Appendix to "Comparing Partial Likelihood and Robust Estimation Methods for the Cox Regression Model"

Bruce A. Desmarais* Jeffrey J. Harden†

Contents

1 Implementing the CVMF Test in R **i**

1.1 An Example ii

2 Additional Monte Carlo Results **iv**

3 Additional Replications: PLM Selected, IRR Less Support **v**

3.1 Martin (2004) v

3.2 Golder (2010) vii

1 Implementing the CVMF Test in R

The IRR estimator and CVMF test are currently only available in R. However, Stata users can follow these steps to use the methods in R. First, for a Cox model estimated in Stata with outcome variable `y` and predictors `x1` and `x2`, save the data in a .dta file:

```
stset y, failure(fail)
stcox x1 x2
save "example.dta"
```

*Assistant Professor, Department of Political Science, University of Massachusetts–Amherst, 420 Thompson Hall, 200 Hicks Way, Amherst, MA 01003, 413.545.1992, desmarais@polsci.umass.edu, http://people.umass.edu/bruced/.

†Ph.D Candidate, Department of Political Science, University of North Carolina at Chapel Hill, 312 Hamilton Hall, CB #3265, Chapel Hill, NC 27599, 919.600.9247, jjharden@unc.edu, http://jjharden.web.unc.edu/.

Then, in R, the data can be imported into R as follows:

```
library(foreign)
example <- read.dta("example.dta")
```

Next, read in the accompanying script file "CVMF.R" to load the `CVMF()` function, which estimates the model with PLM and IRR and performs the CVMF test. Note that doing this requires the `survival` and `coxrobust` packages. Finally, write out the model in R's syntax and feed it into the `CVMF()` function. Here we assign this to an object called `results`. The argument `trunc` is the IRR truncation parameter. The default is 0.95.

```
source(CVMF.R)
form <- Surv(y, event = fail) ~ x1 + x2
results <- CVMF(formula = form, data = example, trunc = .95)
```

1.1 An Example

The file "CVMF.R" contains a basic example of how to use the `CVMF()` function. First, it begins by setting the seed, then creating two independent variables, one of which contains measurement error (`x2e`).

```
## Set the seed for replication purposes
 set.seed(12345)
#
# Create two covariates with measurement error in the second
 x1 <- rnorm(100)
 x2 <- rnorm(100)
 x2e <- x2 + rnorm(100, 0, 0.5)

## Create the dependent variable with the unobserved x2
## Each coefficient has a true value of 1
 y <- rexp(100, exp(x1 + x2))
 y <- Surv(y)
#
## Put the observed variables into a data frame
 dat <- data.frame(y, x1, x2e)
#
## Define the formula
 form <- y ~ x1 + x2e
```

Next, the CVMF test is performed and the results are stored in the object `results`. The selection of the CVMF test is automatically written out on the screen.

```
results <- CVMF(formula = form, data = dat)
IRR supported with a two-sided p-value of 0.021
```

You can also look at the PLM, IRR, and CVMF results in detail by using the dollar sign ($) after the name of the results object.[1] Notice that in this case both of the IRR coefficient estimates are closer to the truth (1) than are the PLM estimates, but the PLM standard errors are smaller than the IRR standard errors.

```
## Take a look at results
results$irr

Call:
coxr(formula = formula, data = data, trunc = trunc)

Partial likelihood estimator
     coef exp(coef) se(coef)        p
x1  0.925      2.52    0.126 2.43e-13
x2e 0.784      2.19    0.122 1.53e-10

Wald test=69.1 on 2 df, p=9.99e-16

Robust estimator
     coef exp(coef) se(coef)        p
x1  0.964      2.62    0.226 1.95e-05
x2e 0.834      2.30    0.193 1.52e-05

Extended Wald test=21.9 on 2 df, p=1.73e-05

## Now the test
 results$cvmf

        Exact binomial test

data:  sum(Cvll.r > Cvll.c) and n
number of successes = 62, number of trials = 100, p-value = 0.02098
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
```

[1]Note that the IRR output automatically reports both PLM and IRR results.

```
 0.5174607 0.7152325
sample estimates:
probability of success
                  0.62
```

2 Additional Monte Carlo Results

Here we present the first aspect of the Monte Carlo study: a comparison of the relative performance of PLM and IRR under the contamination conditions imposed. These results are presented in Figure A.1. The first row (panels a-c) presents the absolute value of the bias in the PLM estimates divided by the absolute value of the bias with IRR. The second row (panels d-f) gives the mean squared error (MSE) of the PLM estimates divided by the IRR MSE.[2] In all six graphs, values greater than 1 indicate that IRR performance is better (smaller absolute bias or smaller MSE) than that of PLM.

[Insert Figure A.1 here]

Note first that when measurement error is introduced to the covariate by adding noise (panels a and d), IRR outperforms PLM in terms of bias and MSE in all cases. The bias of PLM ranges from 110% to 130% of that of IRR and decreases as the variance of the measurement error increases. Sample size has little effect on the relative bias, but the MSE of PLM is higher than IRR MSE in all cases, and especially with the larger sample size (ranging from 105% to 140% of IRR MSE). In all cases with an omitted variable (panels b and e), the bias and MSE of IRR are lower than those of PLM, though the two methods' performances become similar as the correlation between the omitted and included variables increases. In addition, the relative MSE of PLM to IRR is greater when the sample size is 500.

Specification error introduced via heterogeneous effects (panels c and f) presents a similar result. The absolute bias and MSE of PLM are always higher than the bias and MSE of IRR.

[2]Recall that the true parameter value, β, is set to 1. Bias is computed as $\bar{\hat{\beta}} - 1$, where $\bar{\hat{\beta}}$ is the average estimate of the regression coefficient over the 500 iterations from the respective case in the Monte Carlo study, and the mean squared error is computed as $\frac{1}{500}\sum_{i=1}^{500}\left(\hat{\beta}_i - 1\right)^2$.

Also, there does not appear to be a relationship between the variance of the effect and the relative performance of PLM to IRR. Overall, Figure A.1 supports the claims and findings of earlier work on this topic, showing that considerable gains in estimator performance are possible with IRR when the assumptions underlying the Cox model are violated through specification problems.

3 Additional Replications: PLM Selected, IRR Less Support

Here we describe two additional replications. In both cases, the CVMF test selects PLM and IRR shows less support for the original hypotheses.

3.1 Martin (2004)

Martin (2004) examines how government coalitions organize the legislative agenda through an analysis of the sequence and timing of government bills introduced to the legislature. Martin tests his expectations on an original dataset covering government bills introduced in four parliamentary democracies. Specifically, the dependent variable is the number of days between government formation and the introduction of a bill to the legislature in Belgium, Germany, Luxembourg, and Netherlands during the 1980s and 1990s. One main independent variable of interest is a measure of each bill's saliency to the coalition (*Government Issue Saliency*), which is derived from the expert survey data of Laver and Hunt (1992). His theory predicts that, for a bill of average divisiveness, *Government Issue Saliency* shortens the time-to-introduction (a positive effect on the hazard rate), but also that this effect declines over time. Martin reasons that as the end of the parliamentary term draws closer, "it is less likely that *any* type of legislation will be able to make it all the way through the legislative process, 'attractive' or otherwise." (2004, 455, emphasis in original). Accordingly, *Government Issue Saliency* is interacted with the log of the time remaining in the parliamentary term (*ln[CIEP]*).

Martin estimates a Cox model with the PLM method, which, according to the CVMF test, is the better-fitting estimator at a statistically significant level ($p < 0.05$). Thus, we confirm the findings he reports. Nonetheless, we present the differences between the two techniques to highlight the test's usefulness. Panel (a) of Figure A.2 plots standardized coefficients and 95% confidence inter-

vals from the PLM and IRR estimates and panel (b) illustrates the changing effect of *Government Issue Saliency* across the time remaining in the parliamentary term (*ln[CIEP]*).[3]

[Insert Figure A.2 here]

The coefficient plot in panel (a) only shows the effects of the key variables at one point in time (no time remaining in the term), and so it is primarily useful for gauging an initial sense of the variance associated with each estimate. In this regard, note that the PLM confidence intervals are much smaller than those of IRR for all six variables shown. This is consequential for hypothesis testing at the 95% level on *Opposition Issue Divisiveness*, for which the PLM estimate is significant but the IRR estimate is not. Moving to the effect of saliency over time, note that both methods produce a significant negative coefficient on *Government Issue Saliency*, but a positive coefficient on *Government Issue Saliency* $\times$ *ln(CIEP)*, suggesting that the effect of saliency becomes positive when there is more time remaining in the term.

This effect is shown in detail in panel (b). That plot shows the percentage change in the hazard rate for a one standard deviation increase in *Government Issue Saliency* for a bill of average divisiveness across the range of time left in the parliamentary term. Consistent with expectations, the PLM estimate (gray line) shows a positive change in the hazard rate when there is between 1,400 and about 600 days left, and a negative effect after 600 days (though between about 800 and 400 days the effect is not significant). This shows support for the theory: when there is sufficient time remaining, an increase in saliency to the coalition partners increases the chance of a bill being introduced, all else equal. But as the term draws to a close, the effect weakens, and eventually becomes significantly negative, supporting "the idea that coalition members are less concerned about introducing important legislation late in the parliamentary term, when it is less likely that any bill will make its way through the entire policymaking process" (Martin 2004, 457).

In contrast, note that with the IRR estimates, the effect is weaker, is only positive for the first 400 days, and is never positive and statistically significant. In short, with IRR, there is much less support for the expectation. This illustrates the usefulness of the CVMF test. As stated previously,

[3]Policy area and country fixed effects are also included in the specification, but not shown.

the test selects PLM at a statistically significant level, which provides evidence in favor of using that method instead of IRR. Thus, a more complete analysis of these data would result in the same conclusions, but with stronger justification for the chosen modeling technique.

3.2 Golder (2010)

Golder (2010) picks up where the work of Diermeier and van Roozendaal (1998) and Martin and Vanberg (2003) leaves off in the study of Western European government formation duration. Rather than focusing exclusively on uncertainty (Diermeier and van Roozendaal 1998) or on the additive effects of uncertainty and complexity (Martin and Vanberg 2003), she considers the possibility of an interactive effect between the two. Using a new data set that includes 17 democracies from Western Europe from 1944–1998, Golder (2010) hypothesizes that bargaining complexity "should lead to increasing delays in the government formation process as uncertainty increases" (12). As in the other two studies, Golder measures uncertainty with an indicator for whether bargaining occurred *Postelection* (high uncertainty) or during the interelection period (low uncertainty) and measures complexity as the effective number of *Legislative Parties* and *Ideological Polarization* between the parties.

Like the previous studies, Golder (2010) estimates a Cox model with the PLM method. The dependent variable is the number of bargaining days to government formation. According to the CVMF test, PLM is the better-performing estimator at a statistically significant level ($p < 0.05$). Thus, as in the Martin (2004) example we confirm the findings Golder (2010) reports. Note that this is a change from the Martin and Vanberg (2003) model, in which IRR was chosen. This suggests that the interactive effect included Golder (2010) may represent a crucial omitted variable in the Diermeier and van Roozendaal (1998) and Martin and Vanberg (2003) versions of the model, which led to the bias-reducing selection of IRR. Once that variable is included, however, the Golder (2010) model can benefit from the added efficiency of PLM.

We again present the differences between the two techniques to highlight the CVMF test's usefulness. Panel (a) of Figure A.3 plots standardized coefficients and 95% confidence intervals from the PLM and IRR estimates and panel (b) illustrates the interactive effect of *Legislative*

Parties with the level of uncertainty.

[Insert Figure A.3 here]

Note first from panel (a) that Golder's main theoretical expectation is supported with the PLM results. Specifically, the coefficients on *Legislative Parties* $\times$ *Postelection* and *Ideological Polarization* $\times$ *Postelection* are negative and statistically significant at the 95% level. All else equal, the effect of increasing complexity (more parties or more ideological distance between parties) exerts a stronger negative effect on the hazard rate of government formation during a postelection period (more uncertainty) than during an interelection period (less uncertainty). However, notice that those same coefficients are both nonsignificant and smaller in magnitude when estimated with IRR. In that case, the effects of *Legislative Parties* and *Ideological Polarization* during a postelection period cannot be statistically distinguished from the effects during an interelection period.

Panel (b) illustrates this in more detail. That plot shows the percentage change in the hazard rate for a one standard deviation increase in *Legislative Parties* for both levels of uncertainty. Consistent with expectations, the negative effect on the hazard rate is stronger in magnitude during the postelection period than during the interelection period. In addition, the confidence intervals show that the difference between those two periods is just on the edge of statistical significance with PLM (t = 1.95). In contrast, that difference is not statistically significant when IRR is used (t = 1.19). In short, this is another example of divergence between PLM and IRR. However, recall that the CVMF test selects PLM at a statistically significant level, which provides evidence in favor of using PLM instead of IRR. Thus, as in the Martin (2004) model, a more complete analysis of these data would ultimately result in the same conclusions, but with stronger justification for the chosen modeling technique.

References

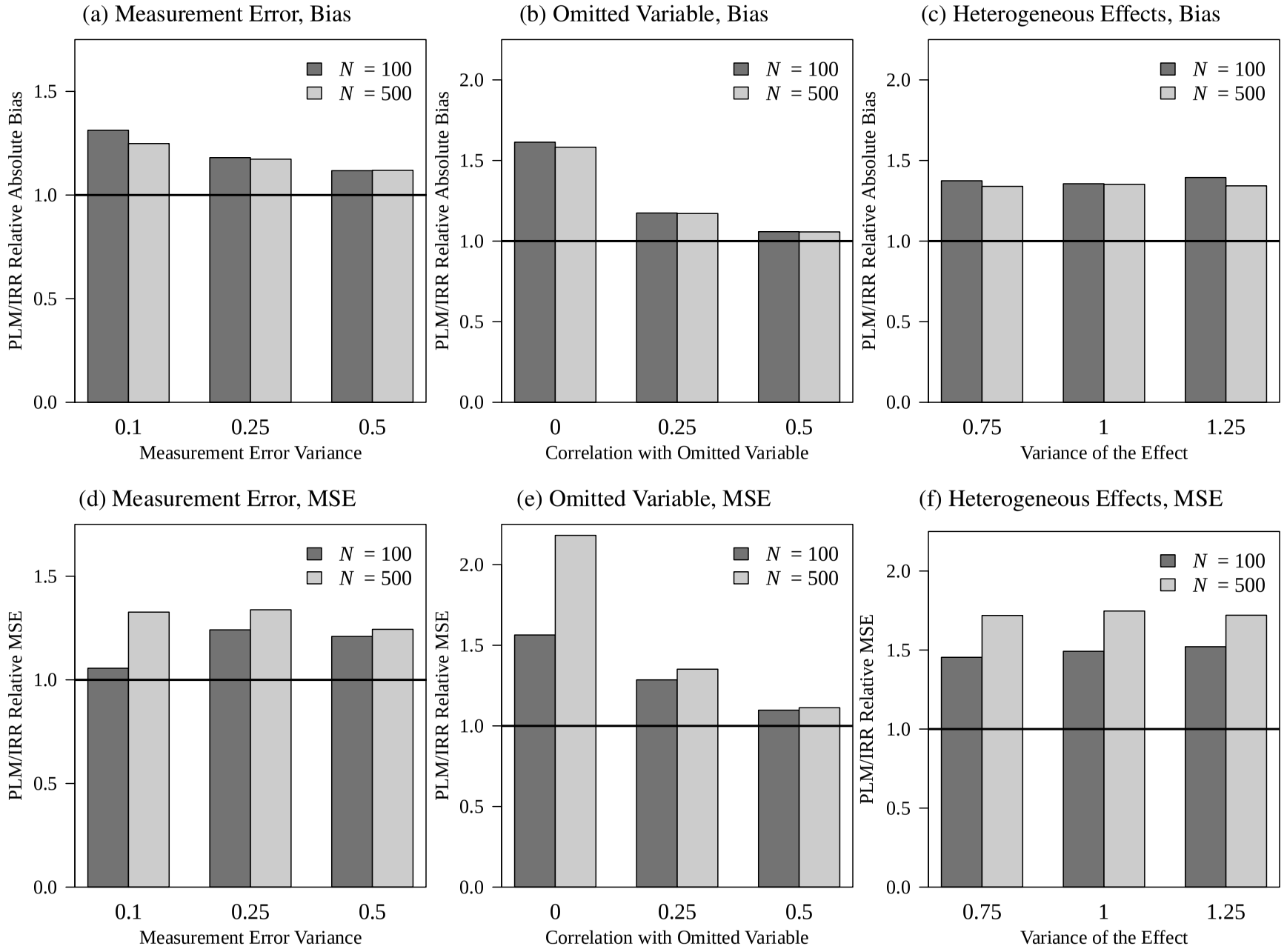
Diermeier, Daniel, and Peter van Roozendaal. 1998. "The Duration of Cabinet Formation Processes in Western Multi-Party Democracies." *British Journal of Political Science* 28(4): 609–626.

Golder, Sona N. 2010. "Bargaining Delays in the Government Formation Process." *Comparative Political Studies* 43(1): 3–32.

Laver, Michael, and W. Ben Hunt. 1992. *Policy and Party Competition*. Ann Arbor, MI: University of Michigan Press.

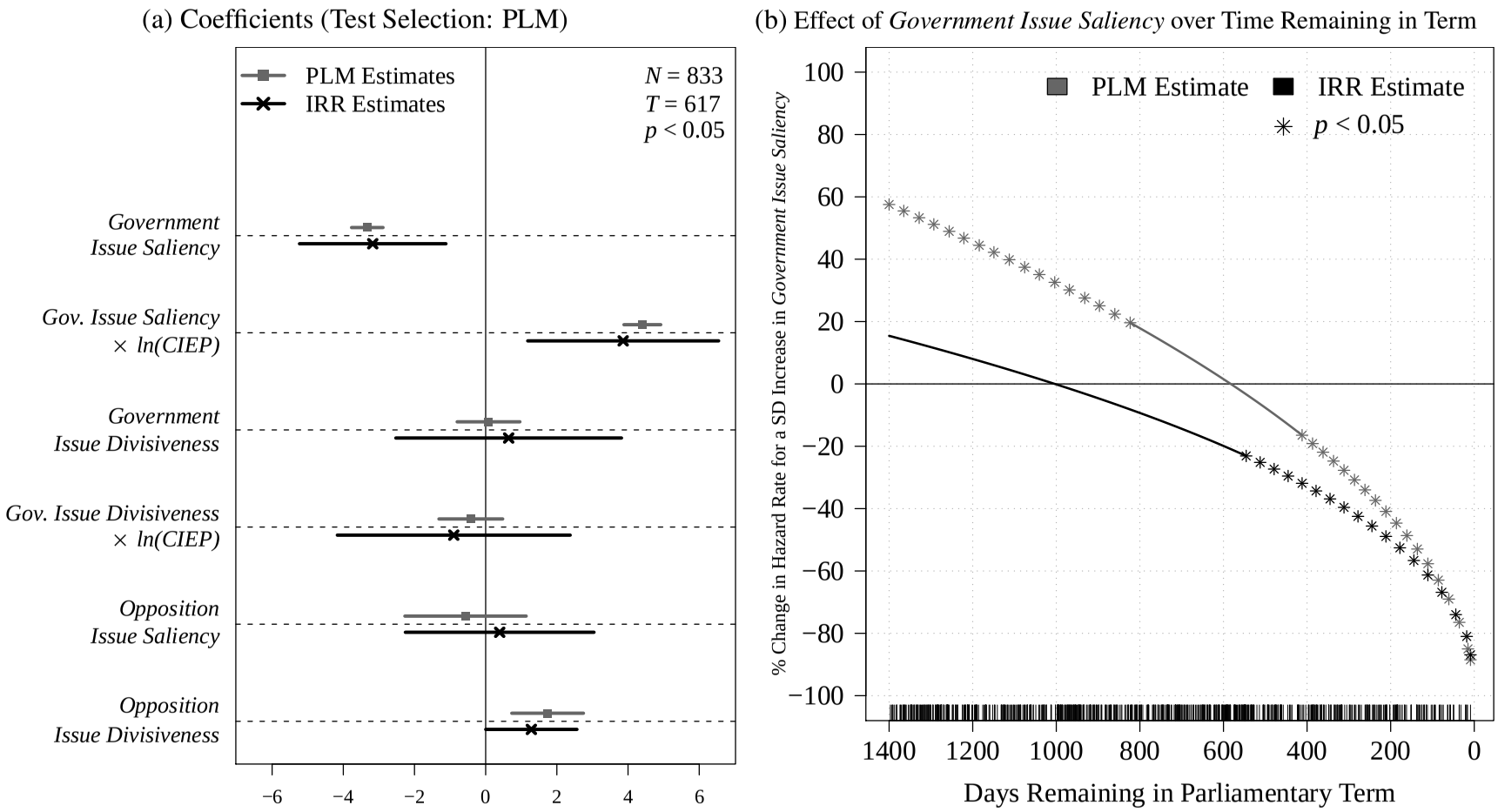
Martin, Lanny W. 2004. "The Government Agenda in Parliamentary Democracies." *American Journal of Political Science* 48(3): 445–461.

Martin, Lanny W., and Georg Vanberg. 2003. "Wasting Time? The Impact of Ideology and Size on Delay in Coalition Formation." *British Journal of Political Science* 33(2): 323–344.



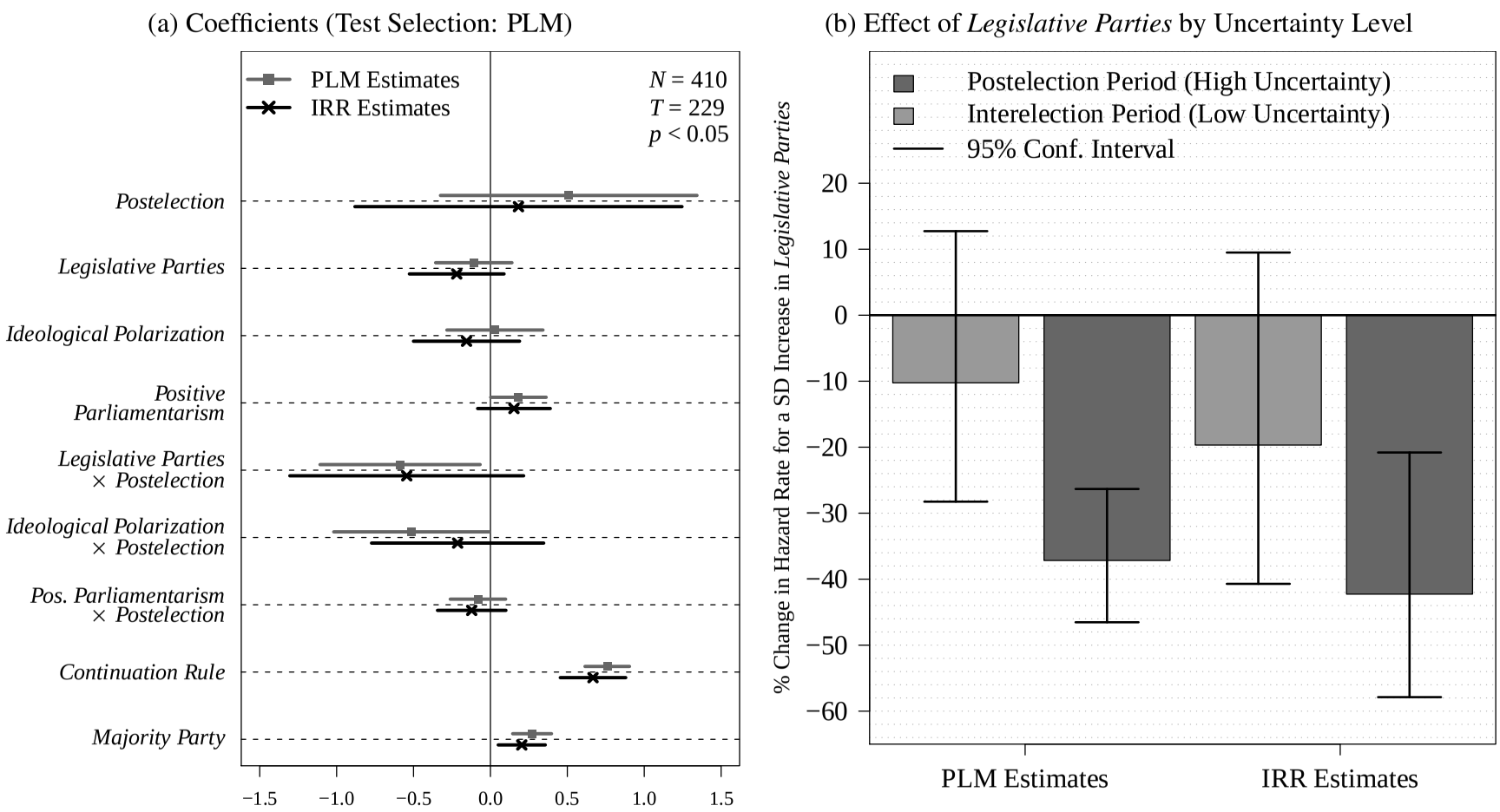
Note: The graphs in panels (a)-(c) present the absolute value of the bias in the PLM estimates divided by the absolute value of the bias with IRR. The graphs in panels (d)-(f) present the mean squared error (MSE) of the PLM estimates divided by the IRR MSE. In all six graphs, values greater than one indicate that IRR performance is better (smaller absolute bias or smaller MSE) than that of PLM. Overall, these graphs show that considerable gains in estimator performance are possible with IRR when the assumptions underlying the Cox model are violated through specification problems.

Figure A.1: Relative Absolute Bias and Relative MSE of PLM and IRR in the Monte Carlo Simulations



Note: The graph in panel (a) reports standardized coefficient estimates and 95% confidence intervals from the PLM and IRR methods. T *is a count of the number of observations for which the PLM cross-validated log-partial-likelihood value is greater than that of IRR. In this case, the PLM values are greater for 617 of 833 observations, which corresponds to a selection of PLM as the better-fitting method (*p < 0.05*). Notice that the confidence intervals are considerably wider when the IRR method is used. This results in a statistically significant estimate for* Opposition Issue Divisiveness *with PLM, but not IRR. Panel (b) plots the percentage change in the hazard rate for a standard deviation increase in* Government Issue Saliency *across the length of the parliamentary term. Note that, as Martin (2004) expects, the effect is positive and statistically significant early in the term with the PLM estimate (gray line), then declines over time. However, this expectation is* not *supported with IRR because the effect is weaker in magnitude and never positive and significant with the IRR estimate (black line).*

Figure A.2: Re-analysis of the Timing of Government Bills (Martin 2004, Table 1)



Note: The graph in panel (a) reports standardized coefficient estimates and 95% confidence intervals from the PLM and IRR methods. T *is a count of the number of observations for which the PLM cross-validated log-partial-likelihood value is greater than that of IRR. In this case, the PLM values are greater for 229 of 410 observations, which corresponds to a selection of PLM as the better-fitting method* ($p < 0.05$). *Notice that, in line with the theoretical expectations of Golder (2010), the coefficients on* Legislative Parties × Postelection *and* Ideological Polarization × Postelection *are negative and statistically significant with the PLM method, but drop in magnitude and become nonsignificant with IRR. Panel (b) plots the PLM (left) and IRR (right) estimates of the percentage change in the hazard rate for a standard deviation increase in* Legislative Parties *for both levels of uncertainty: postelection period (dark gray) and interelection period (light gray). Note that, as Golder (2010) expects, the negative effect on the hazard rate is stronger in magnitude during the postelection period than during the interelection period. In addition, the confidence intervals show that the difference between those two periods is just on the edge of statistical significance with PLM ($t = 1.95$). In contrast, that difference is not statistically significant when IRR is used ($t = 1.19$).*

Figure A.3: Re-analysis of Determinants of the Duration of Government Bargaining Delays (Golder 2010, Table 2)