Online Supplementary Appendix for "More Bang for the Buck: Campaign Spending and Fundraising Success"

Testing for Endogeneity Bias

We performed two types of endogeneity tests. First we performed a Davidson and MacKinnon robust modification of the Hausman test. We instrumented our organization and travel and event spending variables as a function of the log of average spending in each of these categories five and six weeks previously. The first stage regression results showed that at least one instrument was a significant predictor of each spending measure, indicating they were informative instruments. The second stage regression estimates, presented in Table 1, show that our now instrumented measures maintain their positive signs, in fact both have larger coefficients. Despite the noisy instrumentation, travel and event spending retains a significant positive estimate on future contributions, whereas organization spending's effect increases but also shows a much larger standard error. The direction of these estimates gets larger and not smaller, which is the first sign that endogeneity is not driving our results. But, more specifically, the Davidson-MacKinnon test of the relationship between the first stage residuals and the dependent variable is insignificant. Our sample test statistic of 1.374 is less than the critical value of F of 3.73 and has a p-value of .255, such that we cannot reject the null hypothesis of exogeneity.

An alternative test for time series variables like these is a very strict exogeneity test outlined by Jeffrey Wooldridge. To test the extent to which spending is a function of contributions received we can include future values of spending measures within our regression explaining contribution receipts. To the extent spending is driven by recent contributions, then its future values should exhibit a significant positive relationship. In this case we include the logged average of organization and travel and event spending over the future two weeks. As shown in the second column of the Table 1, our estimates for the other variables do not show any clear change and each variable using future values has a positive, but insignificant relationship with contribution amounts. The joint F-test is 2.23 ($p > .1$), such that we fail to reject the null hypothesis of strict exogeneity.

To be sure, the Wooldridge test results are not evidence that spending is not at all a function of future money received, but that the reverse relationship takes longer to occur. It is perhaps likely that the bias is evident when using quarterly measures, but our weekly spending measures do not appear to be a function of immediate or recent contributions within the shorter time period measures we use. In sum, across two different exogeneity tests we find no significant evidence that endogeneity bias is driving our results.

Table 1: Examining Potential Endogeneity in Table 2's Results

Variable	Model Davidson/MacKinnon IV Test	Wooldridge Test
Staff	0.578	0.396*
	(.617)	(0.135)
Travel/Events	0.969*	0.302*
	(.562)	(0.068)
Direct Marketing	-0.255	-0.106
	(.159)	(.043)
Future Staff		0.075
		(.050)
Future Travel/Events		0.056
		(0.047)
Viability	0.016	0.020*
	(.010)	(.005)
Week to 2nd Deadline	-0.058*	-0.090*
	(.029)	(.013)
Week to 3rd Deadline	-0.140*	-0.165*
	(.018)	(.012)
Week to 4th Deadline	-0.096*	-0.112*
	(.015)	(.012)
Week of Deadline	1.425*	1.058*
	(.255)	(.159)
Intercept	-1.954*	4.543*
	(7.299)	(1.325)
F-test (H_0: exogeneity)	1.374	2.23
p-value	0.255	0.109
N	335	320
Candidates	9	9
R^2	0.417	0.627

Dependent variable: Logged weekly big donor contributions (April - December, 2007). Fixed-effects regression coefficients with standard errors in parenthesis; * - indicates $p \leq .05$, one-tailed test. Estimates for the intercept represent the average value of the candidate-specific effect.

Potential Autocorrelation

Two candidates exhibit some residual autocorrelation (for only the first lag) and fail to pass the Q-test of the residuals being white noise: Huckabee and McCain. This presents a common dilemma with dynamic panel data analysis. Although there is some form of autocorrelation, by specifying the lagged AR model we are placing an autocorrelation adjustment on our coefficient estimates that only exists for two of our nine candidates. As a result, we believe the model that does not correct for autocorrelation is a truer representation for seven of our nine candidates. As Table 2 shows below, the results adjusting for autocorrelation are substantively equivalent to the results we present in Table 2 of the article.

Table 2: Alternative Table 2 Estimates, AR1 process

Variable	Overall	Model Frontrunners	Long Shots
Staff	0.326*	0.224	.781*
	(0.142)	(0.145)	(0.285)
Travel/Events	0.159*	0.209*	-0.006
	(0.078)	(0.083)	(0.136)
Direct Marketing	-0.049	0.014	-0.126
	(.056)	(0.064)	(0.089)
Viability	0.026*	0.0134*	0.061*
	(.007)	(0.007)	(0.014)
Week to 2nd Deadline	-0.073*	-0.066*	-0.113*
	(.019)	(0.017)	(0.039)
Week to 3rd Deadline	-0.163*	-0.164*	-0.163*
	(.015)	(0.014)	(0.030)
Week to 4th Deadline	-0.103*	-0.126*	-0.037
	(.014)	(0.014)	(0.025)
Week of Deadline	1.153*	1.277*	0.773*
	(.144)	(0.149)	(0.290)
Intercept	7.682*	8.087*	4.521*
	(.965)	(1.221)	(2.003)
$\hat{\rho}_{AR1}$	.360	.263	.220
N	338	228	101
Candidates	9	6	3
R^2	0.613	0.685	0.601

Dependent variable: Logged weekly big donor contributions (April - December, 2007). Fixed-effects regression coefficients with standard errors in parenthesis; * - indicates $p \leq .05$, two-tailed test. Estimates for the intercept represent the average value of the candidate-specific effect.

Explaining Differences in the National Polling and Futures Market Results

It is likely the case that the differences in the polling model and futures market model results are a function of missing data in the national polling measures, not that one measure performs better or represents something different than the other.

When including the national polling variable a quarter of the observations are not included in the national polling estimates. This reduction in sample size is especially costly for fixed effects estimates because those additional time periods add to the consistency of the estimates for the candidate-specific constants. As we show below, when we only include a measure of viability for the periods in which there is non-missing national polling data, the results are equivalent; i.e., there is no change in the significance of the variable regardless of the viability measure used. The national polling data results differ due to missingness, and therefore we argue that it is better to use the futures markets measure than the polling measure. Table 3 presents the results when using our market "viability" measure for both the overall and the frontrunner model over this same sample (we ignore the long shot model in our interpretation since there are no differences). A couple things are of note here:

1. The estimate of spending effects in Table 3 replicate those results we present in the article's appendix using national polling measures, where staff spending now has a positive significant effect for frontrunners. The model fit is also essentially the same. So, in general, the differences in results are not a function of our measure but of sample selection.

2. In comparison to the national polling results, our viability measure is still a significant predictor for the frontrunner model, suggesting it is a sharper measure of the relationship between campaign performance and campaign contributions.

Therefore, the significant finding for frontrunner organization spending is not a function of which measure we use, but a loss in observations that is a product of sparse national polling data in the early stages of the primary. Since the finding is not consistent when including more observations or across other specifications we do not think the change in the significance of the staff variable merits a discussion in the manuscript, it is not a finding that we can confidently generalize from since it is not found when using more observations (Table 2 in the article) or when excluding national polling or national viability measures entirely (the first 3 columns in the table in the Appendix). In all, it appears to be a single significant result that is not replicated when using different specifications or adding observations.

Table 3: Results with Futures Market Variable when National Polling Measure is Observed

| | Model | |
Variable	Overall	Frontrunners
Staff	0.611*	0.491*
	(0.159)	(0.170)
Travel/Events	0.272*	0.265*
	(0.079)	(0.088)
Direct Marketing	-0.100*	-0.066
	(.050)	(0.066)
Viability	0.026*	0.016*
	(.006)	(0.007)
Week to 2nd Deadline	0.039	0.072
	(.062)	(0.063)
Week to 3rd Deadline	-0.141*	-0.148*
	(.012)	(0.012)
Week to 4th Deadline	-0.085*	-0.112*
	(.012)	(0.012)
Week of Deadline	1.295*	1.411*
	(.157)	(0.162)
Intercept	3.361*	4.732*
	(1.553)	(1.845)
N	266	180
Candidates	9	6
R^2	0.659	0.731

Dependent variable: Logged weekly big donor contributions (June - December, 2007). Fixed-effects regression coefficients with standard errors in parenthesis; * - indicates $p \leq .05$, two-tailed test. Estimates for the intercept represent the average value of the candidate-specific effect.