Consistent Lagrange Multiplier Type Specification Tests for Semiparametric Models*

Ivan Korolev†

Job Market Paper

November 3, 2017
(the latest version is available here)

Abstract

This paper considers specification testing in semiparametric econometric models. It develops a consistent series-based specification test for semiparametric conditional mean models against nonparametric alternatives. Consistency is achieved by turning a conditional moment restriction into a growing number of unconditional moment restrictions using series methods. The test is simple to implement because it requires estimating only the restricted semiparametric model and because the asymptotic distribution of the test statistic is pivotal. The use of series methods in estimation of the null semiparametric model allows me to account for the estimation variance and obtain refined asymptotic results. The test remains valid even if other semiparametric methods are used to estimate the null model as long as they achieve suitable convergence rates. This includes popular kernel estimators for single index or partially linear models. The test demonstrates good size and power properties in simulations. To illustrate the use of my test, I apply it to one of the semiparametric gasoline demand specifications from Yatchew and No (2001) and find no evidence against it.

*I am grateful to my advisors Frank Wolak, Han Hong, and Peter Reiss for their support and guidance, as well as to Svetlana Bryzgalova, Brad Larsen, and Joe Romano for very helpful conversations. I also thank Dmitry Arkhangelsky, Chris Bruegge, Colin Cameron, Pascaline Dupas, Liran Einav, Guido Imbens, Gordon Leslie, Jessie Li, Tom MaCurdy, Onder Polat, and Paulo Somaini for useful comments. I thank Adonis Yatchew for permission to use the Canadian household gasoline consumption dataset from Yatchew and No (2001). I gratefully acknowledge the financial support from the Stanford Graduate Fellowship Fund as a Koret Fellow and from the Stanford Institute for Economic Policy Research as a B.F. Haley and E.S. Shaw Fellow. All remaining errors are mine.

†Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA, 94305. E-mail: ikorolev@stanford.edu. Website: http://web.stanford.edu/~ikorolev/
1 Introduction

Applied economists often want to achieve two conflicting goals in their work. On the one hand, they wish to use the most flexible specification possible, so that their results are not driven by functional form assumptions. On the other hand, they wish to have a model that is consistent with the restrictions imposed by economic theory and can be used for valid counterfactual analysis.

While parametric models are often too restrictive and may not capture heterogeneity in the data well, nonparametric models may violate restrictions imposed by economic theory and suffer from the curse of dimensionality, i.e. become imprecise if the dimensionality of regressors is high. This, together with the fact that economic theory usually specifies one portion of the model but leaves the other unrestricted, makes semiparametric models especially attractive for empirical work in economics. For instance, semiparametric models have been used in estimation of demand functions (Hausman and Newey (1995), Schmalensee and Stoker (1999), Yatchew and No (2001)), production functions (Olley and Pakes (1996), Levinsohn and Petrin (2003)), Engel curves (Gong et al. (2005)), the labor force participation equation (Martins (2001)), the relationship between land access and poverty (Finan et al. (2005)), and marginal returns to education (Carneiro et al. (2011)).

Because many semiparametric models are restricted versions of fully nonparametric models, it is important to check the validity of implied restrictions. If semiparametric models are correctly specified, then using them, as opposed to nonparametric models, typically leads to more efficient estimates and may increase the range of counterfactual questions that can be answered using the model at hand. On the other hand, if semiparametric models are misspecified, then the semiparametric estimates are likely to be misleading and may result in incorrect policy implications.

In this paper I develop a new specification test that determines whether a semiparametric conditional mean model that the researcher has estimated provides a statistically valid description of the data as compared to a general nonparametric model. The test statistic
is based on a quadratic form in the semiparametric model residuals. When the errors are homoskedastic, this quadratic form can be computed as $nR^2$ from the regression of the semiparametric residuals on the series approximating functions. The heteroskedasticity robust version of the quadratic form on which the test is based can also be computed in regression based ways. Thus, the proposed test is simple to implement and avoids kernel smoothing in high dimensions. Moreover, the asymptotic distribution of the test statistic is pivotal, i.e. does not depend on the unknown parameters, so that calculating asymptotically exact critical values for the test is straightforward and does not require the use of resampling methods.

The proposed test uses series methods to turn a conditional moment restriction into a growing number of unconditional moment restrictions. I show that if the series functions can approximate the nonparametric alternatives that are allowed as the sample size grows, the test is consistent. My assumptions and proofs make precise what is required of the approximation and its behavior as the number of series terms and the sample size grow. These arguments differ from standard parametric arguments, when the number of regressors in the model is fixed.

My asymptotic theory allows both the number of parameters under the null as well as the number of restrictions to grow with the sample size. By doing so, I show that the parametric Lagrange Multiplier test can be extended to semiparametric models and serve as a consistent model specification test for these models. Because series methods have a projection interpretation, using series methods to nest the null model in the alternative and estimate the restricted model makes it possible to directly account for the estimation variance and obtain refined asymptotic results. This refinement, which can be viewed as a degrees of freedom correction, allows me to derive the asymptotic distribution of the test statistic under fairly weak rate conditions and leads to very good finite sample performance of the test in simulations.

Though this refinement is unique to series estimation methods, the proposed test, with a slight modification, remains valid even if other semiparametric methods, such as kernels or
local polynomials, are used to estimate the null model. Thus, the test applies to a wide class of semiparametric models, including single index models or partially linear models estimated using the two-step method proposed in Robinson (1988). Because the degrees of freedom correction is not available in that case, I have to impose more restrictive conditions on the convergence rates of semiparametric estimators, as well as an additional high level assumption that may be difficult to verify in practice. Moreover, even though the test statistics for series estimators and for other semiparametric estimators are asymptotically equivalent, my simulations show that the test based on the latter is undersized and low-powered in finite samples.

Intuitively, while the test based on series estimation methods uses the projection property to directly account for the form of the semiparametric residuals, the test based on general estimation methods only requires the semiparametric residuals to be close to the true errors. However, in finite samples, there may be a substantial difference between the residuals and true errors, which the latter approach fails to capture. As a result, even though both approaches are asymptotically valid, the former yields an accurate approximation of the finite sample distribution of the test statistic, while the latter does not work nearly as well.

Specification tests have long played an important role in theoretical econometrics. Several papers have studied specification testing when the null model contains a nonparametric component. Early work on specification testing in semiparametric models required certain ad hoc modifications, such as sample splitting in Yatchew (1992) and Whang and Andrews (1993) or randomization in Gozalo (1993). Fan and Li (1996) solve this problem and develop a kernel-based specification test that can be used to test a semiparametric null hypothesis against a general nonparametric alternative, but their test requires high-dimensional kernel smoothing and cannot be implemented with standard econometric software. Lavergne and Vuong (2000) refine the test of Fan and Li (1996), but they only consider significance testing in nonparametric models. Kernel-based specification tests are also developed in Chen and Fan (1999), Delgado and Manteiga (2001), Ait-Sahalia et al. (2001), and Bravo (2012). In
all these papers, the asymptotic distribution of the test statistic is quite complicated and requires either estimating several nuisance parameters or using the bootstrap, which can be computationally costly.

I circumvent this problem by relying on series methods to construct the test statistic. Because the number of series terms grows with the sample size, the usual asymptotic results for the parametric Lagrange Multiplier no longer apply. However, it is possible to normalize the test statistic so that the resulting normalized statistic is asymptotically standard normal. Therefore, the quantiles of the standard normal distribution can be used as asymptotically exact critical values for the test.

The proposed test is based on a quadratic form in the restricted model residuals. Thus, it can be viewed as a nonparametric generalization of the conventional Lagrange Multiplier test, classical treatments of which include Breusch and Pagan (1980), Engle (1982), and Engle (1984). This generalization is not novel in itself, as Hall (1990) and McCulloch and Percy (2013) also develop Lagrange Multiplier specification tests for parametric models against flexible alternatives. However, they only consider null hypotheses with fully specified parametric distributions, and their asymptotic analysis treats the number of series terms in the alternative model as fixed. As a result, their tests fail to achieve consistency. In contrast, I allow for semiparametric conditional mean models and develop an asymptotic theory for the case when the number of series terms grows with the sample size, which results in a consistent specification test.

My work is closely related to the literature on series-based specification tests, such as de Jong and Bierens (1994), Hong and White (1995), Koenker and Machado (1999), Donald et al. (2003), and Sun and Li (2006). These papers extend the Conditional Moment test of Newey (1985) by considering a growing number of unconditional restrictions and thus achieve consistency. However, they only consider parametric null hypotheses and do not develop a degrees of freedom correction. In contrast, my test can handle a broad class of semiparametric models and involves the degrees of freedom correction in the case when the semiparametric
null model is nested in the nonparametric alternative and estimated using series methods. I show that this correction plays a crucial role in the semiparametric case, because it allows me to weaken the rate conditions and improves the finite sample performance of the test.

Hong and White (1995), among others, investigate the behavior of the test statistic for the parametric null hypotheses under the global alternative and under local alternatives. I repeat this analysis for semiparametric null hypotheses and reach similar conclusions: the test statistic diverges to infinity faster than the parametric rate $n^{1/2}$ under the global alternative, but the test can only detect local alternatives that approach zero slower than the parametric rate $n^{-1/2}$. Moreover, both rates are asymptotically the same as in the parametric case.\footnote{By saying “asymptotically the same,” I mean that the exact rates in the parametric and semiparametric cases are different but the ratio of two rates goes to 1 as $n \to \infty$.}

To my knowledge, there are two studies that develop series-based specification tests that allow for semiparametric null hypotheses. Gao et al. (2002) consider only additive models and do not explicitly develop a consistent test against a general class of alternatives. Their test is based on the estimates from the unrestricted model and can be viewed as a Wald type test for variables significance in nonparametric additive conditional mean models. In contrast, my test is based on the residuals from the restricted semiparametric model and is consistent against a broad class of alternatives for the conditional mean function.

Li et al. (2003) use the approach that was first put forth in Bierens (1982) and Bierens (1990) and develop a series-based specification test based on weighting the moments, rather than considering an increasing number of series-based unconditional moments. Their test can detect local alternatives that approach zero at the parametric rate $n^{-1/2}$, but the asymptotic distribution of their test statistic depends on nuisance parameters, and it is difficult to obtain appropriate critical values. They propose using a residual-based wild bootstrap to approximate the critical values. In contrast, my test statistic is asymptotically standard normal under the null, so calculating critical values is straightforward.

Another attractive feature of the proposed test is that the alternative model does not have to be fully nonparametric. Because series methods make it easy to impose restrictions
on nonparametric models, the proposed test can be used to test a more restricted semiparametric model against a broader semiparametric alternative instead of a fully nonparametric alternative. For instance, a researcher may be willing to compare a partially linear model $Y_i = X_{1i}' \beta + g(X_{2i}) + \varepsilon_i$ with a varying coefficient model $Y_i = X_{1i}' \beta(X_{2i}) + g(X_{2i}) + \varepsilon_i$ or an additive model $Y_i = h(X_{1i}) + g(X_{2i}) + \varepsilon_i$. The proposed test can be modified to handle such comparisons by considering only the unconditional moments based on the series terms that are present under the alternative.

Restricting the class of alternatives will result in the loss of consistency against a general nonparametric alternative, because the semiparametric class of alternatives will be unable to detect certain deviations from the null. However, it will also increase the test power if the true model does lie in the conjectured semiparametric class. It is possible to use my test with several alternatives simultaneously, including semiparametric alternatives to improve power in certain directions but also including a general nonparametric alternative to achieve consistency. The Bonferroni correction can then be used to control the test size. I show in simulations that this approach leads to higher power when the null hypothesis is false but the true model belongs to the restricted class of alternatives without disturbing the size of the test or losing consistency against a general class of alternatives.

Finally, similar to the overidentifying restrictions test in GMM models or other omnibus specification tests, the proposed test is silent on how to proceed if the null is rejected. In this respect, it is clearly a test of a particular model specification but not a comprehensive model selection procedure. I plan to study model selection methods for semiparametric and nonparametric models, such as series-based Bayesian Information Criterion or upward/downward testing procedures based on the proposed test, in future work.

The remainder of the paper is organized as follows. Section 2 presents a motivating example from industrial organization. Section 3 introduces the model and describes how to construct the series-based specification test for semiparametric models. Section 4 develops the asymptotic theory for the proposed test when series methods are used in estimation.
Section 5 extends the asymptotic theory to the case when other semiparametric methods, such as kernels or local polynomials, are used in estimation. Section 6 studies the behavior of the proposed test in simulations. Section 7 applies the proposed test to one of the semiparametric household gasoline demand specifications from Yatchew and No (2001) and shows that it is not rejected by the data. Section 8 concludes.

Appendix A collects all tables and figures. Appendix B provides an intuitive derivation of the proposed test as well as a step-by-step description of how to implement the proposed test. Appendix C contains proofs of technical results.

2 Motivating Example

Suppose that a researcher has cross-section data on total costs $TC_i$, output $Q_i$, firm characteristics $Z_i$, and factor prices $P_{Li}$ and $P_{Ki}$ for firms in a given industry, and wants to estimate a cost function. Any cost function $C(q,p_L,p_K)$ has to satisfy certain properties, such as monotonicity, concavity, and homogeneity of degree 1 in factor prices. Because nonparametric estimation of the cost function may lead to a violation of these restrictions, the researcher may choose a theory-based parametric functional form of the cost function.

If the researcher assumes a Cobb-Douglas production function for firm $i$, $Q_i = A_i L_i^\alpha K_i^\beta$, under certain assumptions on the unobservables (for details, see Reiss and Wolak (2007)) she will derive the following relationship:

$$\ln TC_i = C_1 + \gamma \ln P_{Ki} + (1 - \gamma) \ln P_{Li} + \delta \ln Q_i + \epsilon_i,$$

where $\gamma = \beta/(\alpha + \beta)$ and $\delta = 1/(\alpha + \beta)$, and $E[\epsilon_i| \ln P_{Ki}, \ln P_{Li}, Q_i, Z_i] = 0$.

This specification is restrictive because it does not allow the parameters to vary across firms, while economic theory usually provides no reason to believe that the parameters should be the same for all firms. Hence, a better alternative may be a semiparametric varying coefficient model, which allows the input prices and output elasticities to vary flexibly with
observable firm characteristics $Z_i$: $\gamma = \gamma(z)$ and $\delta = \delta(z)$ for some unknown functions $\gamma(\cdot)$ and $\delta(\cdot)$. Then the labor and capital shares can be recovered from these functions as $\alpha(z) = (1 - \gamma(z))/\delta(z)$ and $\beta(z) = \gamma(z)/\delta(z)$. The total cost relationship becomes:

$$\ln TC_i = C_1(Z_i) + \gamma(Z_i) \ln P_{Ki} + (1 - \gamma(Z_i)) \ln P_{Li} + \delta(Z_i) \ln Q_i + \varepsilon_i,$$

This model may fit the data better and be more realistic than the fully parametric model. Moreover, it can easily satisfy the monotonicity, concavity, and homogeneity restrictions. However, because the semiparametric model is more restricted than the nonparametric model, the researcher may want to guard against possible misspecification of the semiparametric model and check whether it is consistent with the data. The proposed test allows the researcher to test the semiparametric model against a general nonparametric model:

$$\ln TC_i = g(\ln Q_i, \ln P_{Li}, \ln P_{Ki}, Z_i) + U_i,$$

with $E[U_i | \ln Q_i, \ln P_{Li}, \ln P_{Ki}, Z_i] = 0$.

3 Model and Specification Test

This section describes the model, discusses semiparametric series estimators, and introduces the test statistic.

3.1 Model and Null Hypothesis

Let $(Y_i, X_i') \in \mathbb{R}^{1+d_x}$, $d_x \in \mathbb{N}$, $i = 1, \ldots, n$, be independent and identically distributed random variables with $E[Y_i^2] < \infty$. Then there exists a measurable function $g$ such that $g(X_i) = E[Y_i | X_i]$ a.s. Then the nonparametric model can be written as

$$Y_i = g(X_i) + \varepsilon_i, \quad E[\varepsilon_i | X_i] = 0$$
The goal of this paper is to test the null hypothesis that the conditional mean function is semiparametric. A generic semiparametric null hypothesis is given by

\[ H_{0}^{SP} : P_{X} (g(X_i) = f(X_i, \theta_0, h_0)) = 1 \text{ for some } \theta_0 \in \Theta, h_0 \in H, \]  

where \( f : \mathcal{X} \times \Theta \times H \rightarrow \mathbb{R} \) is a known function, \( \theta \in \Theta \subset \mathbb{R}^d \) is a finite-dimensional parameter, and \( h \in H = \mathcal{H}_1 \times ... \times \mathcal{H}_q \) is a vector of unknown functions.

The global alternative is

\[ H_1 : P_{X} (g(X_i) \neq f(X_i, \theta, h)) > 0 \text{ for all } \theta \in \Theta, h \in H \]  

When the semiparametric null hypothesis is true, the model becomes

\[ Y_i = f(X_i, \theta_0, h_0) + \varepsilon_i, \quad E[\varepsilon_i | X_i] = 0 \]

Because the model is semiparametric while the moment condition \( E[\varepsilon_i | X_i] = 0 \) is fully nonparametric, the null model is overidentified and it is possible to test its specification. In order to do so, I turn the conditional moment restriction \( E[\varepsilon_i | X_i] = 0 \) into a sequence of unconditional moment restrictions using series methods. But first, I introduce series approximating functions and semiparametric series estimators.

For any variable \( z \), let \( Q_{l_n} (z) = (q_1(z), ..., q_{l_n}(z))' \) be a \( l_n \)-dimensional vector of approximating functions of \( z \), where the number of series terms \( l_n \) is allowed to grow with the sample size \( n \). Possible choices of series functions include:\(^2\)

(a) Power series. For univariate \( z \), they are given by:

\[ Q_{l_n} (z) = (1, z, ..., z^{l_n-1})' \]

\(^2\)For a more detailed discussion of series methods and for other possible choices of basis functions, see Section 5 in Newey (1997), Section 2 in Donald et al. (2003), or Section 2.3 in Chen (2007).
(b) Splines. Let \( s \) be a positive scalar giving the order of the spline, and let \( t_1, \ldots, t_{k_n-s-1} \) denote knots. Then for univariate \( z \), splines are given by:

\[
Q_{ln}^{\prime}(z) = (1, z, \ldots, z^s, 1\{z > t_1\}(z - t_1)^s, \ldots, 1\{z > t_{k_n-s-1}\}(z - t_{k_n-s-1})^s)
\]

(3.4)

Multivariate power series or splines can be formed from products of univariate ones.

### 3.2 Series Estimators

In this section I introduce additional notation and define semiparametric estimators for the case when series methods are used to estimate the restricted model.\(^3\) Suppose that the researcher writes the semiparametric null model in a series form:

\[
Y_i = f(X_i, \theta, h) + \varepsilon_i \equiv W_{mn}^n(X_i)'\beta_1 + R_{ln}^n + \varepsilon_i \equiv W_{mn}^n(X_i)'\beta_1 + e_i,
\]

(3.5)

where \( W_{mn}^n(X_i) \) are appropriate regressors or basis functions, \( R_{ln}^n \equiv (g(X_i) - W_{mn}^n(X_i)'\beta_1) \) is the approximation error, and \( e_i = \varepsilon_i + R_{ln}^n \) is the composite error term.

**Example 1.** Suppose that the semiparametric null model is partially linear with \( f(X_i, \theta, h) = X_{1i}\theta + h(X_{2i}) \), where \( X_{1i} \) and \( X_{2i} \) are scalars and \( X_i = (X_{1i}, X_{2i})' \). Approximate \( h(x_2) \approx \sum_{j=1}^{l_n} \gamma_j q_j(x_2) = Q_{ln}^{\prime}(x_2)'\gamma \) and rewrite the model as:

\[
Y_i = X_{1i}\theta + h(X_{2i}) + \varepsilon \equiv X_{1i}\theta + Q_{ln}^{\prime}(X_{2i})'\gamma + R_{ln}^n + \varepsilon_i \equiv W_{mn}^n(X_i)'\beta_1 + e_i,
\]

where \( W_{mn}^n(X_i) = (X_{1i}, Q_{ln}^{\prime}(X_{2i}))' \), \( m_n = l_n + 1 \), \( \beta_1 = (\theta, \gamma)' \), and \( e_i = R_{ln}^n + \varepsilon_i \), and \( R_{ln}^n = R_{ln}^l \equiv h(X_{2i}) - Q_{ln}^{\prime}(X_{2i})'\gamma \). If power series are used, then \( W_{mn}^n(X_i) = (X_{1i}, 1, X_{2i}, X_{2i}^2, \ldots, X_{2i}^{l_{n-1}})' \).

In order to test the semiparametric model, I test whether additional series terms should enter the model. These are not the extra series terms used to approximate the semiparametric

\(^3\)For more details, refer to Appendix B.1.
null model better, but rather the series terms that cannot enter the model under the null and are supposed to capture possible deviations from it. Thus, the alternative is given by:

\[ Y_i = W^{mn}(X_i)' \beta_1 + T^{rn}(X_i)' \beta_2 + R^{mn}_i + \varepsilon_i \equiv P^{kn}(X_i)' \beta + \varepsilon_i \quad (3.6) \]

where \( P^{kn}(X_i) = (W^{mn}(X_i)', T^{rn}(X_i)')' \), \( \beta = (\beta_1', \beta_2')' \), and \( T^{rn}(X_i) \) are the series terms that are present only under the alternative.

**Example 1** (continued). In the partially linear model example, a possible choice of \( T^{rn}(X_i) \) is \( T^{rn}(X_i) = (X_{1i}^2, ..., X_{n1}^{l_2-1}, X_{1i}X_{2i}, ..., X_{n1}^{l_2-1}X_{2i}^{l_1-1})' \). These series terms can enter the model if the null hypothesis is false, but they cannot enter the model if it is true.

If the moment condition \( E[Y_i - f(X_i, \theta_0, h_0)|X_i] = 0 \) holds, then the errors \( \varepsilon_i = Y_i - f(X_i, \theta_0, h_0) \) are uncorrelated with any function of \( X_i \), not only with \( W^{mn}(X_i) \), and the additional series terms should be insignificant. Hence, the null hypothesis corresponds to \( \beta_2 = 0 \). The estimate of \( \beta_1 \) under the null becomes \( \tilde{\beta}_1 = (W'W)^{-1}W'Y \), \( 4 \) and the restricted residuals are given by

\[ \tilde{\varepsilon} = Y - W\tilde{\beta}_1 = Y - W(W'W)^{-1}W'Y = M_WY = M_W(\varepsilon + R), \]

where \( M_W = I - P_W \), \( P_W = W(W'W)^{-1}W' \). The residuals satisfy the condition \( W'\tilde{\varepsilon} = 0 \).

\( m_n \) is the number of terms in the semiparametric null model. It has to grow with the sample size in order to approximate nonparametric components of the model sufficiently well.\(^5\) \( r_n \) is the number of series terms that capture possible deviations from the null. \( k_n = m_n + r_n \) is the total number of series terms in the unrestricted model. This number has to grow with the sample size if it is to approximate any nonparametric alternative. Typically the number of terms in the restricted semiparametric model, \( m_n \), is significantly smaller than the number of terms in the unrestricted nonparametric model, \( k_n \), so that \( m_n/k_n \to 0 \) and \( r_n/k_n \to 1 \).

---

\( ^4 \) For any vector \( V_i \), let \( V = (V_1, ..., V_n)' \) be a matrix that stacks all observations together. To simplify notation, I make dependence of \( W^{mn}(X) \) and \( R^{mn} \) on the sample size implicit.

\( ^5 \) The notion of sufficiently well is made precise later.
3.3 Test Statistic

Instead of the conditional moment restriction implied by the null hypothesis: $E[\varepsilon_i | X_i] = E[Y_i - f(X_i, \theta_0, h_0) | X_i] = 0$, the test will be based on the unconditional moment restriction $E[P_k^n(X_i) \varepsilon_i] = 0$. I will show that requiring $k_n$ to grow with the sample size and $P_k^n(x)$ to approximate any unknown function in a wide class of functions will allow me to obtain a consistent specification test based on a growing number of unconditional moment restrictions.

The test statistic is based on the sample analog of the population moment condition $E[P_k^n(X_i) \varepsilon_i] = 0$. In the homoskedastic case, the test statistic resembles the LM test statistic and is given by

$$
\xi = \tilde{\varepsilon}' P (\hat{\sigma}^2 P' P)^{-1} P' \tilde{\varepsilon}, \tag{3.7}
$$

where $P = (P_k^n(X_1), ..., P_k^n(X_n))'$, $\tilde{\varepsilon} = (\tilde{\varepsilon}_1, ..., \tilde{\varepsilon}_n)'$, $\tilde{\varepsilon}_i = Y_i - f(X_i, \tilde{\theta}, \tilde{h})$ are semiparametric residuals, and $\hat{\sigma}^2 = \tilde{\varepsilon}' \tilde{\varepsilon} / n$. Note that when the null model is nested in the alternative and estimated using series methods, the test statistic reduces to

$$
\xi = \tilde{\varepsilon}' \tilde{T} (\hat{\sigma}^2 \tilde{T}' \tilde{T})^{-1} \tilde{T}' \tilde{\varepsilon},
$$

where $\tilde{T} = M W T$ are the residuals from the regression of $T_r^n(X_i)$ on $W_m^n(X_i)$.

In the heteroskedastic case, the test statistic is modified appropriately:

$$
\xi_{HC,1} = \tilde{\varepsilon}' P (P' \tilde{\Sigma} P)^{-1} P' \tilde{\varepsilon}, \tag{3.8}
$$

where $\tilde{\Sigma} = diag(\tilde{\varepsilon}_i^2)$. A further modification is possible when the null model is nested in the alternative and estimated using series methods:

$$
\xi_{HC,2} = \tilde{\varepsilon}' \tilde{T} (\tilde{T}' \tilde{\Sigma} \tilde{T})^{-1} \tilde{T}' \tilde{\varepsilon} \tag{3.9}
$$
All three quadratic forms, $\xi$, $\xi_{HC,1}$, and $\xi_{HC,2}$ can be computed in regression-based ways. These ways are discussed in Section B.5 in the Appendix.\(^6\)

Because the dimensionality $k_n$ of $P$ grows with the sample size, a normalization is needed to obtain a convergence in distribution result. I show that the test statistic

$$t_{\tau_n} = \frac{\xi - \tau_n}{\sqrt{2\tau_n}}$$

for a suitable sequence $\tau_n \to \infty$ as $n \to \infty$ works and is asymptotically pivotal. I will discuss the choice of $\tau_n$ in the next section.

4 Asymptotics with Series Estimation Methods

In this section I derive the asymptotic properties of the proposed test. I focus here on the case when series methods are used to nest the null model in the alternative and estimate the restricted model. This is because series methods have a projection interpretation, which makes it possible to directly account for the estimation variance. As a result, the rate conditions required for the validity of the proposed test are mild. The use of other semiparametric estimators, such as kernels or local polynomials, does not allow for the degrees of freedom correction and leads to more restrictive rate conditions. The results on the asymptotic behavior of the proposed test in a general case are presented in Section 5.

First, I impose basic assumptions on the data generating process.

Assumption 1. $(Y_i, X_i')' \in \mathbb{R}^{1+d_x}$, $d_x \in \mathbb{N}$, $i = 1, \ldots, n$ are i.i.d. random draws of the random variables $(Y, X)$, and the support of $X$, $\mathcal{X}$, is a compact subset of $\mathbb{R}^{d_x}$.

Next, I define the error term and impose two moment conditions.

**Assumption 2.** Let $\varepsilon_i = Y_i - E[Y_i|X_i]$. The following two conditions hold:

\(^6\)For regression-based ways to compute the parametric LM test statistic, see Wooldridge (1987) or Chapters 7 and 8 in Cameron and Trivedi (2005).
(a) \(0 < \sigma^2(x) = E[\varepsilon_i^2 | X_i = x] < \infty\).

(b) \(E[\varepsilon_i^4 | X_i]\) is bounded.

The following assumption deals with the behavior of the approximating series functions. From now on, let \(||A|| = [tr(A'A)]^{1/2}\) be the Euclidian norm of a matrix \(A\).

**Assumption 3** (Donald et al. (2003), Assumption 2). For each \(k\) there is a constant scalar \(\zeta(k)\) and matrix \(B\) such that \(\tilde{P}^k(x) = BP^k(x)\) for all \(x \in \mathcal{X}\), \(\sup_{x \in \mathcal{X}} ||\tilde{P}^k(x)|| \leq \zeta(k)\), \(E[\tilde{P}^k(X_i)\tilde{P}^k(X_i)']\) has smallest eigenvalue bounded away from zero uniformly in \(x\), and \(\sqrt{k} \leq \zeta(k)\).

**Remark 1** (\(\zeta(k)\) for Common Basis Functions). Explicit expressions for \(\zeta(k)\) are available for certain families of basis approximating functions. For instance, it has been shown (see, e.g., Section 15.1.1 in Li and Racine (2007)) that under additional assumptions, \(\zeta(k) = O(k^{1/2})\) for splines and \(\zeta(k) = O(k)\) for power series.

The following lemma shows that a convenient normalization can be used. This normalization is typical in the literature on series methods.

**Lemma 1** (Donald et al. (2003), Lemma A.2). If Assumption 3 is satisfied then it can be assumed without loss of generality that \(\tilde{P}^k(x) = P^k(x)\) and that \(E[P^k(X_i)P^k(X_i)'] = I_k\).

Next, I impose an assumption that requires the approximation error for the semiparametric model to vanish sufficiently fast under the null.

**Assumption 4.** Suppose that \(H_0\) holds. There exists \(\alpha > 0\) such that

\[
\sup_{x \in \mathcal{X}} |f(x, \theta_0, h_0) - W^{mn}(x)' \beta_1| = O(m^{-\alpha})
\]

**Remark 2** (\(\alpha\) for Common Basis Functions). In certain special cases, it is possible to characterize \(\alpha\) explicitly. Suppose that power series or splines are used and that \(f(x, \theta, h) = x_1' \theta + h(x_2)\), where \(h\) has \(r\) continuous derivatives and \(x_2\) is \(d_{x_2}\)-dimensional. Then, similarly to the results in Chapter 15 in Li and Racine (2007), it can be shown that \(\alpha = r/d_{x_2}\).
The following lemma provides the rates of convergence for semiparametric series estimators under the null hypothesis.

**Lemma 2** *(Li and Racine (2007), Theorem 15.1)*. Let \( f(x) = f(x, \theta_0, h_0), f_i = f(X_i), \)
\( \hat{f}(x) = f(x, \hat{\theta}, \hat{h}) = W_m^m(x)\hat{\beta}_1, \) and \( \hat{f}_i = \hat{f}(X_i) \). Under Assumptions 1, 3, and 4, the following is true:

(a) \( \sup_{x \in X} |\hat{f}(x) - f(x)| = O_p \left( \zeta(m_n)(\sqrt{m_n/n} + m_n^{-\alpha}) \right) \)

(b) \( \frac{1}{n} \sum_{i=1}^{n} (\hat{f}_i - f_i)^2 = O_p(m_n/n + m_n^{-2\alpha}) \)

(c) \( \int (\hat{f}(x) - f(x))^2 dF(x) = O_p(m_n/n + m_n^{-2\alpha}) \)

**Remark 3** *(Convergence Rates for Semiparametric Series Estimators)*. The exact rates given in Lemma 2 are derived in Li and Racine (2007) for series estimators in nonparametric models. However, as other examples in Chapter 15 in Li and Racine (2007) show, similar rates can be derived in a wide class of semiparametric models, such as partially linear, varying coefficient, or additive models (see Theorems 15.5 and 15.7 in Li and Racine (2007)). In each of these cases, it is possible to replace the rates in Lemma 2 with the rates for the particular case of interest. With appropriately modified rate conditions and assumptions, the results developed below will continue to hold.

**Remark 4** *(On the Possibility of Using Improved Results on Convergence Rates of Semiparametric Estimators)*. It may be possible to use recent results on the convergence rates for series estimators given on Belloni et al. (2015) instead of the convergence rates in Lemma 2. This may lead to weaker rate conditions in theorems given in subsequent sections. I plan to study this question in future research.

It is now possible to derive the asymptotic distribution of the test statistic 3.10 under the null. I consider the homoskedastic and heteroskedastic cases separately.
4.1 Homoskedastic Case

The limiting distribution of the test statistic under the null is given by the next result:

**Theorem 1.** Assume that series methods are used to estimate the restricted model. Also assume that Assumptions 1, 2, 3, and 4 are satisfied, \( \sigma^2(x) = \sigma^2, \ 0 < \sigma^2 < \infty, \) for all \( x \in X, \) and the following rate conditions hold:

\[
\begin{align*}
\zeta(k_n)^2 k_n r_n^{1/2} / n &\to 0 \quad (4.1) \\
\zeta(r_n) r_n / n^{1/2} &\to 0 \quad (4.2) \\
\zeta(k_n) m_n^{1/2} k_n^{1/2} / n^{1/2} &\to 0 \quad (4.3) \\
m m_n^{-2\alpha} / r_n^{1/2} &\to 0 \quad (4.4) \\
\zeta(r_n)^2 / n^{1/2} &\to 0 \quad (4.5)
\end{align*}
\]

Then

\[
t_{r_n} = \frac{\xi - r_n}{\sqrt{2r_n}} \overset{d}{\to} N(0, 1),
\]

where \( \xi \) is as in Equation 3.7.

**Remark 5** (Rate Conditions under Homoskedasticity). Conditions 4.1–4.5 are sufficient conditions for the result of the theorem to hold. Conditions 4.1 and 4.2 are used to bound the error from replacing \( \tilde{\sigma}^2 T' M_W T \) with \( \sigma^2 E[T_i T'_i] \). Conditions 4.3 and 4.4 are used to bound the error from approximating unknown functions with finite series expansions. Condition 4.5 is used to obtain the convergence in distribution result by applying Lemma A.1 in the Appendix.

**Remark 6** (Examples of Permissible Choices of \( m_n, r_n, \) and \( k_n \) under Homoskedasticity). As discussed above, typically \( \zeta(k) = O(k^{1/2}) \) for splines and \( \zeta(k) = O(k) \) for power series. It can be shown that for splines, the rates \( k_n = O(n^{2/7}), \ r_n = O(n^{2/7}), \ m_n = O(n^{1/4}) \) are permissible in the sense of rate conditions 4.1–4.5 if \( \alpha \geq 4 \). For power series, the rates \( k_n = O(n^{2/9}), \ r_n = O(n^{2/9}), \ m_n = O(n^{1/5}) \) are permissible in the sense of rate conditions 4.1–4.5 if \( \alpha \geq 5 \).
The following corollary shows that using a $\chi^2$ approximation with a growing number of degrees of freedom also results in an asymptotically exact test in the homoskedastic case.

**Corollary 1.** If the conditions of Theorem 1 hold, then

$$P(\xi \geq \chi^2_{r_n}(1-\alpha)) = P\left(\frac{\xi - r_n}{\sqrt{2r_n}} \geq \frac{\chi^2_{r_n}(1-\alpha) - r_n}{\sqrt{2r_n}}\right) \to \alpha$$

Thus, there are two ways to construct an asymptotically exact specification test. Exactly which one might be preferred in terms of the finite sample performance is studied in Section 6.

### 4.2 Heteroskedastic Case

The limiting distribution of the heteroskedasticity robust test statistic under the null is given by the next two results:

**Theorem 2.** Assume that series methods are used to estimate the restricted model. Also assume that Assumptions 1, 2, 3, and 4 are satisfied, and the following rate conditions hold:

\[
\begin{align*}
(m_n/n + m_n^{-2\alpha})\zeta(r_n)^2r_n^{1/2} & \to 0 \\
\zeta(r_n)r_n/n^{1/2} & \to 0 \\
\zeta(k_n)m_n^{1/2}k_n^{1/2}/n^{1/2} & \to 0 \\
nm_n^{-2\alpha}/r_n^{1/2} & \to 0 \\
\zeta(r_n)^2/n^{1/2} & \to 0 
\end{align*}
\]

Also assume that $||\hat{\Omega} - \tilde{\Omega}|| = o_p(r_n^{-1/2})$, where $\tilde{\Omega} = T'\tilde{\Sigma}T/n$ and

$$\hat{\Omega} = T'\hat{\Sigma}T/n - (T'\hat{\Sigma}W/n)(W'\hat{\Sigma}W/n)^{-1}(W'\hat{\Sigma}T/n)$$
Then
\[
t_{r_n, HC, 1} = \frac{\xi_{HC, 1} - r_n}{\sqrt{2}r_n} \overset{d}{\rightarrow} N(0, 1),
\]
(4.12)
where \(\xi_{HC, 1}\) is as in Equation 3.8.

**Theorem 3.** Assume that series methods are used to estimate the restricted model. Also assume that Assumptions 1, 2, 3, and 4 are satisfied, and the rate conditions 4.7–4.11 hold. Also assume that \(||\hat{\Omega} - \tilde{\Omega}|| = o_p(r_n^{-1/2})\), where \(\tilde{\Omega} = T^\prime \tilde{\Sigma} T / n\) and \(\hat{\Omega} = \tilde{T}^\prime \tilde{\Sigma} \tilde{T} / n\).

Then
\[
t_{r_n, HC, 2} = \frac{\xi_{HC, 2} - r_n}{\sqrt{2}r_n} \overset{d}{\rightarrow} N(0, 1),
\]
(4.13)
where \(\xi_{HC, 2}\) is as in Equation 3.9.

**Remark 7** (Rate Conditions under Heteroskedasticity). Conditions 4.7–4.11 are sufficient conditions for the result of the theorem to hold. Conditions 4.7 and 4.8 are used to bound the error from replacing \(T^\prime \tilde{\Sigma} T / n\) with \(E[\varepsilon_i^2 T_i T_i^\prime]\). Conditions 4.9 and 4.10 are used to bound the error from approximating unknown functions with finite series expansions. Condition 4.11 is used to obtain the convergence in distribution result by applying Lemma A.1 in the Appendix.

**Remark 8** (Examples of Permissible Choices of \(m_n, r_n,\) and \(k_n\) under Heteroskedasticity). It can be shown that the same rates as in Remark 6 (\(k_n = O(n^{2/7}), r_n = O(n^{2/7}), m_n = O(n^{1/4})\) for splines if \(\alpha \geq 4; k_n = O(n^{2/9}), r_n = O(n^{2/9}), m_n = O(n^{1/5})\) for power series if \(\alpha \geq 5\) are permissible in the sense of rate conditions 4.7–4.11.

However, it is not clear whether these rates satisfy the high-level assumption \(||\hat{\Omega} - \tilde{\Omega}|| = o_p(r_n^{-1/2})\). It is possible that this assumption imposes stronger rate restrictions that are difficult to verify.

\[\text{\footnotesize \cite{19}}\]

\(\overset{d}{\rightarrow}\) Instead of imposing primitive conditions under which \(\hat{\Omega}\) can be replaced with \(\tilde{\Omega}\), I directly impose the requirement that \(||\hat{\Omega} - \tilde{\Omega}|| = o_p(r_n^{-1/2})\). This is a high level condition, but it is difficult to derive sufficient primitive conditions.
The following corollary shows that using a $\chi^2$ approximation with a growing number of degrees of freedom also results in an asymptotically exact test in the heteroskedastic case.

**Corollary 2.** If the conditions of Theorem 2 (Theorem 3) hold, then for $j = 1$ ($j = 2$)

$$P\left(\xi_{HC,j} \geq \chi^2_{r_n}(1 - \alpha)\right) = P\left(\frac{\xi_{HC,j} - r_n}{\sqrt{2r_n}} \geq \frac{\chi^2_{r_n}(1 - \alpha) - r_n}{\sqrt{2r_n}}\right) \to \alpha,$$

4.3 Behavior of Test Statistic under a Global Alternative

In this subsection, I study the behavior of the test statistic under a global alternative. To obtain a consistent specification test, I will rely on the following result from Donald et al. (2003) that shows that the conditional mean restriction $E[\varepsilon_i|X_i] = 0$ is equivalent to a sequence of unconditional moment restrictions.

**Assumption 5.** (Donald et al. (2003), Assumption 1)

Assume that $E[P^k(X_i)P^k(X_i)']$ is finite for all $k$, and for any $a(x)$ with $E[a(X_i)^2] < \infty$ there are $k \times 1$ vectors $\gamma_k$ such that, as $k \to \infty$,

$$E[(a(X_i) - P^k(X_i)\gamma_k)^2] \to 0$$

**Lemma 3.** (Donald et al. (2003), Lemma 2.1)

Suppose that Assumption 5 is satisfied and $E[\varepsilon_i^2]$ is finite. If $E[\varepsilon_i|X_i] = 0$ then $E[P^k(X_i)\varepsilon_i] = 0$ for all $k$. Furthermore, if $E[\varepsilon_i|X_i] \neq 0$ then $E[P^k(X_i)\varepsilon_i] \neq 0$ for all $k$ large enough.

This is a population result in the sense that it does not involve the sample size $n$. In order to use this result in practice, I require the number of series terms used to construct the test statistic, $k_n$, to grow with the sample size. By doing so, I ensure that the unconditional moment restriction $E[P^{k_n}(X_i)\varepsilon_i] = 0$, on which the test is based, is equivalent to the conditional moment restriction $E[\varepsilon_i|X_i]$. Thus, the test will be consistent against a wide class of alternatives satisfying Assumption 5.
In order to analyze the behavior of the test under a global alternative, I introduce some notation first. The true model is nonparametric:

\[ Y_i = g(X_i) + \varepsilon_i, \quad E[\varepsilon_i|X_i] = 0 \]

An alternative way to write this model is

\[ Y_i = f(X_i, \theta^*, h^*) + \varepsilon^*_i, \]

where \( \theta^* \) and \( h^* \) are pseudo-true parameter values and \( \varepsilon^*_i \equiv \varepsilon_i + (g(X_i) - f(X_i, \theta^*, h^*)) \equiv \varepsilon_i + d(X_i) \) is a composite error term. The pseudo-true parameter values minimize

\[ E[(g(X_i) - f(X_i, \theta, h))^2] \]

over a suitable parameter space, and I assume that the semiparametric estimates under misspecification are consistent for the pseudo-true values.

Note that the model can be written as

\[ Y_i = W^{mn}(X_i)'\beta^*_1 + \varepsilon^*_i + R^*_i, \]

where \( R^*_i \equiv (f(X_i, \theta^*, h^*) - W^{mn}(X_i)'\beta^*_1) \). The pseudo-true parameter value \( \beta^*_1 \) solves the moment condition \( E[W^{mn}(X_i)(Y_i - W^{mn}(X_i)'\beta^*_1)] = 0 \), and the semiparametric estimator \( \tilde{\beta}_1 \) solves its sample analog \( W'(Y - W'\tilde{\beta}_1)/n = 0 \).

The following theorem provides the divergence rate of the test statistic under the global alternative.

**Theorem 4.** Assume that series methods are used in estimation. Let \( \Omega^* = E[\varepsilon^*_i^2T_iT_i'] \). In the homoskedastic case, let \( \hat{\Omega} = \hat{\sigma}^2T'M_WT/n \). In the heteroskedastic case, let

\[ \hat{\Omega} = T'\hat{\Sigma}T/n - (T'\hat{\Sigma}W/n)(W'T\hat{\Sigma}W/n)^{-1}(W'T\hat{\Sigma}T/n), \]

21
where \( \tilde{\Sigma} = \text{diag}(\tilde{\varepsilon}_i^2) \), if the test is based on \( \xi_{HC,1} \), and let \( \hat{\Omega} = \tilde{T}'\tilde{\Sigma}\tilde{T}/n \), where \( \tilde{T} = M_WT \), if the test is based on \( \xi_{HC,2} \).

Suppose that
\[
\sup_{x \in X} |f(x, \theta^*; h^*) - W^{mn}(x)'\beta_1^*| \to 0,
\]

\[
||\hat{\Omega} - \Omega^*|| \overset{p}{\to} 0,
\]

the smallest eigenvalue of \( \Omega^* \) is bounded away from zero, \( m_n \to \infty \), \( r_n \to \infty \), \( r_n/n \to 0 \), \( E[\varepsilon_i'^T\Omega^*^{-1}E[T_i\varepsilon_i]] \to \Delta \), where \( \Delta \) is a constant. Then under homoskedasticity

\[
\frac{\sqrt{r_n}\xi - r_n}{\sqrt{2r_n}} \overset{p}{\to} \Delta/\sqrt{2},
\]

and under heteroskedasticity for \( j = 1, 2 \)

\[
\frac{\sqrt{r_n}\xi_{HC,j} - r_n}{\sqrt{2r_n}} \overset{p}{\to} \Delta/\sqrt{2}
\]

The divergence rate of the test statistic under the alternative is \( n/\sqrt{r_n} \). However, in most cases, the restricted semiparametric model is of lower dimension than the unrestricted nonparametric model, so that \( m_n/k_n \to 0 \) and \( r_n/k_n \to 1 \). Thus, the divergence rate in the semiparametric case discussed here is the same as in the parametric case in Hong and White (1995) and Donald et al. (2003), so the fact that the null hypothesis is semiparametric does not affect the global power of the test.

4.4 Behavior of Test Statistic under Local Alternatives

In this section, I analyze the power of the proposed test against local alternatives of the form

\[
H_{1n} : g_n(X_i) = f(X_i; \theta^*, h^*) + (r_1^{1/4}/n^{1/2})d_0(X_i),
\]

where \( d_0 \) is square integrable on \( X \), \( E[d_0(X_i)] = 0 \), and \( E[f(X_i, \theta^*, h^*)d_0(X_i)] = 0 \). The null hypothesis corresponds to \( d_0 = 0 \). I need to slightly modify the assumptions from the precious sections.

First, I impose the following condition. It is not primitive; however, an analogous result
for $\varepsilon$ instead of $d$ is derived in the proof of Theorem 1. Thus, this assumption is likely to hold under the primitive conditions I impose.

**Assumption 6.** Assume that $\|n^{-1}T'd\| = O_p(\sqrt{r_n/n})$ and $\|n^{-1}W'd\| = O_p(\sqrt{r_n/n})$.

Next, I impose the following assumption that requires the basis functions $T r_n(X_i)$ to approximate $d_0(X_i)$ sufficiently well. Note that because $d_0(X_i)$ is orthogonal to the semiparametric part of the model, $f(X_i, \theta^*, h^*)$, it is also orthogonal to $W_m n(X_i)$, and only $T r_n(X_i)$ is needed to approximate it.

**Assumption 7.** There exists $\alpha_d > 0$ such that

$$\sup_{x \in \mathcal{X}} |d_0(x) - T r_n(x)' \hat{\pi}| = O(r_n^{-\alpha_d})$$

Next, I apply the following result to obtain the convergence rates of series estimators of $d_0(x)$. These estimators are infeasible, because in practice $d_0(x)$ is unknown, but these convergence rates will be used in the proofs.

**Lemma 4** (Li and Racine (2007), Theorem 15.1). Let $d_i = d_0(X_i)$, $\hat{\pi} = (T'T)^{-1}T'd$, $\hat{d}(x) = T r_n(x)' \hat{\pi}$, and $\hat{d}_i = \hat{d}(X_i)$. Under Assumptions 1, 3, and 7, the following is true:

(a) $\sup_{x \in \mathcal{X}} |\hat{d}(x) - d_0(x)| = O_p \left( \zeta(r_n) (\sqrt{r_n/n} + r_n^{-\alpha_d}) \right)$

(b) $\frac{1}{n} \sum_{i=1}^{n} (d_i - \hat{d}_i)^2 = O_p(r_n/n + r_n^{-2\alpha_d})$

(c) $\int (\tilde{d}(x) - d_0(x))^2 dF(x) = O_p(r_n/n + r_n^{-2\alpha_d})$

Next, I impose an assumption that parallels Assumption 4.

**Assumption 8.** There exists $\alpha > 0$ such that

$$\sup_{x \in \mathcal{X}} |f(x, \theta^*, h^*) - W_m n(x)' \beta_1^*| = O(m_n^{-\alpha})$$
Now I impose an assumption that mimics the results of Lemma 2 and requires the semi-parametric estimators to converge to the pseudo-true values under the local alternative at the same rates as to the true values under the null.

**Assumption 9.** Let \( f^*(x) = f_n(x, \theta^*, h^*) \), \( f_i^* = f^*(X_i) \), \( \tilde{f}(x) = f(x, \tilde{\theta}, \tilde{h}) = W^{m_0}(x)'\tilde{\beta}_1 \), and \( \tilde{f}_i = \tilde{f}(X_i) \). For some \( \alpha > 0 \), the following conditions hold:

(a) \( \sup_{x \in X} |\tilde{f}(x) - f^*(x)| = O_p \left( \zeta(m_n)(\sqrt{m_n/n} + m_n^{-\alpha}) \right) \)

(b) \( \frac{1}{n} \sum_{i=1}^{n} (\tilde{f}_i - f_i^*)^2 = O_p(m_n/n + m_n^{-2\alpha}) \)

(c) \( \int (\tilde{f}(x) - f^*(x))^2 = O_p(m_n/n + m_n^{-2\alpha}) \)

The next result gives the behavior of the test under local alternatives. For simplicity, I only treat the homoskedastic case, but a similar result can be obtained in the heteroskedastic case at the expense of additional, less transparent, assumptions and tedious derivations.

**Theorem 5.** Assume that series methods are used to estimate the restricted model. Also assume that Assumptions 1, 2, 3, 6, 7, 8, and 9 are satisfied, \( \sigma^2(x) = \sigma^2 \), \( 0 < \sigma^2 < \infty \), for all \( x \in \mathcal{X} \), and rate conditions 4.1–4.5 hold.

Then

\[
 t_{r_n} = \frac{\xi - r_n}{\sqrt{2r_n}} \overset{d}{\rightarrow} N(\delta, 1),
\]

where \( \xi \) is as in Equation 3.7 and \( \delta = E[d_i^2]/\sigma^2 \).

## 5 Asymptotics with General Semiparametric Methods

One may expect that the same type of specification tests, based on the quadratic form \( \xi = \tilde{\epsilon}'P(\tilde{\sigma}^2P'P)^{-1}P'\tilde{\epsilon} \), may remain valid even if other semiparametric methods, such as kernels, are used to estimate the restricted model, or if the null model is not nested in the alternative. In this section I develop the asymptotic theory for the proposed test in this general case.
Because general semiparametric methods, unlike series methods, make it difficult to explicitly nest the restricted model in the unrestricted one, \( r_n \), the number of restrictions imposed by the null hypothesis on a general nonparametric model, is undefined. Thus, the only possible normalization in this case is \( \tau_n = k_n \), the number of series terms used to construct the test statistic. Moreover, because general semiparametric methods do not have a projection interpretation, it is problematic to directly account for the estimation variance and derive a degrees of freedom correction. Thus, I will have to impose stronger assumptions than in the series estimation case.

First, I impose a generic assumption on the convergence rate of the semiparametric conditional mean function estimator to the true conditional mean function. In many semiparametric models, \( \tilde{\theta} \) is \( \sqrt{n} \)-consistent and \( \tilde{h} \) obeys certain convergence rates, so semiparametric estimators often satisfy this assumption for appropriately chosen \( \eta_n \) and \( \psi_n \).

**Assumption 10.** Suppose that \( H_0 \) holds. Let \( f(x) = f(x, \theta_0, h_0) \), \( f_i = f(X_i) \), \( \tilde{f}(x) = f(x, \tilde{\theta}, \tilde{h}) \), and \( \tilde{f}_i = \tilde{f}(X_i) \). For some \( \eta_n \to 0 \) and \( \psi_n \to 0 \) as \( n \to \infty \), the following conditions hold:

(a) \( \sup_{x \in X} |\tilde{f}(x) - f(x)| = O_p(\eta_n) \)

(b) \( \frac{1}{n} \sum_{i=1}^n (\tilde{f}_i - f_i)^2 = O_p(\psi_n) \)

(c) \( \int (\tilde{f}(x) - f(x))^2 dF(x) = O_p(\psi_n) \)

Next, I impose a high-level assumption on how the error term in the model interacts with the estimation error:

**Assumption 11.** Assume that \( \sum_i \varepsilon_i (f_i - \tilde{f}_i) = o_p(k_n^{1/2}) \).

This assumption is used to bound the difference between the quadratic form in the semiparametric residuals, \( \varepsilon' P(\tilde{\sigma}^2 P' P)^{-1} P' \varepsilon \), and the quadratic form in the true regression errors, \( \varepsilon' P(\sigma^2 P' P)^{-1} P' \varepsilon \). While this is a straightforward task in the parametric case or in the semiparametric case with series estimation methods, this is a much more involved task in the
general case. The primitive conditions for this assumption may differ depending on a semi-parametric null model and a particular method used to estimate it. I plan to rigorously derive the primitive conditions for important special cases in future work. In the current version of the paper, I justify this assumption by providing an outline of the proof for leave-one-out kernel estimators in Remark A.2 in the Appendix.

As before, I separately treat the homoskedastic and heteroskedastic cases.

5.1 Homoskedastic Errors

The limiting distribution of the test statistic under the null is given by the next result:

**Theorem 6.** Assume that Assumptions 1, 2, 3, 10, and 11 are satisfied, \( \sigma^2(x) = \sigma^2 \), \( 0 < \sigma^2 < \infty \), for all \( x \in \mathcal{X} \), and the following rate conditions hold:

\begin{align*}
    \psi_n k_n^{3/2} &\to 0 \tag{5.1} \\
    \zeta(k_n) k_n / n^{1/2} &\to 0 \tag{5.2} \\
    n \psi_n / k_n^{1/2} &\to 0 \tag{5.3} \\
    \zeta(k_n)^2 / n^{1/2} &\to 0 \tag{5.4}
\end{align*}

Then

\[ t_{kn} = \frac{\xi - k_n}{\sqrt{2k_n}} \to N(0, 1), \tag{5.5} \]

where \( \xi \) is as in Equation 3.7.

**Remark 9** (Rate Conditions for General Semiparametric Estimators under Homoskedasticity). Conditions 5.1–5.4 are sufficient conditions for the result of the theorem to hold. Conditions 5.1 and 5.2 are used to bound the error from replacing \( \bar{\sigma}^2 P' P / n \) with \( \sigma^2 E[P_i P_i'] \). Condition 5.3 is used to bound the error from replacing \( \bar{\varepsilon} \) with \( \varepsilon \). Condition 5.4 is used to obtain the convergence in distribution result by applying Lemma A.7 in the Appendix.
Theorem 6 does not explicitly account for the estimation error \((\hat{\varepsilon} - \varepsilon)\); instead, it imposes the rate conditions that make it asymptotically negligible. As a result, the rate conditions imposed in Theorem 6 are substantially stronger than in Theorem 1. Intuitively, because Theorem 6 does not directly account for the form of the semiparametric estimator, it has to impose strong restrictions on its convergence rate, so that asymptotically the estimation error is negligible.

**Remark 10** (Examples of Permissible Choices of \(m_n\) and \(k_n\) under Homoskedasticity in General Case). Suppose that series methods are used for estimation but are paired with rate conditions 5.1–5.4 instead of 4.1–4.5. Then the rates given in Remark 6 will not work. It can be shown that, for splines, \(k_n = O(n^{2/7})\) would require \(m_n = o(n^7)\) and \(\alpha \geq 7\). For example, \(k_n = O(n^{2/7})\) and \(m_n = O(n^{2/15})\) are permissible in the sense of rate conditions 5.1–5.4 if \(\alpha \geq 7\). For power series, \(k_n = O(n^{2/9})\) would require \(m_n = o(n^9)\) and \(\alpha \geq 9\). For example, \(k_n = O(n^{2/7})\) and \(m_n = O(n^{2/19})\) are permissible in the sense of rate conditions 5.1–5.4 if \(\alpha \geq 9\).

An analog of Corollary 1 with \(k_n\) in place of \(r_n\) holds for general semiparametric methods, but I omit its formal statement for brevity.

**Corollary 3.** If the conditions of Theorem 6 hold, then

\[
P(\xi \geq \chi^2_{k_n}(1 - \alpha)) = P\left(\frac{\xi - k_n}{\sqrt{2k_n}} \geq \frac{\chi^2_{k_n}(1 - \alpha) - k_n}{\sqrt{2k_n}}\right) \to \alpha
\]

5.2 Heteroskedastic Errors

Now I deal with heteroskedastic errors. The limiting distribution of the heteroskedasticity robust test statistic under the null is given by the next result:

**Theorem 7.** Assume that Assumptions 1, 2, 3, 10, and 11 are satisfied, and the following
rate conditions hold:

\[ \psi_n k_n^{1/2} \zeta(k_n)^2 \to 0 \]  \hspace{1cm} \text{(5.6)}
\[ \zeta(k_n) k_n/n^{1/2} \to 0 \]  \hspace{1cm} \text{(5.7)}
\[ n\psi_n / k_n^{1/2} \to 0 \]  \hspace{1cm} \text{(5.8)}
\[ \zeta(k_n)^2 / n^{1/2} \to 0 \]  \hspace{1cm} \text{(5.9)}

Then

\[ t_{k_n,HC} = \frac{\xi_{HC,1} - k_n}{\sqrt{2k_n}} \overset{d}{\to} N(0, 1), \]  \hspace{1cm} \text{(5.10)}

where \( \xi_{HC,1} \) is as in Equation 3.8.

**Remark 11** (Rate Conditions for General Semiparametric Estimators under Heteroskedasticity). Conditions 5.6–5.9 are sufficient conditions for the result of the theorem to hold. Conditions 5.6 and 5.7 are used to bound the error from replacing \( \tilde{\sigma}^2 P' P/n \) with \( \sigma^2 E[P_i P_i'] \). Condition 5.8 is used to bound the error from replacing \( \tilde{\varepsilon} \) with \( \varepsilon \). Condition 5.9 is used to obtain the convergence in distribution result by applying Lemma A.7 in the Appendix.

An analog of Corollary 2 with \( k_n \) in place of \( r_n \) holds for general semiparametric methods, but I omit its formal statement for brevity.

### 5.3 Behavior of Test Statistic under a Global Alternative

In this section I consider the same global alternative as in Section 4. It is given by

\[ Y_i = g(X_i) + \varepsilon_i = f(X_i, \theta^*, h^*) + \varepsilon_i^*, \]

where \( \theta^* \) and \( h^* \) are pseudo-true parameter values and \( \varepsilon_i^* \equiv \varepsilon_i + (g(X_i) - f(X_i, \theta^*, h^*)) \equiv \varepsilon_i + d(X_i) \) is a composite error term. As before, let \( \tilde{\theta} \) and \( \tilde{h} \) be the semiparametric estimators.
Then the following result holds:

**Theorem 8.** Let $\Omega^* = E[\varepsilon_i^2 P_i P_i']$. In the homoskedastic case, let $\hat{\Omega} = \tilde{\sigma}^2 P' P / n$, and in the heteroskedastic case, let $\hat{\Omega} = P' \hat{\Sigma} P / n$, where $\hat{\Sigma} = \text{diag}(\tilde{\varepsilon}_i^2)$.

Suppose that $\sup_{x \in \mathcal{X}} |f(x, \hat{\theta}, \hat{h}) - f(x, \theta^*, h^*)| \overset{p}{\to} 0$, $||\hat{\Omega} - \Omega^*|| \overset{p}{\to} 0$, the smallest eigenvalue of $\Omega^*$ is bounded away from zero, $k_n \to \infty$, $k_n/n \to 0$, $E[\varepsilon_i^2 P_i P_i'] \Omega^* - 1 E[P_i \varepsilon_i^2] \to \Delta$, where $\Delta$ is a constant. Then under homoskedasticity

$$\frac{\sqrt{k_n} \xi - k_n}{n \sqrt{2k_n}} \overset{p}{\to} \frac{\Delta}{\sqrt{2}},$$

and under heteroskedasticity

$$\frac{\sqrt{k_n} \xi_{HC,1} - k_n}{n \sqrt{2k_n}} \overset{p}{\to} \frac{\Delta}{\sqrt{2}}.$$

The result of Theorem 8 is very similar to the result of Theorem 4. Given that in most cases of interest $m_n/k_n \to 0$, $r_n/k_n \to 1$, the divergence rates of the statistics $t_{r_n}$ and $t_{k_n}$, $n/\sqrt{r_n}$ and $n/\sqrt{k_n}$ respectively, are asymptotically equivalent. However, Theorem 4 achieves this result under weaker assumptions, as it only requires the series approximation error to go to zero without explicitly requiring the semiparametric estimator to be consistent. In contrast, Theorem 8 requires the semiparametric estimator $f(x, \hat{\theta}, \hat{h})$ to be consistent for the pseudo-true value $f(x, \theta^*, h^*)$.

### 5.4 Behavior of Test Statistic under Local Alternatives

In this section I consider a slightly different family of local alternatives as compared to Section 4. It is given by

$$H_{1n} : g_n(X_i) = f(X_i, \theta^*, h^*) + (k_n^{1/4}/n^{1/2})d_0(X_i),$$

where $d_0$ is square integrable on $\mathcal{X}$, $E[d_0(X_i)] = 0$, and $E[f(X_i, \theta^*, h^*)d_0(X_i)] = 0$. The null hypothesis corresponds to $d_0 = 0$. I impose the following assumptions.
First, I assume that the unknown function \( d(x) \) can be approximated by a finite series expansion sufficiently well.

**Assumption 12.** There exists \( \alpha_d > 0 \) such that

\[
\sup_{x \in \mathcal{X}} |d_0(x) - P^{k_n}(x)'\pi| = O(k_n^{-\alpha_d})
\]

Next, I apply the following result to obtain the convergence rates of series estimators of \( d_0(x) \). These estimators are infeasible, because in practice \( d_0(x) \) is unknown, but these convergence rates will be used in the proofs.

**Lemma 5** (Li and Racine (2007), Theorem 15.1). Let \( d_i = d_0(X_i), \hat{\pi} = (P'P)^{-1}P'd, \hat{d}(x) = P^{k_n}(x)'\hat{\pi}, \) and \( \hat{d}_i = \hat{d}(X_i) \). Under Assumptions 1, 3, and 12, the following is true:

(a) \( \sup_{x \in \mathcal{X}} |\hat{d}(x) - d_0(x)| = O_p(\zeta(k_n)(\sqrt{k_n/n + k_n^{-\alpha_d}})) \)

(b) \( \frac{1}{n} \sum_{i=1}^{n} (\hat{d}_i - d_i)^2 = O_p(k_n/n + k_n^{-2\alpha_d}) \)

(c) \( \int (\hat{d}(x) - d_0(x))^2 dF(x) = O_p(k_n/n + k_n^{-2\alpha_d}) \)

I modify Assumption 10 as follows:

**Assumption 13.** Let \( f^*(x) = f(x, \theta^*, h^*), f_i^* = f^*(X_i), \tilde{f}(x) = f(x, \tilde{\theta}, \tilde{h}), \tilde{f}_i = \tilde{f}(X_i) \). For some \( \eta_n \to 0 \) and \( \psi_n \to 0 \) as \( n \to \infty \), the following conditions hold:

(a) \( \sup_{x \in \mathcal{X}} |\tilde{f}(x) - f^*(x)| = O_p(\eta_n) \)

(b) \( \frac{1}{n} \sum_{i=1}^{n} (\tilde{f}_i - f_i^*)^2 = O_p(\psi_n) \)

(c) \( \int (\tilde{f}(x) - f^*(x))^2 dF(x) = O_p(\psi_n) \)

Assumption 13 requires the semiparametric estimators to converge to the pseudo-true values fast enough. Next, I impose a high-level assumption on how the error term in the model interacts with the estimation error. It parallels Assumption 11 and is discussed in Remark A.2 in the Appendix.
Assumption 14. Assume that

\[ \sum_i \varepsilon_i(f_i^* - \tilde{f}_i^*) = o_p(k_n^{1/2}) \]

The next result gives the behavior of the test under local alternatives:

Theorem 9. Assume that Assumptions 1, 2, 3, 12, 13, and 14 are satisfied, \( \sigma^2(x) = \sigma^2 \), \( 0 < \sigma^2 < \infty \), for all \( x \in \mathcal{X} \), and rate conditions 5.1–5.4 hold.

Then

\[ t_{kn} = \frac{\xi - k_n}{\sqrt{2k_n}} \xrightarrow{d} N(\delta, 1), \]

where \( \xi \) is as in Equation 3.7 and \( \delta = E[d_i^2]/\sigma^2 \).

5.5 Asymptotic Theory: Summary

To summarize, there are two versions of the proposed test that are asymptotically equivalent. One relies on series estimation method to define the number of restrictions imposed by the semiparametric model on a general nonparametric model and uses the projection interpretation of series estimators to derive the degrees of freedom correction. Another one uses general estimation methods, without imposing the series structure and defining the number of restrictions. While the former approach restricts the class of models to which the test applies to the models that can be estimated by series, it allows me to obtain refined asymptotic results under mild rate conditions. In contrast, the latter approach is applicable to a broader range of models, but it results in cruder asymptotic analysis and requires stronger assumptions.

These two approaches differ because they differently cope with a key step in the proof, going from the semiparametric regression residuals \( \tilde{\varepsilon} \) to the true errors \( \varepsilon \). The series-based approach relies on the projection property of series estimators to eliminate the estimation variance and hence only has to deal with the approximation bias. Specifically, it uses the equality \( \tilde{\varepsilon} = M_W \varepsilon + M_W R \), applies a central limit theorem for \( U \)-statistics to the quadratic...
form in $M_W \varepsilon$, and bounds the remainder terms by requiring the approximation error $R$ to be small.

The general approach does not impose any special structure on the model residuals and thus has to deal with both the bias and variance of semiparametric estimators. Specifically, it uses the equality $\bar{\varepsilon} = \varepsilon + (g - \tilde{g})$, which can be written in a series form\(^8\) as $\bar{\varepsilon} = \varepsilon + R + W'(\beta_1 - \tilde{\beta}_1)$, applies a central limit theorem for $U$-statistics to the quadratic form in $\varepsilon$, and bounds the remainder terms by requiring both the bias term $R$ and variance term $W'(\beta_1 - \tilde{\beta}_1)$ to be small. This, in turn, requires the semiparametric estimates of the restricted model to converge to the true values very fast and leads to restrictive rate conditions.

6 Simulations

There are several variants of the test, depending on whether series methods are used in estimation, what normalization is used, and what limiting distribution is used. In this section I study the finite sample performance of different variants of the proposed tests in simulations that mimic the cost estimation example in Section 2. The Monte Carlo studies have several goals: first, to compare the tests based on the $\chi^2$ and normal asymptotic approximations; second, to compare two normalizations of the test statistic; third, to investigate the test behavior with different sample sizes and different numbers of series terms; finally, to study the use of multiple alternatives as a tool to improve the test power in particular directions.

I assume that the researcher wants to estimate returns to scale (or their inverse $\delta(Z_i)$) in the model

$$
\left(\ln TC_i - \ln P_{Li}\right) \approx C_1(Z_i) + \gamma(Z_i) \left(\ln P_{Ki} - \ln P_{Li}\right) + \delta(Z_i) \ln Q_i + \varepsilon_i
$$

and wants to test whether this model is adequate. From now on, I work with the rearranged

---

\(^8\)Even though series methods do not have to be used in estimation, it is convenient to write the model in a series form to facilitate the comparison between two approaches.
model $\widetilde{TC}_i = C_1(Z_i) + \gamma(Z_i)\widetilde{P}_i + \delta(Z_i)\widetilde{Q}_i + \varepsilon_i$.

In the subsequent analysis, I will test the specification of the semiparametric conditional mean model against the nonparametric alternative using the proposed specification test. The varying coefficient null hypothesis is given by

$$H_0^{\text{SP}} : P \left( E[\widetilde{TC}_i | \widetilde{P}_i, \widetilde{Q}_i, Z_i] = C_1(Z_i) + \gamma(Z_i)\widetilde{P}_i + \delta(Z_i)\widetilde{Q}_i \right) = 1$$

for some $C_1(Z_i), \gamma(Z_i), \delta(Z_i)$, while the alternative is

$$H_1 : P \left( E[\widetilde{TC}_i | \widetilde{P}_i, \widetilde{Q}_i, Z_i] \neq C_1(Z_i) + \gamma(Z_i)\widetilde{P}_i + \delta(Z_i)\widetilde{Q}_i \right) > 0$$

I obtain $(\ln P_{Li}, \ln P_{K_i}, \ln Q_i, Z_i)$ as follows. First, I draw four uniform random variables $V_{j,i} \sim U[0,2], j = 1, \ldots, 4$. Then I set

$$\ln P_{K_i} = 1.25 + 0.75(0.7V_{1,i} + 0.15V_{2,i} + 0.1V_{3,i} + 0.05V_{4,i})$$
$$\ln P_{L_i} = 1.4 + 0.6(0.1V_{1,i} + 0.75V_{2,i} + 0.05V_{3,i} + 0.1V_{4,i})$$
$$\ln Q_i = 2 + (0.05V_{1,i} + 0.1V_{2,i} + 0.65V_{3,i} + 0.2V_{4,i})$$
$$Z_i = 1.25 + 0.75(0.025V_{1,i} + 0.025V_{2,i} + 0.15V_{3,i} + 0.8V_{4,i})$$

This way of generating data ensures that all variables are bounded and at the same time correlated with one another. Table 1 shows the pairwise correlations between the variables. Then I simulate data from two data generating processes:

1. Semiparametric varying coefficient, which corresponds to $H_0^{\text{SP}}$:

$$\widetilde{TC}_i = C_1(Z_i) + \gamma(Z_i)\widetilde{P}_i + \delta(Z_i)\widetilde{Q}_i + \varepsilon_i,$$  \hspace{1cm} (6.1)
where
\[
C_1(z) = 20 - 3(z - 2) + 7(z - 2)^2 - 12(z - 2)^3
\]
\[
\alpha(z) = 0.35 + 0.05 \exp(2(z - 2)) + 0.05(z - 2)^2
\]
\[
\beta(z) = 0.3 + 0.075 \exp(2(z - 2)) + 0.025(z - 2)^2
\]
\[
\gamma(z) = \frac{\beta(z)}{\alpha(z) + \beta(z)}, \quad \delta(z) = \frac{1}{\alpha(z) + \beta(z)}
\]

Figure 1 plots the coefficient functions $C_1(z)$, $\gamma(z)$, and $\delta(z)$. This choice of functional forms is motivated by the fact that it results in reasonable and economically interesting parameter values. For instance, returns to scale lie roughly between 0.7 and 1.2, with more than 90% of the firms having decreasing returns to scale and less than 10% having increasing returns to scale. The firms with increasing returns to scale are those with the highest R&D spending.

2. Nonparametric, which corresponds to $H_1$:

\[
\widehat{T}C_i = C_1(Z_i) + \gamma(Z_i)\widetilde{P}_i + \delta(Z_i)\widetilde{Q}_i + \lambda_1 \cos(2(\widetilde{Q}_i - 3)) + \lambda_2 \sin(3\widetilde{P}_i) + \varepsilon_i, \quad (6.2)
\]

where $C_1(z)$, $\gamma(z)$, and $\delta(z)$ are the same as before and $(\lambda_1, \lambda_2) = (0.7, 0.3)$.

Figures 3 and 4 plot the relationship between the total costs and prices or output under the null and under the alternative to illustrate how the nonparametric model compares to the semiparametric one.

Though the deviation from the null is nontrivial, it explains only a small fraction of the dependent variable variation. The variance of the dependent variable $\widehat{T}C_i$ is around 8.45, while the variance of the semiparametric part $C_1(Z_i) + \gamma(Z_i)\widetilde{P}_i + \delta(Z_i)\widetilde{Q}_i$ is around 6.15 and the variance of the deviation from the null $\lambda_1 \cos(2(\widetilde{Q}_i - 3)) + \lambda_2 \sin(3\widetilde{P}_i)$ is around 0.07.
Because the parametric model and the semiparametric models are restricted versions of the nonparametric model, I will use the proposed test to check if they are correctly specified. But first, I demonstrate why specification testing is important.

Suppose that the true model is the varying coefficient model above, but the researcher estimates the parametric model:

$$\widetilde{TC}_i = C_1 + \gamma \widetilde{P}_i + \delta \widetilde{Q}_i + \theta Z_i + \varepsilon_i,$$

or the parametric model with interactions:

$$\widetilde{TC}_i = C_1 + (\gamma_0 + \gamma_1 Z_i) \widetilde{P}_i + (\delta_0 + \delta_1 Z_i) \widetilde{Q}_i + \theta Z_i + \varepsilon_i$$

Figure 2 shows the estimates of $\delta(z)$, the inverse of returns to scale, from OLS, OLS with interactions, and the varying coefficient model, as well as the true function $\delta(z)$. As we can see from the figure, both ordinary OLS, and OLS with interactions, yield very misleading estimates of $\delta(z)$ and thus returns to scale. The true $\delta(z)$ is greater than 1 and slowly decreasing over most of its support, meaning that most firms in the sample have decreasing returns to scale. The OLS results imply that it $\delta(z)$ is smaller than 1 and hence returns to scale are increasing. The OLS with interactions results imply that $\delta(z)$ is increasing in R&D spending, so not only the magnitude of returns to scale is incorrect, but also the pattern of their dependence on R&D spending. If the researcher wants to evaluate a possible merger of two firms, OLS or OLS with interactions estimates will likely lead to very misleading counterfactuals. Thus, using a plausible model is important, and I will show next that the proposed test helps determine if a given model is correctly specified.

Before I move on and discuss the behavior of the proposed test in finite samples, I need to make two choices in order to implement the test. First, I need to choose the family of basis functions; second, choose the number of series terms in the restricted and unrestricted models. I use power series because of their simplicity, but I do not have any data-driven
methods to choose tuning parameters.

The choice of tuning parameters presents a big practical challenge in implementing the proposed test, as well as many other specification tests. If one is interested only in estimating the null or the alternative models, then certain data-driven methods, such as cross-validation, can be used to select the number of series terms. For details, see, e.g., Section 15.2 in Li and Racine (2007). However, it is not clear how these data-driven procedures may affect the proposed test and whether using them would lead to any optimality results in testing. I leave the choice of tuning parameters for the proposed test for future research, and in my simulations choose tuning parameters according to the rate conditions imposed in Section 4. However, the choices I make are still arbitrary, because I can multiply \( m_n \) or \( k_n \) by any constant and still satisfy the rate conditions, which are asymptotic in nature.

The number of terms in the parametric model is \( m_n^{OLS} = 4 \), the number of regressors plus one. The number of terms in the series expansions of the unknown coefficient functions \( \gamma(z) \) and \( \delta(z) \) in the varying coefficient model is \( l_n = \lfloor 2.5n^{0.12} \rfloor \), which leads to \( m_n^{VCM} = 3l_n \). In the nonparametric model, I include \( j_n = \lfloor 3n^{0.06} \rfloor \) series terms in \( \tilde{P}_i \) and \( \tilde{Q}_i \) each, which leads to \( k_n = l_n j_n^2 \). The number of restrictions is given by \( r_n^{OLS} = k_n - m_n^{OLS} \) and \( r_n^{VCM} = k_n - m_n^{VCM} \) correspondingly.

These choices lead to \( m_n = 15 \), \( r_n = 65 \), and \( k_n = 80 \) when \( n = 1,000 \), and to \( m_n = 18 \), \( r_n = 132 \), and \( k_n = 150 \) when \( n = 5,000 \). Because the behavior of the test statistic depends both on the sample size and the number of series terms, I also consider an intermediate setup with \( n = 5,000 \) but \( m_n = 15 \), \( r_n = 65 \), and \( k_n = 80 \) to separate these two effects. In other words, I first fix the number of series terms and increase the sample size, and then fix the sample size and increase the number of series terms. Throughout my analysis, I use \( B = 2,000 \) simulation draws.

---

\(^{9}\)For a discussion of regularization parameters choice in the context of kernel-based tests, see a review by Sperlich (2014).
6.1 Homoskedastic Errors

First, I investigate the performance of the test when the errors are homoskedastic. I use the $\xi$ test statistic directly:

$$\xi = \tilde{\varepsilon}'P(\tilde{\sigma}^2P'P)^{-1}P'\tilde{\varepsilon} \sim \chi^2_{\tau_n},$$

as well as the $t$ statistic:

$$t_{\tau_n} = \frac{\xi - \tau_n}{\sqrt{2\tau_n}} \sim N(0,1),$$

with $\tau_n = k_n$ and $\tau_n = r_n$.

I consider two sample sizes, $n = 1,000$ and $n = 5,000$. The errors are normally distributed: $\varepsilon_i \sim$ i.i.d. $N(0, 2.25)$. With this choice of the distribution of $\varepsilon_i$, the semiparametric part of the model explains about 70% of the dependent variable variance, while the errors account for the remaining 30%. I repeated my analysis with centered exponential errors, which have an asymmetric distribution, and found no substantial difference from the normal case. The results for exponential errors are omitted for brevity.

Table 2 shows the size and power of the nominal 5% level test for three combinations of the sample size and the number of series terms: Setup 1 with $(n = 1,000, k_n = 80)$, Setup 2 with $(n = 5,000, k_n = 80)$, and Setup 3 with $(n = 5,000, k_n = 150)$, and two normalizations of the test statistic: $\tau_n = k_n$ and $\tau_n = r_n$. As we can see from the table, the test with the former normalization is severely undersized, with the size being below 1% even with $n = 5,000$ observations. In contrast, the size of the test with the latter normalization is close to the nominal level of 5%. Moreover, in all three settings, the test based on the latter normalization has better power against the semiparametric varying coefficient null model when the nonparametric model is true.

As far as the choice between the $\xi$ test statistic and the normalized $t$ test statistic is concerned, the latter typically leads to slightly higher rejection probabilities, meaning that the test based on the $t$ statistic has marginally better power but is slightly oversized.

Finally, we can see that increasing the sample size from $n = 1,000$ to $n = 5,000$ while
keeping the series term constant at \( k_n = 80 \) brings the size of the test closer to the nominal level and greatly increases power, while increasing the number of series terms from \( k_n = 80 \) to \( k_n = 150 \) with the sample size of \( n = 5,000 \) has almost no effect on the size but reduces the power of the test. This observation calls for a data-driven method to choose the number of series terms. As my simulations show, including too many series terms when they are not necessary can worsen the test performance; however, including too few series terms can make it difficult to detect certain alternatives,\(^{10}\) which will also lead to low power. Thus, having a data-driven method that would help balance these two considerations and determine the appropriate number of series terms to include is important, and I plan to study this question in future work.

Next, I plot the simulated distribution of the test statistic in different Monte Carlo settings. Figure 5 plots the distribution of the \( t \) test statistic under the null for the three setups I study. As we can see, in all three settings the simulated distribution of the \( t_{\tau_n} \) test statistic is very close to the standard normal. In contrast, the simulated distribution of the \( t_{k_n} \) test statistic is off to the left as compared to the standard normal. This illustrates the importance of the proper normalization, which explicitly accounts for the estimation variance, and helps explain why the tests based on the normalization \( \tau_n = k_n \) are severely undersized.

### 6.2 Heteroskedastic Errors

In this subsection I investigate the performance of the test when the errors are heteroskedastic. The form of heteroskedasticity is \( \varepsilon_i \sim \text{i.n.i.d. } N(0, 0.015 \exp(\tilde{Q}_i + Z_i)) \). Figure 6 illustrates this form of heteroskedasticity. I will consider the test statistic based on

\[
\xi_{HC,1} = \varepsilon' P (P' \tilde{\Sigma} P)^{-1} P' \tilde{\varepsilon},
\]

\(^{10}\)The series-based test cannot detect alternatives that are orthogonal to all series terms used to form the test statistic. The more series terms are used to construct the test, the fewer such alternatives exist.
where \( \hat{\Sigma} = diag(\hat{\varepsilon}_i^2) \), as well as the alternative test statistic based on

\[
\xi_{HC,2} = \hat{\varepsilon}' \bar{T} (\bar{T}' \hat{\Sigma} \bar{T})^{-1} \bar{T}' \hat{\varepsilon},
\]

where \( \bar{T} = M_W T \). Unlike the former test statistic, the latter can be computed only when the null model is nested in the alternative and is estimated using series methods. To make it easier to distinguish between two versions of the test, I will refer to the tests based on \( \xi_{HC,1} \) as unadjusted, and to the tests based on \( \xi_{HC,2} \) as adjusted. This is because the latter test takes into account the use of the series methods in estimation and adjusts the variance matrix accordingly.

The normalized \( t \) statistic is given by

\[
t_{\tau_n, HC,j} = \frac{\xi_{HC,j} - \tau_n}{\sqrt{2\tau_n}},
\]

for \( \tau_n = r_n \) and \( j = 1, 2 \). I do not report the results for \( \tau_n = k_n \), because they are similar to the previous section, i.e. the test based on \( \tau_n = k_n \) is severely undersized and low-powered. Instead of comparing the normalizations \( \tau_n = r_n \) and \( \tau_n = k_n \), I compare the feasible test statistics \( t_{\tau_n, HC,j} \) with the infeasible test statistics

\[
t_{\tau_n, HC,j, inf} = \frac{\xi_{HC,j, inf} - \tau_n}{\sqrt{2r_n}},
\]

where \( j = 1, 2 \) and

\[
\xi_{HC,1, inf} = \hat{\varepsilon}' P (P' \Sigma P)^{-1} P' \hat{\varepsilon},
\]

\[
\xi_{HC,2, inf} = \hat{\varepsilon}' \bar{T} (\bar{T}' \Sigma \bar{T})^{-1} \bar{T}' \hat{\varepsilon},
\]

where \( \Sigma = diag(E[\varepsilon_i^2|\bar{Q}_i, \bar{P}_i, Z_i]) = diag(0.015 \exp(\bar{Q}_i + Z_i)) \). I make this comparison in order to understand whether the behavior of the test under heteroskedastic errors is driven
by heteroskedasticity itself or by using the estimated variance-covariance matrix instead of
the true one. Because of higher computational burden associated with the heteroskedasticity
robust test statistic, I reduce the number of simulation draws from $B = 2,000$ to $B = 1,000$.

Table 3 shows the size and power of the nominal 5% level test for the three setups discussed
above and two test statistics, feasible $t_{rn,HC,1}$ and infeasible $t_{rn,HC,1,inf}$. Figure 7 plots their
simulated distribution under the null for these three setups. Table 4 shows the size and power
of the nominal 5% level test for the three setups discussed above and two test statistics,
feasible $t_{rn,HC,2}$ and infeasible $t_{rn,HC,2,inf}$. Figure 8 plots their simulated distribution under
the null for these three setups.

As we can see, the feasible unadjusted test is oversized in all three setups. The size of the
test based on the $\xi$ test statistic is closer to the nominal level than the size of the test based
on the $t$ test statistic. At the same time, the feasible adjusted test is undersized in all three
setups. Now the size of the test based on the $t$ test statistic is closer to the nominal level than
the size of the test based on the $\xi$ test statistic. Not surprisingly, because the simulated size
of the adjusted test is lower than that of the unadjusted test, the former test is less powerful.
The difference in power between these versions of the test becomes less pronounced as the
sample size grows.

While the simulated distribution of the adjusted test statistic $t_{rn,HC,2}$ somewhat differs
from the standard normal, it is pretty symmetric around zero in all three setups. At the
same time, the simulated distribution of the unadjusted test statistic $t_{rn,HC,1}$ is somewhat
shifted to the right, which explains why the test based on it overrejects.

As for the comparison between the feasible and infeasible test statistics, their simulated
distributions look quite different for the unadjusted test statistic and somewhat less different
for the adjusted one. In both cases, the infeasible test statistic distribution is closer to
the standard normal. As for the rejection probabilities, while the unadjusted infeasible test
noticeably overrejects, the size of the adjusted infeasible test is very close to the nominal
level.
It appears that in the case of the unadjusted test, it is heteroskedasticity itself, and not the variance-covariance matrix estimation, that causes the size distortion. Conversely, in the case of the adjusted test, heteroskedasticity does not seem to lead to any size distortion when the true variance-covariance matrix is known, so the feasible test is conservative due to the variance-covariance matrix estimation.

Based on the evidence presented in this section, I recommend using the adjusted test statistic instead of the unadjusted one. The adjusted test does not overreject, and its size is closer to the nominal level. Moreover, the size distortion of the adjusted test seems to be caused by the variance-covariance matrix estimation, and this distortion may disappear in large samples, when the variance matrix can be estimated more precisely. At the same time, the unadjusted test statistic is oversized even when the form of heteroskedasticity is known, so the size distortion might remain severe even with large sample sizes. Better ability of the adjusted test to control the size comes at the cost of lower power, but the difference in power between the adjusted and unadjusted tests decreases as the sample size grows.

### 6.3 Test Behavior under Local Alternatives

In this section, I study the behavior of the proposed test under local alternatives. More specifically, I use the same setup as before, but the true DGP changes with the sample size:

\[
\begin{align*}
\tilde{T}C_i &= C_1(Z_i) + \gamma(Z_i)\tilde{p}_i + \delta(Z_i)\tilde{Q}_i + (r_{n^1/4}/n^{1/2}) d_0(\tilde{p}_i, \tilde{Q}_i) + \varepsilon_i, \\
d_0(\tilde{p}_i, \tilde{Q}_i) &= \lambda_1 \cos(2(\tilde{Q}_i - 3)) + \lambda_2 \sin(3\tilde{P}_i),
\end{align*}
\]

where \(C_1(z), \gamma(z),\) and \(\delta(z)\) are the same as before and \((\lambda_1, \lambda_2) = (10, 4)\).

In other words, the true model is nonparametric, but it approaches the semiparametric varying coefficient model at the rate \((r_n^{1/4}/n^{1/2})\) as the sample size grows. Based on the result of Theorem 5, the rejection probability should remain the same as the sample size changes.

I gradually increase the sample size from \(n = 500\) to \(n = 25,000\) and compute the
simulated rejection probabilities for the nominal 5% level tests based on the $t_{rn}$ test statistic. Table 5 shows the results. Figure 9 plots the distribution of the $t$ test statistic under local alternatives for $n = 1,000$ and $n = 10,000$.

As we can see, the rejection probabilities for the test based on the $t_{rn}$ statistic lie between 36% and 39% for all sample sizes considered, and the simulated distributions of the $t_{rn}$ statistic for $n = 1,000$ and $n = 10,000$ look very similar. These findings are consistent with the theoretical results established in Section 4.4.

### 6.4 Multiple Alternatives and Bonferroni Correction

If a researcher wants to estimate a model that can be nested in an expanding set of alternatives (e.g. a parametric model can be nested in a semiparametric partially linear model or in a nonparametric model), it is possible to test it against multiple alternatives simultaneously while using the Bonferroni correction.

Namely, if the null model is fully parametric:

$$ H_0^{PL}: P(E[Y_i|X_i] = X_{1i}'\beta_1 + h(X_{2i})) = 1 \text{ for some } \beta_1 \in R^{d_2}, h(\cdot): R^{d_2} \to R, $$

where $X_i = (X_{1i}', X_{2i}')'$, a researcher may consider a semiparametric varying coefficient and a fully nonparametric alternatives simultaneously:

$$ H_1^{VC}: E[Y_i|X_i] = \tilde{X}_{1i}'\beta(X_{2i}) \text{ for some } \beta(\cdot): R^{d_2} \to R^{d_2+1}; $$

$$ H_1^{NP}: E[Y_i|X_i] = g(X_i) \text{ for some } g(\cdot): R^{d_2} \to R. $$

where $\tilde{X}_{1i} = (1, X_{1i}')'$.

Intuitively, using the former alternative may improve power if the true model turns out to be close to a varying coefficient one, while the latter ensures consistency against a general nonparametric alternative. The test statistic should be modified accordingly, by including in $P^{k_n}(X_i)$ only those power series terms that are present under the alternative. For example,
alternative $H_1^{VC}$ does not allow higher powers of $X_{1i}$ to enter the model, so they should be removed from $P^{kn}(X_i)$ when constructing the test statistic for $H_0^{PL}$ against $H_1^{VC}$.

Because now several hypotheses tests are done simultaneously, the Bonferroni correction is needed to control size. Namely, the nominal significance level for each individual test should be $\alpha/2$ (or $\alpha/T$, if there are $T$ tests) instead of $\alpha$. The resulting overall test rejects the null if at least one individual test rejects the null at the $\alpha/2$ level.

Next, I simulate data from three data generating processes:

1. Semiparametric partially linear, which corresponds to $H_0^{PL}$:

$$\tilde{TC}_i = C_1(Z_i) + \gamma \tilde{P}_i + \delta \tilde{Q}_i + \epsilon_i,$$

where $(\gamma, \delta) = (0.5, 1.25)$ and

$$C_1(z) = 20 - 3(z - 2) + 7(z - 2)^2 - 12(z - 2)^3.$$

This DGP is new as compared to the ones considered previously.

2. Semiparametric varying coefficient, which corresponds to $H_1^{VC}$:

$$\tilde{TC}_i = C_1(Z_i) + \gamma(Z_i) \tilde{P}_i + \delta(Z_i) \tilde{Q}_i + \epsilon_i,$$

where $C_1(z)$ is the same as before and

$$\alpha(z) = 0.5 - 0.3 \sin(2(z - 1)) + 0.075 \log(z)$$

$$\beta(z) = 0.3 - 0.1 \cos(1.5(z - 2)) + 0.2 \log(z)$$

$$\gamma(z) = \frac{\beta(z)}{\alpha(z) + \beta(z)}, \delta(z) = \frac{1}{\alpha(z) + \beta(z)}.$$

This DGP resembles the one in equation 6.1 but makes it easier to distinguish between the partially linear and varying coefficient models.
3. Nonparametric, which corresponds to $H_{1}^{NP}$:

$$\tilde{T}C_{i} = C_{1}(Z_{i}) + \gamma \tilde{P}_{i} + \delta \tilde{Q}_{i} + \theta Z_{i} + \lambda \cos(2(\tilde{Q}_{i} - 3)) \sin(3\tilde{P}_{i}) + \varepsilon_{i},$$

where $(\gamma, \delta, \lambda) = (0.5, 1.25, 0.7)$ and $C_{1}(z)$ is the same as before. This DGP resembles the one in equation 6.2 but removes the varying coefficient component from the nonparametric model and modifies the deviation from the null.

Because in all previous simulations the normalization $\tau_{n} = r_{n}$ performs better than $\tau_{n} = k_{n}$, in this section I only consider the former normalization. Table 6 presents the size and power of three nominal 5% level tests. The first one tests the semiparametric partially linear null $H_{0}^{PL}$ against the nonparametric alternative $H_{1}^{NP}$ at the nominal 5% level. The second one tests the semiparametric partially linear null $H_{0}^{PL}$ against the semiparametric varying coefficient alternative $H_{1}^{VC}$ at the nominal 5% level. The third one tests the semiparametric varying coefficient null $H_{0}^{VC}$ against both the nonparametric alternative $H_{1}^{NP}$ and semiparametric varying coefficient alternative $H_{1}^{VC}$ at the 2.5% nominal level each, which by the Bonferroni inequality means that the nominal size of the joint test is no greater than 5%.

I consider two setups: $n = 1,000$ with $k_{n} = 80$ and $n = 5,000$ with $k_{n} = 150$.

As we can see, the test of the partially linear null $H_{0}^{PL}$ against the varying coefficient alternative $H_{1}^{VC}$ has excellent power if the true model is semiparametric varying coefficient, but limited power if the true model is nonparametric. Intuitively, in the former case the series terms used to construct the test can approximate the alternative very well, and there are no irrelevant terms included, which results in a high powered test. In the latter case, the series terms used to construct the test cannot approximate the alternative well, which results in a low powered test.

Conversely, the test of the partially linear null $H_{0}^{PL}$ against the nonparametric alternative $H_{1}^{NP}$ has limited power if the true model is semiparametric varying coefficient, but very good power if the true model is nonparametric. Intuitively, in the former case the series terms
used to construct the test can approximate the alternative very well, but there are many irrelevant terms included, which reduces the power of the test. In the latter case, the series terms used to construct the test can approximate the alternative well, which results in a high powered test.

The Bonferroni procedure combines the best features of the two individual tests. First, it controls the size of the test fairly well. Second, it significantly improves power as compared to the test of the partially linear null $H_{0}^{PL}$ against the nonparametric alternative $H_{1}^{NP}$ when the true model is semiparametric varying coefficient. Third, it loses little power as compared to the test of the partially linear null $H_{0}^{PL}$ against the nonparametric alternative $H_{1}^{NP}$ when the true model is nonparametric. Thus, when several alternative models are available, it may be beneficial to test the null model against different alternatives simultaneously to improve the power against particular alternatives while using the Bonferroni adjustment to control the test size.

A possible direction for future research is to develop a step-up or step-down model selection procedure that would not only test a given model against one or several alternatives, but also would allow the researcher to consider a different, more general, null model if the original null is rejected.

7 Empirical Example

In this section, I apply the proposed test to the Canadian household gasoline consumption data from Yatchew and No (2001). They estimate gasoline demand ($y$), measured as the logarithm of the total distance driven in given month, as a function of the logarithm of the gasoline price ($PRICE$), the logarithm of the household income ($INCOME$), the logarithm of the age of the primary driver of a car ($AGE$), and demographics ($z$), which include the logarithm of the number of drivers in a household ($DRIVERS$), the logarithm of the household size ($HHSIZE$), an urban dummy, and a dummy for singles under 35 years old.
Yatchew and No (2001) use several demand models, including semiparametric specifications. They use differencing (see Yatchew (1997) and Yatchew (1998) for details) to estimate semiparametric models. The relevance of semiparametric models in gasoline demand estimation was first pointed out by Hausman and Newey (1995) and Schmalensee and Stoker (1999). Yatchew and No (2001) follow these papers in using semiparametric specifications and pay special attention to specification testing. However, they only test semiparametric specifications against parametric ones, while I can use series methods to estimate semiparametric specifications and implement the proposed series-based specification test to assess their validity as compared to a general nonparametric model.

I focus on the model that is nonparametric in $AGE$ but parametric in $PRICE$ and $INCOME$ (roughly corresponds to Model 3.4 in Yatchew and No (2001)):

$$y = \alpha_1 PRICE + \alpha_2 INCOME + g(AGE) + z'\beta + \varepsilon$$  \hspace{1cm} (7.1)$$

In this model, the relationship between gasoline demand, price, and household income has a familiar log-log form\textsuperscript{11} that could be derived from a Cobb-Douglas utility function. The age of the primary driver enters the model nonparametrically to allow for possible nonlinearities, while the remaining demographics enter the model linearly. Next, I investigate whether demand model 7.1 is flexible enough as compared to a more general model.

In order to apply my test, I need to choose the series functions $P^{k_n}(x)$ or, equivalently, define a nonparametric alternative that can be approximated by these series functions. Ideally, I would want to use a fully nonparametric alternative $y = h(PRICE, AGE, INCOME, z) + \varepsilon$. However, this is impractical in the current setting: the dataset from Yatchew and No (2001) contains 12 monthly dummies, an urban dummy, and a dummy for singles under 35 years old. The fully nonparametric alternative would require to completely saturate the model with the dummies, i.e. interact all series terms in continuous regressors with a full set of

\textsuperscript{11}Recall that $y$, $PRICE$, and $INCOME$ are the logarithms of gasoline demand, price, and household income correspondingly.
dummies. This would be equivalent to dividing the dataset into $12 \cdot 2 \cdot 2 = 48$ bins and estimating the model within each bin separately. Given that the total number of observations is 6,230, this would leave me with about 125 observations per bin on average and would make semiparametric estimation and testing problematic.

I choose a different approach to deal with the dummies. I assume that the nonparametric alternative still satisfies separability between the continuous variables and the dummies. Separating $z$ into $z_1$, which includes the logarithm of the number of drivers in the household and the logarithm of the household size, and $z_2$, which includes the dummies, I rewrite all models with $z_1$ in place of $z$ and I consider the nonparametric alternative given by

$$y = h(\text{PRICE}, \text{AGE}, \text{INCOME}, z_1) + z_2'\lambda + \varepsilon$$ (7.2)

To construct the regressors $P^{kn}$ used to evaluate the test statistic, I use $l_n = 4$ power series terms in $\text{AGE}$, $\text{PRICE}$, and $\text{INCOME}$, the set of dummies discussed above, $j_n = 2$ power series terms in $	ext{DRIVERS}$ and $\text{HHSIZE}$. I then use pairwise interactions (tensor products) of univariate power series, and add all possible three, four, and five element interactions between $\text{AGE}$, $\text{PRICE}$, $\text{INCOME}$, $\text{DRIVERS}$, and $\text{HHSIZE}$, without using higher powers in these interaction terms to avoid multicollinearity. This gives rise to $m_n = 22$ terms under the null, $k_n = 146$ terms under the alternative, and $r_n = 124$ restrictions.

I estimate specification 7.1 using series methods and then compute the test statistic $t_{r_n}$ given by equation 4.6 and the heteroskedasticity robust test statistic $t_{r_n,HC,2}$ given by equation 4.13. The estimation results are shown in Table 7. I obtain $t_{r_n} = -0.346$ and $t_{r_n,HC,2} = 0.055$. The critical value at the 5% level equals 1.645, so the null hypothesis that model 7.1 is correctly specified is not rejected.

This is in line with the results in Yatchew and No (2001): even though they do not test their semiparametric specifications against a general nonparametric alternative, they find no evidence against a specification similar to 7.1 when compared to the following semiparametric
Next, I investigate what would happen if instead of the semiparametric model 7.1 the researcher estimated the following parametric model:

\[ y = \alpha_1 PRICE + \alpha_2 INCOME + \gamma_0 + \gamma_1 AGE + z'\beta + \varepsilon \] (7.3)

This model leads to \( m_n = 19 \) terms under the null, \( k_n = 146 \) terms under the alternative, and \( r_n = 127 \) restrictions. When testing it against specification 7.2, I obtain \( t_{r_n} = 2.765 \) and \( t_{r_n,HC;2} = 2.876 \), so that specification 7.3 is rejected at the 5% significance level. Thus, it seems that controlling for \( AGE \) flexibly is crucial in this gasoline demand application, and specification 7.1 used in Yatchew and No (2001) is flexible enough yet parsimonious.

Finally, I use the proposed test to compare specification 7.3 against the following semiparametric alternative:

\[ y = \alpha_1 PRICE + \alpha_2 INCOME + h_1(AGE, z_1) + z'_2\lambda + \varepsilon, \] (7.4)

which is more restricted than the nonparametric model 7.2. In this case, \( P^{k_n} \) includes only series terms in \( AGE \) and \( z_1 \) and their interactions. It results in \( k_n = 45 \) terms under the alternative and \( r_n = 26 \) restrictions. I obtain \( t_{r_n} = 20.118 \) and \( t_{r_n,HC;2} = 8.869 \), so that specification 7.3 is again rejected at the 5% significance level. This shows that restricting the class of alternatives can improve the power of the test if the true model is close to the conjectured restricted class. However, this will also result in the loss of consistency against alternatives that do not belong to the restricted set.

For practical purposes, if the null model under consideration can be nested in several more general models, it may make sense to test it against these several alternatives simultaneously using the Bonferroni correction. Using a general nonparametric alternative will result in
8 Conclusion

In this paper, I develop a new specification test for semiparametric conditional mean models. The proposed test achieves consistency by turning a conditional moment restriction into a growing number of unconditional moment restrictions using series methods. Because the number of series terms grows with the sample size, the usual asymptotic theory for the parametric Lagrange Multiplier test is no longer valid. I explicitly allow the number of terms to grow and show that the normalized test statistic converges in distribution to the standard normal. The proposed test has several attractive features compared to the existing tests.

First, the proposed test is simple to implement. The test statistic is based on a quadratic form in the semiparametric regression residuals, so only estimation of the restricted model is required to compute the test statistic. In the homoskedastic case, the quadratic form on which the test is based can be computed as $nR^2$ from the regression of the semiparametric residuals on the series terms used to construct the test. There are also regression based ways to compute the heteroskedasticity robust version of the test statistic. Moreover, the asymptotic distribution of the test statistic is pivotal, which facilitates the calculation of appropriate critical values.

Second, when the null model is nested in the alternative and is estimated by series methods, the projection property of series estimators makes it possible to explicitly account for the estimation variance and obtain refined asymptotic results. This refinement can be thought of as a degrees of freedom correction. Simulations show that because of this adjustment the proposed test behaves well in finite samples.

Third, series methods make it easy to restrict the class of alternatives from a fully non-parametric to certain semiparametric classes, such as additive, varying coefficient, or partially
linear. Doing so will result in the loss of consistency against a general alternative but will improve the power of the test in certain directions. In order to maintain consistency, one can combine tests against various alternatives simultaneously, including a fully nonparametric alternative, and use the Bonferroni correction to control the size of the test.

Finally, the test is not limited to models estimated by series methods. With a slightly different normalization, it remains valid even when the restricted model is estimated using other semiparametric methods, such as kernels or local polynomials. However, because the degrees of freedom correction is not available when these semiparametric estimators are used, stronger assumptions are required, and my simulations show that the test based on a generic normalization is typically undersized and low powered.

I apply the proposed test to the Canadian household gasoline consumption data from Yatchew and No (2001) and find no evidence against one of the semiparametric specifications used in their paper. However, I show that my test does reject a less flexible parametric model.

There are several avenues for future research, such as developing a data-driven procedure to choose tuning parameters (the number of series terms under the null and alternative), extending the proposed test to semiparametric models with endogeneity in the parametric part, and designing a model selection procedure (such as step-up or step-down procedures) based on the proposed test.
Appendices

A  Tables and Figures

Table 1: Correlation Between Regressors

<table>
<thead>
<tr>
<th></th>
<th>ln$p_{Ki}$</th>
<th>ln$p_{Li}$</th>
<th>ln$Q_i$</th>
<th>$Z_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln$p_{Ki}$</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln$p_{Li}$</td>
<td>0.337</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln$Q_i$</td>
<td>0.260</td>
<td>0.232</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$Z_i$</td>
<td>0.113</td>
<td>0.161</td>
<td>0.474</td>
<td>1</td>
</tr>
</tbody>
</table>

The table shows pairwise correlations between the regressors and is based on one realization of the regressor values.

Table 2: Rejection Probabilities, Normal Errors

<table>
<thead>
<tr>
<th></th>
<th>SP</th>
<th>NP</th>
<th>SP</th>
<th>NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi$</td>
<td>test statistic</td>
<td>test statistic</td>
<td>test statistic</td>
<td>test statistic</td>
</tr>
<tr>
<td>-------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>Setup 1</td>
<td>0.003</td>
<td>0.024</td>
<td>0.003</td>
<td>0.030</td>
</tr>
<tr>
<td>Setup 2</td>
<td>0.004</td>
<td>0.711</td>
<td>0.004</td>
<td>0.730</td>
</tr>
<tr>
<td>Setup 3</td>
<td>0.003</td>
<td>0.472</td>
<td>0.004</td>
<td>0.488</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\tau_n = k_n$</th>
<th>$\tau_n = r_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setup 1</td>
<td>0.057</td>
<td>0.199</td>
</tr>
<tr>
<td>Setup 2</td>
<td>0.053</td>
<td>0.952</td>
</tr>
<tr>
<td>Setup 3</td>
<td>0.053</td>
<td>0.811</td>
</tr>
</tbody>
</table>

Setup 1: $n = 1,000$, $k_n = 80$, $r_n = 65$. Setup 2: $n = 5,000$, $k_n = 80$, $r_n = 65$. Setup 3: $n = 5,000$, $k_n = 150$, $r_n = 132$. SP refers to semiparametric DGP, NP refers to nonparametric DGP. Entries in bold correspond to cases when $H_0$ is true. Results are based on $B = 2,000$ simulation draws.
### Table 3: Rejection Probabilities, Normal Heteroskedastic Errors

<table>
<thead>
<tr>
<th></th>
<th>SP</th>
<th>NP</th>
<th>SP</th>
<th>NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ξ test statistic</td>
<td>t test statistic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feasible Test Statistic $t_{n,HC,1}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Setup 1</td>
<td>0.103</td>
<td>0.310</td>
<td>0.120</td>
<td>0.346</td>
</tr>
<tr>
<td>Setup 2</td>
<td>0.076</td>
<td>0.980</td>
<td>0.083</td>
<td>0.983</td>
</tr>
<tr>
<td>Setup 3</td>
<td>0.071</td>
<td>0.921</td>
<td>0.076</td>
<td>0.933</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>SP</th>
<th>NP</th>
<th>SP</th>
<th>NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infeasible Test Statistic $t_{n,HC,1,inf}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Setup 1</td>
<td>0.095</td>
<td>0.326</td>
<td>0.106</td>
<td>0.355</td>
</tr>
<tr>
<td>Setup 2</td>
<td>0.070</td>
<td>0.982</td>
<td>0.075</td>
<td>0.983</td>
</tr>
<tr>
<td>Setup 3</td>
<td>0.071</td>
<td>0.931</td>
<td>0.076</td>
<td>0.941</td>
</tr>
</tbody>
</table>

Setup 1: $n = 1,000$, $k_n = 80$, $r_n = 65$. Setup 2: $n = 5,000$, $k_n = 80$, $r_n = 65$. Setup 3: $n = 5,000$, $k_n = 150$, $r_n = 132$. SP refers to semiparametric DGP, NP refers to nonparametric DGP. Entries in bold correspond to cases when $H_0$ is true. Results are based on $B = 1,000$ simulation draws.

### Table 4: Rejection Probabilities, Normal Heteroskedastic Errors

<table>
<thead>
<tr>
<th></th>
<th>SP</th>
<th>NP</th>
<th>SP</th>
<th>NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ξ test statistic</td>
<td>t test statistic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feasible Test Statistic $t_{n,HC,2}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Setup 1</td>
<td>0.028</td>
<td>0.112</td>
<td>0.036</td>
<td>0.129</td>
</tr>
<tr>
<td>Setup 2</td>
<td>0.035</td>
<td>0.931</td>
<td>0.042</td>
<td>0.940</td>
</tr>
<tr>
<td>Setup 3</td>
<td>0.036</td>
<td>0.799</td>
<td>0.041</td>
<td>0.811</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>SP</th>
<th>NP</th>
<th>SP</th>
<th>NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infeasible Test Statistic $t_{n,HC,2,inf}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Setup 1</td>
<td>0.052</td>
<td>0.194</td>
<td>0.061</td>
<td>0.216</td>
</tr>
<tr>
<td>Setup 2</td>
<td>0.043</td>
<td>0.943</td>
<td>0.050</td>
<td>0.951</td>
</tr>
<tr>
<td>Setup 3</td>
<td>0.052</td>
<td>0.853</td>
<td>0.054</td>
<td>0.866</td>
</tr>
</tbody>
</table>

Setup 1: $n = 1,000$, $k_n = 80$, $r_n = 65$. Setup 2: $n = 5,000$, $k_n = 80$, $r_n = 65$. Setup 3: $n = 5,000$, $k_n = 150$, $r_n = 132$. SP refers to semiparametric DGP, NP refers to nonparametric DGP. Entries in bold correspond to cases when $H_0$ is true. Results are based on $B = 1,000$ simulation draws.

### Table 5: Rejection Probabilities, Local Alternatives

<table>
<thead>
<tr>
<th>$n$</th>
<th>$k_n$</th>
<th>$r_n$</th>
<th>P(reject $H_0$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>80</td>
<td>65</td>
<td>0.359</td>
</tr>
<tr>
<td>1000</td>
<td>80</td>
<td>65</td>
<td>0.370</td>
</tr>
<tr>
<td>2500</td>
<td>96</td>
<td>78</td>
<td>0.369</td>
</tr>
<tr>
<td>5000</td>
<td>150</td>
<td>132</td>
<td>0.379</td>
</tr>
<tr>
<td>10000</td>
<td>175</td>
<td>154</td>
<td>0.380</td>
</tr>
<tr>
<td>25000</td>
<td>200</td>
<td>176</td>
<td>0.395</td>
</tr>
</tbody>
</table>

Results are based on $B = 2,000$ simulation draws.
Table 6: Rejection Probabilities, Normal Errors, Bonferroni

<table>
<thead>
<tr>
<th></th>
<th>PL VC</th>
<th>NP</th>
<th>PL VC</th>
<th>NP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\xi$ test statistic</td>
<td>$t$ test statistic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_{0}^{PL}$ vs. $H_{1}^{NP}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Setup 1</td>
<td>0.047 0.076 0.200</td>
<td>0.057 0.091 0.224</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Setup 2</td>
<td>0.052 0.258 0.846</td>
<td>0.059 0.272 0.862</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_{0}^{PL}$ vs. $H_{1}^{VC}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Setup 1</td>
<td>0.049 0.217 0.090</td>
<td>0.066 0.262 0.121</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Setup 2</td>
<td>0.054 0.809 0.268</td>
<td>0.067 0.841 0.312</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H_{0}^{PL}$ vs. $H_{1}^{NP}$ and $H_{1}^{VC}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Setup 1</td>
<td>0.049 0.161 0.150</td>
<td>0.071 0.229 0.196</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Setup 2</td>
<td>0.049 0.733 0.797</td>
<td>0.078 0.801 0.834</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Setup 1: $n = 1,000, k_n = 80, r_n = 65$. Setup 2: $n = 5,000, k_n = 150, r_n = 132$. PL refers to semiparametric partially linear DGP, VC refers to semiparametric varying coefficient DGP, NP refers to nonparametric DGP. The upper panel presents the results from testing the parametric null hypothesis $H_{0}^{PL}$ against the nonparametric alternative $H_{1}^{NP}$ at the nominal 5% level. The middle panel presents the results from testing the parametric null hypothesis $H_{0}^{PL}$ against the semiparametric alternative $H_{1}^{VC}$ at the nominal 5% level. The lower panel presents the results from testing the parametric null hypothesis $H_{0}^{PL}$ against both nonparametric alternative $H_{1}^{NP}$ and semiparametric alternative $H_{1}^{VC}$ at the nominal 2.5% level each in order to control the size of the joint test. Entries in bold correspond to cases when $H_0$ is true. Results are based on $B = 2,000$ simulation draws.

Table 7: Partially Linear Model Estimates

<table>
<thead>
<tr>
<th>Regressor</th>
<th>Dep. Var.: Gasoline Demand</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRICE</td>
<td>-0.921 (0.096)</td>
</tr>
<tr>
<td>INCOME</td>
<td>0.285 (0.021)</td>
</tr>
<tr>
<td>DRIVERS</td>
<td>0.540 (0.034)</td>
</tr>
<tr>
<td>HHSIZE</td>
<td>0.110 (0.028)</td>
</tr>
<tr>
<td>single, age &lt; 35</td>
<td>0.189 (0.064)</td>
</tr>
<tr>
<td>urban dummy</td>
<td>-0.331 (0.020)</td>
</tr>
<tr>
<td>monthly effects</td>
<td>Yes</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.265</td>
</tr>
<tr>
<td>number of observations</td>
<td>6,230</td>
</tr>
</tbody>
</table>

The table presents the estimates specification 7.1. I use $l_n = 4$ power series terms in $AGE$ to flexibly control for $AGE$, which yields $m_n = 22$ parameters in the semiparametric model. Standard errors are shown in parentheses.
Figure 1: Varying Coefficient Functions

This figure shows the true coefficient functions $C_1(z)$, $\gamma(z)$, and $\delta(z)$ for the varying coefficient semiparametric model from equation 6.1 used in simulations.

Figure 2: Comparison of RTS Estimates

This long dashed line shows the true function $\delta(z)$ for equation 6.1. The short dashed line shows its OLS estimate, the dash-dotted line shows its OLS with interactions estimate, and the solid line shows its semi-parametric varying coefficient estimate. The figure is based on $B = 1,000$ simulation draws with $n = 1,000$ observations in each. The regressors are fixed across the simulation draws, only the errors are redrawn as $\varepsilon_i \sim \text{i.i.d. } N(0, 2.25)$. 
Figure 3: $H_0$ and $H_1$ in 3D

The left figure shows the dependence of $\tilde{TC}_i$ (on the $z$ axis) on $\tilde{P}_i$ (on the $y$ axis) and $\tilde{Q}_i$ (on the $x$ axis) under $H_0$ in equation 6.1. The right figure shows the dependence of $\tilde{TC}_i$ (on the $z$ axis) on $\tilde{P}_i$ (on the $y$ axis) and $\tilde{Q}_i$ (on the $x$ axis) under $H_1$ in equation 6.2. The figure is based on one realization of the regressor values and errors.

Figure 4: $H_0$ and $H_1$ in 2D

The left figure shows the dependence of $\tilde{TC}_i$ on $\tilde{p}_i$ conditional on fixed $\tilde{Q}_i$ and $Z_i$. The right figure shows the dependence of $\tilde{TC}_i$ on $\tilde{Q}_i$ conditional on fixed $\tilde{p}_i$ and $Z_i$. The solid lines show the linear relationship which holds under $H_0$ in equation 6.1. The dashed lines show the nonlinear relationship which holds under $H_1$ in equation 6.2. The figure is based on one realization of the regressor values and errors.
The solid line shows the simulated distribution of the test statistic $t_{r_n}$, the dash-dotted line shows the simulated distribution of the test statistic $t_{k_n}$, and the dashed line shows the standard normal distribution. The results are based on $B = 2,000$ simulation draws, $\varepsilon_i \sim$ i.i.d. $N(0, 2.25)$. In the upper left figure $n = 1,000, k_n = 80, r_n = 65$; in the upper right figure $n = 5,000, k_n = 80, r_n = 65$; in the bottom figure $n = 5,000, k_n = 150, r_n = 132$.

The figure illustrates the form of heteroskedasticity $\varepsilon_i \sim$ i.i.d. $N(0, 0.015 \exp(\tilde{Q}_i + Z_i))$. The coordinates of the points in the scatter plot are given by $(\ln Q_i + Z_i, \varepsilon_i)$. It is based on one realization of the regressor values and errors.
Figure 7: Distribution of $t$ under $H_0$, Normal Heteroskedastic Errors

The solid line shows the simulated distribution of the feasible test statistic $t_{r_n, HC, 1}$, the dash-dotted line shows the simulated distribution of the infeasible test statistic $t_{r_n, HC, 1, \text{inf}}$, and the dashed line shows the standard normal distribution. The results are based on $B = 1,000$ simulation draws, $\varepsilon_i \sim \text{i.n.i.d. } N(0, 0.015 \exp(\tilde{Q}_i + Z_i))$. In the upper left figure $n = 1,000$, $k_n = 80$, $r_n = 65$; in the upper right figure $n = 5,000$, $k_n = 80$, $r_n = 65$; in the bottom figure $n = 5,000$, $k_n = 150$, $r_n = 132$. 

57
The solid line shows the simulated distribution of the feasible adjusted test statistic $t_{r_n,HC;2}$, the dash-dotted line shows the simulated distribution of the infeasible adjusted test statistic $t_{r_n,HC;2,inf}$, and the dashed line shows the standard normal distribution. The results are based on $B = 1,000$ simulation draws, $\varepsilon_i \sim \text{i.n.i.d. } N(0, 0.015 \exp(\tilde{Q}_i + Z_i))$. In the upper left figure $n = 1,000$, $k_n = 80$, $r_n = 65$; in the upper right figure $n = 5,000$, $k_n = 80$, $r_n = 65$; in the bottom figure $n = 5,000$, $k_n = 150$, $r_n = 132$.

The solid line shows the simulated distribution of the test statistic $t_{r_n}$ under local alternatives for $n = 10,000$, $k_n = 175$, $r_n = 154$, the dash-dotted line shows the simulated distribution of the test statistic $t_{r_n}$ under local alternatives for $n = 1,000$, $k_n = 80$, $r_n = 65$, and the dashed line shows the standard normal distribution. The results are based on $B = 2,000$ simulation draws, $\varepsilon_i \sim \text{i.i.d. } N(0, 2.25)$. 

58
B Some Practical Tips

B.1 Series Representation

In order to derive the series-based specification test, I write the restricted and unrestricted models in a series form. For a variable \( z \), let \( Q_l(z) = (q_1(z), ..., q_{l_n}(z))' \) be a sequence of approximating functions of \( z \). Then an unknown function \( h(z) \) can be approximated as

\[
h(z) \approx \sum_{j=1}^{l_n} \gamma_j q_j(z) = Q_l(z)' \gamma
\]

Let \( W_{mn}(x) \) be the sequence of functions which is used to estimate the restricted semi-parametric model. Next, let \( T_{rn}(x) = (t_1(x), ..., t_{r_n}(x),) \)' be the sequence of approximating functions which is used in addition to \( W_{mn}(x) \) to estimate the fully nonparametric model, so that \( P_{kn}(x) = (W_{mn}(x)', T_{rn}(x)')' \). The difference between \( T_{rn}(x) \) and \( W_{mn}(x) \) is that the former is present only in the unrestricted nonparametric model, while the latter is present in both restricted and unrestricted models.

The unrestricted nonparametric model can be written as

\[
Y_i = P_{kn}(X_i)' \beta + R_i + \varepsilon_i = W_{mn}(X_i)' \beta_1 + T_{rn}(X_i)' \beta_2 + R_i + \varepsilon_i,
\]

where \( \beta = (\beta_1', \beta_2')' \).

The null hypothesis that the conditional mean function is semiparametric corresponds to \( r_n \) restrictions \( \beta_2 = 0 \). To test this hypothesis, the researcher first needs to estimate the semiparametric model

\[
Y_i = W_{mn}(X_i)' \beta_1 + \varepsilon_i,
\]

obtain the estimates \( \tilde{\beta}_1 \) and residuals \( \tilde{\varepsilon}_i = Y_i - W_{mn}(X_i)' \tilde{\beta}_1 \), compute \( \tilde{\sigma}^2 = \tilde{\varepsilon}' \tilde{\varepsilon}/n \), and then
use the following statistic as a basis for the test:

$$\xi = \hat{\varepsilon}' P (\hat{\sigma}^2 P' P)^{-1} P' \hat{\varepsilon}$$

As I show below, this test statistic can be derived as a Lagrange Multiplier or Conditional Moment test statistic for the semiparametric model written in a series form if the dependence of the number of terms on the sample size is ignored and the model is treated as parametric. In parametric models with a fixed number of restrictions $r$, this test statistic converges in distribution $\chi^2_r$ under the null. In the present paper, the number of restrictions $r_n$ grows with the sample size, so the usual asymptotic result does not hold. I develop an asymptotic theory for the proposed test in Section 4.

B.2 Proposed Test as LM Test

As shown above, the unrestricted nonparametric model is given by

$$Y_i = P^{kn}(X_i)' \beta + e_i = W^{mn}(X_i)' \beta_1 + T^{rn}(X_i)' \beta_2 + e_i,$$

where $\beta = (\beta_1', \beta_2')'$. The semiparametric null model imposes the restriction that $\beta_2 = 0$. If one ignored the presence of approximation errors and the dependence of the number of series terms on the sample size, this restriction could be tested using the Lagrange Multiplier test.

The (quasi-)log-likelihood for the nonparametric model is given by

$$L_n(\beta, \sigma^2) = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log \sigma^2 - \frac{1}{2\sigma^2 n} (Y - P\beta)'(Y - P\beta)$$

Then the score equals

$$S_n(\beta, \sigma^2) = \left( \begin{array}{c} \frac{\partial L_n(\beta, \sigma^2)}{\partial \beta} \\ \frac{\partial L_n(\beta, \sigma^2)}{\partial \sigma^2} \end{array} \right) = \left( \begin{array}{c} \frac{1}{\sigma^2 n} P'(Y - P\beta) \\ -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4 n} (Y - P\beta)'(Y - P\beta) \end{array} \right),$$
which under the null hypothesis becomes

\[ S_n(\tilde{\beta}, \tilde{\sigma}^2) = \begin{pmatrix} \frac{1}{\tilde{\sigma}^2_n} P' \tilde{\varepsilon} \\ 0 \end{pmatrix}, \]

where \( \tilde{\beta} = (\tilde{\beta}'_1, 0'_{r_n})'. \)

The information matrix evaluated at true parameter values is:

\[ F_n(\beta, \sigma^2) = \begin{pmatrix} E[-\frac{\partial S_n}{\partial \beta}] \\ E[-\frac{\partial S_n}{\partial \sigma^2}] \end{pmatrix} = \begin{pmatrix} E[\frac{1}{\tilde{\sigma}^2_n} P' P] \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1}{2\sigma^2} \end{pmatrix}. \]

Assuming that the information matrix equality holds, the \( \xi \) test statistic for the restricted null model is constructed by evaluating the score and the Hessian at the restricted estimates:

\[ \xi = nS_n(\tilde{\beta}, \tilde{\sigma}^2)' F_n(\tilde{\beta}, \tilde{\sigma}^2)^{-1} S_n(\tilde{\beta}, \tilde{\sigma}^2) = \tilde{\varepsilon}' P(\tilde{\sigma}^2 P' P)^{-1} P' \tilde{\varepsilon}, \]

**B.3 Proposed Test as CM/J Test**

As shown above, the semiparametric model is given by

\[ Y_i = W^{m_n}(X_i)' \beta_1 + R_i + \varepsilon_i, \quad E[\varepsilon_i|X_i] = 0 \]

Ignoring the presence of the approximation error, rewrite it as

\[ Y_i = W^{m_n}(X_i)' \beta_1 + e_i, \quad E[e_i|X_i] = 0 \]

It has \( m_n \) parameters but should satisfy \( k_n = m_n + r_n \) moment conditions, because \( e_i \) should be uncorrelated with any function of \( X_i \), not only with \( W^{m_n}(X_i) \):

\[ E[P^{k_n}(X_i)e_i] = 0 \quad (B.1) \]
Because the model satisfies more unconditional moment restrictions than there are parameters to estimate, it is overidentified. Thus, the specification test can be based on testing the increasing number of moment conditions.

The semiparametric model satisfies the following population moment condition

$$E[W^m(X_i)e_i] = 0$$  \hspace{1cm} (B.2)

It means that the residuals $\tilde{\epsilon}_i$ solve the sample analog of the population moment condition B.2:

$$\frac{1}{n} \sum_{i=1}^{n} W^m(X_i)\tilde{\epsilon}_i = \frac{1}{n} W'\tilde{\epsilon} = 0$$

If the semiparametric conditional model model is correctly specified, we would expect that the sample analog of the moment condition B.1 to be close to zero, i.e.

$$\frac{1}{n} \sum_{i=1}^{n} P^k(X_i)\tilde{\epsilon}_i = \frac{1}{n} P'\tilde{\epsilon} \approx 0$$

The Conditional Moment test statistic could be used to evaluate whether these moment conditions are statistically different from zero:

$$\xi = \tilde{\epsilon}'P(\tilde{\sigma}^2P'P)^{-1}P'\tilde{\epsilon}$$

**Remark A.1 (Conditional Moments and Overidentifying Restrictions Tests).** The proposed test statistic is very similar, though not exactly the same, as the overidentifying restrictions test statistic. Suppose that instead of estimating the parameters by solving the sample analog of the OLS population moment conditions:

$$E[W^m(X_i)e_i] = E[W^m(X_i)(Y_i - W^m(X_i)'\beta_i)] = 0,$$

the researcher estimates the parameters using GMM based on the overidentified population
moment condition:

\[ E[P^k(X_i)e_i] = E[P^k(X_i)(Y_i - W^{mn}(X_i)'\beta_1)] = 0 \]

For a weighting matrix \( M \) and its first step estimate \( \hat{M} \), the GMM estimates \( \hat{\beta}_1 \) solve

\[ \min_{\hat{\beta}_1} ((Y - W\beta_1)'P/n) \hat{M}(P'(Y - W\beta_1)/n) \]

For the GMM residuals \( \hat{\varepsilon}_i = Y_i - W^{mn}(X_i)'\hat{\beta}_1 \), the overidentifying test statistic is

\[ J = n(\hat{\varepsilon}'P/n)\hat{M}(P'\hat{\varepsilon}/n), \]

In general, \( \hat{\beta}_1 \neq \tilde{\beta}_1 \), because they solve different minimization problems, and hence \( \hat{\varepsilon} \neq \tilde{\varepsilon} \). Under the null, however, \( \tilde{\beta}_1 \) and \( \hat{\beta}_1 \), and thus \( \tilde{\varepsilon} \) and \( \hat{\varepsilon} \), should be close to one another, so it should be possible to use both \( \xi \) and \( J \) as a basis for the specification test. The expression for \( J \) may look more familiar from GMM literature; however, I prefer to use \( \xi \) because it can also be viewed as an LM test, because it requires estimating only the restricted semiparametric model, and because it makes it possible to directly account for the estimation variance.

If the researcher estimates 2SLS (i.e. GMM with \( M = (\sigma^2E[P_iP_i'])^{-1} \) and \( \hat{M} = (\hat{\sigma}^2P'P/n)^{-1} \)), then two statistics coincide as long as the second step estimate of \( \sigma^2 \) is used to form the \( J \) statistic. Namely, 2SLS estimates solve

\[ \min_{\hat{\beta}_1} ((Y - W\beta_1)'P/n)(\hat{\sigma}^2P'P/n)^{-1}(P'(Y - W\beta_1)/n) \]

The first order condition is

\[ (W'P/n)(\hat{\sigma}^2P'P/n)^{-1}(P'(Y - W\beta_1)/n) = 0, \]
which yields

\[ \hat{\beta}_1 = (W'P(P'P)^{-1}P'W)^{-1}W'P(P'P)^{-1}P'Y = (W'W)^{-1}WY = \tilde{\beta}_1 \]

Thus, the 2SLS and OLS estimates of \( \beta_1 \) coincide, hence, \( \tilde{\epsilon} = \hat{\epsilon} \) and \( \tilde{\sigma}^2 = \hat{\sigma}^2 \), where \( \hat{\sigma}^2 \) is the updated second-step estimate. The J statistic becomes

\[ J = n(\hat{\epsilon}'P/n)(\hat{\sigma}^2P'P/n)^{-1}(P'\hat{\epsilon}/n) = \hat{\epsilon}'P(\hat{\sigma}^2P'P)^{-1}P'\hat{\epsilon} = \xi \]

### B.4 LM, LR, and Wald Type Test Statistics

In certain cases, it is possible to construct a consistent specification test for semiparametric models by using a Wald type or Likelihood Ratio type test statistic. Let \( \hat{\beta} = (\hat{\beta}_1', \hat{\beta}_2')' \) be the estimates for the unrestricted nonparametric model 3.6. Let \( \hat{\epsilon} = Y - P\hat{\beta} \) and \( \hat{\sigma}^2 = \hat{\epsilon}'\hat{\epsilon}/n \).

The Wald type test statistic is given by

\[ \xi_W = \hat{\beta}_2'\hat{V}^{-1}\hat{\beta}_2, \]

where \( \hat{V} = \text{Var}(\hat{\beta}_2) \).

The Likelihood Ratio type test statistic is given by

\[ \xi_{LR} = n\frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2} \]

I show below that the test statistics \( \xi, \xi_W, \) and \( \xi_{LR} \) are asymptotically equivalent when series methods are used to estimate the model and the errors are homoskedastic. First, note that the LM type statistic \( \xi \) can be written as

\[ \xi = \hat{\epsilon}'P(\hat{\sigma}^2P'P)^{-1}P'\hat{\epsilon} = Y'M_WP(P'P)^{-1}P'M_WY, \]
where \( M_W = I - W(W'W)^{-1}W' \). Using the partitioned matrix inverse formula and the fact that \( P = (W \quad T) \),
\[
\xi = Y'M_W T(T'M_W T)^{-1}T'M_W Y
\]

Next, I consider the LR type test statistic. Note that by the projection property of series estimators, \( \tilde{\varepsilon} = M_W Y \) and \( \hat{\varepsilon} = M_P Y \), where \( M_P = I - P(P'P)^{-1}P' \). Thus,
\[
\xi_{LR} = \frac{n \hat{\sigma}^2 - \bar{\sigma}^2}{\bar{\sigma}^2} = \frac{(Y'M_W Y - Y'M_P Y)/\hat{\sigma}^2}{(Y'P(P'P)^{-1}P'Y - Y'W(W'W)^{-1}W'Y)/\tilde{\sigma}^2}
\]

Using the partitioned matrix inverse formula,
\[
\xi_{LR} = Y'M_W T(T'M_W T)^{-1}T'M_W Y / \hat{\sigma}^2 = \xi \frac{\tilde{\sigma}^2}{\hat{\sigma}^2}
\]

Now I consider the Wald type test statistic. Note that
\[
\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (P'P)^{-1}P'Y = \begin{pmatrix} W'W \\ W'T \\ T'W \\ T'T \end{pmatrix}^{-1} \begin{pmatrix} W'Y \\ T'Y \end{pmatrix}
\]

Using the partitioned matrix inverse formula, \( \hat{\beta}_2 = (T'M_W T)^{-1}T'M_W Y \), which, using the usual OLS arguments, leads to \( \hat{V} = \hat{\sigma}^2 (T'M_W T)^{-1} \). Thus,
\[
\xi_W = \hat{\beta}_2' \hat{V}^{-1} \hat{\beta}_2 = Y'M_W T(T'M_W T)^{-1}(T'M_W T)(T'M_W T)^{-1}T'M_W Y / \hat{\sigma}^2
\]
\[
= Y'M_W T(\hat{\sigma}^2 T'M_W T)^{-1}T'M_W Y = \xi \frac{\hat{\sigma}^2}{\tilde{\sigma}^2}
\]

Typically, \( \hat{\sigma}^2 \overset{p}{\to} \sigma^2 \) and \( \tilde{\sigma}^2 \overset{p}{\to} \sigma^2 \), so that \( \frac{\hat{\sigma}^2}{\tilde{\sigma}^2} \overset{p}{\to} 1 \),\(^{12}\) which means that \( \xi \) is asymptotically equivalent to both \( \xi_{LR} \) and \( \xi_W \). However, using the LM type test statistic has two advantages. First, unlike the Wald type test statistic, it can be used even when the null model is not nested in the alternative or is estimated using general semiparametric methods. Second,\(^{12}\)

\(^{12}\)The derivation of primitive conditions under which this result holds is beyond the scope of this paper.
unlike the LR type test statistic, it can be easily modified to deal with heteroskedasticity. Thus, the LM type specification test is more general than the other two tests. Moreover, in the context of specification testing, the LM test may seem more natural, because it is based on the estimates from the model of interest and does not require the researcher to explicitly specify the alternative. Hence, in this paper, I focus only on the LM type specification test.

B.5 Implementing Test in Practice

If the researcher is willing to nest the null model in the alternative and use series methods to estimate the null model, then the following steps are needed to implement the test:

1. Pick the sequence of approximating functions of \( x \), \( W_m(x) = (w_1(x), \ldots, w_m(x))' \), which will be used to estimate the semiparametric model.

2. Estimate the semiparametric model \( Y_i = f(X_i, \theta, h) + \varepsilon_i \approx W_m(X_i)'\beta_1 + \varepsilon_i \) using series methods. Obtain the estimates \( \tilde{\beta}_1 = (W'W)^{-1}W'Y \) and residuals \( \tilde{\varepsilon}_i = Y_i - W_m(X_i)'\tilde{\beta}_1 \).

3. Pick the sequence of approximating functions of \( x \), \( T_r(x) = (t_1(x), \ldots, t_r(x))' \), which will complement \( W_m(x) \) to form the matrix \( P(x) \), \( k_n = m_n + r_n \), which corresponds to a general nonparametric model. \( P(x) \) should be able to approximate any unknown function sufficiently well. Common choices of basis functions include power series (see Equation 3.3) and splines (see Equation 3.4).

4. Compute the quadratic form \( \xi = \tilde{\varepsilon}'P(\tilde{\sigma}^2P'P)^{-1}P'\tilde{\varepsilon} \).

Note that in the homoskedastic case \( \xi \) can be computed as \( nR^2 \) from the regression of \( \tilde{\varepsilon}_i \) on \( P_i \). The residuals from the regression of \( \tilde{\varepsilon}_i \) on \( P_i \) are given by \( e = \tilde{\varepsilon} - P(P'P)^{-1}P'\tilde{\varepsilon} \), so that

\[ e'e = \tilde{\varepsilon}'\tilde{\varepsilon} - \tilde{\varepsilon}P(P'P)^{-1}P'\tilde{\varepsilon} \]

\[ \text{13} \] If the semiparametric model written in a series form includes a constant, both centered and uncentered \( R^2 \) can be used because they are identical. If the semiparametric model does not include a constant, the uncentered \( R^2 \) should be used.
Then
\[ nR^2 = n \left( 1 - \frac{\bar{e}'e}{\bar{e}'\bar{e}} \right) = n \frac{\bar{e} \bar{P}(\bar{P}'P)^{-1} \bar{P}' \bar{e}}{\bar{e}'\bar{e}} = \bar{e} P(\bar{\sigma}^2 P'P)^{-1} P' \bar{e} \]

Note also that, as shown in Remark A.1, \( \xi \) can be computed as the overidentifying restrictions test statistic from the 2SLS instrumental variables regression of \( Y_i \) on \( W^m(X_i) \) with \( (W^m(X_i)', T^a(X_i))' \) as instruments.

5. Compute the test statistic which is asymptotically standard normal under the null:
\[ t = \frac{\xi - r_n}{\sqrt{2r_n}} \sim N(0, 1) \]

Reject the null if \( t > z_{1-\alpha} \), the \((1 - \alpha)\)-quantile of the standard normal distribution.

Alternatively, use the \( \chi^2 \) approximation directly: \( \xi \overset{a}{\sim} \chi^2_{r_n} \), reject the null if \( \xi > \chi^2_{r_n}(1 - \alpha) \), the \((1 - \alpha)\)-quantile of the \( \chi^2 \) distribution with \( r_n \) degrees of freedom.

If the researcher is willing to use other semiparametric methods, such as kernels, to estimate the null model, then the following steps are needed to implement the test:

1. Estimate the semiparametric model \( Y_i = f(X_i, \theta, h) + \varepsilon_i \) using the preferred estimation method. Obtain the estimates \( \hat{\theta} \) and \( \hat{h} \) and residuals \( \hat{\varepsilon}_i = Y_i - f(X_i, \hat{\theta}, \hat{h}) \).

2. Pick the sequence of approximating functions of \( x \), \( P^k(x) = (p_1(x), ..., p_k(x))' \), which is implicitly used to approximate a general nonparametric model. As in the previous case, \( P^k(x) \) should be able to approximate any unknown function sufficiently well. Common choices of basis functions include power series (see Equation 3.3) and splines (see Equation 3.4).

3. Compute the quadratic form \( \xi = \varepsilon'P(\bar{\sigma}^2 P'P)^{-1} P' \varepsilon \).

As before, in the homoskedastic case \( \xi \) can be computed as \( nR^2 \) from the regression of \( \hat{\varepsilon}_i \) on \( P_i \).
4. Compute the test statistic which is asymptotically standard normal under the null:

\[ t = \frac{\xi - k_n}{\sqrt{2k_n}} \sim N(0, 1) \]

Reject the null if \( t > z_{1-\alpha} \), the \((1-\alpha)\)-quantile of the standard normal distribution.

Alternatively, use the \( \chi^2 \) approximation directly: \( \xi \sim \chi^2_{k_n} \), reject the null if \( \xi > \chi^2_{k_n}(1-\alpha) \), the \((1-\alpha)\)-quantile of the \( \chi^2 \) distribution with \( k_n \) degrees of freedom.

If the researcher suspects that the errors are heteroskedastic, then \( \xi = \tilde{\varepsilon}' P(\tilde{\sigma}^2 P' P)^{-1} P' \tilde{\varepsilon} \) is replaced with \( \xi_{HC,1} = \tilde{\varepsilon}' P(P' \tilde{\Sigma} P)^{-1} P' \tilde{\varepsilon} \) or \( \xi_{HC,2} = \tilde{\varepsilon}' \tilde{T}'(\tilde{T}' \tilde{\Sigma} \tilde{T})^{-1} \tilde{T}' \tilde{\varepsilon} \), where \( \tilde{\Sigma} = \text{diag}(\tilde{\varepsilon}_i^2) \) and \( \tilde{T}_i = T_i - W_i'(W' W)^{-1} W'T \). All other steps remain unchanged.

Note that \( \xi_{HC,1} \) can be used no matter what method is used to estimate the semiparametric model, while \( \xi_{HC,2} \) can be used only when the semiparametric null model is nested in the alternative and estimated using series methods. This is because the researcher needs to distinguish between the series terms \( W^{m_n}(X_i) \) and \( T^{r_n}(X_i) \) to compute \( \xi_{HC,2} \). Whenever possible, I recommend using the test based on \( \xi_{HC,2} \), because it appears to control the size of the test better based on the simulation evidence.

Note that in the heteroskedastic case \( \xi_{HC,1} \) can be computed as \( nR^2 \) from the regression of 1 on \( P_i \tilde{\varepsilon}_i \). The matrix of regressors can be written as \( \tilde{\Sigma}^{1/2} P \), where \( \tilde{\Sigma}^{1/2} = \text{diag}(\tilde{\varepsilon}_i) \). Let \( \iota_n \) be a \( n \)-vector of ones. Note that \( \tilde{\Sigma}^{1/2} \iota_n = \tilde{\varepsilon} \). Then

\[ nR^2 = n \left( 1 - \frac{e'e}{\iota'_n \iota_n} \right) = n \iota'_n \tilde{\Sigma}^{1/2} P(P' \tilde{\Sigma} P)^{-1} P' \tilde{\Sigma}^{1/2} \iota_n = \tilde{\varepsilon}' P(P' \tilde{\Sigma} P)^{-1} P' \tilde{\varepsilon} \]

In turn, \( \xi_{HC,2} \) can be computed as \( nR^2 \) from the regression of 1 on \( \tilde{T}_i \tilde{\varepsilon}_i \). The matrix of regressors can be written as \( \tilde{\Sigma}^{1/2} \tilde{T} \), where \( \tilde{\Sigma}^{1/2} = \text{diag}(\tilde{\varepsilon}_i) \). Let \( \iota_n \) be a \( n \)-vector of ones. Note that \( \tilde{\Sigma}^{1/2} \iota_n = \tilde{\varepsilon} \). Then

\[ nR^2 = n \left( 1 - \frac{e'e}{\iota'_n \iota_n} \right) = n \iota'_n \tilde{\Sigma}^{1/2} \tilde{T}(\tilde{T}' \tilde{\Sigma} \tilde{T})^{-1} \tilde{T}' \tilde{\Sigma}^{1/2} \iota_n = \tilde{\varepsilon}' \tilde{T}(\tilde{T}' \tilde{\Sigma} \tilde{T})^{-1} \tilde{T}' \tilde{\varepsilon} \]

68
C Proofs

C.1 Proof of Theorem 1

Given the projection nature of series estimators, the test statistic becomes

\[
\frac{\xi - r_n}{\sqrt{2r_n}} = \frac{\varepsilon' P(\hat{\sigma}^2 P'P)^{-1} P'\varepsilon - r_n}{\sqrt{2r_n}} = \frac{(\varepsilon + R)' M_W P(\hat{\sigma}^2 P'P)^{-1} P'M_W(\varepsilon + R) - r_n}{\sqrt{2r_n}}
\]

Note that \(P = (W \quad T)\), so that \(M_W P = \begin{pmatrix} 0_{n \times m_n} & M_W T \end{pmatrix}\). Then by the blockwise matrix inverse formula,

\[
M_W P(\hat{\sigma}^2 P'P)^{-1} P'M_W = M_W T(\hat{\sigma}^2 T'M_W T)^{-1} T'M_W
\]

Thus,

\[
\frac{\xi - r_n}{\sqrt{2r_n}} = \frac{(\varepsilon + R)' M_W T(\hat{\sigma}^2 T'M_W T)^{-1} T'M_W(\varepsilon + R) - r_n}{\sqrt{2r_n}} = \frac{(\varepsilon + R)' \tilde{T}(\hat{\sigma}^2 \tilde{T}'\tilde{T})^{-1} \tilde{T}'(\varepsilon + R) - r_n}{\sqrt{2r_n}}
\]

where \(\tilde{T} = M_W T\). The remainder of the proof consists of several steps.

Step 1. Show that \(\|\tilde{T}'\tilde{T}/n - T'T/n\| = o_p(1/\sqrt{r_n})\).

Note that

\[
\tilde{T}'\tilde{T}/n = T'T/n - (T'W/n)(W'W/n)^{-1}(W'T/n)
\]

Thus,

\[
\|\tilde{T}'\tilde{T}/n - T'T/n\| = \|(T'W/n)(W'W/n)^{-1}(W'T/n)\|
\]

By Lemma 15.2 in Li and Racine (2007), \(E[\|P'P/n - I_{k_n}\|^2] = O_p(\zeta(k_n)^2 k_n/n)\) and \(\|P'P/n - I_{k_n}\| = O_p(\zeta(k_n)\sqrt{k_n/n})\). Note that \(P'P/n - I_{k_n} = \begin{pmatrix} W'W/n & W'T/n \\ T'W/n & T'T/n \end{pmatrix} - \begin{pmatrix} I_{m_n} & 0_{m_n \times r_n} \\ 0_{r_n \times m_n} & I_{r_n} \end{pmatrix}\).

Hence, \(\|W'W/n - I_{m_n}\| = O_p(\zeta(k_n)\sqrt{k_n/n})\) and \(\|W'T/n\| = O_p(\zeta(k_n)\sqrt{k_n/n})\). Thus,
the eigenvalues of $W'W/n$ are bounded below and above w.p.a.1, and

\[
\left\| (T'W/n)(W'W/n)^{-1}(W'T/n) \right\| \leq C \left\| (T'W/n)(W'T/n) \right\|
\]

\[
\leq C \left\| (T'W/n) \right\| \left\| (W'T/n) \right\| = O_p(\zeta(k_n)^2 n^2)
\]

where the last inequality is due to the fact that $||AB||^2 \leq ||A||^2||B||^2$ (see, e.g., Trefethen and Bau III (1997), p. 23).

As long as condition 4.1 holds, $\zeta(k_n)^2 k_n \sqrt{r_n} \to 0$, and $||\tilde{T}'\tilde{T}/n - T'T/n|| = o_p(1/\sqrt{n})$. This also implies that the smallest and largest eigenvalues of $\tilde{T}'\tilde{T}/n$ converge to one.

Step 2. Decompose the test statistic and bound the remainder terms.

\[
(\varepsilon + R)'(\hat{\sigma}^2\hat{T}'\hat{T})^{-1}\hat{T}' \varepsilon + 2R'(\hat{\sigma}^2\hat{T}'\hat{T})^{-1}\hat{T}' \varepsilon + R'(\hat{\sigma}^2\hat{T}'\hat{T})^{-1}\hat{T}' R
\]

By the projection inequality and Assumption 4,

\[
R'(\hat{\sigma}^2\hat{T}'\hat{T})^{-1}\hat{T}' R \leq R' R/\hat{\sigma}^2 = O_p(n m^{-2})
\]

Because $\tilde{T}'\tilde{T}/n$ and $\hat{T}\hat{T}'/n$ have the same nonzero eigenvalues and all eigenvalues of $\tilde{T}'\tilde{T}/n$ converge to one, $\lambda_{\text{max}}(\tilde{T}\tilde{T}'/n)$ converges in probability to 1. Thus,

\[
\left| R'(\hat{\sigma}^2\hat{T}'\hat{T})^{-1}\hat{T}' \varepsilon \right| \leq C \lambda_{\text{max}}(\hat{T}\hat{T}'/n) R' \varepsilon \leq C R' \varepsilon = C \sum_i R_i \varepsilon_i
\]

Note that

\[
E \left[ \left( \sum_i R_i \varepsilon_i \right)^2 \right] = E \left[ \sum_i \sum_j \varepsilon_i \varepsilon_j R_i R_j \right] = E \left[ \sum_i R_i^2 \varepsilon_i^2 \right]
\]

\[
= n E[R_i^2 \varepsilon_i^2] = n \sigma^2 E[R_i^2] \leq n \sigma^2 \sup_{x \in X} R(x)^2 \leq O(n m^{-2})
\]

by Assumption 4.

Hence,

\[
\left| R'(\hat{\sigma}^2\hat{T}'\hat{T})^{-1}\hat{T}' \varepsilon \right| \leq C \sum_i R_i \varepsilon_i
\]

by $O_p(n^{1/2} m^{-\alpha})$.
Thus,

\[(\varepsilon + R)\tilde{T}(\tilde{\sigma}^{2n2n})^{-1}\tilde{T}'(\varepsilon + R) = \varepsilon'\tilde{T}(\tilde{\sigma}^{2n2n})^{-1}\tilde{T}'\varepsilon + O_p(nm^{-2\alpha}) + O_p(n^{1/2}m^{-\alpha})\]

Next,

\[\varepsilon'\tilde{T}(\tilde{\sigma}^{2n2n})^{-1}\tilde{T}'\varepsilon = \varepsilon'T(\tilde{\sigma}^{2n2n})^{-1}T'\varepsilon - 2\varepsilon'P_W T(\tilde{\sigma}^{2n2n})^{-1}T'\varepsilon + \varepsilon'P_W T(\tilde{\sigma}^{2n2n})^{-1}T'P_W \varepsilon\]

Note that

\[E[(n^{-1}\varepsilon'T)\Omega^{-1}(n^{-1}T'\varepsilon)] = E[\varepsilon_1^2 T_1'\Omega^{-1}T_1]/n = E[tr(\Omega^{-1}\varepsilon_1^2 T_1 T_1')] = tr(I_{r_n})/n = r_n/n\]

Thus, by Markov’s inequality,

\[||\Omega^{-1}(n^{-1}T'\varepsilon)|| \leq C\sqrt{(n^{-1}\varepsilon'T)\Omega^{-1}(n^{-1}T'\varepsilon)} = O_p(\sqrt{r_n/n})\]

Because the eigenvalues of \(\Omega\) are bounded below and above w.p.a. 1, it is also true that \(||n^{-1}T'\varepsilon|| = O_p(\sqrt{r_n/n})\) and \(||n^{-1}W'\varepsilon|| = O_p(\sqrt{m_n/n})\). Using this result and the inequality \(||AB||^2 \leq ||A||^2 ||B||^2\), get

\[\left\|\varepsilon'P_W T(\tilde{\sigma}^{2n2n})^{-1}T'\varepsilon\right\| = \left\|\varepsilon'W(W'W)^{-1}W'T(\tilde{\sigma}^{2n2n})^{-1}T'\varepsilon\right\|
\]

\[= \left\|n(\varepsilon'W/n)(W'T/n)(\tilde{\sigma}^{2n2n})^{-1}(T'\varepsilon/n)\right\|
\]

\[\leq Cn\left\| (\varepsilon'W/n)(W'T/n)(T'\varepsilon/n) \right\| \leq Cn\left\| (\varepsilon'W/n) \right\| \left\| (W'T/n) \right\| \left\| (T'\varepsilon/n) \right\|
\]

\[= nO_p(\sqrt{m_n/n})O_p(\zeta/(k_n)\sqrt{k_n/n})O_p(\sqrt{r_n/n}) = O_p(\zeta(k_n)\sqrt{m_nr_n/n})\]

In turn,

\[\left\|\varepsilon'P_W T(\tilde{\sigma}^{2n2n})^{-1}T'P_W \varepsilon\right\| = \left\|\varepsilon'W(W'W)^{-1}W'T(\tilde{\sigma}^{2n2n})^{-1}T'W(W'W)^{-1}W'\varepsilon\right\|
\]

\[= \left\|n(\varepsilon'W/n)(W'T/n)(\tilde{\sigma}^{2n2n})^{-1}(T'W/n)(W'W)^{-1}(W'\varepsilon/n)\right\|
\]

\[\leq Cn\left\| (\varepsilon'W/n)(W'T/n)(T'W/n)(W'\varepsilon/n) \right\|
\]

\[\leq Cn\left\| (\varepsilon'W/n) \right\| \left\| (W'T/n) \right\| \left\| (T'W/n) \right\| \left\| (W'\varepsilon/n) \right\|
\]

\[= nO_p(\sqrt{m_n/n})O_p(\zeta/(k_n)\sqrt{k_n/n})O_p(\sqrt{m_n/n})O_p(\sqrt{m_n/n}) = O_p(\zeta(k_n)^2m_nk_n/n)\]
Thus, under conditions 4.3 and 4.4,

\[ \varepsilon' P(\hat{\sigma}^2 P')^{-1} P' \varepsilon = \varepsilon' T(\hat{\sigma}^2 \hat{T}^\top)^{-1} T' \varepsilon + O_p(nm_n^{-2\alpha}) + O_p(n^{1/2} m_n^{-\alpha}) \]  \tag{C.1}

+ \text{ and } \min_{\Omega} \quad \text{ and } \min_{\hat{\Omega}}

Thus, under conditions 4.3 and 4.4,

\[ E[(\varepsilon_i-T_i\Omega^{-1}T_i^\top\varepsilon_i)^2]/(r_n\sqrt{n}) \rightarrow 0, \text{ and } r_n \rightarrow \infty, \text{ then} \]

\[ \frac{n(n^{-1}\varepsilon'T)\Omega^{-1}(n^{-1}T'\varepsilon) - r_n}{\sqrt{2r_n}} \xrightarrow{d} N(0,1) \]  \tag{C.2}

All three conditions of this lemma hold. \( E[T_i\varepsilon_i] = 0 \) and \( r_n \rightarrow \infty \) hold trivially, while

\[ E[(\varepsilon_i-T_i\Omega^{-1}T_i^\top\varepsilon_i)^2] \leq CE[\varepsilon_i^2]||T_i||^2 \leq CE[||T_i||^2] \leq C\zeta(r_n)^2r_n \]

Under condition 4.5, the assumptions of Lemma A.1 hold.

Lemma A.2. Suppose that Assumptions 2, 3, and 4 hold. Moreover, \( \sigma^2(x) = \sigma^2 \) for all \( x \in X \). Then

\[ \frac{n(n^{-1}\varepsilon'T)(\hat{\sigma}^2 n^{-1}\hat{T}^\top)^{-1}(n^{-1}T'\varepsilon) - n(n^{-1}\varepsilon'T)\Omega^{-1}(n^{-1}T'\varepsilon)}{\sqrt{n}} \xrightarrow{p} 0 \]  \tag{C.3}

Proof of Lemma A.2. To prove the lemma, I will need an auxiliary result given below.

Lemma A.3. Let \( \hat{\Omega} = \hat{\sigma}^2 \hat{T}^\top/n, \hat{\Omega} = \hat{\sigma}^2 \hat{T}T/n, \Omega = \sigma^2 T'T/n, \Omega = \sigma^2 E[TiT_i^\top] \). Suppose that Assumptions 2(ii), 3, and 4 are satisfied. Then

\[ ||\hat{\Omega} - \hat{\Omega}|| = O_p(\zeta(k_n)^2k/n) \]

\[ ||\hat{\Omega} - \Omega|| = O_p(r_n/n^{1/2}) \]

\[ ||\Omega - \Omega|| = O_p(\zeta(r_n) r_n^{1/2}/n^{1/2}) \]

If Assumption 2(i) is also satisfied then \( 1/C \leq \lambda_{\min}(\hat{\Omega}) \leq \lambda_{\max}(\Omega) \leq C \), and if \( \zeta(k_n)^2k/n \rightarrow 0 \) and \( \zeta(r_n)r_n^{1/2}/n^{1/2} \rightarrow 0 \), then w.p.a. 1, \( 1/C \leq \lambda_{\min}(\hat{\Omega}) \leq \lambda_{\max}(\hat{\Omega}) \leq C \) and \( 1/C \leq \lambda_{\min}(\hat{\Omega}) \leq \lambda_{\max}(\hat{\Omega}) \leq C \).
Proof of Lemma A.3 is presented after the remainder of the main proof.

Given the result of Lemma A.3,

\[
\left| \frac{n(n^{-1} \varepsilon'T)\tilde{\Omega}^{-1}(n^{-1}T'\varepsilon)}{\sqrt{2r_n}} - \frac{n(n^{-1} \varepsilon'T)\Omega^{-1}(n^{-1}T'\varepsilon)}{\sqrt{2r_n}} \right| = \left| \frac{n(n^{-1} \varepsilon'T)\Omega^{-1}(n^{-1}T'\varepsilon)}{\sqrt{2r_n}} \right|
\]

\[
\leq \frac{n(n^{-1} \varepsilon'T)(\Omega^{-1}(\tilde{\Omega} - \Omega)\tilde{\Omega}^{-1}(\tilde{\Omega} - \Omega)\Omega^{-1})(n^{-1}T'\varepsilon)}{\sqrt{2r_n}} - \frac{n(n^{-1} \varepsilon'T)(\Omega^{-1}(\tilde{\Omega} - \Omega)\Omega^{-1})(n^{-1}T'\varepsilon)}{\sqrt{2r_n}}
\]

\[
\leq \frac{n||\Omega^{-1}n^{-1}T'\varepsilon||^2(||\tilde{\Omega} - \Omega|| + C||\tilde{\Omega} - \Omega||^2)}{\sqrt{2r_n}}
\]

As has been shown above, \(||\Omega^{-1}(n^{-1}T'\varepsilon)|| = O_p(\sqrt{r_n/n})\). Hence,

\[
\frac{n||\Omega^{-1}n^{-1}T'\varepsilon||^2(||\tilde{\Omega} - \Omega|| + C||\tilde{\Omega} - \Omega||^2)}{\sqrt{2r_n}} = \frac{nO_p(r_n/n)O_p(1/\sqrt{r_n})}{\sqrt{2r_n}} = o_p(1/\sqrt{r_n}) = o_p(1),
\]

provided that \(||\tilde{\Omega} - \Omega|| = o_p(1/\sqrt{r_n})\), which holds under rate conditions 4.1–4.2.

The result of Theorem 1 now follows from equations C.1, C.2, and C.3.

Proof of Lemma A.3. It has been shown in the proof of Theorem 1 that \(||\tilde{T}'\tilde{T}/n - T'T/n|| = O_p(\zeta(k_n)^2k_n/n)\). As long as \(\sigma^2 \overset{p}{\to} \sigma^2\), this implies \(||\tilde{\Omega} - \tilde{\Omega}|| = O_p(\zeta(k_n)^2k_n/n)\).

Note that \(\tilde{\varepsilon} = \varepsilon + f - \tilde{f}\). Due to homoskedasticity,

\[
||\tilde{\Omega} - \Omega|| = ||(\tilde{\sigma}^2 - \sigma^2)\sum_i T_iT_i'/n|| \leq |\tilde{\sigma}^2 - \sigma^2|\sum_i ||T_i||^2/n
\]

\[
= |n^{-1}\sum_i (\varepsilon_i^2 - \sigma^2) + 2n^{-1}\sum_i \varepsilon_i(f_i - \tilde{f}_i) + n^{-1}\sum_i (f_i - \tilde{f}_i)^2|\sum_i ||T_i||^2/n
\]

First, by Chebyshev’s inequality, \(n^{-1}\sum_i (\varepsilon_i^2 - \sigma^2) = O_p(n^{-1/2})\).

Second, by Assumption 2, \(n^{-1}\sum_i (f_i - \tilde{f}_i)^2 = O_p(m_n/n + m_n^{-2\alpha})\).

Finally, note that

\[
n^{-1}\sum_i \varepsilon_i(f_i - \tilde{f}_i) = n^{-1}\sum_i \varepsilon_i(W'_i(\beta_1 - \tilde{\beta}_1) + R_i) = n^{-1}(\beta_1 - \tilde{\beta}_1)'W'\varepsilon + n^{-1}R'\varepsilon
\]

73
Using the result proved above, \( n^{-1}R'\varepsilon = O_p(n^{-1/2}m_n^{-\alpha}) \).

Next,

\[
\|n^{-1}(\beta_1 - \tilde{\beta}_1)'W'\varepsilon\| = \|n^{-1}(\beta_1 - \tilde{\beta}_1)'W'\varepsilon\| \leq \|\beta_1 - \tilde{\beta}_1\| \|n^{-1}W'\varepsilon\|
\]

\[
= O_p\left((m_n/n + m_n^{-2\alpha})^{1/2}\right) O_p(n^{-1/2}m_n^{1/2}) = O_p\left(m_n^{1/2}(m_n/n + m_n^{-2\alpha})^{1/2}/n^{1/2}\right)
\]

Then

\[
n^{-1}\sum_i \varepsilon_i(f_i - \tilde{f}_i) = O_p(n^{-1/2}m_n^{-\alpha}) + O_p\left(m_n^{1/2}(m_n/n + m_n^{-2\alpha})^{1/2}/n^{1/2}\right)
\]

\[
= O_p\left(m_n^{1/2}(m_n/n + m_n^{-2\alpha})^{1/2}/n^{1/2}\right)
\]

Combining the results,

\[
\sigma^2 - \sigma^2 = O_p(n^{-1/2}) + O_p(m_n/n + m_n^{-2\alpha}) + O_p\left(m_n^{1/2}(m_n/n + m_n^{-2\alpha})^{1/2}/n^{1/2}\right) = O_p(n^{-1/2}),
\]

because \( m_n^{1/2}(m_n/n + m_n^{-2\alpha})^{1/2}/n^{1/2} = o(n^{-1/2}) \) and \( m_n/n + m_n^{-2\alpha} = o(n^{-1/2}) \).

By Lemma 1, \( E[||T_i||^2] \leq r_n \), which yields \( ||\tilde{\Omega} - \Omega|| = O_p\left(r_n n^{-1/2}\right) \).

Next, \( ||\tilde{\Omega} - \Omega|| = ||\sigma^2(T'T/n - E[T_iT_i'])|| \).

Thus,

\[
E[||\tilde{\Omega} - \Omega||^2] = E[||\sigma^2 \sum_i (T_iT_i' - E[T_iT_i'])/n||^2]
\]

\[
= \sigma^4 E[||T_iT_i' - E[T_iT_i']||^2]/n \leq C E[||T_i||^4]/n = C\zeta(r_n)^2 r_n/n,
\]

and hence \( ||\tilde{\Omega} - \Omega|| = O_p(\zeta(r_n)\sqrt{r_n/n}) \).

The remaining conclusions follow from Lemma A.6 in Donald et al. (2003). \( \blacksquare \)

### C.2 Proof of Theorem 2

Given the projection nature of the series estimators, the test statistic becomes

\[
\frac{\xi - r_n}{\sqrt{2r_n}} = \frac{\varepsilon'P(P'\Sigma P)^{-1}P'\varepsilon - r_n}{\sqrt{2r_n}} = \frac{(\varepsilon + R)'M_W P(P'\Sigma P)^{-1}P'M_W(\varepsilon + R) - r_n}{\sqrt{2r_n}}
\]

Next, note that \( P = (W' T) \), so that \( M_W P = (0_{n \times m_n} M_W T) \). Then by the blockwise
matrix inverse formula,

\[ M_W P (P' \Sigma P)^{-1} P' M_W = M_W T (n\hat{\Omega})^{-1} T' M_W, \]

where

\[ \hat{\Omega} = T' \Sigma T / n - (T' \Sigma W / n) (W' \Sigma W / n)^{-1} (W' \Sigma T / n) \]

Thus,

\[ \xi - r_n = (\varepsilon + R)' M_W T (n\hat{\Omega})^{-1} T' M_W (\varepsilon + R) - r_n = (\varepsilon + R)' \tilde{T} (n\hat{\Omega})^{-1} \tilde{T}' (\varepsilon + R) - r_n \]

where \( \tilde{T} = M_W T \). The remainder of the proof consists of several steps.

Step 1. Assume that \( ||\hat{\Omega} - \tilde{\Omega}|| = o_p(1/\sqrt{r_n}) \), where \( \tilde{\Omega} = T' \Sigma T \). This is not a primitive assumption; however, it is difficult to derive more primitive sufficient conditions.

Step 2. Use the following auxiliary result:

**Lemma A.4.** Let \( \hat{\Omega} = \sum_i \hat{\epsilon}_i^2 T_i T_i' / n, \tilde{\Omega} = \sum_i \epsilon_i^2 T_i T_i' / n, \bar{\Omega} = \sum_i \sigma_i^2 T_i T_i' / n, \Omega = E[\varepsilon_i^2 T_i T_i'] \), where \( \sigma_i^2 = E[\varepsilon_i^2 | X_i] \). Suppose that Assumptions 2(ii), 3, and 4 are satisfied. Then

\[
||\hat{\Omega} - \tilde{\Omega}|| = O_p (\zeta (r_n)^2 (m_n / n + m_n^{-2\alpha})) \\
||\hat{\Omega} - \bar{\Omega}|| = O_p (\zeta (r_n) r_n^{1/2} / n^{1/2}) \\
||\bar{\Omega} - \Omega|| = O_p (\zeta (r_n) r_n^{1/2} / n^{1/2})
\]

If Assumption 2(i) is also satisfied then \( 1/C \leq \lambda_{\min}(\hat{\Omega}) \leq \lambda_{\max}(\hat{\Omega}) \leq C \), and if \( \zeta (r_n)^2 (m_n / n + m_n^{-2\alpha}) \to 0 \) and \( \zeta (r_n)^{1/2} / n^{1/2} \to 0 \), then w.p.a. 1, \( 1/C \leq \lambda_{\min}(\tilde{\Omega}) \leq \lambda_{\max}(\tilde{\Omega}) \leq C \) and \( 1/C \leq \lambda_{\min}(\bar{\Omega}) \leq \lambda_{\max}(\bar{\Omega}) \leq C \).

Moreover, if \( ||\hat{\Omega} - \tilde{\Omega}|| = o_p(1) \), then \( 1/C \leq \lambda_{\min}(\hat{\Omega}) \leq \lambda_{\max}(\hat{\Omega}) \leq C \).

Step 3. Decompose the test statistic and bound the remainder terms.

\[
(\varepsilon + R)' \tilde{T} (n\hat{\Omega})^{-1} \tilde{T}' (\varepsilon + R) = \varepsilon' \tilde{T} (n\hat{\Omega})^{-1} \tilde{T}' \varepsilon + 2R' \tilde{T} (n\hat{\Omega})^{-1} \tilde{T}' \varepsilon + R' \tilde{T} (n\hat{\Omega})^{-1} \tilde{T}' R
\]

Because \( \tilde{T}' \tilde{T} / n \) and \( \tilde{T} \tilde{T}' / n \) have the same nonzero eigenvalues and all eigenvalues of \( \tilde{T}' \tilde{T} / n \) converge to one, \( \lambda_{\max}(\tilde{T} \tilde{T}' / n) \) converges in probability to 1. Moreover, the eigenvalues of \( \hat{\Omega} \)
are bounded below and above. Thus, by Assumption 4,

$$R'\tilde{T}(n\hat{\Omega})^{-1}\tilde{T}'R \leq CR'(n^{-1}\tilde{T}'\tilde{T})R/\tilde{\sigma}^2 \leq CR'\tilde{\sigma}^2 = O_p(nm_n^{-2\alpha})$$

Next,

$$\left| R'\tilde{T}(n\hat{\Omega})^{-1}\tilde{T}'\varepsilon \right| \leq C\lambda_{\text{max}}(\tilde{T}'\tilde{T}/n)R'\varepsilon \leq C\varepsilon' = O_p(n^{1/2}m_n^{-\alpha})$$

In turn,

$$\varepsilon'\tilde{T}(n\hat{\Omega})^{-1}\tilde{T}'\varepsilon = \varepsilon'T(n\hat{\Omega})^{-1}T'\varepsilon - 2\varepsilon'P_WT(n\hat{\Omega})^{-1}T'\varepsilon + \varepsilon'P_WT(n\hat{\Omega})^{-1}T'P_W\varepsilon$$

Next,

$$\left| \varepsilon'P_WT(n\hat{\Omega})^{-1}T'\varepsilon \right| = \left| \varepsilon'W(W'W)^{-1}W'T(n\hat{\Omega})^{-1}T'\varepsilon \right|$$

$$= \left| n(\varepsilon'W/n)(W'W/n)^{-1}(W'T/n)\hat{\Omega}^{-1}(T'\varepsilon/n) \right|$$

$$\leq Cn\left| (\varepsilon'W/n)(W'T/n)(T'\varepsilon/n) \right| \leq Cn\left| (\varepsilon'W/n) \right| \left| (W'T/n) \right| \left| (T'\varepsilon/n) \right|$$

$$= nO_p(\sqrt{m_n/n})O_p(\zeta(k_n)\sqrt{k_n/n})O_p(\sqrt{r_n/n}) = O_p(\zeta(k_n)\sqrt{m_nk_nr_n/n})$$

In turn,

$$\left| \varepsilon'P_WT(n\hat{\Omega})^{-1}T'P_W\varepsilon \right| = \left| \varepsilon'W(W'W)^{-1}W'T(n\hat{\Omega})^{-1}T'W(W'W)^{-1}W'\varepsilon \right|$$

$$= \left| n(\varepsilon'W/n)(W'W/n)^{-1}(W'T/n)\hat{\Omega}^{-1}(T'W/n)(W'W/n)^{-1}(W'\varepsilon/n) \right|$$

$$\leq Cn\left| (\varepsilon'W/n)(W'T/n)(T'W/n)(W'\varepsilon/n) \right|$$

$$\leq Cn\left| (\varepsilon'W/n) \right| \left| (W'T/n) \right| \left| (T'W/n) \right| \left| (W'\varepsilon/n) \right|$$

$$= nO_p(\sqrt{m_n/n})O_p(\zeta(k_n)\sqrt{k_n/n})O_p(\zeta(k_n)\sqrt{m_n/k_n})O_p(\sqrt{m_n/n}) = O_p(\zeta(k_n)^2m_nk_n/n)$$

Thus, under conditions 4.8 and 4.9

$$\varepsilon'P(P'S\hat{\Sigma}P)^{-1}P'\varepsilon = \varepsilon'T(n\hat{\Omega})^{-1}T'\varepsilon + O_p(nm_n^{-2\alpha}) + O_p(n^{1/2}m_n^{-\alpha})$$

$$+ O_p(\zeta(k_n)^2m_nk_n/n) + O_p(\zeta(k_n)\sqrt{m_nk_nr_n/n}) = \varepsilon'T(n\hat{\Omega})^{-1}T'\varepsilon + o_p(\sqrt{r_n}) \quad (C.4)$$

Step 4. Deal with the leading term.

Conditions of Lemma A.1 do not rely on the homoskedasticity assumption. Thus, the
Proof of Lemma A.4. First,

\[
\frac{n(n^{-1} \varepsilon' T) \Omega^{-1}(n^{-1} T' \varepsilon) - r_n}{\sqrt{2r_n}} \overset{d}{\to} N(0, 1) \tag{C.5}
\]

Next, as in the proof of Theorem 1,

\[
\frac{n(n^{-1} \varepsilon' T) \hat{\Omega}^{-1}(n^{-1} T' \varepsilon) - n(n^{-1} \varepsilon' T) \Omega^{-1}(n^{-1} T' \varepsilon)}{\sqrt{2r_n}} \leq \frac{n||\Omega^{-1}n^{-1}T'\varepsilon||^2(||\hat{\Omega} - \Omega|| + C||\hat{\Omega} - \Omega||^2)}{\sqrt{2r_n}}
\]

Similarly to the proof of Theorem 1, \(||\Omega^{-1}(n^{-1}T'\varepsilon)|| = O_p(\sqrt{r_n/n})\). Then

\[
\frac{n||\Omega^{-1}n^{-1}T'\varepsilon||^2(||\hat{\Omega} - \Omega|| + C||\hat{\Omega} - \Omega||^2)}{\sqrt{2r_n}} = \frac{nO_p(r_n/n)O_p(1/\sqrt{r_n})}{\sqrt{2r_n}} = o_p(\sqrt{r_n/n}) = o_p(1),
\]

provided that \(||\hat{\Omega} - \Omega|| = o_p(1/\sqrt{r_n})\), which holds under rate conditions 4.7, 4.8, and 4.8.

Hence,

\[
\frac{n(n^{-1} \varepsilon' T) \hat{\Omega}^{-1}(n^{-1} T' \varepsilon) - n(n^{-1} \varepsilon' T) \Omega^{-1}(n^{-1} T' \varepsilon)}{\sqrt{r_n}} \overset{p}{\to} 0 \tag{C.6}
\]

The result of Theorem 2 now follows from equations C.4, C.5, and C.6. ■

Proof of Lemma A.4. First,

\[
||\hat{\Omega} - \Omega|| = ||\sum T_i T_i' (\varepsilon_i^2 - \hat{\varepsilon}_i^2)/n|| = ||\sum T_i T_i' (\varepsilon_i + f_i - \hat{f}_i)^2 - \hat{\varepsilon}_i^2)/n|| = ||\sum T_i T_i' ((f_i - \hat{f}_i)^2 + 2 \varepsilon_i (f_i - \hat{f}_i))/n|| \leq \sup \|T_i\|^2 \sum ((f_i - \hat{f}_i)^2 + 2 \varepsilon_i (f_i - \hat{f}_i))/n||
\]

\[
= \zeta(r_n)^2 [O_p(m_n/n + m_n^{-2\alpha}) + O_p(n^{-1/2}m_n^{1/2}(m_n/n + m_n^{-2\alpha})^{1/2})] = O_p(\zeta(r_n)^2(m_n/n + m_n^{-2\alpha}))
\]

The following two results can be obtained exactly as in Lemma A.6 in Donald et al.
Finally, the results about the eigenvalues can also be obtained in the same way as in Lemma A.6 in Donald et al. (2003).

\[ ||\hat{\Omega} - \bar{\Omega}|| = || \sum \frac{T_iT_i'(\varepsilon_i^2 - \sigma_i^2)}{n} || = O_p(\sqrt{r_n/n}) \]
\[ ||\bar{\Omega} - \Omega|| = || \sum \frac{T_iT_i\varepsilon_i^2}{n - \Omega} || = O_p(\sqrt{r_n/n}) \]

(C.3) Proof of Theorem 3

The proof is analogous to the proof of Theorem 2. The only difference is that in Step 1 of the proof, the definition of \(\hat{\Omega}\) is different. The result \(||\hat{\Omega} - \bar{\Omega}|| = o_p(1/\sqrt{r_n})\) is imposed directly. I plan to derive the primitive conditions in future work.

(C.4) Proof of Theorem 4

Denote \(\hat{\Omega}_P = (\hat{\sigma}^2 n^{-1}P'P)\) in the homoskedastic case and \(\hat{\Omega}_P = (n^{-1}P'\Sigma P)\) in the heteroskedastic case. Then under homoskedasticity
\[ \xi = n(n^{-1}\varepsilon'P)(\hat{\sigma}^2 n^{-1}P'P)^{-1}(n^{-1}P'\varepsilon) = n(n^{-1}\varepsilon'P)\hat{\Omega}_P^{-1}(n^{-1}P'\varepsilon), \]
while under heteroskedasticity
\[ \xi_{HC,1} = n(n^{-1}\varepsilon'P)(n^{-1}P'\Sigma P)^{-1}(n^{-1}P'\varepsilon) = n(n^{-1}\varepsilon'P)\hat{\Omega}_P^{-1}(n^{-1}P'\varepsilon) \]

Note that
\[ \frac{\sqrt{r_n} n(n^{-1}\varepsilon'P)\hat{\Omega}_P^{-1}(n^{-1}P'\varepsilon) - r_n}{\sqrt{2r_n}} = \frac{1}{\sqrt{2}}(n^{-1}\varepsilon'P)\hat{\Omega}_P^{-1}(n^{-1}P'\varepsilon) + T_2, \]
where \(T_2 = -r_n/(n\sqrt{2}) \rightarrow 0.\)

Hence, it suffices to show that \((n^{-1}\varepsilon'P)\hat{\Omega}_P^{-1}(n^{-1}P'\varepsilon) \overset{p}{\rightarrow} \Delta.\)

Next, note that due to the projection nature of the series estimators, \(\varepsilon = M_WY = \)
$M_W(\varepsilon^* + R^*)$. Hence,

$$(n^{-1}\hat{\varepsilon}'P)\hat{\Omega}_P^{-1}(n^{-1}P'\hat{\varepsilon}) = (n^{-1}(\varepsilon^* + R^*)'M_WT)\hat{\Omega}^{-1}(n^{-1}T'M_W(\varepsilon^* + R^*)),$$

where $\hat{\Omega}$ is defined in the statement of Theorem 4.

Note that

$$\xi_{HC,2} = n(n^{-1}\hat{\varepsilon}'M_WT)(n^{-1}T'M_W\hat{\Sigma}M_WT)^{-1}(n^{-1}T'M_W\hat{\varepsilon}) = n(n^{-1}(\varepsilon^* + R^*)'M_WT)\hat{\Omega}^{-1}(n^{-1}T'M_W(\varepsilon^* + R^*)),$$

for the appropriately defined $\hat{\Omega}$ in the statement of Theorem 4. Thus, for all three test statistics $\xi$, $\xi_{HC,1}$, and $\xi_{HC,2}$ it suffices to show that

$$(n^{-1}(\varepsilon^* + R^*)'M_WT)\hat{\Omega}^{-1}(n^{-1}T'M_W(\varepsilon^* + R^*)) \xrightarrow{p} \Delta$$

Next,

$$
(n^{-1}(\varepsilon^* + R^*)'M_WT)\hat{\Omega}^{-1}(n^{-1}T'M_W(\varepsilon^* + R^*)) = (n^{-1}\varepsilon''M_WT)\hat{\Omega}^{-1}(n^{-1}T'M_W\varepsilon^*) \\
+ (n^{-1}R''M_WM_WT)\hat{\Omega}^{-1}(n^{-1}T'M_W\varepsilon^*) + (n^{-1}R''M_WM_WM_W(\varepsilon^* + R^*))
$$

Similarly to the proofs of Theorems 1 and 2, but using the fact that $\sup_{x\in\mathcal{X}} R^*(x) = o(1)$ instead of $\sup_{x\in\mathcal{X}} R(x) = O(m_{\alpha^2})$,

$$
(n^{-1}R''M_WM_WT)\hat{\Omega}^{-1}(n^{-1}T'M_WM_W\varepsilon^*) \leq CR''R^*/(n\hat{\sigma}^2) \leq O_p(n^{-1/2})o_p(1) = o_p(1)
$$

and

$$
(n^{-1}R''M_WM_WM_WM_W(\varepsilon^* + R^*)) \leq CR''R^*/(n\hat{\sigma}^2) = o_p(1)
$$

Thus,

$$(n^{-1}(\varepsilon^* + R^*)'M_WM_WM_WM_WM_W(\varepsilon^* + R^*)) = (n^{-1}\varepsilon''M_WM_WM_WM_WM_WM_WM_WM_WM_WM_WM_WM_WM_WM_W)\hat{\Omega}^{-1}(n^{-1}T'M_WM_WM_WM_WM_WM_WM_WM_WM_WM_WM_WM_WM_WM_WM_WM_WM_WM_WM_WM_W) + o_p(1)$$


79
eigenvalues of $\hat{\Omega}$ are bounded above and below w.p.a. 1,

$$(n^{-1}e^{-*}M_W T)\hat{\Omega}^{-1}(n^{-1}T'M_W e^*) = (n^{-1}e^{-*}T)\hat{\Omega}^{-1}(n^{-1}T'e^*) + o_p(1).$$

Next,

$$\left|(n^{-1}e^{-*}T)(\hat{\Omega}^{-1} - \Omega^{-1})(n^{-1}T'e^*)\right| \leq \left|(n^{-1}e^{-*}T)(\Omega^{-1}(\hat{\Omega} - \Omega^*)\hat{\Omega}^{-1} - \Omega^{-1})(n^{-1}T'e^*)\right|$$

$$+ \left|(n^{-1}e^{-*}T)(\Omega^{-1}(\hat{\Omega} - \Omega^*)\hat{\Omega}^{-1})(n^{-1}T'e^*)\right| \leq \|\Omega^{-1}n^{-1}T'e^*\|^2(||\hat{\Omega} - \Omega^*|| + C||\hat{\Omega} - \Omega^*||^2) = o_p(1).$$

Thus, $(n^{-1}e^{-*}T)\hat{\Omega}^{-1}(n^{-1}T'e^*) = (n^{-1}e^{-*}T)\Omega^{-1}(n^{-1}T'e^*) + o_p(1)$.

To complete the proof, note that $\text{Var}(T_ie_i^*) \leq \Omega^*$, because $\Omega^* = E[e_i^2T_iT_i']$. Then

$$E[(n^{-1}T'i^* - E[T_i^*])'\Omega^{-1}(n^{-1}T'i^* - E[T_i^*])]$$

$$\leq E[(n^{-1}T'i^* - E[T_i^*])'\text{Var}(T_i^*)^{-1}(n^{-1}T'i^* - E[T_i^*])]$$

$$= E[\text{tr}(\text{Var}(T_i^*)^{-1}(n^{-1}T'i^* - E[T_i^*])(n^{-1}T'i^* - E[T_i^*])')] = \text{tr}(I_{p_n})/n = r_n/n \to 0$$

Thus,

$$\left|(n^{-1}e^{-*}T)\Omega^{-1}(n^{-1}T'e^*) - E[e_i^*T_i']\Omega^{-1}E[T_i^*]\right|$$

$$\leq \left|(n^{-1}T'e^* - E[T_i^*])'\Omega^{-1}(n^{-1}T'e^* - E[T_i^*])\right| + 2\left|E[e_i^*T_i']\Omega^{-1}(n^{-1}T'e^* - E[T_i^*])\right|$$

$$\leq o_p(1) + 2\sqrt{E[e_i^*T_i']\Omega^{-1}E[T_i^*]}\sqrt{(n^{-1}T'e^* - E[T_i^*])'\Omega^{-1}(n^{-1}T'e^* - E[T_i^*])}$$

$$= o_p(1) + 2\sqrt{\Delta}o_p(1) = o_p(1).$$

Combining the results above, $(n^{-1}\varepsilon'P)\hat{\Omega}_P^{-1}(n^{-1}P'\varepsilon) \xrightarrow{P} \Delta.$

## C.5 Proof of Theorem 5

Because series methods are used, it is still the case that $\varepsilon = M_W Y$. Under the local alternative,

$$\varepsilon = M_W Y = M_W(g_n + \varepsilon) = M_W(f^* + (r_n^{1/4}/n^{1/2})d + \varepsilon)$$

$$= M_W (W'\beta^*_1 + (f^* - W'\beta^*_1) + (r_n^{1/4}/n^{1/2})d + \varepsilon) = M_W (\varepsilon + R + (r_n^{1/4}/n^{1/2})d)$$
Thus, the test statistic becomes

\[ \frac{\xi - r_n}{\sqrt{2r_n}} = \frac{\varepsilon' P (\hat{\sigma}^2 P' P)^{-1} P' \varepsilon - r_n}{\sqrt{2r_n}} \]

\[ = \frac{(\varepsilon + R + (r_n^{1/4}/n^{1/2}) d') M_W P (\hat{\sigma}^2 P' P)^{-1} P' M_W (\varepsilon + R + (r_n^{1/4}/n^{1/2}) d) - r_n}{\sqrt{2r_n}} \]

\[ = \frac{(\varepsilon + R + (r_n^{1/4}/n^{1/2}) d') M_W T (\hat{\sigma}^2 T' M_W T)^{-1} T' M_W (\varepsilon + R + (r_n^{1/4}/n^{1/2}) d) - r_n}{\sqrt{2r_n}} \]

\[ = \frac{(\varepsilon + R + (r_n^{1/4}/n^{1/2}) d') T(\hat{\sigma}^2 T' T)^{-1} T' (\varepsilon + R + (r_n^{1/4}/n^{1/2}) d) - r_n}{\sqrt{2r_n}} , \]

where \( \hat{T} = M_W T \). The remainder of the proof consists of several steps.

Step 1. It has been shown in the proof of Theorem 1 that \(||\hat{T}' \hat{T}/n - T' T/n|| = o_p(1/\sqrt{r_n})\) as long as condition 4.1 holds, \( \zeta (k_n)^2 k_n \sqrt{r_n}/n \to 0 \). This also implies that the smallest and largest eigenvalues of \( \hat{T}' \hat{T}/n \) converge to one.

Step 2. Decompose the test statistic and bound the remainder terms.

\[ (\varepsilon + R + (r_n^{1/4}/n^{1/2}) d') T(\hat{\sigma}^2 T' T)^{-1} T' (\varepsilon + R + (r_n^{1/4}/n^{1/2}) d) \]

\[ = \varepsilon' T(\hat{\sigma}^2 T' T)^{-1} T' \varepsilon + 2 R' T(\hat{\sigma}^2 T' T)^{-1} T' \varepsilon + R' T(\hat{\sigma}^2 T' T)^{-1} T' R \]

\[ + 2(r_n^{1/4}/n^{1/2}) d' T(\hat{\sigma}^2 T' T)^{-1} T' \varepsilon + 2(r_n^{1/4}/n^{1/2}) d' T(\hat{\sigma}^2 T' T)^{-1} T' R + (r_n^{1/2}/n) d' T(\hat{\sigma}^2 T' T)^{-1} T' d \]

As in the proof of Theorem 1, \( R' T(\hat{\sigma}^2 T' T)^{-1} T' R = O_p(n m_n^{-2\alpha}) \) and \( R' T(\hat{\sigma}^2 T' T)^{-1} T' \varepsilon \) = \( O_p(n^{1/2} m_n^{-\alpha}) \).

Recall that, because \( \hat{T}' \hat{T}/n \) and \( \hat{T}' \hat{T}/n \) have the same nonzero eigenvalues and all eigenvalues of \( \hat{T}' \hat{T}/n \) converge to one, \( \lambda_{\text{max}}(\hat{T}' \hat{T}/n) \) converges in probability to 1.

Thus, as for the fourth term,

\[ \left| (r_n^{1/4}/n^{1/2}) d' T(\hat{\sigma}^2 T' T)^{-1} T' \varepsilon \right| \leq (r_n^{1/4}/n^{1/2}) \left| C \lambda_{\text{max}}(\hat{T}' \hat{T}/n) d' \varepsilon \right| \]

\[ \leq (r_n^{1/4}/n^{1/2}) \left| C d' \varepsilon \right| = (r_n^{1/4}/n^{1/2}) O_p(n^{1/2}) = O_p(r_n^{1/4}) \]

As for the fifth term,

\[ \left| (r_n^{1/4}/n^{1/2}) d' T(\hat{\sigma}^2 T' T)^{-1} T' R \right| \leq (r_n^{1/4}/n^{1/2}) \left| C \lambda_{\text{max}}(\hat{T}' \hat{T}/n) d' R \right| \leq (r_n^{1/4}/n^{1/2}) \left| C d' R \right| \]
By the Cauchy-Schwartz inequality,
\[
\left| \frac{r_n^{1/4}}{n^{1/2}} d' R \right| \leq \left( \frac{r_n^{1/4}}{n^{1/2}} \right) \sqrt{\left( \sum_i R_i^2 / n \right) \left( \sum_i d_i^2 / n \right)}
\]
\[
= \left( \frac{r_n^{1/4}}{n^{1/2}} \right) \sqrt{O_p(m_n^{-2\alpha}) E[d_i^2/n]} (1 + o_p(1)) = O_p(r_n^{1/4} m_n^{-\alpha} n^{1/2})
\]

As for the last term,
\[
\left( \frac{r_n^{1/2}}{n} \right) d' \tilde{\mathbf{T}} (\tilde{\sigma}^2 \tilde{T}'\tilde{T})^{-1} \tilde{T}' d = \left( \frac{r_n^{1/2}}{n} \right) d' T (\tilde{\sigma}^2 \tilde{T}'\tilde{T})^{-1} T' d
\]
\[
+ \left( \frac{r_n^{1/2}}{n} \right) d' P_W T (\tilde{\sigma}^2 \tilde{T}'\tilde{T})^{-1} T' P_W d - 2 \left( \frac{r_n^{1/2}}{n} \right) d' P_W T (\tilde{\sigma}^2 \tilde{T}'\tilde{T})^{-1} T' d
\]

Using Assumption 6,
\[
\left\| \left( \frac{r_n^{1/2}}{n} \right) d' P_W T (\tilde{\sigma}^2 \tilde{T}'\tilde{T})^{-1} T' d \right\| = \left( \frac{r_n^{1/2}}{n} \right) d' W (W' W)^{-1} W' T (\tilde{\sigma}^2 \tilde{T}'\tilde{T})^{-1} T' d
\]
\[
= \left( \frac{r_n^{1/2}}{n} \right) \left\| d' W (W' W)^{-1} W' T (\tilde{\sigma}^2 \tilde{T}'\tilde{T})^{-1} T' d / n \right\|
\]
\[
\leq C r_n^{1/2} \left\| (d' W / n) (W' T / n) (T' d / n) \right\| \leq C r_n^{1/2} \left\| (d' W / n) \right\| \left\| (W' T / n) \right\| \left\| (T' d / n) \right\|
\]
\[
= r_n^{1/2} O_p(\sqrt{m_n / n}) O_p(\zeta(k_n) \sqrt{k_n / n}) O_p(\sqrt{r_n / n}) = O_p(\zeta(k_n) r_n^{3/2} m_n^{1/2} k_n^{1/2} / n^{3/2})
\]

Next,
\[
\left\| \left( \frac{r_n^{1/2}}{n} \right) d' P_W T (\tilde{\sigma}^2 \tilde{T}'\tilde{T})^{-1} T' P_W d \right\| = \left( \frac{r_n^{1/2}}{n} \right) d' W (W' W)^{-1} W' T (\tilde{\sigma}^2 \tilde{T}'\tilde{T})^{-1} T' W (W' W)^{-1} W' d
\]
\[
= \left( \frac{r_n^{1/2}}{n} \right) \left\| (d' W / n) (W' W / n)^{-1} (W' W / n)^{-1} (T' W / n) (W' W / n)^{-1} (W' d / n) \right\|
\]
\[
\leq C r_n^{1/2} \left\| (d' W / n) (W' T / n) (T' W / n) (W' d / n) \right\|
\]
\[
\leq C r_n^{1/2} \left\| (d' W / n) \right\| \left\| (W' T / n) \right\| \left\| (T' W / n) \right\| \left\| (W' d / n) \right\|
\]
\[
= r_n^{1/2} O_p(\sqrt{m_n / n}) O_p(\zeta(k_n) \sqrt{k_n / n}) O_p(\zeta(k_n) \sqrt{k_n / n}) O_p(\sqrt{m_n / n}) = O_p(\zeta(k_n)^2 r_n^{1/2} m_n k_n / n^2)
\]

Thus, under rate condition 4.3,
\[
\left( \frac{r_n^{1/2}}{n} \right) d' \tilde{\mathbf{T}} (\tilde{\sigma}^2 \tilde{T}'\tilde{T})^{-1} \tilde{T}' d = \left( \frac{r_n^{1/2}}{n} \right) d' T (\tilde{\sigma}^2 \tilde{T}'\tilde{T})^{-1} T' d + o_p(r_n^{1/2})
\]
Next, similarly to the proof of Theorem 1,

\[
\frac{r_n^{1/2}(n^{-1}dT)(n^{-1}\hat{\sigma}^2\tilde{T}'\tilde{T})^{-1}(n^{-1}T'd)}{\sqrt{2r_n}} - \frac{r_n^{1/2}(n^{-1}dT)(n^{-1}\sigma^2T'T)^{-1}(n^{-1}T'd)}{\sqrt{2r_n}}
\]

\[
\leq \frac{r_n^{1/2}||n^{-1}\hat{\sigma}^2\tilde{T}'\tilde{T})^{-1}n^{-1}T'\varepsilon||^2 (||n^{-1}\hat{\sigma}^2\tilde{T}'\tilde{T}) - (n^{-1}\sigma^2T'T)|| + C|| (n^{-1}\hat{\sigma}^2\tilde{T}'\tilde{T}) - (n^{-1}\sigma^2T'T)||^2}{\sqrt{2r_n}}
\]

\[
= \frac{r_n^{1/2}O_p(r_n/n)O_p(1/\sqrt{r_n})}{\sqrt{2r_n}} = o_p(r_n^{1/2}/n) = o_p(1)
\]

Next,

\[
(r_n^{1/2}/n)d'T(T'T)^{-1}T'd/\hat{\sigma}^2 = (r_n^{1/2}/n)(d'T(T'T)^{-1}T'(T(T'T)^{-1}T'd)/\hat{\sigma}^2 = (k_n^{1/2}/n)d\hat{d}/\hat{\sigma}^2,
\]

where \(d\hat{d} = T(T'T)^{-1}T'd\) are the fitted values from the nonparametric series regression of \(d\) on \(T\).

By Lemma 4, \((d\hat{d} - d)'(d\hat{d} - d) = O_p(n(r_n/n + r_n^{-2\alpha}))\).

By the Cauchy-Schwarz inequality and Lemma 4,

\[
|d'(d\hat{d} - d)| \leq n\sqrt{\left(\sum_i d_i^2/n\right) \left(\sum_i (d_i - \hat{d}_i)^2/n\right)}
\]

\[
= n\sqrt{E[d_i^2]/(1 + o_p(1))O_p(n(r_n/n + r_n^{-2\alpha}))} = O_p(n(r_n/n + r_n^{-2\alpha})^{1/2})
\]

Thus,

\[
(r_n^{1/2}/n)d\hat{T}(\sigma^2\tilde{T}'\tilde{T})^{-1}\tilde{T}'d = (r_n^{1/2}/n)d\hat{d}/\hat{\sigma}^2 + O_p\left(r_n^{1/2}(r_n/n + r_n^{-2\alpha})^{1/2}\right)
\]

\[
= r_n^{1/2}E[d_i^2]/(1 + o_p(1))/\sigma^2 + O_p\left(r_n^{1/2}(r_n/n + r_n^{-2\alpha})^{1/2}\right) + o_p(r_n^{1/2})
\]

where the last equality is due to the law of large numbers and \(\hat{\sigma}^2 = \sigma^2 + o_p(1)\).

Thus,

\[
\varepsilon'P(\hat{\sigma}^2\hat{P}'P)^{-1}\hat{P}'\varepsilon = \varepsilon'\tilde{T}(\hat{\sigma}^2\tilde{T}'\tilde{T})^{-1}\tilde{T}'\varepsilon + r_n^{1/2}E[d_i^2]/\sigma^2 + O_p(nm_n^{-2\alpha}) + O_p(n^{1/2}m_n^{-\alpha})
\]

\[
+ O_p(r_n^{1/4}) + O_p(r_n^{1/4}m_n^{-\alpha}n^{1/2}) + O_p(r_n^{1/2}(r_n/n + r_n^{-2\alpha})^{1/2}) + o_p(r_n^{1/2})
\]

83
Next, as has been shown in the proof of Theorem 1,

\[
\tilde{\varepsilon}' P(\tilde{\sigma}^2 P' P)^{-1} P' \tilde{\varepsilon} = \varepsilon'T(\sigma^2 \tilde{T}' \tilde{T})^{-1} T' \varepsilon + O_p(nm_n^{-2\alpha}) + O_p(n^{1/2}m_n^{-\alpha}) \\
+ O_p(\zeta(k_n)^2m_n^2n_k/n) + O_p(\zeta(k_n)\sqrt{m_n}knr_n/n)
\]

Under rate conditions 4.3 and 4.4,

\[
\tilde{\varepsilon}' P(\tilde{\sigma}^2 P' P)^{-1} P' \tilde{\varepsilon} = \varepsilon'T(\tilde{\sigma}^2 \tilde{\tilde{T}}' \tilde{T})^{-1} T' \varepsilon + \tilde{T}' \tilde{\varepsilon} + O_p(nm_n^{-2\alpha}) + O_p(n^{1/2}m_n^{-\alpha}) \\
+ O_p(\zeta(k_n)^2m_n^2n_k/n) + O_p(\zeta(k_n)\sqrt{m_n}knr_n/n)
\]

Step 3. Deal with the leading term.

The result of Lemma A.1 applies, so that

\[
n(n^{-1}\varepsilon'T)\Omega^{-1}(n^{-1}T'\varepsilon) - r_n \xrightarrow{d} N(0,1)
\]

Next, use the following lemma:

**Lemma A.5.** Suppose that Assumptions 2, 3, and 8 hold Then

\[
\frac{n(n^{-1}\varepsilon'T)(\tilde{\sigma}^2 n^{-1}\tilde{T}' \tilde{T})^{-1}(n^{-1}T'\varepsilon) - n(n^{-1}\varepsilon'T)\Omega^{-1}(n^{-1}T'\varepsilon)}{\sqrt{T_n}} \xrightarrow{p} 0
\]

**Proof of Lemma A.5.** To prove the lemma, I will again need an auxiliary result given in the following lemma.

**Lemma A.6.** Let \( \tilde{\Omega} = \tilde{\sigma}^2 \tilde{T}' \tilde{T} / n, \tilde{\tilde{\Omega}} = \tilde{\sigma}^2 \tilde{T}' \tilde{T} / n, \tilde{\Omega} = \sigma^2 T' T / n, \Omega = \sigma^2 E[T_i T_i'] \). Suppose that Assumptions 2(ii), 3, and 8 are satisfied. Then

\[
||\tilde{\Omega} - \tilde{\tilde{\Omega}}|| = O_p(\zeta(k_n)^2k_n/n) \\
||\tilde{\Omega} - \tilde{\Omega}|| = O_p(r_n n^{-1/2}) \\
||\tilde{\Omega} - \tilde{\Omega}|| = O_p(\zeta(r_n)r_n^{1/2}/n^{1/2})
\]

If Assumption 2(i) is also satisfied then \( 1/C \leq \lambda_{\min}(\Omega) \leq \lambda_{\max}(\Omega) \leq C \), and if \( \zeta(k_n)^2k_n/n \to 0 \) and \( \zeta(r_n)r_n^{1/2}/n^{1/2} \to 0 \), then w.p.a. 1, \( 1/C \leq \lambda_{\min}(\tilde{\Omega}) \leq \lambda_{\max}(\tilde{\Omega}) \leq C \) and \( 1/C \leq \lambda_{\min}(\tilde{\tilde{\Omega}}) \leq \lambda_{\max}(\tilde{\tilde{\Omega}}) \leq C \).

Proof of Lemma A.6 is presented after the remainder of the main proof.
Given the result of Lemma A.6, as in the proof of Theorem 1,

\[
\left| \frac{n(n^{-1}\varepsilon'T)\tilde{\Omega}^{-1}(n^{-1}T'\varepsilon)}{\sqrt{2r_n}} - \frac{n(n^{-1}\varepsilon'T)\Omega^{-1}(n^{-1}T'\varepsilon)}{\sqrt{2r_n}} \right| \leq \frac{n||\Omega^{-1}n^{-1}T'\varepsilon||^2(||\tilde{\Omega} - \Omega|| + C||\tilde{\Omega} - \Omega||^2)}{\sqrt{2r_n}}
\]

\[
= \frac{nO_p(r_n/n)o_p(1/\sqrt{r_n})}{\sqrt{2r_n}} = o_p(1),
\]

provided that \(||\tilde{\Omega} - \Omega|| = o_p(1/\sqrt{r_n})\), which holds under rate conditions which holds under rate conditions 4.1–4.2. \(\blacksquare\)

The result of Theorem 5 now follows from equations C.7, C.8, and C.9. \(\blacksquare\)

**Proof of Lemma A.6.** It has been shown in the proof of Theorem 1 that \(||\tilde{T}'T/n - T'T/n|| = O_p(\zeta(\kappa_n)^2\kappa_n/n)\).

As long as \(\tilde{\sigma}^2 \xrightarrow{P} \sigma^2\), this implies \(||\tilde{\Omega} - \tilde{\Omega}|| = O_p(\zeta(\kappa_n)^2\kappa_n/n)\).

Under the local alternative, \(\tilde{\varepsilon} = \varepsilon + (g_n - \tilde{f}) = \varepsilon + (f_n^* - \tilde{f}_n) + (r_1^{1/4}/n^{1/2})d\)

Thus, \(\tilde{\sigma}^2 = \varepsilon'\varepsilon/n = \varepsilon'\varepsilon/n + (f_n^* - \tilde{f}_n)'(f_n^* - \tilde{f}_n)/n + (r_1^{1/2}/n)d'd/n + 2(f_n^* - \tilde{f}_n)'\varepsilon/n + 2(r_1^{1/2}/n^{1/2})(f_n^* - \tilde{f}_n)'d/n + 2(r_1^{1/4}/n^{1/2})d'\varepsilon/n\)

First, by Chebyshev’s inequality, \(n^{-1}\sum_i(\varepsilon_i^2 - \sigma^2) = O_p(n^{-1/2})\).

Second, similarly to the proof of Lemma A.3,

\[
n^{-1}\sum_i\varepsilon_i(f_n^* - \tilde{f}_n) = O_p\left(m_n^{1/2}(m_n/n + m_n^{-2\alpha})^{1/2}/n^{1/2}\right),
\]

and

\[
n^{-1}\sum_i(f_n^* - \tilde{f}_n)^2 = O_p(m_n/n + m_n^{-2\alpha})
\]

Next, by the law of large numbers,

\[
(r_1^{1/2}/n)d'd/n = (r_1^{1/2}/n)E[d(X_i)^2](1 + o_p(1)) = O_p(r_1^{1/2}/n)
\]

85
In turn, 
\[
\left| (r_{n}^{1/4}/n^{1/2})(f^{*}_n - \tilde{f}_n)'d/n \right| = O_p(r_{n}^{1/4}(m_n/n + m_n^{-2\alpha})^{1/2}/n^{1/2})
\]

Finally, 
\[
(r_{n}^{1/4}/n^{1/2})d'\varepsilon/n = (r_{n}^{1/4}/n^{1/2})O_p(n^{-1/2}) = O_p(r_{n}^{1/4}/n)
\]

Combining the results, 
\[
\sigma^2 - \tilde{\sigma}^2 = O_p(n^{-1/2}) + O_p(m_n^{1/2}(m_n/n + m_n^{-2\alpha})^{1/2}/n^{1/2})
\]
\[
+ O_p(r_{n}^{1/2}/n) + O_p(r_{n}^{1/4}(m_n/n + m_n^{-2\alpha})^{1/2}/n^{1/2}) = O_p(n^{-1/2})
\]

Moreover, by Lemma 1, \(E[||T_i||^2] \leq r_n\), which yields \(||\hat{\Omega} - \tilde{\Omega}|| = O_p(r_n n^{-1/2})\).

The remaining conclusions can be proved as in Lemma A.9. \(\blacksquare\)

C.6 Proof of Theorem 6

Recall that 
\[
\xi = \varepsilon P(\hat{\sigma} P' P)^{-1} P\varepsilon
\]

Define \(\Omega = \sigma^2 E[P_i P_i'] = \sigma^2 I_{k_n}\), where \(P_i = P^{k_n}(X_i)\). The proof of the theorem relies on verifying the conditions of the following lemmas:

**Lemma A.7.** (Donald et al. (2003), Lemma 6.2) If \(E[P_i \varepsilon_i] = 0\), \(E[(\varepsilon_i P_i \Omega^{-1} P_i' \varepsilon_i)^2]/(k_n \sqrt{n}) \to 0\), and \(k_n \to \infty\), then
\[
\frac{n(n^{-1} \varepsilon' P)\Omega^{-1}(n^{-1} P' \varepsilon) - k_n}{\sqrt{2k_n}} \overset{d}{\to} N(0, 1) \tag{C.10}
\]

All three conditions of this lemma hold. \(E[P_i \varepsilon_i] = 0\) and \(k_n \to \infty\) hold trivially, while
\[
E[(\varepsilon_i P_i \Omega^{-1} P_i' \varepsilon_i)^2] \leq CE[\varepsilon_i^4||P_i||^4] \leq CE[||P_i||^4] \leq C\zeta(k_n)^2 k_n
\]

As long as conditions 5.4 holds, \(\zeta(k_n)^2/\sqrt{n} \to 0\), so that the condition of the lemma holds.
Lemma A.8. Suppose that Assumptions 2, 3, 10, and 11 hold. Then
\[
\frac{n(n^{-1} \tilde{\varepsilon} P)(\hat{\sigma}^2 n^{-1} P' P)^{-1}(n^{-1} P' \tilde{\varepsilon}) - n(n^{-1} \varepsilon' P) \Omega^{-1}(n^{-1} P' \varepsilon)}{\sqrt{k_n}} \overset{p}{\to} 0 \quad (C.11)
\]

Proof. Step 1. Show that \( \tilde{\varepsilon} \) can be replaced with \( \varepsilon \).

Because \( \tilde{\varepsilon} = Y - \tilde{g} = \varepsilon + (g - \tilde{g}) \),

\[
n(n^{-1} \varepsilon' P)(\hat{\sigma}^2 n^{-1} P' P)^{-1}(n^{-1} P' \varepsilon) - n(n^{-1} \varepsilon' P)(\hat{\sigma}^2 n^{-1} P' P)^{-1}(n^{-1} P' \varepsilon)
= (g - \tilde{g})' P(P' P)^{-1} P'(g - \tilde{g})/\hat{\sigma}^2 + 2(g - \tilde{g})' P(P' P)^{-1} P' \varepsilon/\hat{\sigma}^2
\]

As for the first term,

\[
(g - \tilde{g})' P(P' P)^{-1} P'(g - \tilde{g})/\hat{\sigma}^2 \leq (g - \tilde{g})' (g - \tilde{g})/\hat{\sigma}^2 = O_p(n\psi_n)
\]

by the projection inequality, Assumption 10(b) and \( \hat{\sigma}^2 \xrightarrow{p} \sigma^2 \).

As for the second term,

\[
\left| (g - \tilde{g})' P(P' P)^{-1} P' \varepsilon/\hat{\sigma}^2 \right| \leq \left| \lambda_{\max} \left( (\hat{\sigma}^2 P' P/n)^{-1} \right) n^{-1} (g - \tilde{g})' PP' \varepsilon \right| \leq \left| C \lambda_{\max}(PP'/n)(g - \tilde{g})' \varepsilon \right|
\]

Because \( P' P/n \) and \( PP'/n \) have the same nonzero eigenvalues and all eigenvalues of \( P' P/n \) converge to one, \( \lambda_{\max}(PP'/n) \) converges in probability to 1. Thus,

\[
\left| (g - \tilde{g})' P(P' P)^{-1} P' \varepsilon/\hat{\sigma}^2 \right| \leq \left| C \lambda_{\max}(PP'/n)(g - \tilde{g})' \varepsilon \right| \leq \left| C(g - \tilde{g})' \varepsilon \right| = o_p(k_n^{1/2})
\]

by Assumption 11.

Then

\[
n(n^{-1} \varepsilon' P)(\hat{\sigma}^2 n^{-1} P' P)^{-1}(n^{-1} P' \varepsilon) - n(n^{-1} \varepsilon' P)(\hat{\sigma}^2 n^{-1} P' P)^{-1}(n^{-1} P' \varepsilon) = O_p(n\psi_n) + o_p(k_n^{1/2})
\]

As long as rate condition 5.3 holds, \( n\psi_n/k_n^{1/2} \to 0 \), this implies

\[
\frac{n(n^{-1} \varepsilon' P)(\hat{\sigma}^2 n^{-1} P' P)^{-1}(n^{-1} P' \varepsilon) - n(n^{-1} \varepsilon' P)(\hat{\sigma}^2 n^{-1} P' P)^{-1}(n^{-1} P' \varepsilon)}{\sqrt{k_n}} \overset{p}{\to} 0
\]

Step 2. Show that \( \tilde{\Omega} = \hat{\sigma}^2 P' P/n \) can be replaced with \( \Omega \).
To prove this, I will use the auxiliary result presented in the following lemma.

**Lemma A.9.** Let \( \tilde{\Omega} = \tilde{\sigma}^2 P'P/n, \Omega = \sigma^2 P'P/n, \Omega = \sigma^2 E[P_i P_i] \). Suppose that Assumptions 2(ii), 3, 10, and 11 are satisfied. Then

\[
||\tilde{\Omega} - \Omega|| = O_p(k_n/n^{1/2}) + O_p(k_n \psi_n) + o_p(k_n^{3/2}/n)
\]

\[
||\tilde{\Omega} - \Omega|| = O_p(\zeta(k_n)k_n^{1/2}/n^{1/2})
\]

If Assumption 2(i) is also satisfied then \( 1/C \leq \lambda_{\text{min}}(\Omega) \leq \lambda_{\text{max}}(\Omega) \leq C \), and if \( \psi_n k_n \to 0 \) and \( \zeta(k_n)k_n^{1/2}/n^{1/2} \to 0 \), then w.p.a. \( 1 \), \( 1/C \leq \lambda_{\text{min}}(\tilde{\Omega}) \leq \lambda_{\text{max}}(\tilde{\Omega}) \leq C \) and \( 1/C \leq \lambda_{\text{min}}(\tilde{\Omega}) \leq \lambda_{\text{max}}(\tilde{\Omega}) \leq C \).

Proof of Lemma A.9 is presented after the remainder of the main proof.

Using the result of Lemma A.9,

\[
\frac{n(n^{-1/2}P\tilde{\Omega}^{-1}P(\cdot)\tilde{\Omega}^{-1}P'\cdot)}{\sqrt{2k_n}} - \frac{n(n^{-1/2}P\Omega^{-1}P(\cdot)\Omega^{-1}P'\cdot)}{\sqrt{2k_n}} \leq \frac{n||\tilde{\Omega}^{-1}P'\cdot||^2(||\tilde{\Omega} - \Omega|| + C||\tilde{\Omega} - \Omega||^2)}{\sqrt{2k_n}}
\]

Similarly to the proof of Theorem 1, \( ||\Omega^{-1}P'\cdot|| = O_p(\sqrt{k_n/n}) \).

Then

\[
\frac{n||\Omega^{-1}n^{-1}P'\cdot||^2(||\Omega - \Omega|| + C||\Omega - \Omega||^2)}{\sqrt{2k_n}} = \frac{nO_p(k_n/n)a_p(1/\sqrt{k_n})}{\sqrt{2k_n}} = o_p(\sqrt{k_n}) = o_p(1),
\]

provided that \( ||\tilde{\Omega} - \Omega|| = o_p(1/\sqrt{k_n}) \), which holds under rate conditions 5.1 and 5.2.

The result of Theorem 6 follows directly from combining the results in Equations C.10 and C.11.
Proof of Lemma A.9. Due to homoskedasticity,

\[ ||\hat{\Omega} - \Omega|| = ||(\hat{\sigma}^2 - \sigma^2) \sum_i P_i P_i' / n|| \leq |\hat{\sigma}^2 - \sigma^2| \sum_i ||P_i||^2 / n \]

\[ = |n^{-1} \sum_i (\varepsilon_i^2 - \sigma^2) + 2n^{-1} \sum_i \varepsilon_i (g_i - \bar{g}_i) + n^{-1} \sum_i (g_i - \bar{g}_i)^2| \sum_i ||P_i||^2 / n \]

First, by Chebyshev’s inequality, \( n^{-1} \sum_i (\varepsilon_i^2 - \sigma^2) = O_p(n^{-1/2}) \).

Second, by Assumption 10, \( n^{-1} \sum_i (g_i - \bar{g}_i)^2 = O_p(\psi_n) \).

Finally, by Assumption 11, \( n^{-1} \sum_i \varepsilon_i (g_i - \bar{g}_i) = o_p(k_n^{1/2} / n) \).

Moreover, by Lemma 1, \( E[||P_i||^2] \leq k_n \), which yields

\[ ||\hat{\Omega} - \Omega|| = O_p(k_n / n^{1/2}) + O_p(\psi_n) + o_p(k_n^{3/2} / n) \]

Next, \( ||\hat{\Omega} - \Omega|| = ||\sigma^2(P'P / n - E[P_i P_i'])|| \).

Thus,

\[ E[||\hat{\Omega} - \Omega||^2] = E[||\sigma^2 \sum_i (P_i P_i' - E[P_i P_i'])/n||^2] \]

\[ = \sigma^4 E[||P_i P_i' - E[P_i P_i']||^2] / n \leq CE[||P_i||^4] / n = C(\zeta(k_n)^2 k_n / n, \]

and by Markov inequality \( ||\hat{\Omega} - \Omega|| = O_p(\zeta(k_n) k_n^{1/2} / n^{1/2}) \).

The remaining conclusions follow from Lemma A.6 in Donald et al. (2003).

\[ \blacksquare \]

Remark A.2 (Discussion of Assumptions 11 and 14.). Assumptions 11 and 14 can be justified as follows. Supposed that \( \bar{g}_i \) are leave-one-out kernel estimates of \( g_i \) (i.e., ones that do not use observation \( i \) to estimate \( \bar{g}_i \)). Then a typical form of \( \bar{g}_i \) is

\[ \bar{g}_i = \sum_{j \neq i} Y_j K_{ij} = \sum_{j \neq i} (g_j + \varepsilon_j K_{ij}) = \sum_{j \neq i} g_j K_{ij} + \sum_{j \neq i} \varepsilon_j K_{ij}, \]

where \( K_{ij} = K \left( \frac{A_i - A_j}{h} \right) \) is a \( d \)-dimensional kernel function, and \( A_i \) is chosen appropriately (e.g. \( A_i = X_i \) in usual nonparametric models or \( A_i = X_i' \hat{\alpha} \) in single index models).

Then

\[ \sum_{i=1}^n \varepsilon_i (\bar{g}_i - g_i) = \frac{1}{n h^d} \sum_{i=1}^n \sum_{j \neq i} \varepsilon_i \varepsilon_j K_{ij} + \frac{1}{n h^d} \sum_{i=1}^n \sum_{j \neq i} \varepsilon_i (g_j - g_i) K_{ij} \]
Under appropriate regularity conditions (possibly including trimming to deal with the small denominator problem), \( \frac{1}{nh_d} \sum_{j \neq i} K_{ij} \equiv \hat{p}_i = p_i(1 + o_p(1)) \), where \( p_i \) is the density at observation \( i \). If \( p_i \) is bounded below from zero, then the denominator will be asymptotically bounded. Thus, it is enough to deal with the numerator.

Let \( I_{n1} = \frac{1}{nh_d} \sum_{i=1}^{n} \sum_{j \neq i} \varepsilon_i \varepsilon_j K_{ij} \) and \( I_{n2} = \frac{1}{nh_d} \sum_{i=1}^{n} \sum_{j \neq i} \varepsilon_i (g_j - g_i) K_{ij} \).

First, consider \( I_{n1} \).

\[
E[I_{n1}^2] = E \left[ \frac{1}{n^2 h^{2d}} \sum_{i=1}^{n} \sum_{j \neq i} \sum_{k=1}^{n} \sum_{l \neq k} \varepsilon_i \varepsilon_j \varepsilon_k \varepsilon_l K_{ij} K_{kl} \right]
\]

This expectation is nonzero only when \( i = k \) and \( j = l \) or when \( i = l \) and \( j = k \), so it reduces to

\[
E[I_{n1}^2] = 2E \left[ \frac{1}{n^2 h^{2d}} \sum_{i=1}^{n} \sum_{j \neq i} \varepsilon_i^2 \varepsilon_j^2 K_{ij}^2 \right] = 2 \frac{1}{h^{2d}} E[\varepsilon_1^2 \varepsilon_2^2 K_{12}^2] = 2 \frac{1}{h^{2d}} \sigma^4 E[K_{12}^2]
\]

Next, for some appropriate \( A_i \) (e.g. \( A_i = X_i \) in usual nonparametric models or \( A_i = X_i' \hat{\alpha} \) in single index models)

\[
\frac{1}{h^{2d}} E[K_{12}^2] = \frac{1}{h^{2d}} E \left[ K \left( \frac{A_2 - A_1}{h} \right)^2 \right] = \int \int K \left( \frac{a_2 - a_1}{h} \right)^2 f(a_1) f(a_2) da_1 da_2
\]

\[
= \frac{1}{h^d} \int \int K(u)^2 f(a_1) f(a_1 + hu) du da_1 = h^d \left( \int \int K(u)^2 f(a_1)^2 (1 + o(1)) du \right) da_1
\]

\[
= \frac{1}{h^d} \left( \int K(u)^2 du \right) E[f(A_i)](1 + o(1)) = O(h^{-d})
\]

Hence, \( E[I_{n1}^2] = O(h^{d_a}) \), and consequently \( I_{n1} = O_P(h^{-d/2}) \).

Next, consider \( I_{n2} \).

\[
E[I_{n2}^2] = E \left[ \frac{1}{n^2 h^{2d}} \sum_{i=1}^{n} \sum_{j \neq i} \sum_{k=1}^{n} \sum_{l \neq k} \varepsilon_i \varepsilon_k (g_j - g_i) (g_l - g_k) K_{ij} K_{kl} \right]
\]
This expectation is nonzero only when \( i = k \), so it reduces to

\[
E[T^2_{n2}] = E \left[ \frac{1}{n^2h^{2d}} \sum_{i=1}^{n} \sum_{j \neq i} \sum_{l \neq i} \varepsilon_i^2 (g_j - g_i)(g_l - g_i)K_{ij}K_{il} \right]
\]
\[
= E \left[ \frac{1}{n^2h^{2d}} \sum_{i=1}^{n} \sum_{j \neq i} \sum_{l \neq i,j} \varepsilon_i^2 (g_j - g_i)(g_l - g_i)K_{ij}K_{il} \right] + E \left[ \frac{1}{n^2h^{2d}} \sum_{i=1}^{n} \sum_{j \neq i} \varepsilon_i^2 (g_j - g_i)^2K_{ij}^2 \right]
\]
\[
= \frac{n}{h^{2d}} E[\varepsilon_1^2 (g_2 - g_1)(g_3 - g_1)K_{12}K_{13}] + \frac{1}{h^{2d}} E[\varepsilon_1^2 (g_2 - g_1)^2K_{12}^2]
\]
\[
= \frac{n}{h^{2d}} \sigma^2 E[(g_2 - g_1)(g_3 - g_1)K_{12}K_{13}] + \frac{1}{h^{2d}} \sigma^2 E[(g_2 - g_1)^2K_{12}^2]
\]

As for the first term,

\[
\frac{1}{h^{2d}} E[(g_2 - g_1)(g_3 - g_1)K_{12}K_{13}] = \frac{1}{h^{2d}} E \left[ (g(A_2) - g(A_1))(g(A_3) - g(A_1)) \right]
\]
\[
= \frac{1}{h^{2d}} \int \int \int (g(a_2) - g(a_1))(g(a_3) - g(a_1))K \left( \frac{a_2 - a_1}{h} \right) \left( \frac{a_3 - a_1}{h} \right) f(a_1)f(a_2)f(a_3)da_1da_2da_3
\]
\[
= \int \int \int (g(a_1 + hu) - g(a_1))(g(a_1 + hv) - g(a_1))K(u)K(v) f(a_1)f(a_1 + hu)f(a_1 + hv)da_1du_1dv
\]
\[
= \int \int \int (g'(a_1)hu + g''(a_1)h^2u^2 + o(h^2)) (g'(a_1)hv + g''(a_1)h^2v^2 + o(h^2))K(u)K(v) f(a_1)(f(a_1) + f'(a_1)hu + o(h))
\]
\[
(f(a_1) + f'(a_1)hv + o(h))da_1du_1dv = O(h^4)
\]

if the second order kernel is used. If a \( \nu \)th higher order kernel \( (\nu > 2) \) is used, then this will become \( O(h^{2\nu}) \).

Hence, \( n\sigma^2 E[(g_2 - g_1)(g_3 - g_1)K_{12}K_{13}] = O(nh^4) \).
As for the second term,

\[
\frac{1}{h^{2d}} E[(g_2 - g_1)^2 K^2_{12}] = \frac{1}{h^{2d}} E \left[ (g(A_2) - g(A_1))^2 K \left( \frac{A_2 - A_1}{h} \right)^2 \right]
\]

\[
= \frac{1}{h^{2d}} \int \int (g(a_2) - g(a_1))^2 K \left( \frac{a_2 - a_1}{h} \right)^2 f(a_1)f(a_2) da_1 da_2
\]

\[
= \frac{1}{h^d} \int \int (g(a_1 + hu) - g(a_1))^2 K(u)^2 f(a_1) f(a_1 + hu) da_1 du
\]

\[
= \frac{1}{h^d} \int \int g'(a_1)^2 h^2 u^2 K(u)^2 f(a_1)^2 (1 + o(1)) da_1 du
\]

\[
= \frac{1}{h^d} \left( \int u^2 K(u)^2 du \right) E[g'(A_i)^2 f(A_i)](1 + o(1)) = O(h^{2-d})
\]

Hence, \( \sigma^2 E[(g_2 - g_1)^2 K^2_{12}] = O(h^{2-d}) \).

Thus, \( E[I_{n2}^2] = O(nh^4 + h^{2-d}) \), and consequently \( I_{n2} = O_p(n^{1/2}h^2) + O_p(h^{1-d/2}) \).

Combining the results above,

\[
\sum_{i=1}^{n} \varepsilon_i (\tilde{g}_i - g_i) = O_p(n^{1/2}h^2) + O_p(h^{-d/2})
\]

In order for Assumptions 11 and 14 to hold, the bandwidth \( h \) and the number of series terms \( k_n \) need to satisfy:

\[
n^{1/2}h^2 / k_n^{1/2} \to 0
\]

\[
1 / (k_n^{1/2}h^{d/2}) \to 0
\]

### C.7 Proof of Theorem 7

Conditions of Lemma A.7 do not rely on the homoskedasticity assumption. Thus, the result of the lemma remains valid even under heteroskedasticity, as long as \( \zeta(k_n)^2 / \sqrt{n} \to 0 \):

\[
\frac{n(n^{-1} \varepsilon' P) \Omega^{-1} (n^{-1} P' \varepsilon) - k_n}{\sqrt{2k_n}} \to N(0, 1),
\]

where now \( \Omega = E[\varepsilon_i^2 P_i P_i'] \).
Lemma A.10. Suppose that Assumptions 2, 3, 10, and 11 hold. Then
\[
\frac{n(n^{-1}\epsilon'P)(n^{-1}P'\Sigma P)^{-1}(n^{-1}P'\epsilon) - n(n^{-1}\epsilon'P)(n^{-1}P'\Sigma P)^{-1}(n^{-1}P'\epsilon)}{\sqrt{k_n}} \xrightarrow{p} 0 \quad (C.13)
\]

Proof.

Lemma A.11. Let \( \tilde{\Omega} = \sum_i \hat{\epsilon}_i^2 P_i P_i' / n \), \( \bar{\Omega} = \sum_i \hat{\epsilon}_i^2 P_i P_i' / n \), \( \bar{\Omega} = \sum_i \sigma_i^2 P_i P_i' / n \), \( \Omega = E[\hat{\epsilon}_i^2 P_i P_i'] \), where \( \sigma_i^2 = E[\hat{\epsilon}_i^2 | X_i] \). Suppose that Assumptions 2(ii), 3, 10, and 11 are satisfied. Then
\[
\begin{align*}
||\tilde{\Omega} - \Omega|| &= O_p(\zeta(k_n)^2 \psi_n) + O_p(\zeta(k_n)^2 k_n^{1/2} / n) \\
||\bar{\Omega} - \Omega|| &= O_p(\zeta(k_n) k_n^{1/2} / n^{1/2}) \\
||\bar{\Omega} - \Omega|| &= O_p(\zeta(k_n) k_n^{1/2} / n^{1/2})
\end{align*}
\]

If Assumption 2(i) is also satisfied then \( 1/C \leq \lambda_{\min}(\Omega) \leq \lambda_{\max}(\Omega) \leq C \), and if \( \zeta(k_n)^2 \psi_n \to 0 \) and \( \zeta(k_n) k_n^{1/2} / n^{1/2} \to 0 \), then w.p.a. \( 1, 1/C \leq \lambda_{\min}(\bar{\Omega}) \leq \lambda_{\max}(\bar{\Omega}) \leq C \) and \( 1/C \leq \lambda_{\min}(\bar{\Omega}) \leq \lambda_{\max}(\bar{\Omega}) \leq C \).

Lemma A.10 can now be proven in two steps.

Step 1. Show that \( \tilde{\Omega} \) can be replaced with \( \bar{\Omega} \).

Because \( \hat{\epsilon} = \epsilon + (g - \bar{g}) \),
\[
\left| n(n^{-1}\epsilon'P)(n^{-1}P'\Sigma P)^{-1}(n^{-1}P'\bar{\epsilon}) - n(n^{-1}\epsilon'P)(n^{-1}P'\Sigma P)^{-1}(n^{-1}P'\epsilon) \right|
\]
\[
= \left| n(n^{-1}(g - \bar{g})'P)(n^{-1}P'\Sigma P)^{-1}(n^{-1}P'(g - \bar{g})) + 2n(n^{-1}(g - \bar{g})'P)(n^{-1}P'\Sigma P)^{-1}(n^{-1}P'\epsilon) \right|
\]
\[
\leq Cn(n^{-1}(g - \bar{g})'P)(n^{-1}P'(g - \bar{g})) + 2Cn(n^{-1}(g - \bar{g})'P)(n^{-1}P'\epsilon),
\]

because the eigenvalues of \( n^{-1}P'\Sigma P \) are bounded below and above by Lemma A.11. As was discussed in the proof of Theorem 6, the eigenvalues of \( PP'/n \) are bounded above. Thus,
\[
\left| Cn(n^{-1}(g - \bar{g})'P)(n^{-1}P'(g - \bar{g})) + 2Cn(n^{-1}(g - \bar{g})'P)(n^{-1}P'\epsilon) \right|
\]
\[
= \left| C(g - \bar{g})'(PP'/n)(g - \bar{g}) \right| + 2\left| C(g - \bar{g})'(PP'/n)\epsilon \right| \leq \left| C(g - \bar{g})'(g - \bar{g}) \right| + 2\left| C(g - \bar{g})'\epsilon \right|
\]
\[
= O_p(n\psi_n) + o_p(k_n^{1/2}) = O_p(n\psi_n) + o_p(k_n^{1/2})
\]

93
As long as \( n\psi_n/k_n^{1/2} \to 0 \),

\[
\frac{n(n^{-1}\varepsilon'P)(n^{-1}P'\hat{\Sigma}P)^{-1}(n^{-1}P'\varepsilon) - n(n^{-1}\varepsilon'P)(n^{-1}P'\Sigma P)^{-1}(n^{-1}P'\varepsilon)}{\sqrt{k_n}} \xrightarrow{p} 0
\]

Step 2. Show that \( \tilde{\Omega} = P'\hat{\Sigma}P/n \) can be replaced with \( \Omega \).

Similarly to the proof of Theorem 6,

\[
\frac{n(n^{-1}\varepsilon'P)\hat{\Omega}^{-1}(n^{-1}P'\varepsilon) - n(n^{-1}\varepsilon'P)\Omega^{-1}(n^{-1}P'\varepsilon)}{\sqrt{2k_n}} = \frac{n(n^{-1}\varepsilon'P)(\hat{\Omega}^{-1} - \Omega^{-1})(n^{-1}P'\varepsilon)}{\sqrt{2k_n}}
\]

Similarly to the proof of Theorem 1, \( ||\Omega^{-1}(n^{-1}P'\varepsilon)|| = O_p(\sqrt{k_n/n}) \).

Then

\[
n||\hat{\Omega}^{-1} - \Omega^{-1}||^2( ||\hat{\Omega} - \Omega|| + C||\hat{\Omega} - \Omega||^2 ) = nO_p(k_n/n)O_p(1/\sqrt{k_n}) = o_p(1/\sqrt{k_n}) = o_p(1),
\]

provided that \( ||\hat{\Omega} - \Omega|| = o_p(1/\sqrt{k_n}) \), which holds under rate conditions 5.6 and 5.7. \( \blacksquare \)

The result of Theorem 7 follows directly from combining the results in Equations C.12 and C.13.

\( \blacksquare \)

**Proof of Lemma A.11.** First,

\[
||\tilde{\Omega} - \hat{\Omega}|| = || \sum P_iP_i'((\varepsilon_i^2 - \varepsilon_i^2)/n) || = || \sum P_iP_i'((\varepsilon_i + g_i - \tilde{g}_i)^2 - \varepsilon_i^2)/n ||
\]

\[
= || \sum P_iP_i'((g_i - \tilde{g}_i)^2 + 2\varepsilon_i(g_i - \tilde{g}_i))/n || \leq \sup_i ||P_i||^2 \sum ((g_i - \tilde{g}_i)^2 + 2\varepsilon_i(g_i - \tilde{g}_i))/n
\]

\[
= \zeta(k_n)^2(O_p(\psi_n) + o_p(k_n^{1/2}/n)) = O_p(\zeta(k_n)^2\psi_n) + o_p(\zeta(k_n)^2k_n^{1/2}/n)
\]

The following two results can be obtained exactly as in Lemma A.6 in Donald et al. (2003):

\[
||\tilde{\Omega} - \Omega|| = || \sum P_iP_i'((\varepsilon_i^2 - \sigma_i^2)/n) || = O_p(\zeta(k_n)\sqrt{k_n/n})
\]

\[
||\hat{\Omega} - \Omega|| = || \sum P_iP_i'\varepsilon_i^2/n - \Omega || = O_p(\zeta(k_n)\sqrt{k_n/n})
\]
Finally, the results about the eigenvalues can also be obtained in the same way as in Lemma A.6 in Donald et al. (2003).

C.8 Proof of Theorem 8

Note that

$$\frac{\sqrt{k_n} n(n^{-1} \varepsilon' P) \hat{\Omega}^{-1}(n^{-1} P' \varepsilon) - k_n}{\sqrt{2k_n}} = \frac{1}{\sqrt{2}} (n^{-1} \varepsilon' P) \hat{\Omega}^{-1}(n^{-1} P' \varepsilon) + T_1,$$

where $T_1 = -\frac{k_n}{n\sqrt{2}} \to 0$.

Hence, it suffices to show that $(n^{-1} \varepsilon' P) \hat{\Omega}^{-1}(n^{-1} P' \varepsilon) \overset{p}{\to} \Delta$.

First, note that $\text{Var}(P_i \varepsilon_i^2) \leq \Omega^*$, because $\Omega^* = \text{E}[\varepsilon_i^2 P_i P_i]$. Then

$$E[(n^{-1} P' \varepsilon^* - E[P_i \varepsilon_i^2]) \Omega^*^{-1}(n^{-1} P' \varepsilon^* - E[P_i \varepsilon_i^2])]$$

$$\leq E[(n^{-1} P' \varepsilon^* - E[P_i \varepsilon_i^2]) \text{Var}(P_i \varepsilon_i^2)^{-1}(n^{-1} P' \varepsilon^* - E[P_i \varepsilon_i^2])]$$

$$= E[\text{tr}(\text{Var}(P_i \varepsilon_i^2)^{-1}(n^{-1} P' \varepsilon^* - E[P_i \varepsilon_i^2])(n^{-1} P' \varepsilon^* - E[P_i \varepsilon_i^2])')] = \text{tr}(I_{k_n})/n = k_n/n \to 0$$

Thus,

$$\bigg| (n^{-1} \varepsilon' P) \Omega'^{-1}(n^{-1} P' \varepsilon^*) - E[\varepsilon_i^2 P_i] \Omega'^{-1} E[P_i \varepsilon_i^2] \bigg|$$

$$\leq \bigg| (n^{-1} P' \varepsilon^* - E[P_i \varepsilon_i^2]) \Omega'^{-1}(n^{-1} P' \varepsilon^* - E[P_i \varepsilon_i^2]) \bigg| + 2 \bigg| E[\varepsilon_i^2 P_i] \Omega'^{-1}(n^{-1} P' \varepsilon^* - E[P_i \varepsilon_i^2]) \bigg|$$

$$\leq o_p(1) + 2\sqrt{E[\varepsilon_i^2 P_i] \Omega'^{-1} E[P_i \varepsilon_i^2]} \sqrt{(n^{-1} P' \varepsilon^* - E[P_i \varepsilon_i^2]) \Omega'^{-1}(n^{-1} P' \varepsilon^* - E[P_i \varepsilon_i^2])}$$

$$= o_p(1) + 2\sqrt{\Delta} o_p(1) = o_p(1)$$

Hence, $(n^{-1} \varepsilon' P) \Omega'^{-1}(n^{-1} P' \varepsilon^*) = \Delta + o_p(1)$.

Next,

$$\bigg| (n^{-1} \varepsilon' P) \Omega'^{-1}(n^{-1} P' \varepsilon^*) - (n^{-1} \varepsilon' P) \Omega'^{-1}(n^{-1} P' \varepsilon^*) \bigg|$$

$$\leq \bigg| (n^{-1} P' \varepsilon^* - n^{-1} P' \varepsilon^*) \Omega'^{-1}(n^{-1} P' \varepsilon^* - n^{-1} P' \varepsilon^*) \bigg| + 2 \bigg| (n^{-1} \varepsilon' P) \Omega'^{-1}(n^{-1} P' \varepsilon^* - n^{-1} P' \varepsilon^*) \bigg|$$

95
By the assumption of the theorem, \( \sup_{x \in X} |f(x, \hat{\theta}, \hat{h}) - f(x, \theta^*, h^*)| \to 0 \), so that

\[
n^{-1} P' \varepsilon^* - n^{-1} P' \varepsilon = n^{-1} \sum_i P_i (f(X_i, \hat{\theta}, \hat{h}) - f(X_i, \theta^*, h^*)) = o_p(1)
\]

Thus,

\[
\left| (n^{-1} \varepsilon' P) \Omega^{-1} (n^{-1} P' \varepsilon^*) - (n^{-1} \varepsilon' P) \Omega^{-1} (n^{-1} P' \varepsilon^*) \right| = o_p(1),
\]

and \((n^{-1} \varepsilon' P) \Omega^{-1} (n^{-1} P' \varepsilon^*) = \Delta + o_p(1)\).

Finally,

\[
\left| (n^{-1} \varepsilon' P) (\hat{\Omega}^{-1} - \Omega^{-1}) (n^{-1} P' \varepsilon) \right| \leq \left| (n^{-1} \varepsilon' P) (\Omega^{-1}(\hat{\Omega} - \Omega) \hat{\Omega}^{-1} \Omega^{-1})(n^{-1} P' \varepsilon) \right| \\
+ \left| (n^{-1} \varepsilon' P) (\Omega^{-1} \hat{\Omega}^{-1})(n^{-1} P' \varepsilon) \right| \leq ||\Omega^{-1} n^{-1} P' \varepsilon||^2 (||\hat{\Omega} - \Omega|| + C||\hat{\Omega} - \Omega||^2) \to 0,
\]

It follows from the results above and from the triangle inequality that \((n^{-1} \varepsilon' P) \hat{\Omega}^{-1} (n^{-1} P' \varepsilon) \overset{p}{\to} \Delta.\)

C.9 Proof of Theorem 9

Proof of this theorem follows along the same lines as proof of Theorem 6.

Recall that

\[
\xi = \varepsilon' P (\tilde{\sigma}^2 P' P)^{-1} P' \varepsilon
\]

Define \( \Omega = \sigma^2 E[P_i P'_i] = \sigma^2 I_{k_n} \), where \( P_i = P_{k_n}(X_i) \). First, the conditions of Lemma A.7 hold, so that

\[
\frac{n(n^{-1} \varepsilon' P) \Omega^{-1} (n^{-1} P' \varepsilon) - k_n}{\sqrt{2k_n}} \overset{d}{\to} N(0, 1) \quad \text{(C.14)}
\]

Next, I prove the following result.

Lemma A.12. Suppose that Assumptions 2, 3, 12, 13, 14 hold. Then

\[
\frac{n(n^{-1} \varepsilon' P) (\tilde{\sigma}^2 n^{-1} P' P)^{-1} (n^{-1} P' \varepsilon) - n(n^{-1} \varepsilon' P) \Omega^{-1} (n^{-1} P' \varepsilon)}{\sqrt{k_n}} \overset{p}{\to} E[d_i^2]/\sigma^2 \quad \text{(C.15)}
\]

Proof. Step 1. Show that \( \bar{\varepsilon} \) can be replaced with \( \varepsilon \).
Under the local alternative,

\[ \hat{\varepsilon} = \varepsilon + (g_n - \hat{f}) = \varepsilon + (f_n^* - \hat{f}_n) + (k_n^{1/4}/n^{1/2})d \]

Thus,

\[
n(n^{-1}\varepsilon'P)(\hat{\sigma}^2n^{-1}P'P)^{-1}(n^{-1}P'\hat{\varepsilon}) - n(n^{-1}\varepsilon'P)(\hat{\sigma}^2n^{-1}P'P)^{-1}(n^{-1}P'\varepsilon)
\]

\[
= (f_n^* - \hat{f}_n)'P(P'P)^{-1}P'(f_n^* - \hat{f}_n)/\hat{\sigma}^2 + (k_n^{1/2}/n)d'P(P'P)^{-1}P'd/\hat{\sigma}^2
\]

\[
+ 2(f_n^* - \hat{f}_n)'P(P'P)^{-1}P'\varepsilon/\hat{\sigma}^2 + 2(k_n^{1/4}/n^{1/2})d'P(P'P)^{-1}P'\varepsilon/\hat{\sigma}^2
\]

\[
+ 2(k_n^{1/4}/n^{1/2})(f_n^* - \hat{f}_n)'P(P'P)^{-1}P'd/\hat{\sigma}^2
\]

As for the first term,

\[
(f_n^* - \hat{f}_n)'P(P'P)^{-1}P'(f_n^* - \hat{f}_n)/\hat{\sigma}^2 \leq (f_n^* - \hat{f}_n)'(f_n^* - \hat{f}_n)/\hat{\sigma}^2 = O_p(n\psi_n)
\]

by the projection inequality, Assumption 10(b) and \( \hat{\sigma}^2 \overset{p}{\to} \sigma^2 \).

As for the second term, note that

\[
(k_n^{1/2}/n)d'P(P'P)^{-1}P'd/\hat{\sigma}^2 = (k_n^{1/2}/n)(d'P(P'P)^{-1}P')(P(P'P)^{-1}P'd)/\hat{\sigma}^2 = (k_n^{1/2}/n)d'\hat{d}/\hat{\sigma}^2,
\]

where \( \hat{d} = P(P'P)^{-1}P'd \) are the fitted values from the nonparametric series regression of \( d \) on \( P \). Next,

\[
\left| \hat{d}'\hat{d} - d'd \right| = \left| (\hat{d} - d)'(\hat{d} - d) + 2d'\hat{d}(\hat{d} - d) \right| \leq \left| (\hat{d} - d)'(\hat{d} - d) \right| + 2\left| d'(\hat{d} - d) \right|
\]

By Lemma 5, \( (\hat{d} - d)'(\hat{d} - d) = O_p(n(k_n/n + k_n^{-2\alpha})) \).

By the Cauchy-Schwartz inequality and Lemma 5,

\[
\left| d'(\hat{d} - d) \right| \leq n\sqrt{\left( \sum_i d_i^2/n \right) \left( \sum_i (\hat{d}_i - d_i)^2/n \right) }
\]

\[
= n\sqrt{E[d_i^2](1 + o_p(1))O_p(n(k_n/n + k_n^{-2\alpha}))} = O_p(n(k_n/n + k_n^{-2\alpha})^{1/2})
\]
Thus,

\[ (k_n^{1/2}/n) d' P (P'P)^{-1} P' d / \hat{\sigma}^2 = (k_n^{1/2}/n) d' d / \hat{\sigma}^2 + O_p \left( k_n^{1/2}(k_n/n + k_n^{-2\alpha})^{1/2} \right) \]

\[ = k_n^{1/2} E[d_i^2] (1 + o_p(1)) / \sigma^2 + O_p \left( k_n^{1/2}(k_n/n + k_n^{-2\alpha})^{1/2} \right) \]

where the last equality is due to the law of large numbers and \( \hat{\sigma}^2 = \sigma^2 + o_p(1) \).

As for the third term,

\[
\left| (f_n^* - \tilde{f}_n)' P (P'P)^{-1} P' \varepsilon / \hat{\sigma}^2 \right| \leq \lambda_{\max} \left( (\hat{\sigma}^2 P^2 P/n)^{-1} \right) n^{-1} (f_n^* - \tilde{f}_n)' P P' \varepsilon \\
\leq \left| C \lambda_{\max} (PP'/n) (f_n^* - \tilde{f}_n)' \varepsilon \right|
\]

Because \( P'P/n \) and \( PP'/n \) have the same nonzero eigenvalues and all eigenvalues of \( P'P/n \) converge to one, \( \lambda_{\max}(PP'/n) \) converges in probability to 1. Thus,

\[
\left| (f_n^* - \tilde{f}_n)' P (P'P)^{-1} P' \varepsilon / \hat{\sigma}^2 \right| \leq \left| C \lambda_{\max}(PP'/n) (f_n^* - \tilde{f}_n)' \varepsilon \right| \leq \left| C (f_n^* - \tilde{f}_n)' \varepsilon \right|
\]

By Assumption 14, \( \sum \varepsilon_{i}(f_n^* - \tilde{f}_n) = o_p(k_n^{1/2}) \).

As for the fourth term,

\[
(k_n^{1/4}/n^{1/2}) d' P (P'P)^{-1} P' d / \hat{\sigma}^2 \leq (k_n^{1/4}/n^{1/2}) \lambda_{\max} \left( (\hat{\sigma}^2 P^2 P/n)^{-1} \right) n^{-1} d' P P' \varepsilon \\
\leq (k_n^{1/4}/n^{1/2}) C \lambda_{\max}(PP'/n) d' \varepsilon \leq (k_n^{1/4}/n^{1/2}) C d' \varepsilon \\
= (k_n^{1/4}/n^{1/2}) O_p(n^{1/2}) = O_p(k_n^{1/4})
\]

As for the last term,

\[
(k_n^{1/4}/n^{1/2})(f_n^* - \tilde{f}_n)' P (P'P)^{-1} P' d / \hat{\sigma}^2 \leq (k_n^{1/4}/n^{1/2}) \lambda_{\max} \left( (\hat{\sigma}^2 P^2 P/n)^{-1} \right) n^{-1} (f_n^* - \tilde{f}_n)' P P' d \\
\leq (k_n^{1/4}/n^{1/2}) C \lambda_{\max}(PP'/n) (f_n^* - \tilde{f}_n)' d \leq (k_n^{1/4}/n^{1/2}) C (f_n^* - \tilde{f}_n)' d
\]
By the Cauchy-Schwartz inequality,

\[
\left| \left( \frac{1}{n^{1/2}} \right) (f_n^* - \tilde{f}_n) \right| \leq \left( \frac{1}{n^{1/2}} \right) \sqrt{\sum_i (f_n^* - \tilde{f}_n)^2 / n} \sqrt{\sum_i d_i^2 / n}
\]

\[
= \left( \frac{1}{n^{1/2}} \right) \sqrt{O_p(\psi_n) E[d_i^2]} (1 + o_p(1)) = O_p(k_n^{1/4} \psi_n^{1/2} n^{1/2})
\]

Then

\[
n(n^{-1} \varepsilon' P)(\sigma^2 n^{-1} P' P)^{-1} (n^{-1} P' \varepsilon) - n(n^{-1} \varepsilon' P)(\sigma^2 n^{-1} P' P)^{-1} (n^{-1} P' \varepsilon) \\
= k_n^{1/2} E[d_i^2] / \sigma^2 + O_p(k_n^{1/2} (k_n/n + k_n^{-2})^{1/2}) + O_p(k_n^{1/4}) + O_p(k_n^{1/4} \psi_n^{1/2} n^{1/2}) + o_p(k_n^{1/2})
\]

As long as rate conditions 5.3 and 5.4 hold, this implies that

\[
n(n^{-1} \varepsilon' P)(\sigma^2 n^{-1} P' P)^{-1} (n^{-1} P' \varepsilon) - n(n^{-1} \varepsilon' P)(\sigma^2 n^{-1} P' P)^{-1} (n^{-1} P' \varepsilon) = \frac{P}{\sqrt{k_n}} E[d_i^2] / \sigma^2
\]

Step 2. Show that \( \tilde{\Omega} = \sigma^2 P' P/n \) can be replaced with \( \Omega \).

To prove this, I will use the auxiliary result presented in the following lemma.

Lemma A.13. Let \( \tilde{\Omega} = \sigma^2 P' P/n, \tilde{\Omega} = \sigma^2 P' P/n, \Omega = \sigma^2 P' P/n \). Suppose that Assumptions 2(ii), 3, 13, and 14 are satisfied. Then

\[
||\tilde{\Omega} - \Omega|| = O_p(k_n n^{-1/2}) + O_p(k_n^{3/2} / n) + O_p(k_n^{5/4} \psi_n^{1/2} / n^{1/2}) + o_p(k_n^{3/2} / n)
\]

\[
||\Omega - \Omega|| = O_p(\zeta(k_n) k_n^{1/2} / n^{1/2})
\]

If Assumption 2(i) is also satisfied then \( 1/C \leq \lambda_{\min}(\Omega) \leq \lambda_{\max}(\Omega) \leq C \), and if \( k_n^{5/4} \psi_n^{1/2} / n^{1/2} \to 0 \) and \( \zeta(k_n) k_n^{1/2} / n^{1/2} \to 0 \), then w.p.a. 1, \( 1/C \leq \lambda_{\min}(\tilde{\Omega}) \leq \lambda_{\max}(\tilde{\Omega}) \leq C \) and \( 1/C \leq \lambda_{\min}(\Omega) \leq \lambda_{\max}(\Omega) \leq C \).

Proof of Lemma A.13 is presented after the remainder of the main proof.

Using the result of Lemma A.13, similarly to the proof of Theorem 6,

\[
\left| \frac{n(n^{-1} \varepsilon' P)\tilde{\Omega}^{-1} (n^{-1} P' \varepsilon) - n(n^{-1} \varepsilon' P)\Omega^{-1} (n^{-1} P' \varepsilon)}{\sqrt{2k_n}} \right| \leq \frac{n||\Omega^{-1} n^{-1} P' \varepsilon||^2 (||\tilde{\Omega} - \Omega|| + C||\tilde{\Omega} - \Omega||^2)}{\sqrt{2k_n}}
\]

\[
= nO_p(k_n n) o_p(1 / \sqrt{k_n}) = \frac{o_p(\sqrt{k_n})}{\sqrt{2k_n}} = o_p(1)
\]
provided that $||\tilde{\Omega} - \Omega|| = o_p(1/\sqrt{k_n})$, which holds under rate conditions 5.1 and 5.2. ■

The result of Theorem 9 follows directly from combining the results in Equations C.14 and C.15. ■

Proof of Lemma A.13. Under the local alternative,

$$\tilde{\varepsilon} = \varepsilon + (g - \tilde{f}_n) = \varepsilon + (f_n^* - \tilde{f}_n) + (k_n^{1/4}/n^{1/2})d$$

Thus,

$$\tilde{\sigma}^2 = \tilde{\varepsilon}'\tilde{\varepsilon}/n = \varepsilon'\varepsilon/n + (f_n^* - \tilde{f}_n)'(f_n^* - \tilde{f}_n)/n + (k_n^{1/2}/n)d'd/n$$

$$+ 2(f_n^* - \tilde{f}_n)'\varepsilon/n + 2(k_n^{1/4}/n^{1/2})(f_n^* - \tilde{f}_n)'d/n + 2(k_n^{1/4}/n^{1/2})d'\varepsilon/n$$

First, by Chebyshev’s inequality, $n^{-1}\sum_i(\varepsilon_i^2 - \sigma^2) = O_p(n^{-1/2})$

Second, by Assumptions 13 and 14, $n^{-1}\sum_i\varepsilon_i(f_n^* - \tilde{f}_n) = o_p(k_n^{1/2}/n)$ and $n^{-1}\sum_i(f_n^* - \tilde{f}_n)^2 = O_p(\psi_n)$.

Next, by the law of large numbers, $(k_n^{1/2}/n)d'd/n = (k_n^{1/2}/n)E[d(X_i)^2](1 + o_p(1)) = O_p(k_n^{1/2}/n)$.

In turn, $$\left|(k_n^{1/4}/n^{1/2})(f_n^* - \tilde{f}_n)'d/n\right| = O_p(k_n^{1/4}\psi_n^{1/2}/n^{1/2}).$$

Finally,

$$(k_n^{1/4}/n^{1/2})d'\varepsilon/n = (k_n^{1/4}/n^{1/2})O_p(n^{-1/2}) = O_p(k_n^{1/4}/n)$$

Combining the results,

$$\tilde{\sigma}^2 - \sigma^2 = O_p(n^{-1/2}) + O_p(k_n^{1/2}/n) + O_p(k_n^{1/4}\psi_n^{1/2}/n^{1/2}) + o_p(k_n^{1/2}/n)$$

By Lemma 1, $E[||P_i||^2] \leq k_n$, which yields

$$||\tilde{\Omega} - \bar{\Omega}|| = O_p(k_n^{-1/2}) + O_p(k_n^{3/2}/n) + O_p(k_n^{5/4}\psi_n^{1/2}/n^{1/2}) + o_p(k_n^{3/2}/n)$$

The remaining conclusions can be proved as in Lemma A.9. ■
References


tion and consistent tests with conditional moment restrictions,” Journal of Econometrics,
117, 55 – 93.

of Econometrics, 20, 83 – 104.

Handbook of Econometrics, Volume 2, 775 – 826.

Semiparametric Functional Forms,” Econometrica, 64, 865–890.


on expenditure patterns in rural China: a semiparametric analysis,” Journal of Applied
Econometrics, 20, 509–527.


HALL, A. (1990): “Lagrange Multiplier Tests for Normality against Seminonparametric Al-

HAUSMAN, J. A. AND W. K. NEWEY (1995): “Nonparametric Estimation of Exact Con-
sumers Surplus and Deadweight Loss,” Econometrica, 63, 1445–1476.

Series Regression,” Econometrica, 63, 1133–1159.


