

Measuring Judicial Independence Reconsidered: Survival Analysis, Matching, and Average Treatment Effects

KENTARO FUKUMOTO

Gakushuin University, Japan
Washington University in St. Louis, USA
Kentaro.Fukumoto@gakushuin.ac.jp

MIKITAKA MASUYAMA

National Graduate Institute for Policy Studies, Japan
[mmasuyama@grips.ac.jp](mailto:mamasuyama@grips.ac.jp)

Abstract

This article reconsiders how to judge judicial independence by using the Japanese judicature, one of the allegedly-most dependent judiciary branches. In their influential work, Ramseyer and Rasmusen (2003) argue that judges who once belonged to a leftist group take longer to reach a ‘moderately prestigious status’ under the long-term conservative rule of Japan. Their method does not, however, deal appropriately with the possibility of judges not reaching this position because the judge dies, retires early, or is still at the early stage of her career. Ramseyer and Rasmusen also mistakenly assume that all judges will eventually obtain this position. This article develops a survival analysis model of judicial careers and attempts to solve the problems of censoring, left truncation, and split population. We also offer a way to utilize a matching procedure to estimate average treatment effects on censored time-to-event as well as event occurrence. We re-analyze a corrected version of Ramseyer and Rasmusen’s data using their and our methods. One of the most important findings is that, contrary to what Ramseyer and Rasmusen argue, leftist judges are not discriminated against in terms of the timing of promotion.

An earlier version was presented at the 2006 Annual Meeting of the American Political Science Association, Philadelphia, 31 August–3 September 2006. We appreciate the Japan Society for the Promotion of Science for financial support and Mark J. Ramseyer and Eric B. Rasmusen for making their data available. Our thanks also go to Amy L. Catalinac, Jonathan Marshall, Hatsuru Morita, and Kirk A. Randazzo for their comments. The usual disclaimer is applied.

Introduction

Despite judicial independence being one of the principles embraced in most democratic constitutions, recent studies of judicial politics have made it clear that it is inadequately established in most of these countries. Although the concept has several aspects, '[i]ndependence of the judiciary means: (1) that every judge is free to decide matters . . . without any improper influences . . . (2) that the judiciary is independent of the executive and legislature'.¹ This definition seems to reflect consensus among scholars.² Putting aside the strategic behavior of the court *vis-à-vis* other branches of the government (Epstein and Knight, 1998), the principal-agent theory tells us that one of threats to judicial independence is appointment power. When the president, the prime minister, or legislators can hire and fire judges, the latter end up making decisions to please the former (Ramseyer and Rasmusen, 2003; Ramseyer and Rosenbluth, 1993: chs. 8–9, 1998: ch. 6). Similarly, reappointment and promotion also matter. The 'Universal Declaration on the Independence of Justice' warns that '[t]he assignment of a judge to a post within the court to which he is appointed is an internal administrative function to be carried out by the judiciary'. Otherwise, 'there is a danger of erosion of judicial independence by outside interference. It is vital that the court not make assignment as a result of any bias or prejudice or in response to external pressure' (Shetreet and Deschênes, 1985: 451). The typical test of judicial independence is whether judges of the same opinion as the executive are promoted more often than others (Maitra and Smyth, 2004; Salzberger and Fenn, 1999).

The Japanese judicature is a good case with which to run this test because it is one of the allegedly-most dependent judiciary branches. On the one hand, Article 76 of the Japanese Constitution stipulates that all judges shall be independent in the exercise of their conscience and shall be bound only by the Constitution and the laws. On the other hand, Article 79 provides the Cabinet with the power to designate a Chief Supreme Court Judge (appointed by the Emperor) and to appoint other Supreme Court Judges. According to Article 80, the judges of the inferior courts shall be nominated by the Supreme Court and appointed by the Cabinet. Ramseyer and Rasmusen (2003) argue that the existence of Articles 79 and 80 (stipulating Cabinet control of judicial promotion) means that in practice, even though judicial independence is guaranteed by Article 76, Japanese judges routinely validate what the government has done.³ Thus, judicial independence is questioned. They provide evidence for this claim by conducting a series of statistical tests, which indicate that judges with leftist preferences do worse in their careers. Their argument has been very influential and is frequently cited, not only in Japanese studies but also in the comparative (judicial) politics literature, which refers

¹ See 'International Bar Association Code of Minimum Standard of Judicial Independence' and 'Universal Declaration on the Independence of Justice' (Shetreet and Deschênes, 1985, chs. 32, 39, quote p. 414).

² Examples of recent studies are Ferejohn *et al.* (2007), Gibson (2006), and Hall (2005).

³ Ramseyer and Rasmusen (2006) reaffirm that their findings still hold true even after the 38-year dominance of the Liberal Democratic Party was interrupted in 1993.

to the Japanese judicature as one of the least independent judiciaries (e.g. Ginsburg, 2003: 24; Hall, 2005: 76–7; Solomon, 2007: 127).

We suggest that their dataset is incorrect to hold up their argument, the statistical methods they use are in fact inadequate to assess whether or not Japanese judges are dependent, and their research design is not appropriate for their own research agenda in the first place.

Among their findings, we focus on their first evidence (Ramseyer and Rasmusen, 2003: Table 2.6), that is, the analysis to suggest that a judge who once belonged to a leftist group will take longer to reach the position of ‘division chief’ (*sokatsu* in Japanese; a ‘moderately prestigious status’) during the period of conservative dominance of post-war Japan.⁴ Their findings consist of an OLS model that regresses the time taken for a judge to become division chief on a dummy variable identifying whether the judge once belonged to a leftist group. On closer analysis, however, some of the judges in the dataset used had not yet reached the position of division chief and therefore have no value on the dependent variable. This problem is called censoring. Some retire or die without reaching this position; some are too young to reach this position, and others skip this position by taking an irregular career path. Ramseyer and Rasmusen (2003) either drop those judges from their data set, or keep them in the dataset with arbitrary values substituted. We suspect that this procedure of data manipulation causes bias in their estimates.

In this article, we use the statistical technique of survival analysis. This enables researchers to examine how independent variables affect dependent variables that may or may not be censored. To address the problem that some judges may never become division chiefs, we employ a split population model of survival analysis. We also attempt to deal with the issue of what happens when a judge takes an irregular career path by using left truncation. To our knowledge, for the first time, we propose a survival analysis model that deals simultaneously with the problems of split population and left truncation.

We also extend the causal inference literature to censored, time-to-event data. We match every leftist judge with her non-leftist counterpart so that most variables are controlled and omitted variable bias is unlikely. We estimate the average treatment effects on *two* dependent variables, *Time* and *Event*. In particular, in the case of *Time*, after matching, we drop not only judges whose time is censored but also their corresponding judges so that most variables continue to be controlled.

Our findings indicate that leftist judges are by no means discriminated in terms of the timing and occurrence of promotion, against the argument of Ramseyer and Rasmusen (2003). Note that this result can be interpreted as either judicial independence or dependence; both are observationally equivalent. We are simply unable to tell whether leftist judges behave as they do because they want to, or because they fear the punishment from the conservative government. Though Ramseyer and Rasmusen (2003) present

⁴ A judge who reaches the position of division chief is the head of the division of a court and eligible to preside over court for the cases requiring consultation with associate judges.

other evidence that leftist judges or judges who issue rulings against the government are discriminated against in terms of promotion, this very fact is, against their argument, good evidence that the judiciary is independent of the executive. Judicial independence necessitates that the government acts to promote its followers and punishes those who protest its policies. On the contrary, in systems where government control of the judiciary is stronger, judges may not express *any* conflict with the executive branch. Therefore, the executive need not discriminate any judges for political reasons.

This article proceeds as follows. In the next section, the argument made by Ramseyer and Rasmusen (2003) is briefly reviewed and one of their regression analyses is replicated. The following section reconsiders how to analyze their data. To begin, survival analysis, left truncation, and split population models are explained. Then, we discuss the issue of causal inference and propose a new way to estimate average treatment effects on time and event. The subsequent section reanalyzes their data. First, their data are corrected. Second, we perform the matching procedure, check the balance of covariates and their interactions, and estimate average treatment effects on time and event. Third, we apply survival analysis to the matched data. The final section concludes with a summary of the findings and a brief discussion of judicial independence in general as well as other possible applications of our method.

Replication

Ramseyer and Rasmusen (2003) argue that judges with leftist preferences do worse in their careers. They suggest that Japanese courts use job postings as incentives and judges uphold the conservative positions of the longtime incumbent Liberal Democratic Party. Among various statistical analyses, Ramseyer and Rasmusen (2003) use the career data of 501 judges, and show that judges who once joined a leftist group were promoted more slowly than their peers. In this article, we begin by replicating the statistical analysis of judicial careers reported in Ramseyer and Rasmusen (2003: 41).

Ramseyer and Rasmusen (2003: 39) compiled judicial career data for all Japanese judges hired between 1959 and 1968 using a list that details all the posts held by these judges (Nihon Minshu Horitsuka Kyokai Shiho Seido Iinkai, 1998). Although it would be ideal to have salary information for these judges, there are no such data available. Instead, they focus on the time it takes for a judge to reach the status of division chief. This is somewhat justified as a substitution because it is not until judges become the chief judge that salaries among their cohort begin to differ. The dependent variable, *Time_Chief*, is defined as the year a judge first received a division chief appointment, less the year she graduated from the Legal Research and Training Institute.⁵

Ramseyer and Rasmusen (2003) make two main adjustments. First, they drop those judges who held non-judicial postings (generally regarded as prestigious) in the two years before their first division chief posting (Ramseyer and Rasmusen 2003 call them ‘stars’). In the rare cases where a judge served as *court* chief judge before serving as *division* chief judge, they treat this former position as the division chief posting,

⁵ Ramseyer and Rasmusen call it Time2Sok.

since such an appointment is unambiguously higher than division chief. Second, they drop the group of 164 judges who (i) never obtained a division chief appointment and (ii) quit or died before the mean *Time_Chief* for the rest of the group, which is 20.41 years (we call this group ‘dropouts’). This is because these judges have no value of *Time_Chief*. They kept, however, the 83 judges who quit or died after 20.41 years without an appointment to division chief and treated their death or resignation as their time of first division chief appointment (we call this group ‘losers’). Also, they treat the judges who did not have a division chief appointment as of 1997 (the last year the list of judges reports career data) as becoming a division chief in 1997. In addition to these adjustments, Ramseyer and Rasmusen (2003) drop judges who did not join the court within a year of graduating from the Legal Research and Training Institute (mostly, those who first became a prosecutor or attorney and later joined the court). Of 797 judges listed for the period between 1959 and 1968, 501 remain in the career data set.

To test for political discrimination, Ramseyer and Rasmusen (2003) introduce a key independent variable named *Leftist*, which takes the value 1 if a judge was a member of the communist-leaning Young Jurists League (YJL) in 1969 and 0 otherwise.⁶ They hypothesize that leftist judges are discriminated against in salary, and expect that, in a regression model of *Time_Chief*, the coefficient for *Leftist* would be statistically significant and positive.

Judges may be promoted not because they are favored for political reasons, but because they are talented. In order to control for quality of judge, the authors use three control variables. *Flunks* is defined as the number of times a judge failed the entrance exam to the Legal Research and Training Institute, which is estimated from passage year and birth year.⁷ *Elite_College* takes the value 1 if a judge graduated from either the University of Tokyo or the University of Kyoto, two of the most prestigious national universities in Japan, and 0 otherwise.⁸ *1st_Tokyo* takes the value 1 if a judge started at the Tokyo District Court and 0 otherwise, accounting for the fact that the most promising are generally assigned to the Tokyo District Court.⁹

In order to control for unobservable differences among the cohorts, dummy variables indicating the year in which a judge finished her legal education are incorporated into estimation.¹⁰

We replicate the OLS results in Ramseyer and Rasmusen (2003: 41) and display it in the left half of Table 1.¹¹ From their perspective, what is most notable is the

⁶ Ramseyer and Rasmusen originally name this variable YJL. They use Shiso Undo Kenkyusho (1969), which reproduced the list of members from the YJL’s newsletter.

⁷ The two cases with no birth records are assigned with the mean of *Flunks* for the group as a whole. The calendar year is used to calculate *Flunks*.

⁸ In cases with no school record, judges are assumed to be non-elite college graduates.

⁹ A judge that started at a summary court in the Tokyo area is not considered as a Tokyo starter.

¹⁰ Ramseyer and Rasmusen also examine another estimation model that includes variables identifying whether or not a judge is male and whether or not a judge started at a branch office, although we do not consider these, and focus on their simpler model.

¹¹ Following Ramseyer and Rasmusen, the estimates for the cohort dummies are not reported. The data used to estimate the model were available at <http://rasmusen.org/published/jbook/jbook.htm>.

Table 1. OLS results of Time to division chief

	Ramseyer and Rasmusen data			Corrected data		
	Coef.	S.E.	Sig.	Coef.	S.E.	Sig.
<i>Leftist</i>	0.919	0.538	0.088	0.818	0.540	0.131
<i>1st_Tokyo</i>	-1.383	0.804	0.086	-1.403	0.813	0.085
<i>Flunks</i>	0.014	0.071	0.849	0.014	0.070	0.840
<i>Elite_College</i>	0.086	0.516	0.867	0.098	0.523	0.852
Constant	22.905	0.785	0.000	22.914	0.789	0.000

Note: $N = 501$. The coefficients for cohort dummies are not reported. Significance is based on a two-tailed test. For both models, adjusted R^2 is 0.09.

coefficient of 0.919 for *Leftist*, which has a statistical significance of 0.088. This implies that judges who were members of the YJL in 1969 received their first division chief assignment roughly a year later than their non-YJL peers, controlling for these other factors.

Reconsideration

The way Ramseyer and Rasmusen (2003) handle and analyze the data, which we replicate above, have some problems, as we mentioned in the Introduction. To address these problems, we employ survival analysis, matching, and average treatment effects, which we elaborate on below.

Survival analysis

Latent time and censoring. We define latent time as the unobservable years a judge would have needed to receive a division chief appointment if the judge remained in the court until that moment. We say latent time is ‘censored’ if it ends without reaching a division chief post. Examples are ‘dropouts’ and ‘losers’ who quit or died early in their careers without obtaining a division chief appointment. If judges have not yet obtained a division chief appointment in 1997, the end year the data covers, their latent time is also censored. For judges whose latent time is censored, *Time_Chief* is not observed. Ramseyer and Rasmusen (2003) exclude the ‘dropouts’ from their dataset and assume *Time_Chief* as the time of censoring for the losers and the judges yet to become division chiefs in 1997. These adjustments, however, are problematic. The censored judges would receive a division chief posting some day in future if they had not died or continued to serve as a judge. Their true *Time_Chief* is longer than their presumed *Time_Chief*. Thus, the manipulation by Ramseyer and Rasmusen (2003) underestimates *Time_Chief* and causes bias in estimating their key coefficient, the effect of *Leftist* on *Time_Chief*.

Survival analysis is a set of statistical techniques designed to address censored latent time. It deals simultaneously with the questions of if an event occurs and when it occurs (e.g., if a judge becomes division chief and how long the judge takes

to reach the position).¹² We introduce a set of new variables. First, *Event* takes the value 1 if a judge received a division chief posting and 0 otherwise. While Ramseyer and Rasmusen (2003) dropped the ‘stars’, we keep them in our data set ($Event = 1$) and take into account the time they held non-judicial postings with the method we will explain later. We also keep the ‘dropouts’ in our data ($Event = 0$). Second, *Time* is defined as the years a judge needs to receive a division chief appointment (or to be censored without reaching division chief) after their graduation from the Legal Research and Training Institute. Note that both *Time* and *Time_Chief* measure the time a judge takes to reach the first division chief post, although the former is defined as one year greater than the latter. The former is thus amenable to logarithmic transformation.

Let D , T , and T^* denote *Event*, *Time*, and latent time, respectively. We assume T^* follows a log logistic distribution, whose probability density function is

$$f(T^*) = \frac{\frac{s}{m} \left(\frac{T^*}{m}\right)^{s-1}}{\left(1 + \left(\frac{T^*}{m}\right)^s\right)^2},$$

where m is the scale parameter as well as the median of T^* and s is the shape parameter. We are interested in what factors increase or decrease m , which is always positive. Thus, the logarithm of m is explained by linear combination of covariates W and coefficients b

$$m = \exp(bW).$$

If we observe that a judge becomes division chief ($Event = 1$), its contribution to likelihood is proportional to

$$f(T) = \frac{\frac{s}{\exp(bW)} \left(\frac{T}{\exp(bW)}\right)^{s-1}}{\left(1 + \left(\frac{T}{\exp(bW)}\right)^s\right)^2}.$$

If a judge has not become division chief before censoring ($Event = 0$), its contribution to likelihood is proportional to the probability of survival

$$S(T) = \int_T^\infty f(T^*) dT^* = \frac{1}{1 + \left(\frac{T}{m}\right)^s} = \frac{1}{1 + \left(\frac{T}{\exp(bW)}\right)^s}.$$

Parameters are estimated by maximizing total likelihood.

¹² See Box-Steffensmeier and Jones (2004) for survival analysis in social science. Salzberger and Fenn (1999) apply survival analysis to judicial promotion in UK, although their model does not take into account left truncation and split population.

Left truncation. Suppose that a judge held a non-judicial post (i.e., officers in the Secretariat of the Supreme Court or the Ministry of Justice) at the 10th year of her career, came back to the court in the 14th year, and became division chief in the 20th year. In this case, her latent time as a judge is censored at the 10th year ($Time = 10, Event = 0$), since she had no chance of becoming division chief while holding a non-judicial post. Her second latent time (or observation) resumes at the 14th year and ends with division chief post ($Time = 20, event = 1$). Note that she has no way of reaching a division chief post for the period up to the 14th year in terms of her second latent time. Thus, we should discount the 20 years she needed to reach a division chief post. This is called ‘left truncation’ in survival analysis. We introduce a new variable, $Begin(T_0)$, which is defined as the years a judge takes to resume a judicial posting since graduation from the Legal Research and Training Institute. In the above example, $Begin$ has a value of 14. Thus, a judge is assumed to have more than one latent time if she resumes a judicial career after censoring.

When latent time restarts at T_0 , both survival probability and failure density are conditioned on the fact that an event has not happened by T_0 , whose survival probability is $S(T_0)$. Thus, the conditional survival probability and the conditional probability density at $T > T_0$ are $S(T)/S(T_0)$ and $f(T)/S(T_0)$, respectively.

Split population. Ramseyer and Rasmusen (2003) implicitly assume that all judges eventually become division chief. This assumption is dubious, however. If leftist judges are severely discriminated, some of them should never be on the promotion track. In split population modeling, the population of observations is divided into two groups, one that would end up with promotion to division chief if it is not censored (on the promotion track) and one that would not even without being censored (off the promotion track). If a judge becomes division chief, we know that the judge is on the promotion track. But, if a judge’s career is censored, we are not sure whether the judge is off the promotion track or the judge on the promotion track has not become division chief yet. Thus, we estimate how likely judges are to be on the promotion track, instead of measuring whether they are.

Let r be the probability for a judge i to be on the promotion track. Then, the probability to be censored at T ($D = 0$) is the probability that a judge is off the track, plus the probability that a judge on the track has not become division chief at T : $(1 - r) + rS(T)$. The probability density to become division chief at T is the probability that a judge is on the track, times the probability density that the judge becomes division chief at T : $rf(T)$. Moreover, we explain r by logistic function of linear combination of some covariates Z and coefficients a

$$r = \frac{1}{1 + \exp(-aZ)}.$$

To sum, likelihood of a judge who becomes division chief at T is proportional to $rf(T)/S(T_0)$, namely

$$\frac{1}{1 + \exp(-aZ)} \frac{\frac{s}{\exp(bW)} \left(\frac{T}{\exp(bW)}\right)^{s-1}}{\left(1 + \left(\frac{T}{\exp(bW)}\right)^s\right)^2} \bigg/ \frac{1}{1 + \left(\frac{T_0}{\exp(bW)}\right)^s},$$

and likelihood of a judge who is censored (ceases to be observed without becoming division chief) at T is proportional to $(1-r)+rS(T)/S(T_0)$, that is

$$\frac{1}{1 + \exp(aZ)} + \left(\frac{1}{1 + \exp(-aZ)} \frac{1}{1 + \left(\frac{T}{\exp(bW)}\right)^s} \bigg/ \frac{1}{1 + \left(\frac{T_0}{\exp(bW)}\right)^s} \right).$$

Given the data $D, T, T_0, W,$ and $Z,$ the parameters $(a, b,$ and $s)$ are estimated so that the total likelihood is maximized.¹³

Matching and average treatment effects

Even if we try to specify the model, taking into account every factor we need to, we are unlikely to do it perfectly. Latent time T^* may not follow a log logistic distribution and the population split function may be probit instead of logistic, though we cannot be sure. We simply cannot avoid model-dependent results in the application of any parametric model to raw data.¹⁴

The causal inference literature takes model dependence seriously and aims to ameliorate it. One way to do this is matching the data and estimating the average treatment effect, which we explain below. These methods, however, have rarely been applied to censored time-to-event data. Bland and Altman (1994) lament ‘[s]ometimes there is no suitable method of matched analysis, as in survival analysis’. This is because we cannot observe the values of (unobserved) latent time T^* of all units because of censoring. Below, we propose a new way to estimate the average treatment effect for both (unobserved) latent time T^* and *Event D*.

Average treatment effect on event. Following the potential outcome framework, we assume that *Event D* for judge $i, D_i,$ takes the value of D_{1i} if judge i belongs to the leftist group, YJL, and $D_i = D_{0i}$ otherwise. Hence, the effect of *Leftist* on *Event D_i* for judge i should be $D_{1i} - D_{0i}$. One of the quantities of interest for most scholars is called the

¹³ We use a statistical environment R for estimation and write our own code to deal with the issue of left truncation in the split population model. Box-Steffensmeir *et al.* (2006) offers a STATA program to run the split population model, although their program does not consider left truncation. Clark and Regan (2003) and Hettinger and Zorn (2005) are other applications of the split population model in political science, which do not take into account left truncation, either.

¹⁴ As for strict causal inference in political science, see Diamond and Sekhon (2013), Ho *et al.* (2007), and King and Zeng (2007).

‘average treatment effect’ in the causal inference literature and is expressed as

$$\frac{1}{n} \sum_{i=1}^n D_{1i} - D_{0i}.$$

Here, ‘the fundamental problem of causal inference’ (Holland, 1986) arises because we can only observe D_{1i} or D_{0i} , not both. Let X be a dummy variable of *Leftist* (which is called ‘treatment’ in the literature of causal inference). If $X_i = 1$, we observe D_{1i} but not D_{0i} . If $X_i = 0$, we observe D_{0i} but not D_{1i} . How can we estimate the average treatment effect?

One way is matching. Let M be a vector of all measured covariates. Consider a matching function $j = m(i)$ that satisfies $M_i = M_j$ and $X_i = 1 - X_j$. We match judge $m(i)$ to judge i . For example, YJL Judge Tsutomu Ueno ($X = 1$) is matched to non-YJL Judge Kohei Araki ($X = 0$). Both were born in 1933, graduated from non-elite universities ($Elite_College = 0$), became a judge in 1959 ($Flunks = 1959 - 1933 - 24 = 2$, where 24 is the youngest age for a college graduate to become a judge), started their careers outside of Tokyo ($1st_Tokyo = 0$), and reached division chief without leaving the court ($Begin = 0$).

We estimate the average treatment effect on event by the difference in means between *Event* of YJL judges and that of non-YJL judges

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n [X_i \{D_i(X_i = 1, M_i) - D_{m(i)}(X_{m(i)} = 0, M_{m(i)} = M_i)\} \\ & + (1 - X_i) \{D_{m(i)}(X_{m(i)} = 1, M_{m(i)} = M_i) - D_i(X_i = 0, M_i)\}]. \end{aligned}$$

As both judges have the same values of covariates (M) but YJL membership (X) and, possibly, some omitted variables are different, the difference in D between them cannot be attributed to that in M but rather to that in X or any omitted variables. Now we assume ignorability; that is, conditioned on M_i , D_{1i} , and D_{0i} are assumed to be independent of X_i . Put another way, matching is so successful that once M_i is controlled, no omitted variables are associated with D_{1i} , D_{0i} , and X_i . It follows that $E(D_{1i}|X_i = 1, M_i) = E(D_{1i}|X_i = 0, M_i) = E(D_{1i}|M_i)$, and $E(D_{0i}|X_i = 1, M_i) = E(D_{0i}|X_i = 0, M_i) = E(D_{0i}|M_i)$. Therefore, the difference in D between them can only be attributed to the difference in X and not the difference in either M or any omitted variables, and the difference in means of the matched data is an unbiased estimator of the average treatment effect.

As *Leftist* and *Event* are binary variables, the data can be summarized in a two-by-two table, and a chi-square test can be used to test the null hypothesis of no average treatment effect as well.

As is often the case with survival analysis, we assume censoring is non-informative, namely, censoring time T_i^C is independent of T_i^* where $T_i = T_i^*$ and $D_i = 1$ if $T_i^C > T_i^*$ and $T_i = T_i^C$ and $D_i = 0$ if $T_i^C < T_i^*$. Thus, in matched data, the proportions of judges who become division chiefs should be smaller in the YJL group than in the

non-YJL group if at least one of the following three conditions holds: first, *Leftist* shortens censoring time (for instance, YJL judges retire or die earlier than non-YJL judges); second, the proportion of judges on the promotion track (namely, population split) is smaller in the YJL group; and, third, *Leftist* lengthens latent time T^* . We turn to the method that examines whether the last condition is met or not.

Average treatment effect on time. The average treatment effect on latent time T^* cannot be estimated in the same way as the average treatment effect on event because latent time T^* is not observed if either of the matched judges is (or both are) censored ($D_i = 0$ and/or $D_{m(i)} = 0$). In other words, ‘the fundamental problem of causal inference’ is more severe; not just one but two potential outcomes, T_{1i}^* and T_{0i}^* , may be unobserved. As censoring is a post-treatment variable, matching non-censored judges would only cause bias.

Our solution is that, in the matched data, when we drop a censored judge, we also drop the corresponding judge. Therefore, the balance of measured covariates remains and ignorability holds. In addition, note that we drop censored judges after, not before, matching. We propose to estimate the average treatment effect of X on latent time T^* by difference in means between *Time* of YJL judges and that of non-YJL judges for non-censored pairs alone

$$\frac{1}{n(+)} \sum_{i=1}^n D_i D_{m(i)} [X_i \{T_i(X_i = 1, M_i) - T_{m(i)}(X_{m(i)} = 0, M_{m(i)} = M_i)\} + (1 - X_i) \{T_{m(i)}(X_{m(i)} = 1, M_{m(i)} = M_i) - T_i(X_i = 0, M_i)\}],$$

where $n(+)$ is the number of observations where $D_i = 1$ and $D_{m(i)} = 1$.

Remember that censoring is assumed to be independent of latent time. It follows that $E(T_{1i}|D_i = 1) = E(T_{1i}|D_i = 0) = E(T_{1i})$ and $E(T_{0i}|D_i = 1) = E(T_{0i}|D_i = 0) = E(T_{0i})$. Accordingly, even if we drop a matched pair of judges (not just a judge) where at least one judge is censored, our difference-in-means estimator of the average treatment effect will not be biased (though it will be inefficient). Note that, if we do not use matched data, dropping a censored judge instead of a pair of judges would induce bias (underestimate latent time of censored judges).

Re-analysis

Data correction

In the process of replicating the analysis in Ramseyer and Rasmusen (2003), Masuyama (2005) finds that some of their data do not correspond to the information available in the data sources. In the right half of Table 1, we report the OLS results using the data with corrections.¹⁵ As indicated by the summary statistics in Table 2, the

¹⁵ For the details of corrections, see endnote 10 in Masuyama (2005: 439–40). Our R codes are available at <http://www-cc.gakushuin.ac.jp/~e982440/research/replication/>.

Table 2. Summary statistics

	Ramseyer and Rasmusen data				Corrected data			
	Mean	Std	Min	Max	Mean	Std	Min	Max
<i>Time_Chief</i>	21.711	5.317	11	37	21.681	5.364	11	37
<i>Leftist</i>	0.281	0.450	0	1	0.289	0.454	0	1
<i>1st_Tokyo</i>	0.104	0.305	0	1	0.104	0.305	0	1
<i>Flunks</i>	4.828	3.441	0	18	4.880	3.505	0	18
<i>Elite_College</i>	0.317	0.466	0	1	0.319	0.467	0	1

Note: $N = 501$.

data correction makes no substantial difference. Nonetheless, the estimate of 0.818 for *Leftist* is slightly smaller, with the almost same standard error, but it loses its statistical significance ($p = 0.131$). Thus, even if one uses their methods, their argument is not supported by the information available in the data sources.

Matching and average treatment effects

Unlike Ramseyer and Rasmusen (2003), Masuyama (2005) includes ‘stars’ and ‘dropouts’ in his data. Consequently, the career data set consists of all 797 judges, which generates 1,086 observations.¹⁶ Below, we use this data set with our corrections.¹⁷ The median, minimum, and maximum of *Time* for the enlarged data set is respectively 19, 1, and 38 years.

We match on *1st_Tokyo*, *Flunks*, *Elite_College*, *Begin*, *Strata* (the number of previous censoring (observations) of the same judge),¹⁸ their interactions and squares (except squared *1st_Tokyo* and squared *Elite_College* because they are dummy variables), ten cohort dummies, and interactions between cohort dummies and the other five variables.

It is impossible to match every YJL judge with a non-YJL judge who has the exactly the same covariates (exact matching). As a second best method, we will match every observation to the most similar observation with replacement (nearest neighbor matching). The difficult part is how to detect such pairs, and this is why various

¹⁶ For example, when a judge returns to a judicial post after holding a non-judicial post, this judge appears twice in the dataset (see ‘left truncation’ in the text).

¹⁷ We code seven judges as YJL members (ID in parentheses): Akira Abe (1102), Katsutoshi Kuwata (1127), Takao Oguri (1220), Yasukuni Yoneda (1285), Osamu Sawada (1334), Kouji Hirokawa (1574), and Jiro Inoue (1603). Appendix to Masuyama (2005, 411–35) also incorrectly lists two judges starting at a summary court in the Tokyo area as Tokyo starters: Akira Toyoshi (1846) and Noboru Arakawa (1904). Note that these are irrelevant to the results reported in Tables 1 and 2. The seven YJL judges are not included in the data Ramseyer and Rasmusen (2003) use, and the two non-Tokyo starters are correctly coded in the data Masuyama (2005) uses.

¹⁸ For example, again, suppose that a judge returns to a judicial post after holding a non-judicial post. In this case, time to non-judicial post is the first latent time for this judge, while the resumed latent time is the second one.

Table 3. Cross-tabulation of *Leftist* and *Event* (promotion to division chief) in matched data

		<i>Event</i> (promotion to division chief)			(%)
		0	1	<i>N</i>	
<i>Leftist</i>	0	52.5	47.5	1,866	
	1	51.8	48.2	1,866	

Note: Chi-squared value continuity correction is 0.182. Since degree of freedom is 1, p-value is 0.670.

matching methods are suggested. This article employs a genetic matching method by Diamond and Sekhon (2013), which results in 1,866 matched pairs.¹⁹

Balance of the covariates between YJL judges and non-YJL judges is almost completely achieved. For example, six cohort dummies as well as six interactions between cohort dummies and *1st_Tokyo* are perfectly balanced. Even for the worst balanced variable, *Flunks*, before-matching mean values are 3.56 and 4.71 for YJL and non-YJL judges respectively, while those for after-matching are 3.67 and 4.07; the difference becomes narrower.

In order to summarize the balance of covariate distributions, we compare the distributions of propensity scores between YJL and non-YJL judges before and after matching. Propensity scores are predicted values of logistic analysis where we regress *Leftist* on all matched-on covariates, implying the probability that the observation is an YJL judge. When balance is achieved and a distribution of covariates is independent of *Leftist*, the distributions of propensity scores should not differ between the two judge groups. According to Figure 1, the distributions of propensity scores are considerably different before matching, while they are similar after matching. Clearly, the covariates are much better balanced after matching.

Figure 2 depicts the life tables. The survival curves are indistinguishable between the two groups up to the 25th year, while they differ after that. It is difficult to decide whether *Leftist* affects *Time* to division chief.

Table 3 cross-tabulates the percentage of *Leftist* and *Event* in the matched data. Using all 1,866 pairs, the average treatment effect of *Leftist* on *Event* is 0.007, while its standard error (estimated by using the central limitation theorem) is 0.015. Using uncensored 495 pairs only, the average treatment effect of *Leftist* on *Time* is 0.170, while its standard error is 0.274. Thus, both effects are not significant even at the 10% level and we cannot say that *Leftist* has any effect on *Time* to division chief or *Event* of promotion to division chief. In addition, even at the 10% significance level, a chi-square test with

¹⁹ For matching, we use Matching library. The option for 'estimand' is ATE, while the option for 'boot' is 1000 following recommendation for publication standard. We match on interactions and squares of these variables so as to balance not only marginal distributions of these variables but also joint distribution of them (up to second order). A judge may appear in the matched dataset twice or more.

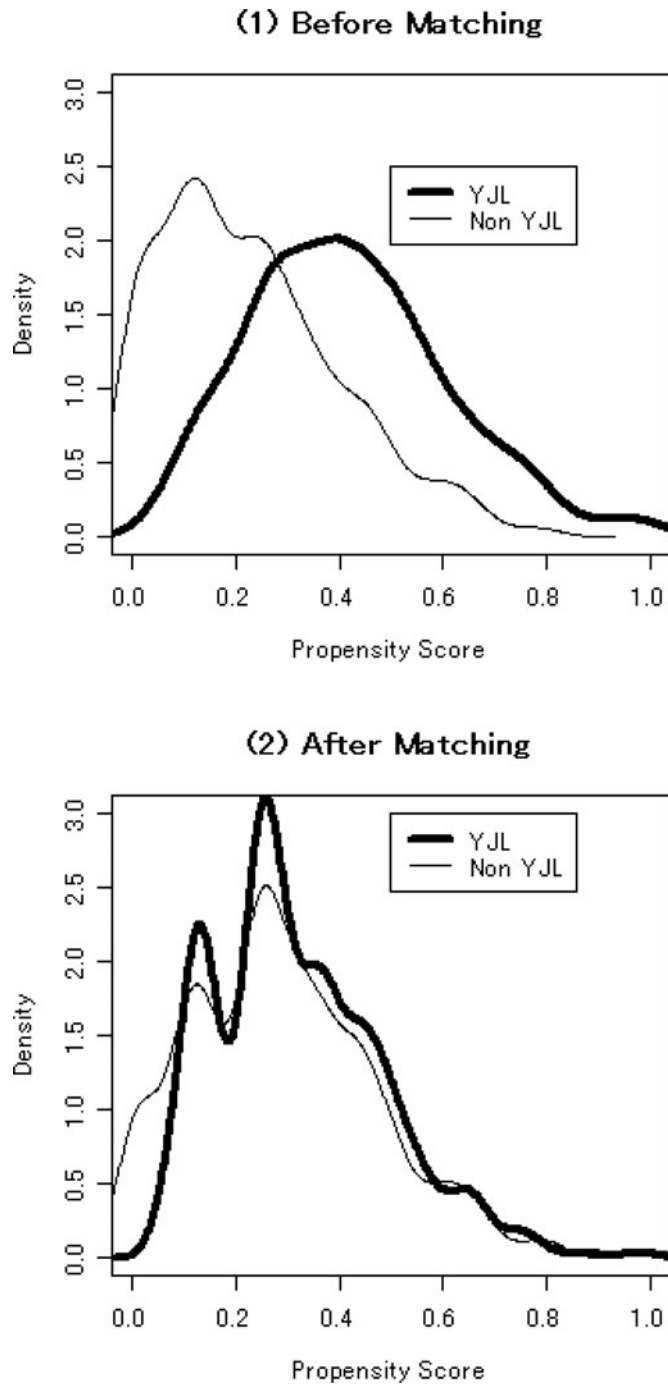


Figure 1. Distributions of propensity scores

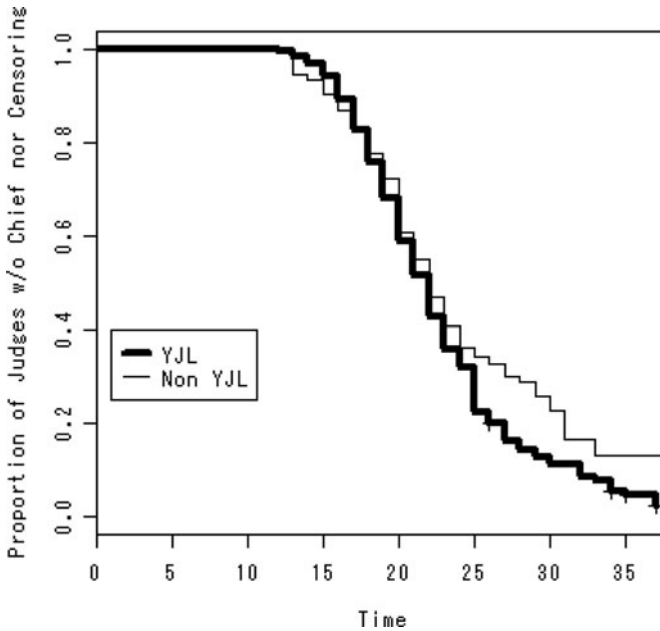


Figure 2. Comparison of life tables between the two judge groups (matched data)

continuity correction also fails to reject the null hypothesis that *Leftist* does not matter whether judges are promoted to division chief before censoring. These results also suggest that *Leftist* does not shorten censoring time and proportions of judges on the promotion track are equal between the YJL group and the non-YJL group. The next subsection considers these points.

Survival analysis

Even if researchers use parametric modeling, it is much better to use matched data instead of raw data. This is because model dependence is reduced. Ho *et al.* (2007) recommends that we use non-parametric matching as ‘preprocessing’, namely, a tool to remove bias of correlated control variables, before we perform any parametric analysis. We still rely on our survival model specification such as a log logistic distribution for latent time and the split population assumption, although we need not worry about which independent variables are included and whether they should be squared or interacted with other variables. Even if matching fails to balance these covariates completely, the parametric models can control them as long as they are correctly specified.

Table 4 reports the result of survival analysis of *Time* to division chief. We calculate standard errors and confidence intervals by bootstrapping the matched data, which is a little different from usual bootstrap. We resample the matched pairs, not the

Table 4. Split population survival analysis of Time to division chief (matched data)

	Coef.	S.E.	95% confidence interval	
Latent time (log logsitic)				
<i>Leftist</i>	0.006	0.013	-0.012	0.040
<i>1st_Tokyo</i>	-0.090	0.026	-0.160	-0.062
<i>Flunks</i>	-0.007	0.003	-0.014	-0.001
<i>Elite_College</i>	0.017	0.014	-0.010	0.045
Constant	3.211	0.020	3.176	3.251
Log(Shape)				
	-2.120	0.033	-2.178	-2.042
Split population				
<i>Leftist</i>	-1.354	5.413	-3.006	0.062
<i>1st_Tokyo</i>	12.965	8.767	-1.916	25.574
<i>Flunks</i>	-0.302	0.576	-0.737	-0.212
<i>Elite_College</i>	-0.188	2.243	-1.079	5.818
Constant	5.094	7.255	4.044	20.543

Notes: $N = 3,732$. The coefficients for cohort dummies are not reported. Standard error and 95% confidence interval are calculated by bootstrap. Log likelihood is -5646.294 .

observations, with replacement in order to retain a balanced joint distribution of covariates between the two groups. We repeat this re-sampling 1,000 times, calculate standard deviances of 1,000 sets of parameter estimates in the survival analysis, and multiply them by $1000/(1000-1)$, which will be our unbiased estimates of standard errors. A 95% confidence interval is derived by 2.5% quantile value and 97.5% ones without assuming any distribution of estimates.

For the covariates in both latent time and split population models, we use the same independent variables introduced by Ramseyer and Rasmusen (2003). The coefficients of *Leftist* are significant in neither latent time part nor split population part. We cannot say that YJL membership matters either for whether judges are promoted to division chief or for whether they are on the promotion track. Thus, this parametric model of survival analysis reaffirms the irrelevance of YJL membership, which the non-parametric analysis of average treatment effects on *Time* and *Event* show in the previous subsection.

To briefly discuss the estimates for the control variables, *1st_Tokyo* and *Flunks* fasten *Time* to division chief. *Flunks* also decreases the chance to be on the track to division chief.²⁰ Since *Flunks* is defined by the age of a judge starting her career, its estimates imply that those who took relatively long to become a judge may reach a division chief post earlier due to their age. We find that *Elite_College* is insignificant in both latent time part and split population part.

²⁰ Note that standard errors are not robust against deviant values and the coefficient of *Flunks* in split the population part is significant despite the relatively larger standard errors.

Conclusion

Is the Japanese judicial system independent of political control? Ramseyer and Rasmusen (2003) studied how long it takes for a judge to reach a division chief post, and argued that judges with leftist preferences do worse in their careers. This article shows that their analysis is problematic as it is based on inaccurate data; takes no account of the issues of censoring, left truncation, and split population; and depends critically on their model (mis-)specification. In this article, we corrected their data; we employed parametric survival analysis to deal with the issues of censoring, left truncation and split population, and we conducted matching to reduce model dependence and estimate non-parametric average treatment effects of YJL membership on time to, and occurrence of, promotion to division chief. As a result, we cannot find any evidence that YJL membership harms the judge's chances of reaching the post of division chief.

We interpret these findings as indicating either that leftist judges *are* independent and the government is not discriminating against them, or that they are dependent and the government *need not* to discriminate them. Moreover, if YJL affiliation damaged the judges' career, it would imply that some judges were independent enough to resist and, thus, be discriminated against by, the government. In other words, judicial independence necessitates that the government acts to promote its followers and punishes those who protest its policies. On the contrary, in systems where the government controls the judiciary, judges may not express any conflict with the executive branch in the first place. Viewed in this light, other analyses in Ramseyer and Rasmusen (2003) which suggest that the government penalized recalcitrant judges, are all problematic. In short, however significant or insignificant the main coefficient is, their research design can never allow us to conclude that leftist judges are dependent. Given that the Japanese judicature is one of the allegedly-most dependent judiciary branches, we suggest that our findings should also cause scholars of comparative judicial politics to reconsider their measurements of judicial independence.

This article also aims to contribute to political methodology. Our new survival analysis model takes into consideration both left truncation and split population. Moreover, for the first time, we extend the causal inference literature to censored time-to-event data which have two dependent variables and propose a method of estimating the average treatment effects on time and event. Needless to say, our methods are not limited to application to judges in Japan. We offer two examples of where they may be applicable below.

One instance is length of legislative deliberation. How long does it take for bills to pass in the legislature? What is important to keep in mind when studying this is that some bills may not pass the legislature by the end of session (censoring: see Becker and Saalfeld, 2004; Martin, 2004; Masuyama, 2001). Bills are introduced not only at the beginning of the session but also during the session (left truncation: Masuyama, 2001). Moreover, some bills will never pass even if there is no end in the legislative session (split population).

Another promising application is to the democratic peace debate. Are democratic dyads less likely to fight than other dyads? There are several problems with the literature. One is that currently peaceful dyads may fight in the future (censoring). A second is that some dyads have no opportunities to fight (split population, Clark and Regan, 2003). Other dyads fight several times (repeated events and left truncation).

In both examples, to reduce model dependence, matching is preferable. Thus, average treatment effects on time and event may help scholars to find rigorous causal effects. We believe that our analysis provides a good example of causal inference of censored time-to-event data for the students of both judicial politics and political methodology.

About the authors

Kentaro Fukumoto is Professor of Political Science at Gakushuin University and visiting fellow at Washington University in St Louis. He received his Ph.D. from University of Tokyo. His research interests include political methodology, legislative studies, and electoral studies. He is the author of *Nihon no Kokkai Seiji: Zen Seifu Rippo no Bunseki* [Politics in the Japanese Diet: A Statistical Analysis of Postwar Government Legislation] (Tokyo Daigaku Shuppan Kai, 2000) and *Rippo no Seido to Katei* [Legislative Institutions and Process] (Bokutaku Sha, 2007). His articles have appeared in *American Political Science Review*, *American Journal of Political Science*, *Legislative Studies Quarterly*, *Journal of American Statistical Association*, and *Japanese Journal of Political Science*.

Mikitaka Masuyama is Professor of Political Science at the National Graduate Institute for Policy Studies. He received his Ph.D. from the University of Michigan. His research interests include legislative institutions, Japanese politics and political methodology. He is the author of *Gikai Seido to Nihon Seiji: Giji Un'ei no Keiryō Seijigaku* [Agenda Power in the Japanese Diet: A Duration Analysis of Lawmaking] (Bokutakusha, 2003). His articles have appeared in *Japanese Journal of Political Science*, *Journal of Legislative Studies and Social Science Japan Journal*.

References

- Becker, Rolf and Thomas Saalfeld (2004), 'The Life and Times of Bills', in Herbert Doring and Mark Hallerberg (eds.), *Patterns of Parliamentary Behavior: Passage of Legislation across Western Europe*, Aldershot: Ashgate, pp. 57–90.
- Bland, J. Martin and Douglas G. Altman (1994), 'Matching', *British Medical Journal*, 309:1128.
- Box-Steffensmeier, Janet and Bradford Jones (2004), *Event History Modeling: A Guide for Social Scientists*, New York: Cambridge University Press.
- Box-Steffensmeier, Janet M., Peter M. Radcliffe, and Brandon L. Bartels (2006), 'The Incidence and Timing of PAC Contributions to Incumbent US House Members, 1993–94', *Legislative Studies Quarterly*, 30(4): 549–79.
- Clark, David H. and Patrick M. Regan (2003), 'Opportunities to Fight: A Statistical Technique for Modeling Unobservable Phenomena', *Journal of Conflict Resolution*, 47(1): 94–115.

- Diamond, Alexis and Jasjeet S. Sekhon (2013), 'Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies', *Review of Economics and Statistics*, 95(3): 932–45.
- Epstein, Lee and Jack Knight (1998), *The Choices Justices Make*, Washington, DC: Congressional Quarterly Inc.
- Ferejohn, John, Frances Rosenbluth, and Charles Shipan (2007), 'Comparative Judicial Politics', in Charles Boix and Susan Carol Stokes (eds.), *The Oxford Handbook of Comparative Politics*, Oxford: Oxford University Press, pp. 727–51.
- Gibson, James L. (2006), 'Judicial Institutions', in R. A. W. Rhodes, Sarah A. Binder, and Bert A. Rockman, *The Oxford Handbook of Political Institutions*, Oxford: Oxford University Press, pp. 514–34.
- Ginsburg, Tom (2003), *Judicial Review in New Democracies: Constitutional Courts in Asian Cases*, Cambridge: Cambridge University Press.
- Hall, Kermit L. (2005), 'Judicial Independence and the Majoritarian Difficulty', in K. L. Hall and K. T. McGuire, *The Judicial Branch*, Oxford: Oxford University Press, pp. 60–85.
- Hettinger, Virginia A. and Christopher Zorn (2005), 'Explaining the Incidence and Timing of Congressional Responses to the US Supreme Court', *Legislative Studies Quarterly*, 30(1): 5–28.
- Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth Stuart (2007), 'Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference', *Political Analysis*, 15(3): 199–236.
- Holland, Paul W. (1986), 'Statistics and Casual Inference', *Journal of American Statistical Association*, 81: 945–60.
- King, Gary and Langche Zeng (2007), 'When Can History be Our Guide? The Pitfalls of Counterfactual Inference', *International Studies Quarterly*, 51(1): 183–210.
- Maitra, Pushkar and Russell Smyth (2004), 'Judicial Independence, Judicial Promotion and the Enforcement of Legislative Wealth Transfers: An Empirical Study of the New Zealand High Court', *European Journal of Law and Economics*, 17(2): 209–35.
- Masuyama, Mikitaka (2001), 'Agenda Power in the Japanese Diet: A Duration Analysis of Lawmaking', Ph. D. Dissertation, University of Michigan.
- Masuyama, Mikitaka (2005), 'Saibankan kyaria no seizon bunseki [Survival Analysis of Judicial Career]', *Seikei Hogaku*, 61: 411–56.
- Martin, Lanny W. (2004), 'The Government Agenda in Parliamentary Democracies', *American Journal of Political Science*, 48(3): 445–61.
- Nihon Minshu Horituka Kyokai Shiho Seido Iinkai (1998), *Zensaibankan keireki soran* [Biographical Information on All Judges], 3rd edn, Tokyo: Kojinsha.
- Ramseyer, J. Mark and Eric B. Rasmusen (2003), *Measuring Judicial Independence: The Political Economy of Judging in Japan*, Chicago: University of Chicago Press.
- Ramseyer, J. Mark and Eric B. Rasmusen (2006), 'The Case for Managed Judges: Learning from Japan after the Political Upheaval of 1993', *University of Pennsylvania Law Review*, 154: 1879–930.
- Ramseyer, J. Mark and Frances M. Rosenbluth (1993), *Japan's Political Marketplace*, Cambridge, MA: Harvard University Press.
- Ramseyer, J. Mark and Frances M. Rosenbluth (1998), *The Politics of Oligarchy: Institutional Choice in Imperial Japan*, Cambridge: Cambridge University Press.
- Salzberger, Eli and Paul Fenn (1999), 'Judicial Independence: Some Evidence from the English Court of Appeal', *Journal of Law and Economics*, 42(2): 831–47.
- Shetreet, Shimon and Jules Deschênes (1985), *Judicial Independence: The Contemporary Debate*, Dordrecht: M. Nijhoff.
- Shiso Undo Kenkyusho (1969), *Osorubeki saiban* [Fearsome Trials], Tokyo: Zenbosha.
- Solomon, Peter H. (2007), 'Courts and Judges in Authoritarian Regimes', *World Politics*, 60(1): 122–45.