Semi-Parametric Estimation of Ballot Stuffing*

Anastasia Burkovskaya†

April 2020

This paper introduces a structural model of electoral choice and ballot stuffing that allows for derivation of the joint distribution of turnout and voter share from the unobservable joint distribution of the costs of voting and preferences over candidates. We assume availability of a "clean" subset of data that provides us with the following results: if political preferences are heterogenous across the country, we are able to identify the density of ballot stuffing at any polling station of interest. If political preferences are rather homogenous, we are able to identify the bounds on average ballot stuffing. We also provide corresponding semi-parametric estimators, which are based on kernel density estimation. In addition, this paper offers an empirical illustration of the model estimation using the 2011 Russia parliamentary election data.

Keywords: electoral preferences, ballot stuffing, identification, semi-parametric estimation, nonclassical measurement error

1 Introduction

Elections and referenda are useful to make political decisions and to observe the preferences of the population. In the absence of strategic voting, electoral data deliver

*I am deeply indebted to Rosa Matzkin for her guidance and support throughout my work on this paper. I appreciate discussions and comments from Kadir Atalay, Peter Exterkate and Stefano Fiorin. I would also like to thank audiences at the 2019 North American Winter Meeting of Econometric Society, Higher School of Economics, USyd, ITAM, PUC-Chile, Universidad Torcuato di Tella, Universidad de los Andes and UCLA for helpful feedback. I am grateful to Vasily Korovkin for sharing the data on independent observers.

†Office 638 Social Sciences building A02, School of Economics, University of Sydney 2006, NSW, Australia. Email: anastasia.burkovskaya@sydney.edu.au

revealed preferences, and thus are the most suitable data available to study the electorate's preferences. However, in some countries, identification of electoral preferences is complicated by the presence of electoral fraud, especially ballot stuffing – the illegal addition of extra ballots. For example, more recent cases of electoral fraud have been documented in Russia (Enikolopov et al., 2013), Turkey (Aksoy, 2016), and Mozambique (Leeffers and Vicente, 2019). In this paper, we propose a structural model that, coupled together with a "clean" (without electoral fraud) subset of the data, allows for identification of ballot stuffing or its bounds in a polling station, depending on the electoral preference's heterogeneity. We provide semi-parametric estimators of the variables of interest and illustrate the estimation procedure using the 2011 Russia parliamentary election data.

This paper supposes a voter who needs to decide between two candidates. Similar to the probabilistic voting model (Lindbeck and Weibull, 1987; Persson and Tabellini, 2000), we assume an individual's political preferences can be separated into three components: personal, local, and regional. Additionally, we explore voter costs in a manner close to Kawai et al. (2020): a voter, who does not find significant differences between candidates, will not be willing to pay a high cost, such as a long wait in line at a polling station. Thus, people who are close to indifferent between candidates will abstain from elections. In addition, we suppose the local component and costs of voting are the same for all people in one polling station, and regional characteristics affect all residents of that region in the same way. From these assumptions, we derive the values of turnout and voter share in the population. Next, we assume the incumbent government has access to ballot stuffing.[1] In this kind of electoral fraud, the administrators of a polling station place additional ballots in the electoral urn. In addition, they adjust corresponding official lists of participants and records of turnout. As a result, the turnout and the number of votes for the incumbent increase. However, even

[1] The model can be extended to include situations in which both parties have the ability to add extra ballots, but the researcher will need to know which polling station is controlled by which candidate for identification. However, we recognize that even with such an extension, we do not cover all possible situations.

though the voter shares are affected as well, the number of votes for the opposition candidate stays unchanged.

Note the presence of ballot stuffing affects only the polling station variables – the local component and the costs of voting. Hence, the truthful number of votes for the opposition should shed some light on the personal and regional components of preferences. However, the identification might be further complicated by the lack of heterogeneity in the dataset. For example, if the overall support for the opposition is low everywhere across the country, even though the number of votes for the opposition is correct, the range of those values will be very narrow and will prevent identification outside of that range. On the other hand, when the country is more heterogenous in electoral preferences, we will observe a significantly broader range of truthful values, implying better identification results. Consequently, we work with two different cases: heterogenous and homogenous political support. In the ideal case of heterogenous polling stations, the truthful data on the opposition allow for the identification of the distribution of personal preferences and the regional component. In the case of a homogenous country, these data allow for the identification of personal and regional components up to multiplication by a constant. In addition, the identified range of the personal component will be restricted by the range of variability of the opposition support.

Next, to recover some information about ballot stuffing, the existing dataset on voter share and turnout is not enough. Hence, we have to assume the existence of additional information. In this paper, we explore the availability of a small randomized subset of the dataset that is known to be "clean," meaning free of ballot stuffing, in one of the regions. Such a validation sample might be available, for example, from randomized independent observers. If these data exist, then in the case of a heterogenous country, we identify the distribution and mean of ballot stuffing at any polling station in the country. In the case of a homogenous country, and depending on the characteristics of the polling station of interest, we might be able to identify bounds on average ballot stuffing.

Finally, we propose semi-parametric estimators of the model that rely on kernel density estimators and two parametric assumptions. First, we shrink the dimensionality of the regional component by restricting it to be linear in the explanatory variables. The reason for this assumption is that a potentially large set of controls will necessarily result in the curse of dimensionality in the non-parametric estimation. Second, we assume a small validation sample, which is often the case (Schennach, 2016), and thus we have to use a parametric distribution to work with these data. To conclude, we illustrate our estimation procedure using the 2011 Russia parliamentary election data.

The estimation procedure offered in this paper has a number of advantages. First, to our knowledge, this paper is the first to propose a structural electoral model that does not rely on assumptions about preference formation or a spatial model for evaluation of electoral fraud. The preferences are exogenous; however it is possible to extend the model to include a specific preference formation and develop further analysis of the detailed preference components. Second, the analysis in this paper does not rely on any previous electoral data, and hence is suitable not only for regular elections, but also for unique events such as a referendum or for studying electoral boycotts. Third, even though we make two parametric assumptions, the semi-parametric estimation is more robust to mistakes in parametric assumptions used in the fully parametric approaches. Fourth, we are able to make inference about any polling station in the country rather than averages across regions. This fact is particularly useful for officials in charge of guaranteeing free and fair elections, because our analysis might indicate polling stations with questionable results requiring additional attention with potential cancellation of the results and/or re-voting. Finally, our approach can be applied more generally in the evaluation of the nonclassical measurement error in the dependent variable in non-linear models, where measurement error is correlated with the observations.

1.1 Related Literature

First, the paper contributes to the literature on estimation of political preferences. Many papers either study turnout and voter share separately (Degan, 2007; Kernell, 2009; Coate and Conlin, 2004; McMurray, 2013; Merlo and de Paula, 2017), or build the model structure based on a spatial model[2] that assumes each voter has a preferred policy and evaluates candidates based on the distances between proposed and preferred policies (Degan and Merlo, 2011). The main difference in this paper is that we work with turnout and voter share and avoid assumptions about preference formation. We are interested in understanding various components (personal, local, and regional) of voter preferences and voting costs in a single election. This fact is especially important in the cases of atypical elections or referenda. Somewhat similar in this sense to our work is the structural model of voting behavior of Kawai et al. (2020); however, the authors concentrate only on the electoral preferences and use parametric estimation.

Second, our paper belongs to the literature on evaluation of electoral fraud. The existing research explores the following main directions: the analysis of different statistical irregularities in the data, such as Benford's law (Mebane, 2008; Breunig and Goerres, 2011; Skovoroda and Lankina, 2017), unusual kurtosis of the distribution of electoral data (Klimek et al., 2012), and spikes in the distribution of votes (Kobak et al., 2016; Rozenas, 2017); the evaluation of fraud through both natural experiments (Cantu, 2014; Casas et al., 2017) and randomized assignment of independent observers (Enikolopov et al., 2013; Asunka et al., 2019); fitting a parametric model (Levin et al., 2009) and training a machine-learning algorithm (Cantu and Saiegh, 2011). The major differences between this work and the previous approaches include evaluation of the amount of ballot stuffing in any given polling station, keeping minimal parametric assumptions, and avoiding reliance on preference stability over time and previous electoral datasets.

Third, this paper is also related to the nonclassical measurement-error literature,

[2]The model was introduced in Downs (1957) and developed in Riker and Ordeshook (1968) and Hinich and Munger (1996).

because ballot stuffing can be interpreted as a measurement error in the dependent variable that might be correlated with the observations. Also, note ballot stuffing cannot be negative, implying non-zero mean and systematic biases in the observations. The "clean" subsample in this case is an example of a validation sample, placing our work close to the papers using additional data in dealing with nonclassical measurement errors (Bound et al., 1989; Hsiao, 1989; Hausman et al., 1991; Pepe and Fleming, 1991; Carroll and Wand, 1991; Lee and Sepanski, 1995; Chen et al., 2005; Carroll et al., 2010; Katz and Katz, 2010; Hu and Ridder, 2012). The closest in spirit to this work is Hu and Ridder (2012), who use the marginal information from a different dataset to deconvolve the true values of the explanatory variable. Even though "clean" data are subsample of the dataset, in our approach, we do not rely on this fact, and generally similar marginal information (e.g., survey data, exit polls) would suffice. Additionally, in our case, the mismeasured variable is one of the dependent variables, and we allow correlation with the explanatory variables. Our assumptions are weak yet we are able to identify the distributions of interest because we have information on the truthful number of votes for the opposition.

This paper proceeds as follows: Section 2 introduces a model of electoral choice and ballot stuffing and discusses its identification. In section 3, we provide estimators of the model. Section 4 consists of empirical illustration. All derivations of estimators and asymptotics are included in the Appendix.

2 Model and identification

2.1 Electoral model

Two candidates, A (incumbent) and B (opposition), are running for office. Each voter has preferences over candidates: Similarly to probabilistic voting in Persson and Tabellini (2000), voter i in a polling station j in region K chooses candidate A over B

if

$$\sigma_A^{ijK} + \delta_A^{jK} + \mu_A^K > \sigma_B^{ijK} + \delta_B^{jK} + \mu_B^K,$$

where σ_t^{ijK} is a parameter of individual "pure" preferences toward candidate t of voter i, δ^{jK} is the popularity of a candidate in the area of the polling station and is the same for one polling station j, and μ_t^K is a regional effect in the popularity of candidate t and is a function of some observable characteristics of the region X_K, such as average income, level of education, share of old population, etc. Thus, $\mu_t^K = \tilde{h}_t(X_K)$. Note that here we are simply trying to decompose a random variable into three independent components.

To obtain the reduced form of the model, we define parameters of difference in preferences between candidates $\sigma^{ijK} = \sigma_B^{ijK} - \sigma_A^{ijK}$, $\delta^{jK} = \delta_B^{jK} - \delta_A^{jK}$ and $\mu^K = \mu_B^K - \mu_A^K = \tilde{h}_B(X_K) - \tilde{h}_A(X_K) \equiv h(X_K)$. Therefore, voter i in a polling station j in region K chooses candidate A over B if

$$\sigma^{ijK} + \delta^{jK} + h(X_K) < 0,$$

where σ^{ijK} is personal "pure" preference for the candidate B, δ^{jK} and μ_K are polling station j and regional effects on preferences, correspondingly. Moreover, $\mu^K = h(X_K)$, where $h(\cdot)$ is a continuous and monotone in all arguments function.

To include turnout in the model, we follow the empirical evidence that suggests voting costs affect electoral participation (Fujiwara et al., 2016; Leon, 2017). Similar to Kawai et al. (2020), a voter chooses to participate in elections if the difference in her preferences from other candidates is higher than the costs of participation:

$$|\sigma^{ijK} + \delta^{jK} + h(X_K)| \geq c^{jK}.$$

Participation costs c^{jK} are random and the same for all voters in the same polling station, but might be different across different polling stations. Costs might represent the length of the line to vote, the weather, the difficulty of obtaining a voter card, etc.

Such representation of participation implies that if a voter's preferences are close to indifference between the candidates, she does not attend elections. And, by contrast, if a person has very strong preferences toward one candidate or another, she comes to the polling station even when costs are high.[3]

Generally, a polling station represents a small geographical area in which the distance from a voter's home to the polling station is similar for all voters. This fact motivates the assumption about the fixed level of costs of voting at a polling station. We understand that several factors, even weather conditions, might feel different for individual participants. However, the electoral data are aggregated at a polling-station level, and disentangling the personal costs of voting from the preferences without personal-level data or/and strong assumptions about the model structure and behavior would be impossible.

Note the above assumption applies well if people are more or less homogenous in terms of the costs of voting in each polling station, whereas variations can exist across polling stations. For example, different age groups may prefer different suburbs. If the homogeneity does not hold, the personal component of the preferences will account for the unique part of the voting costs for each person. This issue does not cause any problems for electoral fraud estimation. However, one must be careful while evaluating counterfactuals related to change in the costs of voting, because the results might be under-/overstated.

Two points are worth mentioning in the context of this model. First, strategic voting is not a problem in our modelling, because we deal only with two candidates. Second, we do not specify closeness of the electoral outcome as a separate variable in our preferences, because we study a single election. Note the model accounts for this variable as part of the preference components, and if necessary, the model can be extended to include closeness directly in a multiple-election context.

[3]We recognize this model does not account for "marginal voter" thinking, i.e., when voters believe their vote does not matter and abstain from elections. The data show people vote, and we do not attempt to contribute to the divisive question of why.

In addition, we assume individual "pure" preferences σ^{ijK} are independent identically distributed variables with density $g(\cdot)$ and cumulative distribution $G(\cdot)$. Costs of voting c^{jK} and local preferences δ^{jK} are independent identically distributed variables with joint density $f_{\delta,c}(\cdot,\cdot)$. Moreover, we normalize personal component's support to start at 0. This normalization does not affect evaluation of fraud; however, it allows for full identification of other elements of the model and simplifies some derivations.

Assumption 1. *Personal preference σ is independent on X, δ, and c, its support in $\mathbb{R}$ is compact, and it has continuously differentiable density $g(\cdot)$, strictly increasing on the support cumulative distribution function $G(\cdot)$, and the minimum possible value of σ is $\sigma_0 = 0$.*

Assumption 2. *Local preferences δ and costs of voting c have continuously differentiable joint density $f_{\delta,c}(\cdot,\cdot)$, cumulative distribution function $F_{\delta,c}(\cdot,\cdot)$, and are independent on regional characteristics X.*

Next, we introduce "swing voters", σ_A^{jK} and σ_B^{jK}, in every polling station j of region K, who are indifferent between participating and not participating in elections:

$$\sigma_A^{jK} = -\delta^{jK} - \mu^K - c^{jK} \text{ and } \sigma_B^{jK} = \sigma_A^{jK} + 2c^{jK}.$$

Notice people with "pure" preferences $\sigma^{ijK} < \sigma_A^{jK}$ will vote for candidate A, people with $\sigma^{ijK} > \sigma_B^{jK}$ will vote for candidate B, and everybody in between the swing voters will abstain from elections. As a result, the number of people who vote for A in a polling station j in region K is $n_A^{jK} = \int_{-\infty}^{\sigma_A^{jK}} dG(x) = G(\sigma_A^{jK})$. The same number for candidate B is $n_B^{jK} = \int_{\sigma_B^{jK}}^{+\infty} dG(x) = 1 - G(\sigma_B^{jK})$. Thus, turnout in the polling station is $\tau^{jK} = 1 - G(\sigma_B^{jK}) + G(\sigma_A^{jK})$, and A's share of votes is $\pi_A^{jK} = \frac{n_A^{jK}}{n_A^{jK} + n_B^{jK}} = \frac{G(\sigma_A^{jK})}{1 - G(\sigma_B^{jK}) + G(\sigma_A^{jK})}$.

The only data available in any elections are voter share, π_A^{jK}, and turnout, τ^{jK}, across all polling stations j and all regions K. However, the following electoral variables can be recovered from the data:

$$G(\sigma_A^{jK}) = \pi_A^{jK}\tau^{jK}$$
$$G(\sigma_B^{jK}) = 1 - \tau^{jK} + G(\sigma_A^{jK}) = 1 - \tau^{jK} + \pi_A^{jK}\tau^{jK}.$$

In everything that follows, we denote the observable electoral variables $Y = G(\sigma_A^{jK})$ (the number of votes for the incumbent) and $Z = G(\sigma_B^{jK})$ ($1-$ the number of votes for the opposition), whereas vector $X \in \mathbb{R}^L$ is a vector of L regional characteristics.

2.2 Identification of the basic components

This section discusses how to identify unobservable $g(\cdot)$, $h(\cdot)$, and $f_{\delta,c}(\cdot,\cdot)$ from the observable joint distribution of Y and Z conditional on X.

The joint distribution of Y and Z in region K with characteristics X is

$$\begin{aligned} F_{Y,Z|X}(y,z) &= P(G(\sigma_A) < y, G(\sigma_B) < z|X) = P(\sigma_A < G^{-1}(y), \sigma_B < G^{-1}(z)) \\ &= P\left(-\delta - c - h(X) < G^{-1}(y), -\delta + c - h(X) < G^{-1}(z)\right) \\ &= F_{-\delta-c,-\delta+c}\left(h(X) + G^{-1}(y), h(X) + G^{-1}(z)\right), \end{aligned} \tag{1}$$

from which we derive $f_{Y,Z|X}(y,z) = \frac{f_{-\delta-c,-\delta+c}\left(h(X)+G^{-1}(y),h(X)+G^{-1}(z)\right)}{g(G^{-1}(y))g(G^{-1}(z))}$ and $F_{Z|X}(z) = F_{-\delta+c}\left(h(X) + G^{-1}(z)\right)$.

The cumulative distribution function $F_{Y,Z|X}(\cdot,\cdot)$ of electoral variables conditional on the regional characteristics is observable. The right-hand side of equation 1 is completely unknown and is to be identified. Notice the proposed model is not identified without additional assumptions. To see this, consider the existence of true functions $F_{\delta,c}(\cdot)$, $G^{-1}(\cdot)$ and $h(\cdot)$. Suppose $F_{-\delta-c,-\delta+c}(t_1,t_2) = \frac{t_1}{t_2}$; then another functions $\tilde{h}(X) = 2h(X)$ and $\tilde{G}^{-1}(y) = 2G^{-1}(y)$ will generate the same data. Hence, we have to make additional assumptions to achieve the identification.

Assumption 3. *The regional-characteristics function is linear, i.e., $h(X) = \beta' X$.*

Assumption 3 guarantees the identification. Even though a weaker assumption could deliver a similar result, keeping in mind estimation and the fact that the list of observables can be very large, we choose the linear form to shrink the dimensionality.

Assumption 4. *The density of electoral variable Z is continuous and*

a. $f_Z(z) > 0$ *for all* $0 \leq z \leq 1$.

b. $f_Z(z) > 0$ *for all* $z_0 \leq z \leq 1$ *for some* $0 < z_0 < 1$ *and* $f_Z(z) = 0$ *for all* $0 \leq z < z_0$.

Our identification of ballot stuffing will rely on truthful votes for the opposition related to the electoral variable Z. However, the identification depends highly on variation in Z. Assumption 4a addresses the ideal situation when the polling stations in the country are highly heterogenous in their preferences for opposition, and hence provide enough observations for the identification of the regional and personal components without relying on votes for the incumbent. Assumption 4b addresses the situation in which the polling stations are rather homogenous with their degree of homogeneity defined by z_0. If z_0 is close to 0, the polling stations are heterogeneous, whereas when z_0 is close to 1, the polling stations are extremely homogenous. When polling stations are similar to one another, the available observations are too restrictive to identify the entire structure of personal and regional characteristics; hence, only partial identification is possible.

Theorem 1. *Suppose Assumptions 1–3 hold.*

1. *If Assumption 4a holds, functions* $g(\cdot)$ *and* $G^{-1}(\cdot)$ *and coefficients* β *are identified from the joint density* $f_{X,Z}(x, z)$.

2. *If Assumption 4b holds, coefficients* β *are identified up to a multiplication constant, and function* $G^{-1}(z)$ *for all* $z_0 \leq z \leq 1$ *is identified up to an affine transformation.*

Proof. The proof for both parts is the same until a point, where it relies on Assumption 4. Notice $g(\cdot)$ is a derivative of $G(\cdot)$; thus, $\frac{\partial G^{-1}(x)}{\partial x} = \frac{1}{g(G^{-1}(x))}$, and by taking partial derivatives of equation (1), we obtain the following:

$$f_{Z|X}(z) = \frac{f_{-\delta+c}\left(h(X) + G^{-1}(z)\right)}{g(G^{-1}(z))}$$
$$\frac{\partial F_{Z|X}(z)}{\partial X^i} = f_{-\delta+c}\left(h(X) + G^{-1}(z)\right) h^i(X) = g(G^{-1}(z)) f_{Z|X}(z)\beta_i,$$

where $h^i(X) = \beta_i$ denotes partial derivatives with respect to the i-th element. The last equality is available for different values of X and z, so we integrate all the points over the joint density of X at some point z:

$$\begin{aligned} E_X \frac{\partial F_{Z|X}(z)}{\partial X^i} &= \int \frac{\partial F_{Z|X}(z)}{\partial X^i} f_X(X) dX = \beta_i \int g(G^{-1}(z)) f_{Z|X}(z) f_X(X) dX \\ &= \beta_i g(G^{-1}(z)) f_Z(z) \end{aligned} \tag{2}$$

$$\Rightarrow \frac{\beta_i}{\beta_j} = \frac{E_X \frac{\partial F_{Z|X}(z)}{\partial X^i}}{E_X \frac{\partial F_{Z|X}(z)}{\partial X^j}} \Rightarrow \beta_i = \alpha E_X \frac{\partial F_{Z|X}(z)}{\partial X^i}$$

as long as $f_Z(z) > 0$. Now, by plugging β_i back into equation 2, we have

$$E_X \frac{\partial F_{Z|X}(z)}{\partial X^i} = \alpha E_X \frac{\partial F_{Z|X}(z)}{\partial X^i} g(G^{-1}(z)) f_Z(z),$$

implying $g(G^{-1}(z)) = \frac{1}{\alpha f_Z(z)}$ when $f_Z(z) > 0$.

a. Now consider Assumption 4a. Function $g(\cdot)$ is a density, so its integral should be equal to 1. We use this condition to obtain constant α:

$$\int_0^1 g(G^{-1}(z)) dz = \int_0^1 \frac{1}{\alpha f_Z(z)} dz = 1 \Rightarrow \alpha = \int_0^1 \frac{1}{f_Z(z)} dz.$$

Hence, β_i and $g(G^{-1}(z))$ are identified. Next, notice $g(G^{-1}(x))$ is density at the point where cumulative distribution function $G(\cdot)$ takes value x. This fact implies we know the structure of the density, but we are missing the axis:

$$F_Z(z) = \int_0^z f_Z(u)du = \int_0^z \frac{1}{\alpha\phi(u)}du = \int_0^z \frac{1}{\alpha g(G^{-1}(u))}du = \frac{1}{\alpha}\left(G^{-1}(z) - \sigma_0\right).$$

Assumption 1 normalizes $\sigma_0 = 0$, implying $G^{-1}(z) = \alpha F_Z(z)$.

b. Under Assumption 4b, we have $g(G^{-1}(z)) = \frac{1}{\alpha f_Z(z)}$ for all $z_0 \leq z \leq 1$; and otherwise, it is not identified. Hence, we also cannot use the fact that $g(\cdot)$ is a density to obtain constant α, implying we know only the shape of the density, and not its exact values. This reason also leaves us with the fact that only ratios $\frac{\beta_i}{\beta_j}$ can be recovered.

Next, we have

$$F_Z(z) = \int_{z_0}^z f_Z(u)du = \int_{z_0}^z \frac{1}{\alpha\phi(u)}du = \int_{z_0}^z \frac{1}{\alpha g(G^{-1}(u))}du = \frac{1}{\alpha}\left(G^{-1}(z) - G^{-1}(z_0)\right),$$

implying $G^{-1}(z) = \alpha F_Z(z) + G^{-1}(z_0)$, where α and $G^{-1}(z_0)$ are not identifiable constants.

□

2.3 Ballot stuffing

This section discusses how to use the developed model to evaluate the amount of ballot stuffing. Formally, ballot stuffing is defined as the illegal practice of one person submitting multiple ballots during a vote in which only one ballot per person is permitted. In non-democratic elections, ballot stuffing happens when the official staff that runs the voting in a polling station puts a number of filled ballots into the ballot box either before the polls open or after they close. In addition, after the polling station closes, the staff illegally corrects the official turnout to be consistent with the number of votes in the ballot box.

In this paper, we assume only incumbent candidate A has an opportunity to rig the election. In this case, the observables will be affected as follows: Variable Y denotes the number of votes for candidate A in a polling station. If q is a number of additional ballots due to fraud, we will observe variable $Y^f = Y + q$. Variable Z will not be affected, because $1 - Z$ represents the number of votes for candidate B.

The fact that only variable Y is mismeasured implies that by Theorem 1, we can still either partially or fully identify the density of the personal component $g(\cdot)$ and the regional coefficients β. However, the joint distribution of local preferences and costs $f_{\delta,c}(\cdot,\cdot)$ is not identified, because of presence of fraud at the local level. Thus, we need some additional information.

In this paper, we assume the existence of a validation sample in region m that allows for estimation of truthful density $f_{Y,Z|X^m}(y,z)$. Notice that any marginal information that provides such estimation would suffice as well. For concreteness, we suppose availability of a representable subsample of "clean" electoral data. For example, independent observers might be randomized to some polling stations in one region. If their presence helps deter fraud, and as long as randomization is done carefully, the data available from these polling stations in region m will allow estimation of $f_{Y,Z|X^m}(y,z)$. The theorem below introduces the identification results that rely on the existence of such a subsample.

Theorem 2. *Suppose Assumptions 1–3 hold, $q \perp\!\!\!\perp c|X, Z$ and we know $f_{Y,Z|X^m}(y,z)$ in a region with characteristics X^m.*

a. If Assumption 4a holds, the density of of fraud $f_{q|X,Y^f,Z}(q)$ is identified for any $0 < q < 1$ at every polling station.

b. If Assumption 4b holds, the density of $f_{q|X,Y^f,Z}(q)$ is identified only at polling stations with characteristics X^k and electoral variable Z such that

$$\frac{\beta}{\alpha}(X^m - X^k) \le F_Z(Z) \le \frac{\beta}{\alpha}(X^m - X^k) + 1,$$

$$\textit{and for } 0 \le q \le q_{\max},\ \textit{where } q_{\max} = \begin{cases} Y^f - z_0,\ \textit{if } \frac{\beta}{\alpha}(X^k - X^m) \ge 0 \\ Y^f - F_Z^{-1}\left(-\frac{\beta}{\alpha}(X^k - X^m)\right),\ \textit{otherwise} \end{cases}.$$

Proof. Recall that

$$f_{-\delta-c,-\delta+c}\left(\beta X^m + G^{-1}(\tilde{y}), \beta X^m + G^{-1}(\tilde{z})\right) = f_{Y,Z|X^m}(\tilde{y},\tilde{z}) g(G^{-1}(\tilde{y})) g(G^{-1}(\tilde{z})).$$

At the same time, by the same argument, for another region with characteristics X^k, we have

$$f_{Y,Z|X^k}(y,z) = \frac{f_{-\delta-c,-\delta+c}\left(\beta X^k + G^{-1}(y), \beta X^k + G^{-1}(z)\right)}{g(G^{-1}(y)) g(G^{-1}(z))}.$$

Hence, if we set

$$\beta X^k + G^{-1}(y) = \beta X^m + G^{-1}(\tilde{y})$$

$$\beta X^k + G^{-1}(z) = \beta X^m + G^{-1}(\tilde{z}),$$

we can rewrite

$$\begin{aligned} f_{Y,Z|X^k}(y,z) &= \frac{f_{Y,Z|X^m}(\tilde{y},\tilde{z}) g(G^{-1}(\tilde{y})) g(G^{-1}(\tilde{z}))}{g(G^{-1}(y)) g(G^{-1}(z))} \\ &= \frac{f_{Y,Z|X^m}(\tilde{y},\tilde{z}) f_Z(y) f_Z(z)}{f_Z(\tilde{y}) f_Z(\tilde{z})}. \end{aligned}$$

Under Assumption 4a, all components of the right-hand side of the above formula are identified, so $f_{Y,Z|X^k}(y,z)$ is identified as well. Next, the condition $q \perp\!\!\!\perp c|X, Z$ is equivalent to $Y \perp\!\!\!\perp q|X, Z$.[4] In addition, we observe $Y^f = Y + q$ at every polling station, and we have recovered $f_{Y|X^k,Z}(\cdot)$. Hence, the density of ballot stuffing can be obtained as $f_{q|X^k,Y^f,Z}(t) = f_{Y|X^k,Z}(Y^f - t)$.

Under Assumption 4b, we have restrictive identification and are able to work only with the polling stations such that $0 \le y, \tilde{y}, z, \tilde{z} \le 1$.

[4]Note this condition does not imply unconditional independence of Y and q. The variables might still correlate through Z and/or X. For example, the polling stations with the higher support for the opposition (low Z) might have a lower number of votes for the incumbent (low Y) and a greater amount of fraud (high q).

First, recall that $G^{-1}(y) = \alpha F_Z(y) + G^{-1}(z_0)$, and then

$$\beta X^k + G^{-1}(y) = \beta X^m + G^{-1}(\tilde{y}) \Rightarrow \frac{\beta}{\alpha} X^k + F_Z(y) = \frac{\beta}{\alpha} X^m + F_Z(\tilde{y}).$$

Also, $\frac{\beta_i}{\alpha} = E_X \frac{\partial F_{Z|X}(z)}{\partial X^i}$ is identified, implying recoverability of $\tilde{y}$. By analogy, the condition for $\tilde{z}$ is

$$\frac{\beta}{\alpha} X^k + F_Z(z) = \frac{\beta}{\alpha} X^m + F_Z(\tilde{z}),$$

implying $\tilde{z}$ is identified as well.

Finally, $f_{q|X^k,Y^f,Z}(q) = f_{Y|X^k,Z}(Y^f - q)$ is identified as long as the definitions of $\tilde{y}$ and $\tilde{z}$ satisfy the corresponding constraints. For the formal derivation of the constraints on Z and q, see Lemma 1 in the Appendix.

□

2.4 Ballot-stuffing bounds under partial identification

This section analyzes ballot stuffing under Assumption 4b when the density of the ballot stuffing can be obtained only on some interval of the support, making evaluation of the density as well as the average ballot stuffing impossible. In this section, we discuss two bounds on the average ballot stuffing when the polling station of interest satisfies conditions on Z from Theorem 2.

2.4.1 General bound

By the Taylor expansion to function G around $G^{-1}(\tilde{y})$, we get

$$\begin{aligned} E(q|X^k, Y^f, Z) &= Y^f - E\left(G\left(G^{-1}(\tilde{y}) - (X^k - X^m)\beta\right)|\tilde{z}\right) \\ &= Y^f - \mu_{Y|\tilde{z}} + (X^k - X^m)\beta E(g(y^*)|\tilde{z}), \end{aligned}$$

where $\mu_{Y|\tilde{z}} = E(\tilde{y}|\tilde{z})$ and y^* is some midpoint. Recall that $(\tilde{y}, \tilde{z})$ are electoral variables from the clean subsample region with observable distribution, implying $\mu_{Y|\tilde{z}}$ can be directly evaluated when the polling station of interest satisfies restrictions on Z.

If we would like to avoid any additional assumptions on the personal component of the preferences, we are able to establish either a lower or upper bound of the average ballot stuffing. Note $\alpha > 0$, so when $\frac{\beta}{\alpha}(X^k - X^m) < 0$, we can obtain the most basic upper bound on the amount of fraud: $E(q|X^k, Y^f, Z) \leq Y^f - \mu_{Y|\tilde{z}} = q_g$. Similarly, if $\frac{\beta}{\alpha}(X^k - X^m) > 0$, the proposed bound would be the lower bound.

This basic bound does not rely on the density estimator and is available as long as $\tilde{z}$ is identified.

2.4.2 Lower bound

Theorem 2 allows identification of the ballot-stuffing density on the interval $0 \leq q \leq q_{\max}$. When this interval is not an empty set, we can evaluate the following lower bound on the average fraud:

$$\begin{aligned} E(q|X^k, Y^f, Z) = \int_0^1 q f_{q|X^k,Y^f,Z}(q)dq &\geq \int_0^{q_{\max}} q f_{q|X^k,Y^f,Z}(q)dq + P(q \geq q_{\max})q_{\max} \\ &\geq P(q \geq q_{\max})q_{\max} = q_l. \end{aligned}$$

Note $q_{\max}$ is identified. In addition,

$$P(q \geq q_{\max}) = 1 - P(q \leq q_{\max}) = 1 - \int_0^{q_{\max}} f_{q|X^k,Y^f,Z}(q)dq,$$

implying identification of q_l as well.

3 Estimation of the model

In this section, we propose a semi-parametric estimation procedure of the model. We begin by reviewing the standard kernel density estimators we use. Then, we offer a semi-parametric estimator for the regional component. Next, we fit the "clean" data into the normal distribution and obtain semi-parametric estimators of the density and the bounds of the average ballot stuffing.

3.1 Standard kernel estimators

To simplify calculations, suppose all variables are normalized, and the same bandwidth h is chosen for all of them. The standard kernel estimators of the densities can be obtained as follows (see, e.g., Li and Racine (2007)):

$$\hat{f}_Z(z) = \frac{1}{Nh} \sum_{i=1}^{N} k\left(\frac{Z_i - z}{h}\right)$$

$$\hat{f}_{Z|X}(z) = \frac{N_r}{Nh} \frac{\sum_{i=1}^{N} \tilde{k}\left(\frac{X_i - X}{h}\right) k\left(\frac{Z_i - z}{h}\right)}{\sum_{i=1}^{N_r} \tilde{k}\left(\frac{X_i - X}{h}\right)},$$

where N_r is the number of different regions, N is the number of polling stations, $k(\cdot)$ is a univariate kernel, and $\tilde{k}\left(\frac{X_i - X}{h}\right) = \prod_{j=1}^{L} k\left(\frac{X_i^j - X^j}{h}\right)$ is a multiplicative kernel of L regional characteristics. Also, we denote by N_m the size of the validation sample, t_i the number of polling stations in region with characteristics X_i and its expectation μ_t.

3.2 Assumptions

The following assumptions are the standard set in the kernel estimation that is necessary to guarantee asymptotics and the uniform-convergence results in the paper.

Assumption 5. *The sequences* $\{Y_i, Z_i|X_i\}$, $\{Y_i, Z_i|X_i, t_i\}$, $\{X_i, t_i\}$, *and* $\{X_i\}$ *are i.i.d.*

Assumption 6. $f_{X,Y,Z,t}(X, y, z, t)$ *has compact support in* $\mathbb{R}^{L+3}$ *and is continuously differentiable up to the order* s', *for some* $s' > 2$.

Assumption 7. *The kernel function* $k(\cdot)$ *is differentiable of order* $\tilde{s}$, *the derivatives of* k *of order* $\tilde{s}$ *are Lipscitz,* $k(\cdot)$ *vanishes outside a compact set, integrates to 1, and is of order* s'', *where* $\tilde{s} + s'' \leq s'$.

Assumption 8. *As* $N \to \infty$, $\frac{N}{N_r} \to \mu_t < +\infty$, $h \to 0$, $\frac{\ln N}{Nh^{L+4}} \to 0$, $Nh^2 \to \infty$, $Nh^{2+2s''} \to 0$, $\sqrt{Nh^2}\left(\sqrt{\frac{\ln N}{Nh^{L+4}}} + h^{s''}\right)^2 \to 0$ *and* $\sqrt{N_m}\left(\sqrt{\frac{\ln N}{Nh^{L+4}}} + h^{s''}\right)^2 \to 0$.

Assumption 9. $0 < f_{X,Y,Z,t}(X, y, z, t) < \infty$.

3.3 Estimation of $\frac{\beta}{\alpha}$

Notice that by definition of the electoral variable Z, we have

$$Z_i = G\left(-\delta_i + c_i - \beta' X_i\right) \Rightarrow G^{-1}(Z_i) = -\delta_i + c_i - \beta' X_i.$$

Under Assumption 4, we have $G^{-1}(Z_i) = \alpha F_Z(Z_i) + G^{-1}(z_0)$, where $z_0 = 0$ in the case of Assumption 4a. By taking this fact into account, we obtain

$$F_Z(Z_i) = \frac{-\delta_i + c_i}{\alpha} - \frac{\beta'}{\alpha} X_i,$$

implying $\frac{\beta}{\alpha}$ can be estimated by the most basic OLS regression of $\widehat{F_Z(Z_i)}$ on X. Indeed, $\frac{c_i - \delta_i}{\alpha} = \frac{1}{\alpha} E(c - \delta) + \epsilon_i$, and by Assumption 1, we have $\epsilon \perp\!\!\!\perp X$. In addition, the estimation error $e = \widehat{F_Z(Z)} - F_Z(Z)$ is independent on ϵ and X. Hence, this case is the standard case of the classical measurement error in the dependent variable, meaning the OLS delivers consistent estimation. Thus, we have $\widehat{\frac{\beta}{\alpha}} = (X'X)^{-1} X' \widehat{F_Z(Z)}$ and

$$\sqrt{N}\left(\widehat{\frac{\beta}{\alpha}} - \frac{\beta}{\alpha}\right) \xrightarrow{d} N\left(0, \sigma^2_{e+\epsilon}(X'X)^{-1}\right),$$

where $\sigma^2_{e+\epsilon}$ is the variance of the error term $e + \epsilon$.

3.4 "Clean" density estimator

We assume we have access to a small randomized sample of the "clean" data in region m with observable characteristics X^m. We believe the assumption about the small validation sample size is generally very reasonable; however, the estimation procedure can be adapted to any sample size. Below, we also fit the "clean" subsample into the normal distribution. In practice, any distribution can be chosen, and even a non-parametric estimator can be used if the data size permits. The estimation procedure proposed in this paper can still be employed; however, the asymptotic behavior will have to be re-derived.

Hence, due to the limited sample size, we fit the validation sample into the normal distribution, i.e., implying $\begin{pmatrix} Y \\ Z \end{pmatrix} |X = X^m \sim N(\mu, \Sigma)$, where $\mu = \begin{pmatrix} \mu_Y \\ \mu_Z \end{pmatrix}$ and $\Sigma = \begin{bmatrix} \sigma_Y^2 & \sigma_{YZ} \\ \sigma_{YZ} & \sigma_Z^2 \end{bmatrix}$. As a result, we obtain a parametric estimator of the density $\hat{f}_{Y,Z|X^m}(y,z)$:

$$\hat{f}_{Y,Z|X^m}(y,z) = \frac{1}{2\pi|\hat{\Sigma}|^{0.5}} \exp\left\{ -\frac{1}{2} \left(\begin{pmatrix} y \\ z \end{pmatrix} - \hat{\mu} \right)' \hat{\Sigma}^{-1} \left(\begin{pmatrix} y \\ z \end{pmatrix} - \hat{\mu} \right) \right\}.$$

We estimate the following five parameters from the "clean" data, $\theta = (\mu_Y, \mu_Z, \sigma_Y^2, \sigma_Z^2, \sigma_{YZ})'$, and use the standard unbiased estimators for the moments:

$$\hat{\mu}_Y = \frac{1}{N_m} \sum_{i=1}^{N_m} Y_i = \bar{Y} \text{ and } \hat{\mu}_Z = \frac{1}{N_m} \sum_{i=1}^{N_m} Z_i = \bar{Z}$$

$$\hat{\sigma}_Y^2 = \frac{1}{N_m - 1} \sum_{i=1}^{N_m} (Y_i - \bar{Y})^2 \text{ and } \hat{\sigma}_Z^2 = \frac{1}{N_m - 1} \sum_{i=1}^{N_m} (Z_i - \bar{Z})^2$$

$$\hat{\sigma}_{YZ} = \frac{1}{N_m - 1} \sum_{i=1}^{N_m} (Y_i - \bar{Y})(Z_i - \bar{Z}),$$

where N_m is the "clean" sample size. By plugging the estimators of the parameters back into the density, we obtain a consistent and asymptotically normal estimator of $f_{Y,Z|X^m}(y,z)$: $\sqrt{N_k}(\hat{f}_{Y,Z|X=X^m}(y,z) - f_{Y,Z|X=X^m}(y,z)) \xrightarrow{d} N(0, V_{f^m})$, where $V_{f^m} = (\partial G)'\Sigma(\partial G) + (\partial H)'\Omega(\partial H)$ with

$$\partial G = \left(\frac{\partial f_{Y,Z|X=X^m}(y,z)}{\partial \mu_Y}; \frac{\partial f_{Y,Z|X=X^m}(y,z)}{\partial \mu_Z} \right)'$$

$$\partial H = \left(\frac{\partial f_{Y,Z|X=X^m}(y,z)}{\partial \sigma_Y^2}; \frac{\partial f_{Y,Z|X=X^m}(y,z)}{\partial \sigma_Z^2}; \frac{\partial f_{Y,Z|X=X^m}(y,z)}{\partial \sigma_{YZ}} \right)'$$

$$\Omega = \begin{bmatrix} 2\sigma_Y^4 & 2\sigma_{YZ}^2 & 2\sigma_{YZ}\sigma_Y^2 \\ 2\sigma_{YZ}^2 & 2\sigma_Z^4 & 2\sigma_{YZ}\sigma_Z^2 \\ 2\sigma_{YZ}\sigma_Y^2 & 2\sigma_{YZ}\sigma_Z^2 & \sigma_{YZ}^2 + \sigma_Y^2\sigma_Z^2 \end{bmatrix}.$$

See the Appendix for more details on the derivation.

3.5 Ballot stuffing in the region of the "clean" sample

To evaluate ballot stuffing in the same region as the "clean" sample, we can directly obtain the analog estimator of the fraud density as

$$\hat{f}_{q|X^m,Y^f,Z}(t) = \hat{f}_{Y|X^m,Z}(Y^f - t).$$

Note that under the normality assumption on the "clean" sample, we have $Y|X^m, Z \sim N(\mu_{Y|Z}, \sigma_{Y|Z}^2)$, where $\mu_{Y|Z} = \mu_Y + \frac{\sigma_{YZ}}{\sigma_Z^2}(Z - \mu_Z)$ and $\sigma_{Y|Z}^2 = \sigma_Y^2 - \frac{\sigma_{YZ}^2}{\sigma_Z^2}$. Hence, the average and density of ballot stuffing can be estimated as

$$\widehat{E(q|X^m, Y^f, Z)} = Y^f - \hat{\mu}_{Y|Z}$$

$$\hat{f}_{q|X^m,Y^f,Z}(q) = \frac{1}{\sqrt{2\pi\hat{\sigma}_{Y|Z}^2}} \exp\left(-\frac{(Y^f - q - \hat{\mu}_{Y|Z})^2}{2\hat{\sigma}_{Y|Z}^2}\right).$$

By the delta-method, the proposed estimators are asymptotically normal

$$\sqrt{N_m}(\widehat{E(q|X^m, Y^f, Z)} - E(q|X^m, Y^f, Z)) \xrightarrow{d} N(0, V_q)$$
$$\sqrt{N_m}(\hat{f}_{q|X^m,Y^f,Z}(q) - f_{q|X^m,Y^f,Z}(q)) \xrightarrow{d} N(0, V_{f^q}),$$

with the asymptotic variances $V_q = W(1,1)$ and $V_{f^q} = (\partial F)W(\partial F)'$, where $W = \Gamma V_\theta \Gamma'$,

$$\Gamma = \begin{bmatrix} 1 & -\frac{\sigma_{YZ}}{\sigma_Z^2} & 0 & -\frac{\sigma_{YZ}(Z-\mu_Z)}{\sigma_Z^4} & \frac{Z-\mu_Z}{\sigma_Z^2} \\ 0 & 0 & 1 & -\frac{\sigma_{YZ}^2}{\sigma_Z^4} & -\frac{2\sigma_{YZ}}{\sigma_Z^2} \end{bmatrix}$$

$$V_\theta = \begin{bmatrix} \sigma_Y^2 & \sigma_{YZ} & 0 & 0 & 0 \\ \sigma_{YZ} & \sigma_Z^2 & 0 & 0 & 0 \\ 0 & 0 & 2\sigma_Y^4 & 2\sigma_{YZ}^2 & 2\sigma_{YZ}\sigma_Y^2 \\ 0 & 0 & 2\sigma_{YZ}^2 & 2\sigma_Z^4 & 2\sigma_{YZ}\sigma_Z^2 \\ 0 & 0 & 2\sigma_{YZ}\sigma_Y^2 & 2\sigma_{YZ}\sigma_Z^2 & \sigma_{YZ}^2 + \sigma_Y^2\sigma_Z^2 \end{bmatrix}$$

$$\partial F = \left[\frac{\partial f_{q|X^m,Y^f,Z}(t)}{\partial \mu_{Y|Z}} \; ; \; \frac{\partial f_{q|X^m,Y^f,Z}(t)}{\partial \sigma_{Y|Z}^2} \right].$$

3.6 Ballot stuffing in other regions

This section discusses estimation of the density of fraud in a polling station when the density at a point is fully identified.

Estimation of fraud across the country requires knowledge of the density $f_{Y,Z|X^k}$ for any observable characteristics X^k. The identification strategy suggests the following analogy estimators:

$$\hat{f}_{q|X^k,Y^f,Z}(q) = \hat{f}_{Y|X^k,Z}(Y^f - q) = \frac{\hat{f}_{Y,Z|X^k}(Y^f - q, Z)}{\hat{f}_{Z|X^k}(Z)},$$

where the estimator for the joint density $\hat{f}_{Y,Z|X^k}(Y^f - q, Z)$ is

$$\hat{f}_{Y,Z|X^k}(Y^f - q, Z) = \frac{\hat{f}_{Y,Z|X^m}(\hat{\tilde{y}}, \hat{\tilde{z}}) \hat{f}_Z(Y^f - q) \hat{f}_Z(Z)}{\hat{f}_Z(\hat{\tilde{y}}) \hat{f}_Z(\hat{\tilde{z}})}$$

with

$$\hat{\tilde{y}} = \widehat{F_Z^{-1}} \left(\frac{\hat{\beta}}{\alpha}(X^k - X^m) + \hat{F}_Z(Y^f - q) \right) \text{ and } \hat{\tilde{z}} = \widehat{F_Z^{-1}} \left(\frac{\hat{\beta}}{\alpha}(X^k - X^m) + \hat{F}_Z(Z) \right).$$

First, note the estimators of $\tilde{y}$ and $\tilde{z}$ converge with the speed $\sqrt{N}$ – faster than the other components of the fraud-density estimator. Hence, the asymptotics of the fraud-density estimator depends on the rate of convergence of density $\hat{f}_{Y,Z|X^m}(\cdot, \cdot)$ relative to the speed of the non-parametric density estimator $\hat{f}_{Z|X^k}(\cdot)$. Theoretically, the rate of convergence of $\hat{f}_{Z|X^k}(\cdot)$ should be $\sqrt{Nh^{L+1}}$; however, under the assumption of the linearity in the observables, we have $\hat{f}_{Z|X^k}(z) = \hat{f}_{Z|\frac{\hat{\beta}}{\alpha}X^k}(z)$, implying we can shrink the dimensionality. Also, using the estimator $\frac{\hat{\beta}}{\alpha}$ would not affect the rate of convergence of $\hat{f}_{Z|\frac{\hat{\beta}}{\alpha}X^k}(z)$, because $\frac{\hat{\beta}}{\alpha}$ converges as $\sqrt{N}$, which is significantly faster. Hence, we obtain that the rate of convergence of $\hat{f}_{Z|\frac{\hat{\beta}}{\alpha}X^k}(z)$ is $\sqrt{Nh^2}$. In this paper, we also suppose the non-parametric density converges faster than the parametric "clean" sample estimators, implying the rate of convergence of the density of ballot stuffing would be defined by the size of the "clean" subsample. This assumption is motivated by the available Russian data, where the total sample size is over 90,000 observations, whereas the "clean" subsample is below 80.

Assumption 10. $\frac{N_m}{Nh^2} \to 0$.

To simplify the notation, define $Q = \frac{f_Z^2(Y^f - q) f_Z^2(Z)}{f^2_{Z|\frac{\beta}{\alpha}' X^k}(Z) f_Z^2(\tilde{y}) f_Z^2(\tilde{z})}$. Then, we have the following result.

Theorem 3. *Let the estimator of* $f_{q|X^k,Y^f,Z}(q)$*, where* $k \neq m$*, be as defined above. If Assumptions 1–10 are satisfied, then,* $\sup_q \left| \hat{f}_{q|X^k,Y^f,Z}(q) - f_{q|X^k,Y^f,Z}(q) \right| \xrightarrow{p} 0$ *and* $\sqrt{N_m} \left(\hat{f}_{q|X^k,Y^f,Z}(q) - f_{q|X^k,Y^f,Z}(q) \right) \xrightarrow{d} N\left(0, QV_{f^q}\right)$*.*

Note Theorem 3 works under both Assumptions 4a and 4b; however, under Assumption 4b the applicability of this theorem is limited because of the restrictive conditions on values of Z and q.

In addition, under Assumption 4a, we can evaluate the average amount of fraud in the polling station of interest either by employing density $f_{q|X^k,Y^f,Z}$ or more directly through

$$E(q|\widehat{X^k, Y^f}, Z) = Y^f - E(\widehat{Y|X^k}, Z) = Y^f - \hat{E}\left(\hat{F}_Z^{-1} \left(\hat{F}_Z(\tilde{y}) - (X^k - X^m)\frac{\hat{\beta}}{\alpha} \right) |\hat{\tilde{z}} \right).$$

Theorem 4. *If Assumptions 1–3, 4a, and 5-10 are satisfied, then,*

$$\sqrt{N_m} \left(E(q|\widehat{X^k, Y^f}, Z) - E(q|X^k, Y^f, Z) \right) \xrightarrow{d} N\left(0, \sigma^2_{y|\tilde{z}}\right),$$

where $\sigma^2_{y|\tilde{z}} = \frac{\sigma^2_Y \sigma^2_Z - \sigma^2_{YZ}}{\sigma^2_Z}\left(1 + \frac{(\tilde{z}-\mu_Z)^2}{\sigma^2_Z}\right)$.

3.7 Ballot-stuffing bounds

In this section, we consider the case of partial identification and propose estimators for the general and lower bounds discussed in section 2.4.

3.7.1 General bound

We have defined the general bound as $q_g = Y^f - E(\tilde{y}|\tilde{z})$. Moreover, by definition, $\tilde{y}$ and $\tilde{z}$ are random variables in the region of the "clean" data. Hence, given the normality assumption, we have $\mu_{Y|\tilde{z}} = \mu_Y + \frac{\sigma_{YZ}}{\sigma^2_Z}(\tilde{z} - \mu_Z)$, implying the analogy estimator

$$\hat{q}_g = Y^f - \hat{\mu}_Y - \frac{\hat{\sigma}_{YZ}}{\hat{\sigma_Z}^2}(\hat{\tilde{z}} - \hat{\mu}_Z).$$

Theorem 5. *If Assumptions 1–3, 4b, and 5-10 hold, then,* $\sqrt{N_m}(\hat{q}_g - q_g) \xrightarrow{d} N(0, \sigma^2_{y|\tilde{z}})$.

3.7.2 Lower bound

We have previously defined the lower bound as $q_l = P(q \geq q_{\max}|X^k, Y^f, Z)q_{\max}$. However, note the following:

$$\begin{aligned} P(q \leq q_{\max}|X^k, Y^f, Z) &= \int_0^{q_{\max}} f_{q|X^k,Y^f,Z}(q)dq \\ &= \frac{f_Z(Z)}{f_{Z|\frac{\beta}{\alpha}X}(Z)} \int_0^{q_{\max}} f_{\tilde{y}|\tilde{z},X^m}(\tilde{y}) \frac{f_Z(Y^f - q)}{f_Z(\tilde{y})} dq \\ &= \frac{f_Z(Z)}{f_{Z|\frac{\beta}{\alpha}X}(Z)} \int_{y_0}^{y_{\max}} f_{\tilde{y}|\tilde{z},X^m}(\tilde{y}) d\tilde{y} \\ &= \frac{f_Z(Z)}{f_{Z|\frac{\beta}{\alpha}X}(Z)} \left(F_{\tilde{y}|\tilde{z},X^m}(y_{\max}) - F_{\tilde{y}|\tilde{z},X^m}(y_0)\right), \end{aligned}$$

where $y_{\max} = F_Z^{-1}\left(\frac{\beta}{\alpha}(X^k - X^m) + F_Z(Y^f)\right)$, $y_0 = \begin{cases} z_0, \text{ if } \frac{\beta}{\alpha}(X^k - X^m) < 0 \\ F_Z^{-1}\left(\frac{\beta}{\alpha}(X^k - X^m)\right), \text{ otherwise} \end{cases}$,
and $F_{\tilde{y}|\tilde{z},X^m}(\cdot)$ is the cumulative distribution function of $\tilde{y}|\tilde{z}, X^m$. By taking into account that $\tilde{y}|\tilde{z}, X^m \sim N\left(\mu_{Y|\tilde{z}}, \sigma^2_{Y|Z}\right)$, we can use the analogy estimator

$$\hat{q}_l = \hat{q}_{\max}\left(1 - \frac{\hat{f}_Z(Z)}{\hat{f}_{Z|\frac{\beta}{\alpha}X}(Z)}\left(\Phi\left(\frac{\hat{y}_{\max} - \hat{\mu}_{Y|\tilde{z}}}{\hat{\sigma}_{Y|Z}}\right) - \Phi\left(\frac{\hat{y}_0 - \hat{\mu}_{Y|\tilde{z}}}{\hat{\sigma}_{Y|Z}}\right)\right)\right),$$

where Φ is the cumulative distribution function of a standard normal random variable.

For the asymptotic results, we use the following notation:

$$\frac{\partial t}{\partial \sigma_Y^2} = \begin{bmatrix} -\frac{1}{2\sigma^3_{Y|Z}} \\ -\frac{\mu_{Y|\tilde{z}}}{2\sigma^3_{Y|Z}} \end{bmatrix}; \frac{\partial t}{\partial \sigma_Z^2} = \begin{bmatrix} -\frac{\sigma^2_{YZ}}{2\sigma^3_{Y|Z}\sigma^4_Z} \\ \frac{-\sigma_{YZ}\left(\sigma_{YZ}\mu_{Y|\tilde{z}} + 2\sigma^2_{Y|Z}(\tilde{z} - \mu_Z)\right)}{2\sigma^3_{Y|Z}\sigma^4_Z} \end{bmatrix}; \frac{\partial t}{\partial \sigma_{YZ}} = \begin{bmatrix} \frac{1}{\sigma^3_{Y|Z}\sigma^2_Z} \\ \frac{\mu_{Y|\tilde{z}} + \sigma^2_{Y|Z}(\tilde{z} - \mu_Z)}{\sigma^3_{Y|Z}\sigma^2_Z} \end{bmatrix}$$

$$\partial T = \left[\frac{\partial t}{\partial \sigma_Y^2}, \frac{\partial t}{\partial \sigma_Z^2}, \frac{\partial t}{\partial \sigma_{YZ}}\right] \text{ and } \tilde{W} = (\partial T)\Omega(\partial T)' + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{aligned} V_\Phi = & \phi^2\left(\frac{y_{\max} - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right)\left(y^2_{\max}\tilde{W}_{11} + \tilde{W}_{22} - 2y_{\max}\tilde{W}_{12}\right) \\ & + \phi^2\left(\frac{y_0 - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right)\left(y_0^2\tilde{W}_{11} + \tilde{W}_{22} - 2y_0\tilde{W}_{12}\right) \\ & - 2\phi\left(\frac{y_{\max} - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right)\phi\left(\frac{y_0 - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right)\left(y_{\max}y_0\tilde{W}_{11} + \tilde{W}_{22} - (y_{\max} + y_0)\tilde{W}_{12}\right), \end{aligned}$$

where $\phi(\cdot)$ is the density of a standard normal variable.

Finally, denote $V_{q_l} = q^2_{\max}\frac{f^2_Z(Z)}{f^2_{Z|\frac{\beta}{\alpha}X}(Z)}V_\Phi$; then, we have the following result.

Theorem 6. *If Assumptions 1–3, 4b, and 5-10 hold, then,* $\sqrt{N_m}\left(\hat{q}_l - q_l\right) \xrightarrow{d} N\left(0, V_{q_l}\right)$.

4 Empirical Illustration

In this section, we illustrate the estimation procedure and evaluate ballot-stuffing bounds for the 2011 Russia parliamentary elections in several polling stations. The Russian political system consists of numerous parties. However, the country's regime is authoritarian, and the current political spectrum can be divided into two factions: United Russia, the ruling party, and all other political parties. Hence, we combine all votes given for the various opposition parties to obtain the number of votes for the opposition. The official results report 49.3% as the voter share obtained by United Russia with 60.1% turnout. In the case of the Russian political preferences, the polling stations are somewhat homogenous, which prevents us from full identification. Because of this homogeneity, we rely on Assumption 4b. Note the below analysis is only for illustration purposes, and the obtained results should be taken with caution.[5]

First, we fit the data from a "clean" subsample in Moscow into the normal distribution and evaluate ballot stuffing at a number of polling stations in Moscow as well. Next, we estimate $\frac{\beta}{\alpha}$ from the entire dataset and evaluate the bounds of ballot stuffing for a number of polling stations in Kostroma and Kalmykia.

4.1 Data

The main data set are the results of the 2011 Russia parliamentary elections for each polling station. The data set is openly available on the website of the Central Electoral Commission of the Russian Federation.[6] In total, we use 94,786 observations after exclusion of the voting on the foreign territory.

The choice of the unit for a region should be based on availability of data that are being used as controls. No requirements exist regarding the region size, and a region might consist of one polling station, if detailed information on regional characteristics

[5] The proper estimation would require matching of over 90,000 data points with their municipalities and collection of regional characteristics for over 2,000 municipalities, for which data are often available in different formats. Such data collection is outside the scope of this paper.

[6] http://www.izbirkom.ru/region/izbirkom

is available. We use a Territorial Electoral Commission (TIK) as a region for the non-urban areas and combine the data of several TIKs inside of each city to form a city-region. This procedure left us with 2,483 regions. Note such choice of a region is not the best for Russia, because TIKs do not always coincide with non-urban regional units used by the municipalities. Hence, for proper analysis, one should form a region by including polling stations inside of the municipality rather than using the division suggested by the TIK.

We take the total number of voters and the average size of a polling station in a region as regional characteristics X. We normalize them by dividing by the largest value in the sample; hence, all values of X are between 0 and 1.[7] Electoral variable Y is the number of votes for United Russia divided by the total number of voters[8] in the polling station. To obtain variable Z, we divide the number of pro-opposition votes by the total number of voters and subtract it from 1.

We also use the data on "clean" polling stations from Enikolopov et al. (2013). The authors collect the data from 156 randomly assigned independent observers in Moscow, which revealed even the presence of one independent observer reduced the average share of United Russia from 47% to 36%. We use 75 polling stations that independent observers reported as having "no violations" in order to constitute our validation sample.

4.2 "Clean" data sample

The clean data size is only 75 data points; hence, we have no choice but to rely on a parametric estimator of the joint density of the electoral variables in Moscow. We fit the validation sample into the normal distribution and obtain the estimates of the

[7]Note the proper regional characteristics should include a large number of controls such as average income, share of old population, etc. These data can be obtained from the municipality websites.

[8]To obtain the total number of voters for each polling station, we add up columns $n1$ (the number of registered voters) and $n13$ (the number of voters voted with the absentee certificates) and subtract columns $n12$ (the number of absentee certificates issued by the polling station) and $n15$ (the number of absentee certificates issued by the TIK) in the dataset.

parameters shown in Table 1.

Table 1: Distribution parameters in Moscow

	$\hat{\mu}_Y$	$\hat{\mu}_Z$	$\hat{\sigma}_Y^2$	$\hat{\sigma}_Z^2$	$\hat{\sigma}_{YZ}$
parameter	0.1258	0.6270	0.0015	0.0023	-0.0001
standard error	0.0044	0.0056	0.0002	0.0004	0.0002

The analysis of the ballot stuffing in the region of the "clean" subsample is straightforward and does not require estimating the structural model. Hence, we calculate the average ballot stuffing in two polling stations in Moscow: one "clean" (UIK 265) and one of unknown quality (UIK 856). See Table 2 for the results. The estimation results are consistent with the knowledge of the quality of UIK 265. By contrast, UIK 856 shows significant levels of ballot stuffing: the administrative staff filled in ballots for 17.49% of the voter population of the polling station, which resulted in the increase in the voter share of United Russia from 32.13% to 53.97% after taking the turnout into account.

	UIK 265	UIK 856
Y^f	0.1275	0.2934
average ballot stuffing	0.0011	0.1749
standard error	0.0045	0.0121
reported voter share of United Russia	25.02%	53.97%
expected voter share of United Russia	24.85%	32.13%

Table 2: Average ballot stuffing in Moscow

4.3 Estimation in other regions

Now we turn to the semi-parametric estimation of the structural model. In this section, we concentrate on the polling stations outside of the validation-sample area. For the

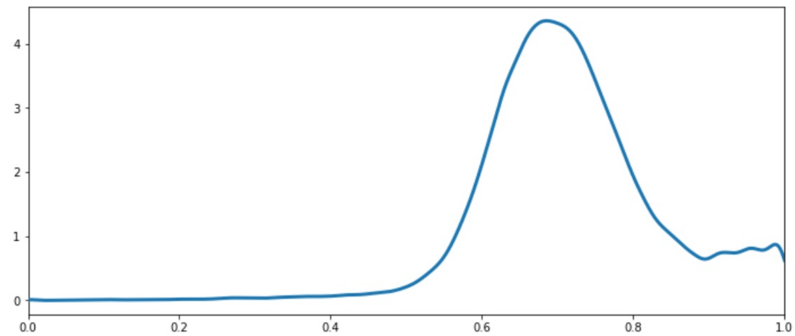

Figure 1: Density of Z

purposes of illustration, we analyze a couple of polling stations in the city of Kostroma[9] and in the Republic of Kalmykia[10].

We begin by estimating the non-parametric density of the electoral variable Z. We use the standard kernel estimator with the normal density as a kernel function. See Figure 1 for the graph. Note Z is non-zero somewhere above 0.5; hence, we have to rely on Assumption 4b, and the model is only partially identified in this case. This fact is the result of overall high support for the incumbent all over the country, coupled with the low voter turnout. If a country had more heterogenous regions, with some of them strongly supporting the opposition, this problem would not arise.

Next, we estimate $\frac{\beta}{\alpha}$ through linear regression of $\hat{F}_Z(Z)$ on X. The estimates are shown in Table 3. The results suggest the total number of votes and the average size of the polling station positively affect the voter share of the opposition, which indicates greater support of the current regime in rural areas.

	const	$\frac{\beta_1}{\alpha}$	$\frac{\beta_2}{\alpha}$
coefficient	1.084	0.367	0.499
standard error	0.010	0.025	0.014

Table 3: Semi-parametric estimators of $\frac{\beta}{\alpha}$

[9]Kostroma is a regional center located 340 kms northeast from Moscow.
[10]Kalmykia is located in the southwest border of Russia with access to the Caspian Sea.

4.3.1 Estimating $\frac{\beta}{\alpha}$ with regular OLS

A natural question to ask at this stage is how different the results would be if instead of regressing $F_Z(Z)$ on X, we estimate $\frac{\beta}{\alpha}$ with the linear regression of Z on X directly. See Table 4 for the results of such estimation. Note the difference in estimation of $\frac{\beta_1}{\beta_2}$: the semi-parametric approach suggests the value is 0.74, whereas the OLS approach evaluates it as 0.51, which is significantly smaller. Hence, the OLS estimator overestimates the relative impact of X_2. In a general case, the results might differ dramatically.

	const	$\frac{\beta_1}{\alpha}$	$\frac{\beta_2}{\alpha}$
coefficient	0.928	0.105	0.207
standard error	0.004	0.010	0.005

Table 4: OLS estimators of $\frac{\beta}{\alpha}$

Additionally, note that by definition $Z_i = G(c_i - \delta_i - X^i\beta)$; hence, the OLS estimation makes an underlying assumption that the personal component is distributed close to uniformly, implying the uniform distribution for Z as well because of the relationship $g(G^{-1}(z)) = \frac{1}{\alpha f_Z(z)}$. The last statement is strongly contradicted by the data (see Figure 1), implying that using the linear approximation is not appropriate with this dataset.

4.3.2 Kostroma

In this subsection, we study polling stations UIK 213 and UIK 299 in the city of Kostroma. We start by evaluating $\frac{\beta}{\alpha}(X^k - X^m) = -0.11 < 0$. This fact implies the general bound will be the upper bound on ballot stuffing. Next, we estimate $\hat{\hat{z}}$ and $\hat{q}_g = Y^f - \mu_{Y|\hat{\hat{z}}}$ for both polling stations. The upper bound on fraud allows us to calculate the lower bound on the voter share of United Russia as well. The results are shown in Table 5.

The results suggest ballot stuffing added up to 3.24% in UIK 213 and up to 8.22% in UIK 299 to the voter share of United Russia. Note UIK 299 reports a higher number

	UIK 213	UIK 299
Y^f	0.1499	0.1964
Z	0.6384	0.6133
upper bound on ballot stuffing	0.0224	0.0643
standard error	0.0051	0.0106
reported voter share of United Russia	29.30%	33.68%
lower bound on voter share of United Russia	26.06%	25.46%

Table 5: Uppers bounds on ballot stuffing in Kostroma

of votes for the opposition than UIK 213. At the same time, the number of votes of United Russia is also greater in UIK 299. The estimation procedure picks up on this fact and reports a greater potential for ballot stuffing in UIK 299.

Finally, we check the condition on the values of Z from Theorem 2b, and unfortunately, $F_Z(Y^f_{213}) = 0.001$, $F_Z(Y^f_{299}) = 0.002$, $-\frac{\beta}{\alpha}(X^k - X^m) = 0.11$, and hence $F_Z(Y^f) < -\frac{\beta}{\alpha}(X^k - X^m)$ for both polling stations. This fact implies the condition is violated and the support for q is empty, so the lower bound cannot be estimated in these cases.

4.3.3 Kalmykia

In this subsection, we analyze ballot stuffing in UIK 183 and UIK 184 in the Republic of Kalmykia. Both polling stations are located in nearby villages of the Chernozemelniy district of Kalmykia.

We start with obtaining $\frac{\beta}{\alpha}(X^k - X^m) = -0.23 < 0$, implying the general bound will provide us with the upper bound on fraud again. However, this time $F_Z(Y^f_{183}) = 0.84$ and $F_Z(Y^f_{184}) = 0.45$, implying $1 - \frac{\beta}{\alpha}(X^k - X^m) > F_Z(Y^f) > -\frac{\beta}{\alpha}(X^k - X^m)$ for both polling stations, so the lower bound can be evaluated as well.

The upper bound $\hat{q}_g$ is estimated as before. To estimate the lower bound, we obtain $\hat{q}_{\max}$ and probability $\widehat{P(q \leq q_{\max}|X^k}, Y^f, Z)$, by using $z_0 = 0.6$, and get the estimator

$\hat{q}_l = \hat{q}_{\max}\left(1 - \frac{\hat{f}_Z(Z)}{\hat{f}_{Z|\frac{\hat{\beta}}{\alpha}X}(Z)} P(q \leq \widehat{q_{\max}|X^k}, Y^f, Z)\right)$. The results are in Table 6.

	UIK 183	UIK 184
Y^f	0.8181	0.6943
Z	0.8760	0.9257
lower bound on ballot stuffing	0.1771	0.0533
standard error	0.0121	0.0062
upper bound on ballot stuffing	0.6996	0.5765
standard error	0.0121	0.0130
reported voter share of United Russia	86.84%	90.33%
upper bound on voter share of United Russia	83.79%	89.61%
lower bound on voter share of United Russia	48.85%	61.33%

Table 6: Bounds on ballot stuffing in Kalmykia

Note the support for the opposition is very low in both polling stations, which is consistent with the prediction of lower opposition support in rural areas. Similarly to Kostroma's polling stations, we again observe UIK 183 has a higher number of votes for both the opposition and United Russia. The estimation procedure picks up this fact for both lower and upper bounds, and results in a greater bounds for UIK 183. As a consequence, ballot stuffing increased the voter share of United Russia by at least 3.05% and by at most 37.99% in UIK 183, and by at least 0.72% and by at most 29% in UIK 184.

References

Aksoy, O. (2016). Regression discontinuity diagnostics reveal statistical anomalies in turkish elections. *Electoral Studies 44*, 284–292.

Asunka, J., S. Brierley, M. Golden, E. Kramon, and G. Ofosu (2019). Electoral fraud or violence: The effect of observers on party manipulation strategies. *British Journal of Political Science 49(1)*, 129–151.

Bound, J., C. Brown, G. Duncan, and W. Rodgers (1989). Measurement error in cross-sectional and longitudinal labor marker surveys: results from two validation studies. *NBER Working Paper 2884*.

Breunig, C. and A. Goerres (2011). Searching for electoral irregularities in an established democracy: Applying benfords law tests to bundestag elections in unified germany. *Electoral Studies 30*, 534–545.

Cantu, F. (2014). Identifying irregularities in Mexican local elections. *American Journal of Political Science 58(4)*, 936–951.

Cantu, F. and S. Saiegh (2011). Fraudulent democracy? An analysis of Argentina's infamous decade using supervised machine learning. *Political Analysis 19(4)*, 409–433.

Carroll, R., X. Chen, and Y. Hu (2010). Identification and estimation of nonlinear models using two samples with nonclassical measurement errors. *Journal of Nonparametric Statistics 22(4)*, 379–399.

Carroll, R. and M. Wand (1991). Semiparametric estimation in logistic measurement error models. *Journal of the Royal Statistical Society B 53*, 573–585.

Casas, A., G. Diaz, and A. Trindade (2017). Who monitors the monitor? Effect of party observers on electoral outcomes. *Journal of Public Economics 145*, 136–149.

Chen, X., H. Hong, and E. Tamer (2005). Measurement error models with auxiliary data. *Review of Economic Studies 72*, 343–366.

Coate, S. and M. Conlin (2004). A group rule-utilitarian approach to voter turnout: Theory and evidence. *American Economic Review 94(5)*, 1476–1504.

Degan, A. (2007). Candidate valence: Evidence from consecutive presidential elections. *International Economic Review 48(2)*, 457–482.

Degan, A. and A. Merlo (2011). A structural model of turnout and voting in multiple elections. *Journal of the European Economic Association 9(2)*, 209–245.

Downs, A. (1957). An economic theory of political action in a democracy. *Journal of Political Economy 65(2)*, 135–150.

Enikolopov, R., V. Korovkin, M. Petrova, K. Sonin, and A. Zakharov (2013). Field experiment estimate of electoral fraud in Russian parliamentary elections. *Proceedings of the National Academy of Sciences 110(2)*, 448–452.

Fujiwara, T., K. Meng, and T. Vogl (2016). Habit formation in voting: Evidence from rainy elections. *American Economic Journal: Applied Economics 8(4)*, 160–188.

Hausman, J., H. Ichimura, W. Newey, and J. Powell (1991). Identification and estimation of polynomial errors-in-variables models. *Journal of Econometrics 50*, 273–295.

Hinich, M. and M. Munger (1996). *Ideology and the Theory of Public Choice.* Ann Arbor: University of Michigan Press.

Hsiao, C. (1989). Consistent estimation for some nonlinear errors-in-variables models. *Journal of Econometrics 41*, 159–185.

Hu, Y. and G. Ridder (2012). Estimation of nonlinear models with mismeasured regressors using marginal information. *Journal of Applied Econometrics 27*, 347–385.

Katz, J. and G. Katz (2010). Correcting for survey misreports using auxiliary information with an application to estimating turnout. *American Journal of Political Science 54(3)*, 815–835.

Kawai, K., Y. Toyama, and Y. Watanabe (2020). Voter turnout and preference aggregation. *American Economic Journal: Microeconomics*, forthcoming.

Kernell, G. (2009). Giving order to districts: Estimating voter distributions with national election returns. *Political Analysis 17(3)*, 215–235.

Klimek, P., Y. Yegorov, R. Hanel, and S. Thurner (2012). Statistical detection of systematic election irregularities. *Proceedings of the National Academy of Sciences 109(40)*, 16469–16473.

Kobak, D., S. Shpilkin, and M. Pshenichnikov (2016). Integer percentages as electoral falsification fingerprints. *Annals of Applied Statistics 10*, 54–73.

Lee, L. and J. Sepanski (1995). Estimation of linear and nonlinear errors-in-variables models using validation data. *Journal of the American Statistical Association 90*, 130–140.

Leeffers, S. and P. Vicente (2019). Does electoral observation influence electoral results? Experimental evidence for domestic and international observers in Mozambique. *World Development 114*, 42–58.

Leon, G. (2017). Turnout, political preferences and information: Experimental evidence from Peru. *Journal of Development Economics 127*, 56–71.

Levin, I., G. Cohn, P. Ordeshook, and R. Alvarez (2009). Detecting voter fraud in an electronic voting context: An analysis of the unlimited reelection vote in Venezuela.

Li, Q. and J. Racine (2007). *Nonparametric Econometrics: Theory and Practice*. The University Press Group Ltd.

Lindbeck, A. and J. Weibull (1987). Balanced-budget redistribution as the outcome of political competition. *Public Choice 52*, 273–297.

McMurray, J. (2013). Aggregating information by voting: The wisdom of the experts versus the wisdom of the masses. *Review of Economic Studies 80*, 277–312.

Mebane, W. J. (2008). Election forensics: The second-digit Benfords law test and recent American presidential elections. pp. 162–181.

Merlo, A. and A. de Paula (2017). Identification and estimation of preference distributions when voters are ideological. *Review of Economic Studies 84(3)*, 1238–1263.

Newey, W. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory 10(2)*, 233–253.

Pepe, M. and T. Fleming (1991). A general nonparametric method for dealing with errors in missing or surrogate covariate data. *Journal of the American Statistical Association 86*, 108–113.

Persson, T. and G. Tabellini (2000). *Political Economics : Explaining Economic Policy.* MIT Press Ltd.

Riker, W. and P. Ordeshook (1968). A theory of the calculus of voting. *American Political Science Review 62(1)*, 25–42.

Rozenas, A. (2017). Detecting election fraud from irregularities in vote-share distributions. *Political Analysis 25(1)*, 41–56.

Schennach, S. (2016). Recent advances in the measurement error literature. *Annual Reviews of Economics 8*, 341–377.

Skovoroda, R. and T. Lankina (2017). Fabricating votes for Putin: new tests of fraud and electoral manipulations from Russia. *Post-Soviet Affairs 33(2)*, 100–123.

A Appendix

Lemma 1. *If Assumptions 1-3 and 4b hold, the density $f_{q|X^k,Y^f,Z}$ is identified at polling stations with characteristics X^k and electoral variable Z such that*

$$\frac{\beta}{\alpha}(X^m - X^k) \leq F_Z(Z) \leq \frac{\beta}{\alpha}(X^m - X^k) + 1,$$

and for any $0 \leq q \leq q_{\max}$*, where* $q_{\max} = \begin{cases} Y^f - z_0, \text{ if } \frac{\beta}{\alpha}(X^k - X^m) \geq 0 \\ Y^f - F_Z^{-1}\left(-\frac{\beta}{\alpha}(X^k - X^m)\right), \text{ otherwise} \end{cases}$.

Proof. First, we start with the constraints on Z. Recall $F_Z(\tilde{z}) = \frac{\beta}{\alpha}(X^k - X^m) + F_Z(Z)$; however, for $\tilde{z}$ to be defined properly, we must have $0 \leq F_Z(\tilde{z}) \leq 1$. Hence, we have the constraint on Z.

Second, we obtain the constraints on q. Recall $F_Z(\tilde{y}) = \frac{\beta}{\alpha}(X^k - X^m) + F_Z(Y^f - q)$. Three restrictions deliver the constraints:

1. $q \geq 0$ because ballot stuffing cannot be negative.

2. $Y^f - q \geq z_0$; otherwise, the original function G is not identified and we cannot use the equality.

3. $0 \leq F_Z(\tilde{y}) \leq 1$, implying

$$-\frac{\beta}{\alpha}(X^k - X^m) \leq F_Z(Y^f - q) \leq 1 - \frac{\beta}{\alpha}(X^k - X^m)$$

$$\Rightarrow Y^f - F_Z^{-1}\left(1 - \frac{\beta}{\alpha}(X^k - X^m)\right) \leq q \leq Y^f - F_Z^{-1}\left(\frac{\beta}{\alpha}(X^k - X^m)\right).$$

However, note $F_Z(Y^f) \leq F_Z(Z) \leq 1 - \frac{\beta}{\alpha}(X^k - X^m)$, implying $Y^f - F_Z^{-1}\left(1 - \frac{\beta}{\alpha}(X^k - X^m)\right) \leq 0$ and making this constraint irrelevant.

Hence, now we have that $0 \leq q \leq \max\left(Y^f - F_Z^{-1}\left(\frac{\beta}{\alpha}(X^k - X^m)\right), Y^f - z_0\right)$. Finally, $z_0 \leq F_Z^{-1}\left(\frac{\beta}{\alpha}(X^k - X^m)\right) \Leftrightarrow \frac{\beta}{\alpha}(X^k - X^m) \geq 0$, and the result follows.

□

We assume $\begin{pmatrix} Y \\ Z \end{pmatrix} | X^m \sim N(\mu, \Sigma)$, where $\mu = \begin{pmatrix} \mu_Y \\ \mu_Z \end{pmatrix}$ and $\Sigma = \begin{bmatrix} \sigma_Y^2 & \sigma_{YZ} \\ \sigma_{YZ} & \sigma_Z^2 \end{bmatrix}$, and we estimate five parameters of the bivariate normal distribution from the validation sample $\theta = (\mu_Y, \mu_Z, \sigma_Y^2, \sigma_Z^2, \sigma_{YZ})'$ as follows:

$$\hat{\mu}_Y = \frac{1}{N_m}\sum_{i=1}^{N_m} Y_i = \bar{Y} \text{ and } \hat{\mu}_Z = \frac{1}{N_m}\sum_{i=1}^{N_m} Z_i = \bar{Z}$$
$$\hat{\sigma}_Y^2 = \frac{1}{N_m - 1}\sum_{i=1}^{N_m}(Y_i - \bar{Y})^2 \text{ and } \hat{\sigma}_Z^2 = \frac{1}{N_m - 1}\sum_{i=1}^{N_m}(Z_i - \bar{Z})^2$$
$$\hat{\sigma}_{YZ} = \frac{1}{N_m - 1}\sum_{i=1}^{N_m}(Y_i - \bar{Y})(Z_i - \bar{Z}).$$

Denote V_θ as

$$V_\theta = \begin{bmatrix} \sigma_Y^2 & \sigma_{YZ} & 0 & 0 & 0 \\ \sigma_{YZ} & \sigma_Z^2 & 0 & 0 & 0 \\ 0 & 0 & 2\sigma_Y^4 & 2\sigma_{YZ}^2 & 2\sigma_{YZ}\sigma_Y^2 \\ 0 & 0 & 2\sigma_{YZ}^2 & 2\sigma_Z^4 & 2\sigma_{YZ}\sigma_Z^2 \\ 0 & 0 & 2\sigma_{YZ}\sigma_Y^2 & 2\sigma_{YZ}\sigma_Z^2 & \sigma_{YZ}^2 + \sigma_Y^2\sigma_Z^2 \end{bmatrix}.$$

Lemma 2. $\sqrt{N_m}(\hat{\theta} - \theta) \xrightarrow{d} N(0, V_\theta)$.

Proof. By multivariate CLT, we have that $\hat{\theta}$ converges to asymptotically normal distribution. We are left with the calculation of variance of that distribution. First, note the following:

$$\hat{\mu}_Y \sim N\left(\mu_Y, \frac{\sigma_Y^2}{N_m}\right) \text{ and } \hat{\mu}_Z \sim N\left(\mu_Z, \frac{\sigma_Z^2}{N_m}\right)$$
$$\frac{(N_m - 1)\hat{\sigma}_Y^2}{\sigma_Y^2} \sim \chi^2(N_m - 1) \Rightarrow V\left(\frac{(N_m - 1)\hat{\sigma}_Y^2}{\sigma_Y^2}\right) = 2(N_m - 1) \Rightarrow V(\hat{\sigma}_Y^2) = \frac{2\sigma_Y^4}{N_m - 1}.$$

By analogy, $V(\hat{\sigma}_Z^2) = \frac{2\sigma_Z^4}{N_m - 1}$. Also, by calculating directly, we have $\text{cov}(\hat{\mu}_Y, \hat{\mu}_Z) = \frac{\sigma_{YZ}}{N_m}$. In addition, due to independence $\text{cov}(\hat{\mu}_Y, \hat{\sigma}_Y^2) = \text{cov}(\hat{\mu}_Z, \hat{\sigma}_Z^2) = 0$ and because of normality $\hat{\mu}_Y \perp\!\!\!\perp (Y_i - \bar{Y})$, $\hat{\mu}_Y \perp\!\!\!\perp (Z_i - \bar{Z})$, $\hat{\mu}_Z \perp\!\!\!\perp (Y_i - \bar{Y})$ and $\hat{\mu}_Z \perp\!\!\!\perp (Z_i - \bar{Z})$. Hence,

$$\text{cov}(\hat{\mu}_Y, \hat{\sigma}_Z^2) = \text{cov}(\hat{\mu}_Z, \hat{\sigma}_Y^2) = \text{cov}(\hat{\mu}_Y, \hat{\sigma}_{YZ}) = \text{cov}(\hat{\mu}_Z, \hat{\sigma}_{YZ}) = 0.$$

Thus, we are left to obtain $V(\hat{\sigma}_{YZ})$, $\text{cov}(\hat{\sigma}^2_Y, \hat{\sigma}^2_Z)$, $\text{cov}(\hat{\sigma}^2_Y, \hat{\sigma}_{YZ})$, and $\text{cov}(\hat{\sigma}^2_Z, \hat{\sigma}_{YZ})$. To calculate them, we need the following elements:

$$\begin{aligned}
E(Y_i Z_i) &= \sigma_{YZ} + \mu_Y \mu_Z \\
E(Y_i^2 Z_i) &= (\sigma^2_Y + \mu^2_Y)\mu_Z + 2\mu_Y \sigma_{YZ} \\
E(Y_i Z_i^2) &= (\sigma^2_Z + \mu^2_Z)\mu_Y + 2\mu_Z \sigma_{YZ} \\
E(Y_i^2 Z_i^2) &= (\sigma^2_Y + \mu^2_Y)\mu^2_Z + (\sigma^2_Y + \mu^2_Y)\sigma^2_Z + 4\mu_Y\mu_Z\sigma_{YZ} + 2\sigma^2_{YZ}.
\end{aligned}$$

Hence, $V(Y_i Z_i) = \sigma^2_Y \mu^2_Z + \sigma^2_Y \sigma^2_Z + \mu^2_Y \sigma^2_Z + 2\mu_Y\mu_Z\sigma_{YZ} + \sigma^2_{YZ}$. By analogy, we obtain

$$V(\bar{Y}\bar{Z}) = \frac{1}{N_m}\sigma^2_Y\mu^2_Z + \frac{1}{N_m^2}\sigma^2_Y\sigma^2_Z + \frac{1}{N_m}\mu^2_Y\sigma^2_Z + \frac{2}{N_m}\mu_Y\mu_Z\sigma_{YZ} + \frac{1}{N_m^2}\sigma^2_{YZ}.$$

Moreover, $\text{cov}(Y_i Z_i, Y_i Z_j) = \sigma^2_Y \mu^2_Z + \mu_Y \mu_Z \sigma_{YZ}$ and

$$\begin{aligned}
\text{cov}(Y_i Z_i, \bar{Y}\bar{Z}) &= \frac{1}{N_m^2}\text{cov}\left(Y_i Z_i, \sum_i \sum_j Y_i Z_j\right) = \frac{1}{N_m^2}\text{cov}\left(Y_i Z_i, Y_i Z_i + \sum_{j \neq i} Y_i Z_j + \sum_{j \neq i} Y_j Z_i\right) \\
&= \frac{1}{N_m^2}(\sigma^2_{YZ} + \sigma^2_Y\sigma^2_Z + N_m\sigma^2_Y\mu^2_Z + N_m\sigma^2_Z\mu^2_Y + 2N_m\mu_Y\mu_Z\sigma_{YZ}).
\end{aligned}$$

Now we are ready to derive $V(\hat{\sigma}_{YZ})$:

$$\begin{aligned}
V(\hat{\sigma}_{YZ}) &= \frac{1}{(N_m - 1)^2}(N_m V(Y_i Z_i) + N_m^2 V(\bar{Y}\bar{Z}) - 2N_m^2 \text{cov}(Y_i Z_i, \bar{Y}\bar{Z})) \\
&= \frac{1}{N_m - 1}(\sigma^2_Y\sigma^2_Z + \sigma^2_{YZ}).
\end{aligned}$$

To find covariances of the estimators of variances, we need the following:

$$\begin{aligned}
\text{cov}(Y_i^2, Z_i^2) &= 4\mu_Y\mu_Z\sigma_{YZ} + 2\sigma^2_{YZ} \\
\text{cov}(Z_i^2, (\bar{Y})^2) &= \frac{1}{N_m^2}\text{cov}\left(Z_i^2, \sum_i Y_i^2 + \sum_j \sum_{k \neq j} Y_j Y_k\right) = \frac{1}{N_m^2}\text{cov}\left(Z_i^2, Y_i^2 + 2Y_i Y_j\right) \\
&= \frac{1}{N_m^2}(4N_m\mu_Y\mu_Z\sigma_{YZ} + 2\sigma^2_{YZ}).
\end{aligned}$$

By analogy, we have $\text{cov}(Y_i^2, (\bar{Z})^2) = \text{cov}((\bar{Y})^2, (\bar{Z})^2) = \frac{1}{N_m^2}(4N_m\mu_Y\mu_Z\sigma_{YZ} + 2\sigma_{YZ}^2)$. Hence,

$$\begin{aligned}
\text{cov}(\hat{\sigma}_Y^2, \hat{\sigma}_Z^2) &= \frac{\big(N_m\text{cov}(Y_i^2, Z_i^2) - N_m^2\text{cov}(Z_i^2, (\bar{Y})^2) - N_m^2\text{cov}(Y_i^2, (\bar{Z})^2) + N_m^2\text{cov}((\bar{Y})^2, (\bar{Z})^2)\big)}{(N_m-1)^2} \\
&= \frac{2}{N_m-1}\sigma_{YZ}^2.
\end{aligned}$$

Finally, for the last covariance, we need to calculate the following values.

$$\begin{aligned}
E(Y_i^3 Z_i) &= (\sigma_Y^2 + \mu_Y^2)\mu_Y\mu_Z + 2\mu_Y\mu_Z\sigma_Y^2 + 3(\sigma_Y^2 + \mu_Y^2)\sigma_{YZ} \\
\text{cov}(Y_i^2, Y_iZ_i) &= 2\mu_Y\mu_Z\sigma_Y^2 + 2(\sigma_Y^2 + \mu_Y^2)\sigma_{YZ} \\
\text{cov}(Y_iZ_i, (\bar{Y})^2) &= \frac{1}{N_m^2}\text{cov}(Y_iZ_i, Y_i^2 + 2(N_m-1)Y_iY_j) \\
&= \frac{1}{N_m^2}(2N_m\mu_Y\mu_Z\sigma_Y^2 + 2N_m\mu_Y^2\sigma_{YZ} + 2\sigma_Y^2\sigma_{YZ})
\end{aligned}$$

$$\begin{aligned}
\text{cov}(Y_i^2, \bar{Y}\bar{Z}) &= \frac{1}{N_m^2}\text{cov}(Y_i^2, Y_iZ_i + (N_m-1)Y_iZ_j + (N_m-1)Y_jZ_i) \\
&= \frac{2}{N_m^2}(N_m\mu_Y\mu_Z\sigma_Y^2 + \sigma_Y^2\sigma_{YZ} + N_m\mu_Y^2\sigma_{YZ}) \\
\text{cov}((\bar{Y})^2, \bar{Y}\bar{Z}) &= \frac{2}{N_m^2}(N_m\mu_Y\mu_Z\sigma_Y^2 + \sigma_Y^2\sigma_{YZ} + N_m\mu_Y^2\sigma_{YZ}).
\end{aligned}$$

Thus, we have

$$\begin{aligned}
\text{cov}(\hat{\sigma}_Y^2, \hat{\sigma}_{YZ}) &= \frac{N_m\text{cov}(Y_i^2, Y_iZ_i) - N_m^2\text{cov}(Y_iZ_i, (\bar{Y})^2) - N_m^2\text{cov}(Y_i^2, \bar{Y}\bar{Z}) + N_m^2\text{cov}((\bar{Y})^2, \bar{Y}\bar{Z})}{(N_m-1)^2} \\
&= \frac{2}{N_m-1}\sigma_Y^2\sigma_{YZ}.
\end{aligned}$$

By analogy, $\text{cov}(\hat{\sigma}_Z^2, \hat{\sigma}_{YZ}) = \frac{2}{N_m-1}\sigma_Z^2\sigma_{YZ}$, and the result follows.

□

Lemma 3. $\sqrt{N_m}(\hat{f}_{Y,Z|X^m}(y,z) - f_{Y,Z|X^m}(y,z)) \xrightarrow{d} N(0, V_{f^m})$.

Proof. The joint density is

$$f_{Y,Z|X^m}(y,z) = \frac{1}{2\pi|\Sigma|^{0.5}} \exp\left\{-\frac{1}{2}\left(\begin{pmatrix} y \\ z \end{pmatrix} - \mu\right)' \Sigma^{-1} \left(\begin{pmatrix} y \\ z \end{pmatrix} - \mu\right)\right\}$$
$$= \frac{1}{2\pi|\Sigma|^{0.5}} \exp\left\{-\frac{1}{2|\Sigma|}((y-\mu_Y)^2\sigma_Z^2 + 2(y-\mu_Y)(z-\mu_Z)\sigma_{YZ} + (z-\mu_Z)^2\sigma_Y^2)\right\},$$

so $f_{Y,Z|X^m}(y,z) = g(\theta)$ and $\hat{f}_{Y,Z|X^m}(y,z) = g(\hat{\theta})$. By the delta-method, we have $\sqrt{N_m}(\hat{f}_{Y,Z|X^m}(y,z) - f_{Y,Z|X^m}(y,z)) \xrightarrow{d} N(0, V_{f^m})$, where the partial derivatives of $g(\theta)$ are

$$\frac{\partial g(\theta)}{\partial \mu_Y} = \frac{1}{|\Sigma|} f_{Y,Z|X=X^m}(y,z)((y-\mu_Y)\sigma_Z^2 + (z-\mu_Z)\sigma_{YZ})$$
$$\frac{\partial g(\theta)}{\partial \mu_Z} = \frac{1}{|\Sigma|} f_{Y,Z|X=X^m}(y,z)((z-\mu_Z)\sigma_Y^2 + (y-\mu_Y)\sigma_{YZ})$$
$$\frac{\partial g(\theta)}{\partial \sigma_Y^2} = \frac{-1}{2|\Sigma|^2} f_{Y,Z|X=X^m}(y,z)((z-\mu_Z)^2(2\sigma_Y^2\sigma_Z^2 - \sigma_{YZ}^2) + (y-\mu_Y)^2\sigma_Z^4$$
$$+ 2(y-\mu_Y)(z-\mu_Z)\sigma_{YZ}\sigma_Z^2 + \sigma_Z^4\sigma_Y^2 - \sigma_Z^2\sigma_{YZ}^2)$$
$$\frac{\partial g(\theta)}{\partial \sigma_Z^2} = \frac{-1}{2|\Sigma|^2} f_{Y,Z|X=X^m}(y,z)((y-\mu_Y)^2(2\sigma_Y^2\sigma_Z^2 - \sigma_{YZ}^2) + (z-\mu_Z)^2\sigma_Y^4$$
$$+ 2(y-\mu_Y)(z-\mu_Z)\sigma_{YZ}\sigma_Y^2 + \sigma_Y^4\sigma_Z^2 - \sigma_Y^2\sigma_{YZ}^2)$$
$$\frac{\partial g(\theta)}{\partial \sigma_{YZ}} = \frac{-1}{|\Sigma|^2} f_{Y,Z|X=X^m}(y,z)((y-\mu_Y)^2\sigma_{YZ}\sigma_Z^2 + (z-\mu_Z)^2\sigma_Y^2\sigma_{YZ}$$
$$+ (y-\mu_Y)(z-\mu_Z)(\sigma_Y^2\sigma_Z^2 + \sigma_{YZ}^2) - \sigma_{YZ}\sigma_Y^2\sigma_Z^2 + \sigma_{YZ}^3).$$

Hence, denote $\partial g = \left(\frac{\partial g(\theta)}{\partial \mu_Y}, \frac{\partial g(\theta)}{\partial \mu_Z}, \frac{\partial g(\theta)}{\partial \sigma_Y^2}, \frac{\partial g(\theta)}{\partial \sigma_Z^2}, \frac{\partial g(\theta)}{\partial \sigma_{YZ}}\right)'$, then $V_{f^m} = (\partial g)' V_\theta (\partial g)$.

□

Lemma 4.

$$\sqrt{N_m}\left(\hat{\mu}_Y + \frac{\hat{\sigma}_{YZ}}{\hat{\sigma}_Z^2}(z - \hat{\mu}_Z) - \mu_Y - \frac{\sigma_{YZ}}{\sigma_Z^2}(z - \mu_Z)\right) \xrightarrow{d} N\left(0, \sigma_{y|z}^2\right),$$

where $\sigma_{y|z}^2 = \frac{\sigma_Y^2\sigma_Z^2 - \sigma_{YZ}^2}{\sigma_Z^2}\left(1 + \frac{(z-\mu_Z)^2}{\sigma_Z^2}\right)$.

Proof. We apply the delta-method again to θ with $g(\theta) = \mu_Y + \frac{\sigma_{YZ}}{\sigma_Z^2}(z - \mu_Z)$. Denote

$$\partial g = \left(1, -\frac{\sigma_{YZ}}{\sigma_Z^2}, -\frac{\sigma_{YZ}(z-\mu_Z)}{\sigma_Z^4}, \frac{z-\mu_Z}{\sigma_Z^2}\right)',$$

then $\sqrt{N_m}(g(\hat{\theta}) - g(\theta)) \xrightarrow{d} N(0, V)$, where $V = (\partial g)' V_\theta (\partial g) = \sigma^2_{y|z}$. □

Lemma 5. *Suppose Assumptions 1-9 hold, then,*

$$\sup_{z\in[0,1]} |\hat{f}_Z(z) - f_Z(z)| \xrightarrow{p} 0$$

$$\sqrt{Nh}(\hat{f}_Z(z) - f_Z(z)) \xrightarrow{d} N\left(0, R(k) f_Z(z) \frac{\mu_{t|z}}{\mu_t}\right),$$

where $\mu_{t|z} = \int t f_{t|z}(t)dt$ *and* t *is the number of polling stations in the region.*

Proof. The dependence between Z_i in the same region will affect only the asymptotic variance of $\hat{f}_Z(z)$ in comparison with the standard non-parametric density estimation. The CLT will still hold, the mean is not affected by the correlation, and the uniform-convergence result stays the same (see, e.g., Li and Racine (2007)). Hence, we have to re-evaluate the asymptotic variance. For simplicity of notation, we denote regions with index j and corresponding polling stations with index i.

First, we derive some basic components of the asymptotic variance:

$$\frac{1}{h} E\left(k\left(\frac{Z_i - z}{h}\right)|X_j, t_j\right) = \frac{1}{h}\int k\left(\frac{Z_i - z}{h}\right) f_{Z|X_j,t_j}(Z_i) dZ_i \to f_{Z|X_j,t_j}(z)$$

$$\frac{1}{h} E\left(k^2\left(\frac{Z_i - z}{h}\right)|X_j, t_j\right) = \frac{1}{h}\int k^2\left(\frac{Z_i - z}{h}\right) f_{Z|X_j,t_j}(Z_i) dZ_i \to R(k) f_{Z|X_j,t_j}(z)$$

$$\frac{1}{h} E\left(\sum_{i=1}^{t_j} k\left(\frac{Z_i - z}{h}\right)|X_j, t_j\right) \to t_j f_{Z|X_j,t_j}(z)$$

$$\frac{1}{h}E\left(\sum_{i=1}^{t_j} k\left(\frac{Z_i - z}{h}\right)\right) \to E(t_j f_{Z|X_j,t_j}(z)) = \int\int t_j f_{Z|X_j,t_j}(z) f_{X,t}(X_j, t_j) dX_j dt_j$$

$$= \int t_j f_{Z,t}(z, t_j) dt_j = f_Z(z)\mu_{t|Z}.$$

Next, we derive the conditional variance:

$$\begin{aligned}
V\left(\sum_{i=1}^{t_j} k\left(\frac{Z_i - z}{h}\right)|X_j, t_j\right) &= t_j V\left(k\left(\frac{Z_i - z}{h}\right)|X_j, t_j\right) \\
&= t_j E\left(k^2\left(\frac{Z_i - z}{h}\right)|X_j, t_j\right) - t_j\left(E\left(k\left(\frac{Z_i - z}{h}\right)|X_j, t_j\right)\right)^2 \\
&= ht_j R(k) f_{Z|X_j,t_j}(z) + o(h).
\end{aligned}$$

The above expression implies the following:

$$\begin{aligned}
E\left(\sum_{i=1}^{t_j} k\left(\frac{Z_i - z}{h}\right)\right)^2 &= E\left(E\left(\left(\sum_{i=1}^{t_j} k\left(\frac{Z_i - z}{h}\right)\right)^2 |X_j, t_j\right)\right) \\
&= E\left(V\left(\sum_{i=1}^{t_j} k\left(\frac{Z_i - z}{h}\right)|X_j, t_j\right) + \left(E\left(\sum_{i=1}^{t_j} k\left(\frac{Z_i - z}{h}\right)|X_j, t_j\right)\right)^2\right) \\
&= hR(k)E\left(t_j f_{Z|X_j,t_j}(z) + o(h)\right) = hR(k) f_Z(z)\mu_{t|z} + o(h).
\end{aligned}$$

Hence, the unconditional variance of the sum is

$$\begin{aligned}
V\left(\sum_{i=1}^{t_j} k\left(\frac{Z_i - z}{h}\right)\right) &= E\left(\sum_{i=1}^{t_j} k\left(\frac{Z_i - z}{h}\right)\right)^2 - \left(E\left(\sum_{i=1}^{t_j} k\left(\frac{Z_i - z}{h}\right)\right)\right)^2 \\
&= hR(k) f_Z(z) E(t|z) + o(h).
\end{aligned}$$

Finally, the asymptotic variance of the estimator is

$$\begin{aligned}
NhV\left(\hat{f}_Z(z)\right) &= NhV\left(\frac{1}{Nh}\sum_{j=1}^{N_r}\sum_{i=1}^{t_j} k\left(\frac{Z_i - z}{h}\right)\right) = \frac{1}{h}\frac{N_r}{N} V\left(\sum_{i=1}^{t_j} k\left(\frac{Z_i - z}{h}\right)\right) \\
&= \frac{N_r}{N}\mu_{t|Z} R(k) f_Z(z) + o(1) \to \frac{\mu_{t|Z}}{\mu_t} R(k) f_Z(z).
\end{aligned}$$

□

Lemma 6. *Suppose Assumptions 1-9 hold, then,*

$$\sup_{z,\frac{\beta}{\alpha}X}\left|\hat{f}_{Z,\frac{\beta}{\alpha}X}\left(z,\frac{\beta}{\alpha}X\right)-f_{Z,\frac{\beta}{\alpha}X}\left(z,\frac{\beta}{\alpha}X\right)\right|\xrightarrow{p}0$$

$$\sqrt{Nh^2}\left(\hat{f}_{Z,\frac{\beta}{\alpha}X}\left(z,\frac{\beta}{\alpha}X\right)-f_{Z,\frac{\beta}{\alpha}X}\left(z,\frac{\beta}{\alpha}X\right)\right)\xrightarrow{d}N\left(0,R(k)f_{Z,\frac{\beta}{\alpha}X}\left(z,\frac{\beta}{\alpha}X\right)\frac{\mu_{t|z,\frac{\beta}{\alpha}X}}{\mu_t}\right).$$

Proof. This proof follows completely Lemma 5. The only thing we have to do is derive the asymptotic variance; the asymptotic normality and uniform convergence are standard. Denote for simplicity of the notation for this proof only $W=\frac{\beta}{\alpha}X$.

First, we derive some basic components of the asymptotic variance:

$$\frac{1}{h}E\left(k\left(\frac{W_j-w}{h}\right)\sum_{i=1}^{t_j}k\left(\frac{Z_i-z}{h}\right)|W_j,t_j\right)\to t_jf_{Z|W_j,t_j}(z)k\left(\frac{W_j-w}{h}\right)$$

$$\begin{aligned}E\left(k\left(\frac{W_j-w}{h}\right)\sum_{i=1}^{t_j}k\left(\frac{Z_i-z}{h}\right)\right)&=hE\left(t_jf_{Z|W_j,t_j}(z)k\left(\frac{W_j-w}{h}\right)\right)+o(h^2)\\&=h\int\int t_jk\left(\frac{W_j-w}{h}\right)f_{Z|W_j,t_j}(z)f_{W,t}(W_j,t_j)dW_jdt_j+o(h^2)\\&=h^2\int t_jf_{Z,W,t}(z,w,t_j)dt_j+o(h^2)=h^2f_{Z,W}(z,w)\mu_{t|z,w}+o(h^2).\end{aligned}$$

Next, we derive the conditional variance:

$$\begin{aligned}&V\left(k\left(\frac{W_j-w}{h}\right)\sum_{i=1}^{t_j}k\left(\frac{Z_i-z}{h}\right)|W_j,t_j\right)=t_jk\left(\frac{W_j-w}{h}\right)V\left(k\left(\frac{Z_i-z}{h}\right)|W_j,t_j\right)\\&=t_jk\left(\frac{W_j-w}{h}\right)E\left(k^2\left(\frac{Z_i-z}{h}\right)|W_j,t_j\right)-t_jk\left(\frac{W_j-w}{h}\right)\left(E\left(k\left(\frac{Z_i-z}{h}\right)|W_j,t_j\right)\right)^2\\&=ht_jR(k)f_{Z|W_j,t_j}(z)k\left(\frac{W_j-w}{h}\right)+o(h).\end{aligned}$$

The above expression implies the following:

$$
\begin{aligned}
E\left(k\left(\frac{W_j-w}{h}\right)\sum_{i=1}^{t_j}k\left(\frac{Z_i-z}{h}\right)\right)^2 &= E\left(E\left(\left(k\left(\frac{W_j-w}{h}\right)\sum_{i=1}^{t_j}k\left(\frac{Z_i-z}{h}\right)\right)^2|W_j,t_j\right)\right) \\
&= E\left(V\left(k\left(\frac{W_j-w}{h}\right)\sum_{i=1}^{t_j}k\left(\frac{Z_i-z}{h}\right)|W_j,t_j\right)+\left(E\left(k\left(\frac{W_j-w}{h}\right)\sum_{i=1}^{t_j}k\left(\frac{Z_i-z}{h}\right)|W_j,t_j\right)\right)^2\right) \\
&= hR(k)E\left(k^2\left(\frac{W_j-w}{h}\right)t_jf_{Z|W_j,t_j}(z)+o(h)\right) \\
&= hR(k)\int\int k^2\left(\frac{W_j-w}{h}\right)t_jf_{Z,W,t}(z,W_j,t_j)dW_jdt_j+o(h^2) \\
&= h^2R^2(k)f_{Z,W}(z,w)\mu_{t|z,w}+o(h^2).
\end{aligned}
$$

Hence, the unconditional variance of the sum is

$$
\begin{aligned}
&V\left(k\left(\frac{W_j-w}{h}\right)\sum_{i=1}^{t_j}k\left(\frac{Z_i-z}{h}\right)\right) \\
&\qquad = E\left(k\left(\frac{W_j-w}{h}\right)\sum_{i=1}^{t_j}k\left(\frac{Z_i-z}{h}\right)\right)^2-\left(E\left(k\left(\frac{W_j-w}{h}\right)\sum_{i=1}^{t_j}k\left(\frac{Z_i-z}{h}\right)\right)\right)^2 \\
&\qquad = h^2R^2(k)f_{Z,W}(z,w)\mu_{t|z,w}+o(h^2).
\end{aligned}
$$

Finally, the asymptotic variance of the estimator is

$$
\begin{aligned}
Nh^2V\left(\hat{f}_{Z,W}(z,w)\right) &= Nh^2V\left(\frac{1}{Nh^2}\sum_{j=1}^{N_r}k\left(\frac{W_j-w}{h}\right)\sum_{i=1}^{t_j}k\left(\frac{Z_i-z}{h}\right)\right) \\
&= \frac{1}{h^2}\frac{N_r}{N}V\left(k\left(\frac{W_j-w}{h}\right)\sum_{i=1}^{t_j}k\left(\frac{Z_i-z}{h}\right)\right) \\
&= \frac{N_r}{N}\mu_{t|z,w}R^2(k)f_{Z,W}(z,w)+o(1) \\
&\to \frac{\mu_{t|z,w}}{\mu_t}R^2(k)f_{Z,W}(z,w).
\end{aligned}
$$

□

Lemma 7. *Suppose Assumptions 1-9 hold, then,*

$$
\sup_{z,\frac{\beta}{\alpha}x}\left|\hat{f}_{Z|\frac{\beta}{\alpha}X}(z)-f_{Z|\frac{\beta}{\alpha}X}(z)\right|\xrightarrow{p}0
$$

$$\sqrt{Nh^2}\left(\hat{f}_{Z|\frac{\beta}{\alpha}X}(z) - f_{Z|\frac{\beta}{\alpha}X}(z)\right) \xrightarrow{d} N\left(0, V_{Z|X}\right),$$

where $V_{Z|X} = \frac{R^2(k) f_{Z|\frac{\beta}{\alpha}X}\left(z, \frac{\beta}{\alpha}X\right)}{f_{\frac{\beta}{\alpha}X}\left(\frac{\beta}{\alpha}X\right)} \frac{\mu_{t|z,\frac{\beta}{\alpha}X}}{\mu_t}$.

Proof. First, note $\hat{f}_{\frac{\beta}{\alpha}X}$ has the standard non-parametric asymptotics, i.e.,

$$\sqrt{N_r h}\left(\hat{f}_{\frac{\beta}{\alpha}X}\left(\frac{\beta}{\alpha}X\right) - f_{\frac{\beta}{\alpha}X}\left(\frac{\beta}{\alpha}X\right)\right) \xrightarrow{d} N\left(0, R(k) f_{\frac{\beta}{\alpha}X}\left(\frac{\beta}{\alpha}X\right)\right)$$

and it also converges uniformly.

Next, by Lemma 6, $\hat{f}_{Z,\frac{\beta}{\alpha}X}\left(z, \frac{\beta}{\alpha}X\right)$ converges with the rate $\sqrt{Nh^2}$. Note the following: $\frac{Nh^2}{N_r h} = \frac{N}{N_r} h \to \mu_t * 0 = 0$, implying $\hat{f}_{Z,\frac{\beta}{\alpha}X}\left(z, \frac{\beta}{\alpha}X\right)$ is the slowest component of the estimator

$$\hat{f}_{Z|\frac{\beta}{\alpha}X}\left(z, \frac{\beta}{\alpha}X\right) = \frac{\hat{f}_{Z,\frac{\beta}{\alpha}X}\left(z, \frac{\beta}{\alpha}X\right)}{\hat{f}_{\frac{\beta}{\alpha}X}\left(\frac{\beta}{\alpha}X\right)}.$$

The rest of the derivation of the asymptotics is standard for the non-parametric conditional density (see, e.g., Li and Racine (2007)), whereas the uniform convergence follows from the uniform convergence of its elements.

□

Denote $m(y) = F_Z^{-1}\left(F_Z(y) - (X^k - X^m)\frac{\beta}{\alpha}\right)$ and $\hat{m}(y) = \hat{F}_Z^{-1}\left(\hat{F}_Z(y) - (X^k - X^m)\frac{\hat{\beta}}{\alpha}\right)$.

Lemma 8. *Suppose Assumptions 1–9 hold, then,* $\sqrt{N}(\hat{m}(y) - m(y)) \xrightarrow{d} N(0, V_y)$, *where* V_y *is a some positive number.*

Proof. Denote $\hat{w} = \hat{m}(y)$ and $w = m(y)$ for simplicity of the notation in this proof. By definition, $F_Z(w) = \frac{\beta}{\alpha}'(X^k - X^m) + F_Z(y)$, implying

$$\widehat{F_Z}(\hat{w}) - F_Z(w) = \left(\frac{\hat{\beta}}{\alpha} - \frac{\beta}{\alpha}\right)'(X^k - X^m) + \widehat{F_Z}(y) - F_Z(y).$$

In addition, the left-hand side can be rewritten as

$$\widehat{F_Z}(\hat{w}) - F_Z(w) = \widehat{F_Z}(\hat{w}) - \widehat{F_Z}(w) + \widehat{F_Z}(w) - F_Z(w).$$

Hence,

$$\widehat{F_Z}(\hat{w}) - \widehat{F_Z}(w) = -\left(\widehat{F_Z}(w) - F_Z(w)\right) + \left(\frac{\hat{\beta}}{\alpha} - \frac{\beta}{\alpha}\right)' (X^k - X^m) + \widehat{F_Z}(y) - F_Z(y).$$

By analogy with Lemma 5, it can be shown that the asymptotics of $\widehat{F_Z}(y)$ is standard with rate $\sqrt{N}$ except a change in the variance. Because we do not need the exact variance of this estimator in the paper, and for the sake of saving space, we do not provide the formal derivations. In addition, $\frac{\hat{\beta}}{\alpha}$ converge with rate $\sqrt{N}$, implying that for some positive number $\tilde{V}$, we must have $\sqrt{N}(\widehat{F_Z}(\hat{w}) - \widehat{F_Z}(w)) \xrightarrow{d} N(0, \tilde{V})$. On the other hand, by the Taylor expansion, we get

$$\widehat{F_Z}(\hat{w}) - \widehat{F_Z}(w) = f_Z(w)(\hat{w} - w) + h_Z(w)(\hat{w} - w) + Res.$$

Note $h_Z(w) = \hat{f}_Z(w) - f_Z(w)$ converges with rate $\sqrt{Nh}$, implying

$$\sqrt{N}(\hat{w} - w) \approx \sqrt{N}\frac{\widehat{F_Z}(\hat{w}) - \widehat{F_Z}(w)}{f_Z(w)} \xrightarrow{d} N\left(0, V_y\right),$$

where $V_y = \frac{\tilde{V}}{f_Z^2(w)} = \frac{\tilde{V}}{f_Z^2(m(y))}$.

□

Lemma 9. *Suppose Assumptions 1–9 hold, then,* $\sqrt{N}(\hat{\tilde{y}} - \tilde{y}) \xrightarrow{d} N(0, V_{\tilde{y}})$ *and* $\sqrt{N}(\hat{\tilde{z}} - \tilde{z}) \xrightarrow{d} N(0, V_{\tilde{z}})$*, where* $V_{\tilde{y}}$ *and* $V_{\tilde{z}}$ *are some positive numbers.*

Proof. The lemma is the direct corollary of Lemma 8, simply note $\tilde{y} = m(Y^f - q)$ and $\tilde{z} = m(Z)$. □

Lemma 10. *Suppose Assumptions 1–9 hold, then,*

$$\sup_{\tilde{y} \in \mathbb{R}} |\hat{f}_Z(\hat{\tilde{y}}) - f_Z(\tilde{y})| \xrightarrow{p} 0$$

$$\sqrt{Nh}(\hat{f}_Z(\hat{\tilde{y}}) - f_Z(\tilde{y})) \xrightarrow{d} N\left(0, R(k) f_Z(\tilde{y}) \frac{\mu_{t|z=\tilde{y}}}{\mu_t}\right).$$

Proof. Note

$$\begin{aligned}\hat{f}_Z(\hat{\tilde{y}}) - f_Z(\tilde{y}) &= \hat{f}_Z(\hat{\tilde{y}}) - \hat{f}_Z(\tilde{y}) + \hat{f}_Z(\tilde{y}) - f_Z(\tilde{y}) \\ &= f'_Z(\tilde{y})(\hat{\tilde{y}} - \tilde{y}) + \frac{\partial h_Z(\tilde{y})}{\partial \tilde{y}}(\hat{\tilde{y}} - \tilde{y}) + \hat{f}_Z(\tilde{y}) - f_Z(\tilde{y}) + Res.\end{aligned}$$

Denote $R(\phi, h_Z, \tilde{y}, h_{\tilde{y}}) = \frac{\partial h_Z(\tilde{y})}{\partial \tilde{y}}(\hat{\tilde{y}} - \tilde{y}) + Res$; then, for some $a > 0$, $|R(\phi, h_Z, \tilde{y}, h_{\tilde{y}})| \leq a||h||$ due to the Taylor Theorem and the fact that $\left|\frac{\partial h_Z(\tilde{y})}{\partial \tilde{y}}\right|$ is bounded and converges uniformly by Lemma B.3 in Newey (1994) under Assumption 8. Thus, $\sup_{\tilde{y}\in\mathbb{R}} |\hat{f}_Z(\hat{\tilde{y}}) - f_Z(\tilde{y})| \xrightarrow{p} 0$.

Now also by Lemma B.3 in Newey (1994), we have $\sqrt{Nh}R(\phi, h_Z, \tilde{y}, h_y) \xrightarrow{p} 0$. By Lemma 9, $h_{\tilde{y}}$ converges as $\sqrt{N}$ and, by Lemma 5, h_Z converges as $\sqrt{Nh}$; thus, the result follows: $\sqrt{Nh}(\hat{f}_Z(\hat{\tilde{y}}) - f_Z(\tilde{y})) \xrightarrow{d} N\left(0, R(k) f_Z(\tilde{y}) \frac{\mu_{t|z=\tilde{y}}}{\mu_t}\right)$. □

Lemma 11. *If Assumptions 1–9 hold, then,*

$$\sup_{(\tilde{y},\tilde{z})} |\hat{f}_{Y,Z|X^m}(\hat{\tilde{y}}, \hat{\tilde{z}}) - f_{Y,Z|X^m}(\tilde{y}, \tilde{z})| \xrightarrow{p} 0$$

$$\sqrt{N_m}(\hat{f}_{Y,Z|X^m}(\hat{\tilde{y}}, \hat{\tilde{z}}) - f_{Y,Z|X^m}(\tilde{y}, \tilde{z})) \xrightarrow{d} N(0, V_{f^m}).$$

Proof. Note

$$\begin{aligned}\hat{f}_{Y,Z|X^m}(\hat{\tilde{y}}, \hat{\tilde{z}}) - f_{Y,Z|X^m}(\tilde{y}, \tilde{z}) =& \hat{f}_{Y,Z|X^m}(\hat{\tilde{y}}, \hat{\tilde{z}}) - \hat{f}_{Y,Z|X^m}(\tilde{y}, \tilde{z}) + \hat{f}_{Y,Z|X^m}(\tilde{y}, \tilde{z}) - f_{Y,Z|X^m}(\tilde{y}, \tilde{z}) \\ =& \frac{\partial f_{Y,Z|X^m}(\tilde{y}, \tilde{z})}{\partial \tilde{y}} h_{\tilde{y}} + \frac{\partial f_{Y,Z|X^m}(\tilde{y}, \tilde{z})}{\partial \tilde{z}} h_{\tilde{z}} + h_{Y,Z|X^m}(\tilde{y}, \tilde{z}) \\ &+ R(f_{Y,Z|X^m}, h_{Y,Z|X^m}, \tilde{y}, h_{\tilde{y}}, \tilde{z}, h_{\tilde{z}}).\end{aligned}$$

By the Taylor Theorem, there is $a > 0$ such that $|R(f_{Y,Z|X^m}, h_{Y,Z|X^m}, \tilde{y}, h_{\tilde{y}}, \tilde{z}, h_{\tilde{z}})| \leq a||h||$; hence, the uniform convergence of the estimator follows.

Also, we have $\sqrt{N_m}R(f_{Y,Z|X^m}, h_{Y,Z|X^m}, \tilde{y}, h_{\tilde{y}}, \tilde{z}, h_{\tilde{z}}) \xrightarrow{p} 0$, and by Lemma 9, $h_{\tilde{y}}$ and $h_{\tilde{z}}$ converge with the rate $\sqrt{N}$, whereas $h_{Y,Z|X^m}(y, z)$ has slower speed $\sqrt{N_m}$, and the result follows. □

Lemma 12. *Let* $\hat{f}_{Z|\frac{\hat{\beta}}{\alpha}'X}(z)$ *be the estimator of* $f_{Z|\frac{\beta}{\alpha}'X}(z)$*. Then, under Assumptions 1–9,*

$$\sup_{z,\frac{\beta}{\alpha}'X}\left|\hat{f}_{Z|\frac{\hat{\beta}}{\alpha}'X}(z)-f_{Z|\frac{\beta}{\alpha}'X}(z)\right|\xrightarrow{p}0$$

$$\sqrt{Nh^2}\left(\hat{f}_{Z|\frac{\hat{\beta}}{\alpha}'X}(z)-f_{Z|\frac{\beta}{\alpha}'X}(z)\right)\xrightarrow{d}N(0,V_{Z|X}).$$

Proof. Note that by the Taylor expansion,

$$\begin{aligned}\hat{f}_{Z|\frac{\hat{\beta}}{\alpha}'X}(z)-f_{Z|\frac{\beta}{\alpha}'X}(z)&=\hat{f}_{Z|\frac{\hat{\beta}}{\alpha}'X}(z)-\hat{f}_{Z|\frac{\beta}{\alpha}'X}(z)+\hat{f}_{Z|\frac{\beta}{\alpha}'X}(z)-f_{Z|\frac{\beta}{\alpha}'X}(z)\\&=\sum_{i=1}^{L}\frac{\partial f_{Z|\frac{\beta}{\alpha}'X}(z)}{\partial\frac{\beta_i}{\alpha}}\left(\frac{\hat{\beta}_i}{\alpha}-\frac{\beta_i}{\alpha}\right)+\sum_{i=1}^{L}\frac{\partial h_{Z|\frac{\beta}{\alpha}'X}(z)}{\partial\frac{\beta_i}{\alpha}}\left(\frac{\hat{\beta}_i}{\alpha}-\frac{\beta_i}{\alpha}\right)\\&+\hat{f}_{Z|\frac{\beta}{\alpha}'X}(z)-f_{Z|\frac{\beta}{\alpha}'X}(z)+Res.\end{aligned}$$

Denote $R\left(f_{Z|\frac{\beta}{\alpha}'X},h_{Z|\frac{\beta}{\alpha}'X}\right)=\sum_{i=1}^{L}\frac{\partial h_{Z|\frac{\beta}{\alpha}'X}(z)}{\partial\frac{\beta_i}{\alpha}}\left(\frac{\hat{\beta}_i}{\alpha}-\frac{\beta_i}{\alpha}\right)+Res$. Note $\left|\frac{\partial h_{Z|\frac{\beta}{\alpha}'X}(y,z)}{\partial\frac{\beta_i}{\alpha}}\right|$ is bounded by $||h||$; thus, $|Res|$ and $\left|R(f_{Z|\frac{\beta}{\alpha}'X},h_{Z|\frac{\beta}{\alpha}'X})\right|$ are bounded by $||h||^2$ due to the Taylor Theorem. As a result, we obtain uniform convergence of $\hat{f}_{Z|\frac{\hat{\beta}}{\alpha}'X}(z)-f_{Z|\frac{\beta}{\alpha}'X}(z)$.

Now recall that $\left(\frac{\hat{\beta}}{\alpha}-\frac{\beta}{\alpha}\right)$ converge with speed $\sqrt{N}$, whereas, by Lemma 7, $\hat{f}_{Z|\frac{\beta}{\alpha}'X}(z)-f_{Z|\frac{\beta}{\alpha}'X}(z)$ has rate $\sqrt{Nh^2}$. Additionally, under Assumption 8, $\sqrt{Nh^2}R\left(f_{Z|\frac{\beta}{\alpha}'X},h_{Z|\frac{\beta}{\alpha}'X}\right)\xrightarrow{p}0$ by Lemma B.3 in Newey (1994). Thus, the result follows.

□

Proof of Theorem 3. Recall that

$$\hat{f}_{q|X^j,Y^f,Z}(q)=\frac{\hat{f}_{Y,Z|X^m}(\hat{\tilde{y}},\hat{\tilde{z}})}{\hat{f}_{Z|\frac{\hat{\beta}}{\alpha}'X^k}(Z)}\frac{\hat{f}_Z(Y^f-q)\hat{f}_Z(Z)}{\hat{f}_Z(\hat{\tilde{y}})\hat{f}_Z(\hat{\tilde{z}})}.$$

Denote by $p(\tilde{y},\tilde{z},q,Y^f,Z)=\frac{f_Z(Y^f-q)f_Z(Z)}{f_Z(\tilde{y})f_Z(\tilde{z})}$. Define the functional

$$\Phi(p,G)=p(\tilde{y},\tilde{z},q,Y^f,Z)\frac{g_{Y,Z|X^m}(\tilde{y},\tilde{z})}{g_{Z|\frac{\beta}{\alpha}'X^k}(Z)}.$$

Then, $\Phi(\hat{p}, \hat{F}) = \hat{f}_{q|X^j,Y^f,Z}(q)$ and $\Phi(p, F) = f_{q|X^j,Y^f,Z}(q)$. For any h_Z such that $||h||$ is sufficiently small: $|h_Z(y)| \leq a||h||$ and $|h_{Y,Z|X^m}(y,z)| \leq a||h||$ for some $0 < a < \infty$. Also,

$$\Phi(p + h, F + H) - \Phi(p, F) = D\Phi(p, F, h, H) + R\Phi(p, F, h, H),$$

where

$$\begin{aligned} D\Phi(p, F, h, H) = &\frac{f_{Y,Z|X^m}(\tilde{y}, \tilde{z})}{f_{Z|\frac{\beta}{\alpha}'X^k}(Z)} h_p(\tilde{y}, \tilde{z}, q, Y^f, Z) + \frac{p(\tilde{y}, \tilde{z}, q, Y^f, Z)}{f_{Z|\frac{\beta}{\alpha}'X^k}(Z)} h_{Y,Z|X^m}(\tilde{y}, \tilde{z}) \\ &- \frac{f_{Y,Z|X^m}(\tilde{y}, \tilde{z}) p(\tilde{y}, \tilde{z}, q, Y^f, Z)}{f^2_{Z|\frac{\beta}{\alpha}'X^k}(Z)} h_{Z|\frac{\beta}{\alpha}'X^j}(Z). \end{aligned}$$

Thus, for some $c < \infty$:

$$|D\Phi(p, F, h, H)| \leq c||h|| \text{ and } |R\Phi(p, F, h, H)| \leq c||h||^2,$$

implying uniform convergence. In addition, by Lemma 5, $h_p(\tilde{y}, \tilde{z}, q, Y^f, Z)$ converges with rate $\sqrt{Nh}$; by Lemma 11, $h_{Y,Z|X^m}(\tilde{y}, \tilde{z})$ converges as $\sqrt{N_m}$; and by Lemma 12, $h_{Z|\frac{\beta}{\alpha}'X^k}(Z)$ converges as $\sqrt{Nh^2}$. Assumption 10 guarantees $h_{Y,Z|X^m}(\tilde{y}, \tilde{z})$ converges slower than $h_{Z|\frac{\beta}{\alpha}'X^k}(Z)$; thus,

$$\sqrt{N_m} D\Phi(p, F, h, H) \xrightarrow{d} N(0, QV_{f^m}).$$

Now, by Lemma B.3 in Newey (1994), $\sqrt{N_m} R\Phi(p, F, h, H) \xrightarrow{p} 0$ and the result follows. □

Lemma 13. *Suppose Assumptions 1-9 hold, then,* $\sqrt{N_m}(\hat{\mu}_{Y|\hat{\tilde{z}}} - \mu_{Y|\tilde{z}}) \xrightarrow{d} N(0, \sigma^2_{\tilde{y}|\tilde{z}})$.

Proof. Recall that $\hat{\mu}_{Y|\hat{\tilde{z}}} = \hat{\mu}_Y + \frac{\hat{\sigma}_{YZ}}{\hat{\sigma}^2_Z}(\hat{\tilde{z}} - \hat{\mu}_Z)$ and rewrite

$$\begin{aligned} \hat{\mu}_{Y|\hat{\tilde{z}}} - \mu_{Y|\tilde{z}} = &\left(\hat{\mu}_Y + \frac{\hat{\sigma}_{YZ}}{\hat{\sigma}^2_Z}(\hat{\tilde{z}} - \hat{\mu}_Z)\right) - \left(\hat{\mu}_Y + \frac{\hat{\sigma}_{YZ}}{\hat{\sigma}^2_Z}(\tilde{z} - \hat{\mu}_Z)\right) \\ &+ \left(\hat{\mu}_Y + \frac{\hat{\sigma}_{YZ}}{\hat{\sigma}^2_Z}(\tilde{z} - \hat{\mu}_Z)\right) - \left(\mu_Y + \frac{\sigma_{YZ}}{\sigma^2_Z}(\tilde{z} - \mu_Z)\right). \end{aligned}$$

First, note

$$\left(\hat{\mu}_Y + \frac{\hat{\sigma}_{YZ}}{\hat{\sigma}_Z^2}(\hat{\tilde{z}} - \hat{\mu}_Z)\right) - \left(\hat{\mu}_Y + \frac{\hat{\sigma}_{YZ}}{\hat{\sigma}_Z^2}(\tilde{z} - \hat{\mu}_Z)\right) = \frac{\hat{\sigma}_{YZ}}{\hat{\sigma}_Z^2}(\hat{\tilde{z}} - \tilde{z}),$$

which by Lemma 9 converges with the rate $\sqrt{N}$, whereas by Lemma 4, the second component converges with the slower rate $\sqrt{N_m}$ and defines the asymptotic distribution.

□

Proof of Theorem 4. Recall that

$$E(q|\widehat{X^k, Y^f}, Z) = Y^f - \hat{E}\left(\hat{F}_Z^{-1}\left(\hat{F}_Z(\tilde{y}) - (X^k - X^m)\frac{\hat{\beta}}{\alpha}\right)|\hat{\tilde{z}}\right).$$

For simplicity of notation, denote $m(y) = F_Z^{-1}\left(F_Z(y) - (X^k - X^m)\frac{\beta}{\alpha}\right)$ and $\hat{m}(y) = \hat{F}_Z^{-1}\left(\hat{F}_Z(y) - (X^k - X^m)\frac{\hat{\beta}}{\alpha}\right)$. The analogy estimator for $\hat{E}(\hat{m}(\tilde{y})|\hat{\tilde{z}})$ would require generating T observations from the distribution of $\tilde{y}|\hat{\tilde{z}}$ and taking the average of $\hat{m}(\cdot)$ at these observations. More formally, denote by $\widehat{\tilde{y}_i}$ an observation drawn from the distribution of $\tilde{y}|\hat{\tilde{z}}$; then,

$$\hat{E}\left(\hat{F}_Z^{-1}\left(\hat{F}_Z(\tilde{y}) - (X^k - X^m)\frac{\hat{\beta}}{\alpha}\right)|\hat{\tilde{z}}\right) = \frac{1}{T}\sum_{i=1}^{T}\hat{m}(\widehat{\tilde{y}_i}).$$

We use a hat over $\widehat{\tilde{y}_i}$ to denote the fact that it is drawn from an approximated distribution due to conditioning on $\hat{\tilde{z}}$ and using a distribution of $(\tilde{y}, \tilde{z})$ with the estimated parameters. As for the asymptotic distribution, note the following:

$$\frac{1}{T}\sum_{i=1}^{T}\hat{m}(\widehat{\tilde{y}_i}) - \mu_{Y|\tilde{z}} = \frac{1}{T}\sum_{i=1}^{T}\left(\hat{m}(\widehat{\tilde{y}_i}) - m(\widehat{\tilde{y}_i})\right) + \frac{1}{T}\sum_{i=1}^{T}m(\widehat{\tilde{y}_i}) - \hat{\mu}_{Y|\tilde{z}} + \hat{\mu}_{Y|\tilde{z}} - \mu_{Y|\tilde{z}}.$$

In this case, by Lemma 8, $\hat{m}(\widehat{\tilde{y}_i}) - m(\widehat{\tilde{y}_i})$ converges with the rate $\sqrt{N}$; by the CLT, $\frac{1}{T}\sum_{i=1}^{T}m(\widehat{\tilde{y}_i}) - \hat{\mu}_{Y|\tilde{z}}$ converges with the rate $\sqrt{T}$, which is the variable of choice of the researcher and can be made as large as needed. Finally, by Lemma 13, $\hat{\mu}_{Y|\tilde{z}} - \mu_{Y|\tilde{z}}$

converges with the rate $\sqrt{N_m}$, which is the slowest component of the sum. Hence, the asymptotic distribution will be driven by the latest element, implying the result. $\square$

Proof of Theorem 5. The result follows directly from Lemma 13. $\square$

Lemma 14. *Suppose Assumptions 1-9 hold, then,*

$$\sqrt{N_m}\left(\begin{bmatrix}\frac{1}{\hat{\sigma}_{Y|Z}}\\ \frac{\hat{\mu}_{Y|\tilde{z}}}{\hat{\sigma}_{Y|Z}}\end{bmatrix}-\begin{bmatrix}\frac{1}{\sigma_{Y|Z}}\\ \frac{\mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\end{bmatrix}\right)\xrightarrow{d}N(0,\tilde{W})$$

Proof. From Lemma 2, we have $\sqrt{N_m}(\hat{\theta}-\theta)\xrightarrow{d}N(0,V_\theta)$. Hence, we can apply the Δ-method to derive the desired result. For the purposes of this proof only, denote $\begin{bmatrix}\frac{1}{\sigma_{Y|Z}}\\ \frac{\mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\end{bmatrix}=g(\theta)$. The matrix of partial derivatives ∂G consists of the following vectors:

$$\frac{\partial g(\theta)}{\partial\mu_Y}=\begin{bmatrix}0\\ \frac{1}{\sigma_{Y|Z}}\end{bmatrix};\ \frac{\partial g(\theta)}{\partial\mu_Z}=\begin{bmatrix}0\\ -\frac{\sigma_{YZ}}{\sigma_Z^2\sigma_{Y|Z}}\end{bmatrix};\ \frac{\partial g(\theta)}{\partial\sigma_Y^2}=\begin{bmatrix}-\frac{1}{2\sigma_{Y|Z}^3}\\ -\frac{Y|\tilde{z}}{2\sigma_{Y|Z}^3}\end{bmatrix}$$

$$\frac{\partial g(\theta)}{\partial\sigma_Z^2}=\begin{bmatrix}-\frac{\sigma_{YZ}^2}{2\sigma_{Y|Z}^3\sigma_Z^4}\\ -\frac{\sigma_{YZ}(\sigma_{YZ}\mu_{Y|\tilde{z}}+2\sigma_{Y|Z}^2(\tilde{z}-\mu_Z))}{2\sigma_{Y|Z}^3}\sigma_Z^4\end{bmatrix};\ \frac{\partial g(\theta)}{\partial\sigma_{YZ}}=\begin{bmatrix}\frac{1}{\sigma_{Y|Z}^3\sigma_Z^2}\\ \frac{\mu_{Y|\tilde{z}}+\sigma_{Y|Z}^2(\tilde{z}-\mu_Z)}{\sigma_{Y|Z}^3\sigma_Z^2}\end{bmatrix}.$$

Note V_θ consists of 2 blocks, so

$$(\partial G)V_\theta(\partial G)'=\left[\frac{\partial g(\theta)}{\partial\mu_Y},\frac{\partial g(\theta)}{\partial\mu_Z}\right]\Sigma\left[\frac{\partial g(\theta)}{\partial\mu_Y},\frac{\partial g(\theta)}{\partial\mu_Z}\right]'+(\partial T)\Omega(\partial T)'.$$

The first component simplifies to $\begin{bmatrix}0&0\\0&1\end{bmatrix}$; hence, $\sqrt{N_m}\left(\begin{bmatrix}\frac{1}{\hat{\sigma}_{Y|Z}}\\ \frac{\hat{\mu}_{Y|\tilde{z}}}{\hat{\sigma}_{Y|Z}}\end{bmatrix}-\begin{bmatrix}\frac{1}{\sigma_{Y|Z}}\\ \frac{\mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\end{bmatrix}\right)\xrightarrow{d}N(0,\tilde{W})$.

$\square$

Lemma 15. *Suppose Assumptions 1-9 hold, then,*

$$\sqrt{N_m}\left(\begin{bmatrix}\frac{1}{\hat{\sigma}_{Y|Z}}\\ \frac{\hat{\mu}_{Y|\hat{\tilde{z}}}}{\hat{\sigma}_{Y|Z}}\end{bmatrix}-\begin{bmatrix}\frac{1}{\sigma_{Y|Z}}\\ \frac{\mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\end{bmatrix}\right)\xrightarrow{d}N(0,\tilde{W}).$$

Proof. Note only $\mu_{Y|\tilde{z}}$ depends on $\tilde{z}$; hence,

$$\begin{bmatrix} \frac{1}{\hat{\sigma}_{Y|Z}} \\ \frac{\hat{\mu}_{Y|\hat{\tilde{z}}}}{\hat{\sigma}_{Y|Z}} \end{bmatrix} - \begin{bmatrix} \frac{1}{\sigma_{Y|Z}} \\ \frac{\mu_{Y|\tilde{z}}}{\sigma_{Y|Z}} \end{bmatrix} = \left(\begin{bmatrix} \frac{1}{\hat{\sigma}_{Y|Z}} \\ \frac{\hat{\mu}_{Y|\hat{\tilde{z}}}}{\hat{\sigma}_{Y|Z}} \end{bmatrix} - \begin{bmatrix} \frac{1}{\hat{\sigma}_{Y|Z}} \\ \frac{\hat{\mu}_{Y|\tilde{z}}}{\hat{\sigma}_{Y|Z}} \end{bmatrix} \right) + \left(\begin{bmatrix} \frac{1}{\hat{\sigma}_{Y|Z}} \\ \frac{\hat{\mu}_{Y|\tilde{z}}}{\hat{\sigma}_{Y|Z}} \end{bmatrix} - \begin{bmatrix} \frac{1}{\sigma_{Y|Z}} \\ \frac{\mu_{Y|\tilde{z}}}{\sigma_{Y|Z}} \end{bmatrix} \right)$$

$$= \begin{bmatrix} 0 \\ \frac{\sigma_{YZ}}{\sigma_{Y|Z}\sigma_Z^2} \end{bmatrix} (\hat{\tilde{z}} - \tilde{z}) + \left(\begin{bmatrix} \frac{1}{\hat{\sigma}_{Y|Z}} \\ \frac{\hat{\mu}_{Y|\tilde{z}}}{\hat{\sigma}_{Y|Z}} \end{bmatrix} - \begin{bmatrix} \frac{1}{\sigma_{Y|Z}} \\ \frac{\mu_{Y|\tilde{z}}}{\sigma_{Y|Z}} \end{bmatrix} \right) + Res.$$

By Lemma 9, $\hat{\tilde{z}}$ converges with the rate $\sqrt{N}$, whereas by Lemma 14, the second component converges with the slower rate $\sqrt{N_m}$ and defines the asymptotic distribution.

□

Lemma 16. *Suppose Assumptions 1-9 hold, then,*

$$\sqrt{N_m}\left(\Phi\left(\frac{y - \hat{\mu}_{Y|\hat{\tilde{z}}}}{\hat{\sigma}_{Y|Z}}\right) - \Phi\left(\frac{y_0 - \hat{\mu}_{Y|\hat{\tilde{z}}}}{\hat{\sigma}_{Y|Z}}\right) - \Phi\left(\frac{y - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right) + \Phi\left(\frac{y_0 - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right)\right) \xrightarrow{d} N(0, V_\Phi).$$

Proof. Notice

$$\begin{aligned}
&\Phi\left(\frac{y - \hat{\mu}_{Y|\hat{\tilde{z}}}}{\hat{\sigma}_{Y|Z}}\right) - \Phi\left(\frac{y_0 - \hat{\mu}_{Y|\hat{\tilde{z}}}}{\hat{\sigma}_{Y|Z}}\right) - \Phi\left(\frac{y - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right) + \Phi\left(\frac{y_0 - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right) \\
&\quad =\phi\left(\frac{y - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right)\left(\frac{y - \hat{\mu}_{Y|\hat{\tilde{z}}}}{\hat{\sigma}_{Y|Z}} - \frac{y - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right) - \phi\left(\frac{y_0 - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right)\left(\frac{y_0 - \hat{\mu}_{Y|\hat{\tilde{z}}}}{\hat{\sigma}_{Y|Z}} - \frac{y_0 - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right) + Res \\
&\quad = \left(y\phi\left(\frac{y - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right) - y_0\phi\left(\frac{y_0 - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right)\right)\left(\frac{1}{\hat{\sigma}_{Y|Z}} - \frac{1}{\sigma_{Y|Z}}\right) \\
&\qquad - \left(\phi\left(\frac{y - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right) - \phi\left(\frac{y_0 - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right)\right)\left(\frac{\hat{\mu}_{Y|\hat{\tilde{z}}}}{\hat{\sigma}_{Y|Z}} - \frac{\mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right) + Res.
\end{aligned}$$

By Lemma 15, $\sqrt{N_m}\left(\begin{bmatrix} \frac{1}{\hat{\sigma}_{Y|Z}} \\ \frac{\hat{\mu}_{Y|\hat{\tilde{z}}}}{\hat{\sigma}_{Y|Z}} \end{bmatrix} - \begin{bmatrix} \frac{1}{\sigma_{Y|Z}} \\ \frac{\mu_{Y|\tilde{z}}}{\sigma_{Y|Z}} \end{bmatrix} \right) \xrightarrow{d} N(0, \tilde{W})$. Hence, by denoting $\tilde{W}_{ij}$ the (ij)-th element of the matrix $\tilde{W}$, we obtain the following result:

$$\sqrt{N_m}\left(\Phi\left(\frac{y - \hat{\mu}_{Y|\hat{\tilde{z}}}}{\hat{\sigma}_{Y|Z}}\right) - \Phi\left(\frac{y_0 - \hat{\mu}_{Y|\hat{\tilde{z}}}}{\hat{\sigma}_{Y|Z}}\right) - \Phi\left(\frac{y - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right) + \Phi\left(\frac{y_0 - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right)\right) \xrightarrow{d} N(0, V_\Phi),$$

where

$$V_\Phi =\phi^2\left(\frac{y-\mu_{Y|\tilde z}}{\sigma_{Y|Z}}\right)(y^2\tilde W_{11}+\tilde W_{22}-2y\tilde W_{12})+\phi^2\left(\frac{y_0-\mu_{Y|\tilde z}}{\sigma_{Y|Z}}\right)(y_0^2\tilde W_{11}+\tilde W_{22}-2y_0\tilde W_{12})$$
$$-2\phi\left(\frac{y-\mu_{Y|\tilde z}}{\sigma_{Y|Z}}\right)\phi\left(\frac{y_0-\mu_{Y|\tilde z}}{\sigma_{Y|Z}}\right)(yy_0\tilde W_{11}+\tilde W_{22}-\tilde W_{12}(y+y_0)).$$

□

Lemma 17. *Suppose Assumptions 1-9 hold, then,*

$$\sqrt{N_m}\left(\Phi\left(\frac{\hat y_{\max}-\hat\mu_{Y|\hat{\tilde z}}}{\hat\sigma_{Y|Z}}\right)-\Phi\left(\frac{\hat y_0-\hat\mu_{Y|\hat{\tilde z}}}{\hat\sigma_{Y|Z}}\right)-\Phi\left(\frac{y_{\max}-\mu_{Y|\tilde z}}{\sigma_{Y|Z}}\right)+\Phi\left(\frac{y_0-\mu_{Y|\tilde z}}{\sigma_{Y|Z}}\right)\right)\xrightarrow{d}N(0,V_\Phi).$$

Proof. Note that

$$\Phi\left(\frac{\hat y_{\max}-\hat\mu_{Y|\hat{\tilde z}}}{\hat\sigma_{Y|Z}}\right)-\Phi\left(\frac{\hat y_0-\hat\mu_{Y|\hat{\tilde z}}}{\hat\sigma_{Y|Z}}\right)-\Phi\left(\frac{y_{\max}-\mu_{Y|\tilde z}}{\sigma_{Y|Z}}\right)+\Phi\left(\frac{y_0-\mu_{Y|\tilde z}}{\sigma_{Y|Z}}\right)$$
$$=\Phi\left(\frac{\hat y_{\max}-\hat\mu_{Y|\hat{\tilde z}}}{\hat\sigma_{Y|Z}}\right)-\Phi\left(\frac{\hat y_0-\hat\mu_{Y|\hat{\tilde z}}}{\hat\sigma_{Y|Z}}\right)-\Phi\left(\frac{y_{\max}-\hat\mu_{Y|\tilde z}}{\hat\sigma_{Y|Z}}\right)+\Phi\left(\frac{y_0-\hat\mu_{Y|\tilde z}}{\hat\sigma_{Y|Z}}\right)$$
$$+\Phi\left(\frac{y_{\max}-\hat\mu_{Y|\hat{\tilde z}}}{\hat\sigma_{Y|Z}}\right)-\Phi\left(\frac{y_0-\hat\mu_{Y|\hat{\tilde z}}}{\hat\sigma_{Y|Z}}\right)-\Phi\left(\frac{y_{\max}-\mu_{Y|\tilde z}}{\sigma_{Y|Z}}\right)+\Phi\left(\frac{y_0-\mu_{Y|\tilde z}}{\sigma_{Y|Z}}\right)$$
$$=\phi\left(\frac{y_{\max}-\mu_{Y|\tilde z}}{\sigma_{Y|Z}}\right)(\hat y_{\max}-y_{\max})-\phi\left(\frac{y_0-\mu_{Y|\tilde z}}{\sigma_{Y|Z}}\right)(\hat y_0-y_0)+Res$$
$$+\Phi\left(\frac{y_{\max}-\hat\mu_{Y|\hat{\tilde z}}}{\hat\sigma_{Y|Z}}\right)-\Phi\left(\frac{y_0-\hat\mu_{Y|\hat{\tilde z}}}{\hat\sigma_{Y|Z}}\right)-\Phi\left(\frac{y_{\max}-\mu_{Y|\tilde z}}{\sigma_{Y|Z}}\right)+\Phi\left(\frac{y_0-\mu_{Y|\tilde z}}{\sigma_{Y|Z}}\right).$$

First, recall that $y_{\max}=F_Z^{-1}\left(\frac{\beta}{\alpha}(X^k-X^m)+F_Z(Y^f)\right)$, and hence by Lemma 8, $\hat y_{\max}$ converges with the rate $\sqrt{N}$. In addition, if $\beta(X^k-X^m)<0$, then y_0 is a constant, implying $\hat y_0-y_0=0$. However, if $\beta(X^k-X^m)\geq 0$, then $y_0=F_Z^{-1}(\frac{\beta}{\alpha}(X^k-X^m))$ and it would converge with $\sqrt{N}$ as well. Hence, the residual from the Taylor expansion would converge as $o(\sqrt{N})$.

Next, by Lemma 16, the last component converges with the slowest rate $\sqrt{N_m}$ and it also defines the asymptotic variance of the original estimator, which is V_Φ.

□

Proof of Theorem 6. Recall

$$\hat{q}_l = \hat{q}_{\max}\left(1 - \frac{\hat{f}_Z(Z)}{\hat{f}_{Z|\frac{\hat{\beta}}{\alpha}X}(Z)}\left(\Phi\left(\frac{\hat{y}_{\max} - \hat{\mu}_{Y|\hat{\tilde{z}}}}{\hat{\sigma}_{Y|Z}}\right) - \Phi\left(\frac{\hat{y}_0 - \hat{\mu}_{Y|\hat{\tilde{z}}}}{\hat{\sigma}_{Y|Z}}\right)\right)\right).$$

Hence,

$$\begin{aligned}
\hat{q}_l - q_l = &\left(1 - \frac{f_Z(Z)}{f_{Z|\frac{\beta}{\alpha}X}(Z)}\left(\Phi\left(\frac{y_{\max} - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right) - \Phi\left(\frac{y_0 - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right)\right)\right)(\hat{q}_{\max} - q_{\max}) \\
&- q_{\max}\frac{f_Z(Z)}{f_{Z|\frac{\beta}{\alpha}X}(Z)}\left(\Phi\left(\frac{\hat{y}_{\max} - \hat{\mu}_{Y|\hat{\tilde{z}}}}{\hat{\sigma}_{Y|Z}}\right) - \Phi\left(\frac{\hat{y}_0 - \hat{\mu}_{Y|\hat{\tilde{z}}}}{\hat{\sigma}_{Y|Z}}\right) - \Phi\left(\frac{y_{\max} - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right) + \Phi\left(\frac{y_0 - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right)\right) \\
&- q_{\max}\left(\Phi\left(\frac{y_{\max} - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right) - \Phi\left(\frac{y_0 - \mu_{Y|\tilde{z}}}{\sigma_{Y|Z}}\right)\right)\left(\frac{\hat{f}_Z(Z)}{\hat{f}_{Z|\frac{\hat{\beta}}{\alpha}X}(Z)} - \frac{f_Z(Z)}{f_{Z|\frac{\beta}{\alpha}X}(Z)}\right) + Res.
\end{aligned}$$

First, if $\beta(X^k - X^m) \geq 0$, then $q_{\max} = Y^f - z_0$, implying it is a constant conditional on observables, so $\hat{q}_{\max} - q_{\max} = 0$. If $\beta(X^k - X^m) < 0$, then $q_{\max} = Y^f - F_Z^{-1}\left(-\frac{\beta}{\alpha}(X^k - X^m)\right)$, implying the rate of convergence $\sqrt{N}$.

Second, by Lemmas 5 and 12, $\frac{\hat{f}_Z(Z)}{\hat{f}_{Z|\frac{\hat{\beta}}{\alpha}X}(Z)}$ converges with the rate $\sqrt{Nh^2}$.

Third, by Lemma 17, $\Phi\left(\frac{\hat{y}_{\max} - \hat{\mu}_{Y|\hat{\tilde{z}}}}{\hat{\sigma}_{Y|Z}}\right) - \Phi\left(\frac{\hat{y}_0 - \hat{\mu}_{Y|\hat{\tilde{z}}}}{\hat{\sigma}_{Y|Z}}\right)$ converges with the rate $\sqrt{N_m}$, implying it is the slowest component. In addition, by Lemma B.3 in Newey (1994), $\sqrt{N_m}Res \xrightarrow{p} 0$. Thus, we obtain that $\sqrt{N_m}(\hat{q}_l - q_l) \xrightarrow{d} N(0, V_{q_l})$, where $V_{q_l} = q_{\max}^2 \frac{f_Z^2(Z)}{f^2_{Z|\frac{\beta}{\alpha}X}(Z)} V_\Phi$. □