

# USC Gould

School of Law

## Measuring Agency Attributes with Attitudes Across Time: A Method and Examples Using Large-Scale Federal Surveys

(Forthcoming in the *Journal of Public Administration Research and Theory*)

*Anthony M. Bertelli*

USC Price School of Public Policy

*Dyana P. Mason*

USC Price School of Public Policy

*Jennifer M. Connolly*

USC Price School of Public Policy

*David A. Gastwirth*

USC Price School of Public Policy

**Center in Law, Economics and Organization  
Research Papers Series No. C13-12**

**Legal Studies Research Paper Series No. 13-13**

June 17, 2013

## Measuring Agency Attributes with Attitudes Across Time: A Method and Examples Using Large-Scale Federal Surveys

Anthony M. Bertelli  
Professor and C.C. Crawford Chair in  
Management and Performance  
USC Price School of Public Policy  
University of Southern California

Professor of Public Policy Analysis  
University of Birmingham

email: bertelli@usc.edu

Dyana P. Mason  
Ph.D Candidate  
USC Price School of Public Policy  
University of Southern California  
email: dyanamas@usc.edu

Jennifer M. Connolly  
Ph.D Candidate  
USC Price School of Public Policy  
University of Southern California  
email: jennifer.connolly@usc.edu

David A. Gastwirth  
Ph.D Candidate  
USC Price School of Public Policy  
University of Southern California  
email: gastwirt@usc.edu

Forthcoming in the *Journal of Public Administration Research and Theory*

Previous versions of this paper were presented at the Annual Meeting of the American Political Science Association, Seattle, WA, September 2011 and the Public Management Research Conference, Syracuse, NY, June 2011. We thank Dan Mazmanian, Shui Yan Tang and the Bedrosian Center on Governance and the Public Enterprise for intellectual and financial support of this project. Preston Canter contributed substantially to earlier stages of the project. Rachel Finfer provided invaluable research assistance. We thank Klaus Brösamle, Scott Lamothe, Gerhard Hammerschmid, Thad Kousser, Pamela Clouser McCann, Nolan McCarty, Koen Verhoest and seminar participants at the Hertie School of Governance, the University of Washington and Katholieke Universiteit Leuven for helpful comments as well as Tim Johnson for assistance with the raw data. Upon publication, the measures in this paper will be made available at <http://agencydata.wordpress.com>.

## **Measuring Agency Attributes with Attitudes Across Time: A Method and Examples Using Large-Scale Federal Surveys**

### **Abstract**

Public management researchers are interested in many characteristics of organizations that cannot be directly captured, making aggregated attitudes from surveys an attractive proxy. Yet difficulties in measuring meaningful attributes over time and across organizations have frequently limited statistical designs to a single organization or time. We offer a method for creating such statistical measures across agencies and time using item response theory. Focusing our attention on U.S. federal administrative agencies, we marshal a variety of questions from surveys commissioned by the Office of Personnel Management and Merit Systems Protection Board and employ statistical models to measure three important attributes — autonomy, job satisfaction, and intrinsic motivation — for 71 agencies between 1998-2010. Our study provides a wealth of data for quantitative public management research designs as well as an adaptable framework for measuring a wide range of concepts.

Public management researchers seek to employ a wide variety of characteristics of public organizations in their efforts to uncover patterns in the structure and practice of public management within them. Yet many such attributes are not directly observable, for instance, employee motivation or satisfaction. Given this problem, surveys of employee attitudes have been an appealing source of proxy information (cf. Lee et. al 2010, Kim 2005, Sowa and Selden 2003, Whitford 2002, Wright and Davis 2003, Yang and Kassekert 2010). However, difficulties in measuring such attributes over time and across organizations remain, and has resulted in a large literature that uses attitudes to measure attributes, but makes inferences based on limited samples, for single organizations or from limited periods of time. Data availability has hampered cross-organizational designs that are vital to understanding the nature of bureaucratic governance. We aim to help public administration researchers measure attributes that are difficult to observe at the organizational, or agency level. Focusing our attention on federal administrative agencies in the U.S., we marshal a variety of questions from surveys commissioned by the Office of Personnel Management and the Merit Systems Protection Board, employing statistical models to measure three important attributes at the agency level and across time that are of great interest to public administration scholars: autonomy, job satisfaction, and intrinsic motivation. Our method for measuring these three attributes, which includes data for 71 agencies from 1998 to 2010, provides a general framework for recovering agency-level characteristics from survey response data that we hope will be of general interest to a wide variety of scholars in their pursuit of answers to multitudinous research questions.

It is certainly not novel in public administration research to use attitudes to capture agency-level concepts. For instance, efforts to develop and test models of organizational performance based on attitudes about work in agencies are prevalent (cf. Kim 2005; Brewer and Selden 2000; Ostroff 1992). While some have raised concerns about linking individual

attitudes to organizational level factors, Ostroff (1992, 965, 969) contends that, “[o]rganizational effectiveness measures can reflect, at least in part, the cumulative responses and interactions among employees.” Furthermore, U.S. separation-of-powers scholarship situates federal public managers within a system of institutions that shapes attitudes within administrative agencies (cf. Brehm and Gates 1997; Krause 1999; Bertelli and Feldmann 2007; Gailmard and Patty 2007; Lewis 2008; Lavertu and Moynihan 2012). Empirical research designs in these and other lines of inquiry would benefit substantially from measures that systematically and meaningfully aggregate attitudes to reflect practices and experiences in agencies.<sup>1</sup> Difficulties in aggregating these measures across time exist, and surveys used to measure attributes important to students of federal agencies had until recently been conducted sporadically.

We rely on advances in the measurement of latent traits in developing the strategy described below. Scholars in the separation-of-powers literature have made considerable progress in capturing the political ideology of administrative officials across federal agencies (e.g., Clinton and Lewis 2008, Clinton, Bertelli, Grose, Lewis, and Nixon 2012) and across both agencies and time (e.g., Bertelli and Grose 2011, 2009, Nixon 2004). This paper employs Bayesian methods (for a public administration contextual introduction, see Gill and Witko 2013) for dynamic ideal point estimation developed by Martin and Quinn (2002) in a novel way to develop our principal measurement strategy, offering valuable guidance to public management scholars interested in capturing latent attributes of agencies beyond ideology (see also Bertelli 2006, 2007). We implement our strategy using publicly available responses to

---

<sup>1</sup> Public administration scholars interested in explaining the role of attitudes in shaping behavioral considerations such as turnover intention and trust in bureaucratic agencies have relied on reduced-dimensional of attitudes in their work at the individual level (e.g., Rubin 2009, 133-34; Bertelli 2006; 2007b; Brehm and Gates 1997) utilize individual-level data from government-administered surveys to examine bureaucratic functional preferences. Nonetheless, work that bridges each of these literatures would be fruitful given appropriate aggregates of individual attitudes.

surveys commissioned by the Office of Personnel Management (OPM) and Merit Systems Protection Board (MSPB), namely, the Merit Principles Survey [2000, 2005], the National Partnership for Reinventing Government Employee Survey [1998, 1999, 2000], and the Federal Human Capital, now known as the Federal Employee Viewpoint Survey [2004, 2006, 2008, 2010]. Our method respects the limitations of the public use response data these surveys provide. In the pages that follow, we demonstrate how our estimation strategy can incorporate some important information for conducting analyses across organizations and time even when individual respondent data is not comparable over time.

Problems of comparability across organizations have also limited the amount we can learn about public management in federal bureaucratic organizations. Reflecting on the performance literature, for instance, Brewer and Selden (2000, 688) feel that the bulk of studies consider “only on a few agencies or bureaus [and], consider only a few factors” influencing the outputs and outcomes in which their authors are interested. However, these differences across organizations can provide essential information and detail how our strategy captures this kind of variation in a meaningful way. We believe that the measures developed using this method will help scholars to examine a wide variety of questions in quantitative public management research. Yet we hasten to add that the measures we present are not meant to be definitive means of capturing the concepts we study—autonomy, job satisfaction or intrinsic motivation. While we believe our measures are valuable in future public management research, we carefully explain our method of capturing them so that future analysts can work with a variety of analytic attributes across agencies and time.

We next describe the attributes of agencies measured in this study and the items from the aforementioned federal surveys used to inform them. The statistical measurement procedure we employ is then explained, as are our data and their sources. Our attention then

turns to a descriptive exploration of the measures we generate with an eye toward enhancing public administration research agendas. The paper concludes with implications of our analysis as well as thoughts about future research that is now possible using the method and measures we offer.

### **Organizational Portraits of Bureaucratic Attitudes**

This paper offers a method of using attitudinal data to measure concepts at the agency level that are difficult to capture more objectively. Our method brings together a variety of sources of information about the perceptions of agency employees by considering agreement and disagreement among respondents as well as the number of responses in each agency. To illustrate the approach, we offer some examples. While this measurement method can be used by future scholars to measure a wide variety of theoretically significant organizational attributes, we hope to provide some useful ones in the principal analysis.<sup>2</sup> We select three attributes that fit the limited observability criterion to measure across federal agencies and time: autonomy, job satisfaction, and intrinsic motivation.

#### *Autonomy*

Delegation of discretionary authority to bureaucrats is a strategic policy decision on the part of politicians, but it can also be “an unavoidable consequence of the political and institutional contexts” of particular public policies (Huber and Shipan 2002, 9). Actions clearly in excess of formal discretionary authority are democratically unsound and unaccountable, yet bureaucrats must in practice develop subjective expectations of the autonomy they enjoy in

---

<sup>2</sup> To identify salient constructs in public management, a team of graduate students conducted a content analysis of more than 100 published studies relying on the OPM and MSPB surveys. Common themes and latent traits were identified through a saturation approach, with themes entered as search terms in online archives of *Public Administration Review* and the *Journal of Public Administration Research and Theory*. Abstract and full-text searches were conducted for articles published between 1991-2011. Accountability is the most represented construct in these journals, associated attributes include managerial and employee autonomy or discretion as well as goal clarity and ambiguity. The frequency of occurrence in both scholarly journal articles and consistency of inclusion in the administrative surveys we employ were jointly considered as criteria for inclusion as examples of our method.

relation to a particular policymaking effort. Bertelli and Lynn (2003, 262) contend, “in all but the most routine tasks . . . administrator[s] . . . must strike a balance among competing interests, values, and interpretations of fact” (see also Sowa and Selden 2003, 700).

While related, autonomy is not what political scientists have considered administrative discretion. Statutory discretion, such as measured by Epstein and O’Halloran (1999) or Huber, Shipan and Pfahler (2001), includes measures of statutes delegating tasks to bureaucrats. These were analyzed for explicit discretion in the delegation. Instead, our measure of autonomy captures the perceptions bureaucrats have about the autonomy they enjoy when performing their work. These attitudes have been considered to play an essential role in shaping their on-the-job behavior (cf. Brehm and Gates 1997; Huber and Shipan 2002; Gailmard and Patty 2007). For instance, Brehm and Gates (1997, 98) claim, “the more a person feels in control of his or her surroundings, the more we should see the employee work rather than shirk,” which has implications for organizational performance of profound political significance. Bureaucrats also have substantial expertise and experience that can only be leveraged if they are given sufficient autonomy on the job (e.g., Schneider, Teske, and Mintrom 1995; Whitford 2002). Perceptions of autonomy can also tap non-pecuniary incentives for policy-motivated bureaucrats to choose public service rather than more lucrative careers outside government (Gailmard and Patty 2007).

### *Job Satisfaction*

Employees who have positive feelings towards their professional activities are thought to work harder and to be more committed to their organizations (Wright and Davis 2003). Job satisfaction is related to productive and responsible on-the-job bureaucratic activity that is not entirely the product of institutional design. Wright and Davis (2003, 70) define job satisfaction as “the congruence between what employees want from their jobs and what employees feel



they receive.” Attitudes have been historically linked to personnel stability and performance at the organizational level because public sector employees are thought to be less satisfied with their jobs than employees in private and nonprofit sector organizations; and because there is insufficient information to guide the increasingly commonplace strategic efforts of agencies to enhance bureaucratic job satisfaction (cf. Wright and Davis 2003; Kim 2005; Ritz 2009).

Satisfaction with compensation specifically is a rich area of management research (Williams, McDaniel and Nguyen 2006). Compensation (dis)satisfaction, mostly based on private sector research, is linked to turnover intention, organizational commitment, and employee citizenship behaviors (Miceli and Mulvey 2000). Scholars have also been substantially interested in compensation satisfaction in public sector contexts. Brehm and Gates (1997, 77), for instance, contend that the “fundamental pecuniary incentive of any organization is the base pay” (see also Currall, Towler, Judge, and Kohn 2005). Our primary measure of job satisfaction includes employee compensation satisfaction as an aspect of an individual’s level of total satisfaction with his or her job. We estimate a second measure without compensation and discuss the statistical relationship between these measures below.

### *Intrinsic Motivation*

The most basic distinction in the workplace motivation literature is between intrinsic motivation, the pursuit of enjoyment or self-expression in work, and extrinsic motivation, or pursuit of goals beyond work itself, such as financial gain or personal recognition (Amabile 2003). Some scholars see this motivation dichotomy as problematic, for public management reforms introduce pecuniary incentives that may crowd-out intrinsic motivation (cf. Deci and Ryan 1985; Bertelli 2006). Rainey and Steinbauer (1999) include three dimensions of motivation in a theory of effective government organizations – public service motivation (working for a collective good), mission motivation (the purpose of the work), and task

motivation (the mechanics of the work). These three components are related to the attribute of intrinsic motivation that we measure.

## **Data Sources**

To capture such concepts across federal agencies and time, large-scale administrative surveys are an alluring source of data. Three surveys of federal government employees, spanning nearly fifteen years, are utilized in this study. Each of these surveys employs a random sample of the U.S. Federal workforce, which produces a very large number of responses. However, surveys in different years sampled from a different list of agencies, and identifying information tying individuals to agencies in different implementations of the surveys is severely limited. Given the limitations in public use data faced by many public management researchers, we develop a method that takes them into account in the next section. Before doing so, we describe the federal surveys from which we take our response data and the survey items used to measure each attribute.

The first survey we enlist — the Merit Principles Survey — is administered by the MSPB. In these surveys, participation is voluntary and each individual's responses are kept confidential. The Merit Principles Survey is distributed to a random sample of full-time, permanent employees, supervisors, and managers representing a wide spectrum of the federal government. The first survey was completed in 1983 and the most recent in 2010. MSPB staff provided the 2000 and 2005 data.<sup>3</sup> The response rate for the Merit Principles Surveys was 43% in 2000 (N=6,958) and 50% in 2005 (N=36,926).

OPM administers the Federal Human Capital Survey (FHCS), now named the Federal Employee Viewpoint Survey. The FHCS is distributed using a stratified random sample of full-

---

<sup>3</sup> Data from iterations prior to 2000 could not be included in the study due to data conversion issues. Archival formats proved incompatible with various methods of data importation.

time federal employees; approximately 550,000 employees were be invited to participate in the survey in 2011, which will allow future extensions of our measures. The 2004, 2006, 2008, and 2010 data were obtained from the OPM website.<sup>4</sup> The FHCS in 2004 enjoyed a 54% response rate (N=147,914), in 2006 there was a response rate of 57% (N=221,425), in 2008 51% (N=212,223) and in 2010 52% (N=263,475)

The National Partnership for Reinventing Government was a task force created during the Clinton administration to identify ways that the federal government could modernize and streamline operations. One tool designed to advance this goal was a survey distributed to a random sample of federal employees representing 48 different agencies and organizations. The survey was created in conjunction with MPSB, OPM, and the Federal Aviation Administration. OPM was the primary agency tasked with overseeing the distribution of the surveys. Surveys were taken in 1998, 1999, and 2000. The 1998 Reinventing Government survey had a 40% response rate (N=13,657), 1999 response rate was 40% (N=12,755) and the 2000 survey had a response rate of 42% (N=21,157). The data was obtained from the Inter-University Consortium for Political and Social Research database (Study Nos. 3419, 3420, and 3421).

While evaluating the data, it became clear that different agencies and sub-agencies were included in each survey and occasionally in each survey year. For example, in the 2004 FHCS survey, 35 agencies were included in the sample. By 2010, that had increased to 76 agencies as they expanded the reach of the survey into more agencies across government. The number of respondents among the surveys we employ also varied and ranged from 6,958 to 263,475. However, to produce comparable measures across agencies and years, and to ensure that employees were not inadvertently double-counted among different agencies, the research team instituted three rules.

---

<sup>4</sup> The 2002 data is not available through public sources and could not be provided by OPM.

1. If a bureau (or sub-agency) was missing for more than two survey years in the data set, it was combined into its parent agency. For example, Marine Corps respondents identified in one survey for one year (one survey year) were combined with the Department of the Navy. The only exception occurred where the gap was spread over time (i.e., included in 1998 and 2005). In such cases, we “smoothed” the measure over the missing years, treating the missingness as described below.
2. If a sub-agency was not included for at least two survey years, then it was either combined or dropped altogether, i.e., if it is an independent agency. This primarily affected smaller independent agencies surveyed only once or twice with the FHCS.
3. A sub-agency must be included in all survey-years to stand alone, otherwise it was combined with its parent agency.<sup>5</sup>

Limitations of the data required that we assume that many individuals who work for a sub-agency were likely included in those years where only the agency was represented in the data provided. In other words, FEMA employees are assumed to have been included in efforts to survey Department of Homeland Security employees in any years following the merger with DHS. The data also do not capture individuals who moved between agencies, and there is no way to know if the same individuals were surveyed in all the surveys provided, or even if they were asked to respond in subsequent years of the same survey (i.e., MPS 2000 and 2005). This is not surprising considering the methodology used by government to randomly select who would receive the survey instrument, but it leaves some statistical contrasts unavailable for

---

<sup>5</sup> The only interesting, and outstanding item was the creation of the Department of Homeland Security (DHS) following the terrorist attacks of 2001. The Immigration and Naturalization Service was renamed Immigration and Customs Enforcement and moved from the Department of Justice (DOJ) to DHS. Although this makes measuring the impact of the individual agency more difficult, the data were combined using the rules above. For example, INS/ICE was combined with the DOJ prior to 2002, and DHS after. We implicitly assume that bureaus were already subsumed in their parent agencies in those surveys that had a narrower scope. For instance, we assume that Federal Aviation Administration (FAA) employee responses were indicated as coming from the Department of Transportation (DOT) in those surveys where the DOT was included but not FAA.

exploitation in measurement models. Although the raw data we use are individual survey responses, those individuals are not meaningfully identifiable, except as members of the agencies in which they work. Because we cannot match these individual responses to actual individuals since there is not identifying information and since the individuals responding to these surveys may have changed from year to year, multi-level modeling or a panel design would not be appropriate. We discuss these limitations and our efforts to address them in detail below.

### *Survey Items*

To select relevant survey questions to measure the latent attributes in this study, the research team reviewed all questions included in OPM, MSPB, and Reinventing Government surveys of the federal bureaucracy since 1979 based on literatures reviewed in connection with autonomy, job satisfaction, and intrinsic motivation. The selected list of survey items which were used to measure job satisfaction, autonomy, and intrinsic motivation is arranged by attribute, question text, survey, year, and question number in Table 1.

Individual responses to 17 questions asked over nine survey years were used to construct two measures of job satisfaction, with questions regarding compensation satisfaction included in one model but not another. Job satisfaction can be defined for measurement purposes as the “degree to which an employee has positive emotions toward the work role” (Currivan 1999, 501) and questions were selected that relate strongly to that definition. FHCS administrations in all years used the questions “Considering everything, how satisfied are you with your job?” (2004 Q65; 2006 Q60; 2008 Q61; 2010 Q69) and “Considering everything, how satisfied are you with your pay?” (2004 Q66; 2006 Q61; 2008 Q62; 2010 Q70). The 2000 Merit Principles Survey used the questions “I would recommend the government as a place to work” (2000 Q9; 2005 Q1k); “Overall, I am satisfied with my current pay” (2000 Q26); “Overall, I am

satisfied with my pay” (2005 Q20d) and “In general, I am satisfied with my job.” (2000 Q27; 2005 Q2m). Reinventing Government surveys included the following question in all three years, “Considering everything, how satisfied are you with your job?” (1998 Q29; 1999 Q28; 2000 Q28).

#### TABLE 1 ABOUT HERE

Twenty-six questions over nine survey years were used to measure autonomy. A literature in organizational behavior provides some guidance on selecting items that relate to autonomy in theory. In offering a perceptual basis for autonomy, Carpenter and Golden (1997, 190) claim that an employee’s “locus of control” — or her generalized perceptions of the extent to which she may control, or is controlled by, her environment (Rotter 1966) — is related to extent to which she perceives her general level of autonomy on the job. As such, we selected questions for this attribute that related to an individual’s locus of control. Thus our measure is not of policy discretion, but more general functional autonomy. The FHCS survey used the questions “I feel encouraged to come up with new and better ways of doing things.” (2004 Q4; 2006 Q4; 2008 Q4; 2010 Q3) “Employees have a feeling of personal empowerment with respect to work processes.” (2004 Q26; 2006 Q24; 2008 Q24; 2010 Q30), “Creativity and innovation are rewarded.” (2004 Q29; 2006 Q26; 2008 Q26; 2010 Q32), and “How satisfied are you with your involvement in decisions that affect your work?” (2004 Q59; 2006 Q54; 2008 Q55; 2010 Q63). The Merit Principles survey used “I have been given more flexibility in how I accomplish my work” (2000 Q6) and “Creativity and innovation are important” (2005 Q2h). Reinventing Government asked “Creativity and innovation are rewarded” (1998 Q11; 1999 Q11; 2000 Q11) and “How satisfied are you with your involvement in decisions that affect your work?” (1998 Q30, 1999 Q29; 2000 Q29) and “In the past 2 years, I have been given more flexibility in how I accomplish my work” (1999 Q18; 2000 Q18).

The measurement of intrinsic motivation involves ten questions from two surveys administered over six years. Once again, guidance from a literature in organizational behavior was available. These relate to a set of sample questions developed by Lawler and Hall (1970) to capture attitudes about “the motivating effects of the work itself” that federal bureaucrats perform (Rainey 1997, 202). Brehm and Gates (1997, 75) present the related idea of functional preferences in a rational choice framework; such preferences capture utility a bureaucrat derives from performing the tasks delegated to her. The FHCS used “My work gives me a feeling of personal accomplishment” (2004 Q6; 2006 Q5; 2008 Q5; 2010 Q4) and “The work I do is important” (2004 Q21; 2006 Q20; 2008 Q20; 2010 Q13). The Merit Principles survey asked “The work I do is meaningful to me” (2000 Q10; 2005 Q2k).

### **Measuring Agency Attributes with Attitudes Using Item Response Theory**

Our measurement strategy begins by constructing a comparable set of data across the agencies in our sample. This requires aggregation of a wide array of individual responses in a consistent and meaningful way and begins with creating a consistent directionality for the underlying survey items. Once again, our aim is to deal with specific limitations of the public use data in providing a framework for analysis by future scholars. FHCS and Reinventing Government surveys used one of two response sets.

- 5 “Strongly Agree”, 4 “Agree”, 3 “Neither Agree nor Disagree”, 2 “Disagree”, 1 “Strongly Disagree”
- 5 “Very Satisfied”, 4 “Satisfied”, 3 “Neither Satisfied nor Dissatisfied”, 2 “Dissatisfied”, 1 “Very Dissatisfied”<sup>6</sup>

---

<sup>6</sup> Three questions relating to the job satisfaction construct employ this satisfaction scale: “Considering everything, how satisfied are you with your job?”; “In General, I am satisfied with my job.”; “Considering everything, how satisfied are you with your pay?” One item relating to the autonomy construct uses this satisfaction scale: “How satisfied are you with decisions that affect your work?” All other items employ the agreement scale.

Merit Principles surveys used the following response sets:

- 5 “Strongly Disagree”, 4 “Disagree”, 3 “Neither Agree nor Disagree”, 2 “Agree”, 1 “Strongly Agree”
- 5 “Very Dissatisfied”, 4 “Dissatisfied”, 3 “Neither Satisfied nor Dissatisfied”, 2 “Satisfied”, 1 “Very Satisfied”

Merit Principles Survey response sets were reversed to be consistent with those employed in the FHCS and Reinventing Government questionnaires. Consequently, in all cases, when an individual working in agency  $j$  responds to item  $i$  in year  $t$  with a higher valued element of the response set, he or she expresses sentiment that is positively related to the trait we are measuring — for instance, higher valued responses are associated with higher levels of job satisfaction. Because they are not used consistently in the surveys and are somewhat rare, we exclude “don’t know” responses. An accompanying appendix (not for publication) provides full summary statistics for all questions used in our measurement models. Those questions that used “don’t know” options in the response set or had missing data (possibly due to coding or clerical errors in the original surveys) are described by summary statistics.

We do not intend to make interpersonal, but, rather, interagency comparisons. We therefore must devise a method to aggregate individual responses to the agency level. Unequivocally, our strategy reduces information from individual respondents, but we do so due to the nature of the data sources we employ. The primary problems arising from these comparisons are a paucity of information about individual respondents; differences in question wording, ordering and response sets; points in time at which no surveys were administered; and varying numbers of responses across agencies. We will discuss our means of dealing with each of these issues below. Because these sources are commonly used in public management research and because administrative surveys share similar characteristics, we offer a method



that provides comparable estimates across agencies while addressing the limitations of the data.

Thinking of the measurement problem as a multilevel, or hierarchical one is useful in revealing the sources of bias that could exist in attempting to utilize these surveys to compare agencies across time. In such an arrangement, we can consider individual responses as nested within administrative agencies *and* within years. If we were able to identify each respondent uniquely in every survey, we may be able to construct an unbalanced panel of response data, where each respondent provides data for the agency in which that respondent works for the year in which that respondent answered one of the surveys in our sample. For instance, suppose respondent A worked in the Department of the Interior in 2000, and responded to Q10 of the MPS “The work I do is meaningful to me” and then transferred jobs due to legislation that transferred a number of people in similar functions to the Department of Agriculture by 2004, responding to Q21 of the FHCS “The work I do is important” before leaving federal employment. Knowing her identity across time, we could record observations in 2000 and 2004 for respondent A in the appropriate agency for which she worked. Given the publicly available data for the MPS and FHCS, we can determine neither whether respondent A, given her identifier in the 2000 MPS, nor whether she responded to the 2004 FHCS and what agency she represented when making that response.

Two types of bias could occur from this situation. First, suppose we could identify her in two separate surveys. Because respondent A transferred from one department to another, treating her as two separate respondents would only be reasonable if the intertemporal correlation between her responses was not due to her unobserved personal characteristics. Addressing this problem would require a reasonable set of exogenous (non-survey response) individual-level variables, which are not present in these public use datasets. For example, the

Reinventing Government surveys do not ask any individual-level questions, such as age, race, gender, seniority or even supervisory status. Other surveys ask about race, but a follow-up question about Latino/Hispanic status has over 50% non-response. Some surveys ask about tenure, or rank, with categories, but others offer an open ended question. Education level is included in the MPS in 2000 and 2005, but not by any other surveys. Second, and more realistically, suppose we could identify respondent A in 2000, but not across time, such that we do not even know whether and in what agency she participated in 2004. If a sufficiently rich set of exogenous individual-level variables existed, respondent A might be “matched” with another respondent in 2004 using, for instance, a propensity score (Rosenbaum and Rubin 1983). A propensity score in this case would represent the probability of respondent A being a respondent in the Department of the Interior in 2004 conditional on set of exogenous variables available for respondents in both years in both surveys. If the propensity score were very high, the researcher might match respondent A with a respondent in the Department of Interior in 2004. If that set of variables did not include information about the legislation that precipitated her move, the match would generate systematic errors that could bias resulting measures. Many other problems along these lines can be considered, and, as we have discussed, exogenous individual-level information is extremely limited in such datasets.

Continuing with the example of respondent A, a second problem emerges from the wording of the questions and the construction of the response sets. The 2000 MPS question (“The work I do is meaningful to me”) asks for a self-referential response, whereas the 2004 FHCS (“The work I do is important”) drops the explicit self-reference and replaces “meaningful” with “important.” While they seem to tap similar attitudes related intrinsic motivation, it may pose a problem to consider them to be the same item for measurement purposes. A look at table 1, which provides question wording for each of the items we use, reveals a range of

differences in question wording. Moreover, changes to the ordering of questions of identical wording occur in nearly all cases. Nine items bear the same question number in multiple implementations of the same survey, but only two questions—one in regarding job satisfaction and one regarding autonomy—have the same wording across surveys, where ordering is far from consistent. For example, in the Federal Human Capital Survey the question, “Creativity and innovation are rewarded” was included in various years as question 26, as question 29 and as question 32. While the same question—“Creativity and innovation are rewarded”—was included as question 11 in all three years of the Reinventing Government Survey, it only appeared once in the 2005 Merit Principles Survey as question 2H. Another question, “In general, I am satisfied with my job” was question 27 in the 2000 Merit Principles Survey, but question 2M in 2005. These changes in the placement of these identical questions in varying years of the surveys could lead to question order effects, where individual responses to some questions may be influenced by the preceding questions (or lack thereof) (Schuman and Presser 1996). For example, in the FHCS of 2004, the question, “Creativity and innovation are rewarded” was asked after questions such as “How good a job do you feel is being done by your immediate supervisor/team leader?” and “My talents are well used in the workplace.” Such ordering could influence how a respondent reacts to the question of interest which follows. By contrast, in the Merit Principles survey, “Creativity and innovation are rewarded.” is the second item in the questionnaire. Another example of the potential for question order effects is the placement of questions related to job satisfaction. In the 2004 FHCS, the question, “Considering everything, how satisfied are you with your job?” is question 65, and is preceded by “I receive the training I need to perform my job” and “Employees are protected from health and safety hazards on the job.” By prompting employees to think about specific issues such as training and

protection from health and safety hazards prior to asking about their overall job satisfaction, there is a potential for response patterns unrelated to the attitudes the items seek to capture.

What is more, the response sets for these items, as we have noted above, differ across surveys. Item response models, such as the one we employ below, assume local independence of the items analyzed. This means that item responses are uncorrelated when conditioned on the latent trait (de Ayala 2009, 132). Put differently, there are other sources of correlation between the items besides the latent variable when local dependence is present. A violation of this local independence is threatened because the same question is being asked at different times. Importantly in this context, the question is being asked of different individuals, and we have no way of knowing whether responses from the same individual respondent are included in multiple surveys. Given the potential issues for comparability, we choose not to use any information that links these items together, allowing their relationship to the latent attributes to be studied separately.<sup>7</sup>

To approach the problem of differing response sets across surveys and due to limitations about what we know about individual respondents, we create an agency-level response variable in the following three steps and do so separately for each construct.

1. We sum all of the responses by employees of agency  $j$  to produce a raw score  $s_j$  for each agency in a given year.<sup>8</sup>

---

<sup>7</sup> If local independence proves to be a problem, one way that it is possible to assess this is through examination of what are called item difficulty parameters. We will return to this issue below.

<sup>8</sup> This calculation assumes that each respondent in an agency contributes equally to his or her employer's raw score. We make this assumption for two reasons. First, the administrative surveys we employ provide little and inconsistent information about a respondent's position in the agency available to the secondary analyst. The Reinventing Government 1998, 1999, and 2000 surveys do not include any information about the position of respondents in the agency hierarchy. The Merit Principles survey in 2005 provides only an indicator of supervisory status, while the earlier 2000 version of that survey provided both the supervisory status indicator and the GS-level of the respondent. The FHCS 2004, 2006, 2008, & 2010 provided both supervisory status AND GS-level. Beyond this, no information about the hierarchy is offered. Second, extant survey research at the agency level has shown that weighting schemes do not make much difference in the aggregate latent measures generated through item response theory. Clinton, et al. (Forthcoming 2012), for example, produce influence weights for latent

2. We calculate an inter-agency median  $\bar{s}$  raw score from the sums, which is simply the median of the raw scores for all agencies in the sample.
3. We create two groups – low and high satisfaction – by placing agencies above the median in the “high” group and those at or below the median in the “low” group.

The following dichotomous indicator records the median-split scores from step three and serves as our agency-level response variable for item  $i$  in agency  $j$  at time  $t$  in the measurement model described below.

$$r_{i,j,t} = \begin{cases} 1 & \text{if } s_j > \bar{s} \\ 0 & \text{if } s_j \leq \bar{s} \end{cases}$$

While this approach reduces variation in the response, it recognizes that differences in the response sets might make cutpoints in an ordered probability model less identifiable and meaningful. Agreement and disagreement of agency employees on the items we study informs these agency averages, and similar variation across agencies and time guides our estimation of latent attributes. Finally, because there is very little that we know consistently about individual respondents in each survey implementation and nothing comparable across time given the nature of the public use response data, it is difficult to gain any leverage in the measurement endeavor by incorporating individual respondents beyond their contribution to the agency averages we study. Such limits are informational, not artifacts of available estimation technology.

### *Statistical Measurement Model*

Armed with this response variable, our statistical problem is to model agency  $j$ 's probability of being in the “high” group on item  $i$  in year  $t$ ,  $\Pr(r_{i,j,t} = 1)$ , conditional on the

---

measures of ideology on the basis of agency aggregate responses to a survey question about the influence of appointees and appointee status. Their resulting weighted ideology estimates correlated very highly (0.7) with unweighted estimates. While we would like to provide such correlations for our measures, it is impossible to create weights without ignoring large numbers of items and responses over time.

unobserved value of agency  $j$ 's latent behavioral attribute in year  $t$ ,  $\theta_{j,t}$ , for instance, job satisfaction in the Department of Energy in 1998. To do so, we use what is known as a two-parameter item response model from the psychometric literature. Formally, we model the response variable ( $r_{j,t}$ ) as a linear function of the latent trait ( $\theta_{j,t}$ ).

$$r_{i,j,t} = \alpha_i + \beta_i \theta_{j,t} + \epsilon_{i,j,t} \quad (1)$$

The intercept in equation (1)  $\alpha_i$  is known as the difficulty parameter and  $\beta_i$  is the discrimination parameter for item  $i$  and the error term is independent and identically distributed standard normal (Martin and Quinn 2002, 138-39). If the difficulty parameter takes high values, agencies having a high degree of the latent attribute being studied are more likely to be in the high group of agencies on the item. By contrast, a small difficulty parameter indicates that agencies with relatively low levels of the attribute have improved chances of being in the high group. The slope in equation (1), or discrimination parameter, indicates the extent to which the item sorts low versus high levels of the attribute into low or high groups of responses ( $r_{i,j,t}$ ) respectively. Steeper slopes sort levels of a latent attribute between high and low response groups more effectively than shallower ones. It is through this specification that we can allow each item be studied separately in regard to how it informs the latent attribute as we discussed earlier in relation to local independence and question comparability. Operationally, we do this by giving each item  $i$  in table 1 separate difficulty and discrimination parameters ( $\alpha_i, \beta_i$ ). We do this as noted above due to questions about the comparability of items due to differences in wording and ordering of substantively similar items across surveys and years. While adding to the parameters to be estimated, this approach allows for the assessment of differences across similar items rather than assuming it away, as would be the case if similar items received the same difficulty and discrimination parameters. The two-parameter model is common in the ideal point estimation literature in political science, where

the latent trait is ideology and the response data are positions taken on issues, i.e., roll-call votes in legislatures (Clinton, Jackman and Rivers 2004); positions in judicial opinions (Martin and Quinn 2002), or a combination of these with response data for other actors on at least some of the same issues (cf. Bailey 2007; Bertelli and Grose 2009; 2011; Clinton, et al. 2012).

Because of the problem of intertemporal comparability of agency attitudes that we wish to address, we estimate this model in a Bayesian framework due to Martin and Quinn (2002).<sup>9</sup> Bayesian inferences are drawn from a summary of the following posterior density.

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta} | \mathbf{R}) \propto p(\mathbf{R} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) \quad (2)$$

In relationship (2),  $\boldsymbol{\alpha}, \boldsymbol{\beta}$  are “stacked” vector representations of the item parameters,  $\boldsymbol{\theta}$  are the latent attributes, and  $\mathbf{R}$  represents the observed response indicators (the  $r_{i,j,t}$ , that is) across agencies and time.<sup>10</sup>

#### *Priors*

We take a Bayesian approach; the term  $p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})$  in (2) represents our prior beliefs about the parameters which are specified in the following way. The item parameters are given a standard normal prior distribution  $\begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix} \sim \mathcal{N}_2(0,1)$ . The attributes have the prior distribution  $\theta_{j,t} \sim \mathcal{N}(\theta_{j,t-1}, \Delta_{\theta_{j,t}})$  where the mean is the attribute in the last period and  $\Delta_{\theta_{j,t}}$  is an “evolution variance parameter” that determines “how much borrowing of strength (or smoothing) takes

---

<sup>9</sup> Specifically, we use the MCMCdynamicIRT1d routine for MCMCpack in the R statistical computing environment. This package makes it straightforward for researchers to employ our method to measure other attributes. Bayesian and non-Bayesian alternatives are available in other software environments, but this framework combines ease of use with appropriateness of modeling to address the agency and time comparability issues we confront in this paper. All information about our specification in MCMCpack is included in this section and the data that we employed as well as the measures we have generated will be available freely via a web upon publication. A nice treatment of the dynamic measurement problem using another computing environment, JAGS, is offered by Jackman (2009, 471-88).

<sup>10</sup> The sampling density is  $p(\mathbf{R} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) \propto \prod_{t=1}^T \prod_{i=1}^{N_t} \prod_{j=1}^{M_t} \Phi(\alpha_i + \beta_i \theta_{j,t})^{r_{i,j,t}} (1 - \Phi(\alpha_i + \beta_i \theta_{j,t}))^{1-r_{i,j,t}}$  where  $N_t$  represents the total number of items for which there are responses at time  $t$  and  $M_t$  is the total number of agencies for which a response has been observed at time  $t$ .

place from one time period to the next” (Martin and Quinn 2002, 140). When  $\Delta_{\theta_{j,t}} = 0$ , the behavioral attribute is fixed across time; by contrast, values of the evolution variance parameter that approach infinity would construct attributes that are independent in each year of our study. Neither of these extremes is realistic. It is natural to believe that some temporal dependence is present given the persistence of personnel and even survey respondents from one year to the next. Importantly, years in which surveys are not taken provide no response data and a prior indicating dependence is required to smooth the estimates for missing years.<sup>11</sup> This leaves us with a tradeoff: we do not want to build excessive dependence into our prior specification, but we must use information from past survey implementations to smooth the data over the missing observations. We choose a middle ground of  $\Delta_{\theta_{j,t}} = 0.5$ .<sup>12</sup> This value borrows considerably less from the past than does the Martin and Quinn (2002, 147) application to judicial ideal points, which sets the evolution variance at 0.1 for the majority of justices. Their problem is different, however, in that the same justices are revealing positions in opinions in several years of their sample in the vast majority of cases, so strong borrowing from past positions allows them to hypothesize that observable changes in ideology are more likely to be real ones. We do not want to make as strong an assumption about agencies, which change personnel and even direction, for instance, under different presidential administrations. Thus, we opt for less smoothing. To be sure, this is a choice for the researcher, and scholars applying our method can make appropriate selections. An advantage of the Martin and Quinn (2002) estimator is that the choice is, in fact, possible.

---

<sup>11</sup> To retain years in which no questions were included, an indicator for a non-survey year was included with the values for all survey items in that year indicated as missing. These indicators represent the equivalent of a question asked in that year to which no agency provided responses. As expected, these items have discrimination and difficulty parameters equal to zero and do not bias the scale construction in the years for which responses are present.

<sup>12</sup> We experimented with a variety of evolution variance parametric choices and did not choose a value below 0.5 for reasons of numerical stability.



The Martin and Quinn (2002, 140) estimator also requires us to anchor each time series of latent attributes to an initial value at the unobserved period  $t = 0$ . That is, we make the assumption that  $\theta_{0,j} \sim \mathcal{N}(0, \frac{1}{z_j})$  where  $z_j$  is the total number of individual responses to all survey items related to the latent attribute for agency  $j$  across time.<sup>13</sup> We center the estimates around zero with this prior specification to make the scales of our attribute measures more sensible to interpret, but the variance specification allows us to incorporate uncertainty into the measurement model given variation in responses collected across agencies for each latent attribute, another of the data problems we have noted. While the response rate to these administrative surveys is high, the variation across agencies in the number of responses used to create  $r_{i,j,t}$  is substantial.<sup>14</sup> In our specification, as the number of responses from an agency increases, the prior variance decreases to reflect the fact that we are basing our estimates on information from more employees of the agency. Conversely, when an agency has a small number of respondents, we impose a wider variance on the latent attribute due to limitations in our information from that agency's employees. An appendix (not for publication) provides detailed information on responses to each item.

---

<sup>13</sup> We intend for this prior restriction to relate to our confidence in the measured attributes from a statistical perspective, introducing information about cross-agency response frequencies into the model. While full information on each question is available in the appendix (not for publication), an example is warranted here. From the 2004 FHCS, response data for Q4 is used in our measurement model for autonomy. That question received 147,899 total responses; responses from the Office of Management and Budget totaled 249 while 10,402 responses were logged from the Department of Agriculture. Such differential patterns exist in all of the item responses we use in this study. While it is possible to bring more information into the model, for example, by considering total responses in proportion to agency size or by considering responses relative to respondents' position hierarchy, it is difficult to reliably collect such data. Thus we focus on a proxy for the richness of information within an agency. We simply want to provide more conservative uncertainty estimates when we base the attributes on fewer responses (less information) and provide a prior that can be overwhelmed by the data rather than a constraint that cannot.

<sup>14</sup> Response rates for the Reinventing Government surveys in 1998 (40 percent), 1999 (56 percent) and 2000 (42 percent). The Merit Principles surveys had rates as high as 74 percent in 1989, falling to 64 percent in 1992, 53 percent in 1996, 43 percent in 2000 and 50 percent in 2005. Exemplary Federal Human Capital Survey response rates for 2004 (54 percent), 2006 (57 percent), 2008 (51 percent) and 2010 (52 percent) were also quite high.

Completing the prior specification requires identification of the direction of the latent scales, we used the highest and lowest raw agency item sums ( $s_j$ ) to inform the estimation by holding constant the behavioral attribute for agency with the highest mean at an arbitrarily high value in the first time period  $\theta_{\bar{s}_j,1} = 3$  and for the agency with the lowest mean an arbitrarily low value of  $\theta_{\underline{s}_j,1} = -3$  for in the first time period. Low and high values used in this way for the autonomy measure were Department of Labor, and the Department of State respectively. For both job satisfaction measures (with and without the pay questions), the Consumer Product Safety Commission had the lowest, while the National Aeronautic and Space Administration had the highest levels of  $s_j$ . The intrinsic motivation model used the Department of the Housing and Urban Development for the prior restriction for informing the low and the Social Security Administration for the high end of the scale. Where agencies tied for high or low values, agencies with the greatest coverage across years were selected for identification restrictions. This rule only applied to the Department of State in the autonomy measure.

### *Estimation*

For each agency-level attribute, our estimates are based on 1.5 million draws from the posterior distribution in (2), with the first 150,000 discarded as “burn-in” with every fiftieth observation used for inference (a total of 27,000 draws). Strong evidence of convergence was gleaned from an examination of trace and autocorrelation plots. We observed failure of the Geweke (1992) diagnostic in less than 5 percent of the parameters for all four attributes. Such results would be expected by chance at conventional levels of statistical significance.

FIGURES 1-3 ABOUT HERE

Figures 1 through 3 graphically depict the discrimination parameters and 95 percent HPD intervals for each measurement model. As noted above, positive discrimination

parameters indicate that higher values of the latent variable are associated with a higher probability of being in the “high” group of agencies ( $r_{i,j,t} = 1$ ) on the item indicated and negative values indicate the inverse relationship.<sup>15</sup> Across all latent traits, all but seven of the indicators significantly discriminate between high and low values of the latent traits in the measurement models.<sup>16</sup> Figure 2 shows that all items used to measure the autonomy attribute have a positive valued discrimination parameter. In figure 3, we see that the FHCS items have positive discrimination parameters as does Q20d from the 2005 Merit Principles Survey that asked for agreement with the statement “Creativity and innovation are important.” In figure 4, only the Merit Principles Survey items asking for agreement with the statement “The work I do is meaningful to me” did not positively discriminate. Beyond an examination of the importance of items in our scale construction, this item analysis provides some information for researchers interested in using individual questions to examine agency-level attributes in the various years of the surveys. For instance, FHCS items appear to be better suited to measuring job satisfaction and intrinsic motivation at the agency-level than any of the other surveys. We also do not observe discrimination parameters values that run counter to our measurement expectations, in our case, negative values that suggest that higher valued items relate to lower values of the latent attributes being studied. Of course, items should be excluded from a scale if they simply measure a different concept. At low levels of discrimination, items still provide

---

<sup>15</sup> While we have coded the data such that higher values of the question are expected to be associated with higher values of the latent variable, but that relationship is estimated in the measurement models as the discrimination parameter.

<sup>16</sup> Table 2 does not display the “dummy” indicators used to ensure smoothing over missing values as noted in footnote 11.

some additional information and reduce measurement error. As long as that information is not contradictory, concerns about the validity of measures are not significant.<sup>17</sup>

Now that we have estimates of important attributes for federal agencies, we examine them in greater detail. In the following section, we turn to a descriptive analysis of a sample of the estimates for each attribute. We offer this discussion simply to illustrate some patterns that are present in the data and that provide face validity to their measurement. We then conclude with a discussion of how scholars might make use of our measures in future research endeavors.

### **Exploring the Data**

Agency-level estimates are shown in figures 4-6. Each figure shows posterior means ( $\theta_{i,j,t}$ ) for each agency as solid lines with a 95-percent highest posterior density region represented by dashed lines. This region indicates uncertainty in the estimates of the latent traits in a way comparable to a frequentist confidence interval. Lower scores represent lower values of each behavioral attribute and vice versa. We provide below a descriptive exploration of some of the patterns our estimates revealed. In some instances, these patterns are not inconsistent with known policy shocks affecting these agencies over the last decade and half. These observations are suggestive and can and should be further explored in public management research.

A portrait of our autonomy estimates (mean = -0.01, SD = 1.77, number of agencies = 71, number of agency-year observations = 573) for selected agencies appears in figure 4. The Department of Education (EDUC) score remains above the mean until 2001 when the No Child Left Behind legislation was signed into law, then steadily declines over the next decade to a low

---

<sup>17</sup> Researchers who have concerns where discriminant validity is a concern may consider a multidimensional item response approach (see the claim in McDonald 2000, p. 108). The method we offer posits a single latent trait,  $\theta$ , and the discrimination parameters provide a way of understanding the relationship from that single theorized trait and each item providing information about it.

of -2.81 in 2010. The autonomy measure for the Department of Energy (DOE) hovers around 0.4 until 2005, and then the score begins a steady decline to -2.27 in 2010. The Environmental Protection Agency (EPA), on the other hand, begins the study period in 1998 at 1.79, peaks at 2.12 in 2000, the final year of the Clinton Administration, and slowly decreases each year thereafter to a low of 0.30 in 2010. Both the Army and Navy have scores near the mean in 1998, but their scores steadily increase over the next decade, approaching a value of three by 2010. The Equal Employment Opportunity Commission (EEOC) sees a dramatic decline in autonomy beginning in 2001, but levels off by 2009, albeit far below the mean at -3.03.

#### FIGURES 4-6 ABOUT HERE

Our job satisfaction estimates (mean = 0.07, SD = 1.58, number of agencies = 71, number of agency-year observations = 573) also reveal interesting patterns and are shown for selected agencies in figure 5. Department of Education estimates steadily decline over the 2000-2006 period from 1.26 to -0.25, then begin to increase over the next four years to a score above the mean (0.82). The period of decline corresponds to the most intense implementation period of No Child Left Behind policies. OPM begins and ends the period just above the mean. The Department of the Interior (DOI) dramatically declines in employee job satisfaction over the study period, falling from 0.31 in 1998 to -3.31 in 2010. DOJ, on the other hand, displays job satisfaction scores that rise from -0.46 in 1998 to 3.24 in 2010. The General Service Administration (GSA), charged with overseeing the day-to-day tasks of running government buildings, sees a steady rise in its above average job satisfaction estimates — from 1.30 in 1998 to 3.48 in 2010. The Department of Energy (ENGY) figures begin the study period below the mean at -0.80 in 1998, achieves average levels by 2002, and steadily increases to 1.23 in 2010. Interestingly, EPA job satisfaction scores rose from 24 percent during the George W. Bush administration, but declined slightly during President Obama's first two years in office. The

Treasury job satisfaction score is revealing of its importance during the financial meltdown; it is well below the mean in 1998 (-1.09) but becomes markedly above average by 2010 (2.13).

Intrinsic motivation estimates (mean = 0.05, SD = 1.68, number of agencies = 71, number of agency-year observations = 513) are represented in figure 6. As noted, they have a shorter range of years, 2000-2010 than the other attributes we measure because relevant questions were not included in earlier surveys. Furthermore, panels depicting scores for the National Credit Union Administration (NCUA), National Labor Relations Board (NLRB), Federal Trade Commission (FTC), and National Archives and Records Administration (NARA) remind the reader that scores are only estimated beginning in the first year observed responses to a survey item were recorded for an agency's employees. The Department of Education has an below average intrinsic motivation estimate of -1.45 in 2000, but that steadily decreases over the next decade to -3.36 in 2010. The trend for DOI is similar, falling from -1.94 in 2000 to -3.33 in 2000. The EPA intrinsic motivation score shows less of a decline over the decade (-1.14 in 2000 and -1.97 in 2010). Department of Energy intrinsic motivation scores, by contrast, begin just above the mean at 0.23 in 2000 and fall to -2.12 in 2010. The Department of State sees a similar growth pattern with intrinsic motivation, rising from 0.99 in 2000 to 3.31 in 2010. In comparison, Social Security Administration (SSA) scores are always above average, increasing from 1.22 in 2000 to 3.34 in 2010, and Health and Human Services (HHS) begins and ends the period with a score close to the mean. The Department of the Treasury (TREAS) is well below the mean at -1.20 in 2000 and falls to -3.42 by 2010.

#### FIGURES 7-9 ABOUT HERE

While in the preceding paragraphs we noted interesting patterns of attribute trends by agency over time, figures 7-9 graphically describe inter-agency differences using data on agency characteristics compiled for the Administrative Conference of the United States

(ACUS) (Lewis and Selin 2012). The vertical lines represent the overall average value for the attributes across all “bins”. The dots, pluses and circles represent the average mean over time, which is the average value of the attribute measures of each bin. The numbers or labels represent upper bounds of each range (i.e., in figure 7, 10,000 indicates average values for agencies employing between 1,001 and 10,000 staff). The three agency characteristics we examine descriptively—size of agency, agency independence and the percentage of political appointees—are institutional choices that may correlate with agency accountability, responsiveness, control or performance (Kalleberg et al 1996, Ritz 2009, Yesilkagit and Christensen 2010, Lewis and Wood 2011, Lewis 2011, Gill and Waterman 2004). Agency size, for example, given in figure 7, widely varies among U.S. agencies and has been of interest to public management scholars. Most theorizing argues that agency size is negatively correlated to performance, including internal efficiency as larger public agencies are more bureaucratic Kalleberg et al 1996; Ritz 2009). However, agency size has been primarily used as a control variable in studies on other topics (see, for example, Yang and Kassekert 2010 in their discussion on job satisfaction and management reform or Lee, Rainey and Chun 2010 on goal ambiguity and work complexity in federal agencies). Interestingly, our data show that the larger the agency (as measured by staff size), the higher the values for autonomy, job satisfaction and intrinsic motivation, although the relationship starts to reverse itself for very large agencies. For example, agencies with 500-1000 employees scored well below average for autonomy, while agencies between 1,000 and 10,000 employees were just below average. This group of agencies was also slightly above average for both job satisfaction and intrinsic motivation.

Agency independence—whether an agency falls outside the Executive Office of the President or within executive departments (Lewis and Selin 2012)—is an important

consideration in understanding agency autonomy and political control (Christensen and Læg Reid 2007; Yesilkagit and Christensen 2010), or for evaluating the role of agency independence on public accountability (Lewis and Wood 2011). As seen in figure 8, agencies that are not independent score higher than average on all three scores, while those independent agencies score lower than average. In other words, those agencies more tightly coupled with the President show higher levels of all three attributes. The role of political appointees versus careerists in job satisfaction, autonomy and intrinsic motivation also provides an interesting finding. Studies as those by Lewis (2010) and Gill and Waterman (2004) argue that political appointments can have an effect on performance. Our data show that as agencies have more political appointees (defined as those appointees requiring Senate confirmation, non-career Senior Executive Service, Schedule C exempted appointment, and other appointees not subject to Senate confirmation), their average values are lower as shown in figure 9. While scholars have found that changes in managerial practices in federal agencies are associated with lower employee job satisfaction (Yang and Kassekert 2010) our data perhaps suggest that higher levels of uncertainty in agencies with more turnover in executive positions may have an effect on rank-and-file employees as well, but this is a question to be studied in the future. We underscore that these graphs do not provide statistical evidence of the depicted relationships, but, we hope, inspire scholars to seek to understand them.

Finally, we look more closely at our measure of job satisfaction with and without the questions regarding compensation. Some literature suggests that these are two distinct attributes (Nagy 2002; Scarpello and Campbell 1983). We estimated a fixed effects regression model to explore the relationship between the two job satisfaction measures – the one which includes compensation satisfaction and the one which does not. The resulting correlation between our measures of job satisfaction including and excluding compensation satisfaction is



positive, but not overwhelming in magnitude (Coefficient = 0.571,  $t=4.75$ ). Consistent with previous studies, the introduction of pay satisfaction into the measure appears to tap a different dimension of the attribute. Researchers can employ our estimates to attempt to understand the determinants of this characteristic across federal agencies.

These descriptive portraits of our estimates show patterns that should be of considerable interest for future research. They follow patterns that should be of relevance to scholars who seek to understand the relationship between the attributes we measure and policy shocks or key institutional characteristics.

## **Conclusion**

Many agency-level attributes are not directly observable, and, as result, scholars have turned to attitudinal surveys to measure these attributes (cf. Lee et. Al 2010, Kim 2005, Sowa and Selden 2003, Whitford 2002, Wright and Davis 2003, Yang and Kassekert 2010). Public management research often uses attitudes to measure such characteristics, but the availability of response data has restricted inferences to limited samples, single organizations or short periods of time. We offer a method for generating intertemporal statistical measures to capture key attributes across organizations. Working at the federal agency level, we create samples of such measures for attributes of interest to many public management scholars, namely, autonomy, job satisfaction, and intrinsic motivation, for 71 federal agencies from 1998 through 2010. Although the data we marshal have well known limitations, the method we develop in this paper—a novel utilization of dynamic ideal point estimation—can provide an important tool for public management scholars. Agency and temporal differences provide crucial information regarding the role of concepts of interest in public management and public administration, and with these agency-level measures and the method we offer, scholars can more carefully test competing theories and conventional wisdom in practice.

Research applications using these agency level measures can have important implications for understanding the practice of public administration. We hope that the linkages between the attribute measures we have constructed and a variety of organizational phenomena will be more carefully studied in the future. Numerous possibilities exist for seeking to understand variation in agency attributes. Studies of how institutional factors influence these concepts and studies of how they influence agency outputs and outcomes would be of considerable interest. One fruitful avenue may be to examine the impact of policy changes across the agencies and time periods our measures reflect. The influence of executive succession and turnover, both in terms of the nature and frequency of senior leadership change would also provide important insights into political and administrative transitions. Other institutional characteristics, such as the concentration of political appointees in an agency, may also have influences on employee perception, performance and organizational outputs (e.g., Lewis 2008). So, too, could identifying the impact of budgetary changes on the attributes of federal agencies could also provide new and important insights. Linking agency job satisfaction levels to quit rates may help to identify the potential long-term impact of continuous employment of dissatisfied personnel. The measures of uncertainty for our attribute estimates can also provide measures of the clarity or ambiguity of attributes. For example, ambiguity about autonomy in an agency as measured through our uncertainty estimates might have an important impact on performance metrics or agency outcomes. Our measures make these and many other topics available for quantitative study across agencies and time.

By creating this set of intertemporal statistical measures to capture autonomy, job satisfaction, and intrinsic motivation at the federal agency level from 1998 through 2010 we hope to foster research projects that span across agencies and across time. While many informative papers have examined the attributes we captured, our measures will allow

analysts to broaden the scope of research on these topics across federal agencies and time. The method by which we measured them can have even more interesting future applications.

## References

- Amabile, Teresa M. 1993. "Motivational Synergy: Toward New Conceptualizations of Intrinsic and Extrinsic Motivation in the Workplace." *Human Resource Management Review* 3(3): 185-201.
- Bailey, Michael A. "Comparable preference estimates across time and institutions for the court, congress, and presidency." *American Journal of Political Science* 51.3 (2007): 433-448.
- Bailey, Kenneth. 2007. *Methods of social research*. New York: Simon and Schuster.
- Bertelli, Anthony M. 2006. "Motivation crowding and the federal civil servant: Evidence from the US Internal Revenue Service." *International Public Management Journal* 9(1): 3-23.
- Bertelli, Anthony M. 2007. "Determinants of Bureaucratic Turnover Intention: Evidence from the Department of the Treasury." *Journal of Public Administration Research and Theory* 17(2): 235-258.
- Bertelli, Anthony M. 2012. *The Political Economy of Public Sector Governance*. New York: Cambridge University Press.
- Bertelli, Anthony M., and Sven E. Feldmann. 2007. "Strategic Appointments." *Journal of Public Administration Research and Theory* 17: 19-38.
- Bertelli, Anthony M., and Christian R. Grose. 2009. "Secretaries of Pork? A New Theory of Distributive Public Policy." *Journal of Politics* 71: 926-45.
- Bertelli, Anthony M., and Christian R. Grose. 2011. "The Lengthened Shadow of Another Institution: Ideal Point Estimates for the Executive Branch and Congress." *American Journal of Political Science* 55(4): 767-781.
- Bertelli, Anthony M., and Laurence E. Lynn. 2003. "Managerial Responsibility." *Public Administration Review* 63(3): 259-268.
- Brehm, John, and Scott Gates. 1997. *Working, Shirking, and Sabotage*. Ann Arbor, MI: University of Michigan Press.
- Brewer, Gene, and Sally Coleman Selden. 2000. "Why Elephants Gallop: Assessing and Predicting Organizational Performance in Federal Agencies." *Journal of Public Administration Research and Theory* 10(4): 685-711.
- Carpenter, Mason A., and Brian R. Golden. 1997. "Managerial Discretion: A Study of Cause And Effect." *Strategic Management Journal* 18(3): 187-206.
- Christensen, Tom, and Per Lægveid. 2007. "Regulatory agencies—The challenges of balancing agency autonomy and political control." *Governance* 20(3): 499-520.

- Clinton, Joshua D., Anthony M. Bertelli, Christian R. Grose, David E. Lewis, David C. Nixon. Forthcoming 2012. "Separated Powers in the United States: The Ideology of Agencies, Presidents, and Congress." *American Journal of Political Science* 56(32): 341-354.
- Clinton, Joshua D., and David E. Lewis. 2008. "Expert opinion, agency characteristics, and agency preferences." *Political Analysis* 16(1): 3-20.
- Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. "The statistical analysis of roll call data." *American Political Science Review* 98(02): 355-370.
- Currall, Steven C., Annette J. Towler, Timothy A. Judge, and Laura Kohn. 2005. "Pay Satisfaction and Organizational Outcomes." *Personnel Psychology* 58: 613-640.
- Currivan, Douglas B. 1999. "The Causal Order of Job Satisfaction and Organizational Commitment in Models of Employee Turnover." *Human Resource Management Review* 9(4): 495-524.
- De Ayala, Rafael Jaime. 2009. *The theory and practice of item response theory*. New York: Guilford Press.
- Deci, Edward L., and Richard M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum.
- Epstein, David, and Sharyn O'Halloran. 1999. *Delegating Powers: A Transaction Cost Politics Approach to Policy Making Under Separate Powers*. New York: Cambridge University Press.
- Gailmard, Sean, and John W. Patty. 2007. "Slackers and Zealots: Civil Service, Policy Discretion, and Bureaucratic Expertise." *American Journal of Political Science* 51: 873-89.
- Geweke, John. 1992. "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion)." In *Bayesian Statistics 4*, J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M Smith. Oxford: Oxford University Press, pp. 169-193.
- Gill, Jeff, and Richard W. Waterman. 2004. "Solidary and functional costs: Explaining the presidential appointment contradiction." *Journal of Public Administration Research and Theory* 14(4): 547-569.
- Gill, Jeff and Christopher Witko. 2013. "Bayesian Analytical Methods: A Methodological Prescription for Public Administration." *Journal of Public Administration Research and Theory* 23(2): 457-494.
- Huber, John D., and Charles R. Shipan. 2002. *Deliberate Discretion? The Institutional Foundations of Bureaucratic Autonomy*. New York: Cambridge University Press.
- Huber, John D., Charles R. Shipan, and Madeleine Pfahler. 2001. "Legislatures and statutory control of bureaucracy." *American Journal of Political Science* 45(2): 330-345.

- Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences*. Chichester: John Wiley.
- Kalleberg, Arne L., David Knoke, Peter V. Marsden, and Joe L. Spaeth, eds. 1996. *Organizations in America: Analysing their structures and human resource practices*. Thousand Oaks, CA: Sage Publications, Inc.
- Krause, George A. 1999. *A Two-Way Street: The Institutional Dynamics of the Modern Administrative State*. Pittsburgh: University of Pittsburgh Press.
- Lavertu, Stéphane and Donald Moynihan. Forthcoming. "Agency Political Ideology and Reform Implementation: Performance Management in the Bush Administration" *Journal of Public Administration Research and Theory*.
- Lawler, Edward E., III and Douglas T. Hall. 1970. "Relationship of Job Characteristics to Job Involvement, Satisfaction and Intrinsic Motivation." *Journal of Applied Psychology* 54: 305-312.
- Lee, Jung Wook, Hal G. Rainey, and Young Han Chun. 2010 "Goal ambiguity, work complexity, and work routineness in federal agencies." *The American Review of Public Administration* 40(3): 284-308.
- Lewis, David E. 2008. *The Politics of Presidential Appointments: Political Control and Bureaucratic Performance*. Princeton: Princeton University Press.
- Lewis, David E. 2010. *The politics of presidential appointments: Political control and bureaucratic performance*. Princeton: Princeton University Press.
- Lewis, David E. and Jennifer L. Selin. 2012. *Administrative Conference of the United States: Sourcebook of United States Executive Agencies*. Washington, D.C.: Administrative Conference of the United States.
- Lewis, David, and Abby Wood. 2011. "Judicial Deference and Agency Accountability: A Federal FOIA Experiment." Unpublished manuscript available at SSRN 1884392.
- Kim, Sangmook. 2005. "Individual-Level Factors and Organizational Performance in Government Organizations." *Journal of Public Administration Research and Theory* 15(2): 245-261.
- Martin, Andrew D. and Quinn, Kevin M. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999." *Political Analysis* 10(2):134-153.
- McCubbins, Mathew D., Roger G. Noll, and Barry R. Weingast. 1987. "Administrative procedures as instruments of political control." *Journal of Law, Economics, & Organization* 3(2): 243-277.

- McCubbins, Matthew D., Roger G. Noll, and Barry R. Weingast. 1989. "Structure and process, politics and policy: Administrative arrangements and the political control of agencies." *Virginia Law Review* 75(2): 431-482.
- McDonald, Roderick P. 2000. "A basis for multidimensional item response theory." *Applied Psychological Measurement* 24(2): 99-114.
- Miceli, Marcia P., and Paul W. Mulvey. 2000. "Consequences of Satisfaction with Pay Systems: Two Field Studies." *Industrial Relations* 39(1): 62-87.
- Moe, Terry M. 2005. "Power and political institutions." *Perspectives on Politics* 3(2): 215-233.
- Nagy, Mark S. 2002. "Using a single-item approach to measure facet job satisfaction." *Journal of Occupational and Organizational Psychology* 75(1): 77-86.
- Nixon, David C. 2004. "Separation of Powers and Appointee Ideology." *Journal of Law, Economics, and Organization* 20:438-57.
- Ostroff, Cheri. 1992. "The relationship between satisfaction, attitudes, and performance: An organizational level analysis." *Journal of Applied Psychology* 77(6): 963.
- Rainey, Hal G. 1997. *Understanding and Managing Public Organizations. 2nd Edition*. San Francisco: Jossey-Bass.
- Rainey, Hal G. and Paula Steinbauer. 1999. "Galloping Elephants: Developing Elements of a Theory of Effective Organizations." *Journal of Public Administration Research and Theory* 9(1): 1-32.
- Ritz, Adrian. 2009. "Public Service Motivation and Organizational Performance in Swiss Federal Government." *International Review of Administrative Sciences* 75(1): 53-77.
- Rosenbaum, P. and Donald Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70(1): 40-55.
- Rotter, Julian. 1966. "Generalized Expectancies for Internal Versus External Control of Reinforcement." *Psychological Monographs: General and Applied* 80: 609-621.
- Scarpello, Vida and John P. Campbell. 1983. "Job Satisfaction: Are All the Parts There?" *Personnel Psychology* 36(3): 577-600.
- Schneider, Mark, Paul Teske, and Michael Mintrom. 1995. *Public Entrepreneurs: Agents for Change in American Government*. Princeton, NJ: Princeton University Press.
- Schuman, Howard and Stanley Presser. 1996. *Questions & Answers in Attitude Surveys*. Thousand Oaks, CA: Sage Publications, Inc.

- Sowa, Jessica E. and Sally Coleman Selden. 2003. "Administrative Discretion and Active Representation: An Expansion of the Theory of Representative Bureaucracy." *Public Administration Review* 63(6): 700-710.
- Whitford, Andrew B. 2002. "Bureaucratic Discretion, Agency Structure, and Democratic Responsiveness: The Case of the United States Attorneys." *Journal of Public Administration Research and Theory* 12(1): 3-27.
- Williams, Margaret L., Michael A. McDaniel, and Nhung T. Nguyen. 2006. "A Meta-analysis of the Antecedents and Consequences of Pay Level Satisfaction." *Journal of Applied Psychology* 91(2): 392-413.
- Wright, Bradley E., and Brian S. Davis. 2003. "Job Satisfaction in the Public Sector: The Role of Work Environment." *American Review of Public Administration* 33(1): 70-90.
- Yang, Kaifeng, and Anthony Kassekert. 2010. "Linking Management Reform with Employee Job Satisfaction: Evidence from Federal Agencies." *Journal of Public Administration Research and Theory* 20(2): 413-436.
- Yesilkagit, Kutsal, and Jørgen G. Christensen. 2010. "Institutional design and formal autonomy: Political versus historical and cultural explanations." *Journal of Public Administration Research and Theory* 20(1): 53-74.

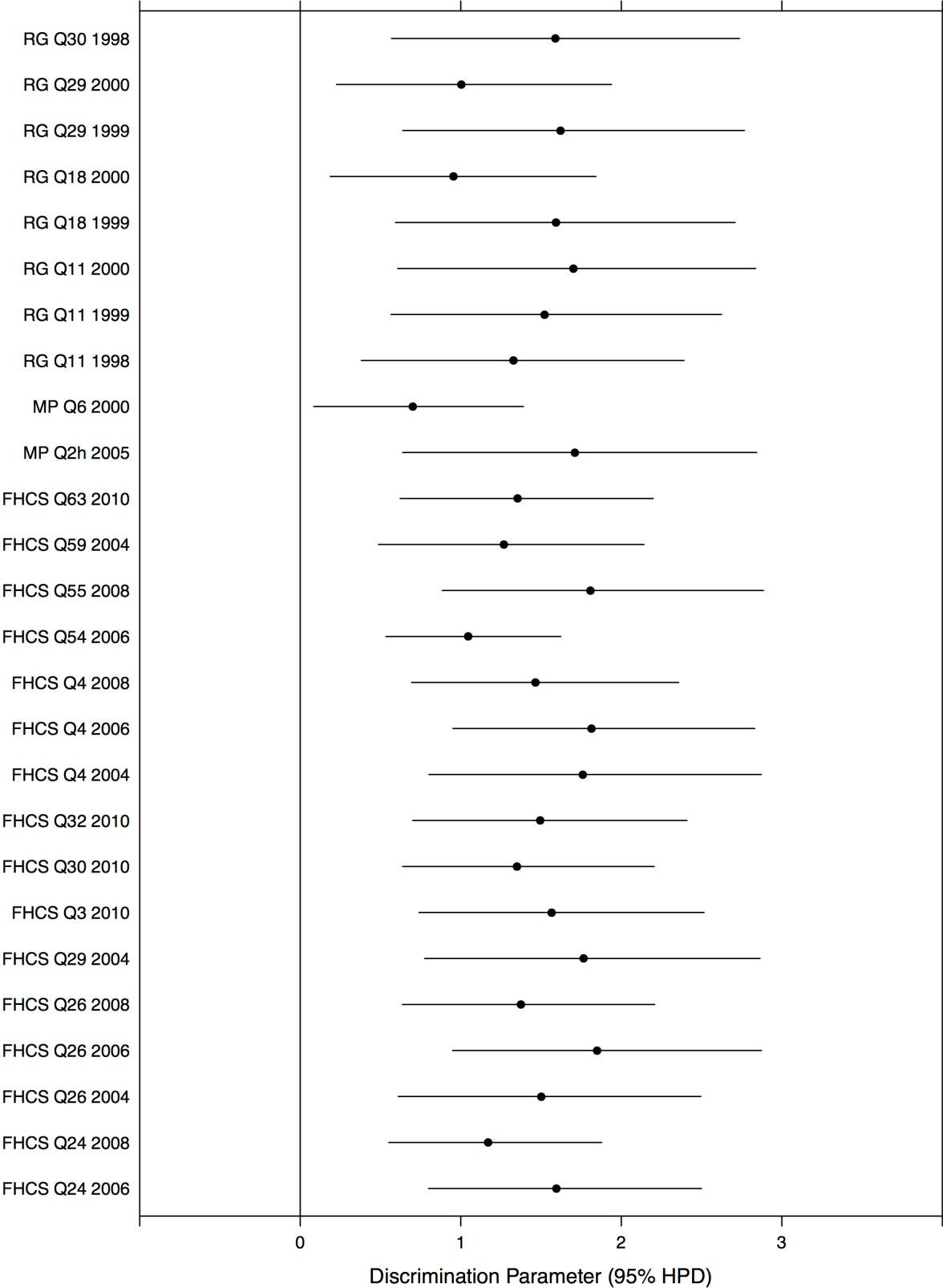


**Table 1: Questions Used for Latent Constructs**

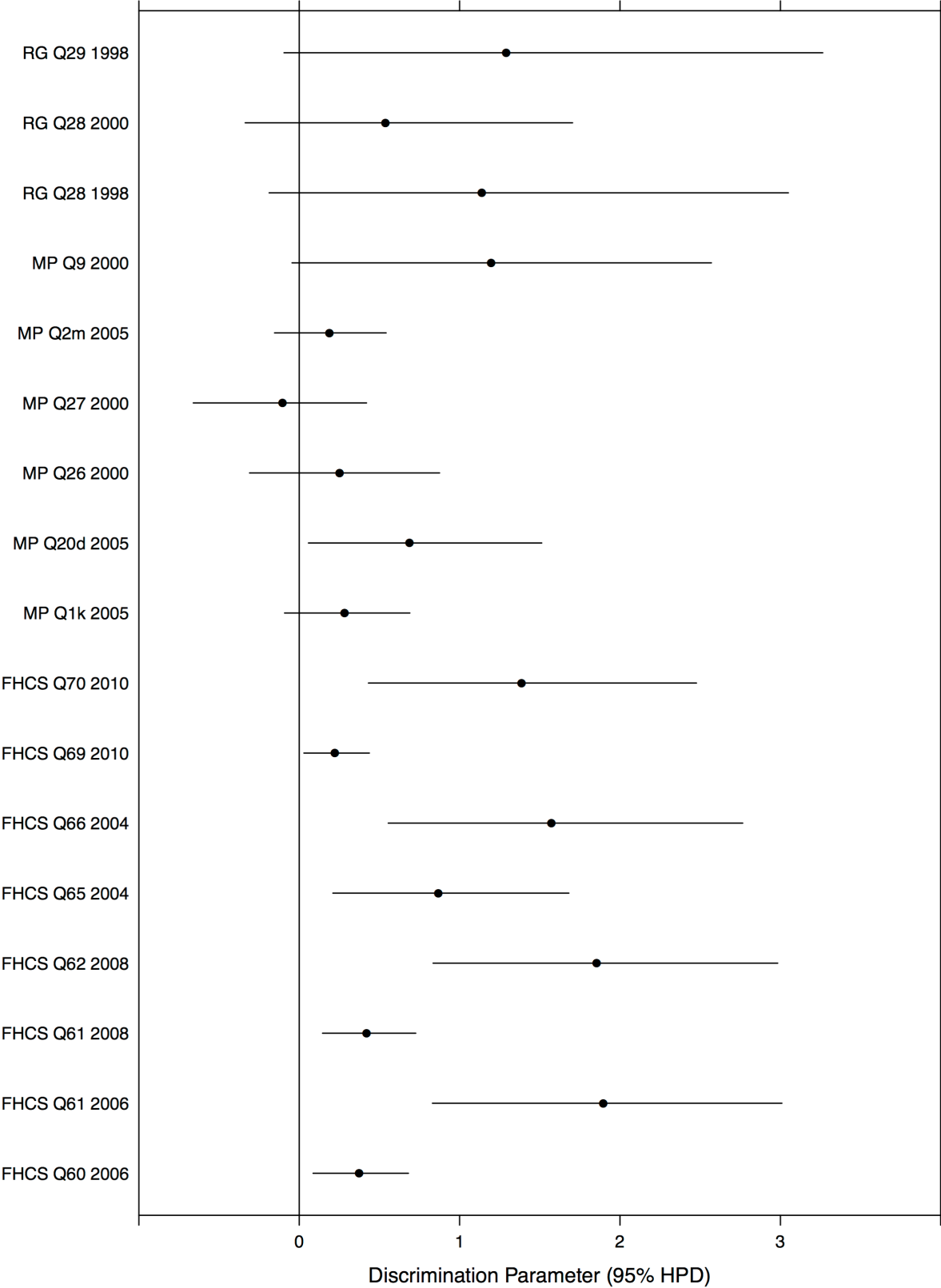
| Attribute                        | Question Text  | Fed. Human Capital Survey                        | Merit Principles       | Reinventing Government              |
|----------------------------------|--|--|------------------------|-------------------------------------|
|                                  |  | Year (Q.)  | Year (Q.)              | Year (Q.)                           |
| <b>Job Satisfaction</b>          | Considering everything, how satisfied are you with your job?                     | 2004 (65)<br>2006 (60)<br>2008 (61)<br>2010 (69) |                        | 1998 (29)<br>1999 (28)<br>2000 (28) |
|                                  | In general, I am satisfied with my job.  |  | 2000 (27)<br>2005 (2M) |                                     |
|                                  | I would recommend the government as a good place to work.                        |  | 2000 (9)<br>2005 (1K)  |                                     |
| <b>Compensation Satisfaction</b> | Considering everything, how satisfied are you with your pay?                     | 2004 (66)<br>2006 (61)<br>2008 (62)<br>2010 (70) |                        |                                     |
|                                  | Overall, I am satisfied with my current pay.                                     |  | 2000 (26)              |                                     |
|                                  | Overall, I am satisfied with my pay.   |  | 2005 (20D)             |                                     |
| <b>Autonomy</b>                  | I feel encouraged to come up with new and better ways of doing things.           | 2004 (4)<br>2006 (4)<br>2008 (4)<br>2010 (3)     |                        |                                     |
|                                  | Employees have a feeling of personal empowerment with respect to work processes. | 2004 (26)<br>2006 (24)<br>2008 (24)<br>2010 (30) |                        |                                     |
|                                  | Creativity and innovation are rewarded.  | 2004 (29)<br>2006 (26)<br>2008 (26)<br>2010 (32) |                        | 1998 (11)<br>1999 (11)<br>2000 (11) |
|                                  | How satisfied are you with decisions that affect your work?                      | 2004 (59)<br>2006 (54)<br>2008 (55)<br>2010 (63) |                        |                                     |
|                                  | How satisfied are you with your involvement in decisions that affect your work?  |  |                        | 1998 (30)<br>1999 (29)<br>2000 (29) |
|                                  | I have been given more flexibility in how I accomplish my work.                  |  | 2000(6)                |                                     |
|                                  | Creativity and innovation are important.   |  | 2005 (2H)              |                                     |

|                             |  |  |                        |                        |
|-----------------------------|--|--|------------------------|------------------------|
|                             | In the past two years, I have been given more flexibility in how I accomplish my work. |  |                        | 1999 (18)<br>2000 (18) |
| <b>Intrinsic Motivation</b> | My work gives me a feeling of personal accomplishment.                                 | 2004 (6)<br>2006 (5)<br>2008 (5)<br>2010 (4) |                        |                        |
|                             | The work I do is important.  | 2004 (6)<br>2006 (5)<br>2008 (5)<br>2010 (4) |                        |                        |
|                             | The work I do is meaningful to me.   |  | 2000 (10)<br>2005 (2K) |                        |

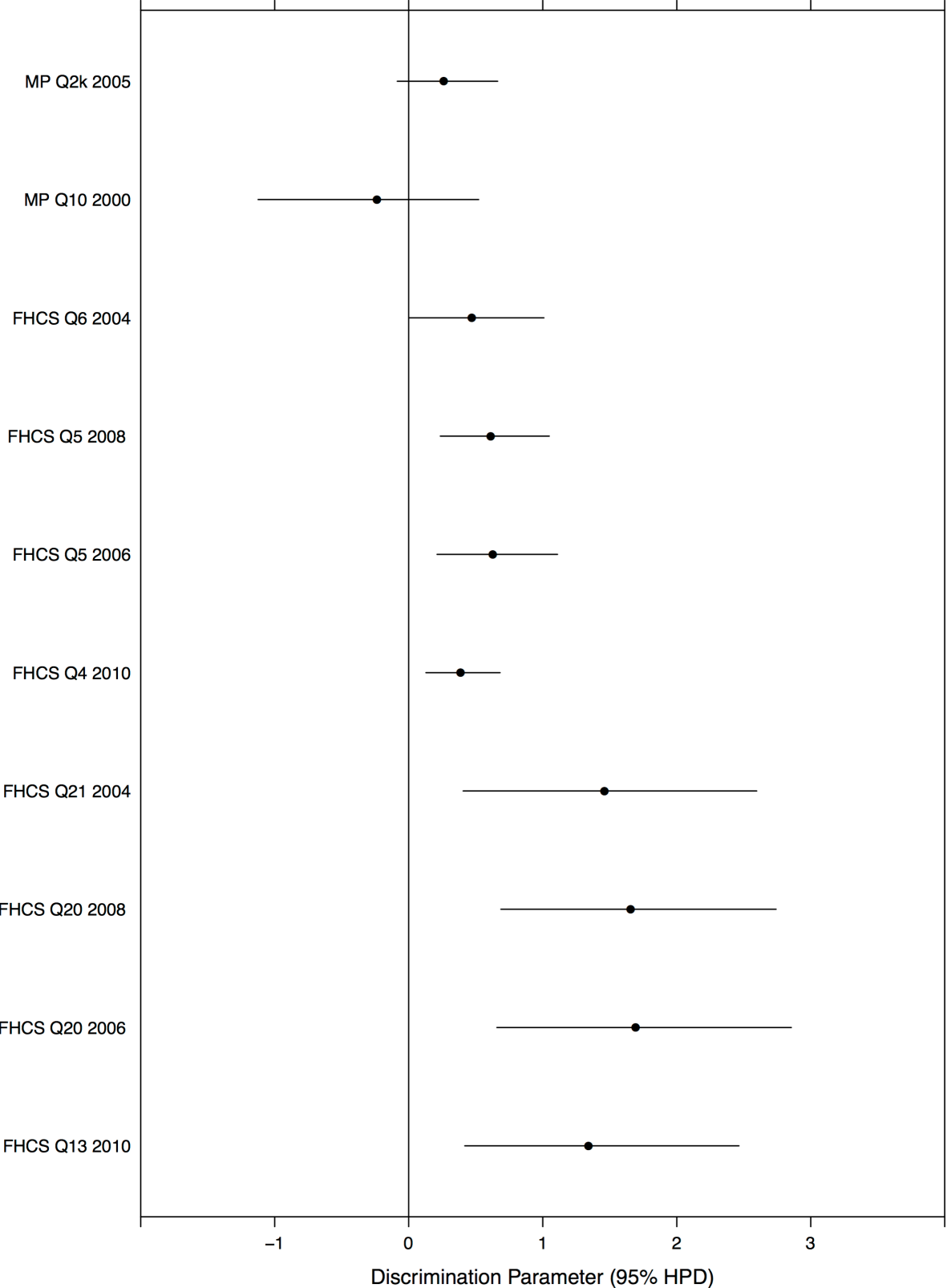
**Figure 1: Discrimination Parameters for Autonomy Measure**



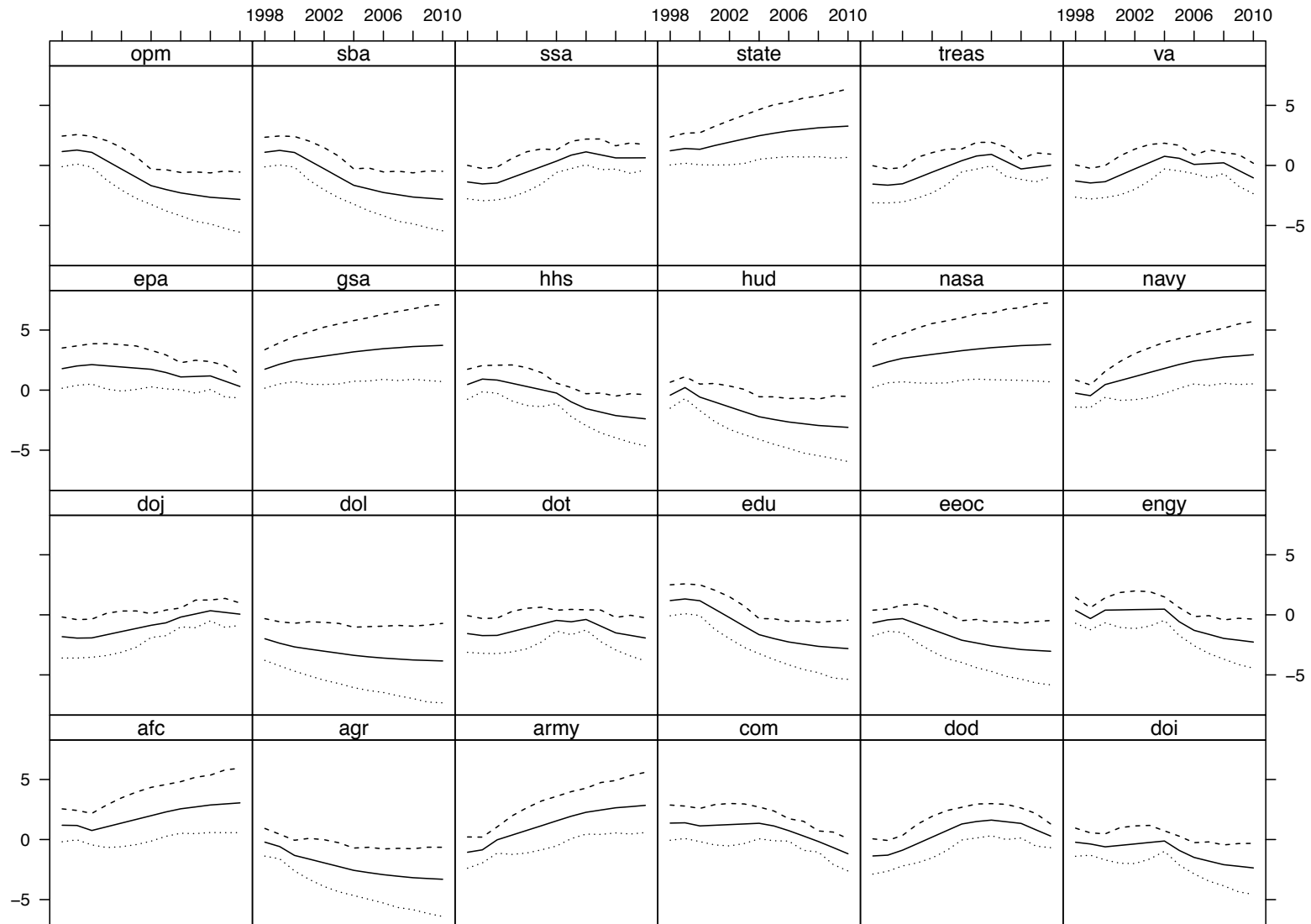
**Figure 2: Discrimination Parameters for Job Satisfaction Measure**



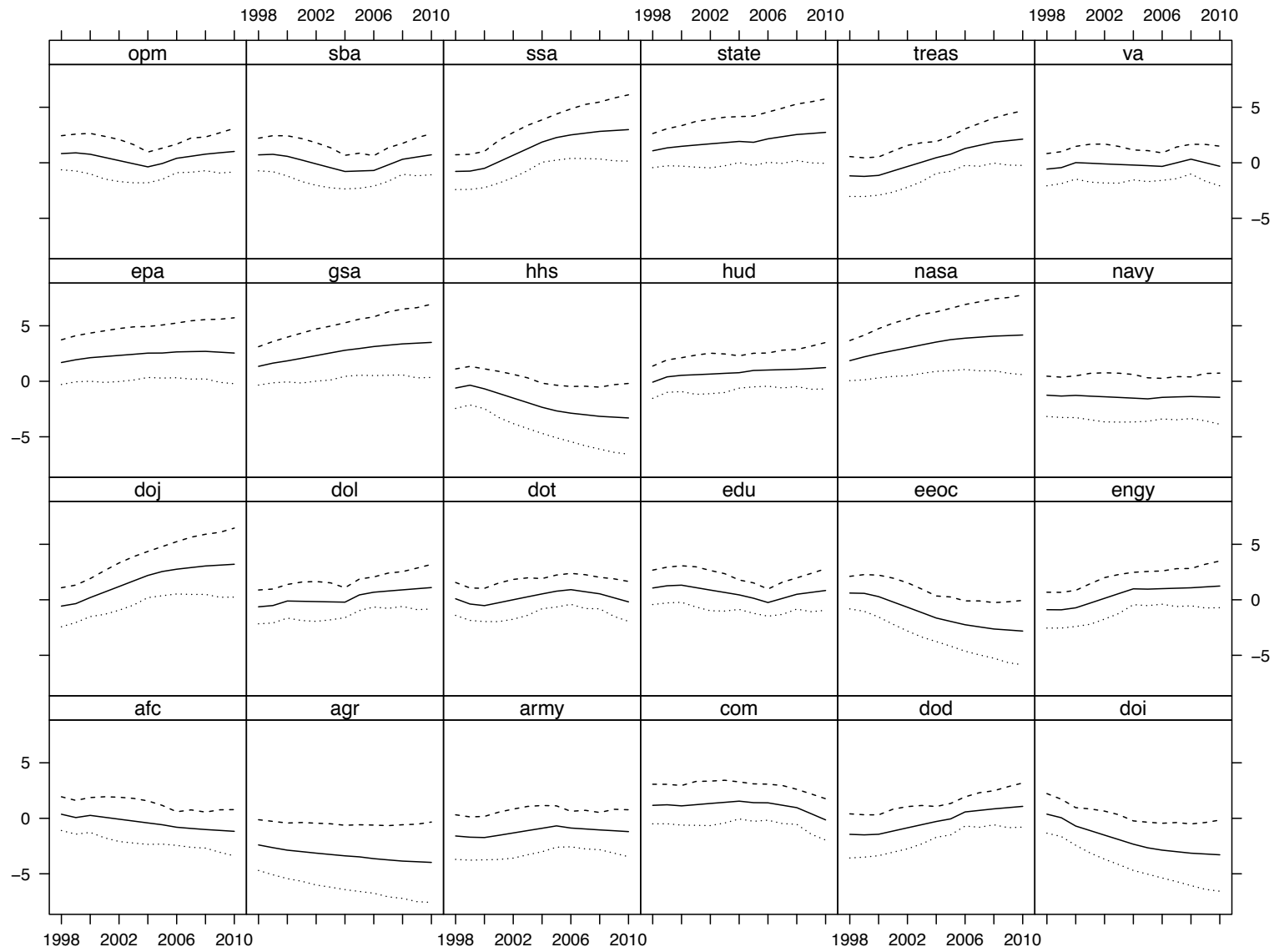
**Figure 3: Discrimination Parameters for Intrinsic Motivation Measure**



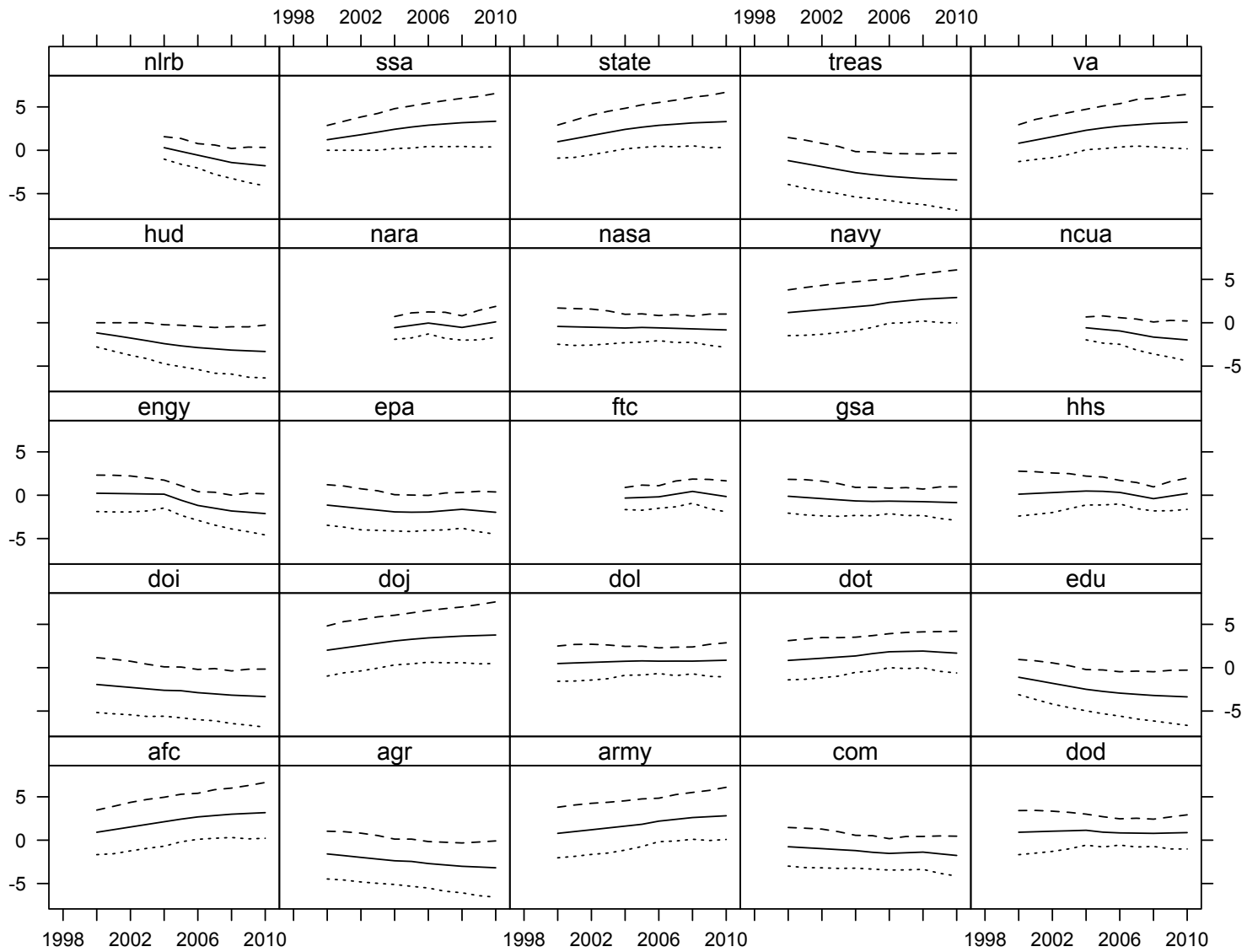
**Figure 4: Time Trends in Autonomy, 1998-2010**



**Figure 5: Time Trends in Job Satisfaction, 1998-2010**

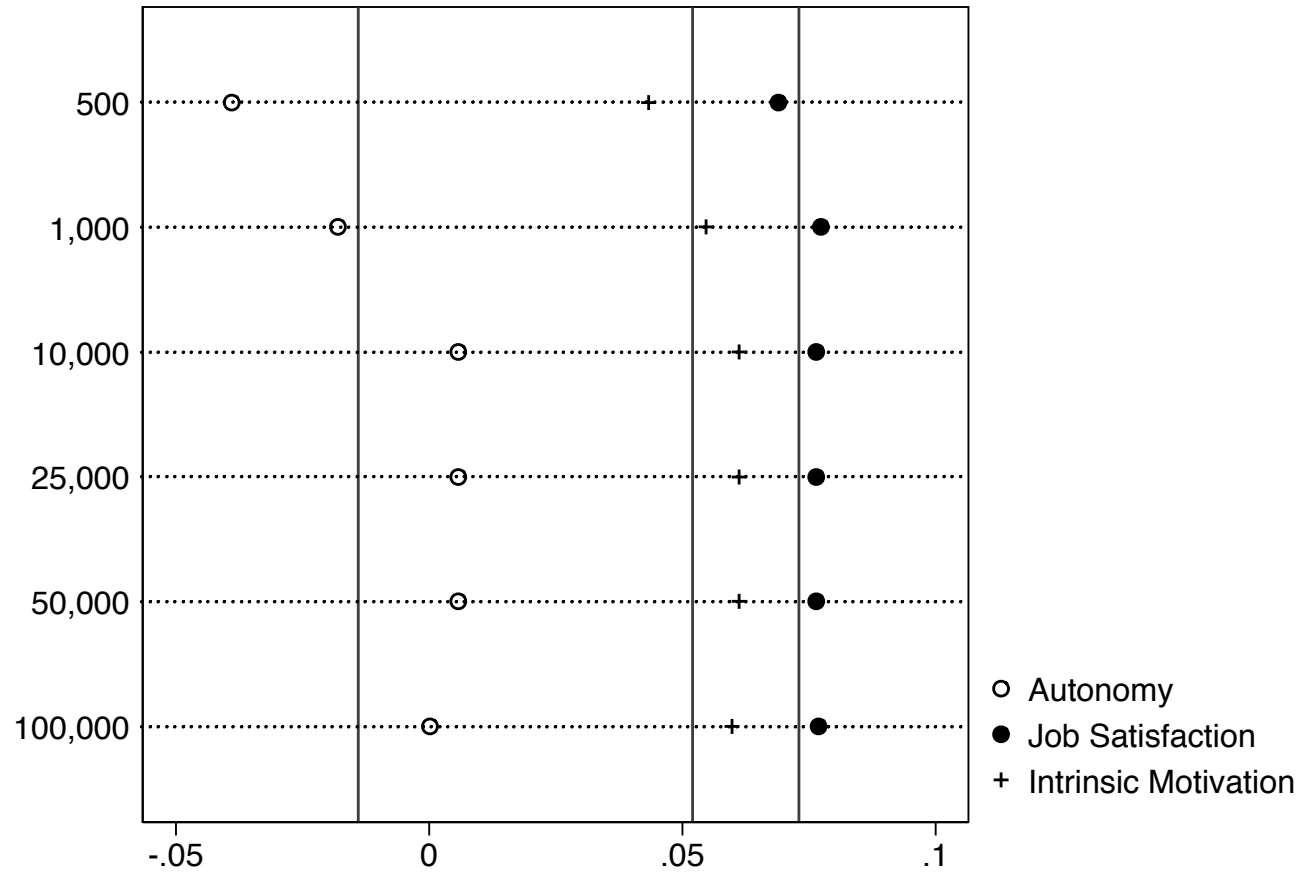


**Figure 6: Time Trends in Intrinsic Motivation, 2000-2010**



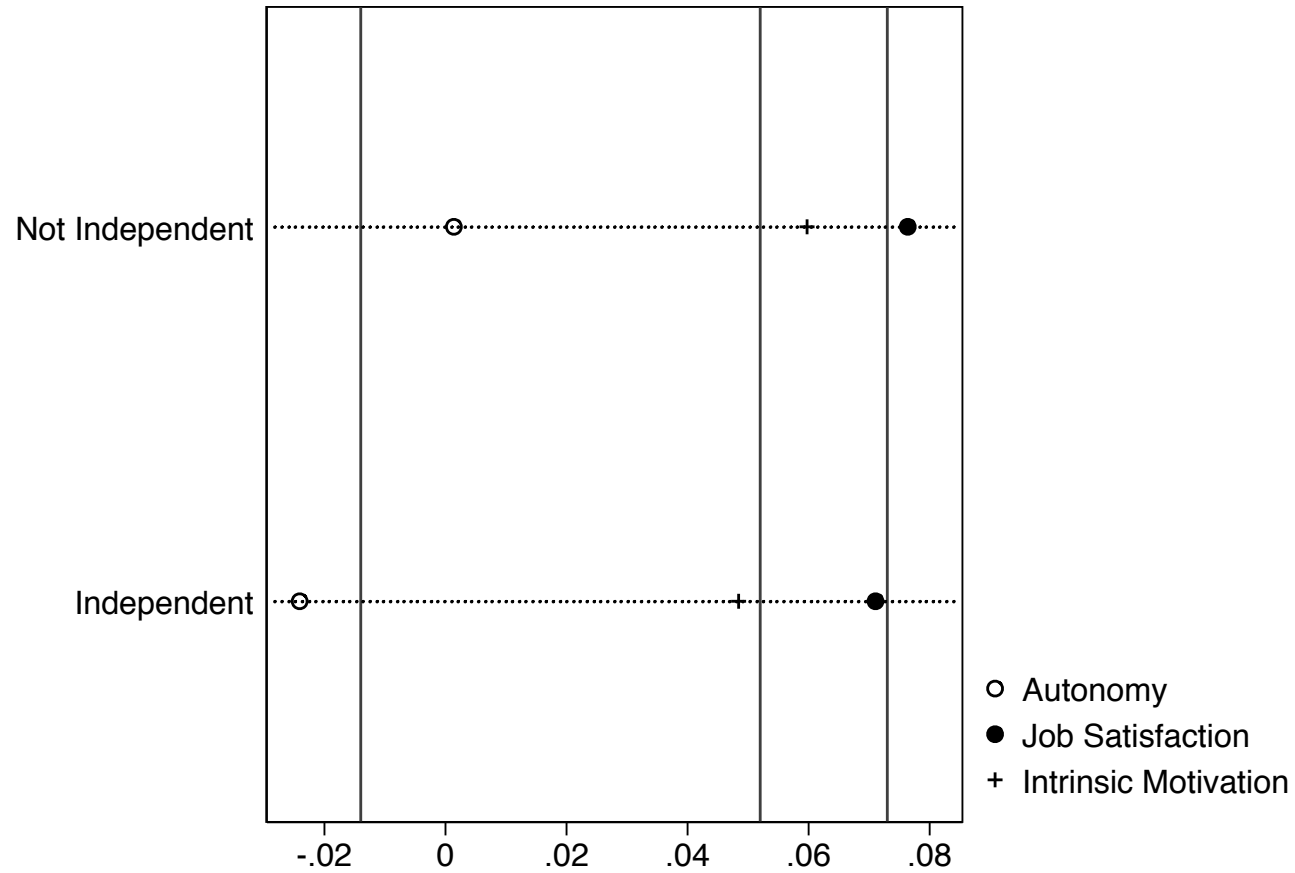


**Figure 7: Intertemporal Average Attribute Measures by Organization Size**



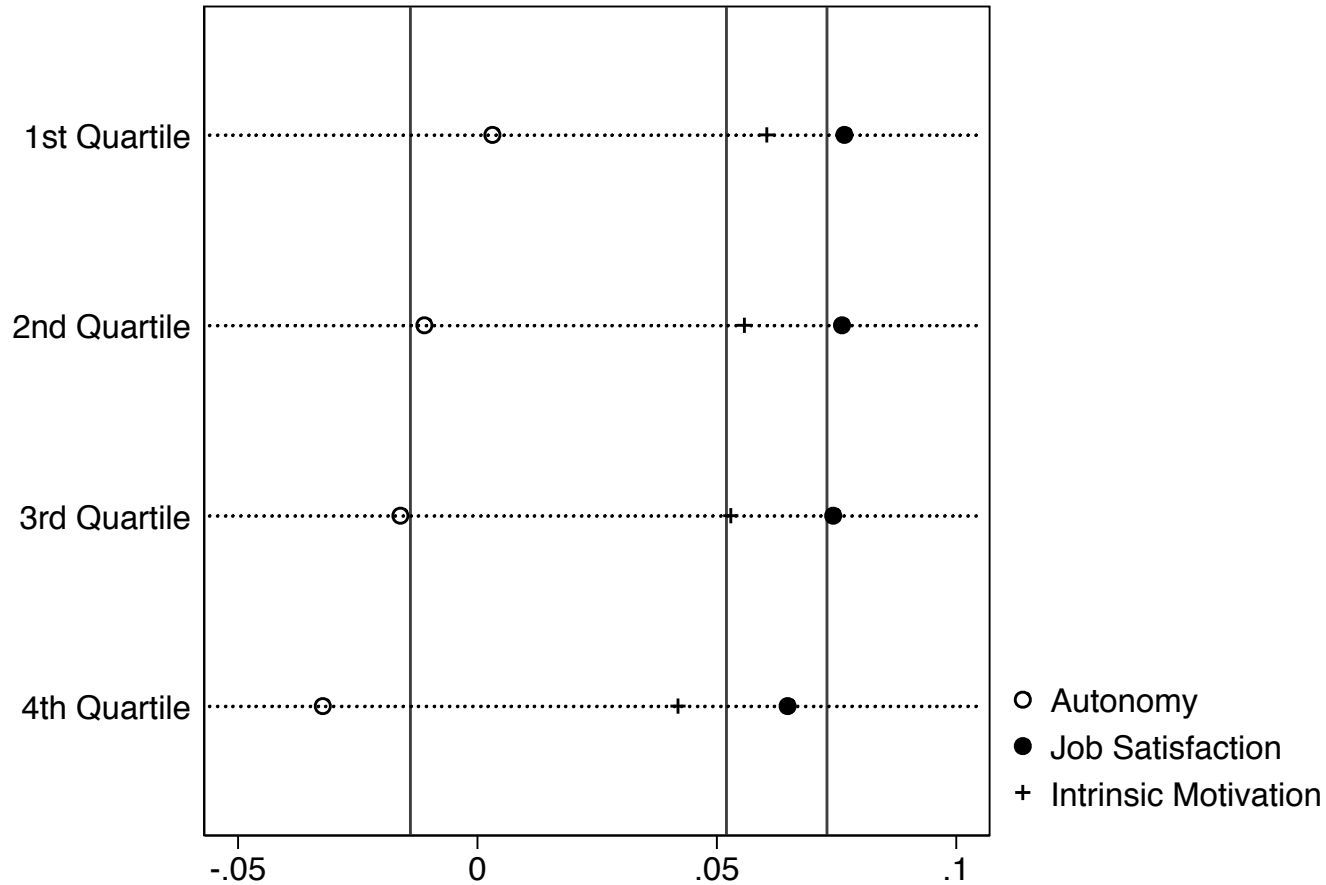
Note: Vertical lines represent average values of autonomy (-0.13), job satisfaction (0.073) and intrinsic motivation (0.052) over agencies and time. Employee numbers represent lower bounds of each range (i.e., 10,000 indicates average values for agencies employing between 1,001 and 10,000 staff).

**Figure 8: Intertemporal Average Attribute Measures by Agency Independence**



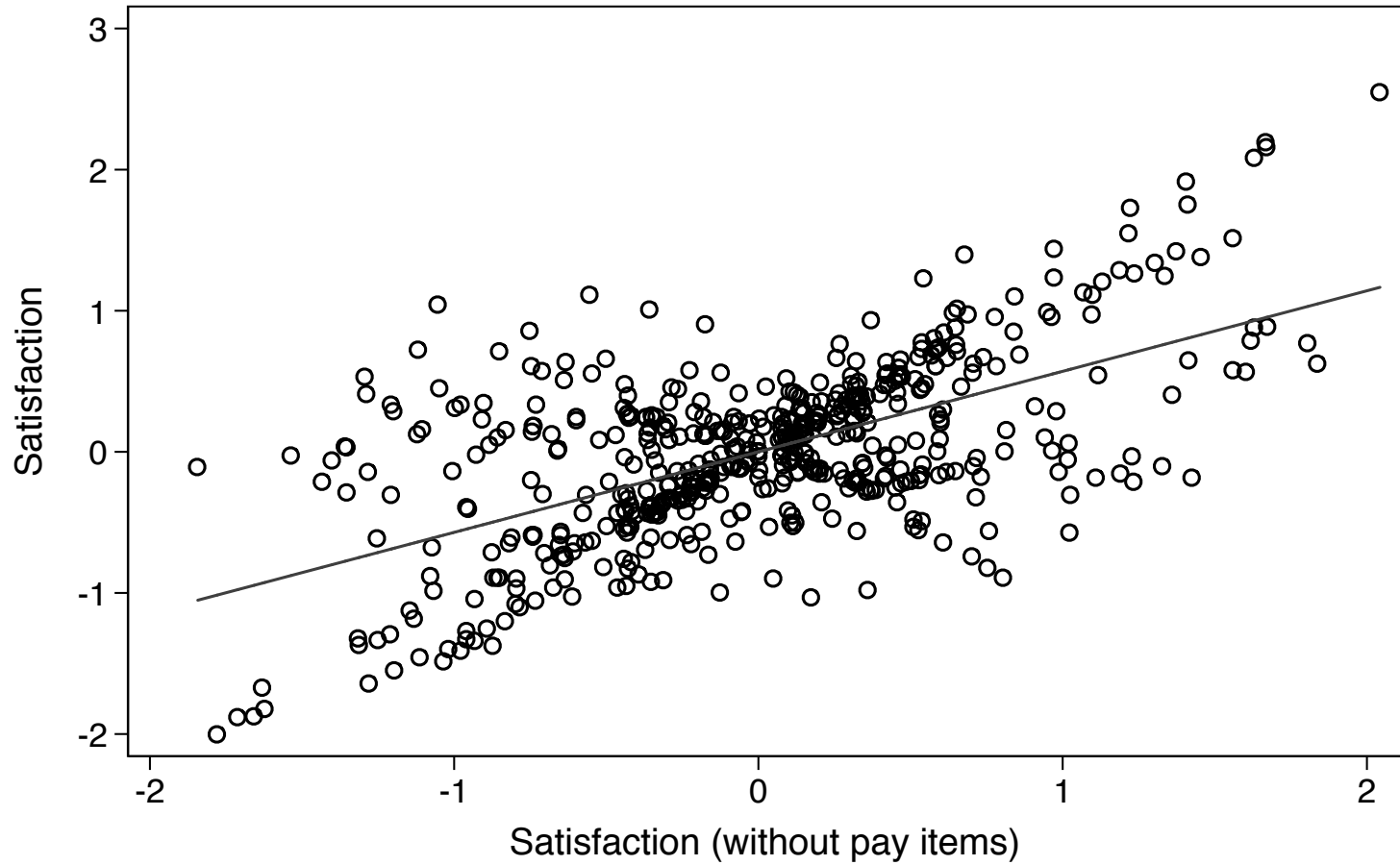
Note: Vertical lines represent average values of autonomy (-0.14), job satisfaction (0.073) and intrinsic motivation (0.052) over agencies and time.

**Figure 9: Intertemporal Average Attribute Measures by Percentage of Political Appointees**



Note: Vertical lines represent average values of autonomy (-0.14), job satisfaction (0.073) and intrinsic motivation (0.052) over agencies and time. Political appointees include Senate confirmable presidential appointments (PAS), Non-career Senior Executive Service, Schedule C excepted appointment, and other presidential and policy or supporting appointees not subject to Senate confirmation.

**Figure 10: Relationship Between Job Satisfaction Measures with and without Pay Items, Fixed-effects Regression.**



Coefficient = 0.571, Robust SE = 0.121, t = 4.75