs}'\iota\right] \\
= & \frac{1}{n}\left[\mathbf{xs2}'P\sigma\mathbf{2}+\mathbf{xs2}'P\mathbf{v}-\mathbf{xs}'\iota\right] \\
= & \frac{1}{n}\left[\mathbf{xs2}'\sigma\mathbf{2}-\mathbf{xs2}'\left(I-P\right)\sigma\mathbf{2}+\mathbf{xs2}'\mathbf{v}-\mathbf{xs2}'\left(I-P\right)\mathbf{v}-\mathbf{xs}'\iota\right] \\
= & \frac{1}{n}\left[-\mathbf{xs2}'\left(I-P\right)\sigma\mathbf{2}+\mathbf{xs2}'\mathbf{v}-\mathbf{xs2}'\left(I-P\right)\mathbf{v}\right] \\
= & O\left(\|f\|\left\|\sigma^2\right\|\right)+O\left(\frac{1}{\sqrt{n}}\right)+O\left(\frac{\|f\|}{\sqrt{n}}\right) \\
= & \frac{1}{n}\sum_i\frac{x_i^{*2}}{\sigma_i^2}\left(\overline{\sigma}_i^2-\sigma_i^2\right)+\frac{1}{n}\sum_i\frac{x_i^{*2}}{\sigma_i^2}v_i.
\end{aligned}$$

We know that

$$\begin{aligned}
-\frac{1}{n}\left|\mathbf{xs2}'\left(I-P\right)\sigma\mathbf{2}\right| & \leq \left[\frac{\mathbf{xs2}'\left(I-P\right)\mathbf{xs2}}{n}\right]^{1/2}\left[\frac{\sigma\mathbf{2}'\left(I-P\right)\sigma\mathbf{2}}{n}\right]^{1/2} \\
& = O\left(K^{-\alpha_{xs2/\sigma}}\right)\cdot O\left(K^{-\alpha_{\sigma\mathbf{2}/\sigma}}\right) \\
& = O\left(\|f\|\left\|\sigma^2\right\|\right).
\end{aligned}$$

Also we have

$$\begin{aligned}
& \frac{1}{n}\mathbf{xs2}'P\mathbf{x}\epsilon \\
= & \frac{1}{n}\mathbf{xs2}'\mathbf{x}\epsilon-\frac{1}{n}\mathbf{xs2}'\left(I-P\right)\mathbf{x}\epsilon \\
= & O\left(\frac{1}{\sqrt{n}}\right)+O\left(\frac{\|f\|}{\sqrt{n}}\right).
\end{aligned}$$

The order of the term in (A.5) will be

$$2\left(\beta-\hat{\beta}\right)\cdot\frac{1}{n}\left[\mathbf{xs2}'P\mathbf{x}\epsilon\right]=O\left(\frac{1}{n}\right)+O\left(\frac{\|f\|}{n}\right).$$

Next,

$$\begin{aligned}
\frac{1}{n}\mathbf{xs2}'P\mathbf{x2} &= \frac{1}{n}\mathbf{xs2}'\mathbf{x2}-\frac{1}{n}\mathbf{xs2}'\left(I-P\right)\mathbf{x2} \\
&= O\left(1\right)+O\left(\|f\|^2\right).
\end{aligned}$$

The order of the term in (A.6) will be

$$\left(\beta-\hat{\beta}\right)^2\cdot\frac{1}{n}\left[\mathbf{xs2}'P\mathbf{x2}\right]=O\left(\frac{1}{n}\right)+O\left(\frac{\|f\|^2}{n}\right).$$

Therefore, the order of equation (A.2) will be

$$\begin{aligned}
&\frac{1}{n}X^{*\prime}\left(\widehat{\Sigma}^*-I\right)X^* \\
= &\frac{1}{n}\left[-\mathbf{xs2}'\left(I-P\right)\sigma\mathbf{2}+\mathbf{xs2}'\mathbf{v}-\mathbf{xs2}'\left(I-P\right)\mathbf{v}\right]+O\left(\frac{1}{n}\right) \\
= &\frac{1}{n}\sum_i\frac{x_i^{*2}}{\sigma_i^2}\left(\overline{\sigma}_i^2-\sigma_i^2\right)+\frac{1}{n}\sum_i\frac{x_i^{*2}}{\sigma_i^2}v_i+O\left(\frac{1}{n}\right).
\end{aligned}$$

□

Lemma A.2.2.

$$\begin{aligned}
&\left[X^{*\prime}\left(\widehat{\Sigma}^*-I\right)\left(\widehat{\Sigma}^*-I\right)X^*/n\right] \\
= &\frac{1}{n}\sum\left[\frac{x_i^{*2}}{\sigma_i^4}\left[\overline{\sigma}_i^2-\sigma_i^2\right]^2\right]+\frac{1}{n}\sum\left[\frac{x_i^{*2}}{\sigma_i^4}\left[q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{v}\right]^2\right]+O\left(\frac{1}{n}\right)
\end{aligned}$$

Proof. It could be rewritten as

$$
\begin{aligned}
& \frac{1}{n} X^{*\prime} \left(\widehat{\Sigma}^* - I \right) \left(\widehat{\Sigma}^* - I \right) X^* \\
= \; & \frac{1}{n} \sum \left[x_i^{*2} \left(\widehat{\sigma}_i^{*2} - 1 \right)^2 \right] \\
= \; & \frac{1}{n} \sum \left[x_i^{*2} \left[\frac{1}{\sigma_i^2} q^{K\prime} (x_i) (Q'Q)^{-1} Q' \left[\epsilon \mathbf{2} + 2 \left(\beta - \hat{\beta} \right) \mathbf{x} \epsilon + \left(\beta - \hat{\beta} \right)^2 \mathbf{x2} \right] - 1 \right]^2 \right] \\
= \; & \frac{1}{n} \sum \left[x_i^{*2} \left[\frac{1}{\sigma_i^2} q^{K\prime} (x_i) (Q'Q)^{-1} Q' \epsilon \mathbf{2} - 1 \right]^2 \right] && \text{(A.7)} \\
& + 4 \left(\beta - \hat{\beta} \right)^2 \frac{1}{n} \sum \left[x_i^{*2} \left[\frac{1}{\sigma_i^2} q^{K\prime} (x_i) (Q'Q)^{-1} Q' \mathbf{x} \epsilon \right]^2 \right] && \text{(A.8)} \\
& + \left(\beta - \hat{\beta} \right)^4 \frac{1}{n} \sum \left[x_i^{*2} \left[\frac{1}{\sigma_i^2} q^{K\prime} (x_i) (Q'Q)^{-1} Q' \mathbf{x2} \right]^2 \right] && \text{(A.9)} \\
& + 4 \left(\beta - \hat{\beta} \right) \frac{1}{n} \sum \left[x_i^{*2} \left[\frac{1}{\sigma_i^2} \overline{\sigma}_K^2 (x_i) - 1 \right] \left[\frac{1}{\sigma_i^2} q^{K\prime} (x_i) (Q'Q)^{-1} Q' \mathbf{x} \epsilon \right] \right] && \text{(A.10)} \\
& + 2 \left(\beta - \hat{\beta} \right)^2 \frac{1}{n} \sum \left[x_i^{*2} \left[\frac{1}{\sigma_i^2} \overline{\sigma}_K^2 (x_i) - 1 \right] \left[\frac{1}{\sigma_i^2} q^{K\prime} (x_i) (Q'Q)^{-1} Q' \mathbf{x2} \right] \right] && \text{(A.11)} \\
& + 4 \left(\beta - \hat{\beta} \right)^3 \frac{1}{n} \sum \left[\frac{x_i^{*2}}{\sigma_i^4} \left[q^{K\prime} (x_i) (Q'Q)^{-1} Q' \mathbf{x} \epsilon \right] \left[q^{K\prime} (x_i) (Q'Q)^{-1} Q' \mathbf{x2} \right] \right], && \text{(A.12)}
\end{aligned}
$$

where $\overline{\sigma}^2_K(x_i) = q^{K\prime}(x_i)(Q'Q)^{-1}Q'\epsilon\mathbf{2}$. The order of term (A.7) is

$$\begin{aligned}
& \frac{1}{n}\sum\left[x_i^{*2}\left[\frac{1}{\sigma^2(x_i)}q^{K\prime}(x_i)(Q'Q)^{-1}Q'\epsilon\mathbf{2}-1\right]^2\right] \\
= & \frac{1}{n}\sum\left[\frac{x_i^{*2}}{\sigma^4(x_i)}\left[q^{K\prime}(x_i)(Q'Q)^{-1}Q'(\sigma\mathbf{2}+\mathbf{v})-\sigma_i^2\right]^2\right] \\
= & \frac{1}{n}\sum\left[\frac{x_i^{*2}}{\sigma^4(x_i)}\left[\left(q^{K\prime}(x_i)(Q'Q)^{-1}Q'\sigma\mathbf{2}-\sigma_i^2\right)+q^{K\prime}(x_i)(Q'Q)^{-1}Q'\mathbf{v}\right]^2\right] \\
= & \frac{1}{n}\sum\left[\frac{x_i^{*2}}{\sigma^4(x_i)}\left[q^{K\prime}(x_i)(Q'Q)^{-1}Q'\sigma\mathbf{2}-\sigma_i^2\right]^2\right] \\
& +\frac{1}{n}\sum\left[\frac{x_i^{*2}}{\sigma^4(x_i)}\left[q^{K\prime}(x_i)(Q'Q)^{-1}Q'\mathbf{v}\right]^2\right] \\
& +\frac{2}{n}\sum\left[\frac{x_i^{*2}}{\sigma^4(x_i)}\left[\left(q^{K\prime}(x_i)(Q'Q)^{-1}Q'\sigma\mathbf{2}-\sigma_i^2\right)\left(q^{K\prime}(x_i)(Q'Q)^{-1}Q'\mathbf{v}\right)\right]\right] \\
= & O\left(\left\|\sigma^2\right\|^2\right)+O\left(\frac{K}{n}\right)+O\left(\frac{\zeta(K)^2\left\|\sigma^2\right\|}{n\sqrt{n}}\right).
\end{aligned}$$

Define

$$\begin{aligned}
& q^{K\prime}(x_i)(Q'Q)^{-1}Q'\sigma\mathbf{2} \\
= & \sum_j P_{ij}\sigma_j^2 \\
= & \overline{\sigma}_i^2
\end{aligned}$$

We know that

$$\begin{aligned}
& \frac{1}{n}\sum\left[x_i^{*2}\left[\frac{1}{\sigma_i^2}q^{K\prime}(x_i)(Q'Q)^{-1}Q'\sigma\mathbf{2}-1\right]^2\right] \\
= & \frac{1}{n}\sum\left[\frac{x_i^{*2}}{\sigma_i^4}\left[q^{K\prime}(x_i)(Q'Q)^{-1}Q'\sigma\mathbf{2}-\sigma_i^2\right]^2\right] \\
\leq & \Delta\frac{1}{n}\sum\left[\overline{\sigma}_i^2-\sigma_i^2\right]^2 \\
= & \Delta\frac{1}{n}\sigma\mathbf{2}'(I-P)\sigma\mathbf{2} \\
= & O\left(\left\|\sigma^2\right\|^2\right).
\end{aligned}$$

$$
\begin{aligned}
& \frac{1}{n}\sum\left[\frac{x_i^{*2}}{\sigma^4(x_i)}\left[q^{K\prime}(x_i)(Q'Q)^{-1}Q'\mathbf{v}\right]^2\right] \\
\leq\ & \frac{1}{n}\sum\left[q^{K\prime}(x_i)(Q'Q)^{-1}Q'\mathbf{v}\right]^2 \\
=\ & \frac{1}{n}\mathbf{v}'P\mathbf{v} \\
=\ & O\left(\frac{K}{n}\right).
\end{aligned}
$$

Let

$$
\begin{aligned}
\varphi &= \frac{2}{n}\sum_i\left[\frac{x_i^{*2}}{\sigma^4(x_i)}\left[\left(q^{K\prime}(x_i)(Q'Q)^{-1}Q'\sigma\mathbf{2}-\sigma_i^2\right)\left(q^{K\prime}(x_i)(Q'Q)^{-1}Q'\mathbf{v}\right)\right]\right] \\
&= \frac{2}{n}\sum_{ij}\left[\frac{x_i^{*2}}{\sigma_i^4}\left[\left(\overline{\sigma}_i^2-\sigma_i^2\right)P_{ij}v_j\right]\right].
\end{aligned}
$$

It is easy to verify that $E[\varphi]=0$. The variance of φ is

$$
\begin{aligned}
& \frac{4}{n^2}E\left[\sum_{ij}\frac{x_i^{*4}}{\sigma_i^8}\left[\left(\overline{\sigma}_i^2-\sigma_i^2\right)^2P_{ij}^2v_j^2\right]\right] \\
=\ & \frac{4}{n^2}\sum_{ij}\left[\frac{x_i^{*4}}{\sigma_i^8}\left(\overline{\sigma}_i^2-\sigma_i^2\right)^2P_{ij}^2E\left[v_j^2\right]\right] \\
\leq\ & \Delta\frac{4}{n^2}\sum_{ij}\left[\frac{x_i^{*4}}{\sigma_i^8}\left(\overline{\sigma}_i^2-\sigma_i^2\right)^2P_{ij}^2\right] \\
=\ & \Delta\frac{4}{n^2}\sum_{i}\left[\frac{x_i^{*4}}{\sigma_i^8}\left(\overline{\sigma}_i^2-\sigma_i^2\right)^2P_{ii}^2\right]+\Delta\frac{4}{n^2}\sum_{i\neq j}\left[\frac{x_i^{*4}}{\sigma_i^8}\left(\overline{\sigma}_i^2-\sigma_i^2\right)^2P_{ij}^2\right] \\
\leq\ & \Delta\frac{4}{n^2}\left|\sup P_{ii}^2\right|\sum_i\left(\overline{\sigma}_i^2-\sigma_i^2\right)^2+\Delta\frac{4}{n^2}\left|\sup P_{ii}^2\right|\sum_i\left[\left(\overline{\sigma}_i^2-\sigma_i^2\right)^2\right] \\
=\ & O\left(\frac{\zeta(K)^4\left\|\sigma^2\right\|^2}{n^3}\right).
\end{aligned}
$$

Therefore,

$$
\varphi = O\left(\frac{\zeta(K)^2\left\|\sigma^2\right\|}{n\sqrt{n}}\right).
$$

The order of term (A.7) is

$$O\left(\left\|\sigma^2\right\|^2\right)+O\left(\frac{K}{n}\right)+O\left(\frac{\zeta\left(K\right)^2\left\|\sigma^2\right\|}{n\sqrt{n}}\right).$$

To calulate the order of term (A.8) rewrite as

$$\begin{aligned}
& \frac{1}{n}\sum\left[x_i^{*2}\left[\frac{1}{\sigma_i^2}q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{x}\epsilon\right]^2\right] \\
\leq\ & \Delta\frac{1}{n}\sum\left[q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{x}\epsilon\right]^2 \\
=\ & \Delta\frac{1}{n}\sum\left[\mathbf{x}\epsilon'Q\left(Q'Q\right)^{-1}q^{K}\left(x_i\right)q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{x}\epsilon\right] \\
=\ & \Delta\frac{1}{n}\mathbf{x}\epsilon'P\mathbf{x}\epsilon \\
=\ & O\left(\frac{K}{n}\right).
\end{aligned}$$

Hence, the order of term (A.8) is

$$\begin{aligned}
& 4\left(\beta-\hat{\beta}\right)^2\frac{1}{n}\sum\left[x_i^{*2}\left[\frac{1}{\sigma_i^2}q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{x}\epsilon\right]^2\right] \\
=\ & O\left(\frac{K}{n^2}\right).
\end{aligned}$$

The order of term (A.9) is

$$\begin{aligned}
& \left(\beta-\hat{\beta}\right)^4\frac{1}{n}\sum\left[x_i^{*2}\left[\frac{1}{\sigma_i^2}q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{x2}\right]^2\right] \\
\leq\ & \Delta\left(\beta-\hat{\beta}\right)^4\frac{1}{n}\sum\left[q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{x2}\right]^2 \\
=\ & \Delta\left(\beta-\hat{\beta}\right)^4\frac{1}{n}\mathbf{x2}'P\mathbf{x2} \\
=\ & \Delta\left(\beta-\hat{\beta}\right)^4\frac{1}{n}\left(\mathbf{x2}'\mathbf{x2}-\mathbf{x2}'\left(I-P\right)\mathbf{x2}\right) \\
=\ & O\left(\frac{1}{n^2}\right)\left[O\left(1\right)+O\left(\|f\|^2\right)\right] \\
=\ & O\left(\frac{1}{n^2}\right)+O\left(\frac{\|f\|^2}{n^2}\right).
\end{aligned}$$

To calculate the order of term (A.10), decompose it as

$$\begin{aligned}
& \sum \frac{x_i^{*2}}{\sigma_i^4} \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\epsilon\mathbf{2} - 1\right] \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{x}\epsilon\right] \\
= & \sum \frac{x_i^{*2}}{\sigma_i^4} \left(\overline{\sigma}_i^2 - \sigma_i^2\right) \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{x}\epsilon\right] \\
& + \sum \frac{x_i^{*2}}{\sigma_i^4} \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{v}\right] \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{x}\epsilon\right].
\end{aligned}$$

The first term is

$$\begin{aligned}
& \sum \frac{x_i^{*2}}{\sigma_i^4} \left(\overline{\sigma}_i^2 - \sigma_i^2\right) \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{x}\epsilon\right] \\
= & O\left(\sqrt{n}\left\|\sigma^2\right\|\right) O\left(\sqrt{K}\right)
\end{aligned}$$

and the second term is[3]

$$\begin{aligned}
& \sum \frac{x_i^{*2}}{\sigma_i^4} \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{v}\right] \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{x}\epsilon\right] \\
\leq & \mathbf{v}' P \mathbf{x}\epsilon \\
= & O\left(\sqrt{K}\right).
\end{aligned}$$

The order of term (A.10) is

$$\begin{aligned}
& 4\left(\beta - \hat{\beta}\right) \frac{1}{n} \sum \left[x_i^{*2} \left[\frac{1}{\sigma_i^2} q^{K\prime}(x_i)(Q'Q)^{-1} Q'\epsilon\mathbf{2} - 1\right] \left[\frac{1}{\sigma_i^2} q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{x}\epsilon\right]\right] \\
= & O\left(\frac{1}{n\sqrt{n}}\right) \left[O\left(\sqrt{n}\left\|\sigma^2\right\|\right) O\left(\sqrt{K}\right) + O\left(\sqrt{K}\right)\right] \\
= & O\left(\frac{\left\|\sigma^2\right\| \sqrt{K}}{n}\right) + O\left(\frac{\sqrt{K}}{n\sqrt{n}}\right).
\end{aligned}$$

[3]Note that $E\left[\mathbf{v}'\mathbf{x}\epsilon\right] = 0$.

To compute the order of term (A.11), similarly,

$$
\begin{aligned}
& \sum x_i^{*2} \left[\frac{1}{\sigma_i^2} q^{K\prime}(x_i)(Q'Q)^{-1} Q'\epsilon\mathbf{2} - 1 \right] \left[\frac{1}{\sigma^2(x_i)} q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{x2} \right] \\
= & \sum \frac{x_i^{*2}}{\sigma_i^4} \left(\overline{\sigma}_i^2 - \sigma_i^2 \right) \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{x2} \right] \\
& + \sum \frac{x_i^{*2}}{\sigma_i^4} \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{v} \right] \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{x2} \right].
\end{aligned}
$$

Now we have the first term being

$$
\begin{aligned}
& \sum \frac{x_i^{*2}}{\sigma_i^4} \left(\overline{\sigma}_i^2 - \sigma_i^2 \right) \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{x2} \right] \\
= & O\left(\sqrt{n} \left\| \sigma^2 \right\| \right) \left(\mathbf{x2}' P \mathbf{x2} \right)^{1/2} \\
= & O\left(\sqrt{n} \left\| \sigma^2 \right\| \right)
\end{aligned}
$$

and second term

$$
\begin{aligned}
& \sum \frac{x_i^{*2}}{\sigma_i^4} \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{v} \right] \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{x2} \right] \\
= & O(1) + O(\|f\|).
\end{aligned}
$$

The order of term (A.11) is

$$
\begin{aligned}
& 2\left(\beta - \hat{\beta} \right)^2 \frac{1}{n} \sum \frac{x_i^{*2}}{\sigma_i^4} \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\epsilon\mathbf{2} - 1 \right] \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{x2} \right] \\
= & O\left(\frac{1}{n^2} \right) \left[O\left(\sqrt{n} \left\| \sigma^2 \right\| \right) + O(1) + O\left(\|f\|^2 \right) \right] \\
= & O\left(\frac{\left\| \sigma^2 \right\|}{n\sqrt{n}} \right) + O\left(\frac{1}{n^2} \right) + O\left(\frac{\|f\|^2}{n^2} \right).
\end{aligned}
$$

The order of term (A.12) is

$$\begin{aligned}
& 4\left(\beta-\hat{\beta}\right)^{3}\frac{1}{n}\sum\left[x_i^{*2}\left[\frac{1}{\sigma_i^2}q_i^{K\prime}Q'Q^{-1}Q'\mathbf{x}\epsilon\right]\left[\frac{1}{\sigma_i^2}q_i^{K\prime}\left(Q'Q\right)^{-1}Q'\mathbf{x2}\right]\right] \\
\leq\ & \Delta 4\left(\beta-\hat{\beta}\right)^{3}\left|\frac{1}{n}\sum\left[q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{x}\epsilon\right]\left[q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{x2}\right]\right| \\
\leq\ & \Delta'\left(\beta-\hat{\beta}\right)^{3}\left[\frac{1}{n}\sum\left[q_i^{K\prime}\left(Q'Q\right)^{-1}Q'\mathbf{x}\epsilon\right]^{2}\right]^{1/2}\left[\frac{1}{n}\sum\left[q_i^{K\prime}\left(Q'Q\right)^{-1}Q'\mathbf{x2}\right]^{2}\right]^{1/2} \\
=\ & \sqrt{O\left(\frac{1}{n^6}\right)}\sqrt{O\left(\frac{K}{n}\right)}\sqrt{O\left(1\right)+O\left(\|f\|^2\right)} \\
=\ & O\left(\frac{1}{n}\right).
\end{aligned}$$

The order of $X^{*\prime}\left(\widehat{\Sigma}^{*}-I\right)\left(\widehat{\Sigma}^{*}-I\right)X^{*}/n$ is

$$\begin{aligned}
& \frac{1}{n}X^{*\prime}\left(\widehat{\Sigma}^{*}-I\right)\left(\widehat{\Sigma}^{*}-I\right)X^{*} \\
=\ & O\left(\left\|\sigma^2\right\|^2\right)+O\left(\frac{K}{n}\right)+O\left(\frac{\zeta\left(K\right)^2\left\|\sigma^2\right\|}{n\sqrt{n}}\right)+O\left(\frac{1}{n}\right) \\
=\ & \frac{1}{n}\sum\left[\frac{x_i^{*2}}{\sigma_i^4}\left[\overline{\sigma}_i^2-\sigma_i^2\right]^2\right] \\
& +\frac{1}{n}\sum\left[\frac{x_i^{*2}}{\sigma_i^4}\left[q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{v}\right]^2\right]+O\left(\frac{1}{n}\right).
\end{aligned}$$

□

Lemma A.2.3.

$$\begin{aligned}
& \left[X^{*\prime}\left(\widehat{\Sigma}^{*}-I\right)\epsilon^{*}/\sqrt{n}\right] \\
=\ & \frac{1}{\sqrt{n}}\sum\frac{x_i^{*}\epsilon_i^{*}}{\sigma_i^2}\left[\overline{\sigma}_i^2-\sigma_i^2\right]+\frac{1}{\sqrt{n}}\sum\frac{x_i^{*}\epsilon_i^{*}}{\sigma_i^2}\left[q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{v}\right] \\
& +2\left(\beta-\hat{\beta}\right)\frac{1}{\sqrt{n}}\sum\frac{x_i^{*}\epsilon_i^{*}}{\sigma_i^2}\left[q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{x}\epsilon\right]+O\left(\frac{1}{n}\right).
\end{aligned}$$

Proof. We could rewrite $X^{*\prime}\left(\widehat{\Sigma}^{*}-I\right)\epsilon^{*}/\sqrt{n}$ as

$$\frac{X^{*\prime}\left(\widehat{\Sigma}^{*}-I\right)\epsilon^{*}}{\sqrt{n}}=\sum\left(\frac{1}{\sqrt{n}}x_i^{*}\epsilon_i^{*}\left(\frac{1}{\sigma_i^2}\widehat{\sigma}_i^2-1\right)\right). \tag{A.13}$$

Plugging equation $(A.3)$ into equation $(A.13)$ gives us

$$\begin{aligned}
& \frac{1}{\sqrt{n}}\sum\left[x_i^{*}\epsilon_i^{*\prime}\left(\frac{1}{\sigma_i^2}\widehat{\sigma}_i^2-1\right)\right] \\
= & \frac{1}{\sqrt{n}}\sum\left[x_i^{*}\epsilon_i^{*\prime}\left[\frac{1}{\sigma_i^2}q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\left[\epsilon\mathbf{2}+2\left(\beta-\hat{\beta}\right)\mathbf{x}\epsilon+\left(\beta-\hat{\beta}\right)^2\mathbf{x2}\right]-1\right]\right] \\
= & \frac{1}{\sqrt{n}}\left[\mathbf{xs}\epsilon P\epsilon\mathbf{2}-\mathbf{x}^{*\prime}\epsilon^{*}\right] && \text{(A.14)} \\
& +2\left(\beta-\hat{\beta}\right)\cdot\frac{1}{\sqrt{n}}\left[\mathbf{xs}\epsilon' P\mathbf{x}\epsilon\right] && \text{(A.15)} \\
& +\left(\beta-\hat{\beta}\right)^2\cdot\frac{1}{\sqrt{n}}\left[\mathbf{xs}\epsilon' P\mathbf{x2}\right], && \text{(A.16)}
\end{aligned}$$

where $\mathbf{xs}\epsilon=\left(x_1^{*}\epsilon_1^{*}/\sigma^2\left(x_1\right),...,x_n^{*}\epsilon_n^{*}/\sigma^2\left(x_n\right)\right)'$. White the term in (B.1) as

$$\begin{aligned}
& \frac{1}{\sqrt{n}}\sum\left[x_i^{*}\epsilon_i^{*}\left[\frac{1}{\sigma_i^2}q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\epsilon\mathbf{2}-1\right]\right] \\
= & \frac{1}{\sqrt{n}}\sum\frac{x_i^{*}\epsilon_i^{*}}{\sigma_i^2}\left[\overline{\sigma}_i^2-\sigma_i^2\right] \\
& +\frac{1}{\sqrt{n}}\sum\frac{x_i^{*}\epsilon_i^{*}}{\sigma_i^2}\left[q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{v}\right]
\end{aligned}$$

$$\varpi=\frac{1}{\sqrt{n}}\sum\frac{x_i^{*}\epsilon_i^{*}}{\sigma_i^2}\left[\overline{\sigma}_i^2-\sigma_i^2\right].$$

We have

$$E\left[\varpi\right]=0.$$

The variance of ϖ is

$$\begin{aligned} Var[\varpi] &= E\left\{\frac{1}{n}\sum_i \left[\frac{x_i^{*2}\epsilon_i^{*2}}{\sigma_i^4}\left[\overline{\sigma}_i^2 - \sigma_i^2\right]^2\right]\right\} \\ &= \frac{1}{n}\sum_i \left[\frac{x_i^{*2}}{\sigma_i^4}\left[\overline{\sigma}_i^2 - \sigma_i^2\right]^2\right] \\ &= O\left(\left\|\sigma^2\right\|^2\right) \end{aligned}$$

Hence,

$$\frac{1}{\sqrt{n}}\sum \frac{x_i^*\epsilon_i^*}{\sigma_i^2}\left[\overline{\sigma}_i^2 - \sigma_i^2\right] = O\left(\left\|\sigma^2\right\|\right).$$

Also,

$$\begin{aligned} &\frac{1}{\sqrt{n}}\sum \frac{x_i^*\epsilon_i^*}{\sigma_i^2}\left[q^{K\prime}(x_i)(Q'Q)^{-1}Q'\mathbf{v}\right] \\ =\ &\frac{1}{\sqrt{n}}\mathbf{x}\epsilon' P\mathbf{v} \\ =\ &\frac{\sqrt{K}}{\sqrt{n}}. \end{aligned}$$

The order of term in $(B.1)$ will be

$$\frac{1}{\sqrt{n}}\left[\mathbf{xs}\epsilon' P\epsilon\mathbf{2} - \mathbf{x}^{*\prime}\epsilon^*\right] = O\left(\left\|\sigma^2\right\|\right) + O\left(\frac{\sqrt{K}}{\sqrt{n}}\right).$$

The order of the term in $(B.2)$ will be

$$\begin{aligned} &2\left(\beta - \hat{\beta}\right)\frac{1}{\sqrt{n}}\left[\mathbf{xs}\epsilon' P\mathbf{x}\epsilon\right] \\ =\ &O\left(\frac{1}{\sqrt{n}}\right)O\left(\frac{K}{\sqrt{n}}\right) \\ =\ &O\left(\frac{K}{n}\right). \end{aligned}$$

The order of the term in $(A.16)$ will be

$$
\begin{aligned}
& \left(\beta-\hat{\beta}\right)^2 \frac{1}{\sqrt{n}}\left[\mathbf{xs}\epsilon' P\mathbf{x2}\right] \\
= & \left(\beta-\hat{\beta}\right)^2 \frac{1}{\sqrt{n}}\left[\mathbf{x2}'\mathbf{xs}\epsilon-\mathbf{x2}'\left(I-P\right)\mathbf{xs}\epsilon\right] \\
= & O\left(\frac{1}{n}\right)\left[O\left(1\right)+O\left(\|f\|\right)\right] \\
= & O\left(\frac{1}{n}\right)+O\left(\frac{\|f\|}{n}\right).
\end{aligned}
$$

Therefore, the order of equation (A.13) will be

$$
\begin{aligned}
& \frac{X^{*\prime}\left(\widehat{\Sigma}^*-I\right)\epsilon^*}{\sqrt{n}} \\
= & O\left(\left\|\sigma^2\right\|\right)+O\left(\frac{\sqrt{K}}{\sqrt{n}}\right)+O\left(\frac{K}{n}\right)+O\left(\frac{1}{n}\right) \\
= & \frac{1}{\sqrt{n}}\sum\frac{x_i^*\epsilon_i^*}{\sigma_i^2}\left[\overline{\sigma}_i^2-\sigma_i^2\right]+\frac{1}{\sqrt{n}}\sum\frac{x_i^*\epsilon_i^*}{\sigma_i^2}\left[q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{v}\right] \\
& +2\left(\beta-\hat{\beta}\right)\frac{1}{\sqrt{n}}\sum\frac{x_i^*\epsilon_i^*}{\sigma_i^2}\left[q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{x}\epsilon\right]+O\left(\frac{1}{n}\right).
\end{aligned}
$$

□

Lemma A.2.4.

$$
\begin{aligned}
& \left[X^{*\prime}\left(\widehat{\Sigma}^*-I\right)\left(\widehat{\Sigma}^*-I\right)\epsilon^*/\sqrt{n}\right] \\
= & \frac{1}{\sqrt{n}}\sum\left[\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\left[\overline{\sigma}_i^2-\sigma_i^2\right]^2\right] \\
& +\frac{1}{\sqrt{n}}\sum\left[\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\left[q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{v}\right]^2\right]+O\left(\frac{1}{n}\right).
\end{aligned}
$$

Proof. It could be rewritten as

$$\frac{1}{\sqrt{n}}X^{*\prime}\left(\widehat{\Sigma}^{*}-I\right)\left(\widehat{\Sigma}^{*}-I\right)\epsilon^{*}$$

$$= \frac{1}{\sqrt{n}}\sum\left[x_i^*\epsilon_i^*\left(\frac{1}{\sigma_i^2}\widehat{\sigma}_i^2-1\right)^2\right]$$

$$= \frac{1}{\sqrt{n}}\sum\left[\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\left[q^{K\prime}(x_i)(Q'Q)^{-1}Q'\left[\epsilon\mathbf{2}+2\left(\beta-\hat{\beta}\right)\mathbf{x}\epsilon+\left(\beta-\hat{\beta}\right)^2\mathbf{x2}\right]-\sigma_i^2\right]^2\right]$$

$$= \frac{1}{\sqrt{n}}\sum\left[x_i^*\epsilon_i^*\left[\frac{1}{\sigma_i^2}q^{K\prime}(x_i)(Q'Q)^{-1}Q'\epsilon\mathbf{2}-1\right]^2\right] \tag{A.17}$$

$$+4\left(\beta-\hat{\beta}\right)^2\frac{1}{\sqrt{n}}\sum\left[x_i^*\epsilon_i^*\left[\frac{1}{\sigma_i^2}q^{K\prime}(x_i)(Q'Q)^{-1}Q'\mathbf{x}\epsilon\right]^2\right] \tag{A.18}$$

$$+\left(\beta-\hat{\beta}\right)^4\frac{1}{\sqrt{n}}\sum\left[x_i^*\epsilon_i^*\left[\frac{1}{\sigma_i^2}q^{K\prime}(x_i)(Q'Q)^{-1}Q'\mathbf{x2}\right]^2\right] \tag{A.19}$$

$$+4\left(\beta-\hat{\beta}\right)\frac{1}{\sqrt{n}}\sum_i\left[\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\left[\sum_j P_{ij}\epsilon_j^2-\sigma_i^2\right]\left[q^{K\prime}(x_i)(Q'Q)^{-1}Q'\mathbf{x}\epsilon\right]\right] \tag{A.20}$$

$$+2\left(\beta-\hat{\beta}\right)^2\frac{1}{\sqrt{n}}\sum\left[\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\left[\sum_j P_{ij}\epsilon_j^2-\sigma_i^2\right]\left[q^{K\prime}(x_i)(Q'Q)^{-1}Q'\mathbf{x2}\right]\right] \tag{A.21}$$

$$+4\left(\beta-\hat{\beta}\right)^3\frac{1}{\sqrt{n}}\sum\left[\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\left[q^{K\prime}(x_i)(Q'Q)^{-1}Q'\mathbf{x}\epsilon\right]\left[q^{K\prime}(x_i)(Q'Q)^{-1}Q'\mathbf{x2}\right]\right] \tag{A.22}$$

The order of term (A.17) is

$$\begin{aligned}
& \frac{1}{\sqrt{n}}\sum\left[x_i^*\epsilon_i^*\left[\frac{1}{\sigma_i^2}q^{K\prime}(x_i)(Q'Q)^{-1}Q'\epsilon\mathbf{2}-1\right]^2\right] \\
= & \frac{1}{\sqrt{n}}\sum\left[\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\left[q^{K\prime}(x_i)(Q'Q)^{-1}Q'(\sigma\mathbf{2}+\mathbf{v})-\sigma_i^2\right]^2\right] \\
= & \frac{1}{\sqrt{n}}\sum\left[\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\left[\left(q^{K\prime}(x_i)(Q'Q)^{-1}Q'\sigma\mathbf{2}-\sigma_i^2\right)+q^{K\prime}(x_i)(Q'Q)^{-1}Q'\mathbf{v}\right]^2\right] \\
= & \frac{1}{\sqrt{n}}\sum\left[\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\left[\overline{\sigma_i}^2-\sigma_i^2\right]^2\right] \\
& +\frac{1}{\sqrt{n}}\sum\left[\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\left[q^{K\prime}(x_i)(Q'Q)^{-1}Q'\mathbf{v}\right]^2\right] \\
& +\frac{2}{\sqrt{n}}\sum\left[\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\left[\left(\overline{\sigma_i}^2-\sigma_i^2\right)\left(q^{K\prime}(x_i)(Q'Q)^{-1}Q'\mathbf{v}\right)\right]\right] \\
= & O\left(\zeta(K)\left\|\sigma^2\right\|^2\right)+O\left(\frac{\zeta(K)K\sqrt{K}}{n}\right)+O\left(\frac{\zeta(K)^2}{n}\left\|\sigma^2\right\|\right).
\end{aligned}$$

Let

$$\psi=\frac{1}{\sqrt{n}}\sum\left[\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\left[\overline{\sigma_i}^2-\sigma_i^2\right]^2\right].$$

We have $E[\psi]=0$ and

$$\begin{aligned}
V[\psi] &= \frac{1}{n}E\sum\left[\frac{x_i^{*2}\epsilon_i^{*2}}{\sigma_i^8}\left[\overline{\sigma_i}^2-\sigma_i^2\right]^4\right] \\
&= \frac{1}{n}\sum\left[\frac{x_i^{*2}}{\sigma_i^8}\left[\overline{\sigma_i}^2-\sigma_i^2\right]^4\right] \\
&\leq \Delta\frac{1}{n}\sup\left[\overline{\sigma_i}^2-\sigma_i^2\right]^2\sum\left[\overline{\sigma_i}^2-\sigma_i^2\right]^2 \\
&= O\left(\zeta(K)^2\left\|\sigma^2\right\|^2\right)O\left(\left\|\sigma^2\right\|^2\right).
\end{aligned}$$

Hence,

$$\frac{1}{\sqrt{n}}\sum\left[\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\left[\overline{\sigma_i}^2-\sigma_i^2\right]^2\right]=O\left(\zeta(K)\left\|\sigma^2\right\|^2\right).$$

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \sum \left[\frac{x_i^* \epsilon_i^*}{\sigma_i^4} \left[q^{K\prime}(x_i) (Q'Q)^{-1} Q'\mathbf{v} \right]^2 \right] \\
= & \frac{1}{\sqrt{n}} \left\| \mathbf{v}'Q (Q'Q)^{-1} (Q'SQ') (Q'Q)^{-1} Q'\mathbf{v} \right\| \\
\leq & \frac{1}{\sqrt{n}} \left\| \mathbf{v}'Q (Q'Q)^{-1} \right\|^2 \|Q'SQ'\| \\
= & \frac{1}{\sqrt{n}} O\left(\frac{K}{n}\right) O\left(\sqrt{n}\zeta(K)\sqrt{K}\right) \\
= & O\left(\frac{\zeta(K) K \sqrt{K}}{n}\right),
\end{aligned}
$$

where $S = diag(x_i^* \epsilon_i^* / \sigma_i^4)$.[4]

$$
\begin{aligned}
& \frac{2}{\sqrt{n}} \sum \left[\frac{x_i^* \epsilon_i^*}{\sigma_i^4} \left[\left(\overline{\sigma_i}^2 - \sigma_i^2\right) \left(q^{K\prime}(x_i) (Q'Q)^{-1} Q'\mathbf{v} \right) \right] \right] \\
= & \frac{2}{\sqrt{n}} \sum_{ij} \left[\frac{x_i^* \epsilon_i^*}{\sigma_i^4} \left(\overline{\sigma_i}^2 - \sigma_i^2\right) (P_{ij} v_j) \right] \\
= & \tau.
\end{aligned}
$$

---

[4]

$$
\begin{aligned}
\left\| (Q'Q)^{-1} Q'\mathbf{v} \right\|^2 &= \mathbf{v}'Q (Q'Q)^{-2} Q'\mathbf{v} \\
&\leq \lambda_{\max} \left[\left(\frac{Q'Q}{n} \right)^{-1} \right] \frac{1}{n} \mathbf{v}'Q (Q'Q)^{-1} Q'\mathbf{v} \\
&= \frac{1}{n} \mathbf{v}'Q (Q'Q)^{-1} Q'\mathbf{v} \\
&= \frac{1}{n} \mathbf{v}'P\mathbf{v} \\
&= O\left(\frac{K}{n}\right).
\end{aligned}
$$

We have[5] $E[\tau] = 0$ and variance of τ being

$$\begin{aligned}
V[\tau] &= \frac{4}{n} E \sum_{ij} \left[\frac{x_i^{*2} \epsilon_i^{*2}}{\sigma_i^8} \left(\overline{\sigma_i}^2 - \sigma_i^2 \right)^2 P_{ij}^2 v_j^2 \right] \\
&= \frac{4}{n} \sum_i \left[\frac{x_i^{*2}}{\sigma_i^8} \left(\overline{\sigma_i}^2 - \sigma_i^2 \right)^2 P_{ii}^2 E\left[\epsilon_i^{*2} v_i^2 \right] \right] \\
&\quad + \frac{4}{n} \sum_{i \neq j} \left[\frac{x_i^{*2} E\left[\epsilon_i^{*2}\right]}{\sigma_i^8} \left(\overline{\sigma_i}^2 - \sigma_i^2 \right)^2 P_{ij}^2 E\left[v_j^2 \right] \right] \\
&\leq \Delta \frac{4}{n} \sum_i \left(\overline{\sigma_i}^2 - \sigma_i^2 \right)^2 P_{ii}^2 + \Delta' \frac{4}{n} \sum_{i \neq j} \left(\overline{\sigma_i}^2 - \sigma_i^2 \right)^2 P_{ij}^2 \\
&\leq \Delta \frac{4}{n} \left| \sup P_{ii}^2 \right| \sum_i \left(\overline{\sigma_i}^2 - \sigma_i^2 \right)^2 \\
&= O\left(\frac{\zeta(K)^2}{n} \right)^2 O\left(\left\| \sigma^2 \right\|^2 \right).
\end{aligned}$$

$$\frac{2}{\sqrt{n}} \sum_{ij} \left[\frac{x_i^* \epsilon_i^*}{\sigma_i^4} \left(\overline{\sigma_i}^2 - \sigma_i^2 \right) \left(P_{ij} v_j \right) \right] = O\left(\frac{\zeta(K)^2}{n} \left\| \sigma^2 \right\| \right).$$

$$\begin{aligned}
&\frac{2}{\sqrt{n}} \sum \left[\frac{x_i^* \epsilon_i^*}{\sigma_i^4} \left(\overline{\sigma_i}^2 - \sigma_i^2 \right) \left(q^{K\prime}(x_i) (Q'Q)^{-1} Q'\mathbf{v} \right) \right] \\
\leq\ & \frac{2}{\sqrt{n}} \left[\sum_i \left(\frac{x_i^* \epsilon_i^*}{\sigma_i^4} \right)^2 \left(\overline{\sigma_i}^2 - \sigma_i^2 \right)^2 \right]^{1/2} \left[\sum_i \left(q^{K\prime}(x_i) (Q'Q)^{-1} Q'\mathbf{v} \right)^2 \right]^{1/2} \\
\leq\ & \frac{2}{\sqrt{n}} \left[\sum_i \left(\overline{\sigma_i}^2 - \sigma_i^2 \right)^2 \right]^{1/2} \left[\mathbf{v}' P \mathbf{v} \right]^{1/2} \\
\leq\ & O\left(\frac{1}{\sqrt{n}} \right) O\left(\sqrt{n} \left\| \sigma^2 \right\| \right) O\left(\sqrt{K} \right) \\
=\ & O\left(\left\| \sigma^2 \right\| \sqrt{K} \right).
\end{aligned}$$

[5] $E[\epsilon_i v_i] = E\left[\epsilon_i \left(\epsilon_i^2 - \sigma_i^2\right)\right] = E\left(\epsilon_i^3\right) - \sigma_i^2 E(\epsilon_i) = 0.$

To calulate the order of term (A.18) let[6]

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum \frac{x_i^* \epsilon_i^*}{\sigma_i^4} \left[q^{K\prime}(x_i) (Q'Q)^{-1} Q'\mathbf{x}\epsilon \right]^2 \\ = & \frac{1}{\sqrt{n}} \mathbf{x}\epsilon' Q (Q'Q)^{-1} Q'SQ (Q'Q)^{-1} Q'\mathbf{x}\epsilon \\ \leq & \frac{1}{\sqrt{n}} \left\| \mathbf{x}\epsilon' Q (Q'Q)^{-1} \right\|^2 \| Q'SQ' \| \\ = & O\left(\frac{\zeta(K) K\sqrt{K}}{n} \right) \end{aligned}$$

Hence, the order of term (A.18) is

$$\begin{aligned} & 4\left(\beta - \hat{\beta}\right)^2 \frac{1}{\sqrt{n}} \sum \left[x_i^* \epsilon_i^* \left[\frac{1}{\sigma_i^2} q^{K\prime}(x_i) (Q'Q)^{-1} Q'\mathbf{x}\epsilon \right]^2 \right] \\ = & O\left(\frac{\zeta(K) K\sqrt{K}}{n^2} \right). \end{aligned}$$

To calulate the order of term (A.19), first compute

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum x_i^* \epsilon_i^* \left[\frac{1}{\sigma^2(x_i)} q^{K\prime}(x_i) (Q'Q)^{-1} Q'\mathbf{x2} \right]^2 \\ = & \frac{1}{\sqrt{n}} \mathbf{x2}' Q (Q'Q)^{-1} Q'SQ (Q'Q)^{-1} Q'\mathbf{x2} \\ \leq & \frac{1}{\sqrt{n}} \left\| \mathbf{x}\epsilon' Q (Q'Q)^{-1} \right\|^2 \| Q'SQ' \| \\ = & O\left(\frac{1}{\sqrt{n}} \right) O(1)\, O\left(\sqrt{n}\zeta(K) \sqrt{K} \right) \\ = & O\left(\zeta(K) \sqrt{K} \right). \end{aligned}$$

[6]Note that we apply the following inequality.

$$\sum \left[q^{K\prime}(x_i) (Q'Q)^{-1} Q'\mathbf{x}\epsilon \right]^4 \leq \left[\mathbf{x}\epsilon' P \mathbf{x}\epsilon \right]^2$$

The order of term (A.19) is

$$\begin{aligned}
&\left(\beta-\hat{\beta}\right)^{4}\frac{1}{\sqrt{n}}\sum\left[x_i^*\epsilon_i^*\left[\frac{1}{\sigma_i^2}q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{x2}\right]^2\right] \\
= \; & O\left(\frac{1}{n^2}\right)O\left(\zeta\left(K\right)\sqrt{K}\right) \\
= \; & O\left(\frac{\zeta\left(K\right)\sqrt{K}}{n^2}\right).
\end{aligned}$$

To calulate the order of term (A.20) rewrite as

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum\left[\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\left[\overline{\sigma}_i^2-\sigma_i^2\right]\left[q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{x}\epsilon\right]\right] \\
&+\frac{1}{\sqrt{n}}\sum\left[\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\left[q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{v}\right]\left[q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{x}\epsilon\right]\right].
\end{aligned}$$

The first term is

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum\left[\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\left[\overline{\sigma}_i^2-\sigma_i^2\right]\left[q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{x}\epsilon\right]\right] \\
= \; & \frac{1}{\sqrt{n}}\left[\sum\left(\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\right)^2\left[\overline{\sigma}_i^2-\sigma_i^2\right]^2\right]^{1/2}\left[\sum\left[q^{K\prime}\left(x_i\right)\left(Q'Q\right)^{-1}Q'\mathbf{x}\epsilon\right]^2\right]^{1/2} \\
\leq \; & \frac{1}{\sqrt{n}}\left(\sum\left[\overline{\sigma}_i^2-\sigma_i^2\right]^2\right)^{1/2}\left(\mathbf{x}\epsilon' P\mathbf{x}\epsilon\right)^{1/2} \\
= \; & O\left(\frac{1}{\sqrt{n}}\right)O\left(\sqrt{n}\left\|\sigma^2\right\|\right)O\left(\sqrt{K}\right) \\
= \; & O\left(\left\|\sigma^2\right\|\sqrt{K}\right).
\end{aligned}$$

The second term is

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \sum \left[\frac{x_i^* \epsilon_i^*}{\sigma_i^4} \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{v} \right] \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{x}\epsilon \right] \right] \\
= \; & \frac{1}{\sqrt{n}} \mathbf{v}' Q (Q'Q)^{-1} Q' S Q (Q'Q)^{-1} Q'\mathbf{x}\epsilon \\
\leq \; & \frac{1}{\sqrt{n}} \left\| \mathbf{v}' Q (Q'Q)^{-1} \right\| \left\| \mathbf{x}\epsilon' Q (Q'Q)^{-1} \right\| \| Q' S Q' \| \\
= \; & O\left(\frac{\zeta(K) K \sqrt{K}}{n} \right)
\end{aligned}
$$

The order of term (A.20) is

$$
\begin{aligned}
& 4 \left(\beta - \hat{\beta} \right) \frac{1}{\sqrt{n}} \sum \left[\frac{x_i^* \epsilon_i^*}{\sigma_i^2} \left[\overline{\epsilon}_i^2 - \sigma_i^2 \right] \left[\frac{1}{\sigma_i^2} q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{x}\epsilon \right] \right] \\
= \; & O\left(\frac{1}{\sqrt{n}} \right) \left[O\left(\|\sigma^2\| \sqrt{K} \right) + O\left(\frac{\zeta(K) K \sqrt{K}}{n} \right) \right] \\
= \; & O\left(\frac{\|\sigma^2\| \sqrt{K}}{\sqrt{n}} \right) + O\left(\frac{\zeta(K) K \sqrt{K}}{n\sqrt{n}} \right).
\end{aligned}
$$

To calulate the order of term (A.21) rewrite as

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \sum \left[\frac{x_i^* \epsilon_i^*}{\sigma_i^4} \left[\overline{\sigma}_i^2 - \sigma_i^2 \right] \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{x2} \right] \right] \\
& + \frac{1}{\sqrt{n}} \sum \left[\frac{x_i^* \epsilon_i^*}{\sigma_i^4} \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{v} \right] \left[q^{K\prime}(x_i)(Q'Q)^{-1} Q'\mathbf{x2} \right] \right].
\end{aligned}
$$

The first term is

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum \left[\frac{x_i^* \epsilon_i^*}{\sigma_i^4} \left[\overline{\sigma}_i^2 - \sigma_i^2 \right] \left[q^{K\prime} \left(x_i \right) \left(Q'Q \right)^{-1} Q' \mathbf{x2} \right] \right] \\
= & \frac{1}{\sqrt{n}} \left[\sum \left(\frac{x_i^* \epsilon_i^*}{\sigma_i^4} \right)^2 \left[\overline{\sigma}_i^2 - \sigma_i^2 \right]^2 \right]^{1/2} \left[\sum \left[q^{K\prime} \left(x_i \right) \left(Q'Q \right)^{-1} Q' \mathbf{x2} \right]^2 \right]^{1/2} \\
\leq & \frac{1}{\sqrt{n}} \left(\sum \left[\overline{\sigma}_i^2 - \sigma_i^2 \right]^2 \right)^{1/2} \left(\mathbf{x2}' P \mathbf{x2} \right)^{1/2} \\
= & O\left(\frac{1}{\sqrt{n}} \right) O\left(\sqrt{n} \left\| \sigma^2 \right\| \right) O\left(1 \right) \\
= & O\left(\left\| \sigma^2 \right\| \right)
\end{aligned}$$

The second term is

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum \left[\frac{x_i^* \epsilon_i^*}{\sigma_i^4} \left[q^{K\prime} \left(x_i \right) \left(Q'Q \right)^{-1} Q' \mathbf{v} \right] \left[q^{K\prime} \left(x_i \right) \left(Q'Q \right)^{-1} Q' \mathbf{x2} \right] \right] \\
= & \frac{1}{\sqrt{n}} \mathbf{v}' Q \left(Q'Q \right)^{-1} Q' S Q \left(Q'Q \right)^{-1} Q' \mathbf{x2} \\
\leq & \frac{1}{\sqrt{n}} \left\| \mathbf{x2}' Q \left(Q'Q \right)^{-1} Q' S Q \left(Q'Q \right)^{-1/2} \right\| \left\| \left(Q'Q \right)^{-1/2} Q' \mathbf{v} \right\| \\
= & O\left(\frac{1}{\sqrt{n}} \right) O\left(\sqrt{K} \right) O\left(\sqrt{K} \right) \\
= & O\left(\frac{K}{\sqrt{n}} \right)
\end{aligned}$$

The order of term (A.21) is

$$\begin{aligned}
& 2 \left(\beta - \hat{\beta} \right)^2 \frac{1}{\sqrt{n}} \sum \left[\frac{x_i^* \epsilon_i^*}{\sigma_i^2} \left[\bar{\epsilon}_i^2 - \sigma_i^2 \right] \left[\frac{1}{\sigma_i^2} q^{K\prime} \left(x_i \right) \left(Q'Q \right)^{-1} Q' \mathbf{x2} \right] \right] \\
= & O\left(\frac{1}{n} \right) \left[O\left(\left\| \sigma^2 \right\| \right) + O\left(\frac{K}{\sqrt{n}} \right) \right] \\
= & O\left(\frac{\left\| \sigma^2 \right\|}{n} \right) + O\left(\frac{K}{n\sqrt{n}} \right).
\end{aligned}$$

To calculate the order of term (A.22) rewrite as

$$\begin{aligned}
& \frac{1}{\sqrt{n}}\sum\left[\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\left[q^{K\prime}(x_i)\left(Q'Q\right)^{-1}Q'\mathbf{x}\epsilon\right]\left[q^{K\prime}(x_i)\left(Q'Q\right)^{-1}Q'\mathbf{x2}\right]\right] \\
= & \frac{1}{\sqrt{n}}\mathbf{x}\epsilon' Q\left(Q'Q\right)^{-1}Q'SQ\left(Q'Q\right)^{-1}Q'\mathbf{x2} \\
\leq & \frac{1}{\sqrt{n}}\left\|\mathbf{x2}'Q\left(Q'Q\right)^{-1}Q'SQ\left(Q'Q\right)^{-1/2}\right\|\left\|\left(Q'Q\right)^{-1/2}Q'\mathbf{x}\epsilon\right\| \\
= & O\left(\frac{1}{\sqrt{n}}\right)O\left(\sqrt{K}\right)O\left(\sqrt{K}\right) \\
= & O\left(\frac{K}{\sqrt{n}}\right).
\end{aligned}$$

The order of term (A.22) is

$$\begin{aligned}
& 4\left(\beta-\hat{\beta}\right)^3\frac{1}{n}\sum\left[x_i^*\epsilon_i^*\left[q^{K\prime}(x_i)\left(Q'Q\right)^{-1}Q'\mathbf{x}\epsilon\right]\left[q^{K\prime}(x_i)\left(Q'Q\right)^{-1}Q'\mathbf{x2}\right]\right] \\
= & O\left(\frac{1}{n\sqrt{n}}\right)O\left(\frac{K}{\sqrt{n}}\right) \\
= & O\left(\frac{K}{n^2}\right).
\end{aligned}$$

To sum up, the order of $X^{*\prime}\left(\widehat{\Sigma}^*-I\right)\left(\widehat{\Sigma}^*-I\right)\epsilon^*/\sqrt{n}$ is

$$\begin{aligned}
& \frac{1}{\sqrt{n}}X^{*\prime}\left(\widehat{\Sigma}^*-I\right)\left(\widehat{\Sigma}^*-I\right)\epsilon^* \\
= & O\left(\zeta(K)\left\|\sigma^2\right\|^2\right)+O\left(\frac{\zeta(K)K\sqrt{K}}{n}\right)+O\left(\frac{\zeta(K)^2}{n}\left\|\sigma^2\right\|\right)+O\left(\frac{1}{n}\right) \\
= & \frac{1}{\sqrt{n}}\sum\left[\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\left[\overline{\sigma_i}^2-\sigma_i^2\right]^2\right] \\
& +\frac{1}{\sqrt{n}}\sum\left[\frac{x_i^*\epsilon_i^*}{\sigma_i^4}\left[q^{K\prime}(x_i)\left(Q'Q\right)^{-1}Q'\mathbf{v}\right]^2\right]+O\left(\frac{1}{n}\right).
\end{aligned}$$

□

Apply Lemma A.2.1-A.2.4 to $\widehat{H}$ and $\widehat{h}$ defined before then drop out terms of small order. We summarize the decomposition of $\widehat{H}$ and $\widehat{h}$ in Table D.13.

To verify lemma 1 in Donald and Newey (2001) we first guess

$$\rho_{K,n} = \frac{K}{n} + O\left(\|f\| \left\|\sigma^2\right\|\right) + O\left(\|f\|\right) + O\left(\left\|\sigma^2\right\|\right).$$

It is obvious that $\left\|T^H\right\|^2$ is $o\left(\rho_{K,n}\right)$. To check small order for $\left\|T^H\right\| \left\|T^h\right\|$, we have to check the orders of $\left\|T_i^H\right\| \left\|T_j^h\right\|$ $(i = 1, .., 4; j = 1, .., 5)$. The requirement of uniform convergence which is $\|\sigma^2\| \zeta(K) \longrightarrow 0$ implies $\left\|T_1^H\right\| \left\|T_4^h\right\| = o\left(\rho_{K,n}\right)$. To get $\left\|T_1^H\right\| \left\|T_5^h\right\| = o\left(\rho_{K,n}\right)$ the extra condition that $\zeta(K) K/\sqrt{n} \longrightarrow 0$ will be needed. Note that $\zeta(K) K/\sqrt{n} \longrightarrow 0$ implies $\zeta(K) \sqrt{K}/\sqrt{n} \longrightarrow 0$. $\left\|T_2^H\right\| \left\|T_5^h\right\| = o\left(\rho_{K,n}\right)$ is ensured. Other terms are easy to check. Combining the above results gives $\left\|T^H\right\| \left\|T^h\right\| = o\left(\rho_{K,n}\right)$.

Next, compute $E\left[\widehat{A}(K)\right]$ and note that

$$\begin{aligned}
& \left(h + T^h\right)\left(h + T^h\right)' - hh'H^{-1}T^{H\prime} - T^H H^{-1}hh' \\
= \;& hh' + T^h h' + hT^{h\prime} + T^h T^{h\prime} - hh'H^{-1}T^{H\prime} - T^H H^{-1}hh' \qquad \text{(A.23)} \\
= \;& A(K) + Z(K).
\end{aligned}$$

To save the space, we list the the results of the expectation of terms in (A.23) sequentially.

$$E\left(hh'\right) = \frac{1}{n} X^{*\prime} X^*$$

$$E\left(T_1^h h'\right) = -\frac{1}{n} \sum_i \frac{x_i^{*2}}{\sigma_i^2} \left[\overline{\sigma}_i^2 - \sigma_i^2\right] = O\left(\|f\| \left\|\sigma^2\right\|\right)$$

$$E\left(T_2^h h'\right) = -\frac{1}{n}\sum_i \frac{x_i^{*2}}{\sigma_i^4} P_{ii} E\left[v_i\right] = O\left(\frac{K}{n}\right)$$

$$E\left(T_3^h h'\right) = \frac{2}{n}\sum_i x_i^{*2} P_{ii} = O\left(\frac{K}{n}\right)$$

$$E\left(T_4^h h'\right) = \frac{1}{n}\sum_i \frac{x_i^{*2}}{\sigma_i^4}\left[\overline{\sigma}_i^2 - \sigma_i^2\right]^2 = O\left(\left\|\sigma^2\right\|^2\right)$$

$$\begin{aligned} E\left(T_5^h h'\right) &= \frac{1}{n}\sum_i \frac{x_i^{*2}}{\sigma_i^4} P_{ii}^2 E\left(v_i^2 \epsilon_i^{*2}\right) + \frac{1}{n}\sum_{i\neq j} \frac{x_i^{*2}}{\sigma_i^4} P_{ij}^2 E\left(v_j^2\right) \\ &\leq \Delta \frac{1}{n}\sum_i P_{ii}^2 + \Delta' \frac{1}{n}\sum_{i\neq j} P_{ij}^2 \\ &= o\left(\frac{K}{n}\right) + O\left(\frac{K}{n}\right) \end{aligned}$$

As for the term $E\left(T^h T^{h\prime}\right)$, by inspection the orders of $E\left(T_3^h T_3^{h\prime}\right)$, $E\left(T_4^h T_4^{h\prime}\right)$ and $E\left(T_5^h T_5^{h\prime}\right)$ are negligible. The only two terms which will matter is

$$E\left(T_1^h T_1^{h\prime}\right) = \frac{1}{n}\sum_i \frac{x_i^{*2}}{\sigma_i^4}\left[\overline{\sigma}_i^2 - \sigma_i^2\right]^2 = O\left(\left\|\sigma^2\right\|^2\right)$$

$$\begin{aligned} E\left(T_2^h T_2^{h\prime}\right) &= \frac{1}{n}\sum_i \frac{x_i^{*2}}{\sigma_i^4} P_{ii}^2 E\left(\epsilon_i^{*2} v_i^2\right) + \frac{1}{n}\sum_{i\neq j} \frac{x_i^{*2}}{\sigma_i^4} P_{ij}^2 E\left(v_j^2\right) \\ &\leq \Delta \frac{1}{n}\sum_i P_{ii}^2 + \Delta' \frac{1}{n}\sum_{i\neq j} P_{ij}^2 \\ &= O\left(\frac{K}{n}\right) \end{aligned}$$

$$E\left(hh' H^{-1} T_1^{H\prime}\right) = -\frac{1}{n}\sum_i \frac{x_i^{*2}}{\sigma_i^2}\left[\overline{\sigma}_i^2 - \sigma_i^2\right] = O\left(\|f\|\left\|\sigma^2\right\|\right)$$

$$
\begin{aligned}
E\left(hh'H^{-1}T_2^{H\prime}\right) &= -\frac{H^{-1}}{n^2}\sum_i \frac{x_i^{*4}}{\sigma_i^4}E\left(v_i\epsilon_i^{*2}\right) \\
&\leq \Delta\frac{H^{-1}}{n^2}\sum_i \frac{x_i^{*4}}{\sigma_i^4} \\
&= O\left(\frac{1}{n}\right)
\end{aligned}
$$

$$
E\left(hh'H^{-1}T_3^{H\prime}\right) = H^{-1}\left[\frac{1}{n}\sum_i \frac{x_i^{*2}}{\sigma_i^4}\left[\overline{\sigma}_i^2 - \sigma_i^2\right]^2\right]H = O\left(\left\|\sigma^2\right\|^2\right)
$$

$$
\begin{aligned}
E\left(hh'H^{-1}T_4^{H\prime}\right) &= \frac{H^{-1}}{n^2}\sum_{ij}\frac{x_i^{*2}x_j^{*2}}{\sigma_i^4}P_{ij}^2E\left(v_j^2\epsilon_j^{*2}\right) + \frac{1}{n}\sum_{ij}\frac{x_i^{*2}}{\sigma_i^4}P_{ij}^2E\left(v_j^2\right) \\
&= O\left(\frac{K}{n^2}\right) + O\left(\frac{K}{n}\right)
\end{aligned}
$$

Combining above results we have[7]

$$
\begin{aligned}
& E\left[\widehat{A}(K)\right] \\
= & \frac{1}{n}X^{*\prime}X^{*} - 2\frac{1}{n}\sum_{i}\frac{x_i^{*2}}{\sigma_i^2}\left[\overline{\sigma}_i^2-\sigma_i^2\right] - 2\frac{1}{n}\sum_{i}\frac{x_i^{*2}}{\sigma_i^4}P_{ii}E\left[v_i^2\right] \\
& +4\frac{1}{n}\sum_{i}x_i^{*2}P_{ii} + 2\frac{1}{n}\sum_{i}\frac{x_i^{*2}}{\sigma_i^4}\left[\overline{\sigma}_i^2-\sigma_i^2\right]^2 \\
& +2\frac{1}{n}\sum_{i\neq j}\frac{x_i^{*2}}{\sigma_i^4}P_{ij}^2E\left[v_j^2\right] + \frac{1}{n}\sum_{i}\frac{x_i^{*2}}{\sigma_i^4}\left[\overline{\sigma}_i^2-\sigma_i^2\right]^2 \\
& +\frac{1}{n}\sum_{i\neq j}\frac{x_i^{*2}}{\sigma_i^4}P_{ij}^2E\left[v_j^2\right] + 2\frac{1}{n}\sum_{i}\frac{x_i^{*2}}{\sigma_i^2}\left[\overline{\sigma}_i^2-\sigma_i^2\right] \\
& -2\frac{1}{n}\sum_{i}\frac{x_i^{*2}}{\sigma_i^4}\left[\overline{\sigma}_i^2-\sigma_i^2\right]^2 - 2\frac{1}{n}\sum_{i\neq j}\frac{x_i^{*2}}{\sigma_i^4}P_{ij}^2E\left[v_j^2\right] \\
= & \frac{1}{n}X^{*\prime}X^{*} - 2\frac{1}{n}\sum_{i}\frac{x_i^{*2}}{\sigma_i^4}P_{ii}\left(\overline{\kappa}_i-\sigma_i^4\right) + 4\frac{1}{n}\sum_{i}x_i^{*2}P_{ii} \\
& +\frac{1}{n}\sum_{i}\frac{x_i^{*2}}{\sigma_i^4}\left[\overline{\sigma}_i^2-\sigma_i^2\right]^2 + \frac{1}{n}\sum_{i\neq j}\frac{x_i^{*2}}{\sigma_i^4}P_{ij}^2\left(\overline{\kappa}_j-\sigma_j^4\right) \\
= & \frac{1}{n}X^{*\prime}X^{*} - 2\frac{1}{n}\sum_{i}\frac{x_i^{*2}}{\sigma_i^4}P_{ii}\overline{\kappa}_i + 6\frac{1}{n}\sum_{i}x_i^{*2}P_{ii} \\
& +\frac{1}{n}\sum_{i}\frac{x_i^{*2}}{\sigma_i^4}\left[\overline{\sigma}_i^2-\sigma_i^2\right]^2 + \frac{1}{n}\sum_{i\neq j}\frac{x_i^{*2}}{\sigma_i^4}P_{ij}^2\overline{\kappa}_j - \frac{1}{n}\sum_{i\neq j}\frac{x_i^{*2}}{\sigma_i^4}P_{ij}^2\sigma_j^4.
\end{aligned}
$$

[7] We'll use the fact that

$$
\begin{aligned}
E\left(v_i^2\right) &= E\left(\epsilon_i^4\right)-\sigma_i^4 \\
&= \overline{\kappa}_i-\sigma_i^4 \\
&= \sigma_i^4\kappa_i-\sigma_i^4 \\
&= \sigma_i^4\left(\kappa_i-1\right)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
& \Omega^{*-1}\left[MSE_{Series}\right]\Omega^{*-1} \\
= \; & \Omega^{*-1} E\left[n\left(\widehat{\beta}_{Series}-\beta\right)\left(\widehat{\beta}_{Series}-\beta\right)^{\prime}\right]\Omega^{*-1} \\
= \; & \frac{X^{*\prime}X^{*}}{n}+\frac{1}{n}\sum_{i}\frac{x_i^{*2}}{\sigma_i^4}\left[\overline{\sigma}_i^2-\sigma_i^2\right]^2-2\frac{1}{n}\sum_{i}\frac{x_i^{*2}}{\sigma_i^4}P_{ii}\kappa_i \\
& +\frac{1}{n}\sum_{i\neq j}\frac{x_i^{*2}}{\sigma_i^4}P_{ij}^2\kappa_j+6\frac{1}{n}\sum_{i}x_i^{*2}P_{ii}-\frac{1}{n}\sum_{i\neq j}\frac{\sigma_j^4}{\sigma_i^4}x_i^{*2}P_{ij}^2+o\left(\frac{K}{n}\right).
\end{aligned}
$$

Q.E.D.

A.3 Proof of Proposition 2.3.3

We are actually regressing $e2$ on Q. We know that

$$e2=\epsilon 2+\left(\widehat{\beta}-\beta\right)^2 x-2\left(\widehat{\beta}-\beta\right)x\epsilon.$$

The predicted value is

$$
\begin{aligned}
\widetilde{\sigma 2} &= Q\left(Q'Q\right)^{-1}Q'e2 \\
&= Q\left(Q'Q\right)^{-1}Q'\epsilon 2+\left(\widehat{\beta}-\beta\right)^2 Q\left(Q'Q\right)^{-1}Q'x \\
&\quad -2\left(\widehat{\beta}-\beta\right)Q\left(Q'Q\right)^{-1}Q'x\epsilon \\
&= P\epsilon 2+\left(\widehat{\beta}-\beta\right)^2 Px-2\left(\widehat{\beta}-\beta\right)Px\epsilon.
\end{aligned}
$$

We do know the behavior of $P\epsilon 2$. The sample MSE of $\widetilde{\sigma 2} - \sigma 2$ is

$$
\begin{aligned}
& \frac{1}{n} E\left[\left(\widetilde{\sigma 2}-\sigma 2\right)^{\prime}\left(\widetilde{\sigma 2}-\sigma 2\right)\right] \\
= \; & \frac{1}{n} \sigma 2^{\prime}(I-P) \sigma 2+\frac{1}{n} E\left(v^{\prime} P v\right)+\frac{1}{n} E\left[\left(\widehat{\beta}-\beta\right)^{2} x^{\prime} P x\right] \\
& +\frac{4}{n} E\left[\left(\widehat{\beta}-\beta\right)^{2} x \epsilon^{\prime} P x \epsilon\right]+\frac{2}{n} E\left[\left(\widehat{\beta}-\beta\right)^{2} v^{\prime} P x\right] \\
& -\frac{4}{n} E\left[\left(\widehat{\beta}-\beta\right) v^{\prime} P x \epsilon\right]-\frac{4}{n} E\left[\left(\widehat{\beta}-\beta\right)^{3} x^{\prime} P x \epsilon\right] .
\end{aligned} \tag{A.24}
$$

It is easy to verify that only the first two terms in (A.24) matter. Hence, we have

$$
\frac{1}{n} E\left[\left(\widetilde{\sigma 2}-\sigma 2\right)^{\prime}\left(\widetilde{\sigma 2}-\sigma 2\right)\right]=\frac{1}{n} E\left[\left(\widehat{\sigma 2}-\sigma 2\right)^{\prime}\left(\widehat{\sigma 2}-\sigma 2\right)\right].
$$

Q.E.D.

A.4 Proof of Proposition 2.3.4

The proof follows the proposition 1A of Linton (1996).

Q.E.D.

A.5 Proof of Proposition 2.3.5

By $\sigma_i^2 = \sigma^2$, the terms of Cragg estimator are as followed.

$$
\begin{aligned}
\frac{1}{n} X^{*\prime} \left(I - P^*\right) X^* &= \frac{1}{n} \sum_i \left[\overline{x}_i^* - x_i^*\right]^2 \\
&= \frac{1}{n} \sum_i \left[\sum_j P_{ij}^* x_j^* - x_i^*\right]^2 \\
&= \frac{1}{n} \sum_i \left[\sigma_i Q_i' \left(\sum_i Q_i Q_i' \sigma_i^2\right)^{-1} \sigma_j Q_j \frac{x_j}{\sigma_j} - \frac{x_i}{\sigma_i}\right]^2 \\
&= \frac{1}{n} \sum_i \left[\sigma Q_i' \left(\sum_i Q_i Q_i' \sigma^2\right)^{-1} \sigma Q_j \frac{x_j}{\sigma} - \frac{x_i}{\sigma}\right]^2 \\
&= \frac{1}{n\sigma^2} \sum_i \left[Q_i' \left(Q'Q\right)^{-1} Q_j x_j - x_i\right]^2 \\
&= \frac{1}{n\sigma^2} \sum_i \left[\sum_j P_{ij} x_j - x_i\right]^2 \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
& -\frac{1}{n} \sum_i x_i^{*2} P_{ii}^* \kappa_i^* \\
= & -\frac{1}{n} \sum_i \frac{x_i^2}{\sigma_i^2} \sigma_i Q_i' \left(\sum_i Q_i Q_i' \sigma_i^2\right)^{-1} \sigma_i Q_i \frac{\kappa_i}{\sigma_i^4} \\
= & -\frac{1}{n\sigma^6} \sum_i x_i^2 Q_i' \left(Q'Q\right)^{-1} Q_i \kappa_i \\
= & -\frac{1}{n\sigma^6} \sum_i x_i^2 P_{ii} \kappa_i
\end{aligned}
$$

$$
\begin{aligned}
& \frac{5}{n}\sum_i x_i^{*2} P_{ii}^* \\
= \ & \frac{5}{n}\sum_i \frac{x_i^2}{\sigma_i^2}\sigma_i Q_i' \left(\sum_i Q_i Q_i' \sigma_i^2\right)^{-1} \sigma_i Q_i \\
= \ & \frac{5}{n\sigma^2}\sum_i x_i^2 P_{ii}
\end{aligned}
$$

The terms of series based FGLS estimator are as followed.

$$
\begin{aligned}
& \frac{1}{n}\sum_i \frac{x_i^{*2}}{\sigma_i^4}\left[\overline{\sigma}_i^2 - \sigma_i^2\right]^2 \\
= \ & \frac{1}{n}\sum_i \frac{x_i^2}{\sigma_i^6}\left[\sum_j P_{ij}\sigma_j^2 - \sigma_i^2\right]^2 \\
= \ & \frac{1}{n\sigma^2}\sum_i x_i^2 \left[\sum_j P_{ij} - 1\right]^2 \\
= \ & \frac{1}{n\sigma^2}\sum_i \left[\sum_j P_{ij} x_i - x_i\right]^2 \\
= \ & 0
\end{aligned}
$$

$$
\begin{aligned}
& -2\frac{1}{n}\sum_i \frac{x_i^{*2}}{\sigma_i^4} P_{ii}\kappa_i \\
= \ & -2\frac{1}{n\sigma^6}\sum_i x_i^2 P_{ii}\kappa_i
\end{aligned}
$$

$$
\begin{aligned}
& \frac{1}{n}\sum_{i\neq j}\frac{1}{\sigma_i^4}x_i^{*2}P_{ij}^2\kappa_j \\
= & \frac{1}{n\sigma^6}\sum_{i\neq j}x_i^2P_{ij}^2\kappa_j \\
= & \frac{1}{n\sigma^6}\left(\sum_{i,j}x_i^2P_{ij}^2\kappa_j - \sum_i x_i^2P_{ii}^2\kappa_i\right) \\
= & \frac{1}{n\sigma^6}\sum_{i,j}x_i^2Q_i'\left(Q'Q\right)^{-1}\left(Q_j\kappa_jQ_j'\right)\left(Q'Q\right)^{-1}Q_i \\
& -\frac{1}{n\sigma^6}\sum_i x_i^2Q_i'\left(Q'Q\right)^{-1}\left(Q_i\kappa_iQ_i'\right)\left(Q'Q\right)^{-1}Q_i \\
= & \frac{1}{n\sigma^6}\sum_i x_i^2Q_i'\left(Q'Q\right)^{-1}\left(\sum_j Q_j\kappa_jQ_j'\right)\left(Q'Q\right)^{-1}Q_i \\
& -\frac{1}{n\sigma^6}\sum_i x_i^2Q_i'\left(Q'Q\right)^{-1}\left(Q_i\kappa_iQ_i'\right)\left(Q'Q\right)^{-1}Q_i \\
= & \frac{1}{n\sigma^6}\sum_i x_i^2P_{ii}\kappa_i - \frac{1}{n\sigma^6}\sum_i x_i^2P_{ii}^2\kappa_i
\end{aligned}
$$

$$
\begin{aligned}
& 6\frac{1}{n}\sum_i x_i^{*2}P_{ii} \\
= & 6\frac{1}{n\sigma^2}\sum_i x_i^2P_{ii}
\end{aligned}
$$

$$
\begin{aligned}
& -\frac{1}{n}\sum_{i\neq j}\frac{\sigma_j^4}{\sigma_i^4}x_i^{*2}P_{ij}^2 \\
= & -\frac{1}{n\sigma^2}\sum_{i\neq j}x_i^2P_{ij}^2 \\
= & -\frac{1}{n\sigma^2}\left(\sum_{i,j}x_i^2P_{ij}^2-\sum_i x_i^2P_{ii}^2\right) \\
= & -\frac{1}{n\sigma^2}\sum_{i,j}x_i^2Q_i'\left(Q'Q\right)^{-1}\left(Q_jQ_j'\right)\left(Q'Q\right)^{-1}Q_i \\
& +\frac{1}{n\sigma^2}\sum_i x_i^2Q_i'\left(Q'Q\right)^{-1}\left(Q_iQ_i'\right)\left(Q'Q\right)^{-1}Q_i \\
= & -\frac{1}{n\sigma^2}\sum_i x_i^2Q_i'\left(Q'Q\right)^{-1}\left(\sum_j Q_jQ_j'\right)\left(Q'Q\right)^{-1}Q_i \\
& +\frac{1}{n\sigma^2}\sum_i x_i^2Q_i'\left(Q'Q\right)^{-1}\left(Q_iQ_i'\right)\left(Q'Q\right)^{-1}Q_i \\
= & -\frac{1}{n\sigma^2}\sum_i x_i^2P_{ii}+\frac{1}{n\sigma^2}\sum_i x_i^2P_{ii}^2.
\end{aligned}
$$

Dropping out the terms of small orders gives the results.

Q.E.D.

A.6 Proof of Proposition 2.3.6

If we think about the special case of homoskedasticity, the term B_i will be zero as well as $\widetilde{B}_i$. The order of B_i is not $O\left(h^2\right)$ anymore.

$$
B_i=\sum_{j\neq i}\left(\sigma_j^2-\sigma_i^2\right)w_{ij}=\sum_{j\neq i}\left(\sigma^2-\sigma^2\right)w_{ij}=0=o\left(1\right).
$$

Let's consider the MSE of $M_n\sqrt{n}\left(\widehat{\beta}_{FGLS}-\beta\right)=\Omega^{*-1}\sqrt{n}\left(\widehat{\beta}_{FGLS}-\beta\right)$. First,

$$\begin{aligned}
&\Omega^{*-1}\sqrt{n}\left(\widehat{\beta}_{FGLS}-\beta\right)\\
=\;&\{X_{N0}-\left[L_{N1}-\frac{b'_{D1}M_n^{-1}X_{N0}}{\sqrt{n}}\right]+\left[L_{N2}-\frac{b'_{D2}M_n^{-1}X_{N0}}{\sqrt{n}}\right]\\
&+\left[L_{N3}-\frac{b'_{D3}M_n^{-1}X_{N0}}{\sqrt{n}}\right]-Q_{N1}+Q_{N2}+C_{N1}-\frac{X'_{D0}M_n^{-1}X_{N0}}{\sqrt{n}}\}+o_p\left(n^{-2\mu}\right)
\end{aligned}$$

Then, MSE of $\Omega^{*-1}\sqrt{n}\left(\widehat{\beta}_{FGLS}-\beta\right)$ is

$$\begin{aligned}
&Var\left[X_{N0}\right]+Var\left[L_{N1}-\frac{b'_{D1}M_n^{-1}X_{N0}}{\sqrt{n}}\right]+Var\left[Q_{N1}\right]\\
=\;&M_n+h^4\left[M_n^{-1}\left[\Gamma_2-\Gamma_1M_n^{-1}\Gamma_1\right]M_n^{-1}\right]+n^{-1}h^{-d}\left[\left(\kappa_3^2+2+\kappa_4\right)M_n^{-1}M_n^*M_n^{-1}\right]\\
=\;&M_n+n^{-1}h^{-d}\left(\kappa_3^2+2+\kappa_4\right)M_n^*\\
=\;&\frac{1}{n\sigma^2}\sum_{i=1}^{n}x_ix_i'+\left(\kappa_3^2+2+\kappa_4\right)\frac{1}{n\sigma^2}\sum_{i=1}^{n}\sum_{j\neq i}x_ix_i'w_{ij}^2\\
=\;&\frac{1}{n\sigma^2}\sum_{i=1}^{n}x_ix_i'+\frac{1}{n\sigma^6}\sum_{i\neq j}x_ix_i'w_{ij}^2\kappa_j-\frac{1}{n\sigma^2}\sum_{i\neq j}x_ix_i'w_{ij}^2.
\end{aligned}$$

We use our previous setting that

$$
\begin{aligned}
\kappa_3^2 &= 0,\ E\left[\epsilon_j^4\right] = \kappa_j \\
Var\left[Q_{N1}\right] &= \sum \sum_{j\neq i} \sum_{k\neq l} \sum \rho_{ij}\rho'_{lk} E\left[\epsilon_i \xi_j \epsilon_k \xi_l\right] \\
&= \sum_{j\neq i} \sum \rho_{ij}\rho'_{ij}\sigma_i^2 \left[E\left[\epsilon_j^4\right] - \sigma_j^4\right] \\
&= n^{-1} \sum_{i=1}^{n} \sum_{j\neq i} w_{ij}^2 x_i x_i' \overline{\sigma}_i^{-8} \sigma_i^2 \left[E\left[\epsilon_j^4\right] - \sigma_j^4\right] \\
&\approx n^{-1} \sum_{i\neq j} x_i x_i' \sigma_i^{-2} \left[\frac{E\left[\epsilon_j^4\right]}{\sigma_j^4} - 1\right] w_{ij}^2 \\
&= n^{-1} \sum_{i\neq j} x_i x_i' \sigma^{-6} E\left[\epsilon_j^4\right] w_{ij}^2 - n^{-1} \sum_{i\neq j} x_i x_i' \sigma^{-2} w_{ij}^2 \\
&= \frac{1}{n\sigma^6} \sum_{i\neq j} x_i x_i' w_{ij}^2 \kappa_j - \frac{1}{n\sigma^2} \sum_{i\neq j} x_i x_i' w_{ij}^2 .
\end{aligned}
$$

Q.E.D.

Appendix B

Appendix-Chapter 3

B.1 Lemmata

Let C denote the generic constants throughout this Appendix. The Eucledean norm $\|\cdot\|$ for a matrix A is defined as $\|A\| = [tr(A'A)]^{1/2}$. Let C.S. denote Cauchy-Schwartz inequality. According to the notaion by Robinson (1988), for scalar or column vector sequences A_i and B_i, we define $S_{A,B} = n^{-1}\sum_{i=1}^{n} A_i B_i'$ and $S_A = S_{A,A}$. The following lemmata of Li (2000) are useful in the proof of our theorems. The proofs are referred to Li (2000: p.1089-1090).[1]

Lemma B.1.1. $\widehat{Q} - I = O_p\left(\zeta(K)\sqrt{K}/\sqrt{n}\right)$, *where* $\widehat{Q} = (P'P/n)$.

Lemma B.1.2. $\|\widetilde{\pi}_f - \pi_f\| = O_p(K^{-\alpha})$, *where* $\widetilde{\pi}_f = (P'P)^{-1} P'f$, *and* $f = g$ *or* $f = h$.

Lemma B.1.3. $(Q'\eta/n) = O_p(\zeta(K)/\sqrt{n}) = o_p(1)$.

Lemma B.1.4. $S_{f-\widetilde{f}} = O_p(K^{-2\alpha}) = o_p\left(n^{-1/2}\right)$, *where* $f = g$ *or* $f = h$.

[1] Note that we adopt the notation from Li (2000). In our paper, since we don't consider additive partial linear model, there is no need to decompose ϵ as $v + \eta$. However, the order related to $\widetilde{\epsilon}$ in our proof is similar to the order of $\widetilde{v}$ in Li's paper. For instance, it is trivial to see $S_{\widetilde{\epsilon}} = O_p(K/n)$.

Lemma B.1.5. *(i)* $S_{\widetilde{v}} = O_p(K/n)$, *(ii)* $S_{\widetilde{u}} = O_p(K/n)$, *(iii)* $S_{\widetilde{\eta}} = o_p(1)$.

Corollary B.1.1. *If we replace the approximating function p_K by the normalized version (say $p_K^* = (p_{K1}/\sigma_1, p_{K2}/\sigma_2, ..., p_{Kn}/\sigma_n)'$), Lemma B.1.1-B.1.5 still hold.*

Corollary B.1.2. *If we replace the random variables by the normalized version (e.g. $f = (f_1/\sigma_1, f_2/\sigma_2, ..., f_n/\sigma_n)'$), Lemma B.1.1-B.1.5 still hold.*

B.2 Proof of Theorem 3.3.1

We can write $\sqrt{n}\left(\widehat{\beta}_{GLS} - \beta\right)$ as

$$\begin{aligned}\sqrt{n}\left(\widehat{\beta}_{GLS} - \beta\right) &= \left[\frac{x^{*\prime}(I - Q^*)\,x^*}{n}\right]^{-1}\sqrt{n}\left[\frac{x^*(I - Q^*)(g^* + u^*)}{n}\right] \\ &= S_{x^*-\widetilde{x}^{**}}^{-1}\sqrt{n}S_{x^*-\widetilde{x}^{**},g^*-\widetilde{g}^{**}+u^*-\widetilde{u}^{**}}. \qquad \text{(B.1)}\end{aligned}$$

What we want is to prove that the first term in (B.1) coverges in probability by Law of Large Number and the second term converges in distribution by Lindberg-Levi Central Limit Theorem. We use the following propositions to prove the results.

Proposition B.2.1. $x^{*\prime}(I - Q^*)\,x^*/n = S_{x^*-\widetilde{x}^{**}} = E\left[\epsilon_i\epsilon_i'/\sigma_i^2\right] + o_p(1)$.

Proof. Let $\widetilde{x}_i^* = p_{Ki}^* \left(p_K^{*\prime} p_K^*\right)^{-1} p_K^{*\prime} x$. Using the definition of x_i and $\widetilde{x}_i^*$ gives

$$\begin{aligned}
& \frac{1}{n} x^{*\prime} \left(I - Q^*\right) x^* \\
= \; & \frac{1}{n} \sum_{i=1}^{n} \frac{\epsilon_i \epsilon_i'}{\sigma_i^2} + \frac{1}{n} \sum_{i=1}^{n} \frac{\left[\left(h_i - \widetilde{h}_i^*\right) - \widetilde{\epsilon}_i^*\right]\left[\left(h_i - \widetilde{h}_i^*\right) - \widetilde{\epsilon}_i^*\right]'}{\sigma_i^2} \\
& + \frac{2}{n} \sum_{i=1}^{n} \frac{\epsilon_i \left[\left(h_i - \widetilde{h}_i^*\right) - \widetilde{\epsilon}_i^*\right]'}{\sigma_i^2} \\
= \; & S_{\epsilon^*} + S_{\left(h^* - \widetilde{h}^{**}\right) - \widetilde{\epsilon}^{**}} + S_{\widetilde{\epsilon}^{**}, \left(h^* - \widetilde{h}^{**}\right) - \widetilde{\epsilon}^{**}}. \qquad \text{(B.2)}
\end{aligned}$$

Note that the variables with single "*" represent nomalization by σ_i and the variables with double "**" stand for normalized variables which are premutiplied by normalized projection matrix Q^*. By LLN, the first term in (B.2) will converge to $E\left[\epsilon_i \epsilon_i' / \sigma_i^2\right] + o_p(1)$. We also have the inequality $S_{\left(h^* - \widetilde{h}^{**}\right) - \widetilde{\epsilon}^{**}} \leq 2\left[S_{h^* - \widetilde{h}^{**}} + S_{\widetilde{\epsilon}^{**}}\right] = o_p(1)$ by Corollary B.1.1, Corollary B.1.2, Lemma B.1.4, and Lemma B.1.5 (i) and (iii). Applying CS on the last term in (B.2) gives

$$S_{\widetilde{\epsilon}^{**}, \left(h^* - \widetilde{h}^{**}\right) - \widetilde{\epsilon}^{**}} \leq \left(S_{\widetilde{\epsilon}^{**}} S_{\left(h^* - \widetilde{h}^{**}\right) - \widetilde{\epsilon}^{**}}\right)^{1/2} = \left(O_p(1)\, o_p(1)\right)^{1/2} = o_p(1).$$

□

Proposition B.2.2. $S_{x^* - \widetilde{x}^{**}, g^* - \widetilde{g}^{**}} = o_p\left(n^{-1/2}\right)$

Proof. Using definition of $x^*$and $\widetilde{x}^{**}$ gives

$$S_{x^* - \widetilde{x}^{**}, g^* - \widetilde{g}^{**}} = S_{\epsilon^*, g^* - \widetilde{g}^{**}} + S_{h^* - \widetilde{h}^{**}, g^* - \widetilde{g}^{**}} - S_{\widetilde{\epsilon}^{**}, g^* - \widetilde{g}^{**}}.$$

1. $S_{\epsilon^*, g^* - \widetilde{g}^{**}} \leq \left(S_{\epsilon^*} S_{g^* - \widetilde{g}^{**}}\right)^{1/2} = O_p\left(K^{-\alpha}\right)$ by C.S., Proposition B.2.1 and Lemma B.1.4.

2. $S_{h^*-\widetilde{h}^{**},g^*-\widetilde{g}^{**}} \leq \left(S_{h^*-\widetilde{h}^{**}} S_{g^*-\widetilde{g}^{**}}\right)^{1/2} = O_p\left(K^{-2\alpha}\right)$ by C.S., and Lemma B.1.4.

3. $S_{\widetilde{\epsilon}^{**},g^*-\widetilde{g}^{**}} \leq \left(S_{\widetilde{\epsilon}^{**}} S_{g^*-\widetilde{g}^{**}}\right)^{1/2} = o_p(1)\, O_p\left(K^{-\alpha}\right)$ by C.S., Lemma B.1.5 (i) and Lemma B.1.4.

□

Proposition B.2.3. $S_{x^*-\widetilde{x}^{**},\widetilde{u}^{**}} = o_p\left(n^{-1/2}\right).$

Proof. Using definition of x^* and $\widetilde{x}^{**}$ gives

$$S_{x^*-\widetilde{x}^{**},g^*-\widetilde{g}^{**}} = S_{\epsilon^*,\widetilde{u}^{**}} + S_{h^*-\widetilde{h}^{**},\widetilde{u}^{**}} - S_{\widetilde{\epsilon}^{**},\widetilde{u}^{**}}.$$

1. $E\left[\|S_{\epsilon^*,\widetilde{u}^{**}}\|^2 |\mathcal{Z}\right] = n^{-2} tr\left[Q^* \epsilon^* \epsilon^{*\prime} Q^* E\left[u^* u^{*\prime}|\mathcal{Z}\right]\right] \leq C n^{-2} tr\left[\widetilde{\epsilon}^{**} \widetilde{\epsilon}^{**\prime}\right]$

   $= C n^{-1} tr\left(S_{\widetilde{\epsilon}^{**}}\right) = O_p\left(K/n^2\right)$ by C.S. and Lemma B.1.5 (i).

2. $S_{h^*-\widetilde{h}^{**},\widetilde{u}^{**}} \leq \left(S_{h^*-\widetilde{h}^{**}} S_{\widetilde{u}^{**}}\right)^{1/2} = O_p\left(K^{-\alpha}\right) O_p\left(\sqrt{K}/\sqrt{n}\right)$ by C.S., Lemma B.1.4 and Lemma B.1.5 (ii).

3. $S_{\widetilde{\epsilon}^{**},\widetilde{u}^{**}} \leq \left(S_{\widetilde{\epsilon}^{**}} S_{\widetilde{u}^{**}}\right)^{1/2} = O_p\left(K/n\right)$ by C.S., Lemma B.1.5 (i) and Lemma B.1.5 (ii).

□

Proposition B.2.4. $\sqrt{n} S_{x^*-\widetilde{x}^{**},u^*} \xrightarrow{d} N\left(0, E\left[\epsilon_i \epsilon_i'/\sigma_i^2\right]\right).$

Proof. Using definition of x^* and $\widetilde{x}^{**}$ gives

$$S_{x^*-\widetilde{x}^{**},u^*} = S_{\epsilon^*,u^*} + S_{h^*-\widetilde{h}^{**},u^*} - S_{\widetilde{\epsilon}^{**},u^*}.$$

1. $\sqrt{n}S_{\epsilon^*,u^*} = \sum_{i=1}^{n}\left[\epsilon_i u_i/\sigma_i^2\right]/\sqrt{n} \xrightarrow{d} N\left(0, E\left[\epsilon_i\epsilon_i'/\sigma_i^2\right]\right)$ by Lindberg-Levi Central Limit Theorem.

2. $E\left[\left\|S_{h^*-\widetilde{h}^{**},u^*}\right\|^2|\mathcal{Z}\right] = n^{-2}tr\left[\left(h^*-\widetilde{h}^{**}\right)\left(h^*-\widetilde{h}^{**}\right)' E\left(u^*u^{*\prime}|\mathcal{Z}\right)\right]$
$\leq Cn^{-1}tr\left[S_{h^*-\widetilde{h}^{**}}\right] = o_p\left(n^{-1}\right)$ by C.S. and Lemma B.1.4.

3. $E\left[\left\|S_{\widetilde{\epsilon}^{**},u^*}\right\|^2|\mathcal{Z}\right] = n^{-2}tr\left[Q^*\epsilon^*\epsilon^{*\prime}Q^*E\left(u^*u^{*\prime}|\mathcal{Z}\right)\right] \leq Cn^{-1}tr\left[S_{\widetilde{\epsilon}^{**}}\right] = o_p\left(n^{-1}\right)$ by C.S. and Lemma B.1.5 (i).

□

Combining Proposition B.2.1-B.2.4 proves Theorem 1.

B.3 Proof of Theorem 3.3.2

Our new estimator could be written as

$$\begin{aligned}\sqrt{n}\left(\widetilde{\beta}_{GLS}-\beta\right) &= \left[\frac{x(I-Q)\Sigma^{-1}(I-Q)x}{n}\right]^{-1}\sqrt{n}\left[\frac{x(I-Q)\Sigma^{-1}(I-Q)\left(g+u\right)}{n}\right] \\ &= S_{x^*-\widetilde{x}^*}^{-1}\sqrt{n}S_{x^*-\widetilde{x}^*,g^*-\widetilde{g}^*+u^*-\widetilde{u}^*}. \qquad \text{(B.3)}\end{aligned}$$

What we want is to prove that the first term in (B.3) coverges in probability by Law of Large Number and the second term converges in distribution by Central Limit Theorem. We use the following propositions to prove the results.

Proposition B.3.1. $x^{*\prime}\left(I-Q\right)x^*/n = S_{x^*-\widetilde{x}^*} = E\left[\epsilon_i\epsilon_i'/\sigma_i^2\right] + o_p\left(1\right).$

Proof. Let $\widetilde{x}_i = p_{Ki}\left(p_K' p_K\right)^{-1} p_K' x$. Using the definition of x_i and $\widetilde{x}_i$ gives

$$\begin{aligned} & \frac{1}{n} x^{*\prime}\left(I-Q\right) x^* \\ = & \frac{1}{n}\sum_{i=1}^{n}\frac{\epsilon_i\epsilon_i'}{\sigma_i^2} + \frac{1}{n}\sum_{i=1}^{n}\frac{\left[\left(h_i-\widetilde{h}_i\right)-\widetilde{\epsilon}_i\right]\left[\left(h_i-\widetilde{h}_i\right)-\widetilde{\epsilon}_i\right]'}{\sigma_i^2} \\ & + \frac{2}{n}\sum_{i=1}^{n}\frac{\epsilon_i\left[\left(h_i-\widetilde{h}_i\right)-\widetilde{\epsilon}_i\right]'}{\sigma_i^2} \\ = & S_{\epsilon^*} + S_{\left(h^*-\widetilde{h}^*\right)-\widetilde{\epsilon}^*} + S_{\epsilon^*,\left(h^*-\widetilde{h}^*\right)-\widetilde{\epsilon}^*}. \end{aligned} \tag{B.4}$$

Note that here we only have the variables with single "*" representing nomalization by σ_i. And the projection matrix Q is not normalized by σ_i. By LLN, the first term in (B.4) will converge to $E\left[\epsilon_i\epsilon_i'/\sigma_i^2\right] + o_p(1)$. We also have the inequality $S_{\left(h^*-\widetilde{h}^*\right)-\widetilde{\epsilon}^*} \leq 2\left[S_{h^*-\widetilde{h}^*} + S_{\widetilde{\epsilon}^*}\right] = o_p(1)$ by Corollary B.1.2, Lemma B.1.4, and Lemma B.1.5 (i) and (iii). Applying CS on the last term in (B.4) gives

$$S_{\epsilon^*,\left(h^*-\widetilde{h}^*\right)-\widetilde{\epsilon}^*} \leq \left(S_{\epsilon^*} S_{\left(h^*-\widetilde{h}^*\right)-\widetilde{\epsilon}^*}\right)^{1/2} = \left(O_p(1)\, o_p(1)\right)^{1/2} = o_p(1).$$

□

Proposition B.3.2. $S_{x^*-\widetilde{x}^*,g^*-\widetilde{g}^*} = o_p\left(n^{-1/2}\right)$

Proof. Using definition of $x^*$and $\widetilde{x}^*$ gives

$$S_{x^*-\widetilde{x}^*,g^*-\widetilde{g}^*} = S_{\epsilon^*,g^*-\widetilde{g}^*} + S_{h^*-\widetilde{h}^*,g^*-\widetilde{g}^*} - S_{\widetilde{\epsilon}^*,g^*-\widetilde{g}^*}.$$

1. $S_{\epsilon^*,g^*-\widetilde{g}^*} \leq \left(S_{\epsilon^*} S_{g^*-\widetilde{g}^*}\right)^{1/2} = O_p\left(K^{-\alpha}\right)$ by C.S., Proposition B.2.1 and Lemma B.1.4.

2. $S_{h^*-\widetilde{h}^*,g^*-\widetilde{g}^*} \leq \left(S_{h^*-\widetilde{h}^*}S_{g^*-\widetilde{g}^*}\right)^{1/2} = O_p\left(K^{-2\alpha}\right)$ by C.S., and Lemma B.1.4.

3. $S_{\widetilde{\epsilon}^*,g^*-\widetilde{g}^*} \leq \left(S_{\widetilde{\epsilon}^*}S_{g^*-\widetilde{g}^*}\right)^{1/2} = o_p(1)\, O_p\left(K^{-\alpha}\right)$ by C.S., Lemma B.1.5 (i) and Lemma B.1.4.

□

Proposition B.3.3. $S_{x^*-\widetilde{x}^*,\widetilde{u}^*} = o_p\left(n^{-1/2}\right).$

Proof. Using definition of x^* and $\widetilde{x}^*$ gives

$$S_{x^*-\widetilde{x}^*,g^*-\widetilde{g}^*} = S_{\epsilon^*,\widetilde{u}^*} + S_{h-\widetilde{h}^*,\widetilde{u}^*} - S_{\widetilde{\epsilon}^*,\widetilde{u}^*}.$$

1. $E\left[\left\|S_{\epsilon,\widetilde{u}^*}\right\|^2|\mathcal{Z}\right] = n^{-2}tr\left[Q\epsilon\epsilon' QE\left[u^*u^{*\prime}|\mathcal{Z}\right]\right] \leq Cn^{-2}tr\left[\widetilde{\epsilon}\widetilde{\epsilon}'\right] = Cn^{-1}tr\left(S_{\widetilde{\epsilon}}\right) = O_p\left(K/n^2\right)$ by C.S. and Lemma B.1.5 (i).

2. $S_{h-\widetilde{h}^*,\widetilde{u}^*} \leq \left(S_{h-\widetilde{h}^*}S_{\widetilde{u}^*}\right)^{1/2} = O_p\left(K^{-\alpha}\right) O_p\left(\sqrt{K}/\sqrt{n}\right)$ by C.S., Lemma B.1.4 and Lemma B.1.5 (ii).

3. $S_{\widetilde{\epsilon}^*,\widetilde{u}^*} \leq \left(S_{\widetilde{\epsilon}^*}S_{\widetilde{u}^*}\right)^{1/2} = O_p\left(K/n\right)$ by C.S., Lemma B.1.5 (i) and Lemma B.1.5 (ii).

□

Proposition B.3.4. $\sqrt{n}S_{x-\widetilde{x}^*,u} \xrightarrow{d} N\left(0, E\left[\epsilon_i\epsilon_i'/\sigma_i^2\right]\right).$

Proof. Using definition of x and $\widetilde{x}^*$ gives

$$S_{x^*-\widetilde{x}^*,u^*} = S_{\epsilon^*,u^*} + S_{h^*-\widetilde{h}^*,u^*} - S_{\widetilde{\epsilon}^*,u^*}.$$

1. $\sqrt{n}S_{\epsilon^*,u^*} = \sum_{i=1}^n \left[\epsilon_i u_i/\sigma_i^2\right]/\sqrt{n} \xrightarrow{d} N\left(0, E\left[\epsilon_i \epsilon_i'/\sigma_i^2\right]\right)$ by Lindberg-Levi Central Limit Theorem.

2. $E\left[\left\|S_{h^*-\widetilde{h}^*,u^*}\right\|^2 |\mathcal{Z}\right] = n^{-2}tr\left[\left(h^* - \widetilde{h}^*\right)\left(h^* - \widetilde{h}^*\right)' E\left(u^* u^{*\prime}|\mathcal{Z}\right)\right]$
$\leq Cn^{-1}tr\left[S_{h^*-\widetilde{h}^*}\right] = o_p\left(n^{-1}\right)$ by C.S. and Lemma B.1.4.

3. $E\left[\left\|S_{\widetilde{\epsilon}^*,u^*}\right\|^2 |\mathcal{Z}\right] = n^{-2}tr\left[Q\epsilon^*\epsilon^{*\prime}QE\left(u^*u^{*\prime}|\mathcal{Z}\right)\right] \leq Cn^{-1}tr\left[S_{\widetilde{\epsilon}^*}\right] = o_p\left(n^{-1}\right)$ by C.S. and Lemma B.1.5 (i).

□

Combining Proposition B.3.1-B.3.4 proves Theorem 2.

B.4 Proof of Theorem 3.3.3

We can write $\sqrt{n}\left(\widehat{\beta}_{FGLS} - \beta\right)$ as

$$\sqrt{n}\left(\widehat{\beta}_{FGLS} - \beta\right) = \left[\frac{1}{n}\sum_{i=1}^n \frac{\left[x_i - \widetilde{x}_i^*\right]^2}{\widehat{\sigma}_i^2}\right]^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\left[x_i - \widetilde{x}_i^*\right]\left[g_i - \widetilde{g}_i^* + u_i - \widetilde{u}_i^*\right]}{\widehat{\sigma}_i^2}\right].$$

One could get

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \frac{\left[x_i - \widetilde{x}_i^*\right]^2}{\widehat{\sigma}_i^2} &= \frac{1}{n}\sum_{i=1}^n \frac{\left[x_i - \widetilde{x}_i^*\right]^2}{\sigma_i^2} - \frac{1}{n}\sum_{i=1}^n \frac{\left[x_i - \widetilde{x}_i^*\right]^2}{\sigma_i^4}\left(\widehat{\sigma}_i^2 - \sigma_i^2\right) + ... \\ &= \frac{1}{n}\sum_{i=1}^n \frac{\left[x_i - \widetilde{x}_i\right]^2}{\sigma_i^2} + ...\end{aligned}$$

and

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\left[x_i - \widetilde{x}_i^*\right]\left[g_i - \widetilde{g}_i^* + u_i - \widetilde{u}_i^*\right]}{\widehat{\sigma}_i^2} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\left[x_i - \widetilde{x}_i\right]\left[g_i - \widetilde{g}_i + u_i - \widetilde{u}_i\right]}{\sigma_i^2} + ...$$

Therefore, based on the proof in Theorem 3.3.4, the result follows.

B.5 Proof of Theorem 3.3.4

We can write $\sqrt{n}\left(\widetilde{\beta}_{FGLS}-\beta\right)$ as

$$\begin{aligned}\sqrt{n}\left(\widetilde{\beta}_{FGLS}-\beta\right) &= \left[\frac{x'M\widehat{\Sigma}^{-1}Mx}{n}\right]^{-1}\left[\frac{x'M\widehat{\Sigma}^{-1}M\left(g+u\right)}{\sqrt{n}}\right] \\ &= \left[\frac{1}{n}\sum_{i=1}^{n}\frac{\left[x_i-\widetilde{x}_i\right]^2}{\widehat{\sigma}_i^2}\right]^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\left[x_i-\widetilde{x}_i\right]\left[g_i-\widetilde{g}_i+u_i-\widetilde{u}_i\right]}{\widehat{\sigma}_i^2}\right].\end{aligned}$$

Here $M=I-P$. $P=p_K\left(p_K'p_K\right)^{-1}p_K'$. $p_K\left(z\right)$ is the approximating function to approximate unknown function $g\left(z\right)$. We go on estimating σ_i^2 by $\widehat{\sigma}_i^2$, which is defined as

$$\widehat{\sigma}_i^2=P_{2i}\mathbf{e2},$$

where $\mathbf{e2}$ is OLS squared residuals and

$$P_2=p_i^K\left(x,z\right)\left(p^K\left(x,z\right)'p^K\left(x,z\right)\right)^{-1}p^K\left(x,z\right)',$$

where $p^K\left(x,z\right)$ is the approximating function to approximate $\sigma^2\left(x,z\right)$ with heteroskedasticity of unknown form. According to asymptotic expansion, we have

$$\begin{aligned}&\sqrt{n}\left(\widetilde{\beta}_{FGLS}-\beta\right)\\ =\;& \widehat{D}^{-1}\widehat{d}\\ =\;& D^{-1}\left[\widehat{d}+\left(D-\widehat{D}\right)D^{-1}d+\left(D-\widehat{D}\right)D^{-1}\left(\widehat{d}-d\right)+...+R_n\right].\end{aligned}$$

The first order result will depend on the term D and d. The following propositions tell us what do D and d look like.

Proposition B.5.1. $D = \frac{1}{n}\epsilon'\Sigma^{-1}\epsilon = O_p(1)$

Proof.

$$\begin{aligned}\widehat{D} &= \frac{1}{n}\sum_{i=1}^{n}\frac{[x_i-\widetilde{x}_i]^2}{\widehat{\sigma}_i^2} = \frac{1}{n}\sum_{i=1}^{n}\frac{[x_i-\widetilde{x}_i]^2}{\sigma_i^2} - \frac{1}{n}\sum_{i=1}^{n}\frac{[x_i-\widetilde{x}_i]^2}{\sigma_i^4}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)+... \\ &= \widehat{D}_1+\widehat{D}_2+...\end{aligned}$$

$$\begin{aligned}\widehat{D}_1 &= \frac{1}{n}\sum_{i=1}^{n}\frac{[x_i-\widetilde{x}_i]^2}{\sigma_i^2} \\ &= \frac{1}{n}\left[x'(I-P)\Sigma^{-1}(I-P)x\right] = \frac{1}{n}(h+\epsilon)'(I-P)\Sigma^{-1}(I-P)(h+\epsilon) \\ &= \frac{1}{n}h'(I-P)\Sigma^{-1}(I-P)h+\frac{2}{n}h'(I-P)\Sigma^{-1}(I-P)\epsilon \\ &\quad +\frac{1}{n}\epsilon'(I-P)\Sigma^{-1}(I-P)\epsilon.\end{aligned}$$

$$\left|\frac{1}{n}h'(I-P)\Sigma^{-1}(I-P)h\right| \le \left|\frac{1}{n}h'(I-P)h\right| = O\left(\|h\|^2\right)$$

$\frac{2}{n}h'(I-P)\Sigma^{-1}(I-P)\epsilon$ is mean zero and its variance is

$$\begin{aligned}Var[\cdot] &= \frac{4}{n^2}h'(I-P)\Sigma^{-1}(I-P)E[\epsilon\epsilon'](I-P)\Sigma^{-1}(I-P)h \\ &\le \frac{4}{n^2}h'(I-P)\Sigma^{-1}(I-P)\Sigma^{-1}(I-P)h \\ &\le \frac{4}{n^2}h'(I-P)\Sigma^{-2}(I-P)h \\ &\le \frac{4}{n^2}h'(I-P)h \\ &= O\left(\frac{\|h\|^2}{n}\right).\end{aligned}$$

We have

$$\frac{2}{n}h'(I-P)\Sigma^{-1}(I-P)\epsilon = O\left(\frac{\|f\|}{\sqrt{n}}\right)$$

W_i	Term	Order
D	$\frac{1}{n}\epsilon'\Sigma^{-1}\epsilon$	$O(1)$
W_1	$\frac{1}{n}\epsilon' P^{*}\epsilon$	$O\left(\frac{K}{n}\right)$
W_2	$-\frac{2}{n}\epsilon' P^{*}\epsilon^{*}$	$O\left(\frac{K}{n}\right)$
W_3	$\frac{1}{n}h'(I-P)\Sigma^{-1}(I-P)h$	$O\left(\Vert h\Vert^2\right)$
W_4	$\frac{2}{n}h'(I-P)\Sigma^{-1}(I-P)\epsilon$	$O\left(\frac{\Vert h\Vert}{\sqrt{n}}\right)$

Table B.1: Terms of $\widehat{D}_1$

$$\begin{aligned}
& \frac{1}{n}\epsilon'(I-P)\Sigma^{-1}(I-P)\epsilon \\
= & \frac{1}{n}\epsilon'\Sigma^{-1}\epsilon+\frac{1}{n}\epsilon' P\Sigma^{-1}P\epsilon-\frac{2}{n}\epsilon' P\Sigma^{-1}\epsilon \\
= & \frac{1}{n}\epsilon'\Sigma^{-1}\epsilon+\frac{1}{n}\epsilon' P^{*}\epsilon-\frac{2}{n}\epsilon' P^{*}\epsilon^{*} \\
= & O(1)+O\left(\frac{K}{n}\right)+O\left(\frac{K}{n}\right)
\end{aligned}$$

Now we obtain $\widehat{D}_1$.

$$\widehat{D}_1 = D+W_1+W_2+W_3+W_4+o\left(\frac{K}{n}\right)$$

□

Proposition B.5.2. $d=\frac{1}{\sqrt{n}}\epsilon'\Sigma^{-1}u=O_p(1)$

Proof.

$$\begin{aligned}
\widehat{d} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{[x_i-\widetilde{x}_i][g_i-\widetilde{g}_i+u_i-\widetilde{u}_i]}{\widehat{\sigma}_i^2}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{[x_i-\widetilde{x}_i][g_i-\widetilde{g}_i+u_i-\widetilde{u}_i]}{\sigma_i^2} \\
&\quad -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{[x_i-\widetilde{x}_i][g_i-\widetilde{g}_i+u_i-\widetilde{u}_i]}{\sigma_i^4}\left(\widehat{\sigma}_i^2-\sigma_i^2\right)+... \\
&= \widehat{d}_1+\widehat{d}_2
\end{aligned}$$

$$
\begin{aligned}
\widehat{d_1} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{[x_i - \widetilde{x}_i][g_i - \widetilde{g}_i + u_i - \widetilde{u}_i]}{\sigma_i^2} \\
&= \frac{1}{\sqrt{n}} x' (I-P) \Sigma^{-1} (I-P) g + \frac{1}{\sqrt{n}} x' (I-P) \Sigma^{-1} (I-P) u
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{\sqrt{n}} x' (I-P) \Sigma^{-1} (I-P) g &= \frac{1}{\sqrt{n}} h' (I-P) \Sigma^{-1} (I-P) g \\
&\quad + \frac{1}{\sqrt{n}} \epsilon' (I-P) \Sigma^{-1} (I-P) g
\end{aligned}
$$

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \left| h' (I-P) \Sigma^{-1} (I-P) g \right| \\
\leq & \frac{1}{\sqrt{n}} \left(h' (I-P) \Sigma^{-1} (I-P) h \right)^{1/2} \left(g' (I-P) \Sigma^{-1} (I-P) g \right)^{1/2} \\
= & \sqrt{n} O(\|h\|) O(\|g\|)
\end{aligned}
$$

$$
\frac{1}{\sqrt{n}} \epsilon' (I-P) \Sigma^{-1} (I-P) g = O(\|g\|)
$$

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} x' (I-P) \Sigma^{-1} (I-P) u \\
= & \frac{1}{\sqrt{n}} h' (I-P) \Sigma^{-1} (I-P) u + \frac{1}{\sqrt{n}} \epsilon' (I-P) \Sigma^{-1} (I-P) u
\end{aligned}
$$

$$
\frac{1}{\sqrt{n}} h' (I-P) \Sigma^{-1} (I-P) u = O(\|h\|)
$$

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \epsilon' (I-P) \Sigma^{-1} (I-P) u \\
= & \frac{1}{\sqrt{n}} \epsilon' \Sigma^{-1} u + \frac{1}{\sqrt{n}} \epsilon' P \Sigma^{-1} P u - \frac{1}{\sqrt{n}} \epsilon' P \Sigma^{-1} u - \frac{1}{\sqrt{n}} \epsilon' \Sigma^{-1} P u \\
= & \frac{1}{\sqrt{n}} \epsilon' \Sigma^{-1} u + \frac{1}{\sqrt{n}} \epsilon' P^* u - \frac{1}{\sqrt{n}} \epsilon' P^* u^* - \frac{1}{\sqrt{n}} \epsilon^{*\prime} P^* u \\
= & O(1) + O\left(\frac{\sqrt{K}}{\sqrt{n}} \right) + O\left(\frac{\sqrt{K}}{\sqrt{n}} \right) + O\left(\frac{\sqrt{K}}{\sqrt{n}} \right).
\end{aligned}
$$

W_i	Term	Order
d	$\frac{1}{\sqrt{n}}\epsilon'\Sigma^{-1}u$	$O(1)$
W_6	$\frac{1}{\sqrt{n}}\epsilon' P^* u$	$O\left(\frac{\sqrt{K}}{\sqrt{n}}\right)$
W_7	$-\frac{1}{\sqrt{n}}\epsilon' P^* u^*$	$O\left(\frac{\sqrt{K}}{\sqrt{n}}\right)$
W_8	$-\frac{1}{\sqrt{n}}\epsilon^{*\prime} P^* u$	$O\left(\frac{\sqrt{K}}{\sqrt{n}}\right)$
W_9	$\frac{1}{\sqrt{n}}h'(I-P)\Sigma^{-1}(I-P)u$	$O(\lVert h\rVert)$
W_{10}	$\frac{1}{\sqrt{n}}\epsilon'(I-P)\Sigma^{-1}(I-P)g$	$O(\lVert g\rVert)$
W_{11}	$\frac{1}{\sqrt{n}}h'(I-P)\Sigma^{-1}(I-P)g$	$O(\sqrt{n}\lVert h\rVert)O(\lVert g\rVert)$

Table B.2: Terms of $\widehat{d}_1$

Combining the orders gives

$$\widehat{d}_1 = d + W_6 + W_7 + W_8 + W_9 + W_{10} + W_{11} + o_p\left(\frac{K}{n}\right)$$

$\square$

To get the first order asymptotics, we use the result of Proposition B.5.1 and B.5.2, and take the product of $D^{-1}d$. By WLLN,

$$D = \frac{1}{n}\epsilon'\Sigma^{-1}\epsilon = \frac{1}{n}\sum_i \frac{\epsilon_i\epsilon_i'}{\sigma_i^2} \xrightarrow{p} E\left[\frac{\epsilon_i\epsilon_i'}{\sigma_i^2}\right].$$

By Lindberg-Levi CLT,

$$\begin{aligned} d &= \frac{1}{\sqrt{n}}\epsilon'\Sigma^{-1}u = \frac{1}{\sqrt{n}}\sum_i \frac{\epsilon_i u_i}{\sigma_i^2} \xrightarrow{d} N\left(0, E\left[\frac{\epsilon_i\epsilon_i' u_i^2}{\sigma_i^4}\right]\right) \\ &= N\left(0, E\left[\frac{\epsilon_i\epsilon_i' E\left[u_i^2|x_i, z_i\right]}{\sigma_i^4}\right]\right) \\ &= N\left(0, E\left[\frac{\epsilon_i\epsilon_i'}{\sigma_i^2}\right]\right). \end{aligned}$$

Therefore,

$$\begin{aligned} D^{-1}d \xrightarrow{d} N\left(0, E\left[\frac{\epsilon_i\epsilon_i'}{\sigma_i^2}\right]^{-1} \cdot E\left[\frac{\epsilon_i\epsilon_i'}{\sigma_i^2}\right] \cdot E\left[\frac{\epsilon_i\epsilon_i'}{\sigma_i^2}\right]^{-1}\right) \\ = \quad N\left(0, E\left[\frac{\epsilon_i\epsilon_i'}{\sigma_i^2}\right]^{-1}\right), \end{aligned}$$

which attains semiparametric efficiency bound.

Appendix C

Appendix-Chapter 4

C.1 Lemmata

The following Lemmata are useful in the proof of Proposition 4.2.1.

Lemma C.1.1. $\frac{1}{n}\sum\left(\widetilde{g}_i^2 - g_i^2\right) = o_p\left(1\right).$

Proof. It follows from MSE convergence of $\widetilde{g}$ to g, boundedness of g. □

Lemma C.1.2. $\frac{1}{n}\sum x_i\left(\widetilde{g}_i - g_i\right) = o_p\left(1\right)$

Proof. By CS,

$$\begin{aligned}
\frac{1}{n}\sum x_i\left(\widetilde{g}_i - g_i\right) &\leq \left(\frac{1}{n}\sum x_i^2\right)^{1/2}\left(\frac{1}{n}\sum\left(\widetilde{g}_i - g_i\right)^2\right)^{1/2} \\
&= O_p\left(1\right)\cdot\|g_K^r\|_{0,\infty,Z} \\
&= O_p\left(1\right)\cdot o_p\left(1\right) \\
&= o_p\left(1\right).
\end{aligned}$$

□

Lemma C.1.3. $\frac{1}{\sqrt{n}}\sum\left(\widetilde{g}_i - g_i\right)u_i = \frac{1}{\sqrt{n}}\left(s'P_z u - g'u\right) = o_p\left(1\right).$

Proof. Using the fact that $s = g(z) + \epsilon$ gives

$$\begin{aligned}
\frac{1}{\sqrt{n}}\left(s' P_z u - g' u\right) &= \frac{1}{\sqrt{n}}\left(g' P_z u + \epsilon' P_z u - g' u\right) \\
&= \frac{1}{\sqrt{n}}\left[\epsilon' P_z u - g'\left(I - P_z\right) u\right] \\
&= O\left(\frac{K}{\sqrt{n}}\right) + \sqrt{n}\left(\frac{u'u}{n}\right)^{1/2}\left(\frac{1}{n}\left\|\left(I - P_z\right) g\right\|^2\right)^{1/2} \\
&\leq O\left(\frac{K}{\sqrt{n}}\right) + O(1)\sqrt{n}\left\|g_K^r\right\|_{0,\infty,Z} \\
&= o_p(1).
\end{aligned}$$

□

Lemma C.1.4. $\frac{1}{\sqrt{n}} \sum \left(\widetilde{g}_i - g_i\right)^2 = o_p(1)$.

Proof.

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum \left(\widetilde{g}_i - g_i\right)^2 &= \sqrt{n}\frac{1}{n}\left\|\widetilde{g} - g\right\|^2 \\
&= \sqrt{n}\left\|g_K^r\right\|_{0,\infty,Z}^2 \\
&= o_p(1).
\end{aligned}$$

□

C.2 Proof of Proposition 4.2.1

We need to show

$$\sqrt{n}\left(\widehat{\gamma} - \gamma\right) = A^{-1}\left(B_1^n + B_2^n\right),$$

where

$$
\begin{aligned}
B_1^n &= \frac{1}{\sqrt{n}} \sum_i w_i u_i \\
B_2^n &= \frac{\alpha}{\sqrt{n}} \sum_i w_i \left(\widetilde{g}_i - g_i\right).
\end{aligned}
$$

From (4.3) we have

$$
\begin{aligned}
\sqrt{n}\left(\widehat{\gamma} - \gamma\right) &= \sqrt{n}\left(w^{\dagger\prime} w^{\dagger}\right)^{-1} w^{\dagger\prime}\left[\left(g - \widetilde{g}\right)\alpha + u\right] \\
&= \left(\frac{w^{\dagger\prime} w^{\dagger}}{n}\right)^{-1}\left[\frac{w^{\dagger\prime}\left(g - \widetilde{g}\right)\alpha + w^{\dagger\prime} u}{\sqrt{n}}\right].
\end{aligned}
$$

It is sufficient to show the followings.

$$
\frac{1}{n}\sum w_i^{\dagger} w_i^{\dagger\prime} - \frac{1}{n}\sum w_i w_i' = o_p(1) \tag{C.1}
$$

$$
\frac{1}{\sqrt{n}}\sum \left(w_i^{\dagger} - w_i\right) u_i = o_p(1) \tag{C.2}
$$

$$
\frac{1}{\sqrt{n}}\sum \left(w_i^{\dagger} - w_i\right)\left(\widetilde{g}_i - g_i\right) = o_p(1) \tag{C.3}
$$

Showing (C.1) is equivalent to show that

$$
\frac{1}{n}\sum \widetilde{g}_i^2 - g_i^2 = o_p(1) \tag{C.4}
$$

and

$$
\frac{1}{n}\sum x_i\left(\widetilde{g}_i - g_i\right) = o_p(1). \tag{C.5}
$$

To show (C.2) we need to prove

$$
\frac{1}{\sqrt{n}}\sum \left(\widetilde{g}_i - g_i\right) u_i = o_p(1). \tag{C.6}
$$

One could claim that showing (C.3) is identical to show

$$\frac{1}{\sqrt{n}}\sum\left(\widetilde{g}_i-g_i\right)^2=o_p\left(1\right). \tag{C.7}$$

(C.4) to (C.7) could be shown by Lemma C.1.1 to C.1.4. Now deal with the variance term of B_1^n and B_2^n and the covariance term between B_1^n and B_2^n.

$$Var\left[B_1^n\right]=\frac{1}{n}\sum_i w_i'w_iE\left[u_i^2|x_i,z_i\right]$$

$$\begin{aligned}Var\left[B_2^n\right] &= Var\left[\frac{\alpha}{\sqrt{n}}\sum_i w_i\left(\widetilde{g}_i-g_i\right)\right]\\ &= \frac{\alpha^2}{n}\sum_i\left(e\left(z_i\right)'g_i\right)'\left(e\left(z_i\right)'g_i\right)E\left[\epsilon_i^2|z_i\right]\end{aligned}$$

$$Cov\left[B_1^n,B_2^n\right]=\frac{\alpha}{n}\sum_i E\left[\epsilon_iu_iw_i'e\left(z_i\right)\right].$$

The asymptotic normality result of $\sqrt{n}\left(\widehat{\gamma}-\gamma\right)$ could be easily combined. □

C.3 Lemmata

Before proceeding the proof of Proposition 4.3.1, we need the following lemmata. Here are some notations.

$$\begin{aligned}s &= E\left[s|z\right]+\epsilon=g\left(z\right)+\epsilon\\ x &= E\left[x|z\right]+v=e\left(z\right)+v\\ g &= E\left[g\left(z\right)|x\right]+\eta=h\left(x\right)+\eta\\ \eta &= E\left[\eta|z\right]+\omega=\psi\left(z\right)+\omega\\ E\left[\epsilon|z\right] &= 0,\quad E\left[v|z\right]=0,\quad E\left[\eta|x\right]=0,\quad E\left[\omega|z\right]=0\end{aligned}$$

Lemma C.3.1. $\widetilde{g}'\left(I-P_x\right)\theta/\sqrt{n}-g'\left(I-P_x\right)\theta/\sqrt{n}=o_p\left(1\right)$

Proof.

$$\begin{aligned}\frac{1}{\sqrt{n}}\left[\widetilde{g}'\left(I-P_x\right)\theta-g'\left(I-P_x\right)\theta\right] &= \frac{1}{\sqrt{n}}\left[\left(\widetilde{g}-g\right)'\left(I-P_x\right)\theta\right] \\ &\leq \sqrt{n}\left\|\frac{\widetilde{g}-g}{\sqrt{n}}\right\|\left\|\frac{\left(I-P_x\right)\theta}{\sqrt{n}}\right\| \\ &\leq \sqrt{n}\left\|g_K^r\right\|_{0,\infty,Z}\left\|\theta_K^r\right\|_{0,\infty,X}\to 0\end{aligned}$$

□

Lemma C.3.2. $g'\left(I-P_x\right)\theta/\sqrt{n}=o_p\left(1\right)$

Proof.

$$\begin{aligned}&\frac{1}{\sqrt{n}}g'\left(I-P_x\right)\theta \\ \leq\; & \sqrt{n}\left(\frac{g'g}{n}\right)^{1/2}\left\|\frac{\left(I-P_x\right)\theta}{\sqrt{n}}\right\| \\ =\; & O_p\left(1\right)\sqrt{n}\left\|\theta_K^r\right\|_{0,\infty,X}\to 0\end{aligned}$$

□

Lemma C.3.3. $\widetilde{g}'\left(I-P_x\right)\left(\widetilde{g}-g\right)/\sqrt{n}-g'\left(I-P_x\right)\left(\widetilde{g}-g\right)/\sqrt{n}=o_p\left(1\right)$

Proof.

$$\begin{aligned}\frac{1}{\sqrt{n}}\left[\widetilde{g}'\left(I-P_x\right)\left(\widetilde{g}-g\right)-g'\left(I-P_x\right)\left(\widetilde{g}-g\right)\right] &= \frac{1}{\sqrt{n}}\left[\left(\widetilde{g}-g\right)'\left(I-P_x\right)\left(\widetilde{g}-g\right)\right] \\ &\leq \sqrt{n}\left\|\frac{\widetilde{g}-g}{\sqrt{n}}\right\|^2 \\ &\leq \sqrt{n}\left\|g_K^r\right\|_{0,\infty,Z}^2\to 0\end{aligned}$$

□

Lemma C.3.4. $g'(I-P_x)(P_z g - g)/\sqrt{n} = o_p(1)$

Proof.

$$\begin{aligned}
\left|\frac{1}{\sqrt{n}} g'(I-P_x)(P_z g - g)\right| &\leq \sqrt{n}\left(\frac{g'(I-P_x)g}{n}\right)^{1/2}\left(\frac{g'(I-P_z)g}{n}\right)^{1/2} \\
&\leq \sqrt{n}\left(\frac{g'(I-P_x)g}{n}\right)^{1/2}\|g_K^r\|_{0,\infty,Z} \to 0
\end{aligned}$$

□

Lemma C.3.5. $h'(I-P_x)(P_z\epsilon)/\sqrt{n} = o_p(1)$

Proof.

$$\begin{aligned}
\left|\frac{1}{\sqrt{n}} h'(I-P_x)(P_z\epsilon)\right| &\leq \sqrt{n}\left(\frac{h'(I-P_x)h}{n}\right)^{1/2}\left(\frac{\epsilon' P_z\epsilon}{n}\right)^{1/2} \\
&\leq \sqrt{n}\,\|h_K^r\|_{0,\infty,X}\, O_p\left(\frac{\sqrt{K_z}}{\sqrt{n}}\right) \to 0
\end{aligned}$$

□

Lemma C.3.6. $\eta' P_x P_z \epsilon/\sqrt{n} = o_p(1)$

Proof.

$$\begin{aligned}
\left|\frac{1}{\sqrt{n}}\eta' P_x P_z\epsilon\right| &\leq \sqrt{n}\left(\frac{\eta' P_x\eta}{n}\right)^{1/2}\left(\frac{\epsilon' P_z\epsilon}{n}\right)^{1/2} \\
&\leq \sqrt{n} O_p\left(\frac{\sqrt{K_x}}{\sqrt{n}}\frac{\sqrt{K_z}}{\sqrt{n}}\right) \to 0
\end{aligned}$$

□

Lemma C.3.7. $\omega' P_z\epsilon/\sqrt{n} = o_p(1)$

Proof.

$$\begin{aligned}\left|\frac{1}{\sqrt{n}}\omega' P_z \epsilon\right| &\leq \sqrt{n}\left(\frac{\omega' P_z \omega}{n}\right)^{1/2}\left(\frac{\epsilon' P_z \epsilon}{n}\right)^{1/2} \\ &\leq \sqrt{n} O_p\left(\frac{\sqrt{K_z}}{\sqrt{n}}\frac{\sqrt{K_z}}{\sqrt{n}}\right) \to 0\end{aligned}$$

□

Lemma C.3.8. $\psi' P_z \epsilon/\sqrt{n} \xrightarrow{d} N\left(0, \psi' E\left[\epsilon\epsilon'|z\right]\psi\right)$

Proof. First we can write $\psi' P_z \epsilon/\sqrt{n}$ as $\psi'\epsilon/\sqrt{n} - \psi'\left(I - P_z\right)\epsilon/\sqrt{n}$. It is easy to verify that $\psi'\left(I - P_z\right)\epsilon/\sqrt{n} = o_p\left(1\right)$. Applying the Liapunov's Central Limit Theorem on $\psi'\epsilon/\sqrt{n}$ completes the proof. □

Lemma C.3.9. $g'\left(I - P_z\right)\left(I - P_x\right)u/\sqrt{n} = o_p\left(1\right)$

Proof.

$$\begin{aligned}\left|\frac{g'\left(I - P_z\right)\left(I - P_x\right)u}{\sqrt{n}}\right| &\leq \sqrt{n}\left(\frac{\left(\widetilde{g} - g\right)'\left(I - P_x\right)\left(\widetilde{g} - g\right)}{n}\right)^{1/2}\left(\frac{u'u}{n}\right)^{1/2} \\ &\leq \sqrt{n}\left\|g_K^r\right\|_{0,\infty,Z} \cdot O_p\left(1\right) \to 0\end{aligned}$$

□

Lemma C.3.10. $\epsilon' P_z\left(I - P_x\right)u/\sqrt{n} = o_p\left(1\right)$

Proof.

$$\begin{aligned}\frac{\epsilon' P_z\left(I - P_x\right)u}{\sqrt{n}} &= \frac{\epsilon' P_z u}{\sqrt{n}} - \frac{\epsilon' P_z P_x u}{\sqrt{n}} \\ &= O_p\left(\frac{K_z}{\sqrt{n}}\right) + O_p\left(\frac{\sqrt{K_z}\sqrt{K_x}}{\sqrt{n}}\right) \\ &= o_p\left(1\right)\end{aligned}$$

□

Lemma C.3.11. $g'(I-P_x)u/\sqrt{n} \xrightarrow{d} N\left(0, E\left[(g-h)' E\left[uu'|x,z\right](g-h)\right]\right)$

Proof. Applying Liapunov's Central Limit Theorem gives the result. □

C.4 Proof of Proposition 4.3.1

One could write $\sqrt{n}\,(\widehat{\alpha}-\alpha)$ as

$$\begin{aligned}\sqrt{n}\,(\widehat{\alpha}-\alpha) &= \left[\frac{\widetilde{g}'(I-P_x)\widetilde{g}}{n}\right]^{-1}\left[\frac{\widetilde{g}'(I-P_x)(\theta+(g-\widetilde{g})\alpha+u)}{\sqrt{n}}\right] \\ &= \left[\frac{s'P_z(I-P_x)P_z s}{n}\right]^{-1}\left[\frac{\widetilde{g}'(I-P_x)[\theta+(g-P_z s)\alpha+u]}{\sqrt{n}}\right].\end{aligned}$$

Define

$$u^* = \theta + (g-P_z s)\alpha + u.$$

The limiting distribution of $\sqrt{n}\,(\widehat{\alpha}-\alpha)$ depends on the new error structure u^*, which contains the original error term u, approximation error $g-P_z s$ and θ. We can show later that θ is not important in building the asymptotic result. By Lemma C.3.1 to C.3.2, we can ignore the term $\widetilde{g}'(I-P_x)\theta/\sqrt{n}$. By Lemma C.3.3, it reduces to deal with the term

$$g'(I-P_x)(\widetilde{g}-g)/\sqrt{n} = g'(I-P_x)(P_z g-g)/\sqrt{n} + g'(I-P_x)(P_z\epsilon)/\sqrt{n}.$$

By Lemma C.3.4, $g'(I-P_x)(P_z g-g)/\sqrt{n}$ is small order. Now we could decompose g as $g = h(x)+\eta$ and η as $\psi(z)+\omega$. By Lemma C.3.5-C.3.8, $g'(I-P_x)(P_z\epsilon)/\sqrt{n}$ will converge to normal distribution with variance $\psi' E[\epsilon\epsilon'|z]\,\psi$.

The last step is to decompose the term $\widetilde{g}'(I-P_x)u/\sqrt{n}$ as follows.

$$\frac{\widetilde{g}'(I-P_x)u}{\sqrt{n}} = \frac{g'(I-P_x)u}{\sqrt{n}} - \frac{g'(I-P_z)(I-P_x)u}{\sqrt{n}} + \frac{\epsilon' P_z(I-P_x)u}{\sqrt{n}}. \quad \text{(C.8)}$$

Lemma C.3.9 and C.3.10 verify the second and third terms in (C.8) are small orders. Lemma C.3.11 gives the asymptotic normality of the first term in (C.8). Combining the Lemma C.3.1 to C.3.11 proves the Proposition 4.3.1. □

Appendix D

Tables-Chapter2

		$\widehat{\beta}_{OLS}^{True}$	$\widehat{\beta}_{OLS}^{B}$	$\widehat{\beta}_{OLS}^{W}$	$\widehat{\beta}_{OLS}^{MW}$
n	Nominal size		Empirical size		
50	.100	.104	.560	.289	.150
	.075	.080	.529	.244	.114
	.050	.052	.496	.202	.093
	.025	.020	.453	.150	.064
	.010	.009	.382	.103	.037
100	.100	.107	.640	.243	.143
	.075	.078	.606	.214	.108
	.050	.047	.573	.171	.076
	.025	.019	.516	.113	.044
	.010	.004	.449	.066	.026
150	.100	.107	.650	.234	.144
	.075	.083	.615	.192	.109
	.050	.054	.588	.146	.080
	.025	.022	.544	.106	.061
	.010	.007	.503	.071	.033

Note: $x \sim \log normal, u \sim normal, \sigma_i^2 \sim abs$

Table D.1: Nominal and Empirical Sizes for OLS estimators

		$\widehat{\beta}_{FGLS}^{True}$	$\widehat{\beta}_{FGLS}^{S}$	$\widehat{\beta}_{FGLS}^{K}$
n	Nominal size	Empirical size		
50	.100	.101	.235	.239
	.075	.076	.191	.205
	.050	.053	.157	.166
	.025	.022	.114	.121
	.010	.010	.062	.077
100	.100	.109	.227	.249
	.075	.087	.194	.209
	.050	.065	.169	.169
	.025	.036	.112	.119
	.010	.014	.066	.079
150	.100	.102	.221	.233
	.075	.078	.186	.193
	.050	.055	.150	.155
	.025	.029	.099	.102
	.010	.008	.052	.056

Note: x ~lognormal, $u \sim normal, \sigma_i^2 \sim abs$

Table D.2: Nominal and Empirical Sizes for FGLS estimators

		$\widehat{\beta}_{Cragg}^{True}$	$\widehat{\beta}_{Cragg}^{W}$	$\widehat{\beta}_{Cragg}^{MW}$
n	Nominal size	Empirical size		
50	.100	.093	.156	.101
	.075	.069	.116	.076
	.050	.044	.079	.054
	.025	.026	.057	.038
	.010	.009	.031	.022
100	.100	.099	.149	.120
	.075	.077	.118	.088
	.050	.049	.079	.059
	.025	.024	.048	.037
	.010	.008	.026	.017
150	.100	.105	.131	.110
	.075	.078	.108	.085
	.050	.046	.076	.061
	.025	.022	.035	.023
	.010	.006	.016	.013

Note: $x \sim \log normal, u \sim normal, \sigma_i^2 \sim abs$

Table D.3: Nominal and Empirical Sizes for Cragg estimators

	$\widehat{\beta}_{FGLS}^{True}$		$\widehat{\beta}_{FGLS}^{S}$		$\widehat{\beta}_{FGLS}^{K}$
$n = 50$	.1110	$K = 1$	.2659	$h = .49$	.1552
		$K = 2$	.1445	$h = .59$	.1534
		$K = 3$	.1443	$h = .69$	.1521
		$K = 4$	.1524	$h = .79$	.1511
		$K = 5$	.1757	$h = .89$	.1505
$n = 100$	.0817	$K = 1$	.2251	$h = .49$	.1046
		$K = 2$	.1026	$h = .59$	.1036
		$K = 3$	.1012	$h = .69$	.1030
		$K = 4$	.1115	$h = .79$	.1023
		$K = 5$	.1274	$h = .89$	.1022
$n = 150$	.0637	$K = 1$	.2108	$h = .49$	.0888
		$K = 2$	.0855	$h = .59$	.0879
		$K = 3$	.0845	$h = .69$	.0873
		$K = 4$	.0924	$h = .79$	.0871
		$K = 5$	.1059	$h = .89$	.0869

Note: $x \sim$lognormal, $u \sim normal, \sigma_i^2 \sim abs$

Table D.4: RMSEs for FGLS estimators

		$\widehat{\beta}_{Cragg}^{True}$	$\widehat{\beta}_{Cragg}^{W}$	$\widehat{\beta}_{Cragg}^{MW}$
$n = 50$	$K = 1$	.2659	.2659	.2659
	$K = 2$	.1174	.1421	.1267
	$K = 3$	.1133	.1433	.1247
	$K = 4$	.1113	.1514	.1338
	$K = 5$	.1112	.1608	.1427
$n = 100$	$K = 1$	.1053	.1053	.1053
	$K = 2$	.0888	.0973	.0921
	$K = 3$	.0843	.0939	.0874
	$K = 4$	.0828	.0982	.0910
	$K = 5$	.0819	.1037	.0957
$n = 150$	$K = 1$	.0886	.0886	.0886
	$K = 2$	.0717	.0793	.0748
	$K = 3$	.0666	.0771	.0713
	$K = 4$	.0651	.0804	.0735
	$K = 5$	.0639	.0840	.0768

Note: $x \sim \log normal, u \sim normal, \sigma_i^2 \sim abs$

Table D.5: RMSEs for Cragg estimators

	Model 1		Model 2		Model 3	
	Cragg	FGLS	Cragg	FGLS	Cragg	FGLS
$K = 1$	192.37	1.91	193.52	3.05	183.23	-7.24
$K = 2$	7.23	7.23	11.58	11.58	-27.49	-27.49
$K = 3$	14.16	14.16	22.66	22.66	-53.81	-53.81
$K = 4$	20.62	20.62	33.00	33.00	-78.37	-78.37
$K = 5$	26.89	26.89	43.02	43.02	-102.18	-102.18
$K = 6$	33.48	33.48	53.57	53.57	-127.24	-127.24
$K = 7$	39.60	39.60	63.36	63.36	-150.48	-150.48
$K = 8$	45.48	45.48	72.76	72.76	-172.81	-172.81
$K = 9$	51.72	51.72	82.75	82.75	-196.52	-196.52
$K = 10$	57.70	57.70	92.32	92.32	-219.25	-219.25

Model 1: $x \sim normal, u \sim normal, \sigma_i^2 \sim$constant
Model 2: $x \sim normal, u \sim uniform, \sigma_i^2 \sim$constant
Model 3: $x \sim normal, u \sim$logistic, $\sigma_i^2 \sim$constant

Table D.6: Theoretical RMSE of Cragg and FGLS estimators under Homoskedasticity

	Model 1		Model 2		Model 3	
	Theory	Criteria	Theory	Criteria	Theory	Criteria
$K = 1$	50.74	46.76	51.07	6.52	48.08	35.42
$K = 2$	2.71*	12.79*	3.58*	3.88*	-4.21	20.89
$K = 3$	2.91	13.36	4.19	5.93	-7.33	20.40
$K = 4$	2.86	13.98	4.56	7.93	-10.80	19.80
$K = 5$	3.55	14.73	5.68	9.91	-13.48	19.14
$K = 6$	4.25	15.48	6.80	11.88	-16.15	18.33
$K = 7$	4.95	16.25	7.92	13.85	-18.82	17.43
$K = 8$	5.65	17.05	9.04	15.82	-21.48	16.50
$K = 9$	6.35	17.86	10.16	17.77	-24.13	15.58
$K = 10$	7.05	18.69	11.27	19.72	-26.78*	14.71*

Model 1: $x \sim$uniform$, u \sim$normal$, \sigma_i^2 \sim \exp$
Model 2: $x \sim$uniform$, u \sim$uniform$, \sigma_i^2 \sim \exp$
Model 3: $x \sim$uniform$, u \sim$logistic$, \sigma_i^2 \sim \exp$

Table D.7: RMSE of Cragg estimator under errors with different kurtosis

	Model 1		Model 2		Model 3	
	Theory	Criteria	Theory	Criteria	Theory	Criteria
$K = 1$	2.10	4.31	2.42	7.83*	-.41	4.05
$K = 2$	2.65	3.19	3.50	11.24	-4.11	1.02
$K = 3$	2.06*	1.94*	3.31*	10.91	-7.99	-.56
$K = 4$	2.75	2.59	4.44	14.68	-10.74	-.90
$K = 5$	3.45	3.24	5.55	18.43	-13.42	-1.27
$K = 6$	4.15	3.89	6.68	22.14	-16.10	-1.68
$K = 7$	4.85	4.53	7.80	25.86	-18.77	-2.10
$K = 8$	5.55	5.18	8.92	29.56	-21.46	-2.61
$K = 9$	6.25	5.83	10.04	33.25	-24.08	-3.08
$K = 10$	6.95	6.47	11.16	36.93	-26.73*	-3.57*

Model 1: $x \sim uniform, u \sim normal, \sigma_i^2 \sim \exp$
Model 2: $x \sim uniform, u \sim uniform, \sigma_i^2 \sim \exp$
Model 3: $x \sim uniform, u \sim$logistic$, \sigma_i^2 \sim \exp$

Table D.8: RMSE of series based FGLS estimator under errors with different kurtosis

	Model 1		Model 2		Model 3	
	Theory	Criteria	Theory	Criteria	Theory	Criteria
$K=1$	15.68	22.98	13.86	1.52	30.22*	298.18
$K=2$	14.62	14.43	10.43	1.18	48.16	228.45
$K=3$	10.92*	7.96*	4.46*	.29*	62.75	142.51*
$K=4$	14.57	10.56	5.85	.38	84.32	189.39
$K=5$	18.24	13.16	7.30	.47	105.77	236.41
$K=6$	21.92	15.78	8.77	.56	127.13	283.56
$K=7$	25.60	18.40	10.24	.65	148.45	330.71
$K=8$	29.27	21.02	11.71	.74	169.75	377.88
$K=9$	32.93	23.65	13.17	.84	190.98	425.10
$K=10$	36.58	26.28	14.63	.93	212.17	472.36

Model 1: $x \sim uniform, u \sim normal, \sigma_i^2 \sim \exp$
Model 2: $x \sim uniform, u \sim uniform, \sigma_i^2 \sim \exp$
Model 3: $x \sim uniform, u \sim$logistic, $\sigma_i^2 \sim \exp$

Table D.9: RMSE of series based estimator in variance regression under errors with different kurtosis

White test	P-value	Breusch-Pagan test	P-value
23.39, Chi-sq(8)	.0029	23.91, Chi-sq(2)	6.4e-06

Table D.10: Heteroskedasticity tests

# of K	Cragg	FGLS	Variance	Variance-CV
$K=1$	-6733378	2777861.4	.0740	1.8784
$K=2$	-7953344	1656619.7	.0756	1.8814
$K=3$	-14383004	7140911.5	.0444	1.8532
$K=4$	-16700076	789443.95*	.0141*	1.8239
$K=5$	-16541297	946018.65	.0177	1.8277
$K=6$	-17643387	1311052.2	.0213	1.8235*
$K=7$	-17905310*	1598746.6	.0249	1.8272
$K=8$	-17830572	1748223.5	.0292	1.8310
$K=9$	-17717161	1977208.3	.0329	1.8360
$K=10$	-17586057	2184831.6	.0366	1.8413

Table D.11: Optimal K using criteria introduced in chapter 2

	Constant	Education	Experience	Experience^2
OLS	0.12800	0.09037	0.04101	-0.00071
t-ratio (wrong s.e.)	1.20830	12.10041	7.89161	-6.16389
	(0.10593)	(0.00747)	(0.00520)	(0.00012)
t-ratio (White s.e.)	1.19940	11.65556	8.19434	-6.52501
	(0.10672)	(0.00775)	(0.00500)	(0.00011)
t-ratio (M-W s.e.)	1.18042	11.45501	8.11618	-6.44709
	(0.10843)	(0.00789)	(0.00505)	(0.00011)
FGLS-Series	0.12422	0.08822	0.04402	-0.00076
t-ratio	1.31631	12.53150	9.21773	-6.99317
	(0.09437)	(0.00704)	(0.00478)	(0.00011)
Cragg-White	0.05836	0.09426	0.04329	-0.00075
t-ratio	0.55395	12.23446	8.97863	-7.17348
	(0.10534)	(0.00770)	(0.00482)	(0.00010)
Cragg-MW	0.04661	0.09513	0.04331	-0.00075
t-ratio	0.43643	12.16447	8.88793	-7.06698
	(0.10679)	(0.00782)	(0.00487)	(0.00011)

Table D.12: Estimation Results of Wage Equation

	Decomposition of $\widehat{h}$	Order
h	$\frac{1}{\sqrt{n}} X^{*\prime} \epsilon^{*}$	$O(1)$
T_1^h	$-\frac{1}{\sqrt{n}} \sum \frac{x_i^* \epsilon_i^*}{\sigma_i^2} \left[\overline{\sigma}_i^2 - \sigma_i^2 \right]$	$O\left(\left\| \sigma^2 \right\| \right)$
T_2^h	$-\frac{1}{\sqrt{n}} \sum \frac{x_i^* \epsilon_i^*}{\sigma_i^2} \left[q^{K\prime}(x_i) (Q'Q)^{-1} Q' \mathbf{v} \right]$	$O\left(\frac{\sqrt{K}}{\sqrt{n}} \right)$
T_3^h	$2 \left(\hat{\beta} - \beta \right) \frac{1}{\sqrt{n}} \sum \frac{x_i^* \epsilon_i^*}{\sigma_i^2} \left[q^{K\prime}(x_i) (Q'Q)^{-1} Q' \mathbf{x} \epsilon \right]$	$O\left(\frac{K}{n} \right)$
T_4^h	$\frac{1}{\sqrt{n}} \sum \frac{x_i^* \epsilon_i^*}{\sigma_i^4} \left[\overline{\sigma}_i^2 - \sigma_i^2 \right]^2$	$O\left(\zeta(K) \left\| \sigma^2 \right\|^2 \right)$
T_5^h	$\frac{1}{\sqrt{n}} \sum \frac{x_i^* \epsilon_i^*}{\sigma_i^4} \left[q^{K\prime}(x_i) (Q'Q)^{-1} Q' \mathbf{v} \right]^2$	$O\left(\frac{\zeta(K) K \sqrt{K}}{n} \right)$

	Decomposition of $\widehat{H}$	Order
H	$\frac{1}{n} X^{*\prime} X^{*}$	$O(1)$
T_1^H	$-\frac{1}{n} \sum_i \frac{x_i^{*2}}{\sigma_i^2} \left(\overline{\sigma}_i^2 - \sigma_i^2 \right)$	$O\left(\left\| f \right\| \left\| \sigma^2 \right\| \right)$
T_2^H	$-\frac{1}{n} \sum_i \frac{x_i^{*2}}{\sigma_i^2} v_i$	$O\left(\frac{1}{\sqrt{n}} \right)$
T_3^H	$\frac{1}{n} \sum \frac{x_i^{*2}}{\sigma_i^4} \left[\overline{\sigma}_i^2 - \sigma_i^2 \right]^2$	$O\left(\left\| \sigma^2 \right\|^2 \right)$
T_4^H	$\frac{1}{n} \sum \frac{x_i^{*2}}{\sigma_i^4} \left[q^{K\prime}(x_i) (Q'Q)^{-1} Q' \mathbf{v} \right]^2$	$O\left(\frac{K}{n} \right)$

Table D.13: Decomposition of $\widehat{H}$ and $\widehat{h}$

Appendix E

Tables-Chapter3

CASES[a]	$TP = .1$ BIAS[b]	MSE	$TP = .01$ BIAS	MSE	$TP = .001$ BIAS	MSE
$n = 100$						
b_{Lin}	0.2410	**0.0921**	0.1714	0.0471	0.1922	0.0601
b_{Li}	**0.2409**	**0.0921**	**0.1713**	**0.0469**	**0.1899**	**0.0584**
b_0	0.2704	0.1180	0.2704	0.1180	0.2704	0.1180
b_K	0.2584	0.1057	0.2259	0.0754	0.2360	0.0814
$n = 200$						
b_{Lin}	0.1683	0.0443	0.1124	**0.0201**	0.1244	0.0253
b_{Li}	**0.1682**	**0.0442**	**0.1123**	**0.0201**	**0.1240**	**0.0251**
b_0	0.1902	0.0573	0.1902	0.0573	0.1902	0.0573
b_K	0.1796	0.0503	0.1554	0.0351	0.1703	0.0407
$n = 400$						
b_{Lin}	**0.1202**	**0.0223**	**0.0824**	**0.0109**	**0.0929**	0.0135
b_{Li}	**0.1202**	**0.0223**	0.0825	**0.0109**	**0.0929**	**0.0134**
b_0	0.1369	0.0286	0.1369	0.0286	0.1369	0.0286
b_K	0.1306	0.0256	0.1080	0.0174	0.1245	0.0221

[a] b_{Lin}, b_{Li}, b_0 and b_K represent my estimator, Li's (2000) estimator, preliminary estimator and kernel estimator respectively.
[b] Note that the number in boldface is the minimum of the corresponding BIAS or MSE

Table E.1: Simulation results of partial regression model assuming $g(z) = \exp(z)$

CASES[a]	$TP = .1$ BIAS[b]	MSE	$TP = .01$ BIAS	MSE	$TP = .001$ BIAS	MSE
$n = 100$						
b_{Lin}	0.2416	0.0923	0.1731	0.0479	0.1944	0.0615
b_{Li}	**0.2415**	**0.0922**	**0.1719**	**0.0472**	**0.1889**	**0.0578**
b_0	0.2712	0.1182	0.2712	0.1182	0.2712	0.1182
b_K	0.4668	0.2866	0.4753	0.2836	0.4753	0.2836
$n = 200$						
b_{Lin}	0.1687	**0.0444**	0.1135	0.0205	0.1277	0.0267
b_{Li}	**0.1686**	**0.0444**	**0.1128**	**0.0203**	**0.1234**	**0.0248**
b_0	0.1909	0.0575	0.1909	0.0575	0.1909	0.0575
b_K	0.3080	0.1287	0.3141	0.1226	0.3141	0.1226
$n = 400$						
b_{Lin}	**0.1203**	**0.0223**	0.0829	0.0110	0.0934	0.0138
b_{Li}	**0.1203**	**0.0223**	**0.0827**	**0.0109**	**0.0921**	**0.0133**
b_0	0.1371	0.0285	0.1371	0.0285	0.1371	0.0285
b_K	0.2142	0.0617	0.2140	0.0570	0.2140	0.0570

[a] b_{Lin}, b_{Li}, b_0 and b_K represent my estimator, Li's (2000) estimator, preliminary estimator and kernel estimator respectively.
[b] Note that the number in boldface is the minimum of the corresponding BIAS or MSE.

Table E.2: Simulation results of partial regression model assuming $g(z) = (1 + z)^3$

CASES	$n = 100$ True MSE[a]	$n = 100$ Bootstrap MSE	$n = 200$ True MSE	$n = 200$ Bootstrap MSE	$n = 400$ True MSE	$n = 400$ Bootstrap MSE
K_{11}	0.1062	0.1439	0.0519	0.0624	0.0227	0.0265
K_{21}	**0.0559**	0.1354	0.0297	0.0569	0.0140	0.0244
K_{31}	0.0569	**0.1326**	0.0296	**0.0539**	0.0141	**0.0238**
K_{12}	0.0949	0.1617	0.0413	0.0712	0.0172	0.0316
K_{22}	0.0560	0.1578	0.0288	0.0675	**0.0121**	0.0288
K_{32}	0.0561	0.1515	**0.0284**	0.0637	0.0123	0.0278
K_{13}	0.1043	0.2009	0.0481	0.0874	0.0208	0.0392
K_{23}	0.0639	0.1778	0.0313	0.0745	0.0142	0.0319
K_{33}	0.0643	0.1761	0.0313	0.0715	0.0143	0.0314

[a] Note that the number in boldface is the minimum of the corresponding MSE.

Table E.3: Choosing smoothing parameters under partial linear regression model assuming $g(z) = \exp(z)$

Appendix F

Tables-Chapter4

	$\rho = .1$		$\rho = .5$		$\rho = .9$	
CASE	BIAS	MSE	BIAS	MSE	BIAS	MSE
$n = 100$						
True	0.0945	0.0144	0.0447	0.0032	0.0297	0.0014
True_series	0.1237	0.0246	0.1315	0.0272	0.2800	0.1227
True_Kernel	0.1294	0.0268	0.1423	0.0317	0.4036	0.2404
Linear_series	0.1382	0.0319	0.1309	0.0276	0.2818	0.1273
Linear_Kernel	0.1446	0.0348	0.1401	0.0329	0.4002	0.2498
2-step series	0.1216	0.0238	0.1274	0.0253	0.2646	0.1110
Double Kernel	0.1654	0.0440	0.1845	0.0530	0.4398	0.2784
$n = 200$						
True	0.0752	0.0092	0.0337	0.0018	0.0215	0.0007
True_series	0.0957	0.0147	0.0967	0.0150	0.1969	0.0625
True_Kernel	0.0993	0.0158	0.1051	0.0176	0.2944	0.1267
Linear_series	0.1042	0.0182	0.0967	0.0153	0.2004	0.0648
Linear_Kernel	0.1074	0.0196	0.1036	0.0179	0.2921	0.1306
2-step series	0.0916	0.0134	0.0938	0.0141	0.1919	0.0587
Double Kernel	0.1343	0.0284	0.1524	0.0348	0.3508	0.1685
$n = 400$						
True	0.0592	0.0055	0.0260	0.0011	0.0162	0.0004
True_series	0.0741	0.0088	0.0746	0.0086	0.1470	0.0333
True_Kernel	0.0762	0.0092	0.0786	0.0097	0.2141	0.0669
Linear_series	0.0827	0.0111	0.0761	0.0091	0.1494	0.0343
Linear_Kernel	0.0849	0.0118	0.0805	0.0103	0.2130	0.0677
2-step series	0.0719	0.0083	0.0747	0.0086	0.1444	0.0327
Double Kernel	0.1099	0.0181	0.1241	0.0221	0.2828	0.1060

Table F.1: Simulation results of semiparametric regression model with generated regressors assuming $g(z) = \left[(z_1 + z_2)^2 + z_2\right]$

	$\rho = .1$		$\rho = .5$		$\rho = .9$	
CASE	BIAS	MSE	BIAS	MSE	BIAS	MSE
$n = 100$						
True	0.0214	0.0008	0.0081	0.0001	0.0036	0.0000
True_series	0.0307	0.0019	0.0134	0.0003	0.0116	0.0002
True_Kernel	0.0322	0.0021	0.0146	0.0004	0.0164	0.0004
Linear_series	0.0513	0.0045	0.0448	0.0033	0.0989	0.0156
Linear_Kernel	0.0586	0.0058	0.0485	0.0039	0.1058	0.0177
2-step series	0.0352	0.0023	0.0190	0.0006	0.0456	0.0034
Double Kernel	0.2255	0.0652	0.1985	0.0458	0.2722	0.0806
$n = 200$						
True	0.0181	0.0006	0.0068	0.0001	0.0028	0.0000
True_series	0.0254	0.0012	0.0110	0.0002	0.0087	0.0001
True_Kernel	0.0263	0.0013	0.0119	0.0003	0.0133	0.0003
Linear_series	0.0391	0.0026	0.0310	0.0016	0.0726	0.0085
Linear_Kernel	0.0428	0.0031	0.0338	0.0019	0.0818	0.0105
2-step series	0.0281	0.0014	0.0151	0.0004	0.0404	0.0025
Double Kernel	0.1887	0.0440	0.1743	0.0337	0.2469	0.0636
$n = 400$						
True	0.0151	0.0004	0.0059	0.0001	0.0023	0.0000
True_series	0.0224	0.0009	0.0100	0.0002	0.0074	0.0001
True_Kernel	0.0230	0.0009	0.0106	0.0002	0.0106	0.0002
Linear_series	0.0347	0.0020	0.0229	0.0009	0.0567	0.0049
Linear_Kernel	0.0364	0.0023	0.0246	0.0010	0.0677	0.0066
2-step series	0.0249	0.0011	0.0129	0.0003	0.0388	0.0020
Double Kernel	0.1573	0.0296	0.1505	0.0242	0.2203	0.0498

Table F.2: Simulation results of semiparametric regression model with generated regressors assuming $g(z) = \exp\left[(z_1 + z_2)^2 + z_2\right]$

CASES	$\rho = .1$		$\rho = .5$		$\rho = .9$	
	True MSE	BT MSE	True MSE	BT MSE	True MSE	BT MSE
$n = 100$						
K_{11}	0.0351	0.0740	0.0282	0.0343	0.1249	0.1473
K_{12}	**0.0246**[a]	0.0272	**0.0253**	0.0284	0.1181	0.1178
K_{13}	0.0252	**0.0247**	0.0260	**0.0261**	0.1110	**0.0835**
K_{21}	0.0357	0.0754	0.0288	0.0357	0.1234	0.1499
K_{22}	0.0248	0.0281	0.0258	0.0291	**0.1105**	0.1215
K_{23}	0.0256	0.0254	0.0274	0.0267	0.1108	0.0843
K_{31}	0.0381	0.0829	0.0307	0.0390	0.1317	0.1624
K_{32}	0.0262	0.0307	0.0273	0.0318	0.1177	0.1317
K_{33}	0.0271	0.0276	0.0289	0.0290	0.1127	0.0998
$n = 200$						
K_{11}	0.0186	0.0234	0.0150	0.0173	0.0640	0.0701
K_{12}	0.0148	0.0160	0.0141	0.0160	0.0662	0.0689
K_{13}	0.0153	0.0154	0.0144	0.0155	0.0625	0.0567
K_{21}	0.0183	0.0234	0.0148	0.0172	0.0616	0.0686
K_{22}	**0.0146**	0.0158	**0.0141**	0.0156	**0.0591**	0.0615
K_{23}	0.0153	**0.0152**	0.0148	**0.0151**	0.0620	**0.0512**
K_{31}	0.0190	0.0248	0.0152	0.0180	0.0633	0.0709
K_{32}	0.0149	0.0165	0.0144	0.0162	0.0608	0.0636
K_{33}	0.0156	0.0158	0.0153	0.0156	0.0628	0.0558
$n = 400$						
K_{11}	0.0115	0.0124	0.0089	0.0093	0.0341	0.0355
K_{12}	0.0090	0.0094	0.0083	0.0089	0.0350	0.0395
K_{13}	0.0094	0.0093	**0.0082**	0.0089	0.0313	0.0370
K_{21}	0.0112	0.0122	0.0089	0.0092	0.0340	0.0345
K_{22}	**0.0088**	0.0092	0.0083	0.0086	**0.0311**	0.0327
K_{23}	0.0092	**0.0091**	0.0086	**0.0086**	0.0326	**0.0311**
K_{31}	0.0113	0.0125	0.0090	0.0094	0.0342	0.0350
K_{32}	0.0089	0.0094	0.0084	0.0088	0.0313	0.0332
K_{33}	0.0093	0.0092	0.0086	0.0087	0.0317	0.0320

[a] Note that the number in boldface is the minimum of the corresponding MSE.

Table F.3: Choosing smoothing parameters under semiparametric regression model with generated regressors

Bibliography

[1] Adams, R. A. (1975) *Sobolev Spaces*, New York, Academic Press.

[2] Ahn, H. (1995) "Nonparametric Two-Stage Estimation of Conditional Choice Probabilities in Binary Choice Model Under Uncertainty," *Journal of Econometrics* **67**, 337-378.

[3] Ahn, H. (1997) "Semiparametric Estimation of a Single-Index Model with Nonparametrically Generated Regressors," *Econometric Theory* **13**, 3-31.

[4] Amemiya, T. (1985) *Advanced Econometrics*, Cambridge, Mass., Harvard University Press.

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

[6] Andrews, D. (1994) "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity," *Econometrica*, **62**, 43–72.

[7] Barro, R. J. (1977) "Unanticipated Money Growth and Unemployment in the United States," *American Economic Review,* **67**, 101-115.

[8] Carrol, R. J. (1982) "Adapting for Heteroskedasticity in Linear Model," *Annals of Statistics*, **10**, 1224-1233.

[9] Chamberlain, G. (1987) "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, **34**, 305-334.

[10] Chamberlain, G. (1992) "Efficiency Bounds for Semiparametric Regression," *Econometrica*, **60**, 567-596.

[11] Chen Hung (1988) "Convergence Rates for Parametric Component in a Partly Linear Model," *Annal of Statistics*, **16**, 136-146.

[12] Chesher, A (1989) "Hajek Inequality, Measures of Leverage and the Size of Heteroskedasticity Robust Wald Tests," *Econometrica*, **57**, 871-977.

[13] Chesher, A. and G. Austin (1991) "The Finite-Sample Distributions of Heteroskedasticity Robust Wald Statistics," *Journal of Econometrics*, **47**, 153-174.

[14] Chesher, A. and I. Jewitt (1987) "The Bias of a Heteroskedasticity Consistent Covariance Matrix Estimator," *Econometrica*, **55**, 1217-1222.

[15] Cragg, J. G. (1983) "More Efficient Estimation in the presence of Heteroskedasticity of Unknown Form," *Econometrica*, **51**, 571-764.

[16] Cribari-Neto F., Ferrari, S. L. P., Cordeiro, G. M (2000) "Improved Heteroskedasticity-Consistent Covariance Matrix Estimator," *Biometrica*, **87**, 907-918.

[17] Cribari-Neto F. and N. M. S. Galvão (2003) "A Class of Improved Heteroskedasticity-Consistent Covariance Matrix Estimator," *Communications in Statistics: Theory and Methods*, **31**, 1951-1980.

[18] David A. Belsley (2002) "An Investigation of an Unbiased Correction for Heteroskedasticity and the Effects of Misspecifying the Skedastic Function," *Journal of Economic Dynamicc & Control,* **26**, 1379-1396.

[19] Davidson , R. and J. G. MacKinnon (1993) *Estimation and Inference in Economterics*, New York, Oxford University Press.

[20] Donald, S. G. (1987) "Efficient Estimation of Models with Heteroskedasticity of Unknown Form in Small Samples," Working Paper, University of British Columbia, Vancouver, Canada.

[21] Donald, S. G. (1992) "Some Asymptotic Results for Averages of Series Based Non-parametric Estimates," Working paper, Department of Economics, University of Florida.

[22] Donald S. G. and W. K. Newey (1994) "Series Estimation of Semilinear Models," *Journal of Multivariate Analysis*, **50**, 30-40.

[23] Donald, S. G. and W. K. Newey (2001) "Choosing the Number of Instruments," *Econometrica*, **69**, 1161-1191.

[24] Donald S. G., G. Imbens, W. Newey (2003) "Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions," *Journal of Econometrics*, **117**, 55-93.

[25] Eicker, F. (1963) "Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions," *Annals of Mathematical Statistics*, **34**, 447-456.

[26] Efron, B. (1982) *The Jackknife, the Bootstrap and Other Resampling Plans*, Philadelphia, Society for Industrial and Applied Mathematics.

[27] Gawande, K. (1997) "Generated Regressors in Linear and Nonlinear Models," *Economic Letters*, **54**, 119-126.

[28] Greene, W. H. (1997) *Econometric Analysis*, third edition, New York, Prentice-Hall.

[29] Griffiths, W. E. (1999) "Heteroskedasticity," Working Papers in Econometrics and Applied Statistics, No. **104**, University of New England, Australia.

[30] Grossman, G. M. and E. Helpman (1994) "Protection for Sale," *American Economic Review,* **84**, 833-850.

[31] Heckman, N. E. (1986) "Spline Smoothing in a Partially Linear Model," *Journal of Royal Statistics Society B*, **48**, 244-248.

[32] Hinkley D. V. (1977) "Jackknifing in Unbalanced Situation," *Techometrics*, **19**, 285-292.

[33] Horn, S. D., R. A. Horn and D. B. Duncan (1975) "Estimating Heteroskedastic Variances in Linear Model," *Journal of the American Statistical Association*, **70**, 380-385.

[34] Inoue, A. (2003) "A Bootstrap Appraoch to Moment Selection," Working Paper, North Carolina State University.

[35] Jian Shi and Tai-Shing Lau (2000) "Empirical Likelihood for Partial Linear Models," *Journal of Multivariate Analysis*, **72**, 132-148.

[36] Koenker, R., Jose A. F. Machado, Christopher L. Skeels and Alan H. Welsh (1994) "Momentary Lapses: Moment Expansions and the Robustness of Minimum Distance Estimation," *Econometric Theory*, **10**, 172-197.

[37] Li Qi (2000) "Efficient Estimation of Additive Partially Linear Models," *International Economic Review*, **41**, No, **4**, 1073-1092.

[38] Linton O. B. (1995) "Second Order Approximation in the Partially Linear Regression Model," *Econometrica*, **63**, , 1079-1112.

[39] Linton, O. B. (1996) "Second Order Approximation in a Linear Regression With Heteroskedasticity of Unknown Form," *Econometric Review*, **15**, 1-32.

[40] Long, J. S. and Ervin, L. H. (2000) "Using Heteroskedasticity-consistent Standard Errors in the Linear Regression Model," *The American Statistician*, **54**, 217-224.

[41] MacKinnon, J. G. and H. White (1987) "Some Heteroskedasticity-Consistent Covariance Matrix Estimator with Improved Finite Sample Properties," *Journal of Econometrics*, **29**, 305-325.

[42] Nagar, A. L. (1959) ”The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations,” *Econometrica*, **27**, 573-595.

[43] Newey, W. K. (1990) ”Semiparametric Efficiency Bounds,” *Journal of Applied Econometrics*, **5**, 99-135.

[44] Newey, W. K. (1993) ”Efficient Estimation of Models with Conditional Moment Restrictions,” in *Handbook of Statistics*, Vol. **11**, ed. by G. S. Maddala, C. R. Rao and H. D. Vinod. Elsevier Science Publishers, pp. 419-454.

[45] Newey, W. K. (1994a) ”Series Estimation of Regression Functionals,” *Econometric Theory*, **10**, 1-28.

[46] Newey, W. K. (1994b) ”Kernel Estimation of Partial Means and a General Variance Estimator,” *Econometric Theory*, **10**, 233-253.

[47] Newey, W. K. (1997) ”Convergence Rates and Asymptotic Normality for Series Estimator,” *Journal of Econometrics*, **79**, 147-168.

[48] Newey, W. K., J. Powell and J. Walker (1990) “Semiparametric Estimation of Selection Models: Some Empirical Results,” ,*American Economic Review, Papers and Proceedings*, **80**, 324-328.

[49] Newey, W. K. and Richard J. Smith (2004) “Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators,” *Econometrica*, **72**, 219-255.

[50] Pagan A. (1984) "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review,* **25**, 221-247.

[51] Pagan A. (1986) "Two Stage and Related Estimators and their Applications," *Review of Economic Studies*, **53**, 517-538.

[52] Rilstone, P. (1993) "Calculating the (Local) Semiparametric Efficiency Bounds for the Generated Regressors Problem," *Journal of Econometrics*, **56**, 357-370.

[53] Rilstone, P. (1996) "Nonparametric Estimation of Models with Generated Regressors" *International Economic Review*, **37**, 299-313.

[54] Rilestone, P. (1991) "Some Monte Carlo Evidence on the Relative Efficiency of Parametric and Semiparametric EGLS Estimator," *Journal of Business and Economic Statistics*, **9**, 179-187.

[55] Robinson, P. M. (1987) "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity," *Econometrica*, **55**, 875-891.

[56] Robinson P. M. (1988) "Root-N-Consistent Semiparametric Regression," *Econometrica*, **56**, 931-954.

[57] Rothenberg, T. J. (1984) "Approximate Power Functions for Some Robust Tests of Regression Coefficients," *Research paper in economics*, No. **84-1,** University of California, Berkeley, CA.

[58] Sheng-Yan Hong and Ping Cheng (1999) "Normal Approximation Rate of the Kernel Smoothing Estimator in a Partial Linear Model," *Journal of Multivariate Analysis*, **70**, 207-220.

[59] Stengos T. and B. Yan (2001) "Double Kernel Non-parametric Estimation in Semiparametric Econometric Models," *Nonparametric Statistics,* **13**, 883-906.

[60] Surekha, K., and Griffifths, W. E. (1984) "A Monte Carlo Comparison of Some Bayesian and Sampling Theory Estimators in Two Heteroskedastic Error Models," *Communications in Statistics*, *Part B - Simulation and Computation*, **13**, 85-106.

[61] White, H. (1980) "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, **48**, 817-838.

[62] White, H., and M. Stinchcombe (1991) "Adaptive Efficient Weighted Least Squares With Dependent Observations," in the IMA Volumes in Mathematics and its applications, Vol. **34**, Directions in Robust Statistics and Diagnostics Part II, ed. by Stahel, W. and Weisberg, S., Springer-Verlag, New York, 337-363.

[63] Wooldridge, J. M. (2000) *Introductory Econometrics: A Modern Approach*, South-Western Colledge Publishing.

[64] Wu, C. F. J. (1986) "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis," *Annals of Statistics*, **14**, 1216-1295.

[65] Yatchew (1997) "An Elementary Estimator of the Partial Linear Model," *Economics Letters* **57**, 135-143.

[66] Zhang Jian and Iréne Gijbels (2003) "Sieve Empirical Likelihood and Extensions of the Generalized Least Squares," *Scandinavian Journal of Statistics*, **30**, 1-24.

[67] Zhao Quanshui (2001) "Asymptotically Efficient Median Regression in the Presence of Heteroskedasticity of Unknown Form," *Econometric Theory*, **17**, 765-784.

Vita

Shih-Chang Lin was born in Chia-Yi, Taiwan on 17 December 1970, the eldest son of Sheng-Mu Lin and Bi-Chu Tsai. After completing his work at National Chia-Yi High School in 1989, he entered the National Taiwan University, Taipei, and majored in Economics. He received the degree of Bachelor of Art from the National Taiwan University in May 1993. In August 1993, he entered the graduate program of Economics at the National Taiwan University. He earned the Master of Art degree from the National Taiwan University in 1995 with a specialization in Time Series Analysis. After graduation, he served in R. O. C. Army and was commissioned a Lieutenant in Hua-Lien. In May 1997, after finishing two years' military services, he worked for the Board of Foreign Trade, Ministry of Economic Affairs in R. O. C. government. He then continued at the University of Texas at Austin to pursue his Ph.D. in Economics.

Permanent address: 3358 Lake Austin Blvd. Apt.# A
Austin, Texas 78703

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.