Myth of the Contamination Effect?

Small parties' survival under the mixed-member majoritarian electoral system in Japan

UMEDA Michio

Department of Political Science
University of Michigan, Ann Arbor

INTRODUCTION & BACKGROUND

The purpose of the article is to study and provide an accurate estimate of the *contamination effect* – additional proportional representation (PR) vote that small parties attain by nominating candidates in single member districts (SMD) tier under the mixed-member electoral system, based on the case of the Japanese House of the Representatives.

Many democratic countries have adopted the mixed-member electoral system since 1990s, although historically the system has been relatively rare. The system has two electoral tiers, SMD and PR. Reformers have argued that the mixed electoral system is able to represent the best of both worlds - allowing party-centered campaigning and candidate responsiveness. However the functioning of the system is still not well understood, because only few countries have a long history under the system. In addition, electoral outcomes after the reform are inevitably influenced for at least a few elections by the political and party systems established under the previous system, although this point has largely been ignored.

Previous studies have argued that, contrary to Duverger's law, party competition does not converge to the two party contests in the SMD tier under the mixed electoral system, because small parties have different strategic calculation from the pure SMD system due to the *contamination effect* in the PR tier. In the mixed electoral system, political parties have incentives to run candidates in as many districts as possible, even though these candidates have little chance to get elected in these districts, because the parties might be able to mobilize the extra PR votes by nominating these candidates. If parties keep nominating candidates in the SMD contests and spend considerable effort in the campaign despite small chances of the victory in the districts, then the constituents are also less likely to abandon these candidates only out of strategic voting consideration. The effect of SMD candidate nomination to the PR vote share is called as the *contamination effect*. If the contamination effect were substantial, small parties would nominate their candidates in many SMD districts for their survival, rather than establish electoral agreements with major parties and exit from most of the SMD contests. Consequently, as another symptom of the contamination effect, the number of candidates competing in each district would not be two, contrary to Duverger's law, which predicts two-party competition in the single member plurality districts. (Cox & Schoppa 2002[1])

Before introducing the scholarly discussion of the contamination effect in the Japanese context, I briefly describe the Japanese mixed-member electoral system. Japan adopted the mixed-member *majoritarian* electoral system (MMM) to the House of the Representatives (HR) in 1994. The MMM is one of the two major sub-types of the mixed electoral system that employs two electoral tiers in parallel. Each Japanese voter has two votes. She casts one to a SMD candidate of the district and the other to the regional PR party list. Under the Japanese MMM system, votes to the SMD tier decide which candidates get elected from the districts but do not influence the party's seat allocation in the PR tier, and votes to the PR list decide the number of seats allocated to each party in the regional list. This mechanism makes the strong contrast with the mixed electoral proportional system (MMP) at which PR votes primary decide the party's total seat share. The Japanese constituents experienced four elections under the current system (1996, 2000, 2003 and 2005). In addition, a election is forthcoming in September 2009.

Japan has been discussed as a strong evidence of the contamination effect. First, many empirical studies find considerable contamination effects in the Japanese HR contests. (Mizusaki & Mori 1998, Herron & Nishikawa 2001, Cox & Schoppa 2002, Reed 2003, Ferrara 2005) Moreover, the average number of the effective number of electoral parties (ENEP, Taagepera & Shugart 1989) in the SMD contests was much larger than those of the single member plurality system, which is discussed as a consequence of the contamination effect (Herron & Nishilawa 2001, Cox & Schoppa 2002). For example, Cox and Schoppa indicate that the ENEP in the Japanese HR was on average 2.83 in the 1996 and 2000 elections. The Japanese ENEP is indeed considerably larger than other

1 Cox & Schoppa used the term *interaction effects* instead of the *contamination effect*, by focusing more on the influence of PR tiers to the effective number of parties in the SMD contests rather than the amount of PR vote-share that small parties gain by nominating in the SMD tiers.

single member plurality countries, such as India (2.49 in 1957-91), New Zealand (2.56 in 1972-93), Canada (2.40 in 1935-1994), the UK (2.27 in 1922-1997) and the US House (1.81 in 1900-1990). Based on these numbers, Cox and Schoppa conclude that "We believe it to be more likely that under mixed rules, Duverger's incentives do not operate strong enough to move ENEP numbers below the mid-2.0 range seen in the SMD tiers of other mixed system (Cox & Schoppa 2002, pp1044)". In the following of this article, I would like to term the group of researchers who advocates the contamination effect and consequent semi-multiparty SMD competition in Japan as *Contamination theorists* for brevity.

However, recent development of the Japanese politics contradicts the previous findings. The Japanese party system appears to be becoming a two party system, the Liberal Democratic Party (LDP) and the Democratic Party of Japan (DPJ) being the two main forces. Other smaller parties have significantly reduced the number of candidates they nominate in SMD tier in the 2005 HR election. Moreover, in the next election small parties except the Japanese Communist Party (JCP) are expect to nominate their SMD candidates only under the electoral accord with either the LDP or the DPJ in most of the districts. In other words, small parties will nominate only in the districts where a major party with the accord does not run (Asahi Newspaper Jan 2-4, 2009). Furthermore, JCP, which has been nominating candidates in most of the districts until 2005, announced that the party significantly reduced the number of districts to nominate.

As a result, the number of candidates competes in each district has considerably decreased. In the 2003 and 2005 HR election, the ENEP went down to 2.40[2]. Moreover, it is almost certain that the ENEP in the next election will be even smaller than that of the previous elections because of the electoral accords between parties and the JCP withdrawal from many SMD contests. Consequently, the ENEP in Japanese SMD is and will be considerably "below the mid-2.0 range",

[2] Even this number could overestimate the actual number of "party candidates" that compete in the districts. The number does not adjust traditional feuds among the LDP; the LDP nominee and conservative independents who will be nominated ex post once he wins the contest. Moreover, due to the LDP split over the postal service reform in 2005, ex-LDP incumbents and LDP *assassin* contested in some districts, which resulted in temporal increase of the number of competing parties.

contrary to the prediction of Cox and Schoppa in 2002. Because the ENEP in SMD in comparison to the pure SMD countries turns out to be much less salient -- assumed as the symptom of the contamination effect -- scholars started to review the evidences of the contamination effect presented in the previous literatures.

Maeda (2008) challenges the estimates of the Contamination theorists. He criticizes that Contamination theorists systematically overestimate the contamination effect, because those studies fail to control the endogeneity of the nomination to the party support level – parties run candidate where the parties are strong. Based on the estimates from the Heckman's treatment-effects model, Maeda concludes that the contamination effect is almost negligible and much smaller than previous studies expect.

I share many of Maeda's criticism toward the Contamination theorists. As he argues, some of the previous studies do not seem to pay enough attention to the endogeneity of candidates to the party support level in the districts, and as a result their estimates are biased due to the endogeneity. Their conclusions could have been justifiable as a provisional answer while available electoral data is scarce, especially the party's level of support within each district due to limited number of elections under the new electoral system as well party realignment that entailed with the electoral reform. However, I believe that these tentative conclusions should be revised once the party system has stabilized and more electoral data became available.

Nevertheless, I find there are few but serious methodological problems in Maeda's approach. First, the Heckman's treatment-effects model should be used when the treatment assignment – the candidate nomination -- is endogenous to the size of the *treatment effect*. Nonetheless, Maeda appears to use the method to solve the endogeneity of the treatment assignment to the level of *pretreatment* - party support before the SMD nomination, suggested from his statement that "parties tend to nominate SMD candidates in areas where they are strong, which makes the impact of running candidates look larger than it actually is (Maeda 2008, pp723)".

Moreover, there is confusion for the estimand between the Contamination theorists and

Maeda, even if Maeda employs the method to solve the heterogeneity of the treatment effect. The Heckman's treatment-effects model is suppose to estimate the ATE— *the average treatment effect*. In other words, Maeda's estimand is the size of the contamination effect if political parties randomly choose the districts that they nominate candidates[3]. Nevertheless, though not so clear, the Contamination theorists do not necessarily discuss the size of the contamination effect under the random assignment. Instead, those literatures appear to discuss that how much the political parties have attained in the districts that they actually nominated the candidate i.e. the ATT, *the average treatment effect for the treated*. Therefore, the relevance of Maeda's conclusion is dubious even if the treatment-effects model is used as designed to solve the heterogeneity of the treatment effect.

In the next section, I explain the different theoretical background behind each estimand and the treatment-effects model of Heckman in more detail. In addition, I also briefly explain the matching approach I use in this article for the parameter estimation.

METHOD

In this section, I discuss the theoretical background of the discussion over the contamination effect based on the framework of Rubin Causal Model (RCM). In particular, I explain the confusion over the estimands between the authors of previous studies.

In RCM framework, the causal effect is defined as the difference between an observed outcome and its counterfactual. Suppose that the treatment is binomial (0-1) for simplicity. The causal effect for the treated subjects is the difference between the observed outcomes for these subjects given their treated condition on the one hand, and the counterfactual (i.e. unobserved) outcome if they were not treated on the other hand. In the following, I use the term of the ATT, *the*

[3] Another possible interpretation of the ATE is the average treatment effect *if all the subjects in the samples are treated* – the average gain of PR vote-share if the party nominates candidates in all the districts. However, it is hard to justify the ATE in this definition because SUTVA does not hold in this case. I explain SUTVA in the next section.

average treatment effect for the treated for the estimand based on the RCM convention. The causal effect for the not-treated subjects is the difference between their counterfactual outcomes if they would have treated on the one hand, and their observed outcome given their controlled condition. I use the term the ATC, *the treatment effect for the treated* for this estimand also based on the RCM. The ATE, *the average treatment effect* is the average of these two estimands weighted according to the frequencies of the treated and controlled subjects in the population.

Suppose that *Y1/Y0* indicates the (counterfactual) outcome when the subject is treated/ controlled, *D* indicates the bivariate treatment status of the subjects, *X* is observed covariate, and *U* is unobserved covariates.

$ATT = E\ (Y1 - Y0 \mid X, D{=}1)$ *: counterfactual difference of the outcome given D = 1*

$= E\ (Y1 \mid X, D{=}1) - E\ (Y0 \mid X, D{=}1)$

$= [\ E\ (Y1 \mid X, D{=}1) - E\ (Y0 \mid X, D{=}0)\ ] - [E\ (Y0 \mid X, D{=}1) - E\ (Y0 \mid X, D{=}0)\ ]$

$= [\ E\ (Y1 \mid X, D{=}1) - E\ (Y0 \mid X, D{=}0)\ ] - [E\ (U0 \mid X, D{=}1) - E\ (U0 \mid X, D{=}0)\ ]$

This is observable unobservable / bias

$ATC = E\ (Y1 - Y0 \mid X, D{=}0)\ )$ *: counterfactual difference of the outcome given D = 0*

$= E\ (Y1 \mid X, D{=}0) - E\ (Y0 \mid X, D{=}0)$

$= [\ E\ (Y1 \mid X, D{=}1) - E\ (Y0 \mid X, D{=}0)\ ] - [E\ (Y1 \mid X, D{=}1) - E\ (Y1 \mid X, D{=}0)\ ]$

$= [\ E\ (Y1 \mid X, D{=}1) - E\ (Y0 \mid X, D{=}0)\ ] - [E\ (U1 \mid X, D{=}1) - E\ (U1 \mid X, D{=}0)\ ]$

This is observable unobservable / bias

ATE is the weighted average of the ATT and ATC, thus if the fraction of the subjects that gets the treatment is τ, then

$ATE = E\ (Y1 - Y0 \mid X\ )$

$$= E\,(Y1 \mid X\,) - E\,(Y0 \mid X\,)$$

$$= \tau\ ATT + (1-\tau)\ ATC$$

$$= [\ E\,(U1 \mid X, D{=}1) - E\,(U0 \mid X, D{=}\ 0)\ ]$$

$$-\quad \{\ \tau\ [E\,(U0 \mid X, D{=}\ 1) - E\,(U0 \mid X, D{=}\ 0)\ ]$$

$$+\ (1-\tau)\ [E\,(U1 \mid X, D{=}\ 1) - E\,(U1 \mid X, D{=}\ 0)]\ \}$$

Matching method estimates each treatment effect by comparing subjects holding the covariates as similar as possible but different treatment status, though there are various approaches to find "similar" subjects. For example, in order to estimate the ATT -- $E\ (Y1 \mid X, D{=}1) - E\ (Y0 \mid X, D{=}1)$, the matching method tries to find an untreated subject with very similar covariate distribution with the treated subject. The assumption is that the untreated subject that shares the observed covariates with the treated subject have the same outcome with the counterfactual outcome that if the treated subject would not be treated. In the formula, this condition is expressed as $E\ (Y0 \mid X{=}x, D{=}0) = E\ (Y1 \mid X{=}x, D{=}0)$. Similarly, for the estimation of ATC, the counterfactual outcome of the untreated subjects should be the same with the treated subjects if the former would be treated, given the observed covariates X, i.e. $E\ (Y1 \mid X{=}x, D{=}1) = E\ (Y1 \mid X{=}x, D{=}0)$. In other words, the assumption requires that the outcomes are the same between the realized/observed and the counterfactual/unobserved given the treatment status and the observed covariates. In the RCM, the condition is termed as *strong ignorability* and expressed more formally as:

$$(Y1, Y0)\ \perp\ D \mid X$$

This means that the distribution of *Y1* and *Y0* are the same regardless to their treatment status, in other words whether they are realized or counterfactual, given the observed covariates X.

The Heckman's treatment-effects model does not assume the strong ignorability. The model assumes that given the observed covariates X, the outcome for actually treated subjects is not necessarily the same with the outcome if the not-treated subjects would receive the treatment. The treated subjects are self-selected into the treatment group because they knew that the treatment is more beneficial than those self-selected into the controlled group. The method is expressed in the formula as follows. First, the outcome Y given the treatment status is the function of observed covariates X and unobserved covariates U. Therefore, the model for the outcome is;

$Y1 = \mu 1\ (x) + U1$: *the outcome Y given observed X, unobserved U1 and D =1*

$Y0 = \mu 0\ (x) + U0$: *the outcome Y given observed X, unobserved U0 and D = 0*

The potential outcome Y is therefore;

$$
\begin{aligned}
Y &= D\ Y1 + (1 - D)\ Y0 \\
&= D\ (\mu 1(x) + U1) + (1 - D)\ (\mu 0(x) + U0) \\
&= \mu 0(x) + [\mu 1(x) - \mu 0(x) + U1 - U0]\ D + U0
\end{aligned}
$$

Consequently, we observe the ($\mu 1(x) – \mu 0(x) + U1 – U0$) difference between the condition of the treated and untreated. The primary assumption of the Heckman's treatment-effects model is that ($U1–U0$) is from the heterogeneity of the treatment effect between the subjects. Hence, the subjects who choose to receive the treatment knew that the treatment effect is larger than those who do not. However, it is not possible to observe ($U1–U0$) directly. Therefore, Heckman uses the treatment model to estimate $\mu d\ (z)$, which is supposed to be correlated with ($U1–U0$). The treatment model is expressed as follows. The treatment D is a function of the observed covariates for the treatment Z and unobserved covariate $\mu d\ (z)$.

$D^* = \mu d\ (z) + U_D$: *treatment function given observed Z and unobserved* U_D

$D = I\ (D^* > 0) = I\ (U_D > - \mu d\ (z))$: *I is an indicator function*

Given the model, it is possible to assume that a subject, who self-selected into the treatment group but the expected probability of the treatment given Z is small, did so because the treatment effect for the subject is relatively large. On the other hand, if a subject self-selected into the controlled groups even if Z predicts the subject is very likely to receive the treatment, it can be assumed that the person knew that for him the treatment does not have effect. For the topic of this article, ($\mu 1(x) - \mu 0(x)$) is described as the ATE and ($U1 - U0$) is the difference of the treatment effect between the individual and the population average. The purpose of the treatment-effect model is to estimate the ATE by correcting the bias with utilizing the correlation between ($U1 - U0$) and $\mu d\ (z)$. See Heckman (xxx) more detail for the method.

It is worth repeating that ($U1 - U0$) is the difference due to the heterogeneous treatment effect, not the difference due to the imbalance of unobserved covariates between the treated and untreated individuals. If the difference is attributable to the latter parameter, the treatments-effects model does not solve the problem because there is no theoretical basis that ($U1 - U0$) and $\mu d\ (z)$ are correlated. Maeda appears to wrongly use the treatment-effects model to solve the latter imbalance given his description of the method.

However, even if he uses the model to solve the heterogeneity of the treatment effect, there still remain some other concerns, as I briefly explained in the previous section. The followings are similar but theoretically different concepts of the treatment effect, which has been confused in the previous studies.

a) Average additional PR vote-share that the small party has gained in the districts where the party nominated candidates.
b) Expected additional PR vote-share that the small party would have gained if the party nominates candidates in randomly chosen districts.
c) Expected additional PR vote-share that the small party would have gained if the party nominates candidates in all the districts.

The ATT is a), which I use as the primal estimand of the article. The scope of the inference of the estimand is limited to the realized number and combination of the treatment to the subjects. In other words, the ATT estimate infers only to the subjects that actually received the treatment, as described in the definition a). The ATT does not infer what the treatment effect would be if the number of the subjects to receive the treatment increases/decreases, nor what the effect could be if the different subset of the subjects receives the treatment. The estimate is more robust because of the limited scope of the inference.

In contrast, both b) and c) can be the ATE, because theoretically both estimates should be the same if the SUTVA assumption holds, which discussed in the following. However, I do not think that the ATE in the definition b) is the estimand that the Contamination theorists discuss about. Given the fact that the party has discretion to choose the districts to nominate, it does not have much theoretical importance to know what would have happened if the party randomly chooses the districts to nominate. On the other hand, the ATE in the definition c) cannot be estimated with any of the statistical approaches though the estimand itself has some theoretical importance, because the estimand is not able to be estimated without the SUTVA.

The SUTVA, *the stable unit treatment value assumption*, indicates the absence of interaction among units/subjects (SUTVA, Rubin 1980, 1986). Let Yij be the outcome of the ith subject ($i = 1, 2, \dots, n$) if exposed to the treatment j ($j = 0, 1$). SUTVA is the assumption that if subject i is exposed to treatment j, the observed value of Y will be Yij, regardless the treatment for the subjects other than i. In other words, it requires that the outcome for each subject should be the same regardless the treatment status of the other subjects.

However, the assumption does not hold in the context of the contamination effect. For example, the extra PR vote-share in a district i by the contamination effect would be different if the number of candidates the party nominates in the other districts increased from 25 to 250. Therefore, it is not reasonable to use the estimated ATE in the definition c). On the other hand, the SUTVA might be hold if the party nominates the same number of candidates to different subsets of the districts. For example, it is not so unreasonable to assume that the electoral outcome of the district *j* is the same regardless to which 25 districts the party nominates except for the district *j*. If the assumption holds, it is still possible to infer the ATE in the definition b), though ironically the ATE in the definition b) might not be as valuable as the one in the definition c).

Consequently, in this article I choose the ATT as the primal estimand because the ATT is most theoretically relevant for the discussion of the contamination effect, in addition the estimand is able to be estimated without relying on the SUTVA. For estimation of the ATT, I use the genetic matching approach in R (See Sekhon Forthcoming). The algorithm uses generalized Mahalanobis metric which include an additional weight matrix W to find the best match between the treated and not treated observations. The formula is;

$$d(Xi, Xj) = \{(Xi, Xj)' (S^{-1/2})' W S^{-1/2} (Xi, Xj)')\}^{-1/2} ,$$

such that X is observed covariates, S is the sample covariate matrix of X, and W is a *k by k* positive definite weight matrix and $S^{-1/2}$ is the Choleskey decomposition of S which is the variance-covariance matrix of X. To estimate the ATT by matching without replacement, the method matches each treated subject with M closest controlled subjects as defined by this distant measure[4]. The difference between the generic matching and traditional Mahalanobis distance approach is the weight matrix *W*. Genetic matching approach sets W to improve the balance in the covariates between the treated and untreated subjects (See Mebane and Sekhon 1998 for more detail).

In conclusion, in this section I briefly described the theoretical basis of different estimands, and explained why the estimate from the treatment-effects model is not appropriate in respect to the estimand of the interest as well to the violation of SUTVA.

[4] M is set equal to be one in this article.

EMPIRICAL DATA & ANALYSIS

This section re-examines the size of the treatment effect in the Japanese mixed electoral system with the empirical data. I use the electoral outcome of the 2003 and 2005 general elections of the Japanese HR. In this article I focus on Social Democratic Party of Japan (SDPJ) among four parties that Maeda analyzes for the contamination effect – LDP, DPJ, JCP and SDPJ.

There are two reasons for the decision. One is theoretical. LDP and DPJ held 25-50% of the legislative seat-share before the elections, and consequently they did not fit well to the original definition of the contamination effect -- additional PR share that *small parties* gain by nominating candidates in SMD. The other is technical. Small fraction of districts these parties did not nominate disturb to estimate the ATT within a reasonable confidence interval. Those three parties nominated more than 85% of the districts in 2003 and 2005. Given the small fraction of the districts without candidates, it is hard to find appropriate match from a small pool of the districts that do not have candidates but share quite similar attributes with many districts having candidates. Although still theoretically possible to match an "untreated" district to more than single "treated" districts, the option increases error in the estimate because the estimate relies on small number of observations for the estimation. Therefore, it is technically difficult to estimate the ATT for those parties.

In contrast, SDPJ provides a good case to estimate the size of the treatment effect. First, the party is not large during the period. In 1996, Japanese Socialist Party[5], which had been the largest opposition party during the post-war period, renamed to SDPJ to refresh the party's image but lost most of the members to the DPJ within the year, the newly formed the center-left party. Although some of the leftist legislators remained to SDPJ, the party lost most of its traditional support base of JSP such as labor union, and consequently the party could nominate in no more than 20% of districts under the new electoral system. Due to the small fraction of the districts with SPDJ candidates, it is easy to find appropriate match – the districts without candidates but shares the

[5] Japanese Socialist Party changed its Japanese title to SDPJ in 1996, though the title was often translated as SDPJ in English even before 1996 while the moderate-wing is strong within the party.

other attributes of the district with candidates, because now we can choose the former from the large pool of alternatives.

As covariates for the matching, I use the PR vote-share of the previous election and the urbanization level of the districts measured by the DID index[6], in addition to the existence of the SDPJ candidate in the previous election which I explain later in more detail. Previous vote-share and the DID index are also two covariates employed by Maeda for the electoral outcome model. Hence, in essence I use the same covariates with him to make comparison of the models clearer.

In order to use the PR vote-share of the previous election as a covariate, each district should be comparable between two consecutive elections. Owing to the redistricting in 2002, 59 out of 300 districts are not available for 2003 because these districts contain different geographic area between elections. In addition, three districts are not available for the analysis in 2005 because of the reorganization of the municipalities in Ibaraki prefecture[7].

In estimating the contamination effect, it is worth paying attention to the fact that there are two types of districts with the candidates[8]. One is *keep* -- the districts where the party nominated candidates in two consecutive elections, and the other is *participate* -- the districts where the party did not nominate in the last election but decided to participate in this election. Similarly, there are also two types of the districts without candidates, which had candidates only in the previous election (*exit*) and did not have in both of the elections (*keep out*). For the estimation of the contamination effect, *keep* district is comparable only with *exit*, as well *participate* is comparable with

---

6 The urbanization level of the district is measured by Densely Inhabited District (DID) index, the concept of the Japanese census which indicates the fraction of the population living in the densely populated area.

7 In Japan, the electoral results are released by municipality as a unit. Before the 2005 election, some municipalities in Ibaraki prefecture were merged to the municipalities that belong to the other districts, and consequently these municipalities contain areas that consist of different districts. After the election, the electoral committees of these municipalities released the electoral outcome of the PR based on the new municipal boundaries; therefore it was impossible to compare the electoral results of PR for these districts across elections in 2003 and 2005. Therefore, those three districts were excluded from the analysis.

8 Reed mentions this point and estimates the contamination effect separately between *keep* and *exit* on the one hand, and *participate* and *keep out* on the other hand (Reed 2003).

keep out, because the previous vote-share of the party was also influenced by the existence of the party candidate in the district. For example, the vote-share of the party in the *keep* districts should not change between two elections owing to the contamination effect, because the vote-share of both elections should be "contaminated" by the party candidates. Therefore, those *keep* districts can be compared with *exit* districts, which moved from the *contaminated* condition to the *not-contaminated*. Likewise, *exit* districts should be compared with *keep out*, because both districts were *not-contaminated* in the previous election.

Among the 241 districts eligible for the analysis of the 2003 election, SDPJ nominated candidates in 51 districts in 2000. In the 2003 election, the party kept nominating 28 districts out of 51, in addition newly participated in 20 districts. Therefore, the party nominated candidates in total at 48 districts out of 241 in 2003[9]. Therefore, in the 2003 election SPDJ had 28 *keep*, 23 *exit*, 20 *participate* and 170 *keep out* districts. In the 2005 elections, 297 districts are available for the analysis. From 61 districts with the SDPJ candidates in 2003 (that include 13 districts not eligible to the analysis of the 2003 election), the party choose 26 districts to keep nominating candidates in 2005, in addition participated in the SMD contests in 12 districts. Therefore, SPDJ had candidates at 38 districts in the 2005 campaign. Consequently, SPDJ had 26 *keep*, 35 *exit*, 12 *participate* and 236 *keep out* districts in 2005.

I take two strategies to cope with the heterogeneity of the districts. First, I add the existence of the party candidates in the previous election as a covariate for the matching. The matching method compares the districts among those with candidates in the previous election, or among those without candidates. The option provides the ATT for all the districts that SDPJ nominated for each of two elections, and thus enables to compare the estimates with the previous literature.

Next, I divide the districts into two groups -- those with the candidate in the previous election and those without -- and conduct the analysis separately. This option enables to estimate

9 SDPJ nominated candidates in 62 districts out of 300 total districts (Maeda 2008).

possibly different size of the contamination effects between the districts where the party kept the candidate and where the party newly participated. Moreover, the option enables to estimate not only ATT but other estimand for the districts with candidates in the previous elections, because it is not unreasonable to assume the SUTVA for those districts. SPDJ had candidates in the previous election within relatively small fraction of the districts (51 out of 241 in 2000, 61 out of 297). Therefore, it is not so absurd to expect that the size of the contamination effect does not change even if the party changes the number and/or combination of the districts it nominates among the small subsets of the districts. By exploiting the situation, I estimate not only the ATT but also the ATC, the average treatment effect for the treated and the controlled for the group. The ATT in this context indicates how much the party would have lost if it had stopped nomination in those districts that it actually kept the candidates, and the ATC designates how much SDPJ has lost by quit nominating in the districts the party actually stopped.

RESULT & DISCUSSION

Table 1 and 2 indicates the result of the analysis for the 2003 and 2005 elections. The two tables also include the estimates from the conventional OLS model and the treatment-effects model[10] for the comparison. In Table 1, the result indicates that SDPJ attained 2.1% additional PR vote-share in the districts the party nominated candidates. The size of the effect is statistically significant, and even larger than the estimate with the OLS. The contamination effect is similar between the districts where the party had and did not have candidates in 2000. SDPJ gained 2.0% in the former and 2.3% in the latter districts. On the other hand, the ATC among the districts with candidates in 2000 was 1.0%, though the effect is still statistically significant. Hence, SPDJ lost around 1.0% of the vote-share

---

[10] The OLS model was newly analyzed by the author with the same data set I used for the matching. The result of Heckman's treatment-effects model was retrieved from the table in Maeda's article, because some of the variables he used are not currently available for me. As a result, there are small differences in the number of districts employed for the analysis, probably due to our different judgments over which districts are comparable between elections after the redistricting.

by stop nominating candidates in those districts. The number is smaller than the ATT of the groups, though the confidence intervals of the two estimates are overlapped. The party appears to choose the districts to quit that harm less for their PR vote-share, though the evidence is not strong enough.

The analysis of the 2005 election in Table 2 indicates similar results. SDPJ gained 2.7% more PR vote-share in the districts where it nominated the candidates, and the effect size is statistically significant. Among the districts with the candidates in 2003, the party gained 3.7% additional vote-share by keeping candidates, but would have gained only 1.7% vote-share in the districts that the party actually stopped nominating. Finally, SDPJ gained less 1% of the additional share by participating in some districts, and the effect size was not statistically significant. The final result was not consistent with the estimate of the 2003 election.

In general, the results of the analyses indicate that SDPJ surely attained extra PR vote-share in the districts where the party nominated candidates, except of the districts it newly participated in 2005. The size of the contamination effect is even larger than those estimated with the conventional OLS, also the districts the party participated in 2005 as the exception. Therefore, the case of SDPJ in 2003 and 2005 supports the claim of the contamination effect that the small party can mobilize additional PR vote-share by nominating candidates in SMD, contrary to Maeda's criticism regarding to the existence of the contamination effect.

Nevertheless, the implication of analysis in this article could be more supportive to Maeda than to the Contamination theorists. The size of the contamination effects was no more than 3% in the districts the party actually nominated the candidates. Even if we believe the strong assumption that SDPJ would have gained the same additional PR vote-share by nominating candidates in all the other districts – the ATT is the same with the ATE in the population -- the party at best would have gained in total few additional seats from some of the regional PR lists with relatively large district magnitude, such as Kinki with 29 or South Kanto with 22. Moreover, the considerable difference observed between the ATT and the ATC suggests that SDPJ strategically chose the districts that provide the best electoral outcome with minimum electoral resources, and hence the marginal return of nominating additional candidates is not to be as large as the return in the districts it had nominated. Consequently, I conclude that the contamination effect is not large enough to keep small parties in the SMD contests, concurring with Maeda.

In the following, I would like to discuss more general implication of the article for the consequence of electoral reform in Japan, in particular on the electoral competition and party system. It seems that there are some deficiencies in the Contamination theorists, given the fact that the political parties in Japan are now converging toward two-party system, in addition most small parties stop nominating candidates in the SMDs without electoral accords with major parties.

The following constitute possible sources of their failure: First, the group does not seriously try to measure how much small parties are able to gain additional number of PR *seats* based on their estimates of the contamination effect though critical for their survival[11]. Their very optimistic estimates would suggest that the size of the contamination effect is at maximum few additional legislative seats even if the party nominate candidate in all the districts. Consequently, it should have been clear before the result of this article that the effect is not large enough to drive small parties to campaigning alone in the SMDs under the Duvergerian pressure. Moreover, it is not difficult to expect that small parties had difficulty to nominate considerably more candidates than they actually did even though the party leaders knew the benefit of nominating more candidates, given the limited electoral resources of small parties like SDPJ such as campaign finance as well quality and quantity of the candidates. Particularly, the Japanese electoral law imposes financial penalty for minor candidates who fails to achieve 10% of the valid ballots of the district[12]. The clause

---

11 As far as I know, Reed (2003) is only the exception. He mentions that the size of the contamination effect is at most to provide a couple of additional seats for small parties, and not substantial for them under adversarial atmosphere under the mixed member majoritarian system.

12 Candidates of the SMD who fail to achieve 10% of the valid ballot are confiscated 3,000,000 yen (33,000US$) required to deposit when they register to run (Public Office Election Law, Article 92 and 93). Moreover, those candidates lose the eligibility for public funding to their electoral campaign, such as cars (ibid. Article 141 Clause 7), campaign flyers (ibid. Article 142 Clause 10), campaign posters (ibid. Article 143 Clause 14), and sign of the campaign meeting (ibid. Article 164-2 Clause 6).

places extra financial burden for small party if the party tries to nominate candidate where the party support is below the level.

Besides, the group seems to share the assumption which is hard to justify in reality; the political and party system established under the previous electoral system is washed away in the first election under the new electoral system, because all the actors switch to the optimal strategy under the new system without friction. Thus, the electoral outcome of the first election after the reform is perfectly ascribable to the function of the new system. In reality, every political actor, party elites, (potential) candidates, as well ordinal voters know only vaguely about the function of the new electoral system and their optimal strategy given the system. Moreover, they have to bear the cost of adjustment if they shift strategies, especially to do quickly. Therefore, they have to learn the new system and adjust the strategy gradually with trials and errors. It seems that the previous studies rushed to the conclusion based on seemingly inflated ENEP under the mixed electoral system, although it was the remnant of the old system and passes away after few elections, which provide actors time to learn and adjust to the new system. In conclusion, these two overvaluations – the influence of the contamination effect to the party system and the campaign strategy (needless to say the overestimation of the size itself) as well the speed of the new electoral system to clean away the legacy of the previous system – resulted in their failed prediction for the fate of Japanese party politics.

I have few points in this article to be improved in the future. First, there is still a possible source of the bias due to endogeneity of the treatment with *the shift of party support between the elections*, which should be improved with additional covariates for the matching. I used the SDPJ's PR vote-share in the previous election as covariates of the analysis. However, the vote-share at the previous election conducted few years ago is not a perfect indicator of the pretreatment support level before the election. If the party nominates candidates where the party's support is relatively *growing* and stops nominating where the support is *declining*, the estimates of the ATT could be overestimated due to differences in the shift of party support between the *treated* and the *controlled* districts. This potential bias in this article is reduced if the shift of the SDPJ's vote-share is considerably correlated with the urbanization level of districts, because the model included the DID index as a covariate. Nevertheless, the model would be improved with additional covariates that reduce the bias due to the support level shift between the elections.

Next, there are few other candidates of the covariates to improve the model, in addition to the covariates which explains the shift of SDPJ's electoral support between the districts. In this article, I use only the previous vote-share, the DID index and the nomination in the previous election as the covariates for the purpose of the comparison with the Maeda's model. Although it is not easy to employ many covariates for the matching approach given the limited number of available observations, I plan to add few more covariates such as the turnout, closeness between two top contenders, and DPJ and JCP PR vote-share of the districts in the previous election as the extension of the model.

Furthermore, the alternation of the JCP's campaign strategy provides another interesting case to test the contamination effect. As described in the previous section, JCP announced that the party reduces the districts to nominate candidates to less than 100 in the next election, nevertheless the party has been nominating candidates in most of the SMDs without any electoral coordination with the other parties. The campaign strategy provides us around 200 districts where another small party *exits* in the next election. The next Japanese HR election takes place no later than September 2009. The expected electoral coordination between small parties (including SDPJ) and major parties brings difficulty for estimating the contamination effect, because now districts with and without small party candidates could differ not only in the existence of small party candidates, but also in the existence of major party candidates as well campaign coordination of major and small parties. Fortunately to the purpose of the study, JCP continues to be outside of the electoral coordination in the next election. I look forward to test the result of this article attained from the SDPJ case to the electoral data of JCP.

Finally, the source of the contamination could need further examination. Cox and Schoppa ascribe the origin of the contamination effect to the priming of expressive motive of the constituents, and consequently the effect the contamination benefits only to small parties. I accepted the definition in this article from more practical than theoretical reason -- available approaches are not useful to estimate the contamination effect for major parties even if it existed. However, it is possible that major parties are also beneficiary of the contamination effect because they nominated candidates in more districts than small parties did, and probably more experienced and charismatic candidates because of their organizational resources. The decline of small parties in Japan might be also ascribable to the contamination effect of major party candidates in SMD, which attached more attractive and tactful human face to the party ideology than those of small parties. I would like to study this point further in the future study with alternative approach.

Table 1. SDPJ 2003

GenMatch	treatment effect	std.err.	# of obs.	# of treated
full (ATT)	0.0209	0.0042***	241	48
w/ cand 2000 (ATT)	0.0204	0.0070***	51	28
w/ cand 2000 (ATC)	0.0101	0.0046*	51	23
w/o cand 2000 (ATT)	0.0225	0.0041***	190	20
Treatment-Effects Model	0.0056	0.0065	243	
OLS	0.0154	0.0021***	241	

* p< .05, ** p < .01, *** p < .001

Table 2. SDPJ 2005

GenMatch	treatment effect	std.err.	# of obs.	# of treated
full (ATT)	0.0266	0.0068***	297	38
w/ cand 2003 (ATT)	0.0367	0.0099***	61	26
w/ cand 2003 (ATC)	0.0174	0.0087*	61	35
w/o cand 2003 (ATT)	0.0085	0.0076	236	12
Treatment-Effects Model	-0.0141	0.0084	300	
OLS	0.0157	0.0024***	297	

* p< .05, ** p < .01, *** p < .001

Reference

Cox, K.E., Schoppa, L.J., 2002. Interaction effects in mixed-member electoral systems: theory and evidence from Germany, Japan, and Italy. Comparative Political Studies 35 (November), 1027–1053.

Ferrara, F., Herron, E., 2005. Going it alone? Strategic entry under mixed electoral rules. American Journal of Political Science 49 (1), 16–31.

Ferrara, F., Herron, E.S., Nishikawa, M., 2005. Mixed Electoral Systems: Contamination and Its Consequences. Palgrave Macmillan, New York.

Herron, E.S., Nishikawa, M., 2001. Contamination effects and the number of parties in mixed-superposition electoral systems. Electoral Studies 20, 63–86.

Maeda, Ko., 2008. Re-examining the contamination effect of Japan's mixed electoral system using the treatment-effects model, Electoral Studies 27, 723–731

Mizusaki, T., Mori, H., 1998. Tokuhyou deeta kara mita heiritsusei no mekanizumu. Senkyo Kenkyu 13, 50–59.

Reed, S.R., 2003. Heiritsusei ni okeru shousenkyoku kouhosha no hireidaihyou tokuhyouritsu eno eikyou. Senkyo Kenkyu 18, 5–11.

Sekhon, J., (forthcoming). Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching package for R. Journal of Statistical Software.

Taagepera,R., Shugart, MS., 1989. Seats and Votes: The Effects and Determinants of Electoral Systems. Yale University Press