THIS APPENDIX WAS REFEREED AS PART OF THE PUBLISHED VERSION OF THE "HUNT FOR PARTY DISCIPLINE" BY *THE AMERICAN POLITICAL SCIENCE REVIEW*. IT WAS OMITTED FROM THE PUBLISHED VERSION ONLY TO SAVE SPACE. SNYDER AND GROSECLOSE DO NOT DISCUSS THIS APPENDIX IN THEIR REPLY.

Appendix B: Monte Carlo Analysis of the Snyder-Groseclose Approach

To demonstrate that the Snyder-Groseclose method is likely to reject the null hypothesis of preference-based voting when it is true, we conduct a number of Monte Carlo experiments. The Monte Carlo data are generated by one-dimensional spatial voting with error. Snyder and Groseclose use the scaling method of Heckman and Snyder (1997). Our specification of the underlying random-utility model is therefore identical to that assumed by Heckman and Snyder. Legislator i votes **Y**ea rather than **N**ay if and only if

$$-\left(x_i - z_{yj}\right)^2 + \varepsilon_{ijy} \geq -\left(x_i - z_{nj}\right)^2 + \varepsilon_{ijn} \qquad \text{(A3)}$$

where x_i is the ideal point of legislator i, z_{yj} and z_{nj} are the positions of the Yea and Nay voting alternatives, and the ε are random shocks. Let $z_{Mj} = (z_{yj} + z_{nj})/2$ be the midpoint of the roll call and $d_j = (z_{yj} - z_{nj})/2$ be half the (directional) distance between the yea and nay outcomes. To simulate realistic values for the yea and nay positions, we assume that z_{Mj} is distributed on [-1,1] according to the density $f(z) = 1-|z|$ which produces a modal voting margin of 50%-50%. Further we assume that d_j is distributed uniformly on $[-1,-.05] \cup [.05,1]$. The "gap" from -.05 to .05 prevents votes in which the yea and nay outcomes are too similar so that voting is purely random.

We divide the 435 members of our House of Representatives into 218 Democrats and 217 Republicans. The ideal points of the Democrats are distributed uniformly across the interval $[-1,r]$ and those of the Republicans are distributed across $[-r,1]$. (See columns **(a)**, **(b)** of table A1.) The variable r controls the extent of overlap in the ideal points of members of the two parties. If $r = 0$, then the parties are perfectly spatially separated. If $r = 1$, the parties are drawn independently from the same distribution. In general, in expectation, a fraction $2r/(1+r)$ of each party overlaps with the other party. In our experiments, we let $r \in \{0.1, 0.2, 0.3\}$ so that the corresponding correlations between party and preferences take on the values of -0.82, -0.76, and -0.68. These are consistent with measures of party overlap and correlation in the post-war House of Representatives.

Also following Heckman and Snyder, we assume that

$$\eta_{ij} \equiv \varepsilon_{ijy} - \varepsilon_{ijn} \tag{A4}$$

is drawn from $U[-m,m]$. We let $m \in \{0.2, 0.4, 0.6\}$. (See column **(c)** of table A1.) These values are chosen to be consistent with the range of goodness-of-fit measures such as classification success reported in Heckman and Snyder (1997). In the experiments that follow, the correct classification of voting decisions following Heckman-Snyder estimation of the ideal points ranged from 81% to 92%. Finally, the experiments are conducted with 1000 roll calls.[i] This is roughly the number of actual roll calls in recent Houses.[ii]

For each set of experiments, we produced two sets of estimates for ideal points using the Heckman-Snyder scaling method:

1. The Snyder-Groseclose estimates using only roll calls with margins greater than 65-35.

2. “Naïve” estimates using all the votes.[iii]

Note that the Heckman-Snyder method should estimate ideal points very close to the true ideal points when the naïve model is used, because the Monte Carlo experiments generate the artificial data from a preference-based voting model.

Each specification was run ten times, so that 10,000 second-stage regressions were performed for each. In table A1, we present the percentage of times the null hypothesis of no party voting [that is, β_2=0 in equation (1)] was rejected at the 1% level (one-tail) using White’s heteroscedasticity-consistent standard errors. Snyder and Groseclose also used the 1% criterion and White’s standard errors, but whether the tests were one-tailed or two-tailed is unclear. That is, because preference-based voting generated the data, we expect to find a “significant” β_2 in only 1% of the simulated roll call votes. The actual results are strikingly different.

The extent of over-rejection for *close roll calls* (column **(d)**) is enormous. Under the most favorable conditions, shown in the last three rows of table A1 — large party overlap — the Snyder-Groseclose model rejects the null at approximately the expected 1% rate. However, in the least favorable conditions[iv] — less overlap and precise voting — shown in the first row, the over-rejection rate is 73.1%.

The example shown in figure 2 indicates that the naïve method should lead to lower levels of rejection than the Snyder-Groseclose method because the better estimation of ideal points using all votes will leave less room for the party dummy to act as a proxy for the ideal points. This intuition is borne out. In all of the 9 matches of cells in table A1, the rejection rate for *close* roll calls (column **(f)**) is lower using the naïve method than using the Snyder-Groseclose method.[v] In the intermediate cases of rows 4-6 where Snyder-Groseclose rejects over 5 times the expected rate, the naïve method rejects at just about the expected rate.

One explanation for the Snyder-Groseclose bias on close roll calls is that the ideal points are recovered incorrectly as we argued with figure 2. Figure 2 was based on voting without errors. Errors in voting are not sufficient, even with large numbers of roll calls, to permit accurate recovery of legislator positions. Compare columns **(i)** and **(j)** of table A1. The correlations for the middle sixth of the legislature are systematically less using only close votes to estimate the ideal points than using all votes. That is, the effect we illustrated with figure 2 occurs even when both error is present and the number of roll calls is very large.

In many of our simulations the Heckman-Snyder estimates contradict the assumption of the underlying linear probability model. In particular, many of the voting probabilities lie outside the [0,1] interval. We indicate the percentage of these probabilities in column **(h)** of Table A.1. While it is true that the over-rejection rate is increasing in the number of improper probabilities, the variation in these proportions is too small to generate the large variation in the over-rejection rates. Furthermore, the percentage of extreme probabilities is approximately what one finds in applications of the Snyder-Groseclose method to actual roll calls from the House of Representatives.

At this point, the reader may have noticed an apparent anomaly. Under the naïve model, we should expect to get about 1% of the coefficients significant at the 1% level. The results are not too far off both for lopsided votes (column **(g)**) and for close votes (column **(f)**) where there is considerable party overlap. On the other hand, there are far too many significant coefficients for other close votes, particularly those in the first rows of the tables, where there is little overlap and only a small amount of randomness in voting.

There is an additional anomaly in our Monte Carlo experiments. For lopsided votes, both the näive and Snyder-Groseclose methods produce a large number of statistically significant coefficients at the one-tailed 1% level, but with the wrong sign.[vi] Table A2 presents the percentage of "wrong" coefficients for the experiments on 1000 vote legislatures. Note that the problem is worst for the naïve model with little party overlap and lopsided votes.

Both of these anomalies arise because the Snyder-Groseclose second stage provides biased estimates of the party effect, *even when the ideal points have been correctly estimated in the first stage*. The intuition for both anomalies is provided by considering the case of a uniform distribution of ideal points on [-1,+1] with $r = 0$, no overlap. Moreover, assume, errorless voting, that is, $m = 0$. (And continue to assume no party pressure.)

Consider midpoints c in the interval [-1,+1]. A straightforward calculation shows that the coefficient on the party dummy is given by $\beta_2 = -1 + 4|c| - 3c^2$. Thus, β_2 is -1 for $c = 0$, the quintessential close 50-50 vote opposing Ds and Rs. From equation (2), the estimate of the extent of party discipline is $\gamma = -1/0 = -\infty$. On the other hand we get a wrong sign with $\beta_2 = 1/4$ for $c = 1/2$, that is, for a lopsided 75-25 split. More generally, the party coefficient is of the wrong sign for party-pressure voting when $1 > |c| > 1/3$. Although the coefficient should always be zero for preference-based voting, the coefficient is 0 only when the magnitude of c is exactly 1/3. When the magnitude of c is 1/3, we would get a 67-33 split. Like Snyder and Groseclose, we chose 65-35, very close to 67-33, to differentiate lopsided from close votes. The results in the tables, "correct" signs for close votes when the true coefficient should be zero and "wrong" signs for lopsided votes, conform to this theoretical analysis.

The theoretical example can be extended to allow for both overlap in party ideal points and for errors in voting. We focus on c=0, or predicted 50-50 splits, since this is the situation where Snyder and Groseclose expect the greatest party pressure. We begin by showing that allowing for overlap does not eliminate bias.

Introduce overlap in the party positions as follows. Let the left-most 25% of the legislature, those with ideal points in [–1, –

½) be Democrats, the next 25% in [–½,0) be Republicans, then another 25%, in [0, ½), be Democrats and the rightmost 25%, in [½,1] be Republicans. On average, the Democrats are still the left, with a mean position of –¼ and the Republicans are at ¼. This is more overlap than appears in any Congress in the last two decades.

For this overlap case, the coefficient on the dummy is +6/13, showing an incorrect sign when voting is purely preference based.

In the no overlap example, the coefficient on the dummy was –1, indicating strong party pressure when there was none. Obviously, as the overlap increases the coefficient on the party dummy increases. There is some amount of overlap that will make the coefficient on the dummy 0, but this would be knife-edge.

Does error save the day? Yes and No. To see this, let the probability of voting Yes be linear in the ideal point between –w, +w, with

$$\text{Prob(Yes vote}|x \leq -w) = 0$$
$$\text{Prob(Yes vote}|-w < x < w) = .5 + x/(2w)$$
$$\text{Prob(Yes vote}|x \geq w) = 1.$$

Again assume a uniform distribution of (or equally spaced) ideal points and a legislature that is 50% D.

After calculating the appropriate variances and covariances and then plugging into the standard formula for the regression coefficient with two independent variables, we can compute values for the coefficient in both the overlap and no overlap cases. The results appear in Table A3.

In the no overlap case, we have a "correct" sign when the coefficient should be zero. In the overlap case, we have a "wrong" sign. The bias falls as the amount of error increases. For w of 0.8 or 0.9, which correspond to the error levels likely to occur with actual data, the bias is quite small. However, since the expected value is not zero, there still should be a disproportionate number of "significant" coefficients in reasonably large samples—such as the US House.

The no overlap case is more disturbing, since the bias does not fall as fast. With *w* of 0.8, the coefficient is -.04, which corresponds, in the example, to 2% of the legislators being switched by non-existent pressures. Thus, for no overlap or very low levels of overlap, one is quite likely to incorrectly conclude that there is some pressure when none exists.

More generally, the expected value of the party dummy coefficient, for a fixed non-random distribution of true ideal points and party affiliations, is a linear combination of the expected value of two covariances

$$E(\beta) = aE(Cov(vote, dummy)) + bE(Cov(vote, x))$$

The linear coefficients depend on the variances of the ideal points and the party dummy and their covariance. The expected covariances depend on the error process, the distributions of the dummy and the ideal points, and the true cutting line for the roll call. Therefore, the sign and magnitude of the party dummy will depend in a complex way on both the distribution of ideal points and the distribution of errors. Only in special cases will the coefficient on the dummy have an expectation of zero when voting is based solely on spatial preferences and stochastic errors.

To illustrate how the patterns uncovered in the Monte Carlo experiments reappear in actual voting data, we replicate the analysis of Snyder and Groseclose for a number of Congresses.[vii] Table A4 contains those results. In addition to the reported percentages of significant party coefficients, we also report the percentage of "wrong signs". Note that the pattern of the Monte Carlo experiments is echoed in the actual data. The number of "correct" significant coefficients is consistently higher for the Snyder-Groseclose model than the naïve model, and the number of "wrong" coefficients is higher for the naïve model. The differences are most striking with respect to correct signs on close votes. Parallel to the Monte Carlo work, the Snyder-Groseclose method produces many more significant instances of party discipline than does the naïve method.

In conclusion, our results demonstrate that the Snyder-Groseclose technique is heavily biased toward rejecting the null hypothesis of preference-based voting. Even if Snyder and Groseclose were able to estimate ideal points correctly in the first stage, they would get too many "significant" coefficients with the correct sign in the second stage on close votes and too many with the wrong sign on lopsided votes. The bias arises because they use OLS in the second stage. The bias is attenuated and becomes unimportant when there is a high degree of party overlap. Even when there is substantial party overlap, however, the Snyder-Groseclose method is biased toward finding too many "significant" coefficients because the first stage provides biased estimates of ideal points.

Table A1. Preference-Based Monte Carlo Data for Legislatures with 435 Legislators, 1000 Roll Calls, 10 Replications

Distribution of Ideal Points		Voting Error	Ideal Points Estimated by Heckman-Snyder Method Applied to:							Optimal Classification
			Close Votes Only (Snyder/GrosecloseMethod)		All Votes (Naïve Method)		Close Votes	Lopsided Votes	All Votes	
			Percentage of 10000 Roll Calls With Party Pressure Effect Significant at 1% Level				%Vote Probs. 0utside [0,1]	Correlation of True and Estimated Ideal Points, Middle Sixth		% Gain in Correct Classification, Two Cutpoint Model Over One Cutpoint Model
Rep.	Dem.		Close Votes	Lopsided Votes	Close Votes	Lopsided Votes				
(a)	**(b)**	**(c)**	**(d)**	**(e)**	**(f)**	**(g)**	**(h)**	**(i)**	**(j)**	**(k)**
U[-.1, 1]	U[-1, .1]	U[-.2,.2]	73.1	0.2	19.1	0.8	24	0.66	0.97	.10
U[-.1, 1]	U[-1, .1]	U[-.4,.4]	47.8	0.5	9.4	1.8	21	0.80	0.94	.16
U[-.1, 1]	U[-1, .1]	U[-.2,.6]	23.4	0.6	4.7	1.5	17	0.82	0.90	.23
U[-.2,1]	U[-1,.2]	U[-.2,.2]	34.0	0.4	0.7	1.2	24	0.79	0.98	.16
U[-.2,1]	U[-1,.2]	U[-.4,.4]	13.1	0.6	0.9	0.9	20	0.86	0.95	.24
U[-.2,1]	U[-1,.2]	U[-.6,.6]	5.8	0.9	1.3	1.5	16	0.88	0.92	.30
U[-.3,1]	U[-1,.3]	U[-.2,.2]	0.6	0.7	0.5	15.7	23	0.82	0.98	.21
U[-.3,1]	U[-1,.3]	U[-.4,.4]	0.7	0.9	0.5	3.2	19	0.89	0.96	.30
U[-.3,1]	U[-1,.3]	U[-.6,.6]	1.2	0.9	0.9	1.3	16	0.87	0.93	.36

Table A2. “Wrong Sign” Party Coefficients from Application of the Snyder-Groseclose Second Stage to Preference-Based Monte Carlo Data

Distribution of Ideal Points		Voting Error	Percentage of 10000 Roll Calls With “Wrong” Party Pressure Effect Significant at 1% Level			
			Snyder/Groseclose Method in First Stage		Naïve Method in First Stage	
Republicans	Democrats		Close Votes	Lopsided Votes	Close Votes	Lopsided Votes
U[-.1, 1]	U[-1, .1]	U[-.2,.2]	13.2	38.1	0.0	56.1
U[-.1, 1]	U[-1, .1]	U[-.4,.4]	7.2	30.5	0.1	42.8
U[-.1, 1]	U[-1, .1]	U[-.2,.6]	3.7	21.6	0.3	29.1
U[-.2,1]	U[-1,.2]	U[-.2,.2]	7.1	33.0	0.4	10.6
U[-.2,1]	U[-1,.2]	U[-.4,.4]	2.4	20.6	0.2	10.7
U[-.2,1]	U[-1,.2]	U[-.6,.6]	1.2	10.8	0.5	7.5
U[-.3,1]	U[-1,.3]	U[-.2,.2]	0.8	14.1	9.8	0.1
U[-.3,1]	U[-1,.3]	U[-.4,.4]	0.6	4.3	1.4	0.2
U[-.3,1]	U[-1,.3]	U[-.6,.6]	0.9	1.7	0.9	0.8

Table A.3 Value of Party Dummy Coefficient (β_2) When There is No Party Pressure

	Party Coefficient	
w	Overlap	No Overlap
0.0	0.4615	-1.0000
0.1	0.3985	-0.8100
0.2	0.3323	-0.6400
0.3	0.2631	-0.4900
0.4	0.1908	-0.3600
0.5	0.1155	-0.2500
0.6	0.0574	-0.1600
0.7	0.0257	-0.0900
0.8	0.0092	-0.0400
0.9	0.0019	-0.0100
1.0	0.0000	0.0000

Table A4. Replication of Snyder-Groseclose Using Six Dimensional Model

House	*# Roll Calls*	*#Close Votes*	*Proportion Sig. at 1%, Correct Sign, Close Votes*	*Proportion Sig. at 1%, Correct Sign, Lopsided Votes*	*Proportion Sig. at 1%, Wrong Sign, Close Votes*	*Proportion Sig. at 1%, Wrong Sign, Lopsided Votes*
85th (Naïve)	175	99	.101	.013	.121	.158
85th (S-G)	175	99	.394	.039	.141	.144
90th (Naïve)	409	158	.203	.060	.127	.068
90th (S-G)	409	158	.601	.056	.025	.143
95th (Naïve)	1348	649	.068	.064	.085	.043
95th (S-G)	1348	649	.317	.019	.045	.081
100th(Naïve)	803	335	.143	.049	.197	.246
100th (S-G)	803	335	.394	.077	.107	.098
105th(Naïve)	550	275	.000	.015	.021	.113
105th (S-G)	550	275	.767	.127	.054	.189
Total (Naïve)	3285	1516	.088	.050	.105	.116
Total (S-G)	3285	1516	.450	.057	.065	.114

[i] Legislatures with 500 roll calls were also examined, but the results are very similar. These may be found in an earlier version of this paper available at http://porkrind.pols.columbia.edu/discip.pdf.

[ii] The setup of these Monte Carlo experiments is very similar to the Monte Carlo experiments reported in the published version of Snyder and Groseclose (2000) which were conducted in response to our original working paper. The major difference is that Snyder and Groseclose impose restrictions on the distribution of voting error to eliminate probabilities outside the [0,1] interval in their linear probability setup. This arbitrary assumption makes their results a best case for their model by assuming away one of the sources of mistaken inferences that we identify below.

[iii] Like Heckman and Snyder (1997), we excluded all votes with less than 1% voting on the minority side.

[iv] McCarty, Poole, and Rosenthal (1997) present evidence that these conditions are likely to prevail in recent Congresses.

[v] One might argue that in the case of large party overlap, the naïve model somewhat under rejects the null.

[vi] That is, the party coefficient implies that Democrats are under pressure to vote in the conservative way. We coded the coefficients “wrong” if β_1 and β_2 had the same sign (recall preferences are scaled so that conservatives score higher). In the case of quadratic preferences, this is equivalent to a finding of $\gamma > 0$. We found that explicitly testing the hypothesis $\gamma < 0$ produced results substantively similar to coding the expected sign of β_2 based on the sign of β_1.

[vii] Our replications appear to match their results with a few caveats. First, Snyder and Groseclose present their results in a line-graph so verifying an exact match is impossible. Second, they do not indicate how wrong -signed coefficients were treated or how many tails were used in their hypothesis tests. Finally, another potential factor for discrepancy is that they do not indicate the dimensionality of the preference model they used on each roll call. They outline a Monte Carlo procedure for determining the right number of dimensions to retain and indicate they retained a “few more” that this number. Rather than replicate this analysis (which would be imperfect due to the use of a Monte Carlo test statistic), we included six dimensions which is approximately the average used by Snyder and Groseclose.