Standard Length Article

Sociological Methodology
1–24

DOI: 10.1177/00811750231193641
http://sm.sagepub.com
Sage

Trend Analysis with Pooled Data from Different Survey Series: The Latent Attitude Method

Donghui Wang[1], Yu Xie[2,3] and Junming Huang[3]

Abstract

The use of pooled data from different repeated survey series to study long-term trends is handicapped by a measurement difficulty: different survey series often use different scales to measure the same attitude and thus generate scale-incomparable data. In this article, the authors propose the latent attitude method (LAM) to address this scale-incomparability problem, on the basis of the assumption that attitudes measured by ordinal categories reflect a latent attitude with cut points. The method extends the latent variable method in the case of a single survey series to the case of multiple survey series and leverages overlapping years for identification. The authors first assess the validity of the method with simulated data. The results show that the method yields accurate estimates of mean attitudes and cut point values. The authors then apply the method to an empirical study of Americans' attitudes toward China from 1974 to 2019.

Keywords

latent variable approach, ordinal scale, interval scale, incomparable scaling, pooled survey series, attitudes toward China

Public opinion researchers have long sought to understand the causes, trends, and consequences of societal attitudes. The literature cuts across a wide range of domains, including attitudes toward gender roles (Cotter, Hermsen, and Vanneman 2011; Scott, Alwin, and Braun 1996), marriage (Thornton 1985), race (Steeh and Schuman 1992), inequality (Osberg and Smeeding 2006), migrants (Semyonov, Raijman, and Gorodzeisky 2011), abortion (Carter, Carter, and Dodge 2009; Mouw and Sobel 2001), and political affairs (DiMaggio, Evans, and Bryson 1996). The advancements in attitude research have largely been facilitated by the emergence of large-scale repeated cross-sectional attitude surveys (hereafter "attitude survey series"). Nationally representative surveys, such as the General Social Survey (GSS), first launched in 1972, have captured the ebb and flow of Americans' attitudes toward hundreds of issues,

[1]Center for Population and Development Studies, School of Sociology and Population Studies, Renmin University of China, Beijing, China
[2]Department of Sociology, Princeton University, Princeton, NJ, USA
[3]Paul and Marcia Wythes Center on Contemporary China, Princeton University, Princeton, NJ, USA

Corresponding Author:
Donghui Wang, Center for Population and Development Studies, School of Sociology and Population Studies, Renmin University of China, 601 W. Chongde Building, 59 Zhongguancun Street, Beijing, 100872, China
Email: donghuiwangruc2020@ruc.edu.cn

ranging from government spending to income inequality and climate change (National Opinion Research Center 2020). Over the past several decades, the number of attitude survey series has grown exponentially because of computing advancements that greatly reduced the costs associated with survey design, conducting interviews, data storage, and analysis. At present, millions of individuals are surveyed every year by academic institutions, commercial companies, and news media organizations regarding their attitudes toward various topics. These survey series can be used to track changes in public attitudes toward various aspects of contemporary social life.

The accumulation of a large amount of attitude survey series has motivated scholars to move beyond single-survey analysis. Yet the use of such data is handicapped by a measurement difficulty: different survey series use different scales to measure the same attitudes and thus generate scale-incomparable data. For example, two surveys could measure the same attitude with different response scales: one may use a 4-point Likert-type scale and the other a 5-point Likert-type scale, making pooling the two a difficult task. The scale-incomparability problem becomes increasingly common in empirical analysis, as new surveys tend to borrow wordings from old ones. As a result, a large number of survey items use similar wordings. For example, one study that investigated the diversity of happiness measurements revealed 1,250 alternative measurements of the concept “happiness” across existing surveys. The response scales of these happiness measures can range from a simple binary yes/no scale to an 11-point scale (Kalmijn, Arends, and Veenhoven 2011).

The scale-incomparability problem becomes acute when a series of cross-sectional surveys are implemented for a limited duration. Pooling data from different survey series that took place across different periods would allow researchers to study societal attitude changes within a much wider observation window. This ability to track attitudinal change over a long period provides opportunities to address many important sociological inquiries. For example, there is a long-standing debate about whether Americans’ political attitudes have polarized over time (DiMaggio et al. 1996; Kiley and Vaisey 2020; Mouw and Sobel 2001). To study the problem, it would be helpful to have independent measures of important political attitudes over a long period. Yet scholars are often forced to use single survey series for limited time periods because of incomparable scales across survey series.

In this article, we propose a new method for estimating attitudinal trends by pooling data from different survey series. The method, which we call the latent attitude method (LAM), addresses the scale-incomparability problem when pooling multiple survey series. Our method contributes to the research on attitudes in two important ways. First, our method extends the existing latent variable framework to the case of pooled survey series, enabling us to separate the measurement model and theoretical construct model. The extant methods that measure attitudinal change over time, such as the proportional method and integer score method, evaluate attitudinal change on the basis of observed marginal distributions of survey responses. Yet given that attitudes are more difficult to measure than objective variables, it is necessary to have models that specify the linkage between the observed variables and individuals’ true underlying attitudes. Building on the latent variable framework, we construct an individual behavior model using an ordered probit model with unknown mean and survey-specific cut points. The model is conducive to flexible adaptations of theoretical determinants of attitudes given data availability.

Second, in pooling single attitude measures across surveys, we draw insight from the comparable distribution condition in survey design evaluation (Couper 2011). We leverage situations where two survey series overlap in a given year. This allows us to assume that two measurement scales, which represent the same underlying attitude, are given randomly to two independent samples drawn from the same population. In other words, we assume these two survey series share the same statistical properties on the latent attitude in the overlapping year. In so doing, we can reduce the number of unknown parameters to be identified and estimate the best-fit parameters with maximum likelihood. Our model uses full distribution of response categories, thus yielding more accurate results, whereas the extant methods of pooling survey items, such as the dyad ratio method, require dichotomizing the survey items.

TWO MEASUREMENT PRINCIPLES

We start with a simple case in which respondents' attitudes toward a particular social issue are captured by a single survey series. That is, a survey instrument has been administered to different, yet population-representative, samples of respondents at different time points by the same survey institution. Using data from a single survey series to quantify trends in social attitudes is one of the most widely adopted practices in social science. This type of data meets two important principles: *measurement consistency* and *sample representativeness*. In this section, we introduce these two principles in the context of a single-survey instrument; in the next section, we show how they can be fruitfully applied in the context of pooling different data sources for trend analysis.

Measurement consistency dictates that the survey instrument, including question wording and response options, be consistent over time. As Smith (2005) summarized conventional wisdom in survey research, "the way to measure change is not to change the measure" (p. 1). In practice, survey researchers ensure that question wording and responses are consistent over time, enabling a survey series to establish trends in the same attitude over time. If a change is made to the questionnaire or response options, it is recommended that researchers split the data, treating before and after the change as two separate time series (Mouw and Sobel 2001). Many large-scale survey series (e.g., the GSS) have been implemented under this principle.

Sample representativeness dictates that a survey sample be representative of the underlying population at each time point at which it is administered. This principle is ensured by carefully designed and implemented sampling procedures. Any changes made to the sampling process will also yield a split in the survey series. This principle may be straightforward, but it serves as a basic requirement for survey design evaluation, such as testing changes in survey mode (Couper 2011) or measurement (Jaeger 1997). For example, survey methodologists often need to conduct random experiments to evaluate the performances of different survey modes, such as Web surveys versus telephone surveys (Fricker et al. 2005). Random assignment ensures the statistical properties of the variable under study are the same regardless of the survey mode.

These two principles ensure that the same attitudinal concept is measured consistently over time. But what can we do when the first principle is violated? Is there anything we can do methodologically when we attempt to pool data from survey series with different instrument scales? In the next section, we discuss the two most conventional approaches to pooling survey series before proposing our own method.

METHODS OF POOLING SURVEY SERIES

Unlike objective quantities such as income or wealth, researchers of attitudinal change face a particular measurement challenge: individuals' attitudes do not have intrinsic scales and are often operationalized as ordinal variables in surveys. For example, a researcher may use a four-level Likert-type scale such as "completely disagree," "disagree," "agree," and "completely agree" to denote respondents' attitudes toward a statement. The limitation of ordinal variables is well recognized: the numerical response represents a ranking order on the underlying attribute instead of the actual quantity. Thus, it is recommended that researchers use models that account for the ordinal nature of attitudinal data. The extant models include, but are not limited to, the ordered logit model, ordered probit model, log-linear model, and scaling method (Agresti 2002; Powers and Xie 2008).

Although the ordinal nature of attitude variables is generally understood, most research based on trend analysis of social attitudes does not consider this. In practice, researchers in trend analysis tend to apply two heuristic approaches: the *integer-scoring approach* and the *proportional approach*. The popularity of these two approaches has grown out of convenience, yet their underlying assumptions are rarely discussed. In the following, we briefly discuss the advantages of the two approaches, and the potential limitations when applying them to the context of pooling survey series.

The integer-scoring approach assigns consecutive integers to represent rank order (Powers and Xie 2008). It can be seen as a simplified form of scaling. For example, the researcher may assign the values 1, 2, 3, and 4 to a four-level Likert-type scale. A key advantage of this approach is that the researcher can easily calculate summary statistics that describe distributions of responses. For example, to examine whether Americans' attitudes toward abortion have polarized, DiMaggio et al. (1996) calculated the mean, variance, and kurtosis of items on Americans' opinions on domestic social issues in the GSS and the National Election Study. The core of their analysis was treating ordinal variables as if they were interval variables and thus deriving summary statistics.

Although it is plausible to apply the integer-scoring approach to pooled survey series, the method is rarely used in practice. This is because a critical assumption of the integer-scoring approach is that the distance between two adjacent categories is equal. However, the underlying true distribution of an attitude variable is unknown, so this assumption hardly holds in practice. The controversial nature of the equal-interval assumption is well documented (Jamieson 2004; Wu and Leung 2017). Thus, even if the two survey series that measure the same attitude adopt the same number of

Table 1. A Hypothetical Example of Aggregate Attitude Change

		S_1	S_2
Response	Assigned Scale	Number of Respondents	
Completely disagree	1	20	10
Disagree	2	20	30
Neither agree nor disagree	3	20	20
Agree	4	20	30
Completely agree	5	20	10
Sample size		100	100
Sample mean		3	3

response scales, it is still inappropriate to directly pool them together and calculate summary statistics such as mean or variance for trend analysis.

In the proportional approach, researchers harmonize responses into fewer categories, usually two, to show changes in percentages of approval or disapproval toward a statement at each time point. Because the approach is easy to implement, institutions such as Pew and Gallup customarily use the proportional approach to report attitudinal change over time. This approach is also widely used in academic papers to explore trends in a particular social attitude. Examples can be found in the study of trends in attitudes toward family-related issues (Thornton and Young-DeMarco 2001) and toward mass incarceration (Duxbury 2021).

Sobel (1997) identified two major problems with the proportional approach. First, there is an obvious loss of information when responses are collapsed. This is because original ordinal responses contain two dimensions: the direction of the attitude (e.g., "agree" vs. "disagree") and the intensity of the attitude (e.g., "completely [dis]agree" vs. "[dis]agree"). Collapsing responses allows the researcher to track the direction of the attitude at the expense of losing information on attitude intensity. Second, if response categories are odd in number, it is unclear whether responses should be collapsed below or above the midpoint. To illustrate this problem in the case of pooled survey series, Table 1 gives a hypothetical example where attitudes toward a particular social issue are measured in two surveys, S_1 and S_2. The five response categories are assigned the values 1, 2, 3, 4, and 5. From S_1 to S_2, there is a clear shift from the extreme categories toward the middle categories. If one were to collapse the first two and the last two categories, this would lead one to conclude that no changes in responses have taken place.

Besides these two conventional approaches, past studies have attempted to measure attitudinal change in a number of other ways. For example, scholars have used structural equation modeling (SEM) to evaluate the stability of individual attitudes over time (Inglehart 1985; Judd and Milburn 1980). The strength of SEM is that researchers can systematically separate indicators of the construct from extraneous errors. However, the method is best used to track individual attitudinal change with panel data. SEM is limited in its ability to address scale-incomparability problems in the context of pooled survey series, because the method requires scales to be measured consistently over time. Scholars have also developed methods to pool indicators among multiple surveys. For example, Stimson (2018) proposed the dyad ratios algorithm, which is widely used to estimate changes in public opinion. The strength of the method is that it can be applied when survey items are sparse across time (Stimson 2018). However, its robustness is questioned, as the first step of the method requires creating binary indicators from survey marginals. Researchers obtain different results when recoding survey indicators differently (McGann 2014).

THE LAM

In this section, we formally introduce the LAM. Given the myriad meanings of the word "attitude," it is first necessary to provide conceptual clarity of what we mean by "latent attitude." We embrace the commonly adopted definition that an attitude is a latent predisposition to respond to a given topic (Alwin and Krosnick 1991; Oskamp and Schultz 2004). There is an enduring debate on whether individuals hold consistent and stable predispositions, or whether they just give random answers in interviews (Achen 1975; Alwin and Krosnick 1991; Converse 1964; Inglehart 1985; Judd and Milburn 1980; Kiley and Vaisey 2020). Some scholars argue that most people do not hold meaningful attitudes regarding important social and political issues, and they construct their responses on the fly in interviews (Converse 1964). Others argue that people do have meaningful predispositions, and the major problem is that survey questions measure them imperfectly (Achen 1975; Inglehart 1985).

Regardless of the nature of attitude, there is consensus that a model of attitudes should be able to specify individuals' true attitudes and observed survey responses. We thus model an individual's attitude as a latent variable that is inferred from observed ordered categories. The idea of viewing an attitude as a latent variable is not new (McKelvey and Zavoina 1975). As a general framework, the latent variable approach allows us to infer true attitudes from ordinal measurement. In our analysis, we use an ordered probit model, a type of cumulative categorical model. It is important to note, however, there are other classes of ordinal models, such as the continuation model (Fienberg 1980) and adjacent category model (Goodman 1983), both based on the logit link. The adjacent category model uses the form $\log[P(Y = j|x)/P(Y = j + 1|x)]$, and the continuation-ratio model uses form $\log[P(Y = j)/P(Y \geq j + 1)]$. There are specification differences between the cumulative model we use and the above-mentioned two alternative models (Bürkner and Vuorre 2019). We chose the cumulative model for simplicity under the assumption there is only one underlying latent variable to the response categories.

A Single Survey Series

In the case of single survey series, a survey organization draws a fresh sample from the underlying population at multiple time points with a consistent survey instrument to assess population-level attitude changes. We model this process at the individual level from the latent variable perspective. Let y_{it}^* denote individual i's latent attitude at time t, following a normal distribution with a mean μ_t denoting the population mean at time

t, as well as a uniform deviation. H_{it} denotes i's idiosyncratic deviation from the mean. We can specify a simple latent variable model as follows:

$$y_{it}^{*} = \mu_t + \eta_{it}, \eta_{it} \sim N(0, 1). \tag{1}$$

Equation (1) shows an individual's attitude can be decomposed into two components: a structural component and a random component. In our example, the structural component specifies the population mean μ_t, which is our primary quantity of interest. For simplicity, we only specify the population mean for the structural part in equation (1), as we focus here on trends. We can easily incorporate other explanatory variables as determinants of the population mean μ_t (Powers and Xie 2000).

The random component η_{it} represents an error term that we assume to be independent and follows a standard normal distribution with mean of 0 and variance of 1. Note that η_{it} can also be specified in other distribution forms, such as the logistic distribution (Agresti 2002; Powers and Xie 2000). The subscript i denotes the randomness is associated with respondents' idiosyncrasies. For instance, some respondents may hold no opinion toward the given topic, thus giving random answers to survey questions. Or respondents might have opinions on a given topic, but these opinions are conflicting or ambiguous and thus difficult to specify. As noted by prior research, when individuals have ambiguous internal cues, random error is introduced when forcing them to specify a single point on an attitude continuum (Alwin and Krosnick 1991).

Next, we introduce the measurement model. Although an individual's true attitude y_{it}^{*} is not directly observable, it can be inferred from a measurement model for an observed ordinal variable. We assume an attitude is measured with k-level ordinal categories. In practice, this is often achieved by asking k-category Likert-type scale survey questions. We denote an individual's observed attitude levels in year t as y_{it}. The relationship between y_{it} and y_{it}^{*} can be expressed by the following measurement equation:

$$y_{it} = k \text{ if } \tau_{k-1} \leq y_{it}^{*} \leq \tau_k, \tag{2}$$

where $\tau_0 \ldots \tau_k$ denote cut points that partition the continuous latent variable into k ordinal categories, with the normalization conditions $\tau_0 = -\infty$ and $\tau_k = +\infty$. Equation (2) suggests that if the attitude is measured with k ordinal categories, there are $k - 1$ cut points that partition the underlying continuous real line into k segments that correspond to the ordinal categories. For example, if the attitude is measured with a five-level Likert-type scale and $y_{it} = 3$, it corresponds to a segment in the real line where $\tau_2 \leq y_{it}^{*} \leq \tau_3$.In the case of a single survey series, the cut points are unknown parameters but are assumed to be time-invariant (i.e., without subscript t), attributable to the principle of measurement consistency discussed earlier, so as to estimate changes in the mean of the attitude, μ_t. Without this assumption, μ_t would be indeterminate and thus not comparable. Substituting equation (1) for equation (2), the probability of the observing level k can be expressed as follows:

$$P(y_{it} = k) = P(\tau_{k-1} \leq y_{it}^{*} \leq \tau_k) = \Phi(\tau_k - \mu_t) - \Phi(\tau_{k-1} - \mu_t). \tag{3}$$

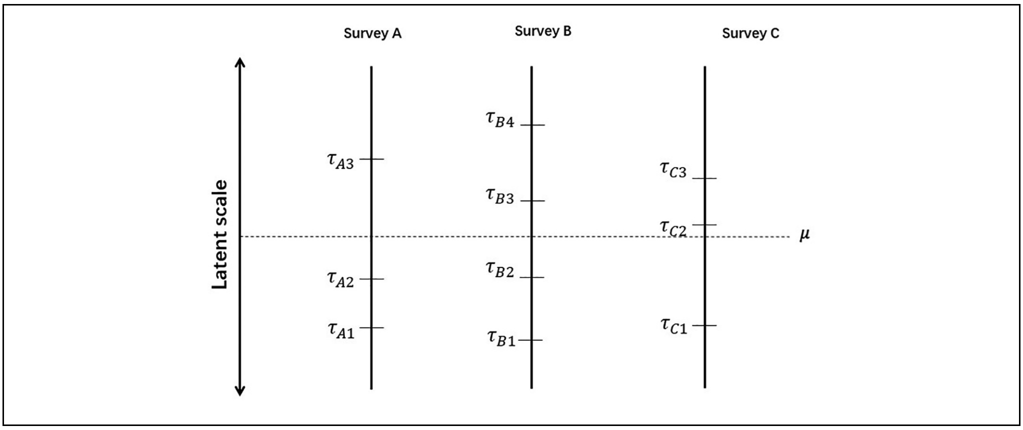

Figure 1. An illustration of two types of scale differences.

In equation (3), Φ is the Gaussian cumulative distribution function. The two types of unknown parameters, the time-dependent mean μ_t and the time-invariant cut points $_\tau$, can be estimated with maximum likelihood.

The model is essentially an application of the ordered probit model to the study of attitude trends. Compared with the integer-scoring approach, the latent variable approach does not assume equal intervals between adjacent categories. It assumes cut points to be unknown (albeit time-invariant) parameters estimable through an iterative procedure by maximizing the likelihood function. Compared with the proportional approach, it uses full response distribution of ordinal attitude measures and thus avoids loss of information.

Pooled Survey Series

In the case of multiple survey series, one or several organizations draw independent samples of the same population at different time points and measure attitudes with different survey instruments. Pooling multiple survey series to draw inferences about trends in attitudes is a challenging task, because the situation violates the principle of measurement consistency. Even when different survey series measure the same attitude in the same population, they may differ in two important ways. First, they may use a different number of categories as response options. Viewed from the latent variable perspective, response-scale differences are equivalent to the differences in the number of cut points. As illustrated in Figure 1, survey A has a four-level response scale, with three cut points partitioning the latent variable distribution on the surveyed sample. Survey B has a five-level response scale with four cut points. It is apparent that if two surveys have different response scales, there is no simple way to pool data from the surveys.

Second, there may be wording differences in response categories across survey series. Survey response patterns are strongly influenced by the features of survey instruments. A slight difference in wording in instruments can yield vast differences in response patterns (Schwarz 1999). We emphasize that wording differences are different from response-scale differences in that the latter are explicitly defined in terms of the number of response categories, whereas the former are often not. Different survey instruments can yield differences in response patterns even if they use exactly the same number of response scales. Schwarz et al.'s (1991) survey experiment provides a classic example: they presented a success-in-life question with an 11-point rating scale ranging either from 0 ("not at all successful") to 10 ("extremely successful"), or from −5 ("not at all successful") to 5 ("extremely successful"). Respondents reacted differently to these two scales: 34 percent of respondents endorsed a value between −5 and 0 on the −5 to 5 scale, yet only 13 percent endorsed values between 0 and 5 on the 0 to 10 scale. Using the language of the latent variable method, we can easily interpret the finding as attributable to different distances between response scales under the two alternative instruments (Clogg 1979). As illustrated in Figure 1, surveys A and C both use four-level response scales, but they differ in the true distances between adjacent cut points. Although such differences are not directly observable, we can rely on a model to uncover the differences. In empirical studies, wording differences also create incomparability issues that prevent researchers from simply pooling data from different surveys.

The proposed LAM is well suited to address the issues of measurement inconsistency in both forms. The key insight of the method is the leveraging of the opportunity presented when two or more survey series overlap in a given year, assuming the distributional properties are the same in the overlapping year. For example, let us imagine two survey series on Americans' life satisfaction: the first was implemented annually from 1990 to 1995, and the second annually from 1995 to 2000. In 1995, when the two survey series were both implemented, they measure the same underlying latent variable because they fulfill the principle of sample representativeness in that year, as the target population of both surveys is the same. The only difference is how survey-specific cut points partition the distributions of the underlying latent life satisfaction. In other words, a scale inconsistency results from either the number of response categories or differences in wording, so the cut points are survey specific. As we will discuss in the next section, the best-fit parameters of the population-level mean and the survey-specific cutoffs can be estimated by maximizing the joint likelihood function across the surveys. In this way, we can pool data collected by the two survey instruments and gain a cohesive understanding of the changes in Americans' mean life satisfaction from 1990 to 2000.

The method can be formally expressed as follows. To begin, equation (1) still holds. Let y^*_{it} represent a given individual's latent attitude in year t. y^*_{it} is expressed as a linear combination of a structural component μ_t and a random component η_{it} that follows standard normal distribution.

Next, consider the measurement model. We extend the single survey series model by specifying survey-specific cut points. Assume there are multiple Likert-type survey series, indexed by q, where $q = 1, 2, \ldots Q$. For each survey series q, we further denote that it has $K_q \in (2 \ldots k_q \ldots K_q)$ levels of ordered response categories. Let y_{itq} denote individual i's observed response category to survey q, so the relationship between latent variable y_{it}^* and observed response categories y_{itq} can be expressed as follows:

$$y_{itq} = k_q \Leftrightarrow \tau_{q,k-1} \leq y_{it}^* \leq \tau_{q,k} \ . \tag{4}$$

Equation (4) is an extension of equation (2). The addition of subscript q in the cut point parameters $\tau_{q,\cdot}$ signifies the cut points are survey specific. The survey-specific cut points account for survey-level design effects. Similar to equation (2), $\tau_{q,\cdot}$ does not have subscript t and thus does not change with t. The only time-varying parameters are the population mean attitude μ_t. We refer to this condition as the constant cut points condition. The constant cut points condition is analogous to the parallel regression assumption in the conventional ordered probit model. This condition denotes that within a single survey series, people assume the same cut points on the underlying continuous variable over time. This condition holds when a single survey item is measured consistently over several years, and people's understanding of this survey item do not change over time. One could also set cut points to vary over time. Prior work has discussed the extension of the parallel regression assumption where cut points are allowed to vary across covariates (Fullerton 2009; Williams 2006). We set cut points to be time constant instead of time varying because of model parsimony considerations. Our goal is to pool multiple survey series, thus the cut points are primarily set to be survey specific to account for any survey-level design effects.[1]

Now, we write the probability of any individual i responding $y_{itq} = k$,

$$P(y_{itq} = k_q) = P(\tau_{q,k-1} - \mu_t \leq \eta_{it} \leq \tau_{q,k} - \mu_t) = \Phi(\tau_{q,k} - \mu_t) - \Phi(\tau_{q,k-1} - \mu_t), \tag{5}$$

where $\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution.

Estimation Procedure

The unknown parameters τ, μ are estimated by maximizing the joint likelihood L expressed as

$$\boldsymbol{L} = \prod_q \prod_{t\in[1,l_q]} \prod_{k\in[1,K_q]} Pr(k)^{N_{qtk}} = \prod_q \prod_{t\in[1,l_q]} \prod_{k\in[1,K_q]} (\Phi(\tau_{q,k} - \mu_t) - \Phi(\tau_{q,k-1} - \mu_t))^{N_{qtk}}, \tag{6}$$

where $N_{qtk} = \sum_i I(y_{itq} = k)$ is the number of individuals responding category k to question q in year t. Note that there exist an infinite number of parallel solutions of μ and τ that equally maximize L. Therefore, we need to normalize the parameters to remove the indeterminacy. Without loss of generality, we set the population mean attitude μ of the first year to 0. This adds a constraint to equation (6), converting the search of τ, μ to the following equivalent optimization problem:

$$max_{\tau,\mu} \prod_q \prod_{t\in[1,l_q]} \prod_{k\in[1,K_q]} (\Phi(\tau_{q,k} - \mu_t) - \Phi(\tau_{q,k-1} - \mu_t))^{N_{qtk}}.$$

$$s.t.\ \mu_0 = 0 \tag{7}$$

To facilitate the optimization process, we feed the model with initial values that draw from a simpler ordered probit model with a homogenous mean (set at 0) across years.

Overlapping Year as the Identification Condition. Our model requires that survey series satisfy the sample representativeness condition; that is, each survey should be representative of the underlying population. For example, if the goal is to pool multiple surveys to study life satisfaction trends of the U.S. population, the surveys used should be representative of the entire U.S. population in the given year.

With the sample representativeness condition, the population-level mean attitude at each time t, μ_t, does not vary across surveys, that is,

$$\mu_{t,q=1} = \mu_{t,q=2} = \cdots = \mu_t \ . \tag{8}$$

This condition allows us to anchor q-specific thresholds τ_q. For this to work, our method further requires that the pooled survey series should overlap with each other. When two survey series (say, $q = 1,2$) overlap in a year (say, year u), the condition $\mu_{t=u,q=1} = \mu_{t=u,q=2}$ means we only need to normalize one mean in one survey series, say, $\mu_{1,1} = 0$, to identify $\tau_q(q = 1,2)$. Formally speaking, we can introduce the following proposition:

> *Proposition 1*: Survey series can be pooled together only when they have at least one overlapping year.

The overlapping year condition draws insights from research practices in survey design and evaluation. When evaluating a particular aspect of survey design, survey researchers often randomly assign respondents to different versions of surveys. The goal is to test various design effects, such as changes in survey mode (Couper 2011) or measurement (Jaeger 1997). The random assignment ensures the distributional properties of the underlying variables should be the same.

In a similar vein, when survey series overlap in a given year, we assume the same latent attitude is measured as if two measurement scales are randomly given to two independent samples drawn from the same population. For example, if two survey series $q = 1$ and $q = 2$ do not overlap in any year, equation (6) can be split into two terms, covering the years of survey $q = 1$ and the years of survey $q = 2$. The two survey series would require two separate normalization conditions so that μ terms are incomparable across the two series. If, however, $q = 1$ and $q = 2$ overlap in a given year, we can reduce the two normalization requirements that are imposed on the two survey series into one, as we can use the information from one survey for normalization in the other.

When two survey series overlap in more than one year, we have an over-identification condition, that is, there is more than enough information in the two series to estimate the unknown parameter τ_q and μ_t. This allows us to test if the mean equality condition in equation (8) holds. For example, if two survey series $q = 1$ and $q = 2$ overlap in two years $t = 1$ and $t = 2$, we can parametrize the two sets of models. The first model assumes the population mean attitudes to be equal in both years in two survey series, that is, $\mu_{t=1,q=1} = \mu_{t=1,q=2} = \mu_1$ and $\mu_{t=2,q=1} = \mu_{t=2,q=2} = \mu_2$. The

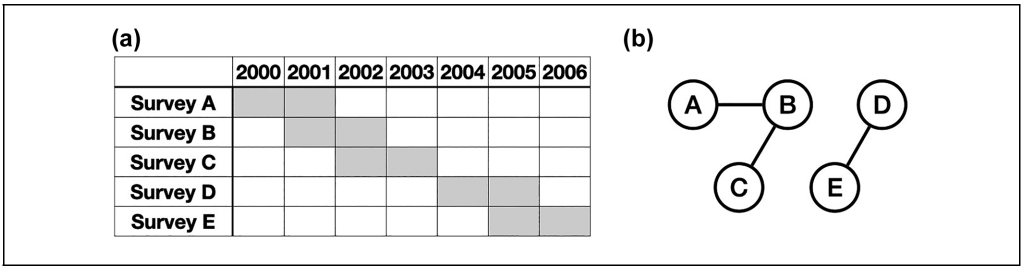

Figure 2. An illustration of the overlapping year condition: (a) five survey series are running from 2001 to 2006 (gray in the left panel) and (b) A and B are directly connected as they overlap in 2001, B and C are directly connected with an overlapping year 2002. A and C are indirectly connected, as there is a path between them. A, B, and C are not connected to survey D or E.

second model assumes population mean attitudes to be different in one of the two years between the two survey series, that is, $\mu_{t=1,q=1} \neq \mu_{t=1,q=2}$ or $\mu_{t=2,q=1} \neq \mu_{t=2,q=2}$. By comparing the performance of the two models via likelihood-ratio statistics, we can test the mean equality hypothesis. When the mean equality hypothesis is supported, we use both overlapping years as a pooled identification condition to improve the efficiency in the estimation of the threshold parameters for the overlapping years.

In the case of pooling more than two survey series, the overlapping year condition requires an undirected network connecting a pair of survey series if they overlap for at least one year. That is, a path exists on this network between them. As illustrated in Figure 2, five survey series are implemented from 2000 to 2006, and each covers a different period. A, B, and C can be pooled together because A and B overlap in 2001, and B and C overlap in 2002. In the same vein, D and E can also be pooled together. However, we cannot pool all five survey series together because there is no connection linking the two clusters {A, B, C} and {D, E}. Our method works for a subset of survey series in which any two survey series are mutually reachable. Similar to the case of two overlapping survey series, for a group of mutually reachable survey series, the normalization conditions can be reduced by borrowing distributional information from one another.

SIMULATION STUDY

So far, we have outlined the LAM. In this section, we conduct a simulation study to evaluate the performance of the LAM. With generated data, we can investigate how successfully the method can recover the true population-level mean attitudes and survey-specific cut points.

To simulate attitudes of the population under study, we generate 10 years of data with 100,000 individual observations per year. Each individual's latent attitude y^*_{it} is drawn from a standard normal distribution with a random mean. Thus, the "true" mean of the population-level attitudes is known. The simulated latent attitude distributions are displayed in Figure 3a.

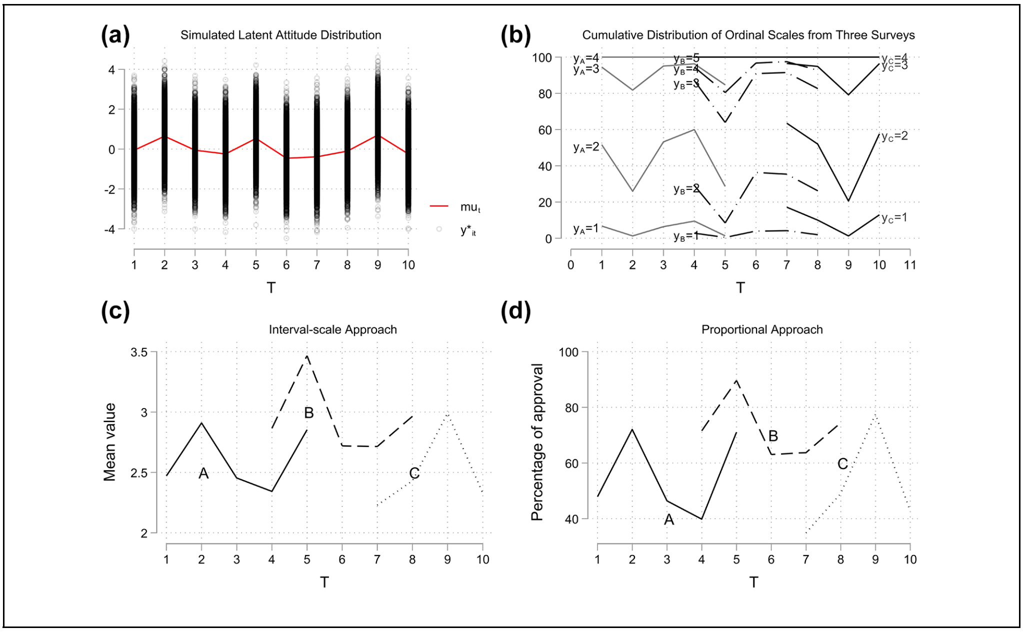

Figure 3. Simulation design: (a) simulated latent attitude distribution, (b) cumulative distribution of ordinal scales from three surveys, (c) interval-scale approach, and (d) proportional approach.

Next, we develop three Likert-type survey series, A, B, and C, with different response scales and time coverage. Each year, a fresh random sample is drawn from the simulated population, satisfying the condition of sample representativeness. Specifically, survey A draws a 20 percent sample from the population each year, and is administered from year 1 to year 5. Survey B draws a 10 percent sample from the population each year, and is administered from year 4 to year 8. Survey C draws a 15 percent sample from the population each year and is administered from year 7 to year 10. As a result, A and B overlap in years 4 and 5, and B and C overlap in years 7 and 8.

The response scales differ across the three survey series. Our goal is to simulate the scenarios described in Figure 1. These scales are set by assigning series-specific cut point values τ, such that the relationships between the latent value y_{it}^* and τ follow equation (2). Survey series A and B are designed to simulate the case of scale number differences, as A has four-level responses and B has a five-level response scale. Survey series A and C are designed to simulate the case of wording differences, as they both have four-level response scales, yet differ in the actual cut point values. The assigned cut point values are presented in columns 1, 2, and 3 in Table 2. With the true cut point values known, we can evaluate the accuracy of the estimated cut point values.

The goal of the LAM approach is to address the problem of measurement inconsistency when pooling data from multiple survey series. To illustrate the problem, we plot the cumulative distribution of the responses across scales (Figure 3b). These three

Table 2. Simulated True Cut Point Values versus Estimated Cut Point Values

	True Value			Estimated Value		
Cut Point	Survey A	Survey B	Survey C	Survey A	Survey B	Survey C
τ_1	−1.577	−2.206	−1.40	−1.529	−2.109	−1.295
τ_2	.020	−.824	−.050	.075	−.691	.023
τ_3	1.604	.867	1.50	1.623	.953	1.561
τ_4	NA	1.414	NA	NA	1.481	NA

Note: NA = not applicable.

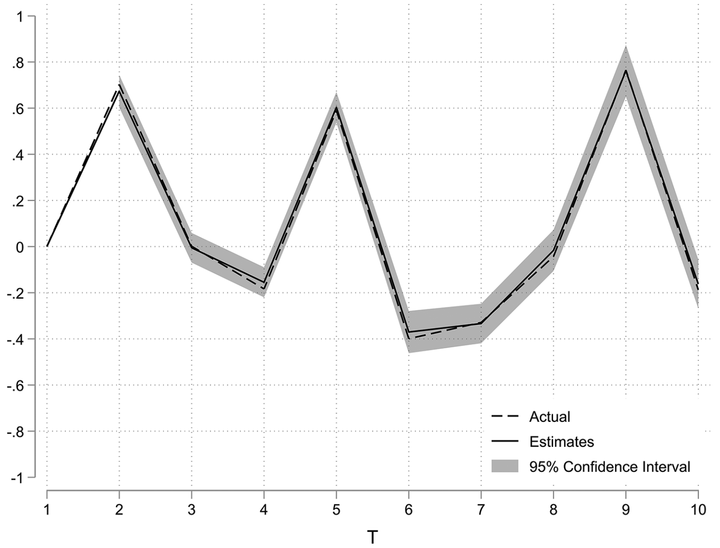

Figure 4. Latent attitude method estimates with simulated data.

survey series measure the same underlying latent attitudes at the overlapping year, but the observed distribution of respondents by each scale differs. As a result, one cannot simply pool these three surveys to infer long-term trends without addressing measurement issues. We further apply the integer-scoring approach and the proportional approach, and the results are presented in Figures 3c and 3d. We see that both methods are inadequate in addressing the problem, as trends appear incomparable across survey series.

Figure 4 presents the estimated mean attitude with the LAM approach. We constructed 95 percent confidence intervals with 500 bootstrapped replicates. The figure shows the LAM estimates successfully recover the simulated true mean values. Table 2 presents the actual and estimated cut points. The estimated cut point values are very close to the actual values, validating the method.

EMPIRICAL EXAMPLE: AMERICANS' ATTITUDES TOWARD CHINA

In this section, we apply the LAM approach to construct a long-term trend in Americans' attitudes toward China. This topic is substantively important today, as Americans' attitudes toward China have become particularly negative in light of the coronavirus disease 2019 pandemic (Cook, Huang, and Xie 2021; He, Zhang, and Xie 2022). The significant decline in Americans' attitudes toward China in recent years has been established (Huang, Cook, and Xie 2021a, 2021b; Silver, Devlin, and Huang 2020; Xie and Jin 2022), but the swings in Americans' attitudes toward China over the long term are not as well determined.

We focus on adult Americans' attitudes toward China from 1974, when the two nations began rapprochement, to 2019, when the most recent data are publicly available. The LAM approach is ideally suited for studying Americans' attitudes toward China over the long term on the basis of survey data: multiple surveys have solicited Americans' attitudes toward China, yet no single one was implemented on a sufficiently long-term basis to provide a comprehensive picture of how Americans viewed China over the past four decades. For example, GSS included one question on attitude toward China beginning in 1974 but dropped this question in 1994. The Pew Research Center's influential survey series on Americans' attitudes toward China was not launched until 2005. Given the complexities of U.S.–China relations from 1974 to 2019, and the associated attitude changes, it would be helpful to pool data from different survey series to construct a single time-series measure of Americans' attitudes toward China over the long term.

Data

Surveys on Americans' attitudes toward China are abundant. To be inclusive, we made a thorough search of such surveys from three large survey archives, the Roper Center for Public Opinion Research, the National Opinion Research Center, and the Inter-University Consortium for Political and Social Research. We include surveys on the basis of the following criteria: (1) The questions' wording should measure general attitudes toward China. We excluded questions asking domain-specific attitudes, such as attitudes toward Chinese people, the Chinese government, or the Chinese economy. (2) The targeted population was the entire U.S. adult population: we thus excluded regional surveys. (3) The same survey question was implemented at least twice over time so we can anchor measurements. We thus excluded surveys that were implemented only once. "Don't know" responses and missing responses are excluded from the final data set.

The final collection includes data from 101 cross-sectional public opinion surveys administered by 10 institutions between 1974 and 2019, shown in Table 3. The surveys came from a wide range of sources, including large national social surveys (e.g., the GSS) and polls conducted by public opinion research companies (e.g., Gallup and Pew) and television networks (e.g., ABC and CBS). We constructed the outcome variables into 13 survey series on the basis of survey organizations, question wordings,

Table 3. Data on Attitudes toward China

Variable Name	Question Wording	Scale	Organization	Available Years
ABC	Would you say you generally have a favorable or unfavorable impression of China?	2	ABC News	2
CBS	Are your feelings toward China generally favorable, generally unfavorable, or neutral?	3	CBS News	5
TRA_4	How about the following countries, please tell me if you have a very favorable, somewhat favorable, somewhat unfavorable or very unfavorable opinion of China.	4	TNS Opinion and Social Institutes	3
PEW	Do you have a very favorable, mostly favorable, mostly unfavorable, or very unfavorable attitude toward China?	4	Pew Research Center	12
ABCWP_4	Would you say you generally have a favorable or unfavorable impression of China? If favorable/ unfavorable, ask: Would you say very or somewhat favorable /unfavorable?	4	ABC/*Washington Post*	2
USORC_4	Is your overall opinion of each of the following countries very favorable, mostly favorable, mostly unfavorable, or very unfavorable? China	4	ORC International	4
PSRA_4	Is your overall opinion of China very favorable, mostly favorable, mostly unfavorable, or very unfavorable?	4	Princeton Survey Research Associates	3
ZOGBY_4	How would you describe your impression of China? Is it very favorable, somewhat favorable, somewhat unfavorable, or very unfavorable?	4	Zogby International	2
GALLUP_4	Is your opinion of China very favorable, mostly favorable, mostly unfavorable or very unfavorable?	4	Gallup	30
TRA_5a[a]	I'd like you to rate your feelings toward some countries, institutions and people, with 100 meaning a very warm, favorable feeling and 0 meaning a very cold, unfavorable feeling, and 50 meaning not particularly cold or warm: China	5	TNS Opinion and Social Institutes	4
USKN_5[a]	Please rate your feelings towards some countries, with 100 meaning a very favorable feeling, 0 meaning a very unfavorable feeling, and 50 meaning a feeling that is neither favorable nor unfavorable. You can use any number between 0 and 100, the higher the number, the more favorable your feelings toward this country: China	5	GfK Knowledge Networks	7
GALLUP_10	You notice that the ten boxes on this card go from the highest position of plus five—for something you have a very favorable opinion of—all the way down to the lower position of minus five—for something you have a very unfavorable opinion of. Please tell me how far up the scale or how far down the scale you rate this nation: China.	10	Gallup	7
GSS	You will notice that the boxes on this card go from the highest position of "plus 5" for a country which you like very much to the lowest position of "minus 5" for a country you dislike very very much. How far up the scale or how far down the scale would you rate the following countries? F. China	10	National Opinion Research Center	12

[a] The 0-to-100 warmth scale was reported as a 1-to-5 scale in the Roper Center for Public Opinion Research archive.

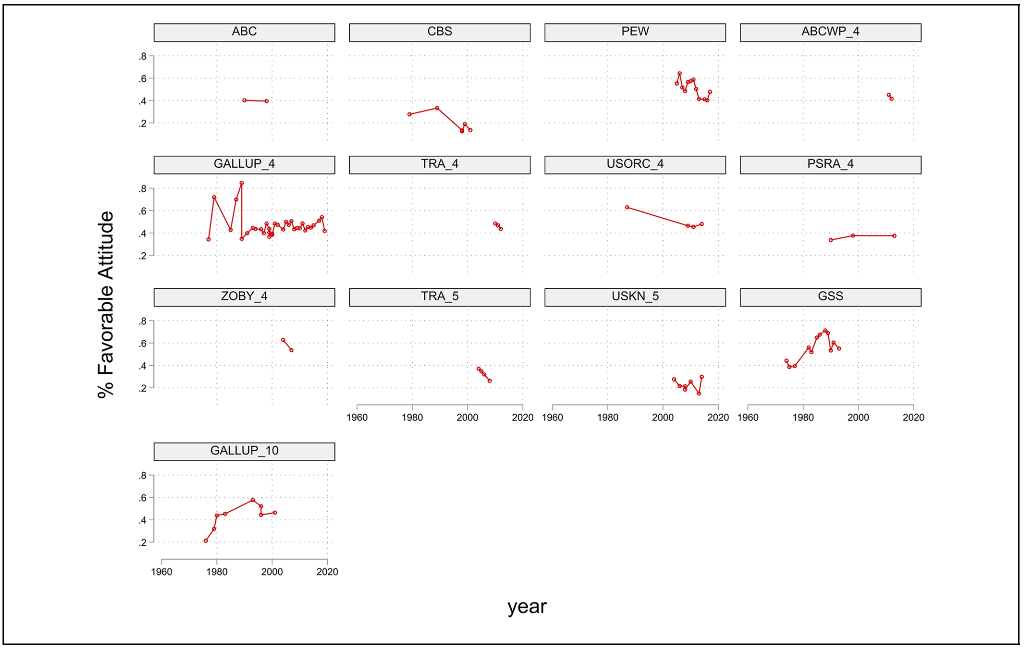

Figure 5. Percentage of respondents with a favorable attitude toward China.

and response categories. Data from different years are considered in the same survey series if survey questions were administered by the same organization with the same question wording and response categories. If surveys were administered by the same survey organization but with different response categories, data are partitioned into different series. For example, Gallup asked two versions of questions about Americans' attitudes toward China. The first asked respondents' level of favorability or unfavorability toward China with a 4-response scale. The second asked to what extent respondents liked China with a 10-response scale. Data from the two questions are treated as different survey-series variables, denoted as GALLUP_4 and GALLUP_10 in Table 3. The final data set includes information on survey institutions, question wording, survey date, response scale, percentage of respondents corresponding with each scale, and the total sample size.

Table 3 presents a detailed description of the data. As shown in the table, the wordings of these questions are very similar, and this consistency assures us they all measure the same latent attitude toward China. However, the response scales vary from a simple binary scale indicating a favorable or unfavorable attitude, to a 0-to-100 warmth scale.[2] The most common is a four-response scale, where respondents rate their attitude toward China as very favorable, somewhat favorable, somewhat unfavorable, or very unfavorable. The longest running question series was administered by Gallup, which asked respondents' attitudes toward China with a four-response scale (GALLUP_4).

Figure 5 provides a first glance at the time-series variables. To better detect the trend, we collapse the response scales and only show the percentage of positive attitudes in each variable. Overall, these polls consistently show the Tiananmen Square protests in 1989 marked a turning point in Americans' attitudes toward China. Before 1990, the proportion of Americans who held favorable views of China was well above 40 percent in most polls. The drastic decline in favorability after 1989 was captured by most polls, including those administered by ABC, CBS, Gallup, and the GSS. However, the trends are less clear thereafter. Some polls, such as Gallup, show Americans' persistent low favorability toward China; others, such as the GSS, show an uptick in favorability in the 2000s. The levels of favorability also differ across polls. For example, CBS and Growth from Knowledge Networks (USKN_5) show persistent lower percentages of favorable attitudes toward China compared with other polls. The inconsistencies across the survey series suggest scale incomparability is a crucial issue and warrants further study.

LAM Estimates of Americans' Attitudes toward China

Much of the prior research on Americans' attitudes toward China draws heavily from the idea of cognitive frameworks (Baum and Potter 2008; Converse 1964; Hurwitz 1987). Cognitive frameworks are the mental shortcuts with which people interpret the information available to them. Fundamentally, human brains have limited capacity and tend to take cognitive shortcuts when evaluating complex issues (Fiske and Taylor 2017). For ordinary people, information regarding foreign affairs is complex and distant from their everyday lives, so they rely on such frameworks to evaluate foreign affairs information. The classic example is the cold war mental framework, wherein people evaluated countries on the basis of their relationships with either the Soviet-led communist bloc or the U.S.-led Western bloc.

Americans' attitudes toward China should also be interpreted in light of different cognitive frameworks the public has adopted in different historical periods. Huang et al. (2021b) posited that Americans' attitudes toward China since the 1970s have been influenced by how the news media has reported major historical events in U.S.–China relations. These historical events set Americans' cognitive frameworks in their evaluation of China.

The LAM facilitates analysis of the long-term changes in Americans' attitudes as the public has adopted different cognitive frameworks. Figure 6 presents the LAM estimates of Americans' attitudes toward China and bootstrapped 95 percent confidence intervals. Compared with individual polls in Figure 5, the LAM estimates provide a clearer picture of the change in attitudes over a much longer period. For example, Americans' attitudes toward China became more positive throughout the 1970s and 1980s. Probably this is because prior to the 1970s, Americans viewed China as a communist, totalitarian regime. Yet beginning in the 1970s, united by a common adversary in the Soviet Union, the two nations moved toward rapprochement and eventually formed a diplomatic relationship in 1979.

Americans' attitudes toward China peaked by the end of the 1980s. Figure 6 shows that Americans' perceptions of China became precipitously more negative immediately after the Tiananmen Square protests in 1989. For many Americans, the protests cued the "communist regime" cognitive framework, resulting in a significant negative

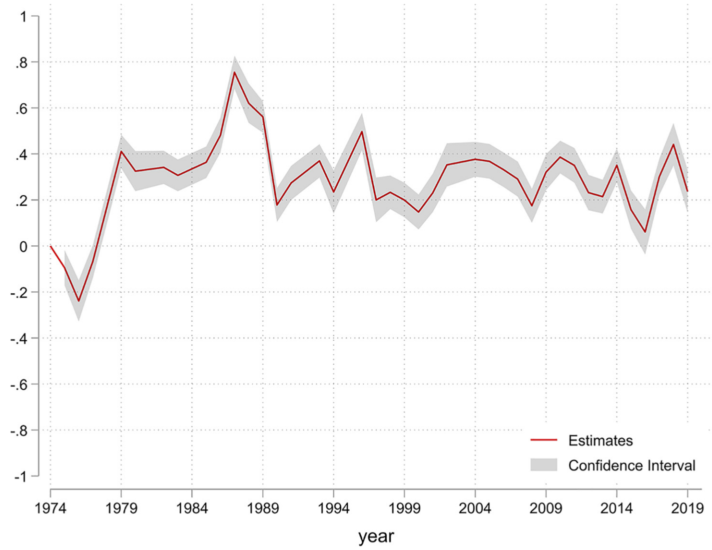

Figure 6. Latent attitude method estimates of Americans' attitudes toward China.

effect on their attitudes toward China. The post-1989 period saw ebbs and flows in attitudes. Although there were some upticks, Americans' attitudes toward China never reached the levels of the pre-1989 period. Scholars have argued that Americans' attitudes toward China were ambivalent in the two decades after the Tiananmen Square protests (Tien and Nathan 2001). The fluctuations in attitudes during this time suggest the U.S. public had not yet formed a conceptual anchoring point with regard to China. Americans were attracted by China's large domestic market but were also threatened by its rising economic power. In recent years, there appears to have been yet another shift in attitudes. During the Trump administration, the U.S.–China relationship experienced rapid deterioration, and this has been reflected in declining attitudes toward China since 2018 (Xie and Jin 2022).

DISCUSSION AND CONCLUSION

The study of attitudinal trends with pooled data is important because of pervasive attitude measures in survey research, and because of the difficulties in measuring attitude as a latent construct. In this study, we developed a new LAM to address the issue of scale incomparability when pooling multiple survey series for trend analysis. We started our analysis with two measurement principles, measurement consistency and sample representativeness. Building on these two principles, we emphasized that attitude surveys usually do not have intrinsic numerical scales and thus cannot be directly subjected to arithmetical operations. We further showed that although the integer-scoring and proportional approaches are widely used in the context of single survey series, not only do these two approaches dismiss the ordinal nature of attitudinal surveys, but they also are incapable of addressing the issue of scale incomparability when pooling multiple survey series.

The challenge of scale incomparability motivated us to propose a latent variable approach to the study of pooling multiple attitudinal survey series. We modeled the process of population attitudinal change by extending the classical latent variable model. A defining feature in our model specification emphasizes that the cut point parameters should be survey specific and time constant. Although it is possible to allow cut points to vary over time in limited forms, this should only be considered with strong theoretical justifications. In general, we recommend a fixed cut point approach unless there are compelling reasons to do otherwise. We rely on overlapping year(s) as the identification condition so the survey series can be conveniently anchored without imposing any additional normalization requirements.

The subsequent simulation analysis demonstrates that the LAM approach successfully recovers the true mean population-level attitudes and survey-specific cut points. We applied the proposed approach to the study of the long-term trend in Americans' attitudes toward China. The empirical results paint a clearer picture of the ebbs and flows of Americans' attitudes toward China over the past four decades. Our method can be fruitfully applied to other research contexts in which multiple surveys measure the same latent constructs, and when scholars are interested in long-term changes in subjective well-being, such as happiness or self-rated health.

Besides trend analysis, the method can also be used when different survey instruments are used in different geographic units. Recall that the key to identifying an unknown cut point and the mean in trend analysis is the overlapping year condition. We suggest that in the case of geographic variation, the same rationale can be applied if two or more surveys overlap in the same geographic unit. In such a case, we can consider the surveys as measuring the same latent attitudes and anchor the survey on the basis of the overlapping geographic location. When applied to the context of estimating geographic variations of attitudes, we stress that our model is different from the existing small-area estimation method in pooling multiple surveys, such as multilevel regression and poststratification. To be precise, the data structure required by the two approaches is different. Multilevel regression and poststratification uses national-level data with subnational geographic unit identifiers, thus the data structure is essentially hierarchical. The first step of the approach, the multilevel regression with random effect, borrows information from the neighboring units. In contrast, our approach does not require a hierarchical data structure.

Finally, one limitation of the LAM is that we assume the survey series under study are representative. The results will not be biased even if survey series adopt different sampling strategies as long as they are representative of the population under study. However, this assumption may not hold in some cases. If sampling bias varies with time within a survey series, it would introduce measurement inconsistency and make data from a survey series incomparable for trend analysis. If sampling bias is present but constant within a survey series, trend analysis is less threatened, but it compromises the pooling of data across survey series through use of the LAM. For example, certain demographic groups or geographic locations may be consistently over- or undersampled by some surveys across all waves. One potential solution is to attain additional external information. For instance, if the distributions of respondents' demographics are known, researchers can address potential survey-series-specific errors by comparing respondents' demographics and reweighting the series to make it nationally representative.

APPENDIX

ALTERNATIVE MODEL SPECIFICATIONS

Table A1 shows model fit statistics that compare the current model against three alternatives that allow cut points to vary by time in different forms. Model 1 is the current model, which assumes cut points are time constant for all surveys. Model 2 assumes cut points differ before and after year 1989 for all surveys. Model 3 allows the second cut point (τ_2) of survey Gallup_4 to vary completely over time, while holding other cut points time-constant. Model 4 assumes all cut points follow a linear function of time.

Table A1. Test Statistics for Model Comparison.

Model	Description	Log Likelihood	d.f.	AIC	BIC
1	Constant cut points	−155,574.5	89	311,327	312,180.6
2	Different cut points before and after 1989	−155,343.7	112	310,911.4	311,985.5
3	Varying $\tau_{Gallup\_4,2}$	−155,407.9	118	311,051.8	312,183.5
4	Linear trend in all cut points	−155,227.7	138	310,731.3	312,054.8

Note: AIC = Akaike information criterion; BIC = Bayesian information criterion.

Acknowledgments

The authors thank Yongai Jin, Bin Tian, Lemeng Liang, and Luyin Zhang for their assistance in compiling data.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Donghui Wang received support from the National Social Science Foundation of China (21CRK001).

ORCID iDs

Donghui Wang https://orcid.org/0000-0002-5403-7947
Yu Xie https://orcid.org/0000-0002-3240-8620

Supplemental Material

Supplemental material for this article is available online.

Notes

1. One could also set cut points to vary over time in limited forms (e.g., over broad periods or linearly). To illustrate, we conducted additional analyses that compare the current model with three alternative models that allow the cut points to vary by time in different forms in the subsequent empirical example (see Appendix A). Because each survey allows distinct cut points, varying cut points by time would lead to a substantial increase in the required degrees of freedom. The results in Table A1 show a very marginal increase in log likelihood of models 2, 3, and 4 over model 1. That is, allowing cut points to vary with time does not significantly improve goodness-of-fit, considering the cost of increased degree of freedom.
2. The 0-to-100 warmth scale was reported as a 1-to-5 scale in the Roper Center for Public Opinion Research archive.

References

Achen, Christopher H. 1975. “Mass Political Attitudes and the Survey Response.” *American Political Science Review* 69(4):1218–31.

Agresti, Alan. 2002. *Categorical Data Analysis*. 2nd ed. Hoboken, NJ: John Wiley.

Alwin, Duane, and Jon A. Krosnick. 1991. “The Reliability of Survey Attitude Measurement.” *Sociological Methods and Research* 20(1):139–81.

Baum, Matthew A., and Philip B. K. Potter. 2008. “The Relationships between Mass Media, Public Opinion, and Foreign Policy: Toward a Theoretical Synthesis.” *Annual Review of Political Science* 11:39–65.

Bürkner, Paul Christian, and Matti Vuorre. 2019. “Ordinal Regression Models in Psychology: A Tutorial.” *Advances in Methods and Practices in Psychological Science* 2(1):77–101.

Carter, J. Scott, Shannon K. Carter, and Jamie Dodge. 2009. “Trends in Abortion Attitudes by Race and Gender: A Reassessment over a Four Decade Period.” *Journal of Sociological Research* 1(1):1–17.

Clogg, Clifford C. 1979. “Some Latent Structure Models for the Analysis of Likert-Type Data.” *Social Science Research* 8(4):287–301.

Converse, Philip E. 1964. “The Nature of Belief Systems in Mass Publics.” *Critical Review* 18(1–3):1–74.

Cook, Gavin, Junming Huang, and Yu Xie. 2021. “How COVID-19 Has Impacted American Attitudes toward China: A Study on Twitter.” *arXiv*. Retrieved August 8, 2023. https://doi.org/10.48550/arXiv.2108.11040.

Cotter, David, Joan M. Hermsen, and Reeve Vanneman. 2011. “The End of the Gender Revolution? Gender Role Attitudes from 1977 to 2008.” *American Journal of Sociology* 117(1):259–89.

Couper, Mick P. 2011. “The Future of Modes of Data Collection.” *Public Opinion Quarterly* 75(5): 889–908.

DiMaggio, Paul, John Evans, and Bethany Bryson. 1996. “Have Americans’ Social Attitudes Become More Polarized?” *American Journal of Sociology* 102(3):690–755.

Duxbury, Scott W. 2021. “Fear or Loathing in the United States? Public Opinion and the Rise of Racial Disparity in Mass Incarceration, 1978–2015.” *Social Forces* 100(2):427–53.

Fienberg, Stephen E. 1980. *The Analysis of Cross-Classified Categorical Data*. Cambridge: Massachusetts Institute of Technology Press.

Fiske, Susan T., and Shelley E. Taylor. 2017. *Social Cognition: From Brains to Culture*. 3rd ed. Thousand Oaks, CA: Sage.

Fricker, Scott, Mirta Galesic, Roger Tourangeau, and Ting Yan. 2005. “An Experimental Comparison of Web and Telephone Surveys.” *Public Opinion Quarterly* 69(3):370–92.

Fullerton, Andrew S. 2009. “A Conceptual Framework for Ordered Logistic Regression Models.” *Sociological Methods and Research* 38(2):306–47.

Goodman, Leo A. 1983. “The Analysis of Dependence in Cross-Classifications Having Ordered Categories, Using Log-Linear Models for Frequencies and Log-Linear Models for Odds.” *Biometrics* 39:149–60.

He, Qian, Ziye Zhang, and Yu Xie. 2022. "The Impact of COVID-19 on Americans' Attitudes toward China: Does Local Incidence Rate Matter?" *Social Psychology Quarterly* 85(1):84–107.

Huang, Junming, Gavin G. Cook, and Yu Xie. 2021a. "Between Reality and Perception: The Mediating Effects of Mass Media on Public Opinion toward China." *Chinese Sociological Review* 53(5):431–50.

Huang, Junming, Gavin G. Cook, and Yu Xie. 2021b. "Large-Scale Quantitative Evidence of Media Impact on Public Opinion toward China." *Humanities and Social Sciences Communications* 8(1):1–8.

Hurwitz, Jon. 1987. "How Are Foreign Policy Attitudes Structured? A Hierarchical Model." *American Political Science Review* 81(4):1099–120.

Inglehart, Ronald. 1985. "Aggregate Stability and Individual-Level Flux in Mass Belief Systems: The Level of Analysis Paradox." *American Political Science Review* 79(1):97–116.

Jaeger, David A. 1997. "Reconciling the Old and New Census Bureau Education Questions: Recommendations for Researchers." *Journal of Business and Economic Statistics* 15(3):300–309.

Jamieson, Susan. 2004. "Likert Scales: How to (Ab)Use Them?" *Medical Education* 38(12):1217–18.

Judd, Charles M., and Michael A. Milburn. 1980. "The Structure of Attitude Systems in the General Public: Comparisons of a Structural Equation Model." *American Sociological Review* 45(4):627–43.

Kalmijn, W. M., L. R. Arends, and R. Veenhoven. 2011. "Happiness Scale Interval Study: Methodological Considerations." *Social Indicators Research* 102(3):497–515.

Kiley, Kevin, and Stephen Vaisey. 2020. "Measuring Stability and Change in Personal Culture Using Panel Data." *American Sociological Review* 85(3):477–506.

McGann, Anthony J. 2014. "Estimating the Political Center from Aggregate Data: An Item Response Theory Alternative to the Stimson Dyad Ratios Algorithm." *Political Analysis* 22(1):115–29.

McKelvey, Richard D., and William Zavoina. 1975. "A Statistical Model for the Analysis of Ordinal Level Dependent Variables." *Journal of Mathematical Sociology* 4(1):103–20.

Mouw, Ted, and Michael E. Sobel. 2001. "Culture Wars and Opinion Polarization: The Case of Abortion." *American Journal of Sociology* 10(4):913–43.

National Opinion Research Center. 2020. "About the GSS." Retrieved August 8, 2023. https://gss.norc.org/About-The-GSS.

Osberg, Lars, and Timothy Smeeding. 2006. "'Fair' Inequality? Attitudes toward Pay Differentials: The United States in Comparative Perspective." *American Sociological Review* 71(6):450–73.

Oskamp, Stuart, and P. Wesley Schultz. 2004. *Attitudes and Opinions*. 3rd ed. New York: Routledge.

Powers, Daniel, and Yu Xie. 2008. *Statistical Methods for Categorical Data Analysis*. Bingley, UK: Emerald Group.

Schwarz, Norbert. 1999. "Self-Reports: How the Questions Shape the Answers." *American Psychologist* 54(2):93–105.

Schwarz, Norbert, Bärbel Knäuper, Hans-J. Hippler, Elisabeth Noelle-Neumann, and Leslie Clark. 1991. "Rating Scales: Numeric Values May Change the Meaning of Scale Labels." *Public Opinion Quarterly* 55(4):570–82.

Scott, Jacqueline, Duane F. Alwin, and Michael Braun. 1996. "Generational Changes in Gender-Role Attitudes: Britain in Cross-National Perspective." *Sociology* 30(3):471–92.

Semyonov, Moshe, Rebeca Raijman, and Anastasia Gorodzeisky. 2011. "The Rise of Anti-foreigner Sentiment in European Societies, 1988–2000." *American Sociological Review* 71(3):426–49.

Silver, Laura, Kat Devlin, and Christine Huang. 2020. "Unfavorable Views of China Reach Historic Highs in Many Countries." *Pew Research Center*. Retrieved August 8, 2023. https://www.pewresearch.org/global/2020/10/06/unfavorable-views-of-china-reach-historic-highs-in-many-countries/.

Smith, Tom W. 2005. "The Laws of Studying Societal Change." Report No. 50. *National Opinion Research Center*. Retrieved August 8, 2023. https://gss.norc.org/Documents/reports/social-change-reports/SC50.pdf.

Sobel, Michael E. 1997. "Modeling Symmetry, Asymmetry, and Change in Ordered Scales with Midpoints Using Adjacent Category Logit Models for Discrete Data." *Sociological Methods and Research* 26(2):213–32.

Steeh, Charlotte, and Howard Schuman. 1992. "Young White Adults: Did Racial Attitudes Change in the 1980s?" *American Journal of Sociology* 98(2):340–67.

Stimson, James A. 2018. "The Dyad Ratios Algorithm for Estimating Latent Public Opinion: Estimation, Testing, and Comparison to Other Approaches." *Bulletin of Sociological Methodology/Bulletin de Methodologie Sociologique* 137–38(1):201–18.

Thornton, Arland. 1985. "Changing Attitudes toward Separation and Divorce: Causes and Consequences." *American Journal of Sociology* 90(4):856–72.

Thornton, Arland, and Linda Young-DeMarco. 2001. "Four Decades of Trends in Attitudes toward Family Issues in the United States: The 1960s through the 1990s." *Journal of Marriage and Family* 63(4): 1009–37.

Tien, Charles, and James A. Nathan. 2001. "The Polls—Trends: American Ambivalence toward China." *Public Opinion Quarterly* 65(1):124–38.

Williams, Richard. 2006. "Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Dependent Variables." *Stata Journal* 6(1):58–82.

Wu, Huiping, and Shing On Leung. 2017. "Can Likert Scales Be Treated as Interval Scales? A Simulation Study." *Journal of Social Service Research* 43(4):527–32.

Xie, Yu, and Yongai Jin. 2022. "Global Attitudes toward China: Trends and Correlates." *Journal of Contemporary China* 31(133):1–16.

Author Biographies

Donghui Wang is an associate professor at the Center for Population and Development Studies, School of Sociology and Population Studies, at Renmin University of China. Her areas of research are demography, stratification, and social attitudes.

Yu Xie is the Bert G. Kerstetter '66 University Professor of Sociology and the Princeton Institute for International and Regional Studies. He is also the chair professor at Center for Social Research at Peking University. His main areas of interest are social stratification, demography, statistical methods, Chinese studies, and sociology of science.

Junming Huang is an associate research scientist at the Paul and Marcia Wythes Center on Contemporary China at Princeton University. His research interests lie in a range of computational social science topics including social media and networks analysis, science of science, and natural language processing.