Can Ideal Point Estimates be Used as Explanatory Variables?*

Andrew D. Martin
Washington University
admartin@wustl.edu

Kevin M. Quinn
Harvard University
kevin_quinn@harvard.edu

October 8, 2005

1 Introduction

For many years scholars have fit measurement models to voting data to recover the latent ideal points of various actors. Poole and Rosenthal (1997), for example, provide a number of different measurement strategies for House members and Senators; Clinton et al. (2004) offer a Bayesian alternative. Martin and Quinn (2002) fit a dynamic item response theory model which provides time-varying ideal points for Supreme Court justices. Can these estimated ideal points be used as explanatory variables in subsequent (oftentimes called second-stage) regression models? In this note we answer this question. Our discussion focuses primarily on the Martin and Quinn (2002) scores for Supreme Court justices, but the theoretical arguments are equally applicable to other modeling strategies. We begin by discussing the Martin-Quinn approach, and presenting the scores. We then outline possible concerns about using the measures, followed by our thoughts about those concerns. We conclude with a set of best practices for the use of Martin-Quinn scores.

2 The Martin-Quinn Scores

Martin and Quinn (2002) posit a measurement model with two estimands: the ideal points of the justices, and two case-specific parameters. The model is unique in that the ideal points of the justices are allowed to vary smoothly across time. To identify the model, some justices are fixed in their first terms of service. That defines the scale with which the other justices' ideal points are measured. The model is estimated using Markov chain Monte Carlo methods. While the case-specific parameters are of interest in certain applications, here we focus on the estimated ideal points. These ideal points are updated annually as the Court decides additional cases, and are made available at `http://adm.wustl.edu/supct.php`. Currently scores are available from the October 1937 term to the October 2003 term. The website provides posterior means; i.e., the ideal points, for each justice in each term in which they served. The website also contains posterior

---

standard deviations, which quantify the uncertainty about each of the measures. We will discuss the use of these below.

In Figures 1-5 we plot the ideal points for the justices. Each figure shares the same y-axis to allow for across-time comparison. Lower numbers on the ideological scale representing liberalism (left); high numbers on the ideological scale represent conservatism (right). It is also important to keep in mind that each of these measures is an estimate, each with a (possibly) different amount of uncertainty associated with it. It is important to take into account this uncertainty when asking questions such as: "Is Justice X more conservative than Justice Y?" These questions can be answered using Monte Carlo methods (see Clinton et al., 2004; Martin et al., 2005).

3   Possible Concerns for Subsequent Regressions

There are a number of criticisms of the use of estimated ideal points as explanatory variables in subsequent regression models.

1. There are two types of subsequent regression models; those with votes as the dependent variable, and those with anything else as the dependent variable. In the former case, researchers may be concerned that they are "using votes to explain votes." Epstein and Mershon (1996) describe this problem as: "the measures of the independent and dependent variables are the same." The measurement work of Segal and Cover (1989) was undertaken to avoid this issue.[1] In the latter case, when votes are not the explanatory variable of interest, circularity is not a concern.

2. The second criticism of these measures is *non-random case selection*. It is well-known that the nature of the agenda is important to consider when modeling observational data. Since agenda-setting on the Court is endogenous, might this yield inaccurate preference measures? And, to what extent should the agenda process be included in the measurement model?

3. The final criticism of that Martin and Quinn (2002) scores is *issue boundedness*. While the measures might be appropriate in some issue areas, the uni-dimensional spatial model might not be appropriate for other, more difficult issues, such as tax, economics, judicial power, etc.

In the following section we discuss each of these concerns in order, and provide some evidence, when possible, as to their applicability.

4   Evidence from the Martin-Quinn Scores

4.1   Votes Explaining Votes

The circularity concern is quite important as a purely technical matter. Strictly speaking, the scores should not be used in this context. What modeling approach would be better? One approach would be to use an exogenous measurement strategy, such as the Segal and Cover (1989) scores. This approach would work fine in some issue areas, such as civil rights and civil liberties, but quite poorly

---

[1]Sometimes this criticism is summarized by claiming we should have measures of revealed preferences that are "independent" of the actual votes. Of course, if these measures were truly independent, they would be unrelated to voting and thus of no use.

in other issue areas (Epstein and Mershon, 1996). Using these scores also requires the assumption of fixed preferences over time, which is inappropriate for some justices, such as Justice Blackmun.

Another approach would be to fit a full structural model, where ideal points were simultaneously estimated along with the regression parameters of interest (Clinton et al., 2004). We discuss this in the concluding section of our *Political Analysis* piece. While this is the principled approach to dealing with the problem, it requires writing custom software, and is thus beyond the reach of many applied researchers.

Still another option that is applicable to the study of votes on the merits within a subset of cases is to estimate ideal points using the data from other cases and to use these ideal point estimates in one's regression model of interest. For instance, if one were interested in analyzing votes on the merits in federalism cases, one could estimate ideal points using data from all cases *except* federalism cases and then use these ideal point estimates in the regression model of votes on federalism cases.

We note that if it turns out that if the publicly available ideal point estimates based on all of the data look essentially the same as the ideal point estimates based on subsetting the data in the manner mentioned above, then the second stage regression using the full data Martin and Quinn scores will be essentially the same as the more principled second stage regression that includes the ideal points estimated from a subset of the data. If this is the case, then there is little to be lost from simply using the publicly available Martin and Quinn scores in second stage regressions.

To assess the extent to which this is the case we have re-estimated the dynamic ideal point model (using the same priors) excluding one issue area at a time. We have re-estimated the model excluding the Spaeth VALUE codes: criminal procedure (1), civil rights (2), first amendment (3), due process (4), privacy (5), attorneys (6), unions (7), economic activity (8), judicial power (9), federalism (10), interstate relations (11), and federal taxation (12). These estimates are labeled as "ButX" in the figures, were X refers to the excluded issue area.

In Figure 6 we compare the ranks of the full data Martin-Quinn estimates with those from the models with excluded issue categories; in Figure 7 we do the same thing, this time comparing the actual estimates. What is clear from these figures is that the excluded issue estimates are very highly correlated with the full data Martin-Quinn estimates. In Figure 8 we compare the estimated location of the median justice for each of the models. The overwhelming pattern in these figures is data falling along the forty-five degree line, indicating that this Court-specific measure changes very little when excluding issues one at a time. Since these correlations are so high, as a practical matter using the full data Martin-Quinn scores when modeling votes in a single issue is perfectly appropriate. While circularity is a technical concern, the resultant measures from purging issues will change very little, and so it is not worth the effort to do so. When modeling votes in a single issue area, circularity is not a practical concern. We summarize these prescriptions in the final section.

4.2   Non-Random Case Selection

Does the fact that the Supreme Court controls its own docket affect the ability of the model to reliably recover ideal points? In many applications this is a germane criticism. One good example is the work of Baum (1988), who demonstrates that looking at statistics such as the percent liberalism in the previous term is problematic because of agenda effects (his so-called Baum correction fixes this under a set of assumptions, one of which is constant preferences).

Unlike most regression models, the item response theory (IRT) model that underlies the Martin and Quinn (2002) model is not nearly as sensitive to selection (particularly for ordinal quantities of interest such as the ranks of the ideal points). Indeed, the IRT model does not treat all cases

equally; some are more informative than others (a 5-4 decision carries more information about the ideal points than an 8-1 decision). If certain coalitions were never observed in the data, selection effects might bias ideal point estimates. But empirically that is not a concern with these data. Moreover, if it were a concern, the ideal points would still be appropriately estimated, but the amount of uncertainty would dramatically increase. See Lynch (2005), who explores how the use of interest group-selected roll calls affects ideal point estimates compared with using all roll calls or just randomly selected roll calls. Finally, Jackman (2001) recommends looking at the estimated item parameters to see if there is support along the ideological continuum. In our data, the space is well-supported.

While extreme cases of agenda control can affect ideal point estimates, this is not a concern for the Martin-Quinn scores. If the agenda process itself is of empirical interest, or if the researcher would like to bring information about agenda setting into the statistical models, this can be accommodated in the IRT framework. See, for example, Clinton and Meriowitz (2004) and Martin and Quinn (2001).

4.3   Issue Boundedness

To what extent are Martin-Quinn scores applicable in areas of the law besides civil rights and civil liberties (the domain of Segal-Cover scores)? In Table 1 we present the percent correctly classified across a number of issue areas. While the scores do better in some area of the law than others, these scores classify well across all issues. In short, a uni-dimensional spatial model performs well across most issues. In Figure 11, we fit the dynamic ideal point model to four single issues, and compare the location of the medial justice. The strongest correlation is between the Martin-Quinn medians and the civil liberties medians (0.91); the weakest is between the Martin-Quinn medians and the economics medians (0.68). Again, this suggests that these measures do quite well across issue areas.

How do the measures compare with existing ones? For the sake of comparison, we compute the term-by-term correlations of our ideal point estimates with other available preference measures. We plot these correlations in Figure 9. Two existing measures are based on multi-dimensional scaling of observed votes: those by Schubert (1974) and Rohde and Spaeth (1976). Schubert (1974) finds two primary dimensions that structure the Court: a "C" scale which comprises civil rights and civil liberties, and an "E" scale focused primarily on economics cases. Not surprisingly our measure correlates highly with the "C" scale. The comparison with the "E" scale is more interesting. Our measure is always positively correlated with the "E" scale, very strongly so in the mid-1940s, 1955 to the early 1960s, and the late 1960s. But there are times when the correlation dips below 0.5. Our measure is thus picking up something slightly different from the "E" scale, which is likely attributable to the dynamic structure of our model. Rohde and Spaeth (1976) find three dimensions – "Freedom," "Equality," and "New Deal" – that structure behavior from the mid 1950s to the late 1970s. But for the "Equality" scale in the mid-1950s, our measure is comparable to *all three of these scales*, including the economics-oriented "New Deal" scale. These findings show that the Martin-Quinn scale is strongly related to the (non-orthogonal) dimensions uncovered by other scholars.

We also correlate our measure with the Segal and Cover (1989) measure in the final cell of Figure 10. The results are important. From 1970 to 1990, and *only during this time period*, does our measure correlate strongly with the Segal and Cover (1989) measure. Indeed, through the 1960s, there is essentially a zero correlation. And, the correlation during the 1990s is modest. This suggests a number of things. First, it is interesting to note that our measure only correlates strongly with the Segal and Cover (1989) measure when the Court is heterogeneous. As the Court became

4

more homogenous in the 1960s, and in the early 1990s, the correlation between the two measures dips significantly. Second, since our measures are essentially summaries of past behavior, this calls into question the validity of the Segal and Cover (1989) scores in many areas.

5 Best Practices

We conclude with what we view as best-practices for the use of Martin-Quinn scores in subsequent regression models. First and foremost, we encourage others to use the scores often and creatively![2] If the dependent variable of interest is not voting, then the scores can be used without any concerns. If the dependent variable is votes on the merits, using Martin-Quinn scores is reasonable, even while recognizing the circularity problem, if the subject of the study is a single issue area. While circularity is still technically a problem, the results in this note demonstrate that as a practical matter it is not a significant concern. Finally, if the focus of the study is votes on the merits on all cases, using Martin-Quinn scores is inappropriate, and a full structural estimation is necessary. Using Segal and Cover (1989) scores as a measure of judicial preferences is also a reasonable approach in some circumstances (modeling aggregated votes in civil rights and civil liberties). However, some of the assumptions on which the measures are based, such as constant preferences, are questionable.

References

Baum, Lawrence. 1988. "Measuring Policy Change in the United States Supreme Court." *American Political Science Review* 82(September):905–912.

Clinton, Joshua D., Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98:355–370.

Clinton, Joshua D., and Adam Meriowitz. 2004. "Testing Explanations of Strategic Voting in Legislatures: A Reexamination of the Compromise of 1790." *American Journal of Political Science* 48:675–689.

Epstein, Lee, and Carol Mershon. 1996. "Measuring Political Preferences." *American Journal of Political Science* 40(February):261–294.

Jackman, Simon. 2001. "Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, Inference and Model Checking." *Political Analysis* 9:227–241.

Lynch, Michael S. 2005. "Are They Asking the Right Questions? Assessing Interest Group Scores Using Item Response Theory." Presented at the 2005 Midwest Political Science Association Meeting.

Martin, Andrew D., and Kevin M. Quinn. 2001. "Bayesian Learning about Ideal Points of Supreme Court Justices, 1953-1999." Presented at the 2001 Political Methodology Summer Meeting.

---

[2]It also is good practice to take into account measure uncertainly when using Martin-Quinn scores. The posterior standard deviations are available, and can be used as weights (although in our experience this matters very little to not-at-all for most models). This uncertainty can also be taken into account using various Monte Carlo methods, but they require developing specialized software.

Martin, Andrew D., and Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999." *Political Analysis* 10:134–153.

Martin, Andrew D., Kevin M. Quinn, and Lee Epstein. 2005. "The Median Justice on the United States Supreme Court." *North Carolina Law Review* 83(5):1275–1321.

Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll-Call Voting*. Oxford: Oxford University Press.

Rohde, David W., and Harold J. Spaeth. 1976. *Supreme Court Decision Making*. San Francisco: W. H. Freeman.

Schubert, Glendon. 1974. *The Judicial Mind Revisited: Psychometric Analysis of Supreme Court Ideology*. London: Oxford University Press.

Segal, Jeffrey A., and Albert D. Cover. 1989. "Ideological Values and the Votes of U.S. Supreme Court Justices." *American Political Science Review* 83:557–565.
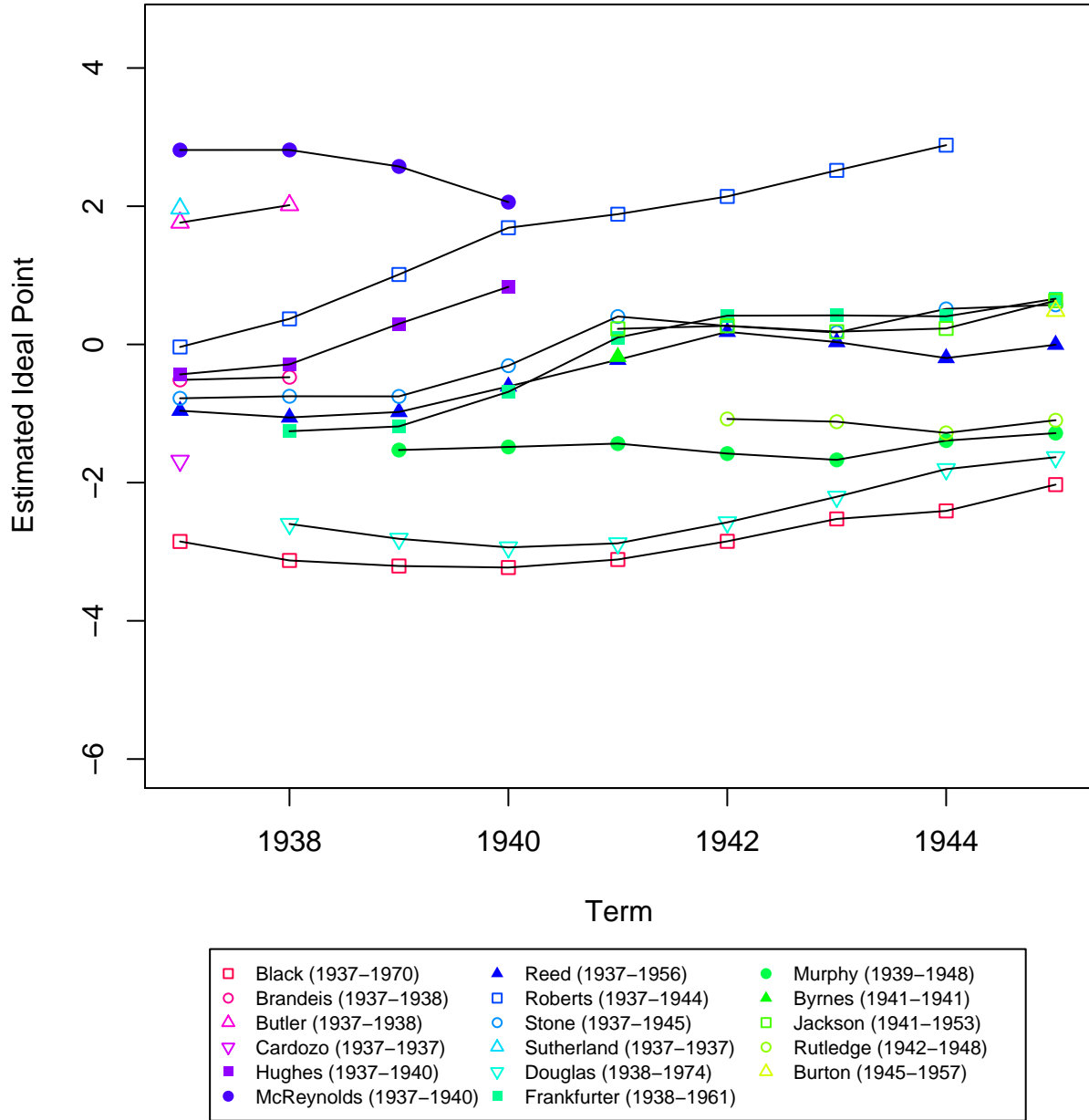
Figure 1: Estimated ideal points for the dynamic ideal point model for the late-Hughes and Stone Courts, 1937-1945.
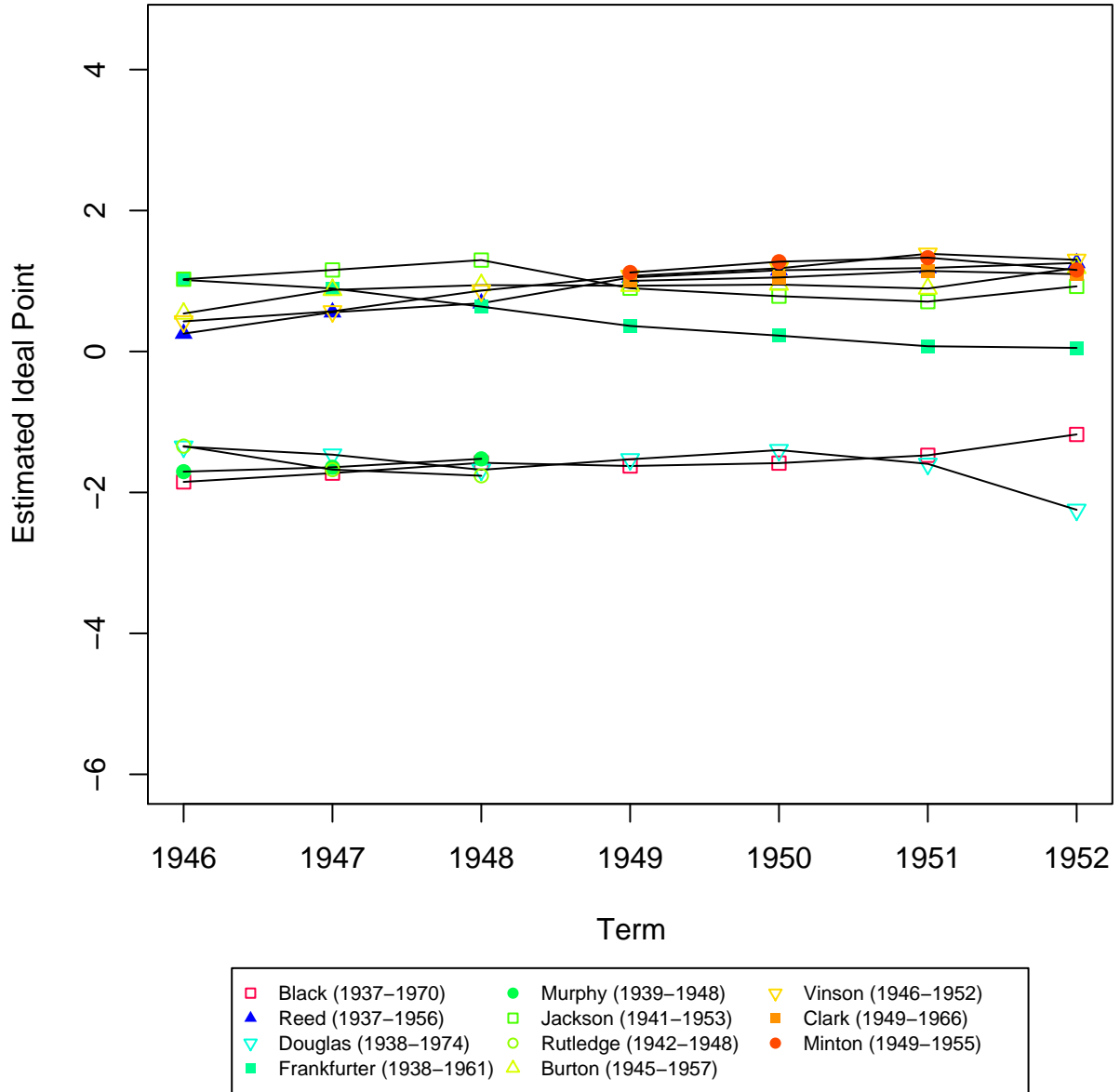
Figure 2: Estimated ideal points for the dynamic ideal point model for the Vinson Court, 1946-1952.
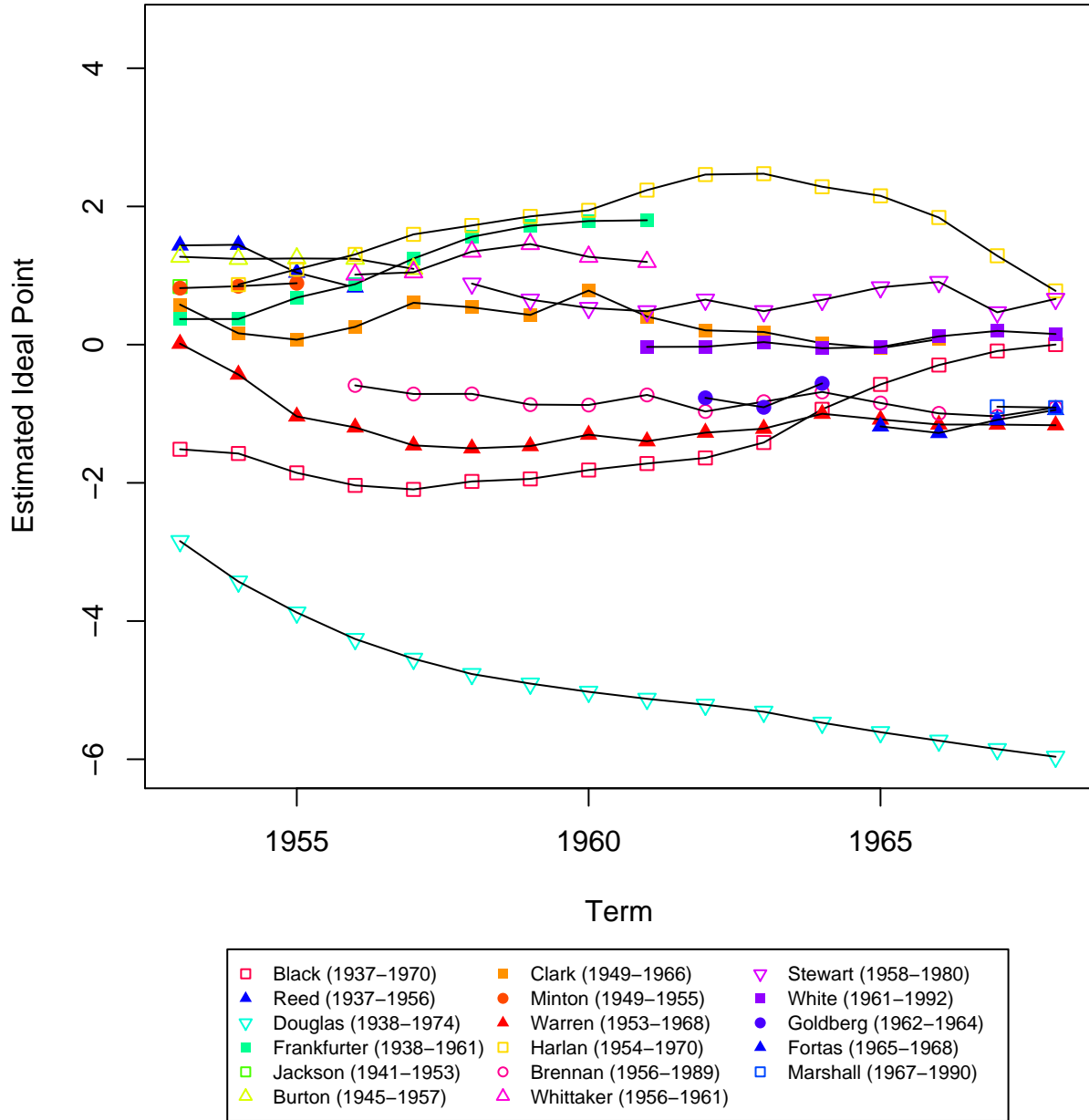
Figure 3: Estimated ideal points for the dynamic ideal point model for the Warren Court, 1953-1968.
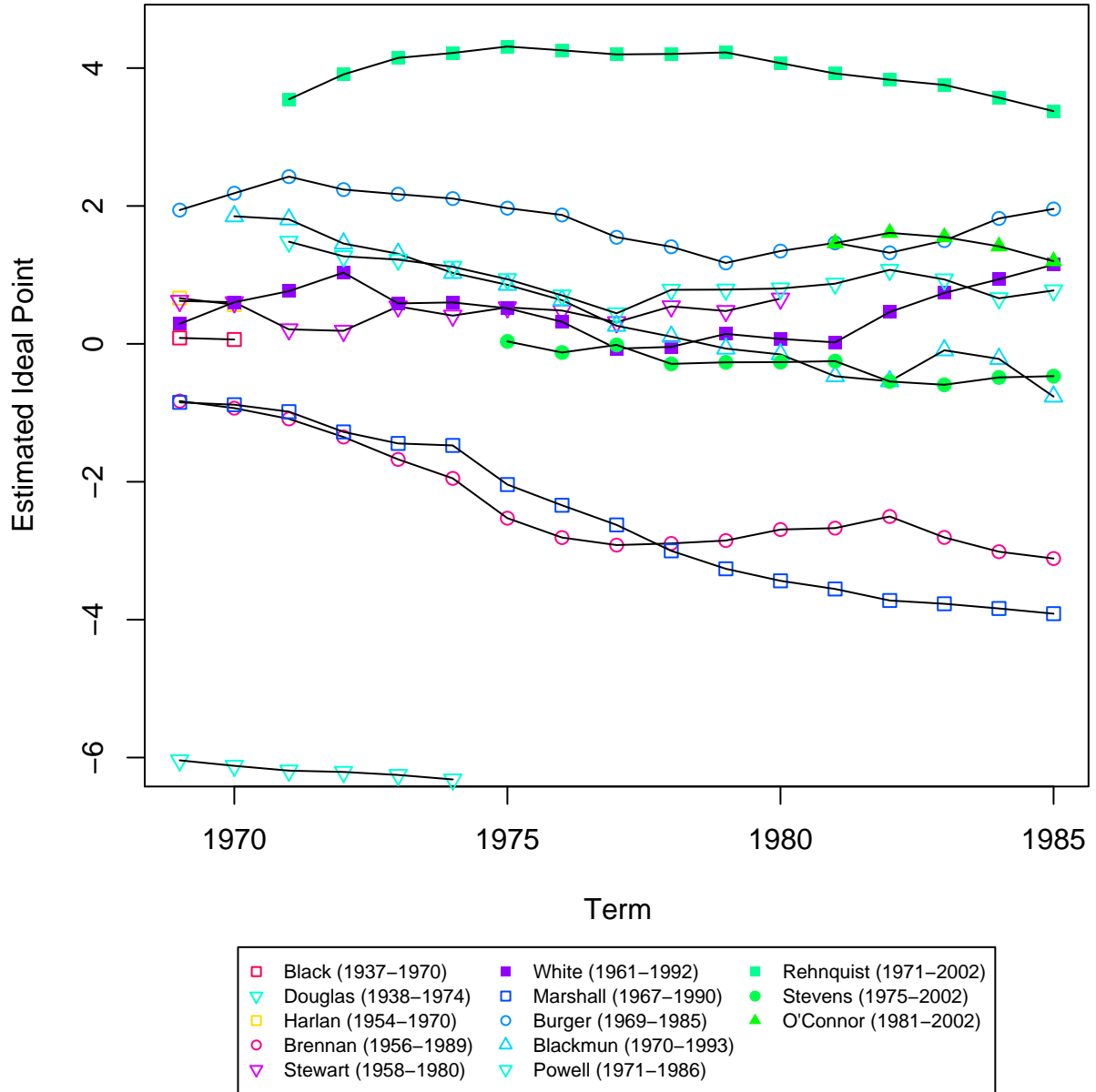
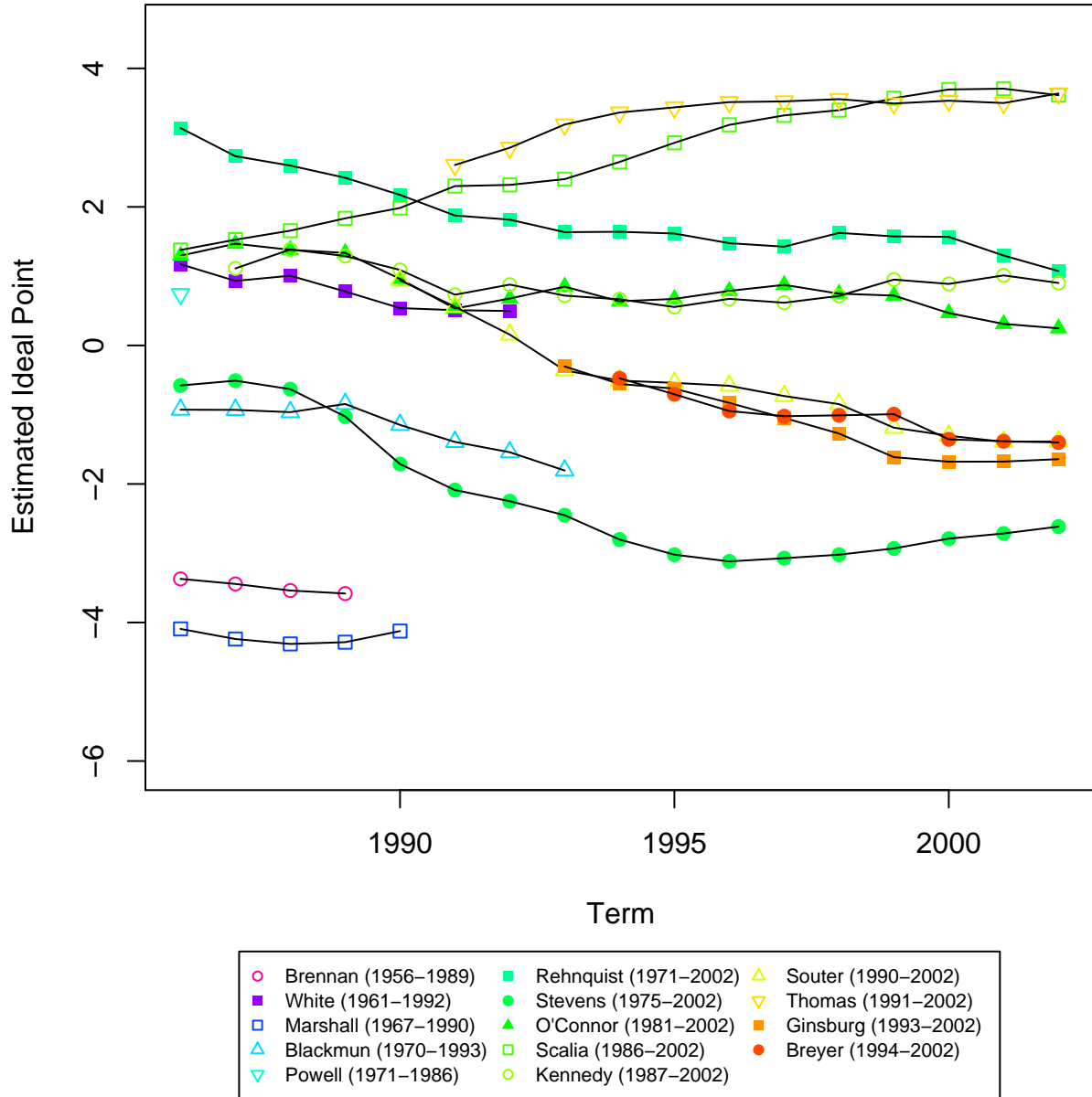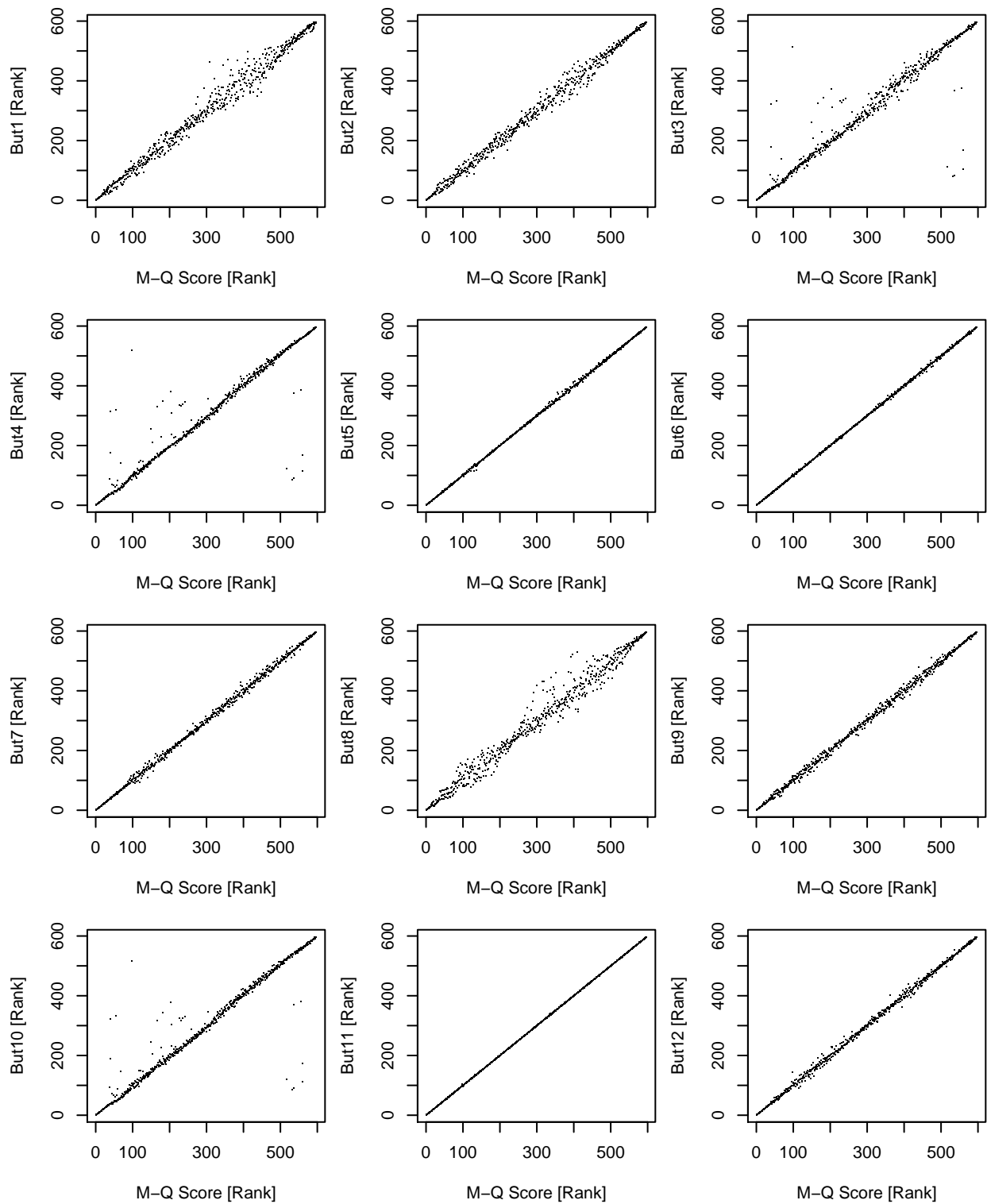Figure 4: Estimated ideal points for the dynamic ideal point model for the Burger Court, 1969-1985.

Figure 5: Estimated ideal points for the dynamic ideal point model for the Rehnquist Court, 1986-2002.

Figure 6: Comparison of estimated ideal point ranks for the dynamic ideal point model with estimates deleting one issue at a time.
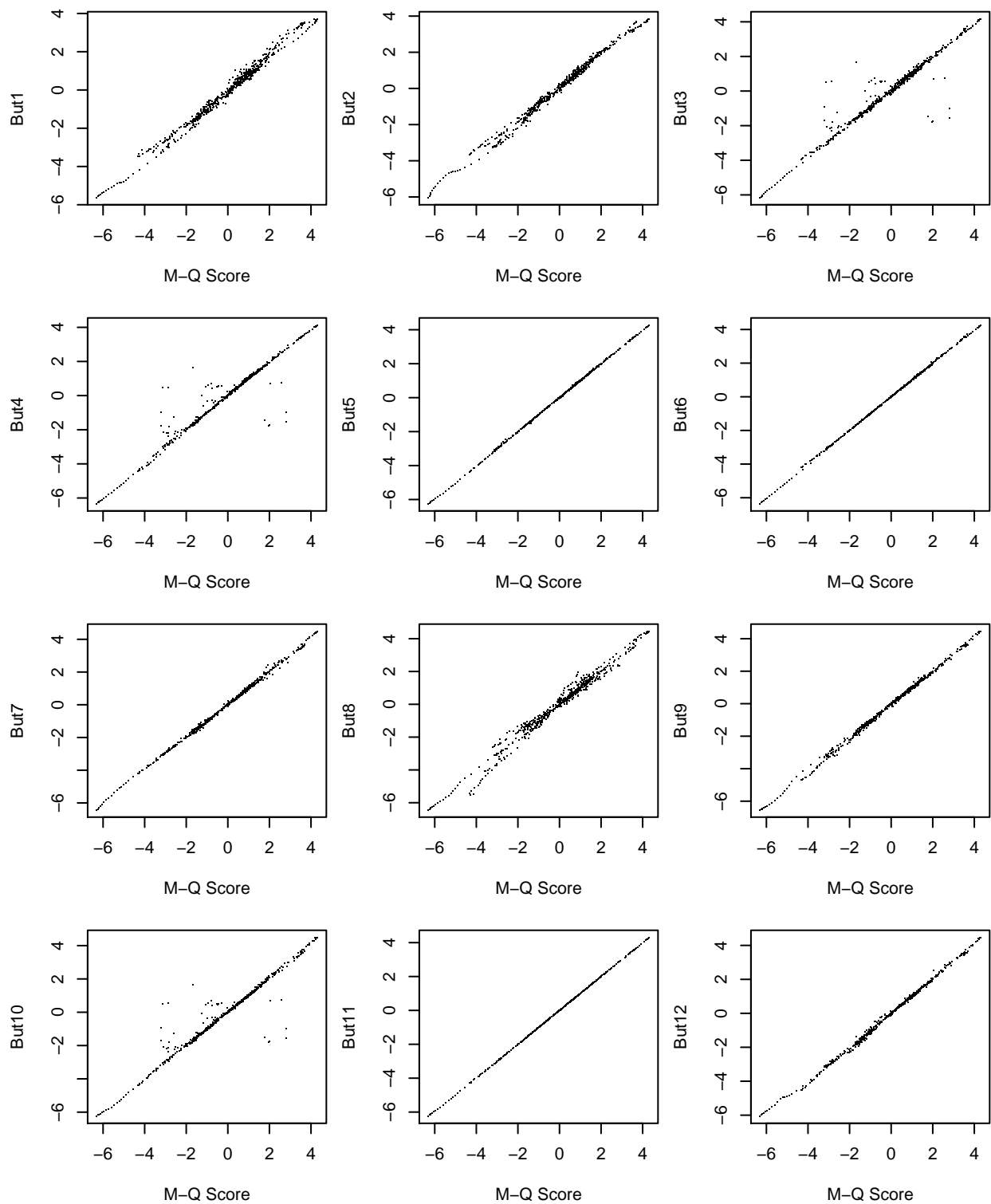
Figure 7: Comparison of estimated ideal points for the dynamic ideal point model with estimates deleting one issue at a time.
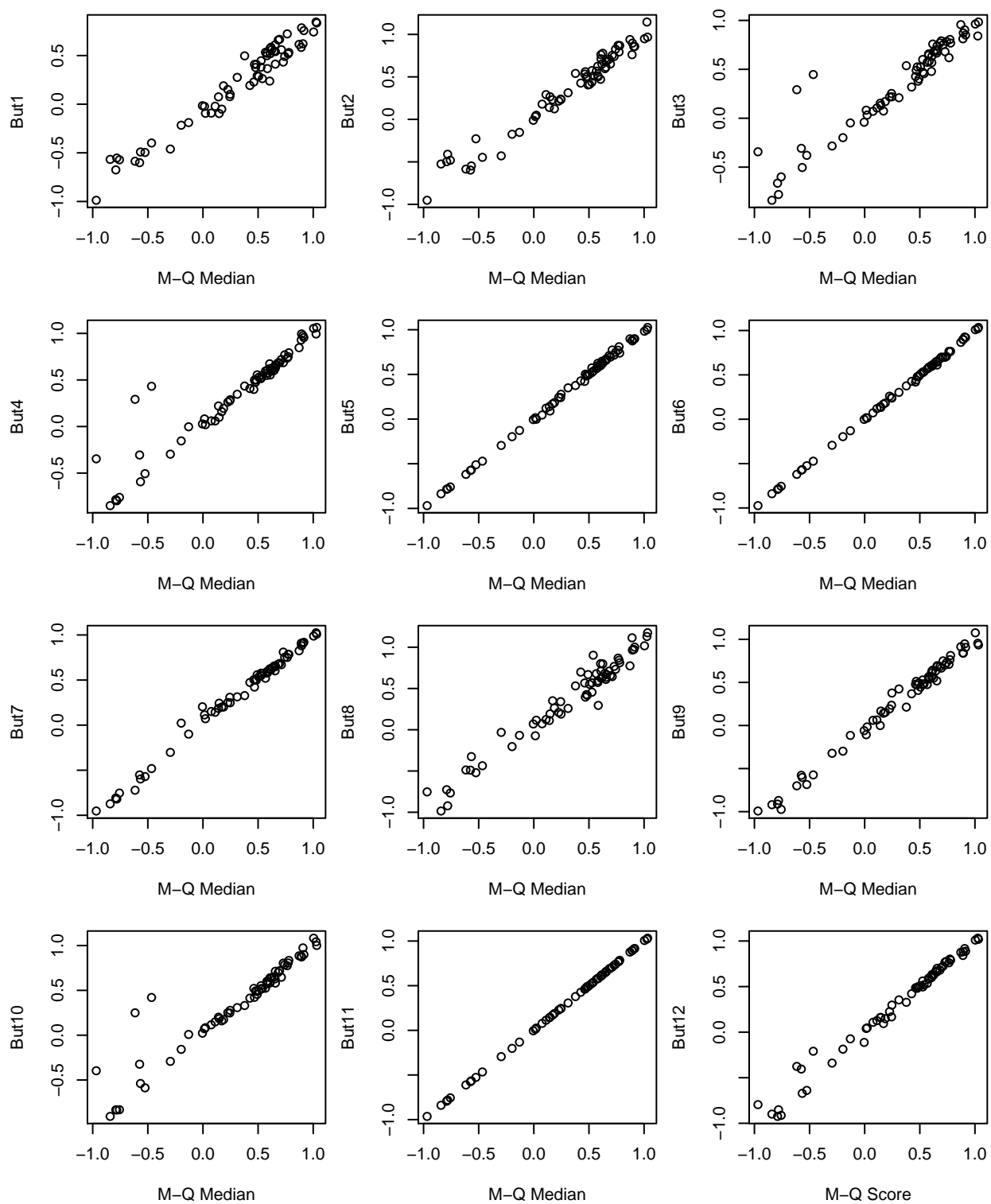
Figure 8: Comparison of estimated Court median for the dynamic ideal point model with estimates deleting one issue at a time.
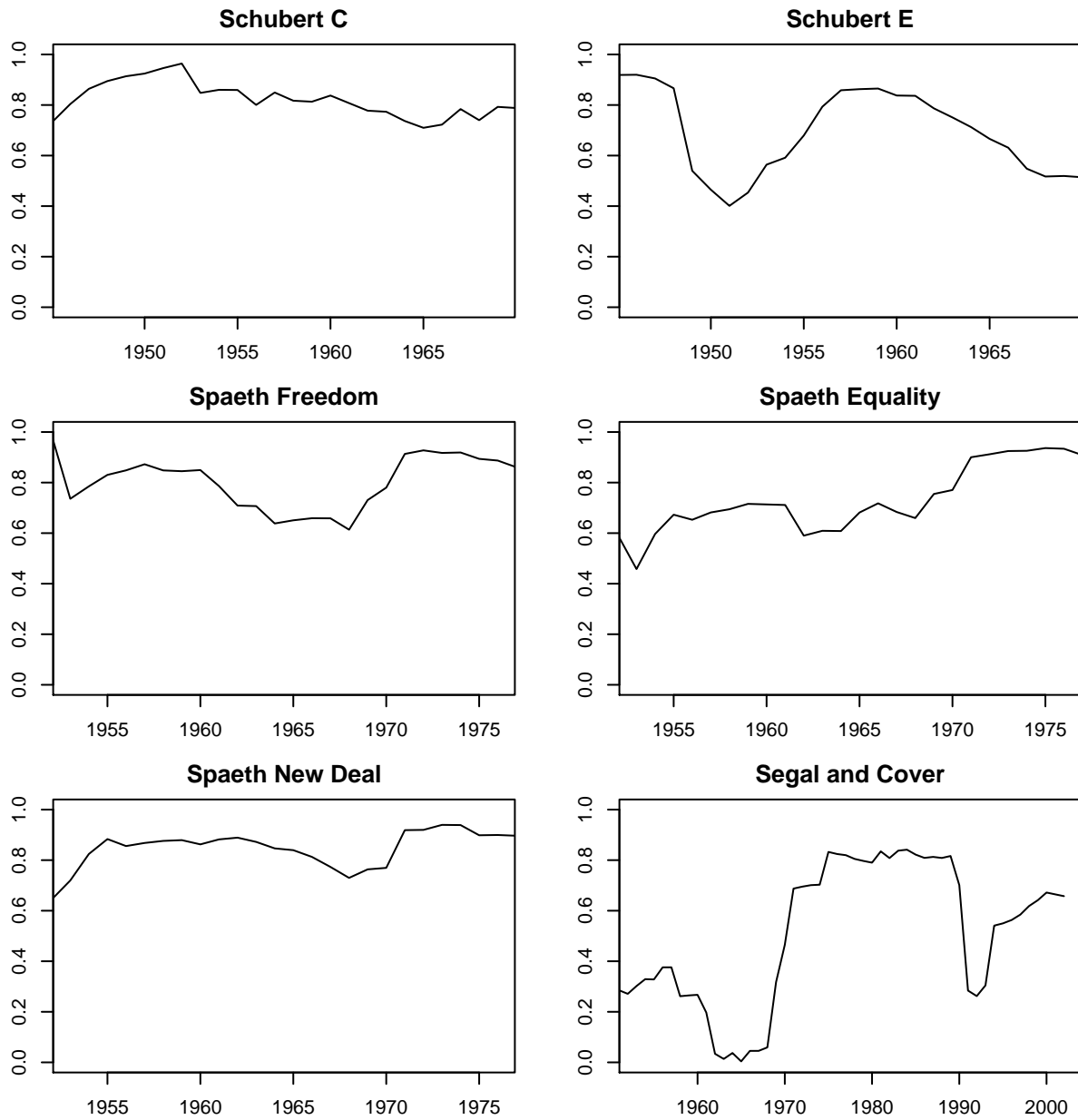
Figure 9: Term-by-term correlations of dynamic ideal point estimates with Schubert (Schubert, 1974), Spaeth (Rohde and Spaeth, 1976), and Segal and Cover (Segal and Cover, 1989) measures.
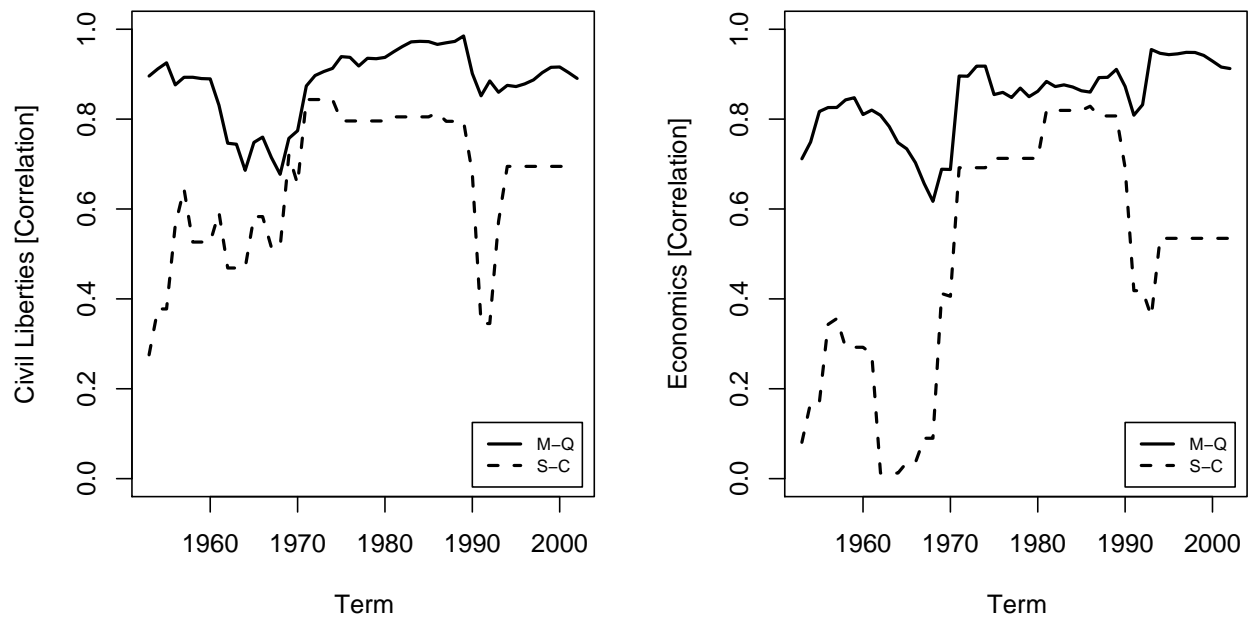
Figure 10: Term-by-term correlations of Martin and Quinn (M-Q) and Segal and Cover (S-C) measures with percent liberal decisions in civil liberties and economics cases.

Issue Area	Percent	Total Votes
Attorneys	78.41	372
Criminal Procedure	77.67	9357
Civil Rights	77.33	6146
First Amendment	76.61	3632
Due Process	75.54	1680
Privacy	75.50	456
Unions	74.48	2243
Economic Activity	74.30	8952
Judicial Power	73.42	4039
Federalism	72.76	1596
Federal Taxation	71.85	2220

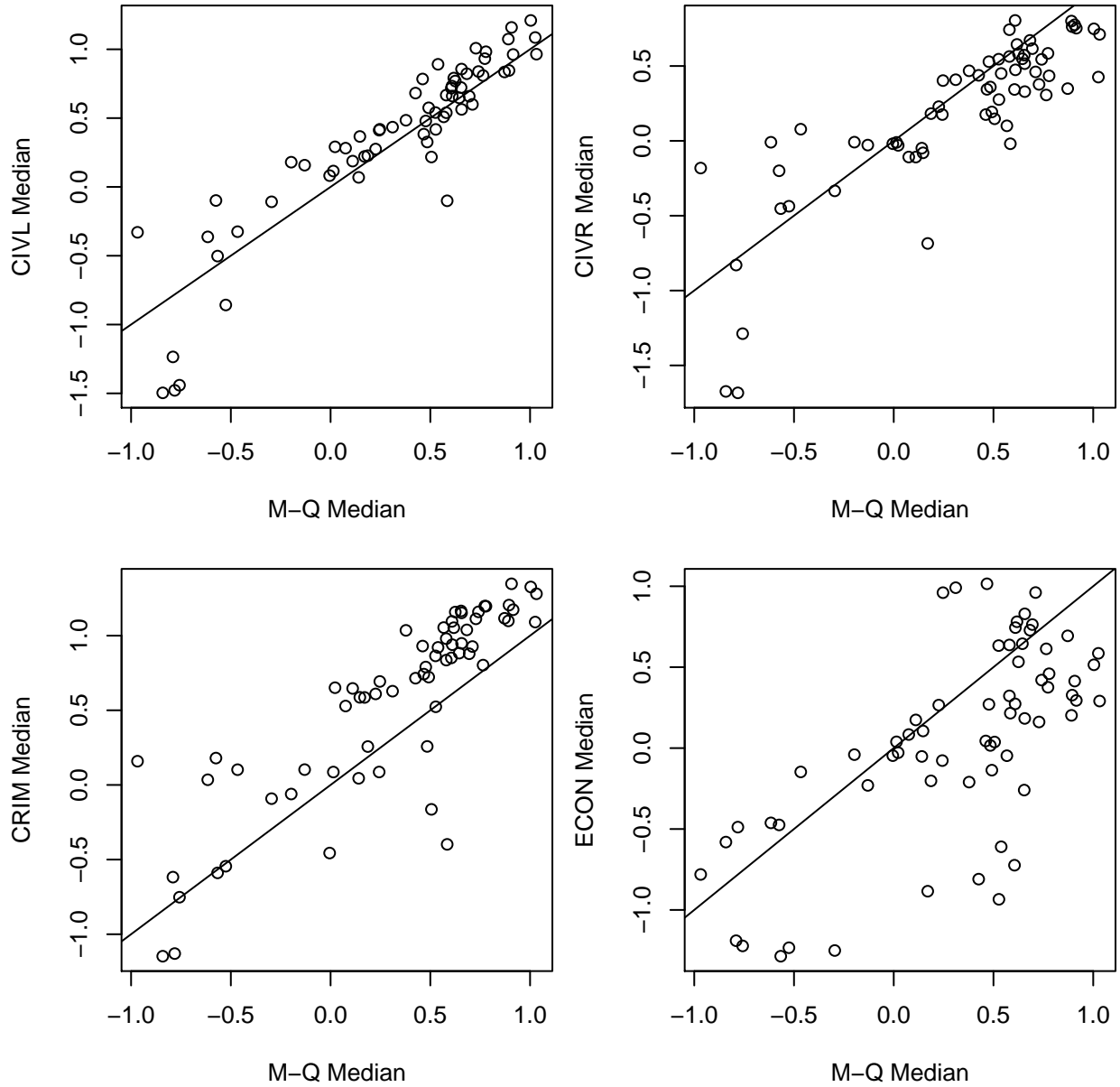Table 1: Mean posterior percent votes classified, by issue area, 1937-2002.

Figure 11: Comparison of estimated Court median for the dynamic ideal point model with single issue estimates for civil liberties (CIVL), civil rights (CIVR), criminal procedure (CRIM), and economics (ECON) cases.