A Specification Test for Semiparametric Models with Generated Regressors

Elia Lapenta*

February 15, 2019

Abstract: In this paper, we provide a test for a conditional moment restriction which is specified in a semiparametric way, and includes some variables that are not observed but can be estimated by using nonparametric techniques. The methodology we present can be employed in order to check the correct specification of several semiparametric models used in Economics, ranging from discrete-choice models with endogenous regressors, to semiparametric regressions with endogeneity and control functions, semiparametric sample-selection models, and incomplete-information games where agents are assumed to play Bayesian-Nash equilibria. Checking the correct specification of these frameworks is relevant from an empirical point of view, as if the model is not specified in a correct way then also the estimates of the effects of a variable and the counterfactual analysis will be wrong. We construct a Wild-bootstrap test: since our statistic converges to a intricate distribution with unknown quantiles, in order to obtain the critical values we propose a wild-bootstrap procedure which is very easy to implement, and prove its the validity under low-level assumptions. Accordingly, our test can be readily applied by the researcher.

*Toulouse School of Economics. I am very grateful to Prof. Pascal Lavergne for many useful conversations and advices on the work behind this paper. This research has received financial support from the European Research Council under the European Community's Seventh Framework Program FP7/2007-2013 grant agreement N°295298.

1 Introduction

In this paper we provide a test for a conditional moment restriction which is specified in a semiparametric way, and includes some variables that are not observed but can be estimated by using nonparametric techniques. In particular, the moment equation we consider in this work writes as

$$\mathbb{E}\{Y|\tilde{X}\} = G_0(\beta_0^T \cdot X, V) \ ,$$

where the variables Y and $\tilde{X}$ are both observed, while the function G_0 and the finite dimensional parameter β_0 are unknown. Both V and some of the components of X are not observed by the researcher, but they are identified and can be estimated in a nonparametric way. If all the variables $(Y, \tilde{X}, X, V)$ were observed, the above moment equation would simply provide a multiple-index restriction that could be tested by using one of the many specification tests available in the statistical literature. On the other hand, to the best of our knoweledge, there is no statistical procedure available in the literature allowing to test the above moment restriction when some of the variables in X and the variable V are unobserved to the researcher, but can be estimated in a nonparametric way. In this paper we fill up this gap, and propose such a specification test.

The methodology we present can be employed in order to check the correct specification of several semiparametric models used in Economics, ranging from discrete-choice models with endogenous regressors, to semiparametric regressions with endogeneity and control functions, to incomplete-information games where agents are assumed to play Bayesian-Nash equilibria. Blundell & Powell (2004), Rothe (2009), and Escanciano et al. estimate a binary-choice model where the discrete choice of an agent depends on both a bunch of covariates and an unobserved error term. They allow such an unobserved error term to be correlated with the explanatory variables and, in order to handle such endogeneity and identify the components of their framework, they introduce a control function which is not observed but can be estimated nonparametrically. These setups have been employed in order to investigate the ralationship between the supply of working hours and non-labour incomes, between households' decisions about buying a house and their permanent income, or to study agents' migration decisions, among others. Since these models give rise to a moment condition of the type written above, our test can be employed in order to check their correct specification. Newey *et al.* (1999) propose an estimation method for nonparametric and semiparametric regression models where the dependent variable is allowed to be continuous, and some of the regressors are allowed to be endogenous. Such endogeneity is handled by using a control function that is not observed but has to be estimated in a nonparametric way. The authors then use this framework in order to study the relationship between the supply of hours worked and the hourly wage rate. Since their setup also gives rise to a moment condition that is a particular case of the one analyzed in this paper, one could use our methodology to check the correct specification of their model. The test presented in this work can also be employed to check the moment condition implied by semiparametric models with sample selection, where the selection mechanism is not restricted to have a specific parametric form, but is specified in a general nonparametric way (Escanciano et al., 2015). Other applications of our test stand in the specification of incomplete-information games, where agents have private information and play according to a Bayesian-Nash behavior (Lewbel & Tang (2015), Aradillas-Lopez (2012)). In these frameworks, the profit function of each agent is assumed to have a linear specification, while the distribution of agents' private information is not restricted to have a specific functional form. As a consequence, these setups give rise to a moment condition of the type we analyze in this work. The empirical application of these types of models is relatively wide: they can be used in order to analyze firms' decisions about entering or not entering an oligopolistic market -similarly to Ciliberto & Tamer (2009)-, or to study the decisions of opening grocery stores -similarly to Grieco (2014)-, or to investigate firms' capital investment strategies -like in Aradillas-Lopez (2010).

Now, checking the correct specification of the frameworks just described is relevant from an empirical point of view. First, if the model is not specified in a correct way, then also the estimates of the effects of a variable will be mispecified, and hence any estimation method might lead to the wrong conclusions. Second, the correct specification of a model is particularly relevant in any couterfactual analysis: if the setup is mispecified and we run a counterfactual based on it, then also our conclusions will be wrong.

The main contribution of this paper stands in the construction of a Wild-bootstrap test. In particular, the test statistic we propose converges to a complicated distribution with unknown quantiles, so that in order to obtain the critical values we set up a wild-bootstrap procedure, and prove its validity under low-level assumptions. Accordingly, our test can be readily applied by the researcher, as essentially he only has to check the convergence rate of a bandwidth sequence. Through a Monte-Carlo simulation study we show that our test has a good performance in small samples.

Beyond those cited above, our work is related to several papers which build up estimation methods for models where some of the regressors are not observed by the researcher but are estimated in a nonparametric way. Ahn (1997), Ahn & Manski (1993), and Li & Wooldridge (2002) propose estimators of single-index models and partially linear models where some of the regressors are estimated in a preliminary step. Mammen *et al.* (2012) and Mammen *et al.* (2016) provide a general framework for estimation of non and semiparametric models where some of the regressors are also nonparametrically estimated in a preliminary step. Finally, Escanciano *et al.* (2014) obtain an expansion (i.e. approximations) of the residuals coming from a regression involving some variables which are estimated in a preliminary step. However, all these papers are mainly concerned with estimation, and do not treat the problem of specification testing in the presence of unobserved regressors which need to be estimated in a nonparametric way.

From a methodological point of view, the paper that is the most related to ours is the one by Escanciano *et al.* (2014) just described. The authors start from a very general framework and derive an approximation of the residuals under some high-level assumptions. Since also our test statistic is based on a sum of the residuals, our setup is linked to the one of these authors. However, there are three main differences between our paper and theirs. First, differently from them, we can avoid the use of high-level conditions as our setup is much less general with respect to theirs, and hence can derive our proofs under low-level assumptions only. Second, in our paper we consider a bootstrap envirnoment, while Escanciano *et al.* (2014) do not tackle the issue of proving the validity of a bootstrap test. Third, from the point of view of the statistical tools, our proofs are mainly based on U-process and U-statistic theory.

The reminder of the paper is organized as follows. Section (3) presents in detail the models our test can be applied to. The following Section (4) sets up the test, describes the statistic we propose, and provides the assumptions at the basis of our proofs. The main results of the paper are reported in Section (5) which hence describes the asymptotic behavior of the statistic, constructs the wild-bootstrap scheme we use to estimate the critical values, and finally proves the validity of the bootstrap test. In Section (6) we discuss some technical issues concerning our test and some of the main ideas behind our proofs. Such a section can be skipped without loss of continuity. Finally, Section (7) concludes.

2 The Test

Recall that the null hypothesis at the center of this paper can be written as

$$H_0 : \mathbb{E}\{Y|\tilde{X}\} = \mathbb{E}\{Y|\beta_{0,1}^T X_1 + \beta_{0,2}^T m_2(X_2)\ ,\ X^e - m_0(Z)\}$$

with $\beta_0 := (\beta_{0,1}^T, \beta_{0,2}^T)^T$, $X = X_1 \cup X_2$, $\tilde{X} = X \cup Z$, $X^e \in X$, $m_0(Z) := \mathbb{E}\{D_0|Z\}$, and $m_2(X_2) := \mathbb{E}\{D_2|X_2\}$. The dimension of each random variable is defined as follows: $p_1 := dim(x_1)$, $p_2 := dim(x_2)$, $p_0 := dim(z)$, $d := 1 + dim(x^e)$. The variables $(Y, \tilde{X}, D_0, D_2)$ are observed, while the functions m_0, m_2 and the finite dimansional parameter β_0 are unknown. We here consider an omnibus test, so the alternative hypothesis H_1 is defined as the logical complement of H_0.

In what follows, we will construct a test for the above null hypothesis and outline how to obtain the bootstrap critical values, by keeping the technical content at the least level possible and adopting a practical standpoint. A complete description of the technical details, the estimators used, and the assumptions is portponed to Section 4.

In order to enlight the notational burdern, we borrow some pieces of notation from Rothe (2009), and define $v(\beta) := (\beta_1^T x_1 + \beta_2^T m_2(x_2), x^e - m_0(z))$ and $V_i(\beta) := (\beta_1^T X_{i,1} + \beta_2^T m_2(X_{i,2}), X_i^e - m_0(Z_i))$ for a vector $\beta \in \Re^{dim(X)}$. Also, define $f_{V(\beta)}$ to be the density function of the random variable $V(\beta)$, and $G_{V(\beta)}(w) := \mathbb{E}\{Y|V(\beta) = w\}$. Notice that the above null hypothesis H_0 is equivalent to

$$H_0 : \mathbb{E}\{[Y - G_{V(\beta_0)}(V(\beta_0))] \cdot f_{V(\beta_0)}(V(\beta_0)) \mid \tilde{X}\} = 0 \ P\text{-almost surely}$$

The first step to build up a test is to transform the above conditional moment into a continuum of unconditional moments. To this end, let $\phi : \Re \mapsto \mathbb{C}$ be an analytical non-polynomial function of a single variable, i.e. ϕ is a univariate function infinitely times continuously differentiable that does not have a polynomial form. $exp(\cdot)$, $exp(\cdot \mathrm{i})$, $cos(\cdot)$, $sin(\cdot)$, $cos(\cdot)$ are examples of such a function. Also, let $\mathcal{T}$ be a compact subset of $\Re^{dim(\tilde{X})}$ encompassing the origin. For the ease of notation, we introduce the collection of functions

$$g_{V(\beta),t}(y,\tilde{x}) := [y - G_{V(\beta)}(v(\beta))] \cdot f_{V(\beta)}(v(\beta)) \cdot \phi(\tilde{x}t) \text{ with } t \in \mathcal{T}$$

Also, in line with the usual notation of Empirical Process Theory, we define the linear operators P and $\mathbb{P}_n$ to be such that $Pg = \int g(y,\tilde{x})\ P(\mathrm{d}y,\tilde{x})$, and $\mathbb{P}_n g = (1/n)\sum_{i=1}^{n} g(Y_i,\tilde{X}_i)$. Notice that if $(y,\tilde{x}) \mapsto g(y,\tilde{x})$ is a nonrandom and deterministic function, $Pg = \mathbb{E}g(Y,\tilde{X})$. By the results in Bierens (1982) and Stinchcombe & White (1998), we can transform the above conditional moment into a continuum of unconditional moments so that

$$H_0 : \ Pg_{V(\beta_0),t} = 0 \ \forall \ t \in \mathcal{T}$$

$$H_1 : \ Pg_{V(\beta_0),t} \neq 0 \text{ form almost all } t \in \mathcal{T}$$

At this stage, in order to build up an Oracle test we could simply take the empirical counterpart of the above moment condition, and compare it to the quantiles of an appropirately rescaled normal distribution. Specifically, by the above display the null hypothesis implies that for any fixed $t \in \mathcal{T}$ the sum $\sqrt{n}\mathbb{P}_n g_{V(\beta_0),t}$ will converge in distribution to a Gaussian with mean zero and variance equal to the variance of $g_{V(\beta_0),t}(Y,\tilde{X})$, while under the alternative H_1 such a sum will explode and converge to $+/-\infty$. Hence, by comparing such a statistic to the quantiles of an appropriately rescaled normal distribution we would obtain a consistent test, in the sense that under the null hypothesis the probability of rejecting H_0 would converge to the nominal size of the test, while under the alternative hypothesis such a probability will converge to one. This procedure, however, will suffer from two drawbacks. First, the value t picked up from $\mathcal{T}$ is completely arbitrary, so different researchers picking up different t's might reach different conclusions by running the same test. Second, such a test is an Oracle one, in the sense that it is based on the true function $g_{V(\beta_0),t}$ which is however unknown to the researcher. In order to overcome the first issue, we can simply take the square of $\sqrt{n}\mathbb{P}g_{V(\beta_0),t}$ and integrate it with respect to t. For tackling the second issue, we can instead replace the true function $g_{V(\beta_0),t}$ with its empirical counterpart, say $\hat{g}_{\hat{V}(\hat{\beta}),t}$, where $\hat{V}(\beta) := (\beta_1 X_1 + \beta_2 \hat{m}_2(X_{i,2}), X_i^e - \hat{m}_0(Z_i))$, and $\hat{m}_2$,$\hat{m}_0$ and $\hat{\beta}$ are estimators for m_2, m_0, and β_0, respectively. Hence, a feasible statistic for our test will be

$$S_n = \int |\sqrt{n}\mathbb{P}_n \hat{g}_{\hat{V}(\hat{\beta}),t}|^2 \mu(\mathrm{d}t)$$

where μ is a measure that is continuous with respect to the Lebesgue measure.

In order to define the estimator $\hat{g}_{\hat{V}(\hat{\beta}),t}$, we need to estimate both the regression function $G_{V(\beta_0)}(\cdot) = \mathbb{E}\{Y|V(\beta_0) = \cdot\}$ and the density $f_{V(\beta_0)}$ of the random variable $V(\beta_0)$, but since we do not observe the regressors $V(\beta_0)$ we will proceed to a "two-stage" estimation: in a first stage, we get the estimates $(\hat{m}_0, \hat{m}_1)$ and $\hat{\beta}$ so as to get an estimate $\hat{V}(\hat{\beta})$ of the regressors $V(\beta_0)$, while in a second stage we construct the estimator for $G_{V(\beta_0)}$ and $f_{V(\beta_0)}$ by replacing the unobserved regressor $V(\beta_0) := (\beta_{0,1}X_1 + \beta_{0,2}m_2(X_2), X^e - m_0(Z))$ with its estimate $\hat{V}(\hat{\beta}) := (\hat{\beta}_1 X_1 + \hat{\beta}_2 \hat{m}_2(X_2), X^e - \hat{m}_0(Z))$. In particular, we can implement the following steps:

Step 1) By definition of m_0 and m_2, regress nonparametrically D_0 onto Z and D_2 onto X_2, so as to get the estimates $\hat{m}_0$ and $\hat{m}_2$, respectively.

Step 2) Given $(\hat{m}_0, \hat{m}_2)$, obtain the estimate $\hat{V}(\beta)$ of the variable $V(\beta)$, and then regress nonparametrically Y onto $\hat{V}(\beta)$ so as to get the estimate $\hat{G}_{\hat{V}(\beta)}$ of $G_{V\beta)}$.

Step 3) In order to obtain an estimate for the finite dimensional parameter $\hat{\beta}$, minimize a (Semiparametric) Least-Square criterion function, so that

$$\hat{\beta} := \arg\min_{\beta \in B} \sum_{i=1}^{n} [Y_i - \hat{G}_{\hat{V}(\beta)}(\hat{V}_i(\beta))]$$

Step 4) Finally, given the estimate $\hat{V}(\hat{\beta})$ of the regressor $V(\beta_0)$, obtain the estimate $\hat{f}_{\hat{V}(\hat{\beta})}$ of the density $f_{V(\beta_0)}$, and then regress Y onto $\hat{V}(\hat{\beta})$ so as to get the estimate $\hat{G}_{\hat{V}(\hat{\beta})}$ of the regression $G_{V(\beta_0)}$.

Step 5) Given the estimates $\hat{G}_{\hat{V}(\hat{\beta})}$ and $\hat{f}_{\hat{V}(\hat{\beta})}$, obtain the test statistic S_n.

Once the test statistic is computed, we have to compare it to a critical value in order to define a decision rule for the test. Since H_0: $\sqrt{n}Pg_{V(\beta_0),t} = 0$ for all $t \in \mathcal{T}$, we expect that whenever the null hypothesis holds true the statistic S_n must be relatively close to zero. Hence, for a suitably chosen critical value $c_{1-\alpha}$, a test at the α significance level can be defined as folows:

$$\text{Reject } H_0 \text{ if } S_n > c_{1-\alpha}$$

Ideally, we would set the critical value $c_{1-\alpha}$ to the $1-\alpha$ quantile of the null distribution of S_n, but since such a distribution is unknown we will approximate such a quantile. Unfortunately, in our context we cannot rely on the usual asymptotic approximation: as we show in Section 4, the cdf of $S_n$will converge to a complicated distribution with unknown quantiles, so that we cannot use these quantiles in order to approximate the critical values of the test statistic. Differently, we will use a bootstrap procedure and simulate the critical values. To this end, we have to first build up a Bootstrap Data Generating Process from which we must resample the observations. In this paper, we use a Wild-bootstrap procedure, so that the bootstrap resampling scheme will rely on the functional forms imposed under the null hypothesis. Intuitively, since the null hypothesis is imposed in the bootstrap resampling scheme, we should obtain a good approximation of the critical values when the null hypothesis holds true. Now, in our context in order to obtain a wild-bootstrap procedure leading to a consistent test, it will not be sufficient to bootstrap only the estimator $\hat{G}$: since the framework we have at hand is featured by the non-observability of the regressors $V(\beta_0)$, we will need to bootstrap also the estimators of these covariates, and this implies that we have to bootstrap both the estimator $\hat{\beta}$ of the finite dimensional parameter, and the nonparametric estimators $\hat{m}_0$ and $\hat{m}_2$.

In more details, consider the sample estimates $\{\hat{G}_{\hat{V}(\hat{\beta})}(\hat{V}_i(\hat{\beta}))\ ,\ \hat{m}_0(Z_i)\ ,\ \hat{m}_1(X_{1,i})\}_{i=1}^{n}$. Furthermore, define $\{\xi_i\}_{i=1}^{n}$ to be a sequence of weights independent from the sample data, randomly drawn in a iid fashion from a distribution $\mathbb{P}^{\xi}$. Then, let

$$D^*_{0,i} := \hat{m}(Z_i) + \xi_i \cdot (D_{0,i} - \hat{m}_0(Z_i)) \ , \ D^*_{2,i} := \hat{m}_2(X_{2,i}) + \xi_i \cdot (D_{2,i} - \hat{m}_2(X_{2,i}))$$

and

$$Y^*_i := \hat{G}_{\hat{V}(\hat{\beta})}(\hat{V}_i(\hat{\beta})) + \xi_i \cdot (Y_i - \hat{G}_{\hat{V}(\hat{\beta})}(\hat{V}_i(\hat{\beta}))) \text{ for } i = 1, .., n$$

The bootstrapped sample (the sample from the bootstrap DGP) will be$\{Y^*_i, X_{1,i}, X_{2,i}, X^e_i, Z_i, D^*_{0,i}, D^*_{2,i}\}^n_{i=1}$. By replacing this sample to the original sample $\{Y_i, X_{1,i}, X_{2,i}, Z_i, D_{0,i}, D_{2,i}\}^n_{i=1}$, we can compute the bootstrap version S^*_n of the statistic S_n by implementing Step 1-Step5 described above. Specifically, the precedure for running the bootstrap test goes as follows:

Step 1B) by regressing D^*_0 onto Z and D^*_2 onto X_2, obtain the bootstrap versions $(\hat{m}^*_0, \hat{m}^*_2)$ of $(\hat{m}_0, \hat{m}_2)$, respectively.

Step 2B) Given the bootstrap version $\hat{V}^*(\beta)$ of $\hat{V}(\beta)$, regress Y^* onto $\hat{V}^*(\beta)$ so as to get the bootstrap version $\hat{G}^*_{\hat{V}^*(\beta)}$ of the estimator $\hat{G}_{\hat{V}(\beta)}$.

Step 3B) Subsequently, the bootstrapped estimate $\hat{\beta}^*$ of the finite dimensional parameter can be computed as

$$\hat{\beta}^* := \arg\min_{\beta \in B} \sum_{i=1}^{n} [Y^*_i - \hat{G}^*_{\hat{V}^*(\beta)}(\hat{V}^*(\beta))]^2 \ .$$

Step 4B) Given $\hat{V}^*(\hat{\beta}^*)$, get the bootstrap version $\hat{f}^*_{\hat{V}^*(\hat{\beta}^*)}$ of the estimated density $\hat{f}_{\hat{V}(\hat{\beta})}$, by regressing Y^* onto $\hat{V}^*(\hat{\beta}^*)$ obtain the bootstrap version $\hat{G}^*_{\hat{V}^*(\hat{\beta}^*)}$ of $\hat{G}_{\hat{V}(\hat{\beta})}$.

Step 5B) Finally obtain the bootstrap version $S^*_n$of the statistic S_n.

Now, by extracting from the bootstrap DGP a very large number of samples, say n_B, each of size n, and performing Step 1B-5B above, we can obtain a sequence $S^*_{n,1}, .., S^*_{n,n_B}$ of bootstrapped statistics. Finally, we consider the $(1 - \alpha)$-quantile of such a distribution as an estimate of the critical value for running the test at the α- significance level.

3 Applications

In this section, we provide several examples of aplications for our test, ranging fom discrete-choice semiparametric models with endogenous regressors, to incomplete information games and sample-selection models.

3.1 Models with endogenous regressors

Semiparametric binary-choice model. The first application we present for our test stands in the specification of a semiparametric binary-choice model, where some of the regressors are allowed to be endogenous, i.e. correlated with an unobserved error term. Let $Y \in \{0,1\}$ denote the discrete choice of an agent, and let us assume that such a choice depends on both a set of covariates $X \in \Re^d$ and an unobserved error term u, according to the model

$$Y = 1\{X_1^T \cdot \beta_{0,1} \geq u\} \tag{1}$$

where $\beta_{0,1} \in \Re^{p_1}$ is a finite-dimensional parameter unknown to the researcher. If we assume that the unobserved component u is independent from the set of regressors X_1, and we do not specify its distribution in a parametric way, the above model will boil-down to a semiparametric single-index equation, for which several estimation and specification methods have been proposed in the literature. On the other hand, following Blundell & Powell (2004) and Rothe (2009), we will allow some of the components of X_1 to be correlated with the unobserved error u, and hence to be endogenous. To this end, let us consider the partition $X_1 = (X^e, Z^{(1)})$, where X^e denotes the set of regressors that are endogenous while $Z^{(1)}$stands for the exogenous regressors. In order to control for such endogeneity, Blundell & Powell (2004) introduce a control function. So, let V be the residual from the nonparametric regresion of X^e onto the vector $Z := (Z^{(1)}, Z^{(2)})$, with $Z^{(2)}$ being another set of exogenous variables, i.e.

$$X^e = m_0(Z) + V \text{ with } \mathbb{E}\{V|Z\} = 0 \tag{2}$$

The function m_0 is not restricted in any parametric way, and is allowed to be nonparametric. In order to identify the components of the model and estimate it, Blundell and Powell impose an exclusion restriction on the conditional distribution of u, so that

$$u|X_1, Z \sim u|X_1, V \sim u|V \tag{3}$$

where the symbol "$\sim$" denotes equality in distribution[1]. The residual V is therefore called "control function", as it is the variable allowing the researcher to control for the presence of endogeneity. Denote with $F_{u|S}(\tilde{u}, s)$ the conditional distribution of u given $S = s$, computed at the value $\tilde{u}$. Indeed, under the above exclusion restriction, we obtain

$$\mathbb{E}\{Y|\tilde{X}\} = \mathbb{E}\{1\{X_1^T \cdot \beta_{0,1} \geq u\}|X_1, Z\} = F_{u|X_1,Z}(X_1^T \cdot \beta_{0,1}, (X_1, Z)) = F_{u|V}(X_1^T \cdot \beta_{0,1}, V) \ , \tag{4}$$

with $\tilde{X} := X \cup Z$, and hence the following moment equation

$$\mathbb{E}\{Y|\tilde{X}\} = \mathbb{E}\{Y|X_1^T \cdot \beta_{0,1}, X^e - m_0(Z)\}$$

which is a specific example of the semiparametric dimensionality-reduction condition at the center of our test.

Remark 1. The finite dimensional-parameter $\beta_{0,1}$ can be indentified by mean of specific exclusion restrictions combined with other conditions, as in Blundell & Powell (2004) and Rothe (2009). This means that some of the exogenous variables -i.e. those in $Z^{(2)}$- in the auxiliary regression function m_0, must not appear in -i.e. must be excluded from- the original equation, and hence cannot be part of X_1. On the other hand, Escanciano et al. (2016) show that in the presence of nonlinearities in the function m_0, these exclusion restrictions are not necessary, and hence identify the components of the model without using "instruments".

[1] The condition in Eq. (3) is a conditional independence restriction requiring that the distribution of the unobserved error u conditonal on X_1 and Z must depend only on the unobserved error term V from the auxiliary regression. In other words, as long as the residual V is kept fixed, the fluctuations of the error u must be independent from the exogenous variables Z, and this allows to exclude the regressors X_1 from the conditional distribution of u.

We now show how to adapt the general testing procedure described in Section 5 to the specific framework at hand. In terms of the general model of Section 5, we can set $D_2 = 0$ and $D_0 = X^e$, so as to get $\beta = \beta_1$and $V(\beta) = (X_1^T \beta_1, X^e - m_0(Z))$. Hence, in order to compute the test statistic S_n we can implement Step 1-Step5 as described in Section 5, with the only difference that we will not need to compute $\hat{m}_2$ in Step 1. As regards the bootstrap critical values, notice that here $D_2 = 0$, so in the bootstrap DGP we do not need to generate the variable D_2^*. Hence, in order to obtain the bootstrapped statistic S_n^* we can implement Step 1B-Step5B as described in Section 5, without computing the estimator $\hat{m}_2^*$ in Step 1B.

Separable-regression model. Our test can also be employed for the specification of a semi-parametric version of the model proposed by Newey *et al.* (1999). Let Y be a continuous variable and X_1 be a set of regressors. A structural model explaining Y as a function of X_1 can be written as

$$Y = G(X_1) + \varepsilon$$

where ε is an unobserved component, while G is a structural regression function the researcher is interested in. If the error term ε is mean independent from the covariates X_1, i.e. $\mathbb{E}\{\varepsilon|X_1\} = 0$, the model can be estimated by classical nonparametric techniques, such as kernel, series, splines, etc. On the other hand, in the presence of correlation between the error term ε and some of the explanatory variables in X_1, say X^e, the mean independence condition breaks down, i.e.$\mathbb{E}\{\varepsilon|X_1\} \neq 0$, and the classical nonparametric methods no longer apply. So, similarly to the previous section, let us partition $X_1 = (X^e, Z^{(1)})$, where X^e is supposed to be the endogenous regressor while $Z^{(1)}$ is the set of regressors which are exogenous. In such a contest, Newey *et al.* (1999) introduce a control function V which can be used in order to handle the presence of endogeneity. This control function is defined in the same way as in Eq. (2). In order to identify the function of interest G, Newey *et al.* (1999) impose the following restriction on the conditional mean of the error term

$$\mathbb{E}\{\varepsilon|X_1, Z\} = \mathbb{E}\{\varepsilon|V\} \tag{5}$$

The above equation represents a mean-independence condition[2]. Such an exclusion restriction together with the condition that $\mathbb{E}\{\varepsilon\} = 0$, allows the authors to identify the function G . Now, the estimation of this structural function can be carried out in a fully nonparametric way, by the method proposed in Newey *et al.* (1999). However, in the presence of an important dimension of the regressor X_1, say 3 or more, the purely nonparametric approach might in practice suffer from the curse of dimensionality. In order to attenuate it, we could adopt the approach used in a large part of the statistical literature -which is also proposed in Newey *et al.* (1999)-, and hence impose a single-index restriction on the functon G, i.e. $G(X_1) = G_0(X_1^T \cdot \beta_{0,1})$, with $\beta_0 \in \Re^d$. Newey *et al.* (1999) show the identification of both this finite dimensional parameter and the function G_0. By the single-index restriction and Eq. (5), we get the following restriction on the distribution of the data

$$\mathbb{E}\{Y|\tilde{X}\} = \mathbb{E}\{Y|X_1^T \cdot \beta_{0,1}, X^e - m_0(Z)\} \text{ with } \tilde{X} := (X_1, Z)$$

which is a specific case of the equation at the center of our test. The general bootstrap test described in Section 5 can be applied to this framework in exactly the same way as described for the semiparametric binary-choice model above.

Binary-choice model with endogeneity and structure on the error term. A further application of our test consists in the specification of an endogenous binary-choice model of the same type as the one presented above, where however the researcher imposes some structure on the unobserved error term in order to reduce the curse of dimensionality. Specifically, consider the same model as in Eq. (1), and partition $X_1 := (X_1^e, X_2^e, Z^{(1)})$, where X_1^e and X_2^e are two endogenous regressor correlated with the unobserved error term u. Since we have two endogenous regressors, we will need two control functions in order to handle such endogeneity and estimate the components of the model. So, define the two control functions V_1 and V_2 as

$$V_1 := X_1^e - m_0(Z),\ V_2 := X_2^e - m_2(Z_2)\ ,\ \mathbb{E}\{V_1|Z\} = 0\ ,\ \mathbb{E}\{V_2|Z_2\} = 0\ .$$

[2]The exclusion restriction in Eq. (5) tells us that the conditional mean of the error ε is allowed to depend on the set of regressors X_1 and the instruments Z only through the residual V. This implies that, as long as the residual V is kept fixed, the instrumental variables Z do not have to impact the mean of the error ε and hence, coditionally on V, the residual ε is mean independent from the covariates X_1.

Define $\tilde{Z} := (Z^{(1)}, Z_2, Z)$ and $\tilde{X} := (X_1 \cup \tilde{Z})$. By imposing a conditional independence restriction on the distribution of th error u similarly as Eq. (3), i.e. $u|X_1, \tilde{Z} \sim u|V_1, V_2$, we obtain $\mathbb{E}\{Y|\tilde{X}\} = \mathbb{E}\{X_1 \cdot \beta_{0,1}, V_1, V_2\}$. Hence, the estimation of the finite dimensional parameter β_0 and the Average Structural Function will require the nonparametric estimation of a function with three arguments, $\mathbb{E}\{X_1 \cdot \beta_1, V_1, V_2\}$, i.e. a triple-index model. In order to reduce such a curse of dimensionality and increase the tractability of the framework, we can impose a further structure on the unobserved error term u. So, assume that

$$u = \gamma_0 V_2 + \varepsilon \text{ with } \varepsilon|X_1, \tilde{Z} \sim \varepsilon|V_1 \tag{6}$$

By replacing the expression of u from Eq. (6) into Eq. (1), and then proceeding in the same way as in Eq. (4),

$$\mathbb{E}\{Y|\tilde{X}\} = \mathbb{E}\{Y|X_1^T \cdot \tilde{\beta}_{0,1} + \gamma_0 m_2(Z_2),\ X_1^e - m_0(Z)\} \tag{7}$$

where $\tilde{\beta}_{0,1} := (\beta_{0,1}^{(1)} - \gamma_0, \beta_{0,1}^{(2)T})^T$, $\beta_{0,1}^{(1)}$ is the first component of $\beta_{0,1}$, while the vector $\beta_{0,1}^{(2)}$ gathers all the remaining $(d_1 - 1)$ components of the vector $\beta_{0,1}$. Notice that Eq. (6) implies that $u|X_1, \tilde{Z} \sim u|V_1, V_2$. If Eq. (7) holds true, then the estimation of the finite dilensional parameter $\beta_{0,1}$ and the Average Structural Function require an estimate of the conditional expectation $\mathbb{E}\{Y|X_1^T \cdot \tilde{\beta}_1 + \gamma m_2(Z_2),\ X_1^e - m_0(Z)\}$, which involves two indeces. Accordingly, the constraint in Eq. (6) can attenuate the curse of dimensionality and therefore simplify the estimation of the structural elements the researcher is interested in.

Now, Eq. (7) is a particular case of the general moment condition at the center of this paper. In terms of the general model of Section 5, we only need to adapt the notation and set $D_2 = X_2^e$, $D_0 = X_1^e$, $X^e := X_1^e$, $\beta_1 := \tilde{\beta}_1$, $\beta_2 := \gamma$, $X_2 := Z_2$, so as to get $\beta = (\tilde{\beta}_1, \gamma)$ and $V(\beta) = (X_1^T \cdot \tilde{\beta}_1 + \gamma \cdot m_2(Z_2),\ X_1^e - m_0(Z))$. Hence, in order to compute the test statistic S_n we can implement Step 1-Step5 as described in Section 5, where in Step 1 in order to compute $\hat{m}_0$ we will regress X_1^e onto Z and in order to compute $\hat{m}_2$ we will regress X_2^e onto Z_2. As regards the bootstrap critical values, we can implement Step 1B-Step5B as described in Section 5, where in

Step 1B we will regress X_1^{e*} onto Z in order to compute $\hat{m}_0^*$, and X_2^{e*} onto Z_2 in order to compute $\hat{m}_2^*$.

3.2 Semiparametric models with nonparametric sample selection

Non-separable semiparametric models with sample-selection constitute another application of our test. Estimation and identification of these models is described in Escanciano et al. (2016) and Rothe (2009). Let $\tilde{Y}$ be a scalar denoting the decision by an agent, and let us assume that such a decision depends on both a set of covariates X_1 and an unobserved component ε, according to the model

$$\tilde{Y} = \varphi_0(X_1^T \cdot \beta_{0,1}, \varepsilon)$$

where φ_0 is a function known by the researcher. In the presence of sample selection, the agents' decision $\tilde{Y}$ will be observed only in the selected sample. Let $\tilde{D}$ denote the selection variable. For the aim of generality, we can specify the selection mechanism in a nonparametric way as

$$\tilde{D} = 1\{H_0(Z) \geq u\} \tag{8}$$

where we assume, without loss of generality, that $u \sim U[0,1]$, so the function H_0 will be identified as the conditional expectation of $\tilde{D}$, i.e. $\mathbb{E}\{\tilde{D}|Z\} = H_0(Z)$. If $\tilde{D} = 1$ the individual is selected and his decision $\tilde{Y}$ is observed, while if $\tilde{D} = 0$ the individual is not selected and hence his decision $\tilde{Y}$ is not observed. This implies that the variable the researcher observes is $Y := \tilde{Y} \cdot \tilde{D}$, and hence

$$Y = \tilde{D} \cdot \varphi_0(X_1^T \cdot \beta_{0,1}, \varepsilon)\ .$$

For exampe , we can set $\varphi_0(X_1^T \cdot \beta_{0,1}, \varepsilon) = 1\{X_1^T \cdot \beta_{0,1} \geq \varepsilon\}$ or $\varphi_0(X_1^T \cdot \beta_{0,1}, \varepsilon) = \max\{0, X_1^T \cdot \beta_{0,1} + \varepsilon\}$. In the former case, we would have $\tilde{Y} \in \{0,1\}$, and hence a binary-choice model with sample selection; in the latter case we would have $\tilde{Y} = \max\{0, X_1^T \cdot \beta_{0,1} + \varepsilon\}$, and hence a truncated regression moldel where $\tilde{Y}$ could be observed only when taking positive values in the selected sample. This truncated regression model with sample selection is also called double hardle-model[3].

In the sample selection literature, the errors (ε, u) are assumed to be jointly independent from the set of covariates (X_1, Z), i.e. $(\varepsilon, u) \perp (X, Z)$, but they are allowed to be dependent from each other.

The model so far described is a semiparametric version of the Heckman sample-selection model which parameterizes the joint distribution of (ε, u). On the other hand, we here leave this distribution as nonparametrically specified, and denote it with $F_{\varepsilon,u}$. By combining the equations so far displayed we get

$$\mathbb{E}\{Y|\tilde{X}\} = \int 1\{u \leq H_0(Z)\} \cdot \max(0, X_1^T \beta_{0,1} + \varepsilon) \, \mathrm{d}F_{\varepsilon,u} =: G(X_1^T \cdot \beta_{0,1}, \ H_0(Z)) \text{ with } \tilde{X} = (X_1, Z) \ ,$$

where $\tilde{X} = (X_1 \cup Z)$, so that the restriction implied on the distribution of the data writes as

$$\mathbb{E}\{Y|\tilde{X}\} = \mathbb{E}\{Y|X_1^T \cdot \beta_{0,1}, \ H_0(Z)\}$$

which is a particular example of the equation we want to test.

Remark 2. The identificaton of the components of the above model can be obtained in the same way as in the semiparametric binary-choice model with endogenous regressors described in the first part of this section, i.e. by using the same derivations as in Blundell & Powell (2004), Rothe (2009), or Escanciano et al.(2016)[4].

[3] On the empirical standpoint, a potential application of the binary-choice model with selection can be, for example, the estimation of the decision about whether to buy a good or not, by using observations on a cross section of individuals. In this case, the variable $\tilde{D}$ will represent whether the good is available in the place where the person leaves, while $\tilde{Y}$ will represent the decision to purchase the good. Differently, for the truncated regression model $\tilde{Y}$ can represent the agents' consumption of durable goods.

[4] Notice that in this model the equation for $\tilde{Y}$ is specified in a semiparametric way, as $\tilde{Y}$ depends on X_1 through a linear-index while, on the other hand, the dependence of $\tilde{D}$ upon Z is allowed to be nonparametric. Such a difference can be explained by the fact that the researcher might have a theory allowing him to justify the (quasi) parametric form of the relationship between $\tilde{Y}$ and X_1, but might not have any justification allowing him to impose a parametric restriction on the relationship between the selection variable $\tilde{D}$ and the covariates Z. Alternatively, this difference can be sustained by the following pragmatic consideration. Normally, researchers are interested in the estimation of the relationship between $\tilde{Y}$ and X_1, and at the same time they wish to give an easy interpretation to the effect the variables X_1 might have on $\tilde{Y}$: the linear index dependence between $\tilde{Y}$ and X_1 is a parametric restriction allowing the researcher to give such an easy interpretation. On the other hand, the impact of Z onto $\tilde{D}$ is not at the center of the attention for the researcher, hence in order to be as safe as possible from the danger of mispecification, the function H_0 can be left nonparametric. These arguments also apply to the contro function models described in the previous section.

We now outline how to adapt the general testing procedure described in Section 5 to the specific model considered in this section. The procedure is pretty similar to the one described for the semiparametric binary-choice model with endogenous regressors. In terms of the general model described in Section 5, we set $D_2 := 0$, $D_0 := -\tilde{D}$, $m_0 := -H_0$, $X^e = 0$, and get $\beta = \beta_1$ and $V(\beta) = (X_1^T \beta_1, H_0(Z))$. For the computation of the test statistic S_n, we can implement Step 1-Step 5 as described in Section 5, by taking care of two modifications of Step 1: we do not need to compute $\hat{m}_2$, and in order to get $\hat{m}_0$ we have to regress $-\tilde{D}$ onto Z. For obtaining the bootstrap critical value, since in this model $D_2 = 0$, in the bootstrap DGP we do not need to generate the variable D_2^*. Hence, we can implement Step 1B-Step5B as described in Section 5, by taking care of two modifications of Step 1B: we do not need to compute $\hat{m}_2^*$ and we can compute $\hat{m}_0^*$ by regressing $-\tilde{D}^*$ onto Z.

3.3 Games of incomplete information

Semiparametric incomplete-information games featured by Bayesian-Nash equilibria are a flexible tool to model environments with agents' interaction, and their specification can be tested by using our procedure. We will describe here the model considered by Aradillas-Lopez (2012) and Lewbel & Tang (2015). In order to simplify the exposition, let us assume to have two players only. Denote each player by the index $p \in \{1, 2\}$, and assume that each of them must take a binary decision, say $a_p \in \{1, 2\}$. For instance, this variable might represent the decision about entering or not entering a market -similarly to Ciliberto & Tamer (2009)-, about having an aggressive or not aggressive capital strategy -like in Aradillas-Lopez (2010)-, about placing radio commercial -as in Sweeting (2009)- etc. Denote with X_p the exogenous covariates relevant for agent p's decision, and assume that these covariates can be observed by each player and the researcher. For notational simplicity, we write $\tilde{X} = (X_1 \cup X_2)$. Each player has a private information, denoted by u_p, that neither the other agent nor the researcher can observe. Such private informations are mutually independent, i.e. $u_1 \perp u_2$, and they are jointly independent from the exogenous covariates, $(u_1, u_2) \perp \tilde{X}$. Although player p does not know the value of the other player's private information, u_{-p}, he is aware about the distribution of such a variable, so that player p will not observe u_{-p} but will know its distribution denoted by F_{-p}. On the other hand, we assume that the researcher does not know this distribution that is therefore nonparametrically specified. The payoff function of player p takes the form

$$\Pi_p(a_p, \tilde{X}, a_{-p}, u_p) = a_p \cdot [x_p \cdot \beta_{0,p} - \alpha_{0,p} \cdot a_{-p} - u_p]$$

where $\beta_{0,p}$ is a finite dimensional parameter. Hence, if player p takes decision $a_p = 1$, he will incur in the profit $x_p \cdot \beta_p - \alpha_{0,p} \cdot a_{-p} - u_p$, while with the decision $a_p = 0$, he will have zero profits, i.e. the profits from decision 0 are normalized to 0. Now, in the presence of simultaneous moves, and since the game is featured by incomplete information and risk-neutral agents, each player must base his decisions on his expected profits, and therefore will decide to maximize this objective function when taking his move. Furthermore, if we assume that agents will play a unique Bayesian-Nash equilibrium, the model will imply the following semiparametric restriction on the distribution of the data (see Aradillas-Lopez (2012) and Lewbel & Tang (2015))

$$\mathbb{E}\{a_p|\tilde{X}\} = \mathbb{E}\{a_p|X_p^T \cdot \beta_{0,p} - \alpha_{0,p} \cdot \mathbb{E}\{a_{-p}|\tilde{X}\}\} \text{ for } p = 1, 2$$

which is a version of the moment condition we wish to test[5].

Remark 3. The identification of the finite dimensional parameter $\beta_{0,p}$ is discussed in Aradillas-Lopez (2012) and Lewbel & Tang (2015), and basically requires the presence of specific profit shifters, in the sense that some of the covariates influencing the profit of agent 1 cannot directly enter the profit function of agent 2, and vice versa.

We now show how to adapt the general testing procedure described in Section 5 to the moment conditions arising from the game theoretical model. Assume wanting to test $\mathbb{E}\{a_1|\tilde{X}\} = \mathbb{E}\{a_1|X_1^T \cdot \beta_{0,1} - \alpha_{0,1} \cdot \mathbb{E}\{a_2|\tilde{X}\}\}$. In terms of the general model of Section 5, we can set $D_2 := a_2$, $X_2 := \tilde{X}$, $m_2(X_2) := \mathbb{E}\{a_2|\tilde{X}\}$, $\beta_{0,2} := -\alpha_1$, and $D_0 = X^e = 0$, so as to get $\beta = (\beta_1, -\alpha_1)$ and $V(\beta) :=$

[5]From the empirical standpoint, this setup can be used in order to model the decision of airline companies about offering or not offering a connection between two different airports, like in Ciliberto & Tamer (2009), by assuming that each company has an incomplete information about the other companies' moves; or to describe firms' capital investment strategies, like in Aradillas-Lopez (2010), so that a_p will denote the decision of using an aggressive capital investment strategy. Other examples include the decisions of broadcasters to place radio commercials, as in Sweeting (2009). More appliations are reported in Lewbel & Tang (2015).

$(X_1^T\beta_1 - \alpha_1 \cdot \mathbb{E}\{a_2|\tilde{X} = \tilde{x}\})$. In order to compute the test statistic S_n we can implement Step 1-Step5 as described in Section 5, with the only difference that we will not need to compute $\hat{m}_0$ in Step 1. So, when implementing Step 1, we only need to regress a_2 onto $\tilde{X}$ so as to get $\hat{m}_2$. As regards the bootstrap critical values, notice that since here $D_0 = 0$, in the bootstrap DGP we do not need to generate the variable D_0^*. Hence, in order to obtain the bootstrapped statistic S_n^* we can implement Step 1B-Step5B as described in Section 5, by taking care of two modifications of Step 1B: we do not have to compute the estimator $\hat{m}_0^*$, and in order to obtain $\hat{m}_2^*$ we have to regress a_2^* onto $\tilde{X}$.

4 The Estimators and the Assumptions

This section builds upon Section 5 and introduces in detail the estimators used and the Assumptions necessary to obtain the properties of the test proposed. So, let us start by contructing the estimator $(\hat{m}_0, \hat{m}_1)$ and $\hat{\beta}$. For the sake of simplicity and in order to obtain tactable proofs we use kernel estimators, so we define the nonparametric estimators for m_0 and m_2 as

$$\hat{m}_0(z) := \frac{\hat{T}_0(z)}{\hat{f}_0(z)} \ , \ \hat{T}_0(z) := \frac{1}{nh_0^{p_0}} \sum_{i=1}^{n} D_{0,i} K_0\Big(\frac{Z_i - z}{h_0}\Big) \ , \ \hat{f}_0(z) := \frac{1}{nh_0^{p_0}} \sum_{i=1}^{n} K_0\Big(\frac{Z_i - z}{h_0}\Big)$$

and

$$\hat{m}_2(x_2) := \frac{\hat{T}_2(z)}{\hat{f}_2(z)} \ , \ \hat{T}_2(x_2) := \frac{1}{nh_2^{p_2}} \sum_{i=1}^{n} D_{2,i} K_2\Big(\frac{X_{2,i} - x_2}{h_2}\Big) \ , \ \hat{f}_2(z) := \frac{1}{nh_2^{p_2}} \sum_{i=1}^{n} K_2\Big(\frac{X_{2,i} - x_2}{h_2}\Big)$$

where h_0 and h_2 are bandwidth rates converging to zero, K_0 and K_2 are kernel functions, while $\hat{f}_0$ and $\hat{f}_2$ can be interpreted as the estimators of the densities of Z and X_2, respectively. In order to estimate the function $G_{V(\beta)}$, we will also use nonparametric kernel methods, where each unobserved regressor is replaced by its estimator , so that

$$\hat{G}_{\hat{V}(\beta)}(w) := \frac{\hat{T}_{\hat{V}(\beta)}(w)}{\hat{f}_{\hat{V}(\beta)}(w)} \ ,$$

$$\hat{T}_{\hat{V}(\beta)}(w) := \frac{1}{nh^d}\sum_{i=1}^{n} Y_i K\Big(\frac{\hat{V}_i(\beta)-w}{h}\Big)\cdot \hat{t}_{n,i}^{(X_2,Z)},\ \hat{f}_{\hat{V}(\beta)}(w) := \frac{1}{nh^d}\sum_{i=1}^{n} K\Big(\frac{\hat{V}_i(\beta)-w}{h}\Big)\cdot \hat{t}_{n,i}^{(X_2,Z)},$$

where

$$\hat{t}_{n,i}^{(X_2,Z)} := 1\{\hat{f}_2(X_2)\geq \tau_n\}\cdot 1\{\hat{f}_0(Z)\geq \tau_n\}$$

with τ_n being a sequence of numbers converging to zero. The sequence $\hat{t}_{n,i}^{(X_2,Z)}$ is a sequence of trimmings whose role is to take care of the random denominators in the functions $\hat{m}_0$ and $\hat{m}_1$: essentially, $\hat{t}_{n,i}^{(X_2,Z)}$ excludes from the computation of the estimator $\hat{G}_{\hat{V}(\beta)}$ all those observations Z_i and $X_{1,i}$ for which $\hat{f}_0(Z_i)$ and $\hat{f}_2(X_{2,i})$ are close to zero and would hence be estimated in a very unprecise way.

In order to estimate the finite dimensional parameter β_0, we minimize the semiparametric least-square criterion function, so that

$$\hat{\beta} = \text{argmin}_{\beta\in B}\frac{1}{n}\sum_{i=1}^{n}[Y_i - \hat{G}_{\hat{V}(\beta)}(\hat{V}_i(\beta))]^2\cdot \hat{t}_{n,i}^{(X_2,Z)}\cdot \hat{t}_{n,i}^{(X,\hat{V})} \tag{9}$$

where B is a compact set containing the true parameter value β_0, $\hat{t}_{n,i}^{(X_2,Z)}$ is the trimmming sequence defined above, and $\hat{t}_{n,i}^{(X,\hat{V})}$ is instead set as

$$\hat{t}_{n,i}^{(X_1,\hat{V})} := 1\{\hat{f}_{X,\hat{V}}(X,\hat{V})\geq \tau_n\}\ ,$$

with

$$\hat{f}_{X,\hat{V}}(x,v) := \frac{1}{nh_3^q}\sum_{i=1}^{n} K_3\Big(\frac{X_i - x}{h_3},\frac{\hat{V}_i - v}{h_3}\Big)\cdot \hat{t}_{n,i}^{(Z)}$$

where, similarly to the kernel estimators above introduced, h_3 is a sequence of bandwidths converging to zero, K_3 is a kernel function, $q = dim(X,V)$, and finally $\hat{t}_{n,i}^{(Z)} := 1\{\hat{f}_0(Z_i)\geq \tau_n\}$. The use of the trimmings $\hat{t}_{n,i}^{(X,\hat{V})}$ and $\hat{t}_{n,i}^{(X_2,Z)}$ is necessary in the semiparametric least square criterion, in order to control for the random denominators of the functions $\hat{G}_{\hat{V}(\beta)}$, $\hat{m}_2$, and $\hat{m}_0$: as already highlighted above, the trimming $\hat{t}_{n,i}^{(X_2,Z)}$ excludes all those observations for which the estimated densities $\hat{f}_2$ and $\hat{f}_0$ -and hence the denominators of $\hat{m}_0$ and $\hat{m}_2$-, are close to zero, while the trimming $\hat{t}_{n,i}^{(X_1,\hat{V})}$ essentially plays the same role for the estimator $\hat{G}_{\hat{V}(\beta)}$.

At this stage, we have all the ingredients to build up the estimator $\hat{g}_{\hat{V}(\hat{\beta}),t}$, i.e. the empirical counterpart of the function $g_{V(\beta_0),t}$. So, given the estimators defined above, let

$$\hat{g}_{\hat{V}(\hat{\beta}),t}(y,\tilde{x}) := \left[y - \hat{G}_{\hat{V}(\hat{\beta})}(\hat{v}(\hat{\beta}))\right] \cdot \hat{f}_{\hat{V}(\hat{\beta})}(\hat{v}(\hat{\beta})) \cdot \hat{t}_n^{(X_2,Z)}(x_2,z)$$

where, again, the trimming $\hat{t}_{n,i}^{(X_1,Z)}$ is employed in order to control for the random denominators present in $\hat{m}_0$ and $\hat{m}_1$, and hence in $\hat{V}(\hat{\beta})$. At this point, it also becomes clear the reason why we have made the density $f_{V(\beta_0)}$ appear in the original null hypothesis: thanks to the presence of such a density, when taking the empirical counterpart of the function $g_{V(\beta_0)}$we can get rid of the random denominator in $\hat{G}_{\hat{V}(\hat{\beta})}$, and hence obtain relatively tractable proofs.

Having defined the test statistic at the basis of our procedure, in order to build up a test we will need to construct the critical values the statistic S_n will be compared to. In order to do so, we will first obtain the asymptotic behavior of S_n and then propose a bootstrap procedure for the computation of the critical values. We therefore close this section by presenting the set of assumptions we will use.

Assumption 1 (IID) $\{Y_i, X_{1,i}, X_{2,i}, X_i^e, Z_i, D_{1,i}, D_{2,i}\}_{i=1}^n$ is an IID sequence of random variables defined over the probability space $(\Omega, \mathcal{A}, P)$.

Assumption 2 (Smoothness) (i) The functions $G_{V(\beta_0)}$, $f_{V(\beta_0)}$ are r-times continuously differentiable in their arguments, with uniformly bounded derivatives; (ii) The random variable $\tilde{X}$ admits a density conditionally on the index $V(\beta_0)$, which is denoted as $(\tilde{x}, v) \mapsto f_{\tilde{X}|V(\beta_0)}(\tilde{x}|V(\beta_0) = v)$ and is r-times continuously differentiable in v, and $(r_1 \vee r_0)$-continuously differentiable in $\tilde{x}$, with uniformly bounded derivatives; (iii) m_j and f_j for $j = 0, 2$, are r_j-times continuously differentiable, respectively, with uniformly bounded derivatives; (iv) the density $f_{X_1,V}$ is r_3-times continuously differentiable with uniformly bounded derivatives.

Assumption 3 (Kernels) (i) The kernel K (used for the estimation of $G_{V(\beta_0)}$ and $f_{V(\beta_0)}$) is a product kernel, i.e. $K(\cdot) = \prod_{j=1}^d k(\cdot)$, where k is a univariate kernel of order r; (ii) the kernel

K_j (employed for the estimation of m_j and f_j, for $j = 0, 2$) is also a product kernel, so that $K_j = \prod_{s=1}^{p_j} k_j$ with k_j being a univariate kernel of order r_j, for $j = 0, 2$; (iii) the kernel K_3 (used for the estimation of $f_{X,V}$) is a product kernel, $K_3 = \prod_{i=1}^{q} k_3$, with k_3 univariate kernel of order r_3.

Assumption 1 is common in the literature on nonparametric estimation and testing, and in general requires to have a cross-section of observations. As regards Assumption 2, it overall imposes a certain degree of smoothness on the functions involved, which is connected with the orders of the kernels described in Assumption 3. Such a smoothness condition is very common in the literature on semiparametric estimation and can basically be considered as a mild one. The part of Assumption 2 that is instead not very common in the statistical literature is the existence of the density $f_{\tilde{X}|V(\beta_0)}$. Notice, however, that as long as the variable $\tilde{X}$ admits a density -i.e. $P^{\tilde{X}}$is dominated by the Lebesgue measure-, by the Radon-Nikodyn Theorem also $\tilde{X}$ conditionally on $V(\beta_0)$ will admit a density, and hence the existence of $f_{\tilde{X}|V(\beta_0)}$ will be ensured: in this contex, therefore, the existence of such a function is not a strong restriction. On the other hand, what Assumption 2 (i) is doing is to impose a smoothnes property on such a density, which can again be considered as a mild requirement.

Now, as already noticed in the part of this section where we have constructed our test statistic, the framework at hand is featured by non-observed regressors. This imples that the bandwidth rates will have to be connected with the rates at which the densities of the observed variables go to zero on the tails. So, define

$$p_n := P\Big(f_{V(\beta_0)}(V(\beta_0)) \geq \tau_n\Big) \; , \; p_{n,0} := P\Big(f_Z(Z) \geq \tau_n\Big)$$

$$, \; p_{n,2} := P\Big(f_{X_2}(X_2) \geq \tau_n\Big) \; , \; p_{n,3} := P\Big(f_{X,V}(X,V) \geq \tau_n\Big) \; .$$

Also, for any integer $\alpha \in \mathbb{N}$, denote with $\underline{Ev}(\alpha)$ the largest even integer weakly smaller than α.

Assumption 4 (Bandwidht for testing) (i) $\frac{(\log n)^2}{n \cdot h_j^{2p_j} \cdot \tau_n^{12}} \to 0$, $n \cdot h_j^{r_j} \cdot \tau_n^{-4} \to 0$, for $j = 0, 2$; (ii) $\frac{(\log n)^2}{n \cdot h^{2d+4} \tau_n^8} \to 0$, $n \cdot h^{2r} \to 0$; (iii) $p_n = o(l_n^{-1})$, $p_{n,0} = o(l_n^{-1})$, $p_{n,2} = o(l_n^{-1})$, where l_n satisfies

$\frac{n^{1/4}}{h^{d+1}\tau_n^2 l_n} \to 0$, $\frac{n^{1/2}}{l_n} \to +\infty$.

Assumption 5 (Bandwidths for estimation) (i) $p_{n,3} = o(l_{n,3}^{-1})$ and the bandwidth h_3 used for the estimation of $f_{X,V}$ satisfies: $n \cdot h_3^{q+5} \cdot \tau_n \to +\infty$, $\frac{\log n}{n \cdot h_3^q \tau_n} \to 0$, $\frac{h_3^{r_3}}{\tau_n} \to 0$, $\tau_n \cdot l_{n,3} \cdot h_3^q \to +\infty$; (ii) the bandwidths $h_0$and h_1 used in the estimation of m_0 and m_1, respectively, satisfy: $\frac{\log n}{n \cdot h_j^{p_j+2s_j} \cdot \tau_n^{2s_j+2}} \to +\infty$, $\frac{h_j^{r_j}}{\tau_n} \to +\infty$ for $s_j = 1 + \underline{Ev}(p_j)/2$, for $j = 0, 2$ (iii) The sequence τ_n satisfies $n \cdot \tau_n^{4\cdot(p_j-1)-2\cdot\underline{Ev}(p_j)} \to \infty$ for $j = 0, 2$, and $n \cdot \tau_n^{4\cdot(d-1)-2\cdot\underline{Ev}(d)} \to \infty$.

From Assumption 4 and Assumption 5, we first notice that the bandwidths for the $\hat{m}_j$'s (with $j = 0, 2$) used in the construction of the test statistic must satisfy different requirements from the bandwidths for the $\hat{m}_j$'s used in the estimation of β_0.

In Assumption 4 the speed at which the bandwidths h, h_0, and h_1 have to go to zero is connected with the rate τ_n appearing in the trimming. This is due to the presence of the unobserved regressors $V(\beta_0)$: since we do not observe these covariates, in order to estimate $G_{V(\beta_0)}$ we replace them with their estimates $\hat{V}(\hat{\beta})$ which involve a random denominator. In order to control for such a random denominator, the bandwidths h, h_1, and h_2 have to converge to zero at a rate that is slower with respect to the case where the variables $V(\beta_0)$ were observed. Similarly, the presence of the rate l_n is due to the trimming, and Assumption 4 is requiring that such a sequence must converge to zero at a sufficiently fast rate in order to avoid a bias coming from the presence of the trimming sequences. However, for the practical implementation of the test here proposed, we can avoid the specification of the trimming rates τ_n and l_n. Hence, from a practical standpoint, if we ignore the presence of the rate τ_n, the role of Asumption 4 is to ensure that the estimators $\hat{m}_j$ (for $j = 0, 2$) are $n^{-1/4}$-consistent for their population counterparts m_j, and that also the function $\partial_2 \hat{G}_{\hat{V}(\hat{\beta})}$ is $n^{-1/4}$-consistent for its target $\partial_2 G_{V(\beta_0)}$. This $n^{-1/4}$ rate is required for the presence of nonobserved regressors, as we argue in more detail in Section **??**.

Assumption 5 (i) is imposing a restriction on the trimming sequence appearing in the criterion function used to estimate the parameter β_0. Assumption 5 (ii) is instead requiring that the bandwidth of the $\hat{m}_j$ (with $j = 0, 2$) used in the estimation of β_0 satisfy a condition different from the one established in Assumption 4 for the construction of the test statistic. Specifically, the bandwidth rates for the functions $\hat{m}_j$ used in the estimation of β_0 are setup so as for each $\hat{m}_j$ to be a sufficiently regular function, i.e. a function that is smooth enough. The same role is essentially played by Assumption 5 (iii). In practice, however, as we show in our simulations reported in Section 6, it is not necessary to specify the bandwidths for the $\hat{m}_j$ used in the estimation of β_0 in a different way from the bandwidths for the $\hat{m}_j$ used in the construction of the test statistic, and hence from a practical point of view only the rates established in Assumption 4 (i)-(ii) are relevant

.

Assumption 6 (Pseudo-true value) The mapping $\beta \mapsto \mathbb{E}\{(Y - G_{V(\beta)}(V(\beta)))^2\}$ admits a unique minimum.

This last assumption is imposing the existence of a unique pseudo-true value of the finite dimensional parameter. Such a condition is common to any specification test for nonlinear models involving finite dimensional parameters estimated by maximizing a nonlinear criterion function (see Bierens (1982) , Lavergne Patilea, Escanciano and Van keilegom ??). Assumption 6 basically plays a double role: under the null hypothesis H_0 it ensures the identification of β_0, while under the alternative H_1 it ensures that the estimator $\hat{\beta}$ has a well defined limit in probability.

Having displayed the assumptions at the basis of our proofs, in the next section we will provide the main results of this paper, which concern the asymptotic behavior of the test statistic S_n and the validity of the bootstrap procedure employed to compute the critical values.

5 The Test: Asymptotics and the validity of the Wild Bootstrap

Before displaying the main results, we define a collection of functions striclty connected with the asymptotic behavior of the random variable $\sqrt{n}\mathbb{P}_n \hat{g}_{\hat{V}(\hat{\beta}),t}$ at the basis of our statistic S_n. So, let

$$\varphi_t(y, \tilde{x}) := \left[y - G_{V(\beta_0)}(v(\beta_0))\right] \cdot \psi_t(\tilde{x}) +$$

$$+\mathbb{E}\Big\{\psi_t(\tilde{X}) \cdot \partial_1 G_{V(\beta_0)}(V(\beta_0)) \mid X_2 = x_2\Big\} \cdot \beta_{0,2}^T \cdot (d_2 - m_2(x_2)) +$$

$$+\mathbb{E}\Big\{\psi_t(\tilde{X})\cdot\partial_2 G_{V(\beta_0)}(V(\beta_0))^T\;\Big|\;Z=z\Big\}\cdot(x^e-m_0(z))$$

where $\psi_t(\tilde{x}) := f_{V(\beta_0)}(v(\beta_0))\cdot\phi_t^{\perp}(\tilde{x}) - a(t)^T\Sigma_{\beta_0}^{-1}\nabla_\beta G_{V(\beta)}(m,\tilde{x},\beta_0)$, $\phi_t^{\perp}(\tilde{x}) := \phi(\tilde{x}^T t) - \iota_{V(\beta_0)}(v(\beta_0))$, $\iota_{V(\beta_0)}(w) := \mathbb{E}\{\phi(\tilde{X}t)\mid V(\beta_0)=w\}$, $\Sigma_{\beta_0} := \mathbb{E}\{\nabla_\beta G_{V(\beta)}(m,\beta_0,\tilde{X})\cdot\nabla_{\beta^T}G_{V(\beta)}(m,\beta_0,\tilde{X})\}$, and finally $a(t) := \int\partial_1 G_{V(\beta_0)}(v(\beta_0))\cdot f_{V(\beta_0)}(v(\beta_0))\cdot(x_1^T, m_2(x_2)^T)^T\cdot\phi_t^{\perp}(\tilde{x})\;\mathrm{d}P(\tilde{x})$. As we show in the Appendix, under the null hypothesis H_0 the random variable $\sqrt{n}\mathbb{P}_n\hat{g}_{\hat{V}(\hat{\beta}),t}$ can be approximated by $\sqrt{n}\mathbb{P}_n\varphi_t$, and hence the latter determines the asymptotic behavior of the statistic S_n at the basis of our test. The next proposition shows that indeed a test based on S_n is consistent.

Proposition 4. *Let Assumptions (1)-(6) hold. Then:*

(i) Under H_0: $S_n \rightsquigarrow \int|\mathbb{G}|^2\mu(dt) = \sum_{j=1}^{\infty}\lambda_j\chi^2_{j,1}$,

(ii) Under H_1, $\frac{S_n}{n}\to^P c$ with $c>0$

where $\mathbb{G}$ is a Gaussian process defined by the collection of covariances $\{\Phi(t_1,t_2)=\mathbb{E}\{\varphi_{t_1}(Y,\tilde{X})\cdot\varphi_{t_2}(Y,\tilde{X})\} : t_1,t_2\in T\}$, $\{\chi^2_{j,1}\}_{j=1}^{\infty}$ is a sequence of mutually independent chi-squared distributions each with one degree of freedom, and $\{\lambda_j\}_{j=1}^{\infty}$ is the collection of the eigenvalues of the covariance matrix operator defining the Gaussian process $\mathbb{G}$, i.e. λ_j's are the solutions of the eigenvalue problem $\lambda\cdot f(t_1)=\int\Phi(t_1,t_2)\cdot f(t_2)\;d\mu(t_2)$ for any $t_1\in T$, where φ is the eigenfunction corresponding to the eigenvalue λ.

According to Proposition (4), under the null hyptothesis H_0 the test statistic S_n will converge in distribution to a functional of a Gaussian process, while under the alternative hypothesis H_1, it will explode and converge in probability to $+\infty$. Hence, in principle, we could build up a test by comparing S_n with the quantiles of the asymptotic distribution above specified: this procedure would deliver a consistent test, in the sense that under the null the probability of rejecting H_0 would converge to the nominal size of the test, while under the alternative H_1 the probability of rejecting the null hypothesis would converge to 1. However, such a decision rule is not feasible in this context, as the quantiles of the asymptotic distribution are unknown: notice that since this distribution is charachterized by the eigenvalues λ_j's, it will also depend on the collection of functions ψ_t, with $t\in T$, and hence on unknown quantities. Accordingly, in order to build up a test, we will need to estimate the critical values. There are several methods proposed in the literature on specification testing that we could adapt to our case. For instance, Bierens & Ploberger (1997) propose to obtain the critical values of a test similar to ours, by estimating an upperbound of the quantiles of the distribution $\sum_{j=1}^{\infty} \lambda_j \chi^2_{j,1}$. Such a procedure, however, would not compare the value of the statistic S_n to an estimate of its asymptotic quantiles, but only to an upperbound for them and might turn out to be too conservative and have a low power. In a related paper, Horowitz (2006) proposed to approximate the asymptotic critical values by estimating a finite collection of the eigenvalues λ_j's: since the number of eigenvalues is potentially infinite, such a procedure requires an arbitrary cut of the sequence $\{\lambda_j\}_{j=1}^{\infty}$. Finally, Delgado & Manteiga (2001) provide a specification test similar to ours, and since their statistic converges to a distribution with a structure similar to $\sum_{j=1}^{\infty} \lambda_j \chi^2_{j,1}$, they estimate the critical values of their test by bootstrapping the statistic S_n, and propose a wild-bootstrap procedure. This is also the approach we will use in this work. In the next lines, we will hence provide all the ingredients we need in order to construct a wild-bootstrap test.

Now, in our context in order to obtain a wild-bootstrap procedure leading to a consistent test, it will not be sufficient to bootstrap only the estimator $\hat{G}$: since the framework we have at hand is featured by the non-observability of the regressors $V(\beta_0)$, we will need to bootstrap also the estimators of these covariates, and this implies that we have to bootstrap both the estimator $\hat{\beta}$ of the finite dimensional parameter, and the nonparametric estimators $\hat{m}_0$ and $\hat{m}_2$. Hence, consider the sample estimates $\{\hat{G}_{\hat{V}(\hat{\beta})}(\hat{V}_i(\hat{\beta}))\ ,\ \hat{m}_0(Z_i)\ ,\ \hat{m}_1(X_{1,i})\}_{i=1}^{n}$. Furthermore, define $\{\xi_i\}_{i=1}^{n}$ to be a sequence of weights independent from the sample data, randomly drawn in a iid fashion from a distribution $\mathbb{P}^{\xi}$. Then, let

$$D^*_{0,i} := \hat{m}(Z_i) + \xi_i \cdot (D_{0,i} - \hat{m}_0(Z_i))\ ,\ D^*_{2,i} := \hat{m}_2(X_{2,i}) + \xi_i \cdot (D_{2,i} - \hat{m}_2(X_{2,i}))$$

and

$$Y^*_i := \hat{G}_{\hat{V}(\hat{\beta})}(\hat{V}_i(\hat{\beta})) + \xi_i \cdot (Y_i - \hat{G}_{\hat{V}(\hat{\beta})}(\hat{V}_i(\hat{\beta}))) \text{ for } i = 1, .., n$$

The bootstrapped sample will be $\{Y^*_i, X_{1,i}, X_{2,i}, X^e_i, Z_i, D^*_{0,i}, D^*_{2,i}\}_{i=1}^{n}$. From this generated sample we can therefore bootstrap the nonparametric estimators entering $\hat{V}(\beta)$ as

$$\hat{m}_0^*(z) = \frac{\hat{T}_0^*(z)}{\hat{f}_0(z)} \text{ , where } \hat{T}_0(z) := \frac{1}{nh_1^{p_1}} \sum_{i=1}^{n} D_{0,i}^* \cdot K_0\Big(\frac{Z_i - z}{h_1}\Big) \text{ ,}$$

$$\hat{m}_2^*(z) = \frac{\hat{T}_2^*(z)}{\hat{f}_2(z)} \text{ , where } \hat{T}_2(z) := \frac{1}{nh_2^{p_2}} \sum_{i=1}^{n} D_{2,i}^* \cdot K_2\Big(\frac{X_{2,i} - x_2}{h_2}\Big)$$

Given the bootstrap version $(\hat{m}_0^*, \hat{m}_1^*)$ of the estimator $(\hat{m}_0, \hat{m}_1)$, we set up $\hat{V}^*(\beta) := (\beta_1 X_1 + \beta_2 \hat{m}_2^*(X_2) \text{ , } X^e - \hat{m}_0^*(Z))$, and $\hat{V}_i^*(\beta) := (\beta_1 X_{1,i} + \beta_2 \hat{m}_2^*(X_{2,i}) \text{ , } X_i^e - \hat{m}_0^*(Z_i))$, so as to get the following bootstrap counterpart of the estimator $\hat{G}_{\hat{V}(\beta)}$:

$$\hat{G}_{\hat{V}(\beta)}^*(w) := \frac{\hat{T}_{\hat{V}^*(\beta)}^*(w)}{\hat{f}_{\hat{V}^*(\beta)}(w)} \text{ , } \hat{T}_{\hat{V}^*(\beta)}^*(w) := \frac{1}{nh^d} \sum_{i=1}^{n} Y_i^* K\Big(\frac{\hat{V}_i^*(\beta) - w}{h}\Big) \text{ , } \hat{f}_{\hat{V}(\beta)}(w) := \frac{1}{nh^d} \sum_{i=1}^{n} K\Big(\frac{\hat{V}_i^*(\beta) - w}{h}\Big)$$

At this stage, we can define the bootstrap version of the estimator $\hat{\beta}$ as

$$\hat{\beta}^* = \text{argmin}_{\beta \in B} \frac{1}{n} \sum_{i=1}^{n} [Y_i^* - \hat{G}_{\hat{V}^*(\beta)}^*(\hat{V}_i^*(\beta))]^2 \cdot \hat{t}_{n,i}^{(X_2,Z)} \cdot \hat{t}_{n,i}^{(X,\hat{V})}$$

Having obtained the bootstrap counterparts of all the components entering the function $\hat{g}_{\hat{V}(\hat{\beta})}$ at the basis of our statistics S_n, we can set up

$$\hat{g}_{\hat{V}^*(\hat{\beta}^*)}^*(y, \tilde{x}) := \Big[y - \hat{G}_{\hat{V}^*(\hat{\beta}^*)}^*(\hat{V}^*(\hat{\beta}^*))\Big] \cdot \hat{f}_{\hat{V}^*(\hat{\beta}^*)}(\hat{V}^*(\hat{\beta}^*)) \cdot \hat{t}_{n,i}^{(X_1,Z)}$$

Similarly to the previous section, we introduce the bootstrap version $\mathbb{P}_n^*$ of the operator $\mathbb{P}_n$, such that $\mathbb{P}_n^* g = (1/n) \cdot \sum_{i=1}^{n} g(Y_i^*, \tilde{X}_i)$. Hence, the bootstrap counterpart of our statistic will be

$$S_n^* := \int |\mathbb{P}_n^* \hat{g}_{\hat{V}^*(\hat{\beta}^*)}^*|^2 \mu(\mathrm{d}t)$$

The estimate of the $(1-\alpha)$-quantile of the null distributon of S_n is defined as

$$\hat{c}(1-\alpha) := \inf\Big\{c : \mathbb{P}^{\xi}(S_n^* \leq c) = 1-\alpha\Big\} ,$$

and therefore the decision rule associated with the bootstrap test at the nominal level of α will lead to reject the null hypothesis H_0 as long as the statistic S_n is larger than the critical value $\hat{c}(1-\alpha)$. The following proposition shows the validity of the wild bootstrap scheme just proposed

Proposition 5. *Let Assumptions (1)-(6) hold. Then:*

(i) Under the null hypothesis H_0, $P(S_n > \hat{c}(1-\alpha)) \to \alpha$

(ii) Under the alternative H_1, $P(S_n > \hat{c}(1-\alpha)) \to 1$

In practice, in order to compute the critical values for our test, we can implement a Monte Carlo procedure. So, we can generate n_B samples of the bootstrap weigths $\{\xi_i\}_{i=1}^n$; for each bootstrap sample we can compute the statistic S_n^*, so as to get the collection $\{S_{n,b}^* : b = 1, .., n_B\}$; finally, we can take the $(1-\alpha)-$ quantile of this collection of values and consider it as an approximation of the critical value $\hat{c}(1-\alpha)$. For a very large number of bootstrap replications n_B, the approximation error will be negligible.

6 On some technical issues

In this section, we will discuss several technical issues concerning our test, and will sketch the main ideas at the basis of our proofs. This exposition will be relatively technical, so the reader interested in the application of the test and not too much in its theoretical aspects can skip this part of the paper without loss of continuity. There are essentially three issues worth discussing: the first has to do with the particular transformation of the null hypothesis we have taken, the second has to do with the procedure we have used in our proofs, while the third is the possibility to avoid the recomputation of the estimator $\hat{\beta}$ in the bootstrap procedure and therefore to speed up the computational aspects of the test.

Alternative testing approaches. When considering the null hypothesis, we have adopted the approach proposed in Bierens (1982) and Stinchcombe & White (1998): first, we have transformed the initial conditional moment restriction into a continuum of unconditional moments, and then we have constructed a test for this continuum of moments. However, such an approach is not the only one that could be employed in our framework and, as an alternative, we could have constructed a double-smoothing test by considering a different transformation of the null hypothesis. In order to analyze this alternative way, let us rewrite the null hypothesis as

$$H_0 : \mathbb{E}\{Y - G_{V(\beta_0)}(V(\beta_0)) \mid \tilde{X}\} = 0$$

Now, define $\varepsilon := Y - G_{V(\beta_0)}(V(\beta_0))$. Instead of transforming the above conditional moment restriction into a continuum of unconditional moments, we could have simply rewritten the null H_0 in the following equivalent way

$$H_0 : \mathbb{E}\{\varepsilon \cdot \mathbb{E}\{\varepsilon|\tilde{X}\} \cdot f_{\tilde{X}}(\tilde{X})\} = 0 \ ,$$

where $f_{\tilde{X}}$ is the density of the random variable $\tilde{X}$. As a second step, we could have constructed a test by proceeding in the same way as in the double smoothing approach (Lavergne & Vuong (1996), Fan & Li (1996), Delgado *et al.* (2006)). This implies that we would have taken the empirical counterpart of the above moment, and then we would have established its asymptotic distribution under H_0 by relying on U-statistic theory. From a conceptual point of view, such a way of proceeding would be simpler than the approach used in the previous section, as we would not have to deal with empirical processes and U-processes as in our context. On the other hand, the power properties of the double smoothing test would suffer from the course of dimensionality: its power will depend on the dimension of the vector $\tilde{X}$, so that the larger such a dimension the lower the power of the test. In more detail, the double-smoothing method is able to detect a sequence of local alrernatives $H_{1,n}$ converging to the null H_0 at a rate equal to $(n \cdot h_{\tilde{X}}^{dim(\tilde{X})})^{-1}$ -where $h_{\tilde{X}}$ is the bandwidth used for the estimation of the density $f_{\tilde{X}}-$, but is not able to detect a sequence of Pitman alternatives, i.e. a sequence of alternatives converging to the null at the speed of $n^{-1/2}$. The same consideration would apply to a test based on Hardle and Mammen's approach -Hardle & Mammen (1993)-, for which we could estimate the conditional expectation $\mathbb{E}\{Y|\tilde{X}\}$ and the function $G_{V(\beta_0)}$, and then construct a test statistic by measuring the distance between these two estimates. On the other hand, a Bierens' type-test like ours is able to detect a sequence of Pitman alernatives. Based on this power considerations, we have therefore preferred the approach proposed in Bierens (1982) and Stinchcombe & White (1998).

Ideas behind the proofs In order to obtain the asymptotic behavior of our statistic and the validity of the bootstrap scheme proposed, we have employed tools from Empirical Proces Theory (in particular Asymptotic Stochastic Equicontinuity), U-Process theory, U-statistic theory and some bias computations generally adopted in nonparametric kernel estimation. In this subsection, we will illustrate the basic ideas behind our proofs, and highlight how we have combined these tools to show our results. For the ease of exposition, let us consider the estimated $\hat{g}_{\hat{V}(\hat{\beta}),t}$ without the trimming. So, the empirical process $\sqrt{n}\mathbb{P}_n\hat{g}_{\hat{V}(\hat{\beta})}$ at the basis of our statistic S_n, can be decomposed as follows

$$\sqrt{n}\mathbb{P}_n\hat{g}_{\hat{V}(\hat{\beta}),t} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[Y_i - \hat{G}_{\hat{V}(\hat{\beta})}(\hat{V}_i(\hat{\beta}))\Big]\cdot\hat{f}_{\hat{V}(\hat{\beta})}(\hat{V}_i(\hat{\beta}))\cdot\phi(\tilde{X}_i t) =$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_i\cdot\hat{f}_{V(\hat{\beta})}(V_i(\beta_0))\cdot\phi(\tilde{X}_i t) + \sqrt{n}(\mathbb{P}_n - P)\Big[G_{V(\beta_0)}(V(\beta_0))\cdot\hat{f}_{\hat{V}(\hat{\beta})}(V(\beta_0)) - \hat{T}_{\hat{V}(\hat{\beta})}(V(\beta_0))\Big]\cdot\phi(\tilde{X}t) +$$

$$+P\Big[G_{V(\beta_0)}(V(\beta_0))\cdot\hat{f}_{\hat{V}(\hat{\beta})}(V(\beta_0)) - \hat{T}_{\hat{V}(\hat{\beta})}(V(\beta_0))\Big]\cdot\phi(\tilde{X}t) +$$

$$+\sqrt{n}(\hat{\beta}-\beta_0)\cdot\mathbb{P}_n\Big[Y\cdot\partial_1\hat{f}_{\hat{V}(\hat{\beta})}(\overline{V(\beta)}) - \partial_1\hat{T}_{\hat{V}(\hat{\beta})}(\overline{V(\beta)})\Big]\cdot(X_1, \overline{m}_2(X_2))\cdot\phi(\tilde{X}t) +$$

$$+\sqrt{n}\cdot\mathbb{P}_n\Big[G_{V(\beta_0)}(V(\beta_0))\cdot\partial_2\hat{f}_{\hat{V}(\hat{\beta})}(\overline{V(\beta)}) - \partial_2\hat{T}_{\hat{V}(\hat{\beta})}(\overline{V(\beta)})\Big]\cdot(\hat{m}_0(Z) - m_0(Z))\cdot\phi(\tilde{X}t)$$

where $\varepsilon := Y - G_{V(\beta_0)}(V(\beta_0))$, $\overline{V(\beta)}$ is an element between $V(\beta_0)$ and $\hat{V}(\hat{\beta})$, and we have used a mean-value expansion of $\hat{T}_{\hat{V}(\hat{\beta})}(\hat{V}(\hat{\beta}))$ and $\hat{f}_{\hat{V}(\hat{\beta})}(\hat{V}(\hat{\beta}))$ around $V(\beta_0)$. Now, the asymptotic behavior of the two addendums appearing in the second line of the above display can be analyzed by mean of asymptotic stochastic equicontinuity. In particular, by the conditions provided in Section 3, the functions $\hat{f}_{\hat{V}(\hat{\beta})}$ and $\hat{T}_{\hat{V}(\hat{\beta})}$will be sufficiently smooth, as the assumptions imply that the derivatives of these two functions are uniformly consistent. Hence, since the two addendums are centered empirical processes under the null H_0 , we can apply asymptotic stochastic equicintinuity and get rid of the estimation errors, so as to get an influence-function representation for these two addendums. Notice that in these two addendums we have no random denominator to handle. The addendum appearig in the third line of the above display can instead be analyzed by using the same tools as those used in the bias computations of nonparametric kernel estimation. For the term appearing in the fourth line, notice that, by the consistency of the kernel estimators involved, the empirical average appearing in such a term will converge to a non-random element. Hence, by replacing $\sqrt{n}(\hat{\beta} - \beta_0)$ with its influence-function representation, we will obtain also an influence-function representation of the addendum appearing in the fourth line of the above display. Finally, in order to analyze the addendum appearing in the fifth line of the above display, recall that as discussed in Section (4), the assumptions imply that botht the nonparametric kernel estimators $(\hat{m}_0, \hat{m}_2)$ and the derivatives $(\partial_2 \hat{f}_{\hat{V}(\hat{\beta})}\,, \partial_2 \hat{T}_{\hat{V}(\hat{\beta})})$ are $n^{-1/4}$-consistent for their respective targets. Therefore, we can replace the term $\left[G_{V(\beta_0)}(V(\beta_0)) \cdot \partial_2 \hat{f}_{\hat{V}(\hat{\beta})}(\overline{V(\beta)}) - \partial_2 \hat{T}_{\hat{V}(\hat{\beta})}(\overline{V(\beta)})\right]$ with its population counterpart , and obtain a U-Process of order 2 that can can be analyzed by using the U-Process theory provided in Sherman (1994) or Delgado & Manteiga (2001).

Speeding-up the bootstrap computations. In this subsection, we discuss the possibiity of speeding up the bootstrap procedure either by avoiding the recomputation of the estimator $\hat{\beta}^*$ at every bootstrap iteration, or by considering estimation methods alternative to the minimization of the semiparametric least-square principle proposed in the previous pages. Now, to better tackle the issue at hand, notice that the bootstrap procedure we have proposed in Section 4 might be relatively demanding from a computational point of view, as it requires the computation of $\hat{\beta}^*$ at every bootstrap iteration , and hence the minimization of the function defined in Eq. (9) at every bootstrap iteration. Since such a function is highly non-linear, its minimum does not have a closed form expression and might take some time to be computed, with the consequence that running the bootstrap test might involve an important amount of time. In order to reduce this computational burden, we might explore different alternatives. The first would be to consider a weighted bootstrap method similar to the one proposed in Delgado & Manteiga (2001) and Huang *et al.* (2016). Such a procedure is based on the influence function representation of the empircal process $\sqrt{n}\mathbb{P}_n\hat{g}_{\hat{V}(\hat{\beta})}$ obtained under H_0. By the results obtained in the Appendix, the influence function of such an empirical process is given by the function ψ_t defined in Section 5, so that under the null hypothesis

$$\sqrt{n}\mathbb{P}_n\hat{g}_{\hat{V}(\hat{\beta})} = \sqrt{n}\mathbb{P}_n\psi_t + o_p(1) \text{ uniformly in } t \in T$$

By following Delgado & Manteiga (2001) and Huang *et al.* (2016), the behavior of $\sqrt{n}\mathbb{P}_n\hat{g}_{\hat{V}(\hat{\beta})}$ under the null hypothesis can be mimiced by

$$\sqrt{n}\mathbb{P}_n\xi\hat{\psi}_t := \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\hat{\psi}_t(Y_i, \tilde{X}_i) \ ,$$

where $\hat{\psi}_t$ is a consistent estimator of ψ_t, while $\{\xi_i\}_{i=1}^n$ is a sequence of iid bootstrap weigths independent from the sample data. Hence, the critical values of the statistic S_n can be estimated by simulating the weights $\{\xi_i\}_{i=1}^n$, computing the element $\int|\sqrt{n}\mathbb{P}_n\xi\psi_t|^2\, \mathrm{d}\mu(t)$ for every bootstrap sample, and finally considering the $(1-\alpha)$ -quantile of such a collection of values (i.e. simulated bootstrap distribution). This procedure would enclose the advantage of avoiding the recomputation of $\hat{\beta}^*$ at every bootstrap iteration, and could hence save a relevant amount of time with respect to the method proposed in Section 4. In a supplementary material not included in this paper, we have shown the validity of such a fast weighted-bootstrap procedure. However, in a set of Monte Carlo simulations that is not reported in this work, we have observed that this method has a very poor performance in small samples, in the sense that it provides very poor approximations of the critical values of S_n. The reason is that such a fast weighted-bootstrap procedure requires the estimation of the function ψ_t and hence, from Section 5, the estimation of the nonparametric derivatives $\partial_1 G_{V(\beta_0)}$ and $\partial_2 G_{V(\beta_0)}$. Since such derivatives are poorly estimated in small samples, it is not surprising that the weighted bootstrap procedure shows a very poor performance in our context.

Another way of reducing the computational burdern and avoiding the recomputation of the estimator $\hat{\beta}^*$ at every bootstrap iteration, is to compute the critical values by an orthogonal projection of the empirical process a the basis of our statistic. This method is based on Wooldridge (). Specifically, define the process

$$\hat{A}_n^*(t) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[Y_i^* - \hat{G}^*_{\hat{V}^*(\hat{\beta})}(\hat{V}_i^*(\hat{\beta})) \right] \cdot \left[\hat{f}_{\hat{V}^*(\hat{\beta})}(\hat{V}_i^*(\hat{\beta})) \cdot \phi(\tilde{X}_i t) - \hat{a}(t)^T \hat{\Sigma}^{-1} \nabla_\beta \hat{G}_{\hat{V}(\hat{\beta})}(\hat{V}_i(\hat{\beta})) \right] ,$$

where the elements $\hat{a}(t)$ and $\hat{\Sigma}$ represents consistent estimators of $a(t)$ and Σ , respectively. In a supplementary material not included in this paper, we have shown that, under some regularity conditions, $\hat{A}_n^*$is able to mimic the null behavior of the empirical process $\sqrt{n}\mathbb{P}_n \hat{g}_{\hat{V}(\hat{\beta})}$ at the basis of our statistic S_n. Notice that the bootstrapped quantity $\hat{A}_n^*$ does not require the recomputation of the $\hat{\beta}^*$ at every bootstrap iteration. Hence, if we approximate the critical values of the statistic S_n by the quantiles of the bootstrap distribution of $\int |\hat{A}_n^*(t)|^2 \, \mathrm{d}\mu(t)$, we would save time with respect to our procedure and obtain a consitent test. In a set of Monte Carlo simulations we have assessed the properties of such a fast bootstrap procedure but, similarly to the wegithed bootstrap method discussed above, we have detected that also the boostrap test based on $\int |\hat{A}_n^*|^2 \, \mathrm{d}\mu(t)$ has poor small sample properties and is not able to approximate in a satisfactory way the critical values of S_n. The reason is essentially the same as in the case of the weighted bootstrap, and has to do with the fact that also this bootstrap scheme requires the estimation of a nonparametric derivative $\nabla_\beta \hat{G}_{\hat{V}(\hat{\beta})}$ which has a slow convergence towards its respective target.

Finally, another way to reduce the computational burden of the bootstrap scheme is to change our the estimation method for β_0, and consider instead estimators having a closed-form solution, like the Average-Derivative estimator (Ahn (1997)), or the Pairwise-Difference estimator (like in Blundell & Powell (2004)). Although such estimators have the advantage of not requiring the numerical maximization of a nonlinear objective function, the wild-bootstrap procedure for these estimators is not consistent, and hence in order to bootstrap consistently the estimators of β_0 we would have to implement a naive (or nonparametric) bootstrap scheme. This implies that also when bootsrapping our statistic S_n we would have to implement a nonparametric bootstrap. In a set of simulations not included in this paper, we have analyzed the small sample performance of the nonparametric bootstrap, and we have found that it has very poor small sample properties.

7 Conclusions

This paper has presented a bootstrap test for conditional moments specified in a semiparametric way, that involve regressors which are not observed but are identified as nonparametric conditional expectations, and hence can be estimated by using nonparametric techniques. Such a test can be employed in order to check the correct specification of setups like: semiparametric models with endogenous regressors, where the endogeneity is handled by mean of control functions; semiparametric binary-choice models with sample selection, where the selection mechanism is specified in a nonparametric fashion; truncated variable models with sample selection where the selection mechanism is again nonparametricaly specified; semiparametric games with incomplete information, where agents are supposed to play a unique Bayesian-Nash equilibrium. We have proposed a bootstrap test and proved its validity under low-level conditions. Essentially, the conditions that must be fullfilled in the empirical applications only have to do with the bandwidth rates, and are very easy to check for the researcher. One drawback of the bootsrap testing procedure proposed in this paper is that it might be requiring from a computational point of view, as we estimate the finite-dimensional parameter entering our initial moment condition by minimizing a semiparametric least-square criterion. This implies that at every bootstrap iteration we will need to minimize a nonlinear function, with the consequence that the application of the bootstrap test might involve a relevant amount of time. Such a drawback, however, is not specific to our methodology, but concerns all those tests for nonlinear models where the finite-dimensional parameter is estimated by minimizing a nonlinear cirterion function. We have also discussed several ways to circumvent this problem. Each of this alternative methods, however, shows a poor small-sample property.

We have built up our test by taking Bierens' approach, but as already discussed in this paper, we could have instead adopted the double-smoothing approach. Such a method, however, is not able to detect a sequence of alternatives converging to the null hypthesis at the rate of $n^{-1/2}$, which instead can be detected by our procedure: in other words, such a test would be less powerful with respect to ours. Nevertheless, the power drawback of a double-smoothing test could be mitigated by using the approach proposed in Lavergne *et al.* (2015), who construct a testing procedure in the double-smoothing fashion that, although is not able to detect alternatives converging to the null at the rate of $n^{-1/2}$, still shows much better power properties than the classical double-smoothing tests. Such a procedure would complement ours, as in such an approach the statistic is usually asymptotically pivotal, and hence the bootstrap would be able to provide refinements over the asymptotic approximation of the critical values, a feature that is not shared by our test. As a practical consequence, such a double-smoothing method might have better size propoerties in small samples with respect to ours, and therefore we believe it would be interesting to provide such a test in order to complement the procedure we have presented in this paper.

References

Ahn, Hyungtaik. 1997. Semiparametric Estimation of a Single-Index Model with Nonparametrically Generated Regressors. *Econometric theory*, **13**(1), 3–31.

Ahn, Hyungtaik, & Manski, Charles F. 1993. Distribution theory for the analysis of binary choice under uncertainty with nonparametric estimation of expectations. *Journal of econometrics*, **56**(3), 291–321.

Aradillas-Lopez, Andres. 2010. Semiparametric estimation of a simultaneous game with incomplete information. *Journal of econometrics*, **157**(2), 409–431.

Aradillas-Lopez, Andres. 2012. Pairwise-difference estimation of incomplete information games. *Journal of econometrics*, **168**(1), 120–140.

Bierens, Herman J. 1982. Consistent model specification tests. *Journal of econometrics*, **20**(1), 105–134.

Bierens, Herman J., & Ploberger, Werner. 1997. Asymptotic Theory of Integrated Conditional Moment Tests. *Econometrica*, **65**(5), 1129–1152.

Blundell, Richard W., & Powell, James L. 2004. Endogeneity in Semiparametric Binary Response Models. *Review of economic studies*, **71**(3), 655–679.

Ciliberto, Federico, & Tamer, Elie. 2009. Market Structure and Multiple Equilibria in Airline Markets. *Econometrica*, **77**(6), 1791–1828.

Delgado, Miguel, Dominguez, Manuel, & Lavergne, Pascal. 2006. Consistent Tests of Conditional Moment Restrictions. *Annales d'economie et de statistique*, **81**(Jan.), 33–67.

Delgado, Miguel A., & Manteiga, Wenceslao GonzÃ¡lez. 2001. Significance testing in nonparametric regression based on the bootstrap. *The annals of statistics*, **29**(5), 1469–1507.

Escanciano, Juan Carlos, Jacho-ChÃ¡vez, David T., & Lewbel, Arthur. 2014. Uniform convergence of weighted sums of non and semiparametric residuals for estimation and testing. *Journal of econometrics*, **178**(P3), 426–443.

Fan, Yanqin, & Li, Qi. 1996. Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms. *Econometrica*, **64**(4), 865–890.

Grieco, Paul L. E. 2014. Discrete games with flexible information structures: an application to local grocery markets. *The rand journal of economics*, **45**(2), 303–340.

Hardle, W., & Mammen, E. 1993. Comparing Nonparametric Versus Parametric Regression Fits. *The annals of statistics*, **21**(4), 1926–1947.

Horowitz, Joel L. 2006. Testing a Parametric Model Against a Nonparametric Alternative with Identification Through Instrumental Variables. *Econometrica*, **74**(2), 521–538.

Huang, Meng, Sun, Yixiao, & White, Halbert. 2016. A Flexible Nonparametric Test for Conditional Independence. *Econometric theory*, **32**(Dec.), 1434–1482.

Kosorok, Michael R. 2007. *Introduction to Empirical Processes and Semiparametric Inference*. Springer. Google-Books-ID: AEnIdEsfMxoC.

Lavergne, Pascal, & Vuong, Quang H. 1996. Nonparametric Selection of Regressors: The Nonnested Case. *Econometrica*, **64**(1), 207–219.

Lavergne, Pascal, Maistre, Samuel, & Patilea, Valentin. 2015. A significance test for covariates in nonparametric regression. *Electronic journal of statistics*, **9**(1), 643–678.

Lewbel, Arthur, & Tang, Xun. 2015. Identification and estimation of games with incomplete information using excluded regressors. *Journal of econometrics*, **189**(1), 229–244.

Li, Qi, & Racine, Jeffrey Scott. 2006. *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Li, Qi, & Wooldridge, Jeffrey M. 2002. Semiparametric Estimation Of Partially Linear Models For Dependent Data With Generated Regressors. *Econometric theory*, **18**(03), 625–645.

Mammen, Enno, Rothe, Christoph, & Schienle, Melanie. 2012. Nonparametric regression with nonparametrically generated covariates. *The annals of statistics*, **40**(2), 1132–1170.

Mammen, Enno, Rothe, Christoph, & Schienle, Melanie. 2016. SEMIPARAMETRIC ESTIMATION WITH GENERATED COVARIATES. *Econometric theory*, **32**(05), 1140–1177.

Newey, Whitney, & Mcfadden, Daniel. 1994. Large Sample Estimation and Hypothesis Tesing. *Handbook of econometrics*, **4**(Dec.), 2111–2245.

Newey, Whitney K., Powell, James L., & Vella, Francis. 1999. Nonparametric Estimation of Triangular Simultaneous Equations Models. *Econometrica*, **67**(3), 565–604.

Pollard, D. 1984. *Convergence of Stochastic Processes*. David Pollard. Google-Books-ID: B2vgGMa9vd4C.

Rothe, Christoph. 2009. Semiparametric estimation of binary response models with endogenous regressors. *Journal of econometrics*, **153**(1), 51–64.

Sherman, Robert P. 1994. Maximal Inequalities for Degenerate U-Processes with Applications to Optimization Estimators. *The annals of statistics*, **22**(1), 439–459.

Stinchcombe, Maxwell B., & White, Halbert. 1998. Consistent Specification Testing With Nuisance Parameters Present Only Under The Alternative. *Econometric theory*, **14**(03), 295–325.

Sweeting, Andrew. 2009. The strategic timing incentives of commercial radio stations: An empirical analysis using multiple equilibria. *Rand journal of economics*, **40**(4), 710–742.

Vaart, A. W. Van Der. 1998a. *Asymptotic Statistics.* Cambridge University Press. Google-Books-ID: udhfQgAACAAJ.

Vaart, A. W. van der. 1998b (Oct.). *Asymptotic Statistics by A. W. van der Vaart.*

Vaart, AW van der, Vaart, Aad van der, Vaart, Adrianus Willem van der, & Wellner, Jon. 1996. *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer Science & Business Media. Google-Books-ID: seH8dMrEgggC.

Appendix

Notation. Recall that $\beta := (\beta_1, \beta_2)$, $\beta_0 := (\beta_{0,1}, \beta_{0,2})$, $\hat{\beta} := (\hat{\beta}_1, \hat{\beta}_2)$, $v(\beta) := (\beta_1 x_1 + \beta_2 m_2(x_2),\ x^e - m_0(z))$, $\hat{v}(\beta) := (\beta_1 x_1 + \beta_2 \hat{m}_2(x_2),\ x^e - \hat{m}_0(z))$, $V(\beta) := (\beta_1 X_1 + \beta_2 m_2(X_2),\ X^e - m_0(Z))$, $\hat{V}(\beta) := (\beta_1 X_1 + \beta_2 \hat{m}_2(X_2),\ X^e - \hat{m}_0(Z))$. Also, $G_{V(\beta)}(w) := \mathbb{E}\{Y \mid V(\beta) = w\}$. For any $\beta \in B$, such a function will have $1 + d_V$ arguments. Denote with $\partial G^T_{V(\beta)} = (\partial_1 G_{V(\beta)}, \partial_{v^T} G_{V(\beta)})$ the partial derivative of $G_{V(\beta)}$ with respect to all its $1 + d_V$ arguments, so that $\partial_1 G_{V(\beta)}$ will denote the partial derivative of $G_{V(\beta)}$ with respect to its first argument, while $\partial_v G_{V(\beta)} = (\partial_2 G_{V(\beta)}, .., \partial_d G_{V(\beta)})^T$ will denote the vector gathering the partial derivatives of $G_{V(\beta)}$ with respect to the other d_V arguments. Notive that $\partial G_{V(\beta)}$ is measurable with respect to (the sigma field generated by) $V(\beta)$.

The conditional expectation $G_{V(\beta)}(w) := \mathbb{E}\{Y \mid V(\beta) = w\}$ can be seen as a mapping $(w, \beta) \mapsto G_{V(\beta)}(w)$. Notice that a change in β will change the shape of the function $w \mapsto G_{V(\beta)}(w)$. Denote the derivative of the mapping $(w, \beta) \mapsto G_{V(\beta)}(w)$ with respect to β as $\partial_\beta G_{V(\beta)}(\cdot)$. Notice that this function is measurable with respect to (the sigma field generated by) $V(\beta)$. On the other hand, for the mapping $(\tilde{x}, \beta) \mapsto G_{V(\beta)}(v(\beta))$, a change in β will impact on both the shape of the function $G_{V(\beta)}(\cdot)$ and the argument $v(\beta)$ where such a function is computed. The differentiation of the maping $(\tilde{x}, \beta) \mapsto G_{V(\beta)}(v(\beta))$ with respect to both occurrences of β will deliver $\partial_\beta G_{V(\beta)}(v(\beta)) + \partial_1 G_{V(\beta)}(v(\beta)) \cdot (x_1^T, m_2(x_2)^T)^T$. Let us denote such a derivative with $\nabla_\beta G_{V(\beta)}(x_1, m_2(x_2), v(\beta))$.

Observe that $\tilde{x} \mapsto \nabla_\beta G_{V(\beta)}(x_1, m_2(x_2), v(\beta))$ is *not* measurable with respect to the sigma field generated by $V(\beta)$.

.

Having specified the notation, we start by presenting three lemmas that will be used througout this appendix. Let $\mathcal{F}$ be a space of real-valued functions defined over $\tilde{\mathcal{X}}$ and metricized by $L_2(P)$, i.e. for any $f, g \in \mathcal{F}$, the distance between f and g is measured by the (pseudo) norm $||f-g||_{L_2(P)} = \{\int |f-g|^2(\tilde{x})\, \mathrm{d}P(\tilde{x})\}^{1/2}$. For any two functions $\tilde{x} \mapsto u(\tilde{x})$ and $\tilde{x} \mapsto l(\tilde{x})$, such that $u \geq l$, define the bracket $[u, l] := \{f \in \mathcal{F} : l \leq f \leq u\}$. We say that the bracket $[u, l]$ has $L_2(P)$- size ϵ if $||u-l||_{L_2(P)} \leq \epsilon$. Hence, define $N_{[]}(\varepsilon, \mathcal{F}, L_2(P))$ to be the bracketing number of the space $(\mathcal{F}, L_2(P))$, i.e. $N_{[]}(\varepsilon, \mathcal{F}, L_2(P))$ represents the minimum number of brackets of $L_2(P)$ size ε that covers the space $\mathcal{F}$. Also, let $J_{[]}(\delta, \mathcal{F}, L_2(P)) := \int_0^\delta \sqrt{\log\, N_{[]}(\varepsilon, \mathcal{F}, L_2(P))}\mathrm{d}\varepsilon$. For a deeper treatment of these concepts, see Kosorok (2007), Pollard (1984), Vaart *et al.* (1996), and Vaart (1998b). The three lemmas that follows are the main ones used in our proofs.

Lemma 6. *(Vaart (1998b)) Let $\mathcal{F}$ be a class of leasurable functions $f : \chi \mapsto \Re$ such that $Pf^2 < \delta^2$ for all $f \in \mathcal{F}$, let F be the envelope function for $\mathcal{F}$, and let $a_{\mathcal{F}}(\delta) := \delta/\sqrt{1 \vee log\ N_{[]}(\delta, \mathcal{F}, L_2(P))}$. Then*

$$\mathbb{E}||\mathbb{G}_n||_{\mathcal{F}} \leq J_{[]}(\delta, \mathcal{F}, L_2(P)) + \sqrt{n}PF\{F > \sqrt{n}a_{\mathcal{F}}(\delta)\}$$

up to a universal constant.

Let k be a positive integer and $\mathcal{F}$ be a class of real-valued functions defined on $\mathcal{S}^k := \mathcal{S}\text{x..x}\mathcal{S}$ (k times). Define $U_n^k$to be the operator such that for all $f \in \mathcal{F}$, $U_n^k f = (n)_k^{-1} \sum_{\tilde{i}_k} f(Z_{i_1}, .., Z_{i_k})$, where $(n)_k := n \cdot (n-1) \cdot .. \cdot (n-k+1)$, and the index $\tilde{i}_k = (i_1, .., i_k)$ ranges over all the permutations of class k of the set $\{1, .., n\}$. Let $N(\varepsilon, \mathcal{F}, ||\cdot||)$ be the covering number of size ε of the class $\mathcal{F}$, i.e. $N(\varepsilon, \mathcal{F}, ||\cdot||)$ is the smallest number of balls of $||\cdot||$- dimension ε that covers the class $\mathcal{F}$ (see Vaart (1998a) or Kosorok (2007)). We say that the class $\mathcal{F}$ is P-degenerate, if $Pf(s_1, .., s_{q-1}, \cdot, s_{q+1}, .., s_k) = 0$ for $q = 1, .., k$. The following lemma is readapted from Sherman (1994).

Lemma 7. *Let $\mathcal{F}$ be a P-degenerate class of functions on $\mathcal{S}^k$, such that for all $f \in \mathcal{F} : |f| \leq \tilde{c}$. Also, assume that $N(\varepsilon, \mathcal{F}, ||\cdot||_\infty) \leq A \cdot \varepsilon^{-d}$. Then,*

$$P\ sup_{f\in\mathcal{F}}|n^{k/2}U_n^k f|^2 \leq \Gamma \cdot \tilde{c}^2$$

where Γ is a fixed constant that depends on k , A, d, and $\tilde{c}$ only.

Let $\mathcal{G}$ be a class of real-valued functions defined on $\mathcal{S}^m := \mathcal{S}\text{x..x}\mathcal{S}$ (m times), and let π_k (with $k \leq m$) be the operator such that $\pi_k g = (\delta_{Z_1} - P)\text{x..x}(\delta_{Z_k} - P)\text{x}P^{m-k}\ g$. In words, the operator π_k applied to a function g takes the Hoeffding decomposition of order k. The following lemma is taken from Delgado and Gonzalo-Manteiga (2001):

Lemma 8. *Let $\mathcal{G}$ be a VC-class of functions on $\mathcal{S}^m$, and let G be an envelope for such a class, with $P^m G^2 < \infty$. Then, there is a constant $\tilde{c}$ that depends only on the VC charachteristics of $\mathcal{G}$such that for al $k = 1, .., m$*

$$\mathbb{E}sup_{g\in\mathcal{G}}\left|n^{k/2}U_n^k(\pi_k g)\right|^{\mathbf{2}} \leq \tilde{c} \cdot P^m G^2$$

.

Lemma 9. *(i) For any class of functions $\mathcal{F}$, $N_{[]}(\varepsilon, \mathcal{F}, L_2(P)) \leq N(\varepsilon/2, \mathcal{F}, ||\cdot||_\infty)$; (ii) If $\mathcal{G} = \{g_t : t \in T\}$ is a class such that $||g_{t_1} - g_{t_2}||_\infty \leq C|t_1 - t_2|$ for any t_1, t_2 in T, then $N(\varepsilon, \mathcal{G}, ||\cdot||_\infty) \leq N(C\varepsilon, T, |\cdot|)$; (iii) $N(\varepsilon, T, |\cdot|) \leq C\varepsilon^{-dim(T)}$; (iv) let $\Phi := \{\tilde{x} \mapsto \phi(t \cdot \tilde{x}) : t \in T\}$ with ϕ analytic non-polynomial function defined over $\Re$, T and $\tilde{\mathcal{X}}$ compact sets, and $logN(\varepsilon, \mathcal{C}, ||\cdot||_\infty) \leq C \cdot \varepsilon^{-\upsilon}$; then the class $\mathcal{C} \cdot \Phi$ is such that $logN(\varepsilon, \mathcal{C}\cdot\Phi, ||\cdot||_\infty) \leq C\varepsilon^{-\upsilon}$, with $\upsilon \in (0, 2)$.*

Proof. To show (i)-(iii), one can proceed in the same way as in Example 19.34 in Vaart (1998b). For point (iv), notice that by the compactness of T and $\tilde{\mathcal{X}}$, Φ is a class satisfying point (ii), so that by point (ii)-(iii) $\log N(\varepsilon, \Phi, ||\cdot||_\infty) \leq \log\varepsilon^{-dim(T)}$. Furthermore, $\log N(\varepsilon, \mathcal{C}, ||\cdot||_\infty) \leq C\varepsilon^{-d}$. Consider now the ε-covers $\mathcal{A}_\mathcal{C} := \{g_I : I = 1, .., N(\varepsilon, \mathcal{C}, ||\cdot||_\infty)\}$ and $\mathcal{A}_\Phi := \{\phi_J : J = 1, .., N(\varepsilon, \Phi, ||\cdot||_\infty)\}$. For the generic element of $\mathcal{C}\cdot\Phi$, say $g \cdot \phi$, with $g \in \mathcal{C}$ and $\phi \in \Phi$, we have $||g - g_I||_\infty < \varepsilon$ and $||\phi - \phi_J||_\infty < \varepsilon$, for some $g_I \in \mathcal{A}_\mathcal{C}$ and $\phi_J \in \mathcal{A}_\Phi$. Hence, $||g\cdot\phi - g_I\cdot\phi_J||_\infty \leq ||(g - g_I)\cdot\phi||_\infty +$

$||g_I \cdot (\phi - \phi_J)||_\infty < C\varepsilon$. So, the collection $\{g_I \cdot \phi_J : g_I \in \mathcal{A}_{\mathcal{C}},\, \phi_J \in \mathcal{A}_\Phi\}$ forms a $C\varepsilon$ cover for the class $\Phi \cdot \mathcal{C}$, and hence $N(C\varepsilon, \Phi{\cdot}\mathcal{C}, ||\cdot||_\infty) \leq \#\mathcal{A}_{\mathcal{C}} \cdot \#\mathcal{A}_\Phi \leq N(\varepsilon, \mathcal{C}, ||\cdot||_\infty) \cdot N(\varepsilon, \Phi, ||\cdot||_\infty)$. □

We follow Vaart *et al.* (1996) and introduce a class of functions with uniformly bounded derivatives that will be used throughout the proofs. For any vector $k = (k_1, .., k_s)$ with s components, denote the differential operator

$$D^k := \frac{\partial^{|k|}}{\partial^{k_1} x_1 .. \partial^{k_s} x_s}$$

where $|k| := k_1 + .. + k_s$. For any function $g : \overline{\mathcal{X}} \mapsto \Re$ that is $|k|$ times continuously differentiable over $\overline{\mathcal{X}}$, define

$$||g||_{\eta,\overline{\mathcal{X}}} := \max_{|k| \leq \eta} \sup_{\overline{x} \in \overline{\mathcal{X}}} |D^k g(\overline{x})| \;, \tag{10}$$

where $\eta \in \mathbb{N}$. The class we will consider in several proofs is defined as

$$C_M^\eta(\overline{\mathcal{X}}) := \left\{ g : \; ||g||_{\eta,\overline{\mathcal{X}}} \leq M \right\} \tag{11}$$

Since M is only a constant we will often drop it, and hence use $C^\eta(\overline{\mathcal{X}})$ in place of $C_M^\eta(\overline{\mathcal{X}})$. Such a class of smooth functions has the following property (see Vaart *et al.* (1996))

$$\log N(\varepsilon, C^\eta(\overline{\mathcal{X}}), ||\cdot||_\infty) \leq C \cdot \varepsilon^{-dim(\overline{X})/\eta}. \tag{12}$$

The next proposition is at the basis of the proofs for the asymptotic behavior of $\sqrt{n}\mathbb{P}_n \hat{g}_{\hat{V}(\hat{\beta})}$. It is based on a repeated application of the above lemmas. Since it is relatively long, it is reported in a supplementary material

Lemma 10. *Let Assumptions (1)-5(i) hold. Then :*

(i)

$$\frac{n^{1/4}}{\tau_n^2}\ sup_{\beta\in B}||\hat{T}_{\hat{V}(\beta)} - T_{V(\beta)}||_\infty = o_p(1)\ ,$$

and the same holds for $\partial_1 \hat{T}_{\hat{V}(\beta)}$, $\partial_v \hat{T}_{\hat{V}(\beta)}$, $\partial_\beta \hat{T}_{\hat{V}(\beta)}$, $\hat{f}_{\hat{V}(\beta)}$, $\partial_v \hat{f}_{\hat{V}(\beta)}$, *and* $\partial_\beta \hat{f}_{\hat{V}(\beta)}$.

Furthermore, $P(\hat{T}_{\hat{V}(\beta)}, \partial_1 \hat{T}_{\hat{V}(\beta)}, \partial_v \hat{T}_{\hat{V}(\beta)}, \partial_\beta \hat{T}_{\hat{V}(\beta)} \in \mathcal{C}\ \forall \beta \in B) \to 1$, *and the same holds for* $\hat{f}_{\hat{V}(\beta)}$.

(ii) $\tau_n^4 \cdot sup_{\beta\in B}||\nabla^2_{\beta\beta^T} \hat{T}_{\hat{V}(\beta)}(\beta,\cdot) - \nabla^2_{\beta\beta^T} T_{V(\beta)}(\beta,\cdot)||_\infty = o_P(1)$, *and the same holds for* $\hat{f}_{\hat{V}(\beta)}$.

Remark 11. The previous Lemma allows us to write that $\sup_{\beta\in B}||\hat{T}_{\hat{V}(\beta)} - T_{V(\beta)}||_\infty = O_P(d_n)$, with $d_n = \epsilon_n n^{-1/4}\tau_n^2$ where ϵ_n is a sequence of numbers converging to zero. The same holds for $\nabla \hat{T}_{\hat{V}(\beta)}$, $\hat{f}_{\hat{V}(\beta)}$, and $\nabla \hat{f}_{\hat{V}(\beta)}$

We next analyze the uniform convergence of $\hat{m}_0$ and $\hat{m}_2$ towards their respective population counterparts, m_0 and m_2. We will make use of the decomposition

$$\hat{m}_0 := \frac{\hat{T}_0}{\hat{f}_0} = \frac{\hat{T}_0 - m_0\cdot \hat{f}_0}{f_0} + \frac{\hat{T}_0 - m_0\cdot \hat{f}_0}{f_0}\cdot \frac{f_0 - \hat{f}_0}{\hat{f}_0} \tag{13}$$

The above decomposition also holds for $\hat{m}_2$.

Lemma 12. *Let Assumption (1)-4(ii) hold. Define* $d_{n,0} = \sqrt{(logn)/(nh_0^{p_0})} + h_0^{r_0}$. *Then,* $||\hat{f}_0 - f_0||_\infty = O_P(d_{n,0})$, $||\hat{T}_0 - T_0||_\infty = O_P(d_{n,0})$, $\left|\left|(\hat{m}_0 - m_0)\cdot t_n^{(Z)}\right|\right|_\infty = O_P(d_{n,0}/\tau_n)$, $||((\hat{T}_0 - m_0\hat{f}_0)\cdot t_n^{(Z)}\cdot(\hat{f}_0 - f_0)/(\hat{f}_0\cdot f_0)||_\infty = O_P(d_{n,0}^2/\tau_n^2)$, *and* $||(\hat{f}_0 - f_0)\cdot t_n^{(Z)}/\hat{f}_0||_\infty = O_P(d_{n,0}/\tau_n^2)$, *with* $d_{n,0}/\tau_n^3 = o(n^{-1/4})$. *The same type of result holds by replacing* t_n *with* $\hat{t}_n$ *and for* $\hat{m}_2$.

Proof. For the ease of notation, drop the index 0 from $\hat{T}_0$, $\hat{f}_0$, and $d_{n,0}$, so that consider $\hat{T}$, $\hat{f}$, and d_n. From Li & Racine (2006), $||\hat{f} - f||_\infty = O_P(d_n)$ and $||\hat{T} - T||_\infty = O_P(d_n)$. Define the event $B_n^{(C)} := \{||\hat{f} - f||_\infty d_n^{-1} < C,\ ||\hat{T} - T||_\infty d_n^{-1} < C\}$. Notice that by choosing C large enough, $P(B_n^{(C)})$ can be made arbitrarily close to one for any n sufficiently high. Over the set $B_n^{(C)}$ and if $t_n(z) = 1$, it must be that $\hat{f}(z) \geq (f(z)/\tau_n - Cd_n/\tau_n)\cdot\tau_n \geq \tau_n/2$ for any large n, so that $|(\hat{T} - m\hat{f})/f\ |(z) \leq Cd_n/\tau_n$ and $|(\hat{f} - f)/\hat{f}|\ (z) \leq Cd_n/\tau_n$. The same reasoning will hold with t_n replaced by $\hat{t}_n$, mutatis mutandis. Conclude from Assumption 4(ii) and the above decomposition for $\hat{m}_0 - m_0$. ☐

Lemma 13. *Let Assumption (1)-(5) hold. Define* $s := (s_1, .., s_{p_0})$, $|s| := s_1 + .. + s_{p_0}$, *and* $d_{n,0}^{(|s|)} = \sqrt{(logn)/(nh_0^{p_0+2|s|})} + h_0^{r_0}$. *Then,* $||D^s\hat{f}_0 - D^s f_0||_\infty = O_P(d_{n,0}^{(|s|)})$, $||D^s\hat{T}_0 - D^sT_0||_\infty = O_P(d_{n,0}^{(|s|)})$, $||(D^s\hat{f}_0 - D^s f_0)\cdot t_n^{(Z)}/f_0||_\infty = O_P(d_{n,0}^{(|s|)}/\tau_n)$, $||(D^s\hat{T}_0 - D^sT_0)\cdot t_n^{(Z)}/f_0||_\infty = O_P(d_{n,0}^{(|s|)}/\tau_n)$, $||(D^s\hat{f}_0 - D^s f_0)\cdot t_n^{(Z)}/\hat{f}_0||_\infty = O_P(d_{n,0}^{(|s|)}/\tau_n)$, $||(D^s\hat{T}_0 - D^sT_0)\cdot t_n^{(Z)}/\hat{f}_0||_\infty = O_P(d_{n,0}^{(|s|)}/\tau_n)$. *The same result also holds for* $\hat{T}_2$ *and* $\hat{f}_2$.

Proof. The results $||D^s\hat{f}_0 - D^s f_0||_\infty = O_P(d_{n,0}^{(|s|)})$, and $||D^s\hat{T}_0 - D^sT_0||_\infty = O_P(d_{n,0}^{(|s|)})$ can be derived by proceeding along the same lines as in , Li & Racine (2006). In order to obtain the remaining results, we can then proceed in the same way as in the proof of Lemma (12). ☐

Lemma 14. *Let Assumptions (1)-(5) hold. Then,* $||(D^s\hat{m}_0 - D^s m_0)t_n^{(Z)}||_\infty = o_p(1)$ *for all* $s := (s_1, .., s_{p_0})$ *such that* $|s| \leq 1 + \underline{Ev}(p_0)/2$. *The same result holds for* $\hat{m}_2$.

Proof. For the ease of notation, let us drop the index 0 from m_0 and $\hat{m}_0$, and hence consider at their place m and $\hat{m}$, respectively. Also, assume wlog that these functions are real-valued and $dim(z) = 1$. Define $d_n := \sqrt{\frac{\log n}{n\cdot h_0^{p_0}}} + h_0^{r_0}$ and $d_n^{(s)} := \sqrt{\frac{\log n}{n\cdot h_0^{p_0+2s}}} + h_0^{r_0}$. Notice that, $D^1 m(z) = \frac{T^{(1)}}{f}(z) - m(z)\cdot\frac{f^{(1)}}{f}(z)$ and $D^1\hat{m}(z) = \frac{\hat{T}^{(1)}}{\hat{f}}(z) - \hat{m}(z)\cdot\frac{\hat{f}^{(1)}}{\hat{f}}(z)$. From these expressions, Lemma (12) and Lemma (13), we have $||(D^1\hat{m} - D^1 m)\cdot t_n^{(Z)}||_\infty = O_P(d_n^{(1)}/\tau_n + d_n/\tau_n^2\Big) = O_P(d_n^{(1)}/\tau_n^2)$. We now proceed by induction. Consider s such that $2 \leq s \leq 1+\underline{Ev}(p_0)$, and assume that for $j = 1, .., s-1$ $||(D^j\hat{m} - D^j m)\cdot t_n||_\infty = O_P(d_n^{(j)}/\tau_n^{j+1})$. Notice that by Assumption 5, $d_n^{(j)}/\tau_n^{j+1} = o(1)$ for all $j = 1, .., s$. By the Liebniz rule for derivation,

$$D^s m = \frac{D^sT}{f} - \sum_{j=0}^{s-1}\binom{s}{j}\frac{D^{s-j}f}{f}\cdot D^j m \text{ and } D^s\hat{m} = \frac{D^s\hat{T}}{\hat{f}} - \sum_{j=0}^{s-1}\binom{s}{j}\frac{D^{s-j}\hat{f}}{\hat{f}}\cdot D^j\hat{m}$$

Now, by Lemma (12) and (13),

$$\Big|\Big|\Big(\frac{D^s\hat{T}}{\hat{f}} - \frac{D^sT}{f}\Big)\cdot t_n^{(Z)}\Big|\Big|_\infty = O_P\Big(\frac{d_n^{(s)}}{\tau_n} + \frac{d_n}{\tau_n^2}\Big) = O_P\Big(\frac{d_n^{(s)}}{\tau_n^2}\Big) \text{ and}$$

$$\Big|\Big|\Big(\frac{D^{s-j}\hat{f}}{\hat{f}} - \frac{D^{s-j}f}{f}\Big)\cdot t_n^{(Z)}\Big|\Big|_\infty = O_P\Big(\frac{d_n^{(s-j)}}{\tau_n} + \frac{d_n}{\tau_n^2}\Big) = O_P\Big(\frac{d_n^{(s-j)}}{\tau_n^2}\Big) .$$

So, by the induction assumption and the previous two displays,

$$||(D^s\hat{m} - D^s m)\cdot t_n^{(Z)}||_\infty = O_P\Big(\frac{d_n^{(s)}}{\tau_n^2}\Big) - \sum_{j=0}^{s-1} O_P\Big(\frac{d_n^{(s-j)}}{\tau_n^2} + \frac{1}{\tau_n}\cdot\frac{d_n^{(j)}}{\tau_n^{j+1}}\Big) \leq$$

$$\leq O_P\Big(\frac{d_n^{(s)}}{\tau_n^2}\Big) + O_P\Big(\frac{d_n^{(s)}}{\tau_n^2} + \frac{d_n^{(s-1)}}{\tau_n^{s+1}}\Big) \leq O_P\Big(\frac{d_n^{(s)}}{\tau_{n^{s+1}}}\Big).$$

Hence, the induction assumption implies that $||(D^s\hat{m} - D^s m)\cdot t_n^{(Z)}||_\infty = O_P(d_n^{(s)}/\tau_n^{s+1})$. Since as shown above the induction assumption holds for $s = 1$, i.e. $||(D^1\hat{m} - D^1 m)\cdot t_n^{(Z)}||_\infty = O_P(d_n^{(1)}/\tau_n^2)$, we conclude that for all $s \geq 1$, $||(D^s\hat{m} - D^s m)\cdot t_n^{(Z)}||_\infty = O_P(d_n^{(s)}/\tau_n^{s+1})$ holds true. By assumption, $d_n^{(s)}/\tau_n^{s+1} = o(1)$, so we conclude. □

Lemma 15. *Define $s_0 := 1 + \frac{\underline{Ev}(p_0)}{2}$ and let Assumption (1)-(5) hold. Then, $P(\hat{m}_0 \in C^{s_0}(\mathcal{Z}_n)) \to 1$, with $\mathcal{Z}_n := \{z : f_0(z) \geq \tau_n\}$ and $p_0/s_0 \in (0, 2)$. The same result also holds for $\hat{m}_2$, mutatis mutandis.*

Proof. Assume wlog that $\hat{m}_0$ and m_0 are unidimensional. From Lemma (14), $||(D^j\hat{m}_0 - D^j m_0)\cdot t_n^{(Z)}||_\infty = o_P(1)$ for all $j = 1, .., s_0$. By definition of the class $C^{s_0}(\mathcal{Z}_n)$, since m_0 has bounded derivatives up to order s_0, we have $P(\hat{m}_0 \in C^{s_0}(\mathcal{Z}_n)) \to 1$. Now, by definition of s_0, $p_0/s_0 < 2$ is equivalent to $p_0 - \underline{Ev}(p_0) < 2$, which always holds true. □

Lemma 16. *Let Assumption (1)-(4) and Assumption (5i) hold.*

(i) Let $\{A_{n,i,t},\ t \in T\}_{i=1}^n$ be a sequence of stochastic processes, such that $sup_{i=1,..,n,\ t\in T}|A_{n,i,t}\cdot \hat{t}_{n,i}^{(X_2,Z)}| = O_P(1)$ and $sup_{i=1,..,n,\ t\in T}|A_{n,i,t}\cdot t_{n,i}^{(X_2,Z)}| = O_P(1)$; then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n A_{n,i,t}\cdot \hat{t}_{n,i}^{(X_2,Z)} = \frac{1}{\sqrt{n}}\sum_{i=1}^n A_{n,i,t}\cdot t_{n,i}^{(X_2,Z)} + o_P(1) \text{ uniformly in } t \in T$$

(ii) Let $\{B_{n,t,i} : t \in T\}_{i=1}^n$ be a sequence of stochastic processes such that $sup_{1\leq i\leq n, t\in T}|B_{n,i,t}\hat{t}_{n,i}^{(X_2,Z)}\hat{t}_{n,i}^{(X,\hat{V})}| = O_P(1)$, $sup_{1\leq i\leq n, t\in T}|B_{n,i,t}\hat{t}_{n,i}^{(X_2,Z)}t_{n,i}^{(X,V)}| = O_P(1)$, $sup_{1\leq i\leq n, t\in T}|B_{n,i,t}t_{n,i}^{(X_2,Z)}t_{n,i}^{(X,V)}| = O_P(1)$. If As-

sumption (1)-(5) hold, then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} B_{i,n,t}\hat{t}_{n,i}^{(X_2,Z)}\hat{t}_{n,i}^{(X,\hat{V})} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} B_{i,n,t}t_{n,i}^{(X_2,Z)}t_{n,i}^{(X,V)} + o_P(1) \text{ uniformly in } t\in T$$

(iii) If $\{D_{i,n,t} : t \in T\}_{i=1}^{n}$ *is a sequence of processes such that* $sup_{1\le i\le n, t\in T}|D_{i,n,t}| = O_P(1)$*, then*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} D_{i,n,t}t_{n,i} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} D_{i,n,t} + o_P(1) \text{ uniformly in } t\in T \text{ , for } t_n \in \{t_n^{(Z)}, t_n^{(X_2)}, t_n^{(X_2,Z)}\}$$

Proof. (i) Since $\hat{t}_n^{(X_2,Z)}$, $t_n^{(X_2,Z)} \in \{0,1\}$, $(\hat{t}_n^{(X_2,Z)} - t_n^{(X_2,Z)}) = (\hat{t}_n^{(X_2,Z)} - t_n^{(X_2,Z)})(\hat{t}_n^{(X_2,Z)} + t_n^{(X_2,Z)})$, so $(1/\sqrt{n})\sum_{i=1}^{n}|A_{n,i,t}\cdot(\hat{t}_{n,i}^{(X_2,Z)} - t_{n,i}^{(X_2,Z)})| = (1/\sqrt{n})\sum_{i=1}^{n}|A_{n,i,t}\hat{t}_{n,i}^{(X_2,Z)} + A_{n,i,t}t_{n,i}^{(X_2,Z)}|\cdot|\hat{t}_{n,i}^{(X_2,Z)} - t_{n,i}^{(X_2,Z)}| = O_P((1/\sqrt{n})\cdot\sum_{i=1}^{n}|\hat{t}_{n,i}^{(X_2,Z)} - t_{n,i}^{(X_2,Z)}|\ )$. Now, consider $(1/\sqrt{n})\cdot\sum_{i=1}^{n}|\hat{t}_{n,i}^{(Z)} - t_{n,i}^{(Z)}|$ and define $B_n^{(C)} := \{||\hat{f}_0 - f_0||_\infty d_{n,0}^{-1} < C\}$, with $d_{n,0} = \sqrt{(\log n)/(nh_0^p)} + h_0^{r_0}$. Since, $||\hat{f}_0 - f_0||_\infty = O_P(d_{n,0})$ (see Lemma (12)), for C large enough, $P(B_n^{(C)})$ can be made arbitrarily close to one for any high n. Notice if $B_n^{(C)}$ holds and $f_0(z) > 3\tau_n/2$, then $t_n^{(Z)}(z) = 1$, and $\hat{f}_0(z) > (f_0(z)/\tau_n - Cd_n/\tau_n)\cdot\tau_n > \tau_n$ for any large enough n, so that $\hat{t}_n^{(Z)}(z) = 1$. So, over the set $B_n^{(C)}$ it must be that $|\hat{t}_n^{(Z)} - t_n^{(Z)}|(z) \le 1\{f_0(z) \le 3\tau_n/2\}$. By Markov's inequality and Assumption 4 (iii), we have $\sqrt{n}\mathbb{P}_n 1\{f_0(Z) \le 3\tau_n/2\} = o_P(1)$, and hence $\sqrt{n}\mathbb{P}_n|\hat{t}_n^{(Z)} - t_n^{(Z)}| = o_P(1)$. By the same reasoning, $\sqrt{n}\mathbb{P}_n|\hat{t}_n^{(X_2)} - t_n^{(X_2)}| = o_P(1)$. So, $(1/\sqrt{n})\cdot\sum_{i=1}^{n}|\hat{t}_{n,i}^{(X_2,Z)} - t_{n,i}^{(X_2,Z)}| = o_P(1)$, and we conclude.

(ii) Define $d_n" := \sqrt{(\log n)/(n\cdot h_3^q)} + h_3^{r_3}$, $d_{n,0} := \sqrt{(\log n)/(n\cdot h_0^{p_0})} + h_0^{r_0}$, $\mathcal{M}_0 := \{m_0(z) : z \in \mathcal{Z}\}$, its δ-enlargment $\mathcal{M}_0^\delta := \{u : |u - u'| < \delta \text{ with } u' \in \mathcal{M}_0\}$, and the event $\mathcal{A}_n^{(C)} := \{||\hat{f}_{X,\hat{V}} - f_{X,V}||_\infty \le Cd_n"\ , ||\hat{t}_n^{(Z)}(\hat{m}_0 - m_0)||_\infty \le Cd_{n,0}/\tau_n\}$. By Assumption (4) and (5), we can consider n large enough so that $C\cdot(d_n"/\tau_n + d_{n,0}/\tau_n^2) < 1/2$ and $Cd_{n,0}/\tau_n < \delta$; if $\mathcal{A}_n^{(C)}$ holds, $\hat{t}_n^{(Z)}(z) = 1$, and $1\{f_{X,V}(x,v) \ge 3\tau_n/2\} = 1$, then: (i) $\hat{m}_0(z) \in \mathcal{M}_0^\delta$, and by a Mean-Value expansion $|\hat{f}_{X,\hat{V}}(x,\hat{v}) - f_{X,V}(x,v)| \le |\hat{f}_{X,\hat{V}}(x,\hat{v}) - f_{X,V}(x,\hat{v})| + C|\hat{m}_0 - m_0|(z) \le C\cdot(d_n' + d_{n,0}/\tau_n)$; so, (ii) $\hat{f}_{X,\hat{V}}(x,\hat{v}) \ge [f_{X,V}(x,v)/\tau_n - C\cdot(d_n"/\tau_n + d_{n,0}/\tau_n^2)]\cdot\tau_n \ge \tau_n$, and hence (iii) $\hat{t}_n^{(X,\hat{V})}(\tilde{x}) = 1$ and $t_n^{(X,V)}(\tilde{x}) = 1$. This reasoning implies that for n large enough such that $C\cdot(d_n"/\tau_n + d_{n,0}/\tau_n^2) < 1/2$ and $Cd_{n,0}/\tau_n < \delta$, and if $\mathcal{A}_n^{(C)}$ holds, then $\hat{t}_n^{(Z)}(z)\cdot|\hat{t}_n^{(X,\hat{V})}(\tilde{x}) - t_n^{(X,V)}(\tilde{x})| \le 1\{f_{X,V}(x,v) < 3\tau_n/2\}$, and hence $(1/\sqrt{n})\sum_{i=1}^{n}\hat{t}_{n,i}^{(Z)}|\hat{t}_{n,i}^{(X,\hat{V})} - t_{n,i}^{(X,V)}| \le (1/\sqrt{n})\sum_{i=1}^{n}1\{f_{X,V}(X_i,V_i) < 3\tau_n/2\}$. Since by selecting C large enough, $P(\mathcal{A}_n^{(C)})$can be made arbitrarily close to 1 for any large

n, from Assumption 5(iii) and Markov's inequality we obtain $(1/\sqrt{n})\sum_{i=1}^{n}\hat{t}_{n,i}^{(Z)}|\hat{t}_{n,i}^{(X,\hat{V})} - t_{n,i}^{(X,V)}| = o_P(1)$. Hence, $|(1/\sqrt{n})\sum_{i=1}^{n} B_{n,i,t}\cdot\hat{t}_{n,i}^{(Z)}\cdot(\hat{t}_{n,i}^{(X,\hat{V})} - t_{n,i}^{(X,V)})| \leq \sup_{1\leq i\leq n, t\in T}|B_{n,i,t}(\hat{t}_{n,i}^{(X,\hat{V})} + t_{n,i}^{(X,V)})\hat{t}_{n,i}^{(Z)}|\cdot (1/\sqrt{n})\cdot\sum_{i=1}^{n}|\hat{t}_{n,i}^{(X,\hat{V})} - t_{n,i}^{(X,V)}|\cdot\hat{t}_{n,i}^{(Z)} = o_P(1)$. Finally, $|(1/\sqrt{n})\sum_{i=1}^{n} B_{n,i,t}t_{n,i}^{(X,V)}\cdot(\hat{t}_{n,i}^{(Z)} - t_{n,i}^{(Z)})| \leq \sup_{i=1,..,n,t\in T}|B_{n,i,t}\cdot t_{n,i}^{(X,V)}\cdot(\hat{t}_{n,i}^{(Z)} + t_{n,i}^{(Z)})|\cdot(1/\sqrt{n})\sum_{i=1}^{n}|\hat{t}_{n,i}^{(Z)} - t_{n,i}^{(Z)}| = o_P(1)$, where the last equality follows from point (i) and by Assumption.

(iii) $(1/\sqrt{n})\sum_{i=1}^{n}|D_{n,i,t}|\cdot|t_{n,i}^{(Z)} - 1| \leq O_P(1)(1/\sqrt{n})\sum_{i=1}^{n}|t_{n,i}^{(Z)} - 1| \leq O_P(1)(1/\sqrt{n})\sum_{i=1}^{n}1\{f(Z_i) < \tau_n\}$. Conclude by Markov's inequality and Assumption 4(ii). □

Lemma 17. *Let* $d_n := \epsilon_n\tau_n n^{-1/4}$, $d_n' =: d_{n,0}\vee d_{n,2}$, $d_{n,0} := \sqrt{(logn)/(n\cdot h_0^{p_0})} + h_0^{r_0}$, $d_{n,2} := \sqrt{(logn)/(n\cdot h_2^{p_2})} + h_2^{r_2}$, *and* $d_n" = \sqrt{\frac{logn}{nh_3^q}} + h_3^{r_3}$. *If Assumption (1)-(4) and Assumption 5(i) hold:*

(i)$\sup_{\beta\in B}||(\hat{f}_{\hat{V}(\beta)}(\hat{v}(\beta) - f_{V(\beta)}(v(\beta))\cdot\hat{t}_n^{(X_2,Z)}(\tilde{x})||_\infty = O_P(d_n + d_n'/\tau_n)$ *and the same holds for* $\hat{T}_{\hat{V}(\beta)}, \partial_1\hat{T}_{\hat{V}(\beta)}, \partial_v\hat{T}_{\hat{V}(\beta)}, \partial_\beta\hat{T}_{\hat{V}(\beta)}, \partial_1\hat{f}_{\hat{V}(\beta)}, \partial_v\hat{f}_{\hat{V}(\beta)}, \partial_\beta\hat{f}_{\hat{V}(\beta)}$, *mutatis mutandis;*

(ii) $||(\hat{f}_{X,\hat{V}}(x,\hat{v}) - f_{X,V}(x,v)\cdot\hat{t}_n^{(Z)}||_\infty = O_P(d_n" + d_{n,0}/\tau_n)$;

(iii) $\sup_{\beta\in B}||[\nabla_\beta\hat{T}_{\hat{V}(\beta)}(x_1,\hat{m}_2(x_2),\hat{v}(\beta)) - \nabla_\beta T_{V(\beta)}(x_1,m_2(x_2),v(\beta))]\hat{t}_n^{(X_2,Z)}||_\infty = O_P(d_n + d_n'/\tau_n)$, *and the same holds by replacing* $\nabla_\beta\hat{T}_{\hat{V}(\beta)}$ *with* $\nabla_\beta\hat{f}_{\hat{V}(\beta)}$

(iv) $\tau_n^4\cdot\sup_{\beta\in B}||[\nabla^2_{\beta\beta^T}\hat{T}_{\hat{V}(\beta)}(x_1,\hat{m}_2(x_2),\hat{v}(\beta)) - \nabla^2_{\beta\beta^T}T_{V(\beta)}(x_1,m_2(x_2),v(\beta))]\cdot\hat{t}_n^{(X,\hat{V})}\cdot\hat{t}_n^{(X_2,Z)}||_\infty = o_P(1)$, *and the same holds for* $\hat{f}_{\hat{V}(\beta)}$, *mutatis mutandis.*

The above results also hold by replacing $\hat{t}_n^{(X_2,Z)}$ *with* $t_n^{(X_2,Z)}$ *and/or* $\hat{t}_n^{(X,\hat{V})}$ *with* $t_n^{(X,V)}$.

Proof. For the ease of notation, define $m_{0,2} := (m_0, m_2)$ and $\hat{m}_{0,2} := (\hat{m}_0,\hat{m}_2)$. Also, define the set $\mathcal{M} := \{m_{0,2}(x_2,z) : x_2\in\mathcal{X}_2\ ,\ z\in\mathcal{Z}\}$, its δ-enlargment $\mathcal{M}^\delta := \{u : |u-u'| < \delta \text{ with } u'\in\mathcal{M}\}$, and the event $\mathcal{A}_n^{(C)} := \{\sup_{\beta\in B}||\hat{f}_{\hat{V}(\beta)} - f_{V(\beta)}||_\infty \leq Cd_n\ ,\ ||\hat{t}_n^{(X_2,Z)}(\hat{m}_{0,2} - m_{0,2})||_\infty \leq Cd_n'/\tau_n\}$. By a large enough n such that $Cd_n'/\tau_n < \delta$, if $\mathcal{A}_n^{(C)}$ holds, whenever $\hat{t}_n^{(X_2,Z)}(x_2,z) = 1$, we have: (a) $\hat{m}_{0,2}(x_2,z)\in\mathcal{M}^\delta$, so that (b)$|\hat{f}_{\hat{V}(\beta)}(\hat{v}(\beta)) - f_{V(\beta)}(\hat{v}(\beta))| \leq \sup_{\beta\in B}||\hat{f}_{\hat{V}(\beta)} - f_{V(\beta)}||_\infty$, and hence by the Mean-Value Theorem, (c)$\sup_{\beta\in B}||[\hat{f}_{\hat{V}(\beta)}(\hat{v}(\beta)) - f_{V(\beta)}(v(\beta))]\hat{t}_n^{(X_2,Z)}||_\infty \leq \sup_{\beta\in B}||[\hat{f}_{\hat{V}(\beta)}(\hat{v}(\beta)) - f_{V(\beta)}(\hat{v}(\beta))]\hat{t}_n^{(X_2,Z)}||_\infty + \sup_{\beta\in B}||[f_{V(\beta)}(\hat{v}(\beta)) - f_{V(\beta)}(v(\beta))]\hat{t}_n^{(X_2,Z)}||_\infty \leq \sup_{\beta\in B}||\hat{f}_{\hat{V}(\beta)} - f_{V(\beta)}||_\infty + C||(\hat{m}_{0,2} - m_{0,2})\hat{t}_n^{(X_2,Z)}||_\infty \leq C\cdot(d_n + d_n'/\tau_n)$. Since by Lemma (10) and Lemma (12), choosing C large enough will make $P(\mathcal{A}_n^{(C)})$ arbitrarily close to one for any large n, we obtain result (i). By the same reasoning obtain the result for $\partial_1\hat{T}_{\hat{V}(\beta)}, \partial_v\hat{T}_{\hat{V}(\beta)}, \partial_\beta\hat{T}_{\hat{V}(\beta)}, \partial_1\hat{f}_{\hat{V}(\beta)}, \partial_v\hat{f}_{\hat{V}(\beta)}, \partial_\beta\hat{f}_{\hat{V}(\beta)}$, mutatis mutandis, and by replacing $\hat{t}_n^{(X_2,Z)}$ with $t_n^{(X_2,Z)}$. Also, by the same kind of reasoning we conclude for point (ii) and (iv). Finally, in order to get point (iii), we can combine the result of point(i) together with Lemma (12). □

In the lemmas that will follow, we will make use of two decompositions for a regression kernel estimator and its derivative, rispectively. Recall that $\hat{G}_{\hat{V}(\beta)} := \hat{T}_{\hat{V}(\beta)}/\hat{f}_{\hat{V}(\beta)}$. So,

$$\hat{G}_{\hat{V}(\beta)}(\hat{v}(\beta)) - G_{V(\beta)}(v(\beta)) =$$

$$= \underbrace{\frac{\hat{T}_{\hat{V}(\beta)}(\hat{v}(\beta)) - G_{V(\beta)}(v(\beta)) \cdot \hat{f}_{\hat{V}(\beta)}(\hat{v}(\beta))}{f_{V(\beta)}(v(\beta))}}_{=:\hat{b}_{n,1,\beta}(\tilde{x})} + \hat{b}_{n,1}(\tilde{x}) \underbrace{\frac{f_{V(\beta)}(v(\beta)) - \hat{f}_{\hat{V}(\beta)}(\hat{v}(\beta))}{\hat{f}_{\hat{V}(\beta)}(\hat{v}(\beta))}}_{=:\hat{b}_{n,2,\beta}(\tilde{x})} \tag{14}$$

From the section about the notation at the beginning of this Appendix, recall that

$\nabla_\beta G_{V(\beta)}(x_1, m_2(x_2), v(\beta)) = \partial_\beta G_{V(\beta)}(v(\beta)) + \partial_1 G_{V(\beta)}(v(\beta)) \cdot (x_1, m_2(x_2))^T$,

$\nabla_\beta \hat{G}_{\hat{V}(\beta)}(x_1, \hat{m}_2(x_2), \hat{v}(\beta)) = \partial_\beta \hat{G}_{\hat{V}(\beta)}(\hat{v}(\beta)) + \partial_1 \hat{G}_{\hat{V}(\beta)}(\hat{v}(\beta)) \cdot (x_1, \hat{m}_2(x_2))^T$, and the same holds for $\hat{T}_{\hat{V}(\beta)}$ and $\hat{f}_{\hat{V}(\beta)}$. So,

$$\nabla_\beta \hat{G}_{\hat{V}(\beta)}(x_1, \hat{m}_2(x_2), \hat{v}(\beta)) - \nabla_\beta G_{V(\beta)}(x_1, m_2(x_2), v(\beta)) =$$

$$= \underbrace{\frac{\hat{a}(x_1, \hat{m}_2(x_2), \hat{v}(\beta)) - \nabla_\beta G_{V(\beta)}(x_1, m_2(x_2), v(\beta)) \cdot \hat{f}_{\hat{V}(\beta)}(\hat{v}(\beta))}{f_{V(\beta)}(v(\beta))}}_{=:\hat{b}_{n,3,\beta}(\tilde{x})} + \hat{b}_{n,3}(\tilde{x}) \cdot \hat{b}_{n,2,\beta}(\tilde{x}) \tag{15}$$

with

$$\hat{a}_{\hat{V}(\beta)}(x_1, \hat{m}_2(x_2), \hat{v}(\beta)) := \nabla_\beta \hat{T}_{\hat{V}(\beta)}(x_1, \hat{m}_2(x_2), \hat{v}(\beta)) - \hat{G}_{\hat{V}(\beta)}(\hat{v}(\beta)) \cdot \nabla_\beta \hat{f}_{\hat{V}(\beta)}(x_1, \hat{m}_2(x_2), \hat{v}(\beta)) \ .$$

Lemma 18. *Let Assumptions (1), (2), (3), (4), and (5i) hold. Then,*

(i)$sup_{\beta\in B,\tilde{x}\in\tilde{\mathcal{X}}} \left| \hat{b}_{n,1,\beta}(\tilde{x}) \cdot t_n^{(X_1,V)} \cdot t_n^{(X_2,Z)} \right| = o_p(\tau_n n^{-1/4})$ *and* $sup_{\beta\in B,\tilde{x}\in\tilde{\mathcal{X}}} \left| \hat{b}_{n,3,\beta}(\tilde{x}) \cdot t_n^{(X_1,V)} \cdot t_n^{(X_2,Z)} \right| = o_p(n^{-1/4})$ *;*

(ii) $sup_{\beta\in B,\tilde{x}\in\tilde{\mathcal{X}}} \left| \hat{b}_{n,1,\beta}(\tilde{x}) \cdot \hat{b}_{n,2,\beta}(\tilde{x}) \cdot t_n^{(X_1,V)} \cdot t_n^{(X_2,Z)} \right| = o_p(\tau_n n^{-1/2})$ *and* $sup_{\beta\in B,\tilde{x}\in\tilde{\mathcal{X}}} \left| \hat{b}_{n,3,\beta}(\tilde{x}) \cdot \hat{b}_{n,2,\beta}(\tilde{x}) \cdot \right.$

$t_n^{(X_1,V)} \cdot t_n^{(X_2,Z)} \Big| = o_p(n^{-1/2})$.

and the same results hold by replacing $t_n^{(X,V)}$ *with* $\hat{t}_n^{(X,V)}$ *and/or* $t_n^{(X_2,Z)}$ *with* $\hat{t}_n^{(X_2,Z)}$.

Proof. Define the rates $d_n = \epsilon_n n^{-1/4}\tau_n$, $d_n" = \sqrt{\frac{\log n}{nh_3^q}} + h_3^{r_3}$, $d_n' = d_{n,0} \vee d_{n,2}$, with $d_{n,0} := \sqrt{\frac{\log n}{nh_0^{p_0}}} + h_0^{r_0}$ and $d_{n,2} := \sqrt{\frac{\log n}{nh_2^{p_2}}} + h_2^{r_2}$ and the events:

$\mathcal{A}_n^{(C)} := \Big\{ ||\hat{t}_n^{(X_2,Z)} \cdot [\hat{f}_{X,\hat{V}}(x,\hat{v}) - f_{X,\hat{V}}(x,v)]||_\infty \leq C \cdot (d_n" + d_n'/\tau_n)$, $\sup_{\beta \in B} ||[\hat{f}_{\hat{V}(\beta)}(\hat{v}(\beta)) - f_{V(\beta)}(v(\beta))]\hat{t}_n^{(X_2,Z)}||_\infty \leq C \cdot (d_n + d_n'/\tau_n)$, $\sup_{\beta \in B} ||[\hat{T}_{\hat{V}(\beta)}(\hat{v}(\beta)) - T_{\hat{V}(\beta)}(v(\beta))]\hat{t}_n^{(X_2,Z)}||_\infty \leq C \cdot (d_n + d_n'/\tau_n) \Big\}$, and

$\mathcal{B}_n^{(C)} := \Big\{ \sup_{\beta \in B} ||[\nabla_\beta \hat{T}_{\hat{V}(\beta)}(x_1, \hat{m}_2(x_2), \hat{v}(\beta)) - \nabla_\beta T_{V(\beta)}(x_1, m_2(x_2), v(\beta))]\hat{t}_n^{(X_2,Z)}||_\infty \leq C \cdot (d_n + d_n'/\tau_n)$, $\sup_{\beta \in B} ||[\nabla_\beta \hat{f}_{\hat{V}(\beta)}(x_1, \hat{m}_2(x_2), \hat{v}(\beta)) - \nabla_\beta f_{V(\beta)}(x_1, m_2(x_2), v(\beta))]\hat{t}_n^{(X_2,Z)}||_\infty \leq C \cdot (d_n + d_n'/\tau_n) \Big\}$.

For large enough n such that $C \cdot (d_n"/\tau_n + d_n'/\tau_n^2) < 1/2$ and $C \cdot (d_n/\tau_n + d_n'/\tau_n^2) < 1/8$, we must have over the set $\mathcal{A}_n^{(C)} \cap \mathcal{B}_n^{(C)}$:

(a) $f_{X,V}(x,v) \geq [\hat{f}_{X,\hat{V}}(x,\hat{v})/\tau_n - C \cdot (d_n"/\tau_n + d_n'/\tau_n^2)]\tau_n \geq \tau_n/2$ whenever $\hat{t}_n^{(X_2,Z)}(\tilde{x}) = \hat{t}_n^{(X,\hat{V})}(\tilde{x}) = 1$, which imples that $f_{V(\beta)}(v(\beta)) \geq \tau_n/4$ for all $\beta \in B$; hence,

(b) $\hat{f}_{\hat{V}(\beta)}(\hat{v}(\beta)) \geq [f_{V(\beta)}(v(\beta))/\tau_n - C \cdot (d_n/\tau_n + d_n'/\tau_n^2)] \cdot \tau_n \geq \tau_n/8$ whenever $\hat{t}_n^{(X_2,Z)}(\tilde{x}) = \hat{t}_n^{(X,\hat{V})}(\tilde{x}) = 1$, and furthermore

(c) $|\hat{f}_{\hat{V}(\beta)}(\hat{v}(\beta)) - f_{V(\beta)}(v(\beta))| \leq C \cdot (d_n + d_n'/\tau_n)$, and $|\hat{T}_{\hat{V}(\beta)}(\hat{v}(\beta)) - T_{V(\beta)}(v(\beta))| \leq C \cdot (d_n + d_n'/\tau_n)$ for all $\beta \in B$ whenever $\hat{t}_n^{(X_2,Z)}(\tilde{x}) = \hat{t}_n^{(X,\hat{V})}(\tilde{x}) = 1$.

The above reasoning implies that for large enough n such that $C \cdot (d_n"/\tau_n + d_n'/\tau_n) < 1/2$ and $C \cdot (d_n/\tau_n + d_n'/\tau_n^2) < 1/8$, over the set $\mathcal{A}_n^{(C)} \cap \mathcal{B}_n^{(C)}$:

(d) $\sup_{\beta \in B} ||\hat{b}_{n,\beta,1}\hat{t}_n^{(X,\hat{V})}\hat{t}_n^{(X_2,Z)}||_\infty \leq C \cdot (d_n/\tau_n + d_n'/\tau_n^2)$ and $\sup_{\beta \in B} ||\hat{b}_{n,\beta,2}\hat{t}_n^{(X,\hat{V})}\hat{t}_n^{(X_2,Z)}||_\infty \leq C \cdot (d_n/\tau_n + d_n'/\tau_n^2)$, and hence by the decomposition in Eq. (14) we also have

(e) $\sup_{\beta \in B} ||[\hat{G}_{\hat{V}(\beta)}(\hat{v}(\beta)) - G_{V(\beta)}(v(\beta))] \cdot \hat{t}_n^{(X,\hat{V})}\hat{t}_n^{(X_2,Z)}||_\infty \leq C \cdot (d_n/\tau_n + d_n'/\tau_n^2)$, so that we must also have

(f) $\sup_{\beta \in B} ||[\hat{a}_{\hat{V}(\beta)}(\hat{v}(\beta)) - \nabla_\beta G_{V(\beta)}(m, \beta, \cdot) \cdot \hat{f}_{\hat{V}(\beta)}(\hat{v}(\beta))] \cdot \hat{t}_n^{(X_2,Z)} \cdot \hat{t}_n^{(X,\hat{V})}||_\infty \leq C \cdot (d_n/\tau_n + d_n'/\tau_n^2)$.

By the above reasoning, for n such that $C \cdot (d_n/\tau_n + d_n'/\tau_n^2) < 1/8$ and $C \cdot (d_n"/\tau_n + d_n'/\tau_n^2) < 1/2$, the event $\mathcal{A}_n^{(C)} \cap \mathcal{B}_n^{(C)}$ implies $|\hat{b}_{n,\beta,3}(\tilde{x})| \leq C \cdot (d_n/\tau_n^2 + d_n'/\tau_n^3)$ for all $\beta \in B$, whenever $\hat{t}_n^{(X_2,Z)}(\tilde{x}) = \hat{t}_n^{(X,\hat{V})}(\tilde{x}) = 1$, so that $\sup_{\beta \in B} ||\hat{b}_{n,\beta,3} \cdot \hat{t}_n^{(X,\hat{V})}\hat{t}_n^{(X_2,Z)}||_\infty \leq C \cdot (d_n/\tau_n^2 + d_n'/\tau_n^3)$. Since by Lemma (17) for large enough C the probability $P(\mathcal{A}_n^{(C)} \cap \mathcal{B}_n^{(C)})$ can be made arbitrarily close to 1 for any large

n, we conclude. □

.

Lemma 19. *Let Assumptions (1)-(4) and Assumpitons (5i) and (5iii) hold. Then,*

$sup_{\beta\in B}||[\nabla^2_{\beta\beta^T}\hat{G}_{\hat{V}(\beta)}(x_1,\hat{m}_2(x_2),\beta)-\nabla^2_{\beta\beta^T}G_{V(\beta)}(x_1,m_2(x_2),\beta)]\cdot\hat{t}_n^{(X,\hat{V})}\cdot\hat{t}_n^{(X_2,Z)}||_\infty=o_P(1)$

Proof. We can proceed in a very similar way for point (ii) of the previous lemma, and making use of the results of Lemma (17) □

.

Lemma 20. *Under Assumptions (1)-(4),* $sup_{\beta\in B}\left|\left|\frac{\hat{T}_{\hat{V}(\beta)}(v(\beta))-G_{V(\beta)}(v(\beta))\cdot\hat{f}_{\hat{V}(\beta)}(v(\beta))}{f_{V(\beta)}(v(\beta))}t_n^{(X,V)}\right|\right|_\infty=o_P(n^{-1/4})$.

Proof. Define the rate $d_n=\epsilon_n n^{-1/4}\tau_n$ and the event

$\mathcal{A}_n^{(C)}:=\{\sup_{\beta\in B}||\hat{T}_{\hat{V}(\beta)}-T_{V(\beta)}||_\infty\le Cd_n\ ,\ \sup_{\beta\in B}||\hat{f}_{\hat{V}(\beta)}-f_{V(\beta)}||_\infty\le Cd_n\}$.

If $t_n^{(X,V)}(x,v)=1$, then $f_{X,V}(x,v)\ge\tau_n$,and hence $f_{V(\beta)}(v(\beta))\ge\tau_n/2$ for all $\beta\in B$. Over the set $\mathcal{A}_n^{(C)}$,

$$\sup_{\beta\in B}\left|\left|\frac{\hat{T}_{\hat{V}(\beta)}(v(\beta))-G_{V(\beta)}(v(\beta))\cdot\hat{f}_{\hat{V}(\beta)}(v(\beta))}{f_{V(\beta)}(v(\beta))}t_n^{(X,V)}\right|\right|_\infty\le Cd_n\sup_{\beta\in B}\left|\left|\frac{t_n^{(X,V)}(x,v)}{f_{V(\beta)}(v(\beta))}\right|\right|_\infty\le Cd_n/\tau_n$$

Conclude by noticing that from Lemma (10) by choosing C large enough, $P(\mathcal{A}_n^{(C)})$ can be made arbitrarily close to one for any large n. □

.

Lemma 21. *Define* $\hat{g}_{n,t}^{(1)}(\tilde{X}):=(\hat{T}_{\hat{V}(\hat{\beta})}-G_{V(\beta_0)}\cdot\hat{f}_{\hat{V}(\hat{\beta})})(V(\beta_0))\cdot\phi(\tilde{X}t)$ *and* $\hat{g}_{n,t}^{(2)}:=\varepsilon\cdot\hat{f}_{\hat{V}(\hat{\beta})}(V(\beta_0))\cdot\phi(\tilde{X}t)$. *Under Assumption (1)-(4),*

$$\mathbb{G}_n\hat{g}_{n,t}^{(1)}=o_p(1)\ \textit{and}\ \mathbb{G}_n\hat{g}_{n,t}^{(2)}=o_p(1),\ \textit{uniformly in}\ t\in T$$

Proof. Define $s_0:=1+\underline{Ev}(d)/2$. From Lemma (10), for any fixed $\delta>0$ arbitrarily small, the event $B_n^{(\delta)}:=\{||\hat{T}_{\hat{V}(\beta)}-G_{V(\beta)}\cdot\hat{f}_{\hat{V}(\beta)}||_\infty<\delta,\ \hat{T}_{\hat{V}(\beta)}-G_{V(\beta)}\cdot\hat{f}_{\hat{V}(\beta)}\in\mathcal{C}^{s_0}\}$ has a probability that converges to one as $n\to\infty$. By definition of the class $\mathcal{C}^{s_0}$, $\log N(\varepsilon,\mathcal{C}^{s_0},||\cdot||_\infty)\le$

$C \cdot \varepsilon^{-d/s_0}$, with $d/s_0 \in (0,2)$. Notice that over the set $B_n^{(\delta)}$, $\sup_{t\in T}|\mathbb{G}_n \hat{g}_{n,t}| \le ||\mathbb{G}_n||_{\mathcal{F}^\delta}$, where $\mathcal{F}^{(\delta)} := \{g \cdot \phi(\cdot t) : g \in \mathcal{C}^{s_0},\ ||g||_\infty < \delta,\ t \in T\}$. Furthermore, since $\mathcal{F}^{(\delta)} \subset \mathcal{F}^{(1)}$ for $\delta < 1$, $\log N_{[]}(\varepsilon, \mathcal{F}^{(\delta)}, L_2(P)) \le \log N_{[]}(\varepsilon, \mathcal{F}^{(1)}, L_2(P))$, and from Lemma (9) $\log N_{[]}(\varepsilon, \mathcal{F}^{(1)}, L_2(P)) \le C\varepsilon^{-d/s_0}$ with $d/s_0 \in (0,2)$. So, $J_{[]}(1, \mathcal{F}^{(1)}, L_2(P)) < \infty$, and therefore for $\delta \to 0$ we must have $J_{[]}(\delta, \mathcal{F}^{(\delta)}, L_2(P)) \le J_{[]}(\delta, \mathcal{F}^{(1)}, L_2(P)) \to 0$. Now, by Lemma (6), $\mathbb{E}||\mathbb{G}_n||_{\mathcal{F}^{(\delta)}} \le J_{[]}(\delta, \mathcal{F}^{(\delta)}, L_2(P)) + \sqrt{n}C\{C > a_{\mathcal{F}^{(\delta)}}(\delta)\sqrt{n}\}$. Since for any $\delta > 0$,

$\limsup_{n\to\infty}\sqrt{n}C\{C > a_{\mathcal{F}^{(\delta)}}(\delta)\sqrt{n}\} \le \limsup_{n\to\infty} a_{\mathcal{F}^{(\delta)}}(\delta)^{-1} C\{C > a_{\mathcal{F}^{(\delta)}}(\delta)\sqrt{n}\} = 0$, by Markov's inequality we conclude. □

Before the next lemma, we introduce another small piece of notation. For a function $(x_1, x_2) \mapsto \alpha(x_1, x_2)$, we denote $\alpha(x_1, P) := \int \alpha(x_1, x)\, \mathrm{d}P(x)$.

Lemma 22. *Let $\Psi := \{\tilde{x} \mapsto \varphi_t(\tilde{x}) : t \in T\}$ be a class of functions that are Lipschitz in t, i.e. $||\varphi_{t_1} - \varphi_{t_2}||_\infty \le C|t_1 - t_2|$, and let Assumption (1)-4(ii) hold. Then,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{m}_{0,i} - m_{0,i}) \cdot \varphi_t(\tilde{X}_i) \cdot t_{n,i}^{(Z)} =$$

$$= \frac{1 + o(1)}{\sqrt{n}} \sum_{i=1}^n \int t_n^{(Z)}(Z_i + uh) \tilde{\varphi}_t(Z_i + uh) K_0(u) \cdot [D_{0,i} - m_0(Z_i + uh)]\ du + o_P(1)$$

uniformly in $t \in T$, with $\tilde{\varphi}_t(z) := \mathbb{E}\{\varphi_t(\tilde{X})|Z = z\}$. The same type of expansion holds for $\hat{m}_2$, mutatis mutandis.

Proof. For the ease of notation, let us drop the index 0 from $\hat{m}_0$ and m_0, and hence let m stand for m_0 and $\hat{m}$ for $\hat{m}_0$. By the decomposition in Eq. (13) and Lemma (12),

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\hat{m}_i - m_i) \cdot \varphi_t(\tilde{X}_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\hat{T}_i - m_i \hat{f}_i}{f_i} t_{n,i}^{(Z)} \varphi_t(\tilde{X}_i) + o_P(1) = \frac{1}{\sqrt{n} n h^p \tau_n} \sum_{i \ne j} \alpha_{n,t}(\tilde{X}_i, \tilde{X}_j) + o_P(1) \tag{16}$$

uniformly in $t \in T$, where $\alpha_{n,t}(\tilde{X}_i, \tilde{X}_j) = \tau_n t_{n,i}^{(Z)} \varphi_t(\tilde{X}_i) K((Z_j - Z_i)/h)[D_j - m_i]/f_i$, and the last inequality in the above display follows from the definitions of $\hat{T}$ and $\hat{f}$, Assumption 4 (ii), and Lemma (12). Define now the symmetric kernel $\tilde{\alpha}_{t,n}(\tilde{X}_i, \tilde{X}_j) := \alpha_{t,n}(\tilde{X}_i, \tilde{X}_j) + \alpha_{t,n}(\tilde{X}_j, \tilde{X}_i)$, and notice that

$$\frac{1}{\sqrt{n}nh^p\tau_n}\sum_{i\neq j}\alpha_{n,t}(\tilde{X}_i,\tilde{X}_j) = \frac{\binom{n}{2}}{\sqrt{n}nh^p\tau_n}\binom{n}{2}^{-1}\sum_{C_{n,2}}\tilde{\alpha}_{t,n}(\tilde{X}_i,\tilde{X}_j), \tag{17}$$

where $C_{n,2}$ denotes all the combinations of class 2 of the elements $\{1,..,n\}$. Furthermore,

$$\frac{2}{n}\sum_{i=1}^{n}\tilde{\alpha}_{t,n}(\tilde{X}_i,P) = \binom{n}{2}^{-1}\sum_{C_{n,2}}[\tilde{\alpha}_{t,n}(\tilde{X}_i,P)+\tilde{\alpha}_{t,n}(\tilde{X}_j,P)]\ , \tag{18}$$

and by the definition of K, $\alpha_{t,n}(P,P) = O(h^{r+p})$. So,

$$\binom{n}{2}^{-1}\sum_{C_{n,2}}\tilde{\alpha}_{t,n}(\tilde{X}_i,\tilde{X}_j) - \frac{2}{n}\sum_{i=1}^{n}\tilde{\alpha}_{t,n}(\tilde{X}_i,P) =$$

$$= \binom{n}{2}^{-1}\sum_{C_{n,2}}\lambda_{t,n}(\tilde{X}_i,\tilde{X}_j) + O(h^{p+r}) = \underbrace{\frac{1}{n(n-1)}\sum_{i\neq j}\lambda_{t,n}(\tilde{X}_i,\tilde{X}_j)}_{=:R_n(t)} + O(h^{p+r})\ , \tag{19}$$

where $\lambda_{t,n}(\tilde{X}_i,\tilde{X}_j) := \tilde{\alpha}_{t,n}(\tilde{X}_i,\tilde{X}_j) - \tilde{\alpha}_{t,n}(\tilde{X}_i,P) - \tilde{\alpha}_{t,n}(\tilde{X}_j,P) + \tilde{\alpha}_{t,n}(P,P)$ is a symmetric degenerate kernel of order 2. Since $||\alpha_{t_1,n} - \alpha_{t_2,n}||_\infty \leq C|t_1 - t_2|$, then also $\tilde{\alpha}_{t,n}$ and hence $\lambda_{t,n}$ must satisfy the same type of Lipschitz condition. So, from Lemma (9), we can apply Lemma (7) with $k = 2$ to get

$$\frac{P\sup_{t\in T}|nR_n(t)|^2}{nh^{2p}\tau_n^2} \leq \frac{\Gamma\tilde{c}^2}{nh^{2p}\tau_n^2}\ . \tag{20}$$

By the rates in Assumption 4 (ii), the right-hand side of the above display is $o(1)$, so that $\sup_{t\in T}\sqrt{n}R_n(t)/(h^p\tau_n) = o_P(1)$. This latter result, together with Eq. (17) and (19) implies that

$$\frac{1}{\sqrt{n}nh^p\tau_n}\sum_{i\neq j}\alpha_{t,n}(\tilde{X}_i,\tilde{X}_j) = \frac{(1+o(1))}{\sqrt{n}}\sum_{i=1}^{n}\frac{[\alpha_{t,n}(\tilde{X}_i,P)+\alpha_{t,n}(P,\tilde{X}_i)]}{h^p\tau_n} + o_P(1) \tag{21}$$

uniformly in $t\in T$. Finally, since K is a kernel of order r, by an $r-$th order Taylor expansion, $\alpha_{t,n}(\tilde{x},P) = O(h^{p+r})$ uniformly in $t,\tilde{x}$. Conclude by this latter result, Eq. (16) and (21). □

.

Lemma 23. *Let Assumption (1)-4(ii) hold. Then,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int t_n^{(Z)}(Z_i+uh)\tilde{\varphi}_t(Z_i+uh)^T\cdot K_0(u)\cdot[D_{0,i}-m_0(Z_i+uh)]\ du=$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\varphi}_t(Z_i)^T\cdot(D_{0,i}-m_{0,i})+o_P(1)$$

uniformly in $t\in T$. The same result also holds by replacing Z , K_0, and D_0 with X_2, K_2, and D_2, respectively.

Proof. For the ease of notation, let us drop the index 0 from D_0, m_0, and K_0. So, at their place let us consider D, m, and K, respectively. Define $g_{n,t}^{(1)}(D,Z):=\int t_n^{(Z)}(Z+uh)\tilde{\varphi}_t(Z+uh)K(u)\cdot[D-m(Z+uh)]\ du$. Notice that by the usual change of variable typical of nonparametric kernel estimation, $\int t_n(z)h^{-p}\tilde{\varphi}_t(z)K((z-Z_i)/h)[D-m(z)]\mathrm{d}u=g_{n,t}(D,Z)$. So, by Tonelli-Fubini's Theorem, $Pg_{n,t}^{(1)}=\int t_n(z)h^{-p}\tilde{\varphi}_t(z)\cdot\{\int K((\tilde{z}-z)/h)[\tilde{d}-m(z)]\mathrm{d}P(\tilde{d},\tilde{z})\}\mathrm{d}z=O(h^r)$ uniformly in $t\in T$, where the last equality follows from applying the law of iterated expectations, i.e. obtaining $\mathbb{E}\{d|Z=\tilde{z}\}$ inside the integral, deriving an $r-$th order Taylor expansion, and finally using the order of the kernel K, in the same fashion as in nonparametric kernel estimation. So, by Assumption 4(ii), $\sqrt{n}\mathbb{P}_n g_{n,t}^{(1)}=\mathbb{G}_n g_{n,t}^{(1)}+o_P(1)$, uniformly in $t\in T$. Define now $g_{n,t}^{(0)}(D,Z):=\int\tilde{\varphi}_t(Z+uh)K(u)[D-m(Z+uh)]\mathrm{d}u$, and the class $\overline{\mathcal{G}_n}:=\{g_{n,t}^{(1)}-g_{n,t}^{(0)}\,:\,t\in T\}$. Since $||g_{n,t_1}^{(1)}-g_{n,t_2}^{(1)}||_\infty\leq C|t_1-t_2|$ for any $t_1,t_2\in T$, and the same holds for $g_{n,t}^{(0)}$, the collection of functions in $\overline{\mathcal{G}_n}$ is Lipschitz in t, so by Lemma (9), $N_{[]}(\varepsilon,\overline{\mathcal{G}_n},L_2(P))\leq C\varepsilon^{-dim(\tilde{X})}$, and hence $\text{limsup}_{n\to\infty}J_{[]}(\delta,\overline{\mathcal{G}_n},L_2(P))$ can be made arbitrarily small by choosing a $\delta>0$ small enough. Now, for z such that $f(z)>0$, $t_n(z+uh)\to 1$ for any fixed u, so by the Lebesgue Dominated Convergence Theorem, $\sup_{t\in T}|g_{n,t}^{(1)}(d,z)-g_{n,t}^{(0)}(d,z)|\to 0$ for any (d,z) such that $f(z)>0$. Hence, $\sup_{t\in T}||(g_{n,t}^{(1)}-g_{n,t}^{(0)})^2||_{2,P}=\sup_{t\in T}||(g_{n,t}^{(1)}-g_{n,t}^{(0)})^2 1\{f>0\}||_{2,P}=:\delta_n\to 0$, again by a Lebesgue Dominated Convergence argument. For any $\delta>0$ such that $\delta>\delta_n$ by Lemma (6), $\mathbb{E}||\mathbb{G}_n||_{\overline{\mathcal{G}_n}}\leq J_{[]}(\delta,\overline{\mathcal{G}}_n,L_2(P))+\sqrt{n}C\{C>\sqrt{n}a_n(\delta)\}$. As already noticed above, $\text{limsup}_{n\to\infty}J_{[]}(\delta,\overline{\mathcal{G}_n},L_2(P))$ can be made arbitrarily small by choosing a δ small enough. Since as noticed above, $N_{[]}(\varepsilon,\overline{\mathcal{G}_n},L_2(P))\leq C\varepsilon^{-dim(\tilde{X})}$, $a_{\overline{\mathcal{G}_n}}(\delta)$ is bounded from above and from below away from zero for any fixed $\delta > 0$, so that for any fixed $\delta > 0$: $\text{limsup}_{n\to\infty}\sqrt{n}C\{C > \sqrt{n}a_{\overline{\mathcal{G}_n}}(\delta)\} \leq \text{limsup}_{n\to\infty}a_{\overline{\mathcal{G}_n}}(\delta)^{-1}C\{C > \sqrt{n}a_{\overline{\mathcal{G}_n}}(\delta)\} = 0$. Deduce by Markov's inequality that $||\mathbb{G}_n||_{\overline{\mathcal{G}_n}} = o_P(1)$. So, conclude by noticing that $g_{n,t}^{(0)}(d,z) = (d - m(z))\tilde{\varphi}_t(z) + O(h^r)$ uniformly in (d,z,t). □

Lemma 24. *Let Assumptions (1)-(4) hold, and let $(y,\tilde{x}) \mapsto \varphi_n(y,\tilde{x})$ be a mapping satisfying $sup_{(y,\tilde{x})\in Supp(Y,\tilde{X})}|\varphi_n(y,\tilde{x})\cdot\tau_n| \leq C$. Then,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_n(Y,\tilde{X}_i)\cdot t_{n,i}^{(Z)}\cdot(\hat{m}_{0,i} - m_{0,i}) =$$

$$= \frac{1+o(1)}{\sqrt{n}}\sum_{i=1}^{n}\int\frac{\mathbb{E}\{\varphi_n(Y,\tilde{X})|Z=z\}t_n^{(Z)}(z)}{f_0(z)h_0^p}\cdot(D_{0,i} - m_0(z))\cdot K_0\Big(\frac{Z_i - z}{h}\Big)\ dP(z) + o_P(1).$$

and the same type of result holds by replacing Z, $\hat{m}_0$, m_0, f_0, K_0, and D_0 with X_2,$\hat{m}_2$, m_2, f_2, K_2, and D_2 , respectively.

Proof. For the ease of notation let us drop the index 0 from D_0, m_0, p_0, h_0, and K_0. So, at their place let us consider D, m, p, h and K, respectively. Also, for notational simplicity, let $\mathcal{X} := (Y,\tilde{X})$ and $\chi := (y,\tilde{x})$. By the decomposition of $\hat{m} - m$ in Eq.(13) and by Lemma (12),

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_n(\mathcal{X}_i)t_{n,i}^{(Z)}\cdot(\hat{m}_i - m_i) =$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_n(\mathcal{X}_i)t_{n,i}^{(Z)}\cdot\frac{\hat{T}_i - m_i\hat{f}_i}{f_i} + o_P(1) = \frac{1}{\sqrt{n}n}\sum_{i\neq j}\alpha_n(\mathcal{X}_i,\mathcal{X}_j) + o_P(1)\ , \tag{22}$$

where $\alpha_n(\mathcal{X}_i,\mathcal{X}_j) = (f_ih^p)^{-1}\cdot\varphi_n(\mathcal{X}_i)\cdot t_{n,i}^{(Z)}\cdot(D_j - m_i)\cdot K((Z_j - Z_i)/h)$, and the last inequality of the above display follows from the definition of $\hat{T}$ and $\hat{f}$, the rates in Assumption 4(ii), and Lemma (12). Define the syummetric kernel $\tilde{\alpha}_n(\mathcal{X}_i,\mathcal{X}_j) := \alpha_n(\mathcal{X}_i,\mathcal{X}_j) + \alpha_n(\mathcal{X}_j,\mathcal{X}_i)$. Then,

$$\frac{1}{\sqrt{n}n}\sum_{i\neq j}\alpha_n(\mathcal{X}_i,\mathcal{X}_j) = \frac{\binom{n}{2}}{\sqrt{n}n}\binom{n}{2}^{-1}\sum_{C_{n,2}}\tilde{\alpha}_n(\mathcal{X}_i,\mathcal{X}_j) \tag{23}$$

Notice that by the usual change of variables, the Taylor expansions, and the order of the kernel K typical of kernel regression estimation, $\mathbb{E}\{\alpha_n(\mathcal{X}_1, \mathcal{X}_2)\} = O(h^r \tau_n^{-1})$, and $\mathbb{E}|\tilde{\alpha}_n(\mathcal{X}_1, \mathcal{X}_2)|^2/n = C \cdot (n\tau_n^4 h^{2p})^{-1} = o(1)$, where the latter equality follows from the rates specified in Assumption 4(ii). Hence, by Lemma B3 in Ahn (1997), we get

$$\binom{n}{2}^{-1} \sum_{C_{n,2}} \tilde{\alpha}_n(\mathcal{X}_i, \mathcal{X}_j) = \frac{2}{n} \sum_{i=1}^{n} \tilde{\alpha}_n(\mathcal{X}_i, P) + O(h^r \tau_n^{-1}) + o_P(n^{-1/2}) \ . \tag{24}$$

By using the classical change of vairable and the order of the kernel K,

$$\alpha_n(\chi, P) = O(h^r \tau_n^{-2}) \text{ and}$$

$$\alpha_n(P, \mathcal{X}_i) = \int \frac{\mathbb{E}\{\varphi_n(\mathcal{X})|Z = z\} \cdot t_n^{(Z)}(z)}{h^p f(z)} (D_i - m(z)) \cdot K\Big(\frac{Z_i - z}{h}\Big) \, \mathrm{d}P(z) \text{ uniformly in } \chi \mathrm{x} \mathcal{X}_i \in \tilde{X}^2$$

Conclude by the above display, Eq. (24), (23), and (22). □

.

Lemma 25. *Let Assumptions (1)-(4) and Assumption (5i) hold. Then,*

(i) sup $_{\beta \in B} \left|\left| \frac{\hat{T}_{\hat{V}(\beta)}(\hat{v}(\beta)) - \hat{T}_{\hat{V}(\beta)}(v(\beta)) - \partial T_{V(\beta)}(v(\beta)) \cdot (\hat{v}(\beta) - v(\beta))}{f_{V(\beta)}(v(\beta))^2} \hat{t}_n^{(X_2,Z)} \hat{t}_n^{(X,\hat{V})} \right|\right|_\infty = o_P(n^{-1/2})$, *and the same holds by replacing* $\hat{t}_n^{(X_2,Z)}$ *with* $t_n^{(X_2,Z)}$ *and/or* $\hat{t}_n^{(X,\hat{V})}$*with* $t_n^{(X,V)}$ *;*

(ii) sup $_{\beta \in B} \left|\left| [\hat{T}_{\hat{V}(\beta)}(\hat{v}(\beta)) - \hat{T}_{\hat{V}(\beta)}(v(\beta)) - \partial T_{V(\beta)}(v(\beta)) \cdot (\hat{v}(\beta) - v(\beta))] \cdot \hat{t}_n^{(X_2,Z)} \right|\right|_\infty = o_P(n^{-1/2})$, *and the same holds by replacing* $\hat{t}_n^{(X_2,Z)}$ *with* $t_n^{(X_2,Z)}$ *.*

(iii) If $\hat{\beta} - \beta_0 = O_P(n^{-1/2})$, *then* $\left|\left| [\hat{T}_{\hat{V}(\hat{\beta})}(v(\hat{\beta})) - \hat{T}_{\hat{V}(\hat{\beta})}(v(\beta_0)) - \partial T_{V(\beta_0)}(v(\beta_0)) \cdot (v(\hat{\beta}) - v(\beta_0))] \cdot \hat{t}_n^{(X_2,Z)} \right|\right|_\infty = o_P(n^{-1/2})$, *and the same holds by replacing* $\hat{t}_n^{(X_2,Z)}$ *with* $t_n^{(X_2,Z)}$ *.*

(iv) If $\hat{\beta} - \beta_0 = O_P(n^{-1/2})$, *then* $\left|\left| \Big[\hat{T}_{\hat{V}(\hat{\beta})}(\hat{v}(\hat{\beta})) - \hat{T}_{\hat{V}(\hat{\beta})}(v(\beta_0)) - \partial T_{V(\beta_0)}(V(\beta_0)) \cdot [v(\hat{\beta}) - v(\beta_0)] - \partial T_{V(\beta_0)}(v(\beta_0)) \cdot [\hat{v}(\beta_0) - v(\beta_0)]\Big] \cdot \hat{t_n}^{(X_2,Z)} \right|\right|_\infty = o_P(n^{-1/2})$, *and the same holds by replacing* $\hat{t}_n^{(X_2,Z)}$ *with* $t_n^{(X_2,Z)}$.

The above first-order expansions also hold for $\hat{f}_{\hat{V}(\beta)}$, *mutatis mutandis.*

Proof. Define the rates $d_n = \epsilon_n n^{-1/4} \tau_n$, $d_n" = \sqrt{\frac{\log n}{n h_3^q}} + h_3^{r_3}$, $d_{n,0} := \sqrt{\frac{\log n}{n h_0^{p_0}}} + h_0^{p_0}$, $d_{n,2} := \sqrt{\frac{\log n}{n h_2^{p_2}}} + h_2^{p_2}$ $d_n' = d_{n,0} \vee d_{n,2}$. For the ease of notation define $m_{0,2} := (m_0, m_2)$, $\hat{m}_{0,2} := (\hat{m}_0, \hat{m}_2)$, the set

$\mathcal{M} := \{m_{0,2}(x_2, z) : x_2 \in \mathcal{X}_2 \ , \ z \in \mathcal{Z}\}$ and its δ-enlargment $\mathcal{M}^\delta := \{u : |u - u'| < \delta$ for some $u' \in \mathcal{M}\}$, with $\delta > 0$. Define the event $\mathcal{A}_n^{(C)} := \{\sup_{\beta \in B} ||\partial\hat{T}_{\hat{V}(\beta)} - \partial T_{V(\beta)}||_\infty \leq C d_n \ , \ ||(\hat{m}_{0,2} - m_{0,2}) \cdot \hat{t}_n^{(X_2,Z)}||_\infty \leq C d'_n/\tau_n \ , \ ||(\hat{f}_{X,\hat{V}}(x,\hat{v}) - f_{X,V}(x,v)) \cdot \hat{t}_n^{(X_2,Z)}||_\infty \leq C \cdot (d_n" + d'_n/\tau_n)\}$. By the Mean-Value theorem,

$$\hat{T}_{\hat{V}(\beta)}(\hat{v}(\beta)) - \hat{T}_{\hat{V}(\beta)}(v(\beta)) - \partial T_{V(\beta)}(v(\beta)) \cdot (\hat{v}(\beta) - v(\beta)) =$$

$$= \left[\partial\hat{T}_{\hat{V}(\beta)}(\tilde{v}(\beta)) - \partial T_{V(\beta)}(v(\beta))\right] \cdot (\hat{v}(\beta) - v(\beta)) \ , \tag{25}$$

where $\tilde{v}(\beta) := (x_1\beta_1 + \beta_2\tilde{m}_2(x_2), x^e - \tilde{m}_0(z))$, $\tilde{m}_2(x_2) \in [\hat{m}_2(x_2), m_2(x_2)]$, and $\tilde{m}_0(z) \in [\hat{m}_0(z), m_0(z)]$. Choose n large enough so that $C d'_n/\tau_n < \delta$ and $C \cdot (d_n"/\tau_n + d'_n/\tau_n^2) < 1/2$. Then, over the set $\mathcal{A}_n^{(C)}$:

(i) $f_{X,V}(x,v) \geq [\hat{f}_{X,\hat{V}}(x,\hat{v})/\tau_n - C \cdot (d_n"/\tau_n + d'_n/\tau_n^2)] \cdot \tau_n > \tau_n/2$, whenever $\hat{t}_n^{(X,\hat{V})}(\tilde{x}) = \hat{t}_n^{(X_2,Z)}(\tilde{x}) = 1$, so that $f_{V(\beta)}(v(\beta)) \geq \tau_n/4$ for all $\beta \in B$ whenever $\hat{t}_n^{(X,\hat{V})}(\tilde{x}) = \hat{t}_n^{(X_2,Z)}(\tilde{x}) = 1$;

(ii) for $\hat{t}_n^{(X_2,Z)}(\tilde{x}) = 1$ we have $\hat{m}_{0,2}(x_2, z) \in \mathcal{M}^\delta$ and hence $(\tilde{m}_0(z), \tilde{m}_2(x_2)) \in \mathcal{M}^\delta$, so that $|\partial\hat{T}_{\hat{V}(\beta)}(\tilde{v}(\beta)) - \partial T_{V(\beta)}(v(\beta))| \leq |\partial\hat{T}_{\hat{V}(\beta)}(\tilde{v}(\beta)) - \partial T_{V(\beta)}(\tilde{v}(\beta))| + C|\hat{v}(\beta) - v(\beta)| \leq ||\partial\hat{T}_{\hat{V}(\beta)} - \partial T_{V(\beta)}||_\infty + C|\hat{v}(\beta) - v(\beta)| \leq C d_n + C d'_n/\tau_n$ for all $\beta \in B$, when $\hat{t}_n^{(X_2,Z)}(\tilde{x}) = 1$.

The above reasoning implies that for n such that $C d'_n/\tau_n < \delta$ and $C \cdot (d_n"/\tau_n + d'_n/\tau_n^2) < 1/2$, over the set $\mathcal{A}_n^{(C)}$ we must have

$$\left|\frac{[\partial\hat{T}_{\hat{V}(\beta)}(\tilde{v}(\beta)) - \partial T_{V(\beta)}(v(\beta))] \cdot [\hat{v}(\beta) - v(\beta)]}{f_{V(\beta)}(v(\beta))^2}\right| \leq C \cdot \left(\frac{d_n}{\tau_n^2} + \frac{d'_n}{\tau_n^3}\right) \cdot \left(\frac{d'_n}{\tau_n}\right)$$

for all $\beta \in B$, whenever $\hat{t}_n^{(X,\hat{V})}(\tilde{x}) = \hat{t}_n^{(X_2,Z)}(\tilde{x}) = 1$. Notice that $\left(\frac{d_n}{\tau_n^2} + \frac{d'_n}{\tau_n^3}\right) \cdot \left(\frac{d'_n}{\tau_n}\right) = o(n^{-1/2})$. Since from Lemma (10), Lemma (12), and Lemma (17) by choosing C large enough $P(\mathcal{A}_n^{(C)})$ can be made arbitrarily close to 1 for any large n, by the above display and Eq. (25) we conclude for point (i).

Point (ii) and (iii) can be proved in essentially the same way. For point (iv), notice that point

(ii) implies that

$$\left|\left|\left[\hat{T}_{\hat{V}(\hat{\beta})}(\hat{v}(\hat{\beta}))-\hat{T}_{\hat{V}(\hat{\beta})}(v(\hat{\beta}))-\partial T_{V(\hat{\beta})}(v(\hat{\beta}))\cdot[\hat{v}(\hat{\beta})-v(\hat{\beta})]\right]\cdot\hat{t}_n^{(X_2,Z)}\right|\right|_\infty=o_P(n^{-1/2}),\tag{26}$$

while from point (iii),

$$\left|\left|[\hat{T}_{\hat{V}(\hat{\beta})}(v(\hat{\beta}))-\hat{T}_{\hat{V}(\hat{\beta})}(v(\beta_0))-\partial T_{V(\beta_0)}(v(\beta_0))\cdot(v(\hat{\beta})-v(\beta_0))]\cdot\hat{t}_n^{(X_2,Z)}\right|\right|_\infty=o_P(n^{-1/2})\tag{27}$$

Now, by the Mean-Value Theorem

$$\left|\left|\partial T_{V(\hat{\beta})}(v(\hat{\beta}))-\partial T_{V(\beta_0)}(v(\beta_0))\right|\right|_\infty=\left|\left|\nabla_\beta\partial T_{V(\beta)}(x_1,m_2(x_2),\tilde{\beta})\cdot[\hat{\beta}-\beta]\right|\right|_\infty=O_P(n^{-1/2})\ ,\tag{28}$$

with $\tilde{\beta}\in[\hat{\beta},\beta_0]$. Furthermore, by Lemma (12) and since $\hat{\beta}-\beta_0=O_P(n^{-1/2})$,

$$||[\hat{v}(\hat{\beta})-v(\hat{\beta})]\cdot\hat{t}_n^{(X_2,Z)}||_\infty=o_P(n^{-1/4}),\ ||[(\hat{v}(\hat{\beta})-v(\hat{\beta}))-(\hat{v}(\beta_0)-v(\beta_0))]\cdot\hat{t}_n^{(X_2,Z)}||_\infty=o_P(n^{-1/2})\tag{29}$$

Conclude for point (iv) by replacing Eq. (27), (28), and (29) into Eq. (26). □

.

Remark 26. Define $d_n:=\epsilon_n\tau_n^2n^{-1/4}$ with $\epsilon_n\to 0$, and consider $d_n/\tau_n^{\frac{1}{2-\upsilon}}$, with $\upsilon=d/(1+\underline{Ev}(d)/2)$. For the next lemma, we will need to analyze the convergence of the sequence $d_n/\tau_n^{\frac{1}{2-\upsilon}}$. Start by noticing that $\frac{d_n}{\tau_n^{\frac{1}{2-\upsilon}}}=\frac{\epsilon_n}{(n\cdot\tau_n^{\frac{4}{2-\upsilon}-8})^{1/4}}$. Hence, in order to have $d_n/\tau_n^{\frac{1}{2-\upsilon}}=o(1)$ it is sufficient to show that $n\cdot\tau_n^{\frac{4}{2-\upsilon}-8}\to\infty$. Now, $2-\upsilon=\frac{4+2\cdot(\underline{Ev}(d)-d)}{2+\underline{Ev}(d)}$. If d is even, then $\underline{Ev}(d)=d$, so that $2-\upsilon=\frac{4}{2+\underline{Ev}(d)}$, and hence $n\cdot\tau_n^{\frac{4}{2-\upsilon}-8}=n\cdot\tau_n^{\underline{Ev}(d)-6}$. Conversely, if d is odd, then $\underline{Ev}(d)=d-1$, so that $2-\upsilon=\frac{2}{2+\underline{Ev}(d)}$, and hence $n\cdot\tau_n^{\frac{4}{2-\upsilon}-8}=n\cdot\tau_n^{2\cdot\underline{Ev}(d)-4}$. If Assumption 5(iii) holds, we will have that $n\cdot\tau_n^{\underline{Ev}(d)-6}\to\infty$ and $n\cdot\tau_n^{2\cdot\underline{Ev}(d)-4}\to\infty$, so that $n\cdot\tau_n^{\frac{4}{2-\upsilon}-8}\to\infty$ and hence $\frac{d_n}{\tau_n^{\frac{1}{2-\upsilon}}}=o(1)$.

Lemma 27. *Let Assumptions (1), (2), (3), (4), and (5, iii) hold. Then,*$\mathbb{G}_n\hat{g}^{(2)}_{\hat{V}(\beta_0)}=o_P(1)$*, where* $\hat{g}^{(2)}_{\hat{V}(\beta_0)}(\tilde{x}):=\frac{\hat{T}_{\hat{V}(\beta_0)}(v(\beta_0))-G_{V(\beta_0)}(v(\beta_0))\cdot\hat{f}_{\hat{V}(\beta_0)}(v(\beta_0))}{f_{V(\beta_0)}(v(\beta_0))}t_n^{(X,V)}(\tilde{x})\cdot\varphi(\tilde{x})$*, and* φ*is a fixed function that is uniformly bounded.*

Proof. Let $s_0 := 1+\underline{Ev}(d)/2$, and define the class $\mathcal{G}_n^{(C)} := \Big\{ \tilde{x} \mapsto \frac{g(v(\beta_0) - G_{V(\beta_0)}(v(\beta_0))\tilde{g}(v(\beta_0))}{f_{V(\beta_0)}(v(\beta_0))} \cdot t_n^{(X,V)}(\tilde{x}) \cdot \varphi(\tilde{x}) : g, \tilde{g} \in \mathcal{C}^{s_0} \text{ and } ||g - G_{V(\beta_0)}\tilde{g}||_\infty \leq Cd_n \Big\}$. From the result on the entropy of $\mathcal{C}^{s_0}$ reported at the beginning of this Appendix, $\log N(\varepsilon, \mathcal{C}^{s_0}, ||\cdot||_\infty) \leq C \cdot \varepsilon^{d/s_0}$, with $d/s_0 \in (0,2)$. Consider an ε-cover for the class $\mathcal{C}^{s_0}$, say $\mathcal{A}_{\mathcal{C}^{s_0}}^{(\varepsilon)} := \{g_I : I = 1, .., N(\varepsilon,\ \mathcal{C}^{s_0},\ ||\cdot||_\infty)\}$. For any element of $\mathcal{G}_n^{(C)}$, say $(g - G_{V(\beta_0)} \cdot \tilde{g})/f_{V(\beta_0)}$ with $g,\ \tilde{g} \in \mathcal{C}^{s_0}$, we must have $||g - g_I||_\infty < \varepsilon$ and $||\tilde{g} - g_J||_\infty < \varepsilon$ for some $g_I, g_J \in \mathcal{A}_{\mathcal{C}^{s_0}}^{(\varepsilon)}$. Hence,

$$\Big| \frac{g - G_{V(\beta_0)} \cdot \tilde{g}}{f_{V(\beta_0)}}(v(\beta_0)) \cdot t_n^{(X,V)}(\tilde{x}) \cdot \varphi(\tilde{x}) - \frac{g_I - G_{V(\beta_0)} \cdot g_J}{f_{V(\beta_0)}}(v(\beta_0)) \cdot t_n^{(X,V)}(\tilde{x}) \cdot \varphi(\tilde{x}) \Big| \leq C \frac{t_n^{(X,V)}(\tilde{x})}{f_{V(\beta_0)}(v(\beta_0))} \cdot \varepsilon \ .$$

Define now

$$u_{I,J}(\tilde{x}) := \frac{g_I - G_{V(\beta_0)} \cdot g_J}{f_{V(\beta_0)}}(v(\beta_0)) \cdot t_n^{(X,V)}(\tilde{x}) \cdot \varphi(\tilde{x}) + C \frac{t_n^{(X,V)}(\tilde{x})}{f_{V(\beta_0)}(v(\beta_0))} \cdot \varepsilon$$

and

$$l_{I,J}(\tilde{x}) := \frac{g_I - G_{V(\beta_0)} \cdot g_J}{f_{V(\beta_0)}}(v(\beta_0)) \cdot t_n^{(X,V)}(\tilde{x}) \cdot \varphi(\tilde{x}) - C \frac{t_n^{(X,V)}(\tilde{x})}{f_{V(\beta_0)}(v(\beta_0))} \cdot \varepsilon$$

with $g_I, g_J \in \mathcal{A}_{\mathcal{C}^{s_0}}^{(\varepsilon)}$. From the previous displays, the collection of brackets $\{[u_{I,J},\ l_{I,J}] : I = 1, .., N(\varepsilon, \mathcal{C}^{s_0},\ ||\cdot||_\infty)$ and $J = 1, .., N(\varepsilon, \mathcal{C}^{s_0}, ||\cdot||_\infty)\}$ forms a cover for $\mathcal{G}_n^{(C)}$. For the $L_2(P)$-size of each bracket, $||u_{I,J} - l_{I,J}||_{L_2(P)} \leq C\varepsilon ||t_n^{(X,V)}/f_{V(\beta_0)}||_{L_2(P)} \leq C\varepsilon \Big\{ \int 1\{f_{V(\beta_0)}(w) \geq \tau_n/2\}/f_{V(\beta_0)}(w)\ \mathrm{d}w \Big\}^{1/2} \leq C\varepsilon\tau_n^{-1/2}$, so that by definition of bracketing number, $N_{[]}(C\varepsilon\tau_n^{-1/2}, \mathcal{G}_n^{(C)}, L_2(P)) \leq N(\varepsilon, \mathcal{C}^{s_0}, ||\cdot||_\infty)^2$, and hence

$$\log N_{[]}(\varepsilon, \mathcal{G}_n^{(C)}, L_2(P)) \leq C \cdot (\varepsilon \cdot \tau_n^{1/2})^{-\upsilon} \text{ with } \upsilon = d/s_0 \in (0,2) \tag{30}$$

Now, for each element $\psi \in \mathcal{G}_n^{(C)}$, we must have $||\psi||_{L_2(P)} \leq Cd_n \cdot \Big\{ \int 1\{f_{V(\beta_0)}(w) \geq \tau_n/2\}/f_{V(\beta_0)}(w)\ \mathrm{d}w \Big\}^{1/2} \leq Cd_n/\tau_n^{1/2}$. Finally, for the envelope of $\mathcal{G}_n^{(C)}$, notice that for each $\psi \in \mathcal{G}_n^{(C)}$, $|\psi(\tilde{x})| \leq C \cdot d_n \cdot 1\{f_{V(\beta_0)}(v(\beta_0)) \geq \tau_n/2\}/f_{V(\beta_0)}(v(\beta_0))$. We can therefore apply Lemma (6) with δ replaced by

$Cd_n/\tau_n^{1/2}$ and the envelope F replaced by $C \cdot d_n \cdot 1\{f_{V(\beta_0)}(v(\beta_0)) \geq \tau_n/2\}/f_{V(\beta_0)}(v(\beta_0))$, so that

$$\mathbb{E}||\mathbb{G}||_{\mathcal{G}_n^{(C)}} \leq J_{[]}(Cd_n/\tau_n^{1/2}, \mathcal{G}_n^{(C)}, L_2(P)) + \sqrt{n} \cdot PF\{F > \sqrt{n} a_{\mathcal{G}_n^{(C)}}(Cd_n/\tau_n^{1/2})\} \tag{31}$$

Now, by definition of $J_{[]}$ and by Eq. (30), and since $\upsilon = d/s_0 \in (0, 2)$,

$$J_{[]}(Cd_n/\tau_n^{1/2}, \mathcal{G}_n^{(C)}, L_2(P)) \leq C \int_0^{Cd_n/\tau_n^{1/2}} \sqrt{(\varepsilon \tau_n^{1/2})^{-\upsilon}} \, \mathrm{d}\varepsilon \leq C \Big(\frac{d_n}{\tau_n^{\frac{1}{2-\upsilon}}} \Big)^{\frac{2-\upsilon}{2}} = o(1) \tag{32}$$

where the last equality is obtained from Assumption 5(iii) and Remark (26). From Eq. (30) and the definition of $a_{\mathcal{G}_n^{(C)}}$, we also get $a_{\mathcal{G}_n^{(C)}}(Cd_n/\sqrt{\tau_n})^{-1} \leq Cd_n^{-(1+\upsilon/2)} \tau_n^{1/2}$, so that by using the expression of the envelope F for the class $\mathcal{G}_n^{(C)}$ defined above, $\sqrt{n} \cdot PF\{F > \sqrt{n} a_{\mathcal{G}_n^{(C)}}(Cd_n/\sqrt{\tau_n})\} \leq a_{\mathcal{G}_n^{(C)}}(Cd_n/\sqrt{\tau_n})^{-1} PF^2 \leq Cd_n^{-(1+\upsilon/2)} \tau_n^{1/2} \int \frac{d_n^2}{f_{V(\beta_0)}(w)} 1\{f_{V(\beta_0)}(w) \geq \tau_n/2\} \, \mathrm{d}w \leq C \Big(\frac{d_n}{\tau_n^{\frac{1}{2-\upsilon}}} \Big)^{\frac{2-\upsilon}{2}} = o(1)$. By this latter result, Eq. (32) and (31), we obtain $\mathbb{E}||\mathbb{G}_n||_{\mathcal{G}_n^{(C)}} = o(1)$ for any $C > 0$. Define now the event $\mathcal{A}_n^{(C)} := \{\hat{T}_{\hat{V}(\beta_0)}, \hat{f}_{\hat{V}(\beta_0)}, \in \mathcal{C}^{s_0}, \ ||\hat{T}_{\hat{V}(\beta_0)} - G_{V(\beta_0)} \hat{f}_{\hat{V}(\beta_0)}||_\infty \leq Cd_n\}$. Since $\{|\mathbb{G}_n \hat{g}^{(2)}_{\hat{V}(\beta_0)}| > \varepsilon\} \cap \mathcal{A}_n^{(C)} \subset \{||\mathbb{G}_n||_{\mathcal{G}_n^{(C)}} > \varepsilon\}$, and from Lemma (10) by choosing C large enough $P(\mathcal{A}_n^{(C)})$ can be made arbitrarily close to one for any large n, by Markov's inequality and $\mathbb{E}||\mathbb{G}_n||_{\mathcal{G}_n^{(C)}} = o(1)$ for any $C > 0$ we conclude. □

.

Remark 28. Let Assumption (4) and (5)(iii) hold. Define $d_n := \epsilon_n n^{-1/4} \tau_n^2$, $d_{n,0} := \sqrt{(\log n)/(nh_0^{p_0})} + h_0^{r_0}$, $d_{n,2} := \sqrt{(\log n)/(nh_2^{p_2})} + h_2^{r_2}$, $\tilde{d}_n := d_n \vee d_{n,0} \vee d_{n,2}$, $\eta := 1 + \underline{Ev}(d)/2$, $\eta_0 := 1 + \underline{Ev}(p_0)/2$, $\eta_2 := 1 + \underline{Ev}(p_2)$, $\upsilon := \frac{d}{\eta} \vee \frac{p_0}{\eta_0} \vee \frac{p_2}{\eta_2}$. For the next lemma, we will need that $\frac{\tilde{d}_n}{\tau_n^{\frac{3-\upsilon}{2-\upsilon}}} = o(1)$. From Assumption (4), it is suficient to show that $\frac{\epsilon_n n^{-1/4} \tau_n^2}{\tau_n^{\frac{3-\upsilon}{2-\upsilon}}} = o(1)$, and hence that $n \cdot \tau_n^{4 \cdot \frac{3-\upsilon}{2-\upsilon} - 8} \to \infty$. Now, if $\upsilon = \frac{d}{\eta}$, then $4 \cdot \frac{3-\upsilon}{2-\upsilon} - 8 = \frac{8 \cdot (d-1) - 4 \cdot \underline{Ev}(d)}{4 + 2\underline{Ev}(d) - 2d}$, so that if d is even then $4 \cdot \frac{3-\upsilon}{2-\upsilon} - 8 = \frac{8 \cdot (d-1) - 4 \cdot \underline{Ev}(d)}{4}$, while if d is odd then $4 \cdot \frac{3-\upsilon}{2-\upsilon} - 8 = 4 \cdot (d-1) - 2 \cdot \underline{Ev}(d)$. Hence, it is sufficient that $n \cdot \tau_n^{4 \cdot (d-1) - 2 \cdot \underline{Ev}(d)} \to \infty$ which is ensured by Assumption 5(iii). Thius reasoning also holds for $\upsilon = \frac{p_0}{\eta_0}$ and $\upsilon = \frac{p_2}{\eta_2}$. So, we conclude that $n \cdot \tau_n^{4 \cdot \frac{3-\upsilon}{2-\upsilon} - 8} \to \infty$, and therefore $\frac{\tilde{d}_n}{\tau_n^{\frac{3-\upsilon}{2-\upsilon}}} = o(1)$

Lemma 29. *Let Assumptions (1), (2), (3), (4), (5) hold. Then,* $\mathbb{G}_n \hat{g}^{(3)}_{\hat{V}(\beta_0)} = o_P(1)$, *where* $\hat{g}^{(3)}_{\hat{V}(\beta_0)}(y, \tilde{x}) := \varphi(\tilde{x}) \cdot t_n^{(X,V)}(\tilde{x}) \cdot t_n^{(X_2,Z)}(\tilde{x}) \cdot \frac{\partial_\beta \hat{T}_{\hat{V}(\beta_0)}(\hat{v}(\beta_0)) - \partial_\beta T_{V(\beta_0)}(v(\beta_0))}{f_{V(\beta_0)}(v(\beta_0))} \cdot \varepsilon$ *and* φ *is a fixed function* *that is uniformly bounded.*

The same type of result holds by replacing $\partial_\beta \hat{T}_{\hat{V}(\beta_0)}$ *with* $\partial_1 \hat{T}_{\hat{V}(\beta_0)}$,$\partial_\beta \hat{f}_{\hat{V}(\beta_0)}$ *, or* $\partial_1 \hat{f}_{\hat{V}(\beta_0)}$.

Proof. Define $\eta := 1 + \underline{Ev}(d)/2$, $\eta_0 := 1 + \underline{Ev}(p_0)/2$, $\eta_2 := 1 + \underline{Ev}(p_2)$. Let $\psi := (\psi^{(0)}, \psi^{(2)})$, $\psi_0 := (\psi_0^{(0)}, \psi_0^{(2)})$, $v(\psi, \tilde{x}) := (x_1\beta_{0,1} + \beta_{0,2}\psi^{(2)}(x_2),\ x^e - \psi^{(0)}(z))$, $\mathcal{Z}_n := \{z : f_0(z) > \tau_n/2\}$, $\mathcal{X}_{n,2} := \{x_2 : f_2(x_2) > \tau_n/2\}$ $d_{n,0} := \sqrt{(\log n)/(nh_0^{p_0})} + h_0^{r_0}$, $d_{n,2} := \sqrt{(\log n)/(nh_2^{p_2})} + h_2^{r_2}$, $d_n := \epsilon_n \tau_n^2 n^{-1/4}$, and $\tilde{d}_n := d_{n,0} \vee d_{n,2} \vee d_n$. Define the class of functions $\mathcal{G}_n := \{\tilde{x} \mapsto \varphi(\tilde{x}) \cdot t_n^{(X,V)}(\tilde{x}) \cdot t_n^{(X_2,Z)}(\tilde{x}) \cdot \frac{g(v(\psi,\tilde{x})) - g_0(v(\psi_0,\tilde{x}))}{f_{V(\beta_0)}(v(\beta_0))} : g \in \mathcal{C}^\eta,\ \psi^{(0)} \in \mathcal{C}^{\eta_0}(\mathcal{Z}_n),\ \psi^{(2)} \in \mathcal{C}^{\eta_2}(\mathcal{X}_{n,2}),\ ||[g(v(\psi,\cdot)) - g_0(v(\psi_0,\cdot))]t_n^{(X_2,Z)}||_\infty < C\tilde{d}_n/\tau_n\}$. We will next compute the bracketing entropy of the class $\mathcal{G}_n$, then use a maximal inequality on the empirical process indexed by such a class, and finally obtain the result of the lemma. So, fix $\epsilon > 0$ and consider the ϵ-covers $\mathcal{A}^{(\epsilon)}_{\mathcal{C}^\eta} := \{g_J : I = 1,..,N(\epsilon, \mathcal{C}^\eta, ||\cdot||_\infty)\}$, $\mathcal{A}^{(\epsilon)}_{\mathcal{C}^{\eta_0}(\mathcal{Z}_n)} := \{\psi_J^{(0)} : J = 1,..,N(\epsilon, \mathcal{C}^{\eta_0}(\mathcal{Z}_n), ||\cdot||_\infty)\}$, $\mathcal{A}^{(\epsilon)}_{\mathcal{C}^{\eta_2}(\mathcal{X}_{n,2})} := \{\psi_S^{(2)} : S = 1,..,N(\epsilon, \mathcal{C}^{\eta_2}(\mathcal{X}_{n,2}), ||\cdot||_\infty)\}$ of the classes $\mathcal{C}^\eta$, $\mathcal{C}^{\eta_0}(\mathcal{Z}_n)$, and $\mathcal{C}^{\eta_2}(\mathcal{X}_{n,2})$, respectively. For a generic element of $\mathcal{G}_n$, say $\tilde{x} \mapsto \varphi(\tilde{x}) \cdot t_n^{(X,V)}(\tilde{x}) \cdot t_n^{(X_2,Z)}(\tilde{x}) \cdot \frac{g(v(\psi,\tilde{x})) - g_0(v(\psi_0,\tilde{x}))}{f_{V(\beta_0)}(v(\beta_0))}$ such that $g \in \mathcal{C}^\eta$, $\psi^{(0)} \in \mathcal{C}^{\eta_0}(\mathcal{Z}_n)$, $\psi^{(2)} \in \mathcal{C}^{\eta_2}(\mathcal{X}_{n,2})$, we must have $||g - g_I||_\infty < \epsilon$, $||(\psi^{(0)} - \psi_J^{(0)})t_n^{(X_2,Z)}||_\infty < \epsilon$, $||(\psi^{(2)} - \psi_S^{(2)})t_n^{(X_2,Z)}||_\infty < \epsilon$ for some $g_I \in \mathcal{A}^{(\epsilon)}_{\mathcal{C}^\eta}$, $\psi_J^{(0)} \in \mathcal{A}^{(\epsilon)}_{\mathcal{C}^{\eta_0}(\mathcal{Z}_n)}$, $\psi_S^{(2)} \in \mathcal{A}^{(\epsilon)}_{\mathcal{C}^{\eta_2}(\mathcal{X}_{n,2})}$ Hence,

$$\left|\varphi(\tilde{x}) \cdot t_n^{(X,V)}(\tilde{x}) \cdot t_n^{(X_2,Z)}(\tilde{x}) \cdot \frac{g(v(\psi,\tilde{x})) - g_0(v(\psi_0,\tilde{x}))}{f_{V(\beta_0)}(v(\beta_0))} - \varphi(\tilde{x}) \cdot t_n^{(X,V)}(\tilde{x}) \cdot t_n^{(X_2,Z)}(\tilde{x}) \cdot \frac{g_I(v((\psi_J^{(0)},\psi_S^{(2)}),\tilde{x})) - g_0(v(\psi_0,\tilde{x}))}{f_{V(\beta_0)}(v(\beta_0))}\right| \leq$$
$$\leq C\frac{t_n^{(X,V)}(\tilde{x})}{f_{V(\beta_0)}(v(\beta_0))}\left\{|g(v(\psi,\tilde{x})) - g_I(v(\psi,\tilde{x}))|t_n^{(X_2,Z)}(\tilde{x}) + |g_I(v(\psi,\tilde{x})) - g_I(v((\psi_J^{(0)},\psi_S^{(2)}),\tilde{x}))|t_n^{(X_2,Z)}(\tilde{x})\right\} \leq$$
$$\leq C\frac{t_n^{(X,V)}(\tilde{x})}{f_{V(\beta_0)}(v(\beta_0))}\left\{||g - g_I||_\infty + C||(\psi^{(0)} - \psi_J^{(0)})t_n^{(X_2,Z)}||_\infty + C||(\psi^{(2)} - \psi_S^{(2)})t_n^{(X_2,Z)}||_\infty\right\} \leq C\epsilon\frac{1\{f_{V(\beta_0)}(v(\beta_0)) \geq \tau_n/2\}}{f_{V(\beta_0)}(v(\beta_0))}.$$

Hence, define the two functions

$$\tilde{x} \mapsto u_{I,J,S}(\tilde{x}) := \varphi(\tilde{x}) \cdot t_n^{(X,V)}(\tilde{x}) \cdot t_n^{(X_2,Z)}(\tilde{x}) \cdot \frac{g_I(v((\psi_J^{(0)},\psi_S^{(2)}),\tilde{x})) - g_0(v(\psi_0,\tilde{x}))}{f_{V(\beta_0)}(v(\beta_0))} + C\epsilon\frac{1\{f_{V(\beta_0)}(v(\beta_0)) \geq \tau_n/2\}}{f_{V(\beta_0)}(v(\beta_0))} \text{ and}$$
$$\tilde{x} \mapsto l_{I,J,S}(\tilde{x}) := \varphi(\tilde{x}) \cdot t_n^{(X,V)}(\tilde{x}) \cdot t_n^{(X_2,Z)}(\tilde{x}) \cdot \frac{g_I(v((\psi_J^{(0)},\psi_S^{(2)}),\tilde{x})) - g_0(v(\psi_0,\tilde{x}))}{f_{V(\beta_0)}(v(\beta_0))} - C\epsilon\frac{1\{f_{V(\beta_0)}(v(\beta_0)) \geq \tau_n/2\}}{f_{V(\beta_0)}(v(\beta_0))},$$

with $g_I \in \mathcal{A}^{(\epsilon)}_{\mathcal{C}^\eta}$, $\psi_J^{(0)} \in \mathcal{A}^{(\epsilon)}_{\mathcal{C}^{\eta_0}(\mathcal{Z}_n)}$, and $\psi_S^{(2)} \in \mathcal{A}^{(\epsilon)}_{\mathcal{C}^{\eta_2}(\mathcal{X}_{n,2})}$. From the above two displays, the collection of brackets $\{[u_{I,J,S}, l_{I,J,S}] : I = 1,..,N(\epsilon, \mathcal{C}^\eta, ||\cdot||_\infty)$, $J = 1,..,N(\epsilon, \mathcal{C}^{\eta_0}(\mathcal{Z}_n), ||\cdot||_\infty)$, and $S = 1,..,N(\epsilon, \mathcal{C}^{\eta_2}(\mathcal{X}_{n,2}), ||\cdot||_\infty)\}$ covers the class $\mathcal{G}_n$. For the $L_2(P)$-size of each bracket, by definition of $u_{I,J,S}$ and $l_{I,J,S}$, we have $||u_{I,J,S} - l_{I,J,S}||_{L_2(P)} \leq C\epsilon||1\{f_{V(\beta_0)}(v(\beta_0)) \geq \tau_n/2\}/f_{V(\beta_0)}(v(\beta_0))||_{L_2(P)} \leq C\epsilon/\tau_n^{1/2}$, so that $N_{[]}(C\epsilon/\tau_n^{1/2}, \mathcal{G}_n, ||\cdot||_{L_2(P)}) \leq N(\epsilon, \mathcal{C}^\eta, ||\cdot||_\infty) \cdot N(\epsilon, \mathcal{C}^{\eta_0}(\mathcal{Z}_n), ||\cdot||_\infty) \cdot N(\epsilon, \mathcal{C}^{\eta_2}(\mathcal{X}_{n,2}), ||\cdot$

$||_\infty$), and hence

$$\log N_{[]}(\epsilon, \mathcal{G}_n, L_2(P)) \leq C \cdot (\epsilon \tau_n^{1/2})^{-\upsilon} \text{ with } \upsilon := \frac{d}{\eta} \vee \frac{p_0}{\eta_0} \vee \frac{p_2}{\eta_2} \in (0,2) \tag{33}$$

The envelope of $\mathcal{G}_n$ can be set to the mapping $\tilde{x} \mapsto F(\tilde{x}) := C \cdot (\tilde{d}_n/\tau_n) \cdot 1\{f_{V(\beta_0)}(v(\beta_0)) \geq \tau_n/2\}/f_{V(\beta_0)}(v(\beta_0))$. Finally, notice that for any element $f \in \mathcal{G}_n$, the object $||f||_{L_2(P)}$ can be upperbounded in the same fashion as done with $||u_{I,J} - l_{I,J}||_{L_2(P)}$ above, so that $||f||_{L_2(P)} \leq C\tilde{d}_n/\tau_n^{3/2}$ This implies that we can apply Lemma (6) with $F = C \cdot (\tilde{d}_n/\tau_n) \cdot 1\{f_{V(\beta_0)} \geq \tau_n/2\}/f_{V(\beta_0)}$ and δ set to $C\tilde{d}_n/\tau_n^{3/2}$. Hence,

$$\mathbb{E}||\mathbb{G}_n||_{\mathcal{G}_n} \leq J_{[]}(C\tilde{d}_n/\tau_n^{3/2}, \mathcal{G}_n, L_2(P)) + \sqrt{n}PF\{F > \sqrt{n} \cdot a_{\mathcal{G}_n}(C\tilde{d}_n/\tau_n^{3/2})\} \tag{34}$$

By definition of $J_{[]}$ and Eq. (33),

$$J_{[]}(C\tilde{d}_n/\tau_n^{3/2}, \mathcal{G}_n, L_2(P)) \leq \int_0^{C\tilde{d}_n/\tau_n^{3/2}} \sqrt{C \cdot (\epsilon\tau_n^{1/2})^{-\upsilon}} \, \mathrm{d}\epsilon = C \cdot \left(\frac{\tilde{d}_n}{\tau_n^{\frac{3-\upsilon}{2-\upsilon}}}\right)^{\frac{2-\upsilon}{2}} = o(1) \tag{35}$$

where the last equality follows from Assumption 5 (iii) and Remark (28). On the other hand, from the definition of $a_{\mathcal{G}_n}$and Eq. (33), we get

$$a_{\mathcal{G}_n}(C\tilde{d}_n/\tau_n^{3/2})^{-1} \leq C\tilde{d}_n^{-(1+\upsilon/2)}\tau_n^{3/2+\upsilon/2}$$

so that by definition of the envelope F,

$\sqrt{n}PF\Big\{F > \sqrt{n} \cdot a_{\mathcal{G}_n}\Big(C\frac{\tilde{d}_n}{\tau_n^{3/2}}\Big)\Big\} \leq a_{\mathcal{G}_n}\Big(C\frac{\tilde{d}_n}{\tau_n^{3/2}}\Big)^{-1} PF^2 \leq C\frac{\tau_n^{3/2+\upsilon/2}}{\tilde{d}_n^{(1+\upsilon/2)}}\frac{\tilde{d}_n^2}{\tau_n^2}\int \frac{t_n^{(V(\beta_0))}}{f^2_{V(\beta_0)}}(v(\beta_0))\,\mathrm{d}P(\tilde{x}) \leq$
$\leq C \cdot \Big(\frac{\tilde{d}_n}{\tau_n^{\frac{3-\upsilon}{2-\upsilon}}}\Big)^{\frac{2-\upsilon}{2}} = o(1).$

From the above display, Eq. (35) and Eq. (34), we obtain that $\mathbb{E}||\mathbb{G}_n||_{\mathcal{G}_n} = o(1)$ for any $C > 0$. Now, consider the sum $\sqrt{n}\mathbb{P}_n\hat{g}^{(5)}_{\hat{V}(\beta_0)}$, where $\hat{g}^{(5)}_{\hat{V}(\beta_0)}(y, \tilde{x}) := \varphi(\tilde{x}) \cdot t_n^{(X,V)}(\tilde{x}) \cdot t_n^{(X_2,Z)}(\tilde{x}) \cdot \frac{\partial_\beta \hat{T}_{\hat{V}(\beta_0)}(\hat{v}(\beta_0)) - \partial_\beta T_{V(\beta_0)}(v(\beta_0))}{f_{V(\beta_0)}(v(\beta_0))} \cdot \varepsilon$, and define the event

$\mathcal{A}_n^{(C)} := \Big\{\partial_\beta\hat{T}_{\hat{V}(\beta_0)} \in \mathcal{C}^\eta,\ \hat{m}_0 \in \mathcal{C}^{\eta_0}(\mathcal{Z}_n),\ \hat{m}_2 \in \mathcal{C}^{\eta_2}(\mathcal{X}_{2,n}),\ ||\partial_\beta\hat{T}_{\hat{V}(\beta_0)} - \partial_\beta T_{V(\beta_0)}||_\infty \leq Cd_n,\ ||(\hat{m}_0 - m_0)t_n^{(X_2,Z)}||_\infty \leq Cd_n'/\tau_n,\ ||(\hat{m}_2 - m_2)t_n^{(X_2,Z)}||_\infty \leq Cd_n'/\tau_n\Big\}$. From Lemma (10), Lemma (12) and

Lemma (15), by choosing C large enough $P(\mathcal{A}_n^{(C)})$ can be made arbitrarily close to one for any large n. Therefore, since $\{|\mathbb{G}_n \hat{g}^{(5)}_{\hat{V}(\beta_0)}| > \epsilon\} \cap \mathcal{A}_n^{(C)} \subset \{||\mathbb{G}_n||_{\mathcal{G}_n^{(C)}} > \epsilon\}$ and $\mathbb{E}||\mathbb{G}||_{\mathcal{G}_n^{(C)}} = o(1)$ for any $C > 0$, by Markov's inequality we conclude that $\sqrt{n}\mathbb{P}_n \hat{g}^{(5)}_{\hat{V}(\beta_0)} = o_P(1)$. By following essentially the same reasoning, we also get that $\sqrt{n}\mathbb{P}_n \hat{g}^{(6)}_{\hat{V}(\beta_0)} = o_P(1)$, with $\hat{g}^{(6)}_{\hat{V}(\beta_0)}(y, \tilde{x}) := \varphi(\tilde{x}) \cdot t_n^{(X,V)}(\tilde{x}) \cdot t_n^{(X_2,Z)}(\tilde{x}) \cdot \frac{\partial_1 \hat{T}_{\hat{V}(\beta_0)}(\hat{v}(\beta_0)) - \partial_1 T_{V(\beta_0)}(v(\beta_0))}{f_{V(\beta_0)}(v(\beta_0))} \cdot \varepsilon$. □

.

Lemma 30. *Let Assumptions (1)-(5) hold. Then* $\mathbb{G}_n \hat{g}^{(7)}_{\hat{V}(\beta_0)} = o_P(1)$, *with* $\hat{g}^{(7)}_{\hat{V}(\beta_0)}(y, \tilde{x}) := \varphi(\tilde{x}) \cdot t_n^{(X,V)}(\tilde{x}) \cdot t_n^{(X_2,Z)}(\tilde{x}) \cdot \frac{\nabla_\beta \hat{T}_{\hat{V}(\beta_0)}(x_1, \hat{m}_2(x_2), \hat{v}(\beta_0)) - \nabla_\beta T_{V(\beta_0)}(x_1, m_2(x_2), v(\beta_0))}{f_{V(\beta_0)}(v(\beta_0))} \cdot \varepsilon$, *where* φ *is a fixed function that is uniformly bounded. The same result also holds for* $\nabla_\beta \hat{T}_{\hat{V}(\beta_0)}$ *replaced by* $\nabla_\beta \hat{f}_{\hat{V}(\beta_0)}$.

Proof. First notice that $\nabla_\beta \hat{T}_{\hat{V}(\beta_0)}(x_1, \hat{m}_2(x_2), \hat{v}(\beta_0)) = \partial_\beta \hat{T}_{\hat{V}(\beta_0)}(\hat{v}(\beta_0)) + \partial_1 \hat{T}_{\hat{V}(\beta_0)}(\hat{v}(\beta_0)) \cdot (x_1^T, \hat{m}_2(x_2)^T)^T$ and $\nabla_\beta T_{V(\beta_0)}(x_1, m_2(x_2), v(\beta_0)) = \partial_\beta T_{V(\beta_0)}(v(\beta_0)) + \partial_1 T_{V(\beta_0)}(v(\beta_0)) \cdot (x_1^T, m_2(x_2)^T)^T$, so that

$$\varphi(\tilde{x}) \cdot t_n^{(X,V)}(\tilde{x}) \cdot t_n^{(X_2,Z)}(\tilde{x}) \cdot \frac{\nabla_\beta \hat{T}_{\hat{V}(\beta_0)}(x_1, \hat{m}_2(x_2), \hat{v}(\beta_0)) - \nabla_\beta T_{V(\beta_0)}(x_1, m_2(x_2), v(\beta_0))}{f_{V(\beta_0)}(v(\beta_0))} \cdot \varepsilon =$$

$$= \varphi(\tilde{x}) \cdot t_n^{(X,V)}(\tilde{x}) \cdot t_n^{(X_2,Z)}(\tilde{x}) \cdot \frac{\partial_\beta \hat{T}_{\hat{V}(\beta_0)}(\hat{v}(\beta_0)) - \partial_\beta T_{V(\beta_0)}(v(\beta_0))}{f_{V(\beta_0)}(v(\beta_0))} \cdot \varepsilon +$$

$$+\varphi(\tilde{x}) \cdot t_n^{(X,V)}(\tilde{x}) \cdot t_n^{(X_2,Z)}(\tilde{x}) \cdot \frac{\left[\partial_1 \hat{T}_{\hat{V}(\beta_0)}(\hat{v}(\beta_0)) - \partial_1 T_{V(\beta_0)}(v(\beta_0))\right]}{f_{V(\beta_0)}(v(\beta_0))} \cdot \left[(x_1^T, \hat{m}_2(x_2)^T)^T - (x_1^T, m_2(x_2)^T)^T\right] \cdot \varepsilon +$$

$$+ \quad \varphi(\tilde{x}) \cdot t_n^{(X,V)}(\tilde{x}) \cdot t_n^{(X_2,Z)}(\tilde{x}) \frac{\left[\partial_1 \hat{T}_{\hat{V}(\beta_0)}(\hat{v}(\beta_0)) - \partial_1 T_{V(\beta_0)}(v(\beta_0))\right]}{f_{V(\beta_0)}(v(\beta_0))} \cdot (x_1^T, m_2(x_2)^T)^T \cdot \varepsilon +$$

$$+ \quad \varphi(\tilde{x}) \cdot t_n^{(X,V)}(\tilde{x}) \cdot t_n^{(X_2,Z)}(\tilde{x}) \cdot \frac{\partial_1 T_{V(\beta_0)}(v(\beta_0))}{f_{V(\beta_0)}(v(\beta_0))} \cdot \left[(x_1^T, \hat{m}_2(x_2)^T)^T - (x_1^T, m_2(x_2)^T)^T\right] \cdot \varepsilon =:$$

$$=: \hat{g}^{(8)}(y, \tilde{x}) + \hat{g}^{(9)}(y, \tilde{x}) + \hat{g}^{(10)}(y, \tilde{x}) + \hat{g}^{(11)}(y, \tilde{x})$$

Now, from Lemma (29), we obtain $\mathbb{G}_n\hat{g}^{(8)} = o_P(1)$, $\mathbb{G}_n\hat{g}^{(10)} = o_P(1)$. From Lemma (24) and since $\mathbb{E}\{\varepsilon|\tilde{X}\} = 0$ under H_0,we obtain $\mathbb{G}_n\hat{g}^{(11)} = o_P(1)$. Finally, from Lemma (12), Lemma (17), and Assumption (4) and (5) we get $\mathbb{G}_n\hat{g}^{(9)} = o_P(1)$. □

;

Remark 31. Consider the mapping $w \mapsto \mathbb{E}\{g_n(\tilde{X}) \mid V(\beta_0) = w\}$, where g_n is a sequence of functions uniformly bounded by a fixed constant. Under Assumption (2), since $\tilde{X}|V(\beta_0)$ admits a density $(\tilde{x}, w) \mapsto f_{\tilde{X}|V(\beta_0)}(\tilde{x}|w)$ that is r-times continuously differentiable in w with uniformly bounded derivatives, then also the mapping $w \mapsto \mathbb{E}\{g_n(\tilde{X}) \mid V(\beta_0) = w\} = \int g_n(\tilde{x}) \cdot f_{\tilde{X}|V(\beta_0)}(\tilde{x}|w)\, \mathrm{d}\tilde{x}$ will be r-times continuously differentiable with uniformly bounded derivatives.

.

Lemma 32. *Let Assumptions (1), (2), (3), and (4) hold.*

Define $\psi_n(w) := \mathbb{E}\{t_n^{(X,V)}(\tilde{X}) \cdot \nabla_\beta G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0)) \mid V(\beta_0)\}$. *Then,*

(i) $\sqrt{n}P\frac{\hat{T}_{\hat{V}(\beta_0)} - G_{V(\beta_0)} \cdot \hat{f}_{\hat{V}(\beta_0)}}{f_{V(\beta_0)}} \cdot \psi_n(V(\beta_0)) = \sqrt{n}\mathbb{P}_n\varepsilon\psi_n(V(\beta_0)) + \sqrt{n}\mathbb{P}_n[\varepsilon\partial\psi_n(V(\beta_0)) - (\partial G_{V(\beta_0)} \cdot \psi_n)(V(\beta_0))] \cdot t_n^{(X_2,Z)} \cdot (\hat{V}(\beta_0) - V(\beta_0)) + o_P(1)$;

(ii) $\sqrt{n}P(\hat{T}_{\hat{V}(\hat{\beta})} - G_{V(\beta_0)}\hat{f}_{\hat{V}(\hat{\beta})})(V(\beta_0)) \cdot \phi(\tilde{X}t) = \sqrt{n}\mathbb{P}_n\varepsilon \cdot \iota_t(V(\beta_0)) \cdot f_{V(\beta_0)}(V(\beta_0)) + \sqrt{n}\mathbb{P}_n[\varepsilon \cdot \partial(\iota_t \cdot f_{V(\beta_0)})(V(\beta_0)) - (\partial G_{V(\beta_0)} \cdot \iota_t \cdot f_{V(\beta_0)})(V(\beta_0))] \cdot (\hat{V}(\hat{\beta}) - V(\beta_0)) \cdot \hat{t}_n^{(X_2,Z)} + o_P(1)$ *uniformly in* $t \in T$.

Proof. By definition of $\hat{T}_{\hat{V}(\beta_0)}$ and $\hat{f}_{\hat{V}(\beta_0)}$, the order of K, the usual change of variable typical of kernel regression estimation, and by taking an r-th order Taylor expansion (see Remark (31)),

$$\sqrt{n}P\frac{\hat{T}_{\hat{V}(\beta_0)} - G_{V(\beta_0)} \cdot \hat{f}_{\hat{V}(\beta_0)}}{f_{V(\beta_0)}} \cdot \psi_n(V(\beta_0)) =$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Y_i\hat{t}_{n,i}^{(X_2,Z)}}{h^d}\int K\left(\frac{\hat{V}_i(\beta_0) - w}{h}\right)\psi_n(w)\, \mathrm{d}w - \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{\hat{t}_{n,i}^{(X_2,Z)}}{h^d}\int K\left(\frac{\hat{V}_i(\beta_0) - w}{h}\right) \cdot G_{V(\beta_0)}(w) \cdot \psi_n(w)\, \mathrm{d}w =$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^n [Y_i - G_{V(\beta_0)}(\hat{V}_i(\beta_0))] \cdot \psi_n(\hat{V}_i(\beta_0)) \cdot \hat{t}_{n,i}^{(X_2,Z)} + o_P(1).$$

Define the rates $d_n = \epsilon_n n^{-1/4}\tau_n$, $d_n" = \sqrt{\frac{\log n}{nh_3^q}} + h_3^{r_3}$, $d_{n,0} := \sqrt{\frac{\log n}{nh_0^{p_0}}} + h_0^{r_0}$, $d_{n,2} := \sqrt{\frac{\log n}{nh_2^{p_2}}} + h_2^{r_2}$, $d_n' = d_{n,0} \vee d_{n,2}$. Also, for notational simplicity define $m_{0,2} := (m_0, m_2)$, $\hat{m}_{0,2} := (\hat{m}_0, \hat{m}_2)$, the set $\mathcal{M} := \{m_{0,2}(x_2, z) : x_2 \in \mathcal{X}_2\ ,\ z \in \mathcal{Z}\}$, its δ-enlargment $\mathcal{M}^\delta := \{u : |u - u'| < \delta \text{ for some } u' \in \mathcal{M}\}$, with $\delta > 0$, and the event $\mathcal{A}_n^{(C)} := \{||(\hat{m}_{0,2} - m_{0,2})t_n^{(X_2,Z)}||_\infty \leq Cd_n'/\tau_n\}$. For n large enough, $Cd_n'/\tau_n < \delta$, so that whenever $\hat{t}_n^{(X_2,Z)}(\tilde{x}) = 1$ and the event $\mathcal{A}_n^{(C)}$ holds, we have $\hat{m}_{0,2}(\tilde{x}) \in$

$\mathcal{M}^\delta$. Hence, over the set $\mathcal{A}_n^{(C)}$, by the Mean-Value Theorem, $||[G_{V(\beta_0)}(\hat{v}(\beta_0)) - G_{V(\beta_0)}(v(\beta_0)) - \partial G_{V(\beta_0)}(v(\beta_0)) \cdot (\hat{v}(\beta_0) - v(\beta_0))] \cdot \hat{t}_n^{(X_2,Z)}||_\infty \leq C||[(\hat{v}(\beta_0) - v(\beta_0)]^2 \cdot \hat{t}_n^{(X_2,Z)}||_\infty$ and the same type of expansion holds for ψ_n, mutatis mutandis. From Lemma (12), $||[(\hat{v}(\beta_0) - v(\beta_0)]^2 \cdot \hat{t}_n^{(X_2,Z)}||_\infty = o_P(n^{-1/2})$. Since from Lemma (12) by choosing C large enough, $P(\mathcal{A}_n^{(C)})$ can be made arbitrarily close to one for any large n, we get

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n[Y_i - G_{V(\beta_0)}(\hat{V}_i(\beta_0))] \cdot \psi_n(\hat{V}_i(\beta_0)) \cdot \hat{t}_{n,i}^{(X_2,Z)} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i \psi_n(V_i(\beta_0))\hat{t}_{n,i}^{(X_2,Z)} +$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^n[\varepsilon_i \partial\psi_n(V_i(\beta_0)) - \partial G_{V(\beta_0)}(V_i(\beta_0)) \cdot \psi_n(V_i(\beta_0))] \cdot (\hat{V}_i(\beta_0) - V_i(\beta_0)) \cdot \hat{t}_{n,i}^{(X_2,Z)} + o_P(1).$$

Conclude by Lemma (12) and Lemma (16) so as to obtain point (i). Deduce point (ii) by proceeding in a very similar way. □

.

Lemma 33. *Let Assumptions (1), (2), (3), (4), and (5i) hold. Then,*

(i) $\sqrt{n}\mathbb{P}_n[\hat{G}_{\hat{V}(\beta_0)}(\hat{V}(\beta_0)) - G_{V(\beta_0)}(V(\beta_0))] \cdot \hat{t}_n^{(X_2,Z)}\hat{t}_n^{(X,\hat{V})} \cdot \nabla_\beta \hat{G}_{\hat{V}(\beta_0)}(X_1, \hat{m}_2(X_2), \hat{V}(\beta_0)) =$
$= \sqrt{n}\mathbb{P}_n\varepsilon\psi_n(V(\beta_0)) + \sqrt{n}\mathbb{P}_n\varphi_n(Y, \tilde{X})^T \cdot (\hat{V}(\beta_0) - V(\beta_0)) \cdot t_n^{(X_2,Z)} + o_P(1)$, *where*
$\varphi_n(Y, \tilde{X}) := \varepsilon\partial\psi_n(V(\beta_0)) - \partial G_{V(\beta_0)}(V(\beta_0)) \cdot \psi_n(V(\beta_0)) + \partial G_{V(\beta_0)}(V(\beta_0)) \cdot \nabla_{\beta^T} G_{V(\beta_0)}(V(\beta_0))$,
and $\psi_n(V(\beta_0)) := \mathbb{E}\{t_n^{(X,V)} \cdot \nabla_\beta G_{V(\beta_0)}(V(\beta_0)) \mid V(\beta_0)\}$.

(ii) For any fixed function $\tilde{x} \mapsto \tilde{\varphi}(\tilde{x})$ *uniformly bounded, we have*
$\sqrt{n}\mathbb{P}_n\varepsilon\frac{\hat{G}_{\hat{V}(\beta_0)}(\hat{V}(\beta_0)) - G_{V(\beta_0)}(V_i(\beta_0))}{f_{V(\beta_0)}(V(\beta_0))} \cdot \tilde{\varphi}(\tilde{X}) \cdot \hat{t}_n^{(X_2,Z)}\hat{t}_n^{(X,\hat{V})} = \sqrt{n}\mathbb{P}_n\varepsilon\frac{\partial G_{V(\beta_0)}(V(\beta_0))}{f_{V(\beta_0)}(V(\beta_0))}t_n^{(X_2,Z)} \cdot t_n^{(X,V)} \cdot \tilde{\varphi}(\tilde{X}) \cdot (\hat{V}(\beta_0) - V(\beta_0)) + o_P(1)$

Proof. From Lemma (18) and the decompositions in Eq. (14) and (15), we get $||[\hat{G}_{\hat{V}(\beta_0)}(\hat{v}(\beta_0)) - G_{V(\beta_0)}(v(\beta_0))]\cdot\hat{t}_n^{(X_2,Z)}\hat{t}_n^{(X,\hat{V})}||_\infty = o_P(n^{-1/4})$ and $||[\nabla_\beta\hat{G}_{\hat{V}(\beta_0)}(x_1, \hat{m}_2(x_2), \hat{v}(\beta_0)) - \nabla_\beta G_{V(\beta_0)}(x_1, m_2(x_2), v(\beta_0))] \cdot \hat{t}_n^{(X_2,Z)}\hat{t}_n^{(X,\hat{V})}||_\infty = o_P(n^{-1/4})$. Also, Lemma (18) ensures the same types of result holds by replacing the trimming $\hat{t}_n^{(X_2,Z)}$ with $t_n^{(X_2,Z)}$ and/or $\hat{t}_n^{(X,\hat{V})}$ with $t_n^{(X,V)}$. Hence, from these rates and and Lemma (16), we obtain

$$\sqrt{n}\mathbb{P}_n[\hat{G}_{\hat{V}(\beta_0)}(\hat{V}(\beta_0)) - G_{V(\beta_0)}(V(\beta_0))] \cdot \hat{t}_n^{(X_2,Z)}\hat{t}_n^{(X,\hat{V})} \cdot \nabla_\beta\hat{G}_{\hat{V}(\beta_0)}(\hat{V}(\beta_0)) =$$
$$= \sqrt{n}\mathbb{P}_n[\hat{G}_{\hat{V}(\beta_0)}(\hat{V}(\beta_0)) - G_{V(\beta_0)}(V(\beta_0))] \cdot t_n^{(X_2,Z)}t_n^{(X,V)} \cdot \nabla_\beta G_{V(\beta_0)}(V(\beta_0)) + o_P(1) =: A_{n,1} + o_P(1)$$

By the decomposition in Eq. (14), Lemma (18), and the expansion in Lemma (25) we also get
$$A_{n,1} = \sqrt{n}\mathbb{P}_n\frac{\hat{T}_{\hat{V}(\beta_0)} - G_{V(\beta_0)}\hat{f}_{\hat{V}(\beta_0)}}{f_{V(\beta_0)}}(V(\beta_0)) \cdot t_n^{(X_2,Z)}t_n^{(X,V)}\nabla_\beta G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0)) +$$

$+\sqrt{n}\mathbb{P}_n\nabla_\beta G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0)) \cdot t_n^{(X_2,Z)} t_n^{(X,V)} \partial G_{V(\beta_0)}(V(\beta_0))^T \cdot (\hat{V}(\beta_0) - V(\beta_0)) + o_P(1)$

By Lemma (20) we can apply Lemma (16) so as to drop $t_n^{(X_2,Z)}$ from the first term on the right hand side of the above expression. Similarly, from Lemma (12) we can apply Lemma (16) so as to drop $t_n^{(X,V)}$ from the second term on the right-hand side of the above expression. Then, by Lemma (27) and Lemma (32), we conclude.

(ii) The proof of result (ii) proceeds in a very similar way as for result (i) above. □

.

Lemma 34. *Let Assumptions (1), (2), (3), (4), (5) hold. Then,*

$$\sqrt{n}\mathbb{P}_n\varepsilon\hat{t}_n^{(X,\hat{V})}\hat{t}_n^{(X_2,Z)}\nabla_\beta\hat{G}_{\hat{V}(\beta_0)}(X_1, \hat{m}_2(X_2), \hat{V}(\beta_0)) = \sqrt{n}\mathbb{P}_n\varepsilon\nabla_\beta G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0)) + o_P(1)$$

Proof. By the decomposition in Eq. (15) and Lemma (18), the term $\varepsilon\nabla_\beta\hat{G}_{\hat{V}(\beta_0)}(\hat{V}(\beta_0))$ respects the same kind of condition as $B_{n,i,t}$ in Lemma (16). So, we can replace the trimmings $\hat{t}_n^{(X,\hat{V})}\hat{t}_n^{(X_2,Z)}$ with $t_n^{(X,V)}t_n^{(X_2,Z)}$ and obtain

$$\sqrt{n}\mathbb{P}_n\varepsilon\hat{t}_n^{(X,\hat{V})}\hat{t}_n^{(X_2,Z)}\nabla_\beta\hat{G}_{\hat{V}(\beta_0)}(X_1, \hat{m}_2(X_2), \hat{V}(\beta_0)) = \sqrt{n}\mathbb{P}_n\varepsilon t_n^{(X,V)}t_n^{(X_2,Z)}\nabla_\beta G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0)) +$$
$$\sqrt{n}\mathbb{P}_n\varepsilon t_n^{(X,V)}t_n^{(X_2,Z)}[\nabla_\beta\hat{G}_{\hat{V}(\beta_0)}(X_1, \hat{m}_2(X_2), \hat{V}(\beta_0)) - \nabla_\beta G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0))] + o_P(1).$$

By Lemma (16), we can drop the trimmings in the first term on the right-hand side of the above expression. Let us consider the second term. Lemma (18) and the decomposition in Eq. (15) deliver

$$\sqrt{n}\mathbb{P}_n\varepsilon t_n^{(X,V)}t_n^{(X_2,Z)}[\nabla_\beta\hat{G}_{\hat{V}(\beta_0)}(X_1, \hat{m}_2(X_2), \hat{V}(\beta_0)) - \nabla_\beta G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0))] =$$
$$= \sqrt{n}\mathbb{P}_n\varepsilon t_n^{(X,V)}t_n^{(X_2,Z)}\frac{\nabla_\beta\hat{T}_{\hat{V}(\beta_0)}(\tilde{X}) - G_{V(\beta_0)}(V(\beta_0))\cdot\nabla_\beta\hat{f}_{\hat{V}(\beta_0)}(\tilde{X}) - \nabla_\beta G_{V(\beta_0)}(\tilde{X})\cdot\hat{f}_{\hat{V}(\beta_0)}(\hat{V}(\beta_0))}{f_{V(\beta_0)}(V(\beta_0))} +$$
$$-\sqrt{n}\mathbb{P}_n\varepsilon t_n^{(X,V)}t_n^{(X_2,Z)}\frac{\hat{G}_{\hat{V}(\beta_0)}(\hat{V}(\beta_0)) - G_{V(\beta_0)}(V(\beta_0))}{f_{V(\beta_0)}(V(\beta_0))}\cdot\nabla_\beta\hat{f}_{\hat{V}(\beta_0)}(\tilde{X}) + o_P(1) =: A_{n,1} + A_{n,2} + o_P(1)$$

where with a slight abuse of notation we have denoted

$\nabla_\beta\hat{T}_{\hat{V}(\beta_0)}(X_1, \hat{m}_2(X_2), \hat{V}(\beta_0))$, $\nabla_\beta\hat{f}_{\hat{V}(\beta_0)}(X_1, \hat{m}_2(X_2), \hat{V}(\beta_0))$, and $\nabla_\beta G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0))$, with $\nabla_\beta\hat{T}_{\hat{V}(\beta_0)}(\tilde{X})$, $\nabla_\beta\hat{f}_{\hat{V}(\beta_0)}(\tilde{X})$, and $\nabla_\beta G_{V(\beta_0)}(\tilde{X})$. Notice that $\nabla_\beta T_{V(\beta_0)} = \nabla_\beta G_{V(\beta_0)}\cdot f_{V(\beta_0)} + G_{V(\beta_0)}\cdot\nabla_\beta f_{V(\beta_0)}$. So, by Lemma (30) $A_{n,1} = o_P(1)$. For the second term, from Lemma (18) and Lemma (17),

$$A_{n,2} = \sqrt{n}\mathbb{P}_n\varepsilon t_n^{(X,V)}t_n^{(X_2,Z)}\frac{\hat{G}_{\hat{V}(\beta_0)}(\hat{V}(\beta_0)) - G_{V(\beta_0)}(V(\beta_0))}{f_{V(\beta_0)}(V(\beta_0))}\cdot\nabla_\beta f_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0)) + o_P(1).$$

By Lemma (33) and the above axpression we get

$$A_{n,2} = \sqrt{n}\mathbb{P}_n\varepsilon \cdot \nabla_\beta f_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0)) \cdot t_n^{(X_2,Z)} \frac{\partial G_{V(\beta_0)}(V(\beta_0))^T}{f_{V(\beta_0)}(V(\beta_0))} \cdot (\hat{V}(\beta_0) - V(\beta_0)) + o_P(1)$$

Conclude by Lemma (24) and $\mathbb{E}\{\varepsilon|\tilde{X}\} = 0$. ☐

.

Lemma 35. *Let Assumption (2) hold. Then,*

(i) $\psi_n(w) \to 0$ *and* $\partial\psi_n(w)^T \to 0$ *for all* $w \in$ *Supp*$(V(\beta_0))$*, where* $\psi_n(w) := \mathbb{E}\{t_n^{(X,V)} \cdot \nabla_\beta G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0)) \mid V(\beta_0) = w\}$*;*

(ii) $\mathbb{E}\{\varphi_n(Y, \tilde{X}) \mid Z = z\} \to \mathbb{E}\{\partial G_{V(\beta_0)}(V(\beta_0)) \cdot \nabla_{\beta^T} G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0)) \mid Z = z\}$ *for all* $z \in$ *Supp*(Z)*, where* $\varphi_n(y, \tilde{x}) := \varepsilon\partial\psi_n(v(\beta_0))^T - \partial G_{V(\beta_0)}(v(\beta_0)) \cdot \psi_n(v(\beta_0))^T + \partial G_{V(\beta_0)}(v(\beta_0)) \cdot \nabla_{\beta^T} G_{V(\beta_0)}(x_1, m_2(x_2), v(\beta_0))$*.*

Proof. Notice first that by the Law of Iterated Expectations, $0 = \mathbb{E}\{1\{f_{X,V}(X,V) = 0\}\} = \mathbb{E}\Big\{\mathbb{E}\{1\{f_{X,V}(X,V) = 0\} \mid V(\beta_0)\}\Big\}$, so that $\mathbb{E}\{1\{f_{X,V}(X,V) = 0\} \mid V(\beta_0) = w\} = 0$ for all $w \in \text{Supp}(V(\beta_0))$. Hence, for any function $\tilde{x} \mapsto \tilde{\varphi}(\tilde{x})$ that is uniformly bounded we must also have $\mathbb{E}\{1\{f_{X,V}(X,V) = 0\} \cdot \varphi(\tilde{X}) \mid V(\beta_0) = w\} = 0$ for all $w \in \text{Supp}(V(\beta_0))$. This implies that $\psi_n(w) = \mathbb{E}\{1\{f_{X,V}(X,V) > 0\} \cdot t_n^{(X,V)} \cdot \nabla_\beta G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0)) \mid V(\beta_0) = w\}$ and $\mathbb{E}\{1\{f_{X,V}(X,V) > 0\} \cdot \nabla_\beta G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0)) \mid V(\beta_0) = w\} = \mathbb{E}\{\nabla_\beta G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0)) \mid V(\beta_0) = w\}$ for all $w \in \text{Supp}(V(\beta_0))$. Since for any $\tilde{x} \in \text{Supp}(\tilde{X})$ $1\{f_{X,V}(x,v) > 0\} \cdot t_n^{(X,V)}(x,v) \cdot \nabla_\beta G_{V(\beta_0)}(x_1, m_2(x_2), v(\beta_0)) \to 1\{f_{X,V}(x,v) > 0\} \cdot \nabla_\beta G_{V(\beta_0)}(x_1, m_2(x_2), v(\beta_0))$, by the Lebesgue Dominated Convergence Theorem $\psi_n(w) \to \mathbb{E}\{\nabla_\beta G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0)) \mid V(\beta_0) = w\}$ for any $w \in \text{Supp}(V(\beta_0))$. From Kline and Spady (1993), $\mathbb{E}\{\nabla_\beta G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0)) \mid V(\beta_0)\} = 0$ P-almost surely, so we conclude that $\psi_n(w) \to 0$ for all $w \in \text{Supp}(V\beta_0))$.

For $\partial\psi_n$, assume wlog that β is unidimensional. Notice that by Lemma (3.6) in Newey & Mcfadden (1994) we can exchange integration and differentiation signs , so that by the same reasoning as above, $\partial\psi_n(w) = \partial \int 1\{f_{X,V}(x,v) > 0\} \cdot t_n^{(X,V)}(x,v) \cdot \nabla_\beta G_{V(\beta_0)}(x_1, m_2(x_2), v(\beta_0)) \cdot f_{X|\tilde{V}(\beta_0)}(\tilde{x}|w) \mathrm{d}\tilde{x} = \int 1\{f_{X,V}(x,v) > 0\} \cdot t_n^{(X,V)}(x,v) \cdot \nabla_\beta G_{V(\beta_0)}(x_1, m_2(x_2), v(\beta_0)) \cdot \partial f_{X|\tilde{V}(\beta_0)}(\tilde{x}|w) \mathrm{d}\tilde{x}$; on the other hand, by a pointwise convergence, $1\{f_{X,V}(x,v) > 0\} \cdot t_n^{(X,V)}(x,v) \cdot \nabla_\beta G_{V(\beta_0)}(x_1, m_2(x_2), v(\beta_0)) \cdot \partial f_{X|\tilde{V}(\beta_0)}(\tilde{x}|w) \to 1\{f_{X,V}(x,v) > 0\} \cdot \nabla_\beta G_{V(\beta_0)}(x_1, m_2(x_2), v(\beta_0)) \cdot \partial f_{X|\tilde{V}(\beta_0)}(\tilde{x}|w)$, so that by the Lebesgue Dominated Convergence Theorem, $\nabla\psi_n(w) \to \int 1\{f_{X,V}(x,v) > 0\} \cdot \nabla_\beta G_{V(\beta_0)}(x_1, m_2(x_2), v(\beta_0)) \cdot$

$\partial f_{X|\tilde{V}(\beta_0)}(\tilde{x}|w)\mathrm{d}\tilde{x} = \partial \int 1\{f_{X,V}(x,v) > 0\} \cdot \nabla_\beta G_{V(\beta_0)}(x_1, m_2(x_2), v(\beta_0)) \cdot f_{X|\tilde{V}(\beta_0)}(\tilde{x}|w)\mathrm{d}\tilde{x}$ for all $w \in \mathrm{Supp}(V(\beta_0))$, where the last inequality follows from exchanging the integral and derivation signs again thanks to Lemma (3.6) in Newey & Mcfadden (1994). Now, notice that by the same reasoning as above $\int 1\{f_{X,V}(x,v) > 0\} \cdot \nabla_\beta G_{V(\beta_0)}(x_1, m_2(x_2), v(\beta_0)) \cdot f_{X|\tilde{V}(\beta_0)}(\tilde{x}|w)\mathrm{d}\tilde{x} = \mathbb{E}\{\nabla_\beta G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0)) \mid V(\beta_0) = w\} = 0$ for all $w \in \mathrm{Supp}(V(\beta_0))$, so we conclude for point (i).

For point (ii), notice that by point (i), $\varphi_n(y,\tilde{x}) \to \partial G_{V(\beta_0)}(v(\beta_0)) \cdot \nabla_{\beta^T} G_{V(\beta_0)}(x_1, m_2(x_2), v(\beta_0))$ for all $(y,\tilde{x}) \in \mathrm{Supp}(Y,\tilde{X})$, so that by a Lebesgue Dominated Convergence Theorem, $\mathbb{E}\{\varphi_n(Y,\tilde{X}) \mid Z = z\} \to \mathbb{E}\{\partial G_{V(\beta_0)}(V(\beta_0)) \cdot \nabla_{\beta^T} G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0)) \mid Z = z\}$ for all $z \in \mathrm{Supp}(Z)$. □

.

Lemma 36. *Let Assumption (1), (2), (3), and (4) hold. Then,*

(i) $\sqrt{n}\mathbb{P}_n \varepsilon \psi_n(V(\beta_0)) = o_P(1)$, *with where* $\psi_n(w) := \mathbb{E}\{t_n^{(X,V)} \cdot \nabla_\beta G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0)) \mid V(\beta_0) = w\}$

(ii) $\sqrt{n}\mathbb{P}_n \varphi_n(Y,\tilde{X})^T \cdot (\hat{V}(\beta_0) - V(\beta_0)) \cdot t_n^{(X_2,Z)} = \sqrt{n}\mathbb{P}_n \mathbb{E}\{\nabla_\beta G_{V(\beta_0)}(V(\beta_0)) \cdot \partial G_{V(\beta_0)}(V(\beta_0))^T \mid Z\} \cdot (\beta_{0,2}^T \cdot (D_2 - m_2(X_2)), -(D_0 - m_0(Z))^T)^T + o_P(1)$, *with* $\varphi_n(y,\tilde{x}) := \varepsilon \partial \psi_n(v(\beta_0))^T - \partial G_{V(\beta_0)}(v(\beta_0)) \cdot \psi_n(v(\beta_0))^T + \partial G_{V(\beta_0)}(v(\beta_0)) \cdot \nabla_{\beta^T} G_{V(\beta_0)}(x_1, m_2(x_2), v(\beta_0))$

Proof. (i) Assume wlog that β is unidimensional. Define $U_{n,i} := \varepsilon_i \psi_n(V_i(\beta_0))/\sqrt{n}$. We will apply the Liendberg-Feller CLT to show that $\sum_{i=1}^n U_{n,i}$converges to a degenerate normal centered at 0. First, $\sum_{i=1}^n \mathbb{E}|U_{n,i}|\{|U_{n,i}| > \epsilon\} \to 0$ for all $\epsilon > 0$. From the previous Lemma, $\psi_n(w)^2 \to 0$ for all $w \in \mathrm{Supp}(V(\beta_0))$, so by the Lebesgue Dominated Convergence Theorem, $\sum_{i=1}^n \mathrm{Cov}(U_{n,i}) = \mathbb{E}\{\mathbb{E}\{\varepsilon^2|V(\beta_0)\} \cdot \psi_n(V(\beta_0))^2\} \to 0$. From the Liendberg Feller CLT, we therefore obtain $\sum_{i=1}^n U_{n,i} \rightsquigarrow \mathcal{N}(0,0)$, so we conclude for point (i).

(ii)

$$\sqrt{n}\mathbb{P}_n \varphi_n(Y,\tilde{X})^T \cdot (\hat{V}(\beta_0) - V(\beta_0)) \cdot t_n^{(X_2,Z)} =$$

$$= \sqrt{n}\mathbb{P}_n \varphi_n^{(1)}(Y,\tilde{X})^T \cdot \beta_{0,2}^T \cdot (\hat{m}_2(X_2) - m_2(X_2)) \cdot t_n^{(X_2,Z)} - \sqrt{n}\mathbb{P}_n \varphi_n^{(v)}(Y,\tilde{X})^T \cdot (\hat{m}_0(Z) - m_0(Z)) \cdot t_n^{(X_2,Z)} \quad (36)$$

where $\varphi_n^{(1)}(Y,\tilde{X}) := \varepsilon\partial_1\psi_n(v(\beta_0))^T - \partial_1 G_{V(\beta_0)}(v(\beta_0))\cdot\psi_n(v(\beta_0))^T + \partial_1 G_{V(\beta_0)}\cdot\nabla_{\beta^T}G_{V(\beta_0)}$ and $\varphi_n^{(v)}(Y,\tilde{X}) := \varepsilon\partial_v\psi_n(v(\beta_0))^T - \partial_v G_{V(\beta_0)}(v(\beta_0))\cdot\psi_n(v(\beta_0))^T + \partial_v G_{V(\beta_0)}\cdot\nabla_{\beta^T}G_{V(\beta_0)}$, and for notational simplicity we have replaced $\partial_1 G_{V(\beta_0)}(v(\beta_0))\cdot\nabla_{\beta^T}G_{V(\beta_0)}(x_1, m_2(x_2), v(\beta_0))$ by $\partial_1 G_{V(\beta_0)}\cdot\nabla_{\beta^T}G_{V(\beta_0)}$

Assume wlog that β, x^e, and d_0 are unidimensional, so that also $x^e - m_0(z)$ will be so. Let us analyze the second term on the right hand side of Eq. (36). From Lemma (12) and Lemma (16) we can replace the trimming $t_n^{(X_2,Z)}$ with the trimming $t_n^{(Z)}$, so that by Lemma (24),

$$\sqrt{n}\mathbb{P}_n\varphi_n^{(v)}(Y,\tilde{X})^T\cdot(\hat{m}_0(Z)-m_0(Z))\cdot t_n^{(X_2,Z)} = \sqrt{n}\mathbb{P}_n\tilde{\psi}_n + o_P(1)\ ,$$

where

$$\tilde{\psi}_n(\tilde{X}) := \int \frac{\mathbb{E}\{\varphi_n^{(v)}(Y,\tilde{X}) \mid Z=z\}\cdot t_n^{(Z)}(z)}{f_0(z)h_0^{p_0}}\cdot(D_0 - m_0(z))\cdot K_0\Big(\frac{z-Z}{h_0}\Big)\ \mathrm{d}P(z)\ .$$

By Tonelli-Fubini's Theorem, the Law of Iterated Expectations, an r-th order Taylor expansion, and the r-th order of the kernel,

$$P\tilde{\psi}_n = \int \frac{\mathbb{E}\{\varphi_n^{(v)}(\tilde{X}) \mid Z=z\}t_n^{(Z)}(z)}{h_0^{p_0}}\Big\{\int \cdot(\tilde{d}-m_0(z))\cdot K_0\Big(\frac{z-\tilde{z}}{h_0}\Big)\ \mathrm{d}P(\tilde{d},\tilde{z})\Big\}\mathrm{d}z = O(h_0^{r_0})\ ,$$

so $\sqrt{n}\mathbb{P}_n\tilde{\psi}_n = \mathbb{G}_n\tilde{\psi}_n + o_P(1)$, by the rates in Assumption 4 (ii). Now,
$\mathbb{G}_n\tilde{\psi}_n = \mathbb{G}_n\int h_0^{-p_0}\cdot(D_0-m_0(z))\cdot\mathbb{E}\{\partial_v G_{V(\beta_0)}\nabla_\beta G_{V(\beta_0)} \mid Z=z\}\cdot K_0\Big(\frac{Z-z}{h_0}\Big)\ \mathrm{d}z +$
$+\mathbb{G}_n\int\mathbb{E}\{\varphi_n^{(v)}(Y,\tilde{X}) - \partial_v G_{V(\beta_0)}\nabla_\beta G_{V(\beta_0)} \mid Z=z\}h_0^{-p_0}\cdot(D_0-m_0(z))\cdot K_0\Big(\frac{Z-z}{h_0}\Big)\ \mathrm{d}z +$
$+\mathbb{G}_n\int\mathbb{E}\{\varphi_n^{(v)}(Y,\tilde{X}) \mid Z=z\}h_0^{-p_0}\cdot(t_n^{(Z)}(z)-1)\cdot(D_0-m_0(z))\cdot K_0\Big(\frac{Z-z}{h_0}\Big)\ \mathrm{d}z := A_{n,1}+A_{n,2}+A_{n,3}$.

By the ususal change of variable, Taylor expansion, the r_0-th order of the kernel, and Assumption 4(ii), $A_{n,1} = \sqrt{n}\mathbb{P}_n(D_0 - m_0(Z))\cdot\mathbb{E}\{\partial_v G_{V(\beta_0)}\cdot\nabla_\beta G_{V(\beta_0)} \mid Z\} + o_P(1)$, and $A_{n,2} = \mathbb{G}_n(D_0-m_0(Z))\cdot\mathbb{E}\{\varphi_n^{(v)}(Y,\tilde{X}) - \partial_v G_{V(\beta_0)}\nabla_\beta G_{V(\beta_0)} \mid Z\} + o_P(1)$. By using Lemma (35) and by applying the Liendberg Feller CLT in a similar way as in point (i), we get $A_{n,2} = o_P(1)$. Finally, for $A_{n,3}$, by the ususal change of variable, $A_{n,3} = \mathbb{G}_n\int\mathbb{E}\{\varphi_n^{(v)}(Y,\tilde{X}) \mid Z = \tilde{Z}+uh_0\}\cdot(t_n^{(Z)}(\tilde{Z}+uh_0)-1)\cdot(D_0 - m_0(\tilde{Z}+uh_0))\cdot K_0(u)\ \mathrm{d}u$. Define $U_{n,i} := n^{-1/2}\int\mathbb{E}\{\varphi_n^{(v)}(Y,\tilde{X}) \mid Z = Z_i + uh_0\}\cdot(t_n^{(Z)}(Z_i+uh_0)-1)\cdot(D_{0,i}-m_0(Z_i+uh_0))\cdot K_0(u)\ \mathrm{d}u$. We will again apply the Liendberg Feller CLT. Notice that $\sum_{i=1}^n \mathbb{E}|U_{n,i}|\{|U_{n,i}| > \epsilon\} \to 0$ for all $\epsilon > 0$. On the other hand, $\sum_{i=1}^n \mathrm{Cov}(U_{n,i}) \leq C \int 1\{f(z) > 0\} \cdot \left\{ \int |t_n^{(Z)}(z + uh_0) - 1| \cdot |K_0(u)| \, du \right\}^2 f_0(z) \, dz$. Since for all z such that $f_0(z) > 0$, $|t_n^{(Z)}(z + uh_0) - 1| \to 0$, by using the Lebesgue Dominated Convergence Theorem we get $\sum_{i=1}^n \mathrm{Cov}(U_{n,i}) \to 0$. So, by the Liendberg Feller CLT, $\sum_{i=1}^n U_{n,i} = o_P(1)$, so that $A_{n,3} = o_P(1)$. We therefore obtain

$$\sqrt{n}\mathbb{P}_n \varphi_n^{(v)}(Y, \tilde{X})^T \cdot (\hat{m}_0(Z) - m_0(Z)) \cdot t_n^{(X_2,Z)} = \sqrt{n}\mathbb{P}_n(D_0 - m_0(Z)) \cdot \mathbb{E}\{\partial_v G_{V(\beta_0)} \cdot \nabla_\beta G_{V(\beta_0)} \mid Z\} + o_P(1)$$

The first term on the right hand side of Eq. (36) can be handled in essentially the same way. So, we conclude. □

.

Lemma 37. *Let Assumptions (1), (2), (3), (4), (6) hold. Then,* $\hat{\beta} = \beta^* + o_P(1)$.

Proof. First, notice that

$\hat{Q}_n(\beta) := \mathbb{P}_n[Y - \hat{G}_{\hat{V}(\beta)}(\hat{V}(\beta))]^2 \hat{t}_n^{(X_2,Z)} \hat{t}_n^{(X,\hat{V})} =$

$= \mathbb{P}_n[Y - G_{V(\beta)}(V(\beta))]^2 \hat{t}_n^{(X_2,Z)} \hat{t}_n^{(X,\hat{V})} - 2\mathbb{P}_n Y[\hat{G}_{\hat{V}(\beta)}(\hat{V}(\beta)) - G_{V(\beta)}(V(\beta))] \hat{t}_n^{(X_2,Z)} \hat{t}_n^{(X,\hat{V})} +$

$+\mathbb{P}_n[\hat{G}_{\hat{V}(\beta)}(\hat{V}(\beta)) - G_{V(\beta)}(V(\beta))]^2 \cdot \hat{t}_n^{(X_2,Z)} \hat{t}_n^{(X,\hat{V})}$.

From Lemma (16), $\mathbb{P}_n[Y - G_{V(\beta)}(V(\beta))]^2 \hat{t}_n^{(X_2,Z)} \hat{t}_n^{(X,\hat{V})} = \mathbb{P}_n[Y - G_{V(\beta)}(V(\beta))]^2 + o_P(1)$ uniformly in $\beta \in B$; from the differentiability of $(\beta, \tilde{x}) \mapsto G_{V(\beta)}(v(\beta))$ ensured by Assumption 2, Example 19.34 in Vaart (1998b), and Glivenk-Cantelli's Theorem, $\mathbb{P}_n[Y - G_{V(\beta)}(V(\beta))]^2 = P[Y - G_{V(\beta)}(V(\beta))]^2 + o_P(1)$ uniformly in $\beta \in B$. On the other hand, from Lemma (18) and the decomposition in Eq. (14), $\mathbb{P}_n Y[\hat{G}_{\hat{V}(\beta)}(\hat{V}(\beta)) - G_{V(\beta)}(V(\beta))] \hat{t}_n^{(X_2,Z)} \hat{t}_n^{(X,\hat{V})} = o_P(1)$ unifomly in $\beta \in B$, and from a similar argument, $\mathbb{P}_n[\hat{G}_{\hat{V}(\beta)}(\hat{V}(\beta)) - G_{V(\beta)}(V(\beta))]^2 \cdot \hat{t}_n^{(X_2,Z)} \hat{t}_n^{(X,\hat{V})} = o_P(1)$ unifomly in $\beta \in B$. Hence, $\sup_{\beta \in B} |\hat{Q}_n(\beta) - Q(\beta)| = o_P(1)$. Conclude by Theorem (2.1) in Newey & Mcfadden (1994). □

.

Lemma 38. *Let Assumptions (1), (2), (3), (4), (5), and (6) hold. Then,*

$\sqrt{n} \cdot (\hat{\beta} - \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i \cdot \Sigma_{\beta_0}^{-1} \cdot \nabla_\beta G_{V(\beta)}(m, \beta_0, \tilde{X}_i) +$

$-\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{E}\{\Sigma_{\beta_0}^{-1}\cdot\nabla_\beta G_{V(\beta)}(m,\beta_0,\tilde{X})\cdot\partial_1 G_{V(\beta_0)}(V(\beta_0)) \mid X_2 = X_{2,i}\}\cdot\beta_{0,2}^T\cdot(D_{i,2}-m_{i,2})+$
$\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{E}\{\Sigma_{\beta_0}^{-1}\cdot\nabla_\beta G_{V(\beta)}(m,\beta_0,\tilde{X}_i)\cdot\partial_v G_{V(\beta_0)}(V(\beta_0))^T \mid Z = Z_i\}\cdot(D_{0,i}-m_{0,i}) + o_P(1),$
where $\Sigma_{\beta_0} := \mathbb{E}\{\nabla_\beta G_{V(\beta)}(m,\beta_0,\tilde{X})\cdot\nabla_{\beta^T} G_{V(\beta)}(m,\beta_0,\tilde{X})\}$.

Proof. Since by definition $\hat{\beta} = \arg\max_{\beta\in B}\hat{Q}_n(\beta)$, by taking the first order conditions,

$$\sqrt{n}\mathbb{P}_n[Y-\hat{G}_{\hat{V}(\hat{\beta})}(\hat{V}(\hat{\beta}))]\hat{t}_n^{(X_2,Z)}\hat{t}_n^{(X,\hat{V})}\nabla_\beta\hat{G}_{\hat{V}(\hat{\beta})}(\hat{V}(\hat{\beta})) = o_P(1)\ ,$$

so by the Mean-Value Theorem,

$$\hat{\Sigma}_{\tilde{\beta}}\sqrt{n}(\hat{\beta}-\beta_0) = -\sqrt{n}\mathbb{P}_n[Y-\hat{G}_{\hat{V}(\beta_0)}(\hat{V}(\beta_0))]\hat{t}_n^{(X_2,Z)}\hat{t}_n^{(X,\hat{V})}\nabla_\beta\hat{G}_{\hat{V}(\beta_0)}(\hat{V}(\beta_0)) + o_P(1)\ , \tag{37}$$

with

$\hat{\Sigma}_{\tilde{\beta}} := -\mathbb{P}_n(\nabla_\beta\hat{G}_{\hat{V}(\beta)}\nabla_{\beta^T}\hat{G}_{\hat{V}(\beta)})\cdot\hat{t}_n^{(X_2,Z)}\hat{t}_n^{(X,\hat{V})}|_{\beta=\tilde{\beta}} + \mathbb{P}_n[Y-\hat{G}_{\hat{V}(\beta)}(\hat{V}(\beta))]\hat{t}_n^{(X_2,Z)}\hat{t}_n^{(X,\hat{V})}\nabla^2_{\beta\beta^T}\hat{G}_{\hat{V}(\beta)}|_{\beta=\tilde{\beta}}$,

where $\tilde{\beta}\in[\beta_0,\hat{\beta}]$ and for notational simplicity we have replaced $\nabla_\beta\hat{G}_{\hat{V}(\beta)}(X_1,\hat{m}_2(X_2),\hat{V}(\beta))$ with $\nabla_\beta\hat{G}_{\hat{V}(\beta)}$ and $\nabla^2_{\beta\beta^T}\hat{G}_{\hat{V}(\beta)}(X_1,\hat{m}_2(X_2),\hat{V}(\beta))$ with $\nabla^2_{\beta\beta^T}\hat{G}_{\hat{V}(\beta)}$. Notice that by Lemma (37),$\tilde{\beta} = \beta_0 + o_P(1)$. Now, by Lemma (18), the decomposition in Eq. (15), Lemma (16) and Lemma (19), we obtain

$$\hat{\Sigma}_{\tilde{\beta}} = -\mathbb{E}\{\nabla_\beta G_{V(\beta)}\nabla_{\beta^T}G_{V(\beta)}\}|_{\beta=\beta_0} + o_P(1) \tag{38}$$

On the other hand,

$\sqrt{n}\mathbb{P}_n[Y-\hat{G}_{\hat{V}(\beta_0)}(\hat{V}(\beta_0))]\hat{t}_n^{(X_2,Z)}\hat{t}_n^{(X,\hat{V})}\nabla_\beta\hat{G}_{\hat{V}(\beta_0)}(\hat{V}(\beta_0)) = \sqrt{n}\mathbb{P}_n\varepsilon\hat{t}_n^{(X_2,Z)}\hat{t}_n^{(X,\hat{V})}\nabla_\beta\hat{G}_{\hat{V}(\beta_0)}(\hat{V}(\beta_0))+$
$+\sqrt{n}\mathbb{P}_n[G_{V(\beta_0)}(V(\beta_0))-\hat{G}_{\hat{V}(\beta_0)}(\hat{V}(\beta_0))]\hat{t}_n^{(X_2,Z)}\hat{t}_n^{(X,\hat{V})}\nabla_\beta\hat{G}_{\hat{V}(\beta_0)}(\hat{V}(\beta_0)) =: A_{n,1}+A_{n,2}$

From Lemma (34),

$$A_{n,1} = \sqrt{n}\mathbb{P}_n\varepsilon\nabla_\beta G_{V(\beta_0)}(V(\beta_0)) + o_P(1)\ , \tag{39}$$

while from Lemma (33),

$$A_{n,2} = -\sqrt{n}\mathbb{P}_n\varepsilon\psi_n(V(\beta_0)) - \sqrt{n}\mathbb{P}_n\varphi_n(Y,\tilde{X})^T\cdot t_n^{(X_2,Z)}\cdot(\hat{V}(\beta_0)-V(\beta_0)) + o_P(1)$$

with $\psi_n(w) := \mathbb{E}\{t_n^{(X,V)} \cdot \nabla_\beta G_{V(\beta_0)} \mid V(\beta_0) = w\}$ and $\varphi_n(y, \tilde{x}) := \varepsilon \partial \psi_n(v(\beta_0)) - \partial G_{V(\beta_0)}(v(\beta_0)) \cdot \psi_n(v(\beta_0)) + \partial G_{V(\beta_0)}(v(\beta_0)) \cdot \nabla_{\beta^T} G_{V(\beta_0)}$. So, from Lemma (36) and the above expression,

$$A_{n,2} = -\sqrt{n}\mathbb{P}_n \mathbb{E}\{\nabla_\beta G_{V(\beta_0)} \cdot \partial G_{V(\beta_0)}(V(\beta_0))^T \mid Z\} \cdot (\beta_{0,2}^T \cdot (D_2 - m_2(X_2)), -(D_0 - m_0(Z))^T)^T + o_P(1) \quad (40)$$

Conclude from Eq. (40), (39), (38), and (37). □

.

Lemma 39. *Let Assumptions (1)-(6) hold. Then,*

$\sqrt{n}\mathbb{P}_n \hat{g}_{\hat{V}(\hat{\beta}),t} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i \cdot \psi(\tilde{X}_i) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E}\{\psi(\tilde{X}) \cdot \partial_1 G_{V(\beta_0)}(V(\beta_0)) \mid X_2 = X_{2,i}\} \cdot \beta_{0,2}^T \cdot (D_{i,2} - m_{i,2}) + \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E}\{\psi(\tilde{X}) \cdot \partial_v G_{V(\beta_0)}(V(\beta_0))^T \mid Z = Z_i\} \cdot (D_{0,i} - m_{0,i}) + o_P(1)$ *uniformly in* $t \in T$,

where

$\psi(\tilde{x}) := f_{V(\beta_0)}(v(\beta_0)) \cdot \phi_t^\perp(\tilde{x}) - a(t)^T \cdot \Sigma_{\beta_0}^{-1} \cdot \nabla_\beta G_{V(\beta_0)}(m, \beta_0, \tilde{x})$, $\Sigma_{\beta_0} = \mathbb{E}\{\nabla_\beta G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0)) \cdot \nabla_{\beta^T} G_{V(\beta_0)}(X_1, m_2(X_2), V(\beta_0))\}$

and $a(t) := \int \partial_1 G_{V(\beta_0)}(v(\beta_0)) \cdot f_{V(\beta_0)}(v(\beta_0)) \cdot \phi_t^\perp(\tilde{x}) \cdot (x_1^T, m_2(x_2)^T)^T \, dP^{\tilde{X}}(\tilde{x})$

Proof. Notice first that

$\sqrt{n}\mathbb{P}_n \hat{g}_{\hat{V}(\hat{\beta}),t} = \sqrt{n}\mathbb{P}_n \varepsilon \cdot \hat{f}_{\hat{V}(\hat{\beta})}(\hat{V}(\hat{\beta})) \cdot \phi(\tilde{X}t) \cdot \hat{t}_{n,i}^{(X_2,Z)} + \sqrt{n}\mathbb{P}_n[G_{V(\beta_0)}(V(\beta_0)) \cdot \hat{f}_{\hat{V}(\beta_0)}(\hat{V}(\hat{\beta})) - \hat{T}_{\hat{V}(\hat{\beta})}(\hat{V}(\hat{\beta}))] \cdot \phi(\tilde{X}t) \cdot \hat{t}_n^{(X_2,Z)}$.

By the above display, Lemma (38), and Lemma (25), after rearranging term,

$\sqrt{n}\mathbb{P}_n \hat{g}_{\hat{V}(\hat{\beta}),t} = \sqrt{n}\mathbb{P}_n \varepsilon \cdot \hat{f}_{\hat{V}(\hat{\beta})}(V(\beta_0)) \cdot \phi(\tilde{X}t) \cdot \hat{t}_{n,i}^{(X_2,Z)} + \sqrt{n}\mathbb{P}_n[G_{V(\beta_0)}(V(\beta_0)) \cdot \hat{f}_{\hat{V}(\hat{\beta})}(V(\beta_0)) - \hat{T}_{\hat{V}(\hat{\beta})}(V(\beta_0))] \cdot \phi(\tilde{X}t) \cdot \hat{t}_n^{(X_2,Z)} +$

$+\sqrt{n}\mathbb{P}_n[\varepsilon \partial f_{V(\beta_0)}(V(\beta_0)) - \underbrace{\partial T_{V(\beta_0)}(V(\beta_0)) + G_{V(\beta_0)}(V(\beta_0)) \partial f_{V(\beta_0)}(V(\beta_0))}_{=\partial G_{V(\beta_0)}(V(\beta_0)) \cdot f_{V(\beta_0)}(V(\beta_0))}]^T \phi(\tilde{X}t) \cdot \hat{t}_n^{(X_2,Z)} \cdot (V(\hat{\beta}) - V(\beta_0)) +$

$+\sqrt{n}\mathbb{P}_n[\varepsilon \partial f_{V(\beta_0)}(V(\beta_0)) - \underbrace{\partial T_{V(\beta_0)}(V(\beta_0)) + G_{V(\beta_0)}(V(\beta_0)) \partial f_{V(\beta_0)}(V(\beta_0))}_{=\partial G_{V(\beta_0)}(V(\beta_0)) \cdot f_{V(\beta_0)}(V(\beta_0))}]^T \phi(\tilde{X}t) \cdot \hat{t}_n^{(X_2,Z)} \cdot (\hat{V}(\beta_0) - V(\beta_0)) + o_P(1) =:$

$=: B_{n,1,t} + B_{n,2,t} + B_{n,3,t} + B_{n,4,t} + o_P(1)$ uniformly in $t \in T$.

By Lemma (10) and Lemma (16), we can drop the trimming $\hat{t}_n^{(X_2,Z)}$ in $B_{n,1,t}$. Notice that under H_0, $\mathbb{E}\{\varepsilon|\tilde{X}\} = 0$, so that $B_{n,1,t}$ is a centerd process under H_0. Then, Lemma (21) will deliver

$$B_{n,1,t} = \sqrt{n}\mathbb{P}_n\varepsilon \cdot f_{V(\beta_0)}(V(\beta_0)) \cdot \phi(\tilde{X}t) + o_P(1) \text{ uniformly in } t \in T \tag{41}$$

Similarly, by Lemma (10) and Lemma (16), we can drop the trimming $\hat{t}_n^{(X_2,Z)}$ in $B_{n,2,t}$. Then, by Lemma (21) we can replace the operator $\sqrt{n}\mathbb{P}_n$ with the operator $\sqrt{n}P$; hence, Lemma (32) will deliver

$$B_{n,2,t} = -\sqrt{n}\mathbb{P}_n\varepsilon \cdot (f_{V(\beta_0)} \cdot \iota_t)(V(\beta_0)) +$$

$$-\sqrt{n}\mathbb{P}_n[\varepsilon\partial(\iota_t \cdot f_{V(\beta_0)})(V(\beta_0)) - (\partial G_{V(\beta_0)} \cdot \iota_t \cdot f_{V(\beta_0)})(V(\beta_0))] \cdot \hat{t}_n^{(X_2,Z)} \cdot (\hat{V}(\hat{\beta}) - V(\hat{\beta}) + V(\hat{\beta}) - V(\beta_0)) +$$

$$+ o_P(1) \text{ uniformly in } t \in T \tag{42}$$

Now, from Lemma (12) and Lemma (38),

$$||[(\hat{v}(\hat{\beta}) - v(\hat{\beta})) - (\hat{v}(\beta_0) - v(\beta_0))] \cdot \hat{t}_n^{(X_2,Z)}||_\infty = o_P(n^{-1/2}) \tag{43}$$

Eq. (41), (42), (43), together with the definitions of $B_{n,3,t}$ and $B_{n,4,t}$ deliver

$B_{n,1,t} + B_{n,2,t} + B_{n,3,t} + B_{n,4,t} =$

$= \sqrt{n}\mathbb{P}_n\varepsilon \cdot f_{V(\beta_0)}(V(\beta_0)) \cdot \phi_t^\perp(\tilde{X}) + \sqrt{n}\mathbb{P}_n\varphi_t(Y,\tilde{X}) \cdot \hat{t}_n^{(X_2,Z)} \cdot (V(\hat{\beta}) - V(\beta_0)) + \sqrt{n}\mathbb{P}_n\varphi_t(Y,\tilde{X}) \cdot \hat{t}_n^{(X_2,Z)} \cdot (\hat{V}(\beta_0) - V(\beta_0)) + o_P(1)$ uniformly in $t \in T$,

with $\varphi_t(Y,\tilde{X}) := \varepsilon\partial f_{V(\beta_0)}(V(\beta_0)) \cdot \phi(\tilde{X}t) - \varepsilon \cdot \partial(\iota_t \cdot f_{V(\beta_0)})(V(\beta_0)) - \partial G_{V(\beta_0)}(V(\beta_0)) \cdot f_{V(\beta_0)}(V(\beta_0)) \cdot \phi_t^\perp(\tilde{X})$.

Now, by Lemma (16), Lemma (38), and the Glivenko-Cantelli Theorem,

$$\sqrt{n}\mathbb{P}_n\varphi_t(Y,\tilde{X}) \cdot \hat{t}_n^{(X_2,Z)} \cdot (V(\hat{\beta}) - V(\beta_0)) = \mathbb{P}_n\varphi_t(Y,\tilde{X}) \cdot (X_1^T, m_2(X_2)^T) \cdot \sqrt{n}(\hat{\beta} - \beta_0) =$$

$$= a(t)^T \cdot \sqrt{n}(\hat{\beta} - \beta_0) + o_P(1) \text{ unif. in } t \in T$$

By using Lemma (12), Lemma (16), Lemma (22), and Lemma (23), together with the fact that $\mathbb{E}\{\varepsilon|\tilde{X}\} = 0$ under H_0, we get

$$\sqrt{n}\mathbb{P}_n \varphi_t(Y, \tilde{X}) \cdot \hat{t}_n^{(X_2, Z)} \cdot (\hat{V}(\beta_0) - V(\beta_0)) =$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{E}\Big\{\partial_1 G_{V(\beta_0)}(V(\beta_0)) \cdot f_{V(\beta_0)}(V(\beta_0)) \cdot \phi_t^{\perp}(\tilde{X}) \mid X_2 = X_{2,i}\Big\} \cdot \beta_{0,2}^T \cdot (D_{2,i} - m_{2,i}) +$$

$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{E}\Big\{\partial_v G_{V(\beta_0)}(V(\beta_0))^T \cdot f_{V(\beta_0)}(V(\beta_0)) \cdot \phi_t^{\perp}(\tilde{X}) \mid Z = Z_i\Big\} \cdot (D_{0,i} - m_{0,i}) + o_P(1) \text{ uniformly in } t \in T$$

Deduce the claim of the Lemma by the previous three displays and Lemma (38). □